

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
Programa de Pós-Graduação em Estudos Linguísticos

Fernanda Carla de Oliveira

**NOMES GERAIS E OUTRAS FORMAS DE INDETERMINAÇÃO DO SUJEITO: um
estudo variacionista sobre o português brasileiro (PB) de Bambuí e sobre o português
europeu (PE) de Coimbra**

Belo Horizonte

2022

Fernanda Carla de Oliveira

NOMES GERAIS E OUTRAS FORMAS DE INDETERMINAÇÃO DO SUJEITO: um estudo variacionista sobre o português brasileiro (PB) de Bambuí e sobre o português europeu (PE) de Coimbra

Tese de doutorado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do título de Doutora em Linguística.

Orientador: Prof. Dr. Eduardo Tadeu Roque Amaral

Área de concentração: Linguística Teórica e Descritiva

Linha de pesquisa: Estudo da Variação e Mudança Linguística

Belo Horizonte

2022

O48n

Oliveira, Fernanda Carla de.

Nomes gerais e outras formas de indeterminação do sujeito [manuscrito] : um estudo variacionista sobre o português brasileiro (PB) de Bambuí e sobre o português europeu (PE) de Coimbra / Fernanda Carla de Oliveira. – 2022.

127 f., enc. : il., maps., grafs., (color), tabs., (p&b)

Orientador: Eduardo Tadeu Roque Amaral.

Área de concentração: Linguística Teórica e Descritiva.

Linha de Pesquisa: Estudo da Variação e Mudança Linguística.

Tese (doutorado) – Universidade Federal de Minas Gerais, Faculdade de Letras.

Bibliografia: f. 113-119.

Apêndices: f. 120-127.

1. Língua portuguesa – Variação – Bambuí (MG) – Teses. 2. Língua portuguesa – Variação – Coimbra (Portugal) – Teses. 3. Mudanças linguísticas – Teses. 4. Língua portuguesa – Etimologia – Nomes – Teses. 5. Língua portuguesa – Lexicologia – Teses. I. Amaral, Eduardo Tadeu Roque. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 469.798



UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

FOLHA DE APROVAÇÃO

NOMES GERAIS E OUTRAS FORMAS DE INDETERMINAÇÃO DO SUJEITO: um estudo variacionista sobre o português brasileiro (PB) de Bambuí e sobre o português europeu (PE) de Coimbra

FERNANDA CARLA DE OLIVEIRA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGUÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGUÍSTICOS, área de concentração LINGUÍSTICA TEÓRICA E DESCRITIVA, linha de pesquisa Estudo da Variação e Mudança Linguística.

Aprovada em 20 de julho de 2022, pela banca constituída pelos membros:

Prof(a). Eduardo Tadeu Roque Amaral - Orientador

UFMG

Prof(a). Alexia Teles Duchowny

UFMG

Prof(a). Janayna Maria da Rocha Carvalho

UFMG

Prof(a). Clézio Roberto Gonçalves

UFOP

Prof(a). Maria Filomena Candeias Gonçalves

Universidade de Évora

Belo Horizonte, 20 de julho de 2022.



Documento assinado eletronicamente por Eduardo Tadeu Roque Amaral, Professor do Magistério Superior, em 20/07/2022, às 13:27, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Maria Filomena Gonçalves, Usuária Externa**, em 20/07/2022, às 14:33, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Janayna Maria da Rocha Carvalho, Professora do Magistério Superior**, em 20/07/2022, às 20:28, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Alexia Teles Duchowny, Professora do Magistério Superior**, em 21/07/2022, às 09:57, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Clézio Roberto Gonçalves, Usuário Externo**, em 22/07/2022, às 09:49, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador 1536538 e o código CRC 999527A3.

Dedico este trabalho a todos os informantes deste estudo, pois, sem a colaboração e disponibilidade de cada um deles, esta pesquisa não teria acontecido.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, pelo dom da vida e por ter me proporcionado a capacidade de realizar esta pesquisa.

À minha mãe Renilda, pelo amor, pela dedicação e principalmente pelo apoio incondicional ao longo da minha história.

Ao meu irmão Júnior, por sempre estar ao meu lado.

Ao meu orientador, professor Dr. Eduardo, que me acompanha desde o início da minha caminhada na pesquisa sociolinguística. Sua paciência e dedicação fizeram toda a diferença.

A minhas tias, primos, sobrinhos e amigos que sempre estão comigo e me dão força para continuar em minhas empreitadas.

Ao meu namorado Luiz Heleno, que esteve comigo nessa reta final e que sempre foi meu grande apoiador para a conclusão deste estudo.

Aos meus professores, que me deram a formação necessária para realizar este projeto de maneira satisfatória.

À FAPEMIG, pelo apoio financeiro.

À Faculdade de Letras e ao Programa de Pós-Graduação em Estudos Linguísticos, pelo ensino de qualidade oferecido.

Finalmente, agradeço a todos aqueles que contribuíram de alguma forma para a realização deste sonho. Deixo registrado aqui meu eterno agradecimento, pois sem vocês esta etapa teria sido muito mais árdua.

“Escrever é fácil. Você começa com uma letra maiúscula e termina com um ponto final. No meio, você coloca ideias.” Pablo Neruda

RESUMO

A indeterminação do sujeito, em sua vertente variacionista, já foi estudada por autores como Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010), Souza (2014), Barbosa (2016) e outros. Todos eles demonstraram que as formas previstas pelas gramáticas tradicionais não são as únicas que servem como recurso de indeterminação no português. Assim, por meio deste estudo, busca-se, a partir de uma comparação entre o português brasileiro (PB) e o português europeu (PE), observar a relevância dos nomes gerais no conjunto das variantes indeterminadoras, verificando a influência de fatores linguísticos e extralinguísticos na realização do fenômeno. Salienta-se que os nomes gerais são definidos por Halliday e Hasan (1995 [1976]) como um caso limítrofe entre um item lexical e um item gramatical, por conterem traços semânticos mínimos e foram selecionados para este trabalho porque pesquisas, como as de Oliveira (2018) e de Amaral e Oliveira (2019), mostram que esses itens são capazes de indeterminar. Os dados analisados pertencem a 48 gravações sociolinguísticas coletadas no Brasil, na cidade de Bambuí, localizada no centro-oeste do estado de Minas Gerais, e na cidade de Coimbra, localizada na região central de Portugal, na província Beira Litoral. Os nomes gerais encontrados no *corpus* são *pessoa, pessoal, pessoas, cara, homem, povo, fulano, gente e ser humano*. As demais formas estudadas referem-se àquelas que já foram discutidas em trabalhos anteriores, isto é, *a gente, você(s)/ocê/cê, nós/nóis/-mos, eles/ês, verbo na terceira pessoa do singular sem sujeito explícito (\emptyset +V3PS), verbo na terceira pessoa do plural sem sujeito explícito (\emptyset +V3PP), verbo no infinitivo sem sujeito explícito (\emptyset +VINF), se e tu*. Os fatores extralinguísticos analisados são sexo e faixa etária e os fatores linguísticos testados são presença ou ausência de locativo, presença ou ausência de anáfora, grau de indeterminação e mais ou menos coletivo. O programa utilizado para a análise de dados é o GoldVarb X. Constata-se que o PB e o PE demonstram panoramas diferentes quanto aos recursos de indeterminação utilizados pelos falantes, pois, enquanto os informantes de Bambuí escolhem formas como *a gente* e *você/ocê/cê*, preferencialmente, para indeterminar o sujeito, os informantes de Coimbra utilizam, preferencialmente, a forma *nós* e as formas de indeterminação previstas pelas gramáticas, isto é, *verbo na terceira pessoa do plural sem sujeito explícito (\emptyset +V3PP)* e *se*. Quanto aos nomes gerais, eles são usados com frequência semelhante pelos informantes brasileiros e portugueses. Em Bambuí, os fatores sexo, presença ou ausência de anáfora e grau de indeterminação favorecem a indeterminação por meio dos nomes gerais. Em Coimbra, foram selecionados como estatisticamente relevantes a presença ou ausência de locativo, a faixa etária, o grau de indeterminação e mais ou menos coletivo. Conclui-se que os informantes do PB e do PE utilizam os nomes gerais, entretanto a distribuição dessas formas dentro do conjunto desses nomes se dá diferentemente nos dois. Ademais, os fatores linguísticos e extralinguísticos que influenciam tal variante também se mostram distintos nas duas localidades. Observa-se também que a indeterminação do sujeito, analisada por meio de graus (total ou parcial), é capaz de sanar a confusão entre os termos indefinição e indeterminação. Em síntese, esta pesquisa demonstra como os nomes gerais são relevantes no conjunto dessas variantes e como a comparação entre duas cidades em países distintos traz resultados significativos para pesquisas na área.

Palavras-chave: Sociolinguística Variacionista. Indeterminação do sujeito. Português de Bambuí e português de Coimbra. Nomes gerais.

ABSTRACT

Subject indetermination, in a variationist framework, has already been studied by authors such as Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010), Souza (2014), Barbosa (2016) among others. All of them demonstrated that the forms provided by traditional grammars are not the only ones that serve as a source of indetermination in Portuguese. Thus, this study seeks to draw a comparison between Brazilian Portuguese (BP) and European Portuguese (EP) to observe the relevance of general nouns in the set of indeterminate variants, verifying the influence of linguistic and extralinguistic factors in the realization of the phenomenon. It should be noted that general nouns are defined by Halliday and Hasan (1995 [1976]) as a borderline case between a lexical item and a grammatical item, as they contain minimal semantic features, and they were selected for this study because some studies, such as those by Oliveira (2018) and Amaral and Oliveira (2019), show that these items are able to indeterminate. The analyzed data belong to 48 sociolinguistic recordings collected in Brazil, in Bambuí city, located in the center-west of the state of Minas Gerais, and in Coimbra city, located in central region of Portugal, in the Beira Litoral province. The general nouns found in the corpus are *pessoa, pessoal, pessoas, cara, homem, povo, fulano, gente* and *ser humano*. The other forms studied refer to those that have already been discussed in previous researches, that is, *a gente, você/ocê/cê, nós/nóis, eles/ês, a verb in third-person singular form without an explicit subject (\emptyset +V3PS), ês, a verb in third-person plural form without an explicit subject (\emptyset +V3PP), a verb in infinitive form without an explicit subject (\emptyset +VINFIN), and *se*. The extralinguistic factors analyzed are sex and age group, and the linguistic factors tested are presence or absence of locative, presence or absence of anaphora, degree of indetermination, and more or less collective. The program used for data analysis is GoldVarb X. It seems that BP and EP demonstrate different panoramas regarding the indetermination resources used by the speakers, because, while the informants from Bambuí choose ways such as *a gente* and *você/ocê /cê*, preferably, to indeterminate the subject, the informants from Coimbra use, preferably, the form *nós* and the indetermination forms foreseen by the grammars, that is, *a verb in the third-person plural without an explicit subject (\emptyset +V3PP)* and *se*. As far as general nouns are concerned, they are used with similar frequency by Brazilian and Portuguese informants. In Bambuí, the factors sex, presence or absence of anaphora, and degree of indetermination favor indetermination through general nouns. In Coimbra, the presence or absence of locative, age group, degree of indetermination, and more or less collective were selected as statistically relevant. It is concluded that informants from BP and EP use general nouns; however, the distribution of these forms within the set of these nouns occurs differently in the two. Furthermore, the linguistic and extralinguistic factors which influence this variant are also different in the two locations. It is also observed that the subject's indeterminacy, analyzed through degrees (total or partial), can solve the confusion between the indefiniteness and indetermination of the term. In summary, this research demonstrates how the general nouns are relevant in the set of these variants and how the comparison between two cities in different countries brings significant results for research in the area.*

Keywords: Variationist Sociolinguistics. Subject Indetermination. Bambuí Portuguese and Coimbra Portuguese. General nouns.

LISTA DE QUADROS

QUADRO 1 – Diferenças entre PB e PE segundo os níveis fonético, morfológico, sintático e lexical	25
QUADRO 2 – Resumo dos resultados obtidos em Oliveira (2018)	45
QUADRO 3 – Distribuição de informantes por variáveis extralinguísticas.....	56
QUADRO 4 – Distribuição por tempo e número de palavras das gravações obtidas	57
QUADRO 5 – Normas adotadas para transcrição das entrevistas	59

LISTA DE FIGURAS

FIGURA 1 – Localização de Bambuí no estado de Minas Gerais	52
FIGURA 2 – Localização de Coimbra em Portugal.....	53

LISTA DE TABELAS

TABELA 1 – Número de ocorrência e porcentagem de cada uma das variantes em Bambuí	73
TABELA 2 – Número de ocorrência e porcentagem de cada uma das variantes em Coimbra	74
TABELA 3 – Ocorrências dos nomes gerais no corpus de Bambuí.....	76
TABELA 4 – Ocorrências dos nomes gerais no corpus de Coimbra	76
TABELA 5 – Influência do fator presença ou ausência de anáfora na indeterminação do sujeito por meio de nomes gerais em Bambuí.....	82
TABELA 6 – Influência do fator grau de indeterminação na indeterminação do sujeito por meio de nomes gerais em Bambuí.....	83
TABELA 7 – Influência do fator sexo na indeterminação do sujeito por meio de nomes gerais em Bambuí.....	84
TABELA 8 – Influência do fator faixa etária na indeterminação do sujeito por meio de nomes gerais em Bambuí	86
TABELA 9 – Influência do fator presença ou ausência de locativo na indeterminação do sujeito por meio de nomes gerais em Bambuí.....	87
TABELA 10 – Influência do fator mais ou menos coletivo na indeterminação do sujeito por meio de nomes gerais em Bambuí.....	88
TABELA 11 – Influência do fator presença ou ausência de locativo na indeterminação do sujeito por meio de nomes gerais em Coimbra	89
TABELA 12 – Influência do fator faixa etária na indeterminação do sujeito por meio de nomes gerais em Coimbra	90
TABELA 13 – Influência do fator grau de indeterminação na indeterminação do sujeito por meio de nomes gerais em Coimbra.....	91
TABELA 14 – Influência do fator mais ou menos coletivo na indeterminação do sujeito por meio de nomes gerais em Coimbra.....	92
TABELA 15 – Influência do fator sexo na indeterminação do sujeito por meio de nomes gerais em Coimbra.....	94
TABELA 16 – Influência do fator presença ou ausência de anáfora na indeterminação do sujeito por meio de nomes gerais em Coimbra	94

LISTA DE GRÁFICOS

GRÁFICO 1 – Influência da variável sexo no conjunto dos nomes gerais em Bambuí.....	96
GRÁFICO 2 – Influência da variável faixa etária no conjunto dos nomes gerais em Coimbra	97
GRÁFICO 3 – Influência da variável presença ou ausência de locativo no conjunto dos nomes gerais em Coimbra	99
GRÁFICO 4 – Influência da variável ausência ou presença de anáfora no conjunto dos nomes gerais em Bambuí.....	101
GRÁFICO 5 – Influência da variável grau de indeterminação no conjunto dos nomes gerais em Coimbra	103
GRÁFICO 6 – Influência da variável grau de indeterminação no conjunto dos nomes gerais em Bambuí.....	103
GRÁFICO 7 – Influência da variável mais ou menos coletivo no conjunto dos nomes gerais em Coimbra	105

LISTA DE SIGLAS

- ANPOLL:** Associação Nacional de Pós-Graduação em Letras e Linguística
- GEVAR:** Grupo de Estudos Variacionistas
- GIEL:** Grupo Interinstitucional de Estudos de Língua(gem): usos, contatos e fronteiras
- LUAL:** A Língua Usada em Alagoas
- NURC:** Projeto de Estudo da Norma Urbana Culta do Brasil
- PB:** Português brasileiro
- PE:** Português europeu
- PEUL:** Programa de Estudos sobre o Uso da Língua
- VALPB:** Projeto Variação Linguística na Paraíba
- VPSA:** voz passiva sem agente
- VPASSINT:** voz passiva sintética
- VARSUL:** Variação Linguística Urbana da Região Sul
- VarX:** Banco de Dados por Classe Social
- Ø+V1PP:** verbo na primeira pessoa do plural sem sujeito explícito
- Ø+V3PS:** verbo na terceira pessoa do singular sem sujeito explícito
- Ø+V3PP:** verbo na terceira pessoa do plural sem sujeito explícito
- Ø+VINF:** verbo no infinitivo sem sujeito explícito

SUMÁRIO

1 INTRODUÇÃO	17
2 REFERENCIAL TEÓRICO	19
2.1 Sociolinguística variacionista.....	19
2.1.1 As três ondas da sociolinguística	22
2.2 A variação no português brasileiro (PB) e no português europeu (PE).....	24
2.3 Impessoalização, indefinição e indeterminação do sujeito.....	30
2.3.1 Impessoalização	30
2.3.2 Indefinição.....	35
2.3.3 Indeterminação do sujeito	37
2.4 Nomes gerais.....	46
3 METODOLOGIA.....	51
3.1 A escolha das cidades.....	51
3.2 História de Bambuí e de Coimbra.....	51
3.3 Bambuí.....	51
3.4 Coimbra	53
3.5 A seleção dos informantes	55
3.6 A gravação e o registro dos dados	57
3.7 A transcrição dos dados	58
3.8 A preparação dos dados para análise estatística.....	59
3.9 As variantes	60
3.9.1 Os nomes gerais	61
3.9.2 As demais formas de indeterminação encontradas	62
3.10 Variáveis extralinguísticas	64
3.10.1 Sexo.....	65
3.10.2 Faixa etária	65
3.11 Variáveis linguísticas	66
3.11.1 Presença ou ausência de anáfora	66
3.11.2 Grau de indeterminação	68
3.11.3 Presença ou ausência de locativo	69
3.11.4 Mais ou menos coletivo	70
3.12 A codificação dos dados.....	71
4 RESULTADOS E ANÁLISE DOS DADOS	73
4.1 Apresentação dos resultados e das variantes	73
4.1.1 A gente	74
4.1.2 Você(s)/ocê/cê.....	75
4.1.3 Nomes gerais	75
4.1.4 Nós/nóis/-mos.....	77
4.1.5 Eles/ês.....	78
4.1.6 Verbo na terceira pessoa do singular sem sujeito explícito (Ø+V3PS)	78

4.1.7 Verbo na terceira pessoa do plural sem sujeito explícito (Ø+V3PP).....	78
4.1.8 Verbo no infinitivo sem sujeito explícito (Ø+VINFINF).....	79
4.1.9 Se.....	79
4.1.10 Tu	80
4.2 Análise dos fatores selecionados estatisticamente pelo programa GoldVarb X em Bambuí.....	81
4.2.1 Presença ou ausência anáfora.....	81
4.2.2 Grau de indeterminação	83
4.2.3 Sexo.....	84
4.3 Análise qualitativa dos fatores que não foram selecionados pelo programa GoldVarb X em Bambuí.....	85
4.3.1 Faixa etária	86
4.3.2 Presença ou ausência de locativo	86
4.3.3 Mais ou menos coletivo.....	88
4.4 Análise dos fatores selecionados estatisticamente pelo programa GoldVarb X de Coimbra	88
4.4.1 Presença ou ausência de locativo	88
4.4.2 Faixa etária	90
4.4.3 Grau de indeterminação	91
4.4.4 Mais ou menos coletivo.....	92
4.5 Análise qualitativa dos fatores que não foram selecionados pelo programa GoldVarb X em Coimbra.....	93
4.5.1 Sexo.....	93
4.5.2 Presença ou ausência de anáfora	94
4.6 Comparação entre os resultados de Bambuí e de Coimbra.....	95
4.6.1 Sexo.....	95
4.6.2 Faixa etária	97
4.6.3 Presença ou ausência de locativo	98
4.6.4 Presença ou ausência de anáfora	100
4.6.5 Grau de indeterminação	102
4.6.6 Mais ou menos coletivo.....	104
5 CONSIDERAÇÕES FINAIS.....	106
REFERÊNCIAS	113
APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA INFORMANTES DE BAMBUÍ	120
APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA INFORMANTES DE COIMBRA.....	122
APÊNDICE C – FICHA DO INFORMANTE.....	124
APÊNDICE D – ROTEIRO DE ENTREVISTAS.....	125
APÊNDICE E – CODIFICAÇÃO DOS DADOS	127

1 INTRODUÇÃO

Este trabalho tem como objetivo principal observar a relevância dos nomes gerais no conjunto das variantes indeterminadoras, verificando a influência de fatores linguísticos e extralinguísticos na realização do fenômeno. Além disso, foi também nosso objetivo propor uma comparação entre o português brasileiro (PB) e o português europeu (PE), a partir de uma amostra de 48 gravações, coletadas no Brasil, na cidade de Bambuí, no estado de Minas Gerais, e em Coimbra, localizada em Portugal. Salienta-se que, para este estudo, o sujeito indeterminado é considerado como aquele que não se pode recuperar por meio do contexto ou de outros recursos disponíveis na fala.

Dessa forma, esta pesquisa busca observar as ocorrências das várias formas de indeterminação do sujeito, tais como *a gente*; *eles*; *nomes gerais*; *nós*; *se*; *você*; *verbo na terceira pessoa do singular* ($\emptyset + V3PS$); *verbo na terceira pessoa do plural* ($\emptyset + V3PS$) e *tu*, tendo em vista alguns trabalhos que discutem o fenômeno, como o de Milanez (1982), de Menon (1994), de Setti (1997), de Godoy (1999), de Duarte, Kato e Barbosa (2001), de Oliveira (2006), de Duarte (2008), de Carvalho (2010) e de Oliveira (2018). O foco da pesquisa é voltado para os itens denominados de nomes gerais, isto é, itens que são definidos por Halliday e Hasan (1995 [1976]) como um caso limítrofe entre um item lexical e um item gramatical e possuem traços mínimos de significado. O exemplo (1), encontrado em nosso *corpus*, demonstra como os nomes gerais são capazes de indeterminar:

(1) portanto eu já não tenho cá muitos conhecimentos não e as **pessoas**¹ hoje são diferentes sabe? (COI11F)²

No exemplo (1), tem-se um caso de indeterminação do sujeito segundo os autores apresentados, porque não é possível recuperar sobre quais pessoas a informante se refere. Nesse trecho, ela relata como as pessoas mais jovens são diferentes das mais antigas, não havendo nenhuma forma de se recuperar sobre quem se fala.

Nesse sentido, são de grande interesse os contextos em que as variantes selecionadas ocorrem e, ainda, visando-se comparar dados de países distintos, busca-se discutir os fatores linguísticos e extralinguísticos que influenciam cada um dos itens analisados. No que diz

¹ Autores como Cunha e Cintra (2008) analisam esse exemplo como sujeito simples, contudo, os autores que servem como referencial teórico deste estudo apontam para a classificação de sujeito indeterminado.

² A sigla de codificação é representada pelo nome da cidade BAM (Bambuí) ou COI (Coimbra), seguida do número de identificação e do sexo do informante.

respeito aos nomes gerais, será explorada a relevância que eles parecem ter no conjunto das variantes.

Ademais, este estudo justifica-se pela necessidade de se compreender melhor como o fenômeno da indeterminação tem ocorrido na fala e, principalmente, se os nomes gerais têm sido utilizados nesses contextos. Alguns estudos apresentados e que serão utilizados como base teórica para este trabalho demonstram resultados muitos semelhantes sobre esse tema, contudo, nenhum deles propôs, até o momento, uma comparação entre o PB e o PE. Esse é, portanto, o fator inédito de nossa pesquisa, uma vez que não foram encontrados trabalhos na área da sociolinguística variacionista que trouxessem essa comparação tendo como foco a indeterminação do sujeito. Diante disso, esta pesquisa busca responder às seguintes questões: Quais são e como se comportam as formas de indeterminação do sujeito presentes na fala de Bambuí e de Coimbra? Qual a relevância dos nomes gerais nesse fenômeno?

Os nomes gerais, já estudados como formas de indeterminação por Milanez (1982), Menon (1994), Setti (1997), Godoy (1999) e Carvalho (2010) e tratados por estes autores como formas nominais, ocupam, em geral, a terceira posição no conjunto das variantes. Além disso, os trabalhos de Duarte, Kato e Barbosa (2001), Oliveira (2006) e Duarte (2008) mostram uma diferenciação encontrada no PE e no PB no que concerne à indeterminação do sujeito. Dessa forma, a nossa hipótese é que em Coimbra se encontrariam nomes gerais diferentes daqueles que poderiam ocorrer em Bambuí e, ainda, que no Brasil os nomes gerais seriam mais usados do que em Portugal, tendo em vista os resultados das várias pesquisas sobre esses elementos na última década (AMARAL, 2017; MIHATSCH, 2017; OLIVEIRA, 2018; AMARAL E MIHATSCH, 2019).

Assim, este texto contém, no capítulo 1, a introdução, seguida do capítulo 2, que trata do referencial teórico e que discorre sobre a sociolinguística variacionista, sobre a variação existente no PB e no PE e sobre a distinção entre impessoalização, indefinição e indeterminação do sujeito. Posteriormente, tem-se, no capítulo 3, a descrição dos processos metodológicos, em que são apresentados os procedimentos adotados para a coleta e análise dos dados. Neste mesmo capítulo, discutem-se os resultados obtidos a partir das rodadas realizadas no *software* GoldVarb X, comparando-se os dados das duas localidades. Por fim, no capítulo 5, têm-se as considerações finais.

2 REFERENCIAL TEÓRICO

O referencial teórico deste trabalho é dividido em quatro seções. A primeira delas apresenta os principais conceitos da Sociolinguística, tendo como base a sociolinguística laboviana (2008 [1972]), além de Weinreich, Labov e Herzog (2006 [1975]) e outros autores que também se dedicaram ao tema, como Lavandera (2014 [1984]), Mollica (2004), Alkmim (2006), Moreno Fernández (2009) e Coelho *et al.* (2015). Além disso, são feitas considerações sobre as três ondas da sociolinguística. A segunda parte trata da variação no PB e no PE com foco na indeterminação do sujeito. A terceira seção propõe uma diferenciação dos conceitos de impessoalização, de indefinição e de indeterminação. Finalmente, têm-se os estudos sobre os nomes gerais, foco desta pesquisa. Nesta última seção, procura-se demonstrar como nomes que possuem traço [+ humano] são capazes de indeterminar o sujeito em português.

2.1 Sociolinguística variacionista

A Sociolinguística Variacionista, também denominada Sociolinguística Laboviana, Sociolinguística Quantitativa ou Teoria da Variação e Mudança, como relatam Coelho *et al.* (2015), constitui uma área de estudo que visa à relação entre língua e sociedade. Segundo Alkmim (2006), esse campo de pesquisa tem como objeto de estudo a língua falada em situações reais de uso e tem como ponto de partida a comunidade linguística, isto é, “um conjunto de pessoas que interagem verbalmente e que compartilham um conjunto de normas com respeito aos usos linguísticos” (ALKMIM, 2006, p. 33).

Embora a relação entre aspectos linguísticos e fatores sociais tenha sido apontada por autores de estudos da Dialetoлогия e da Geografia Linguística (NASCENTES, 1953; NASCENTES, 1958, 1961; SILVA NETO, 1957; ROSSI, 1963) e outros, conforme citado por Ferreira (2019), foi com as pesquisas de Labov (2008 [1972]) que o tema ganhou destaque, a partir de um estudo importante com dados de Martha’s Vineyard, uma ilha no estado de Massachusetts. Neste, analisando a frequência das variantes fonéticas /ay/ e /aw/, o autor considerou fatores extralinguísticos, tais como faixa etária, grupos étnicos etc. No caso da faixa etária, por exemplo, a pesquisa percebeu que a centralização do (ay) e (aw) parece sofrer um aumento sucessivo, alcançando um pico mais alto no grupo de 31 a 45 anos.

Assim, considerando que a heterogeneidade social reflete-se na língua, chegou-se a um dos principais pressupostos da sociolinguística que serão considerados para este trabalho que

diz respeito ao caráter heterogêneo da língua. De acordo com Weinreich, Labov e Herzog (2006 [1975]), a língua é um sistema heterogêneo, porém estruturado, ou seja, a língua é um sistema organizado, mas não homogêneo. Ela possui regras que os falantes obedecem ainda que inconscientemente; contudo, ela sofre variação e mudança e são nesses momentos que se pode observar seu caráter heterogêneo. Sobre isso, os autores afirmam que

o caráter heterogêneo dos sistemas linguísticos discutidos até agora é produto de combinações, alternâncias ou mosaicos de subsistemas distintos, conjuntamente disponíveis. Cada um desses subsistemas é concebido como um corpo coerente e integral de regras do tipo categórico, neogramático: o único aparato teórico adicional necessário é um conjunto de regras que afirmem as condições para a alternância. (WEINREICH, LABOV e HERZOG, 2006 [1975], p. 103).

Vê-se, portanto, que a sociolinguística concebe a língua como um sistema que não ocorre por si só; ela sofre interferência de fatores externos como idade, sexo e classe social e, ainda, de fatores que estão dentro da própria estrutura linguística. Fatores externos são denominados extralinguísticos, e internos, fatores linguísticos.

Outros dois conceitos importantes para essa área de estudo são o de variável e o de variante. De acordo com Coelho *et al.* (2015, p. 17), variável é “o lugar na gramática em que se localiza a variação, de forma mais abstrata” e variantes são “formas individuais que ‘disputam’ pela expressão da variável”. Neste trabalho, a variável é denominada formas de indeterminação do sujeito e as variantes são as dez maneiras distintas em que essa indeterminação se apresenta. São elas: *a gente, você/ocê/cê, nomes gerais, nós/nóis/Ø+VIPP, eles/ês, Ø+V3PS, Ø+V3PP, Ø+VINF, se e tu.*

Acrescenta-se, ainda, que corroborando Lavandera (2014 [1984]), Coelho *et al.* (2015) relatam que é possível um trabalho em uma vertente variacionista quando “duas formas podem ocorrer no mesmo contexto com o mesmo valor referencial/representacional, isto é, com o mesmo significado” (COELHO *et al.*, 2015, p. 16). Desta forma, vale salientar que neste estudo o valor não é referencial/representacional, mas funcional, de acordo com a descrição de Lavandera (2014 [1984]). Conforme demonstra a autora, a noção de variante foi criada, inicialmente, com base em estudos fonológicos e, devido a isso, a definição, muitas vezes, não consegue abarcar pesquisas voltadas para variantes sintáticas e morfológicas. Assim, ela mostra em seu trabalho que, para se definir variantes que não sejam fonológicas, “a proposta é enfraquecer a condição de que o significado deve ser o mesmo para todas as formas alternantes

e substituí-las por uma condição de comparabilidade funcional³ (LAVANDERA, 2014 [1984], p. 47 [Tradução minha]).

Moreno Fernández (2009) evidencia, assim como Lavandera (2014 [1984]), que, no caso de variantes gramaticais, é importante atentar-se para a equivalência existente entre elas. O autor apresenta três tipos de equivalência, sendo elas a referencial, a semântica e a funcional ou pragmática. A que interessa a este estudo é a funcional ou pragmática, pois ela demonstra que variantes podem ser aquelas que possuem a mesma intenção comunicativa, como é o caso de quando um informante utiliza uma ou outra forma para indeterminar o sujeito.

Mollica (2004), ao explicar a relevância das variáveis extralinguísticas, relata que a variação linguística, considerando seu caráter heterogêneo e ordenado, “é controlada por variáveis estruturais e sociais” (MOLLICA, 2004, p. 27). Nesse sentido, essas variáveis não devem ser consideradas isoladamente, mas pensadas que, em conjunto, podem favorecer ou desfavorecer o uso de uma determinada forma. O estudo de Scherre e Yacovenco (2011) retrata essa interligação quando observa que os fatores *sexo* e *faixa etária* trazem considerações significativas sobre o uso de *tu* e *você* pelo país.

No que diz respeito às variáveis linguísticas, Silva (2004) as define como as influências de fatores internos da língua (fonológicos, morfológicos, sintáticos, semânticos etc.). Coelho *et al.* (2015, p. 20) exemplificam esses fatores com “ordem dos constituintes em uma sentença, a classe das palavras envolvidas no fenômeno em variação, aspectos semânticos etc.”. Os autores ainda acrescentam que existe uma correlação muito forte entre os fatores internos e externos da língua nos estudos dessa área, ao afirmarem que

A Sociolinguística assume, portanto, que existe uma forte correlação entre os mecanismos internos da língua e fatores externos a ela, tanto de uma ordem “micro”, envolvendo nosso grau de contato e de identificação com os grupos com os quais interagimos no dia a dia, quanto de uma ordem “macro”, relacionada a uma estratificação mais simples (COELHO *et al.* 2015, p. 22).

Entende-se, portanto, que, para se realizar um trabalho que busque entender a língua e a sociedade, deve-se compreender tanto aquilo que influencia internamente quanto externamente o emprego das variantes.

Nessa explanação, demonstrou-se um pouco da história e dos principais conceitos que se mostram relevantes para esta pesquisa em sociolinguística. Na próxima seção, serão

³No original: “*Mi propuesta es debilitar la condición de que el significado deba ser el mismo para todas las formas alternantes, y reemplazarla por una condición de comparabilidad funcional*”.

apresentadas as três ondas da sociolinguística. Essa denominação é dada para explicar três tendências de estudos dessa área.

2.1.1 As três ondas da sociolinguística

As três ondas da sociolinguística, segundo Eckert (2012), dizem respeito a três vertentes de estudos que serviram para explicar fenômenos linguísticos diversos que envolvem a língua e a sociedade. Sendo assim, o primeiro momento tem como marco inicial os estudos voltados para a estratificação do inglês na cidade de Nova York feitos por Labov. Esses trabalhos estão, na maioria dos casos, voltados para a análise do fator socioeconômico, mostrando, dessa forma, como a língua é um marcador de *status* social. Um exemplo que ilustra essa onda é o do estudo da estratificação social do (r) em lojas de departamentos na cidade de Nova York, no qual Labov (2008 [1972]) concluiu que as funcionárias de uma loja que atende um público de classe mais elevada têm uma pronúncia diferente do (r) quando comparada à pronúncia das funcionárias de lojas que atendem à população de classes mais baixas.

Freitag, Martins e Tavares (2012) demonstram que esses estudos são de natureza quantitativa e que o Brasil contém alguns *corpora* que ilustram essa onda, tais como o Programa de Estudos sobre o Uso da Língua (PEUL), da Universidade do Rio de Janeiro; o Projeto Variação Linguística Urbana na Região Sul do Brasil (VARISUL), formado pelas equipes da Universidade Federal do Paraná, da Universidade Federal de Santa Catarina, da Universidade Federal do Rio Grande do Sul e da Pontifícia Universidade Católica do Rio Grande do Sul; o Projeto Variação Linguística na Paraíba (VALPB), da Universidade Federal da Paraíba; o Banco de Dados Sociolinguístico da Fronteira e da Campanha Sul-Rio-Grandense – BDS-Pampa, da equipe da Universidade Federal de Pelotas e da Pontifícia Universidade de Pelotas; o Banco de Dados por Classe Social – VarX, da Universidade Federal de Pelotas, entre outros.

Salomão (2011), além de apresentar um histórico dos estudos sociolinguísticos, faz também um levantamento de mais de 40 grupos de trabalhos que surgiram a partir de 2000 e que propuseram pesquisas na área da sociolinguística variacionista no Brasil. Alguns deles propõem estudos a partir da montagem de *corpus* e utilizam a análise quantitativa e qualitativa como base, como, por exemplo, Descrição e Análise de Fenômenos do Português do Extremo Sul do Brasil, formado em 2000 pela Universidade Federal de Pelotas (UFPEL); História Social do Português, formado em 2000 pela Universidade Federal de Minas Gerais (UFMG); Vertentes, formado em 2001 pela Universidade Federal da Bahia (UFBA); Grupo

Interinstitucional de Estudos de Língua(gem): usos, contatos e fronteiras (GIEL), formado em 2002 pela Universidade Federal Fluminense (UFF); Acervos Documentais: Memória Linguística e Literária, formado em 2002 pela Universidade Federal de Ouro Preto – UFOP; Grupo de Estudos Linguísticos do Acre, formado em 2007 pela Universidade Federal do Acre (UFAC); VARFON-Minas, formado em 2008 pela Universidade Federal de Minas Gerais (UFMG), Grupo de Estudos Variacionistas – GEVAR, formado em 2009 pela Universidade Federal do Triângulo Mineiro (UFTM); Grupo de Estudos em Fonologia, formado em 2011 pela Universidade Federal da Paraíba (UFPB), entre muitos outros espalhados por todo o Brasil.

Oliveira (2016), ao fazer um levantamento sobre os estudos no campo da sociolinguística, apresenta os projetos pioneiros nessa área, feitos em grupo, e que tinham como objetivo principal constituir um banco de dados com dados de fala. São eles: Projeto de Estudo da Norma Urbana Culta do Brasil (NURC) (1969), Competências Básicas do Português (1977), Programa de Estudos dos Usos da Língua (PEUL) (1980), Confluência Dialeto na Nova Capital Brasileira (1984), Variação Linguística Urbana da Região Sul (VARISUL) (1989), Variação Linguística na Paraíba (VALPB) (1994), Dialeto Sociais Cearenses (1996) e A Língua Usada em Alagoas (LUAL) (1997). Ademais, Oliveira (2016) também apresenta os primeiros trabalhos realizados nessa vertente e que foram orientados por Anthony Naro. Têm-se, nessa relação, estudos que discutem a concordância verbal, as construções relativas, os segmentos nasais, entre outros fenômenos.

No que diz respeito aos rumos da vertente da primeira onda na atualidade, Oliveira (2016) relata um panorama promissor, pois muitas pesquisas e publicações têm sido feitas, além de congressos e do Grupo de Trabalho de Sociolinguística da Associação Nacional de Pós-Graduação em Letras e Linguística (ANPOLL), o qual conta com um número grande de associados e que busca “discutir resultados de pesquisas na área, além de organizar também publicações” (OLIVEIRA, 2016, p. 485).

Ao tratar da segunda onda, Eckert (2012) relata que é possível encontrar estudos voltados para pequenos grupos com foco no vernáculo. Esses trabalhos de natureza etnográfica buscam observar características mais específicas de um grupo ao longo de um período de tempo. Um dos exemplos dados pela autora é sobre o inglês afroamericano na cidade de Nova York, em que Labov observou o uso de emprego de traços vernaculares por meninos pré-adolescentes como forma de indexadores de *status* pelo grupo.

Freitag, Martins e Tavares (2012) mostram que, no Brasil, a segunda onda não tinha sido muito desenvolvida até o período em que fizeram o levantamento e poucos estudos haviam

sido realizados. O de Ferrari (1994), citado por eles como dos poucos trabalhos, analisou 12 traços fonológicos, divididos entre discretos e gradientes, e um traço sintático-semântico (variação de “em” e “ni”) na comunidade de Morro dos Caboclos, no Rio de Janeiro, buscando verificar a relação entre variação e redes sociais no local.

Segundo Eckert (2012), a primeira e a segunda ondas consideravam a variação como uma consequência acidental do espaço social. A terceira, contudo, vê a variação como uma característica essencial da linguagem. Assim, como definem Freitag, Martins e Tavares (2012, p. 923), a terceira onda “centra o foco na variação vista não como o reflexo do lugar social num ponto da escala, mas como um recurso para a construção de significado social”. Além disso, os autores explicam que houve uma mudança da comunidade de fala de Labov para a comunidade de prática. Estudos dessa onda são, por exemplo, o de Podesva (2002) sobre homossexuais, como cita Eckert (2012).

Com isso, pode ser observado que tanto Eckert (2012) quanto Freitag, Martins e Tavares (2012) reconhecem que as três ondas não são delimitadas no tempo, ou seja, uma não substitui a outra, elas coexistem. Assim, a partir dessas considerações, pode-se concluir que a proposta desta pesquisa se enquadra nos estudos de primeira onda por estar mais próximo das pesquisas realizadas por Labov e de um grande conjunto de pesquisas realizadas no Brasil durante as últimas décadas.

2.2 A variação no português brasileiro (PB) e no português europeu (PE)

A variação no PB e no PE é tratada aqui a partir dos trabalhos de Duarte, Kato e Barbosa (2001), Mateus *et al.* (2003), Oliveira (2006), Duarte (2008), Carvalho (2009), Rumeu (2011) e Aragão e Araújo (2017). A discussão é feita por ordem cronológica para que seja possível entender como o estudo comparativo das duas línguas recebeu contribuições com o decorrer dos anos.

Duarte, Kato e Barbosa (2001), ao fazerem um estudo sobre a indeterminação no PB e no PE, apresentam comparações interessantes com o francês, o italiano e o romeno. Os autores salientam que a inexistência do *se* nominativo em francês, por exemplo, estaria relacionada ao parâmetro de sujeito nulo. Além disso, no artigo, os autores concluem que, no PB coloquial, o uso de *se* nominativo tende a ser suprimido e formas pronominais tendem a ser usadas. No PE, no entanto, ocorre o contrário, com o favorecimento da utilização do *se*.

Mateus *et al.* (2003) discorrem sobre as diferenças que podem ser encontradas nas variedades faladas em Portugal e no Brasil nos níveis fonéticos, lexical, morfológico e sintático-semântico. A seguir, apresenta-se um quadro com as principais diferenças apontadas pelas autoras:

Quadro 1 – Diferenças entre PB e PE segundo os níveis fonético, morfológico, sintático e lexical

Nível	PB	PE	Observação
Fonético	<i>m[o]rar</i> <i>l[e]var</i>	<i>m[u]rar</i> <i>l[i]var</i>	As vogais átonas no PB são muito menos reduzidas do que no PE.
Fonético	<i>[t]io</i> <i>[dʒ]irector</i> <i>ba[t]i</i> <i>pe[dʒ]i</i>	<i>[t]io</i> <i>[d]irector</i> <i>bat[e]</i> <i>pe[d]e</i>	Antes de /i/ tônico e átono e antes de /e/ pós-tônico, o /t/ e o /d/ palatalizam e realizam-se como africadas no PB, pronunciando-se respectivamente como [tʃ] e [dʒ], enquanto em PE se mantêm como oclusivas. Vale lembrar que essa variante não é comum em todas as variedades do PB.
Fonético	<i>anima[w]</i> <i>sa[w]tar</i>	<i>anima[ɰ]</i> <i>sa[ɰ]tar</i>	Em final de sílaba e de palavra, o /l/ pronuncia-se como a semivogal [w] no PB e velariza-se em PE (esta velarização representa-se por [ɰ]). No sul do Brasil, é possível encontrar a velarização dessa variante.
Fonético	<i>senhor[r]/senho[x]/</i> <i>senho[h]/senh[ó]</i> <i>ama[r]/ama[x]/</i> <i>ama[h]/am[á]</i>	<i>senho[r]</i> <i>ama[r]</i>	A vibrante final de palavra admite variação de pronúncia no PB, podendo ocorrer como vibrante simples [r], fricativa [x], aspirada [h] ou ainda ser suprimida, enquanto em PE ocorre sempre como vibrante tap, [r], embora no estilo coloquial possa também ser suprimida.
Fonético	<i>me[z]mo</i> <i>pe[s]te</i> <i>menino[s]</i>	<i>me[ʒ]mo</i> <i>pe[ʃ]te</i> <i>menino[ʃ]</i>	As sibilantes em final de sílaba e de palavra mantêm-se como [s] e [z] no PB e pronunciam-se como palatais, [ʃ] e [ʒ], em PE. As

Fonético	<i>cap[i]tura</i> <i>ab[i]surdo</i> <i>p[i]neu</i>	<i>ca[pt]tura</i> <i>a[bs]urdo</i> <i>[pn]eu</i>	variantes encontradas no PE também podem ser encontradas no PB. No PB verifica-se a introdução de um [i] epentético entre duas consoantes enquanto em PE as duas consoantes se mantêm em sequência.
Morfológico e sintático	eu vi <i>ele</i> na rua deixa <i>ela</i> comigo	eu vi- <i>o</i> na rua deixa- <i>a</i> comigo	O PB vernacular perdeu os clíticos de terceira pessoa e apresenta, ao lado do objeto direto nulo, construções com os pronomes <i>ele/ela</i> e <i>lhe</i> . Salienta-se que os clíticos continuam no PB em situações mais formais.
Morfológico e sintático	<i>me</i> diga uma coisa a menina <i>se</i> levantou	diga- <i>me</i> uma coisa a menina levantou- <i>se</i>	Em PB os clíticos são habitualmente colocados em posição proclítica enquanto, nas frases correspondentes, são colocados em posição enclítica no PE.
Morfológico e sintático	<i>Progressivo:</i> estava <i>brincando</i> <i>Predicado secundário:</i> passou um ano <i>ouvindo</i>	<i>Progressivo:</i> estava <i>a</i> <i>brincar</i> <i>Predicado secundário:</i> passou um ano <i>a ouvir</i>	As construções com gerúndio podem ter funções sintáticas de dois tipos: progressivo e de predicados secundários. Em ambos os casos, verifica-se que o PB apresenta construções de gerúndio ao passo que o PE apresenta quase sempre construções com infinitivo.
Morfológico e sintático	Ele está em Paris <i>faz</i> três anos O Paulo parte para Roma <i>em</i> quinze dias.	Ele está em Paris <i>há</i> três anos O Paulo parte para Roma <i>dentro de</i> quinze dias.	Em construções que exprimem distância temporal, o PB utiliza os verbos <i>fazer</i> e <i>ter</i> quando o PE usa o verbo <i>haver</i> . Além dessa diferença de nível lexical, nota-se também uma distinção léxico-semântica, utilizando o PB estruturas inexistentes em PE.
Morfológico e sintático	<i>tem</i> fogo naquela casa	<i>há</i> fogo naquela casa	Os verbos <i>ter</i> e <i>haver</i> têm uso diferente em PB e em PE com o significado de “existir”.

Morfológico e sintático	vou comprar <i>meu</i> vestido	vou comprar <i>o meu</i> vestido	Em PB, habitualmente, antes de possessivo pré-nominal, há ausência de artigo, enquanto em PE este está sempre presente.
Morfológico e sintático	pena que ele não <i>chegou</i> a tempo imaginemos que hoje <i>seja</i> domingo	é pena que ele não <i>tenha</i> <i>chegado</i> a tempo imaginemos que hoje <i>é</i> domingo	Certas construções com predicados (<i>ser</i>) <i>bom</i> , (<i>ser</i>) <i>natural</i> implicam a utilização do modo indicativo em PB e conjuntivo em PE. Pelo contrário, em construções com verbos como <i>supor</i> , <i>imaginar</i> , o PB pode usar o conjuntivo e o PE o indicativo.
Formas de Tratamento	Já te disse que <i>você</i> não vai O <i>senhor/a senhora</i> quer jantar?	Já te disse que <i>tu</i> não vais/Já <i>lhe</i> disse que <i>você</i> não vai A Antônia/O Sr. Dr./o Sr. Diretor/o tio/o meu amigo quer jantar?	No português brasileiro, a utilização de <i>você</i> e suas variantes <i>ocê</i> e <i>cê</i> substitui, na maioria dos dialetos, o <i>tu</i> . Enquanto em PE o tratamento diferente usa o nome próprio, o cargo, o título, ou o grau de parentesco, no PB utiliza-se <i>o senhor, a senhora</i> e, no interior dos grupos profissionais, o cargo ou título.
Léxico	<i>banheiro</i> <i>camisa de dormir</i>	<i>casa de banho</i> <i>camisola</i>	Palavras idênticas com com significados diferentes (“falsos amigos”).
Léxico	<i>esparadrapo</i> <i>concreto</i>	<i>adesivo</i> <i>betão</i>	Palavras diferentes com o mesmo significado.
Léxico	<i>fumante</i> <i>prestativo</i>	<i>fumador</i> <i>prestável</i>	Palavras derivadas com a mesma base e diferentes sufixos, mas com idêntico significado.
Léxico	<i>polpudo</i> <i>encanador</i>	<i>carnudo</i> <i>canalizador</i>	Palavras com o mesmo sufixo e diferentes bases, mas com significado idêntico.

Fonte: Baseado em Mateus *et al.* (2003).

Mateus *et al.* (2003, p. 51), após apresentarem as diferenças entre o PB e o PE, concluem que “a deriva das duas variedades utiliza processos gramaticais comuns, não podendo afirmar-se, portanto, que as variedades do português possuem diferentes gramáticas”. Assim, tendo em

vista esta pesquisa, percebe-se que as construções de indeterminação utilizadas por portugueses e brasileiros são as mesmas e a diferença se dá apenas na preferência de uso das formas. Isso corrobora Mateus e *al.* (2003), pois não é possível falar em diferentes gramáticas em nosso estudo, mas, sim, da diferença de frequência no uso de uma ou outra forma de indeterminação a depender da localidade do falante.

Oliveira (2006) tem como objetivo principal observar as formas pronominais *nós* e *a gente* e o clítico *se* seguido de verbo no infinitivo no PB e no PE. A autora faz uma comparação entre gravações feitas na cidade de Uberlândia-MG e, para a montagem do *corpus* do PE, faz uso de gravações realizadas em diversas regiões de Portugal, inclusive Coimbra. Essas entrevistas foram retiradas do projeto Português Fundamental, constituído de gravações feitas na década de 70. É importante observar o distanciamento temporal entre a coleta dos dados de Coimbra e a realização do trabalho, uma vez que isso pode influenciar nos resultados dos estudos e nas futuras comparações com esta pesquisa. A partir desse levantamento, Oliveira (2006) concluiu que *nós*, *a gente* e *se + infinitivo* são estratégias empregadas tanto no PB quanto no PE, mas que a frequência em que essas estratégias ocorrem são diferentes. Enquanto no PB há preferência por *a gente*, no português europeu as estratégias *nós* e *a gente* são igualmente favorecidas.

Duarte (2008) apresenta um estudo sobre a realização do sujeito de referência indeterminada em sentenças infinitivas com dados de língua oral e escrita no PB e no PE. Nesse trabalho, a autora conclui que, no PB oral, há um favorecimento do uso de formas pronominais nominativas para a representação do sujeito em sentenças infinitivas. No que concerne ao PB escrito e ao PE escrito e oral, a pesquisadora constata uma “preferência pelo sujeito nulo nas infinitivas, uso parcimonioso de *se*, utilizado como uma estratégia marginal, algumas vezes para garantir uma interpretação arbitrária do sujeito” (DUARTE, 2008, p. 23).

Carvalho (2009), ao investigar a realização do sujeito no português europeu, tem como objetivo atestar ou refutar a denominação de [+] *pro-drop*⁴ que essa língua ostenta historicamente. Nesse sentido, a autora selecionou 10 entrevistas do *Corpus* Dialetal para o Estudo da Sintaxe (CORDIAL-SIN). Esse *corpus*, de acordo com Carvalho (2009), ainda não estava totalmente transcrito e suas gravações são divididas de quatro formas: (1) transcrição conservadora; (2) transcrição ortográfica normalizada; (3) texto morfológicamente anotado; (4) texto sintaticamente anotado, esse último a ser disponibilizado depois de concluído o *corpus*.

⁴ Línguas *pro-drop* são aquelas em que a posição de sujeito “pode ser ocupada por uma categoria pronominal não realizada foneticamente”. (TAVARES E SILVA, 2013, p. 15).

Vale salientar que os dados da autora foram todos retirados do grupo “transcrição ortográfica normalizada”, pois é esse o grupo que atende melhor ao objeto pesquisado. As entrevistas analisadas foram gravadas com falantes das seguintes localidades de Portugal: Alcochete, Alvor, Arcos (Viana de Castelo), Cabeço de vide, Camacha, Covo, Figueiró, Outeiro, Ponta Garça e Serpa.

Em sua investigação, Carvalho (2009) obteve um alto índice de sujeitos preenchidos no PE, o que o aproxima do PB. Contudo, como em algumas localidades as características de língua [+] *pro-drop* se mantiveram, pode-se inferir que o preenchimento de sujeito pelo uso do pronome não é um fenômeno uniforme em Portugal. Assim, a autora sugere que “os nossos resultados, além de sugerirem que se fala a ‘mesma língua’ dos dois lados do Atlântico, pelo menos em 70% das localidades investigadas, mostram que o português europeu é sensível aos mesmos contextos de variação apresentados pelo português brasileiro” (CARVALHO, 2009, p. 247).

Rumeu (2011) descreve e analisa as estratégias de indeterminação do argumento externo do verbo na produção escrita das imprensas lusitana e brasileira. Os dados utilizados pela autora foram retirados de crônicas e artigos de opinião publicados por jornais lusitanos e brasileiros no período de 2002 a 2005. As formas utilizadas são verbos na 1ª e na 3ª pessoa do plural, na 3ª pessoa do singular com os pronomes *nós*, *eles*, *a gente* e *você* (expressos ou não) e os dados do pronome *se*. A partir dos dados coletados, a autora pôde concluir que as formas *se* e *nós* são as formas mais produtivas tanto no PB quanto no PE. Além disso, obteve-se como resultado que a forma *se* superou a *nós* no PE, o que não ocorreu no PB, isto é, “o *nós* e o *se* se apresentaram em acirrada concorrência para indeterminar o argumento externo do verbo, na qual o uso do *nós* se sobrepõe ligeiramente ao uso do *se*” (RUMEU, 2011, p. 63). Os resultados observados pela autora também evidenciam que o PB utiliza uma maior variedade de formas indeterminadoras do argumento externo do verbo.

Em uma vertente gerativista, assim como em outros trabalhos já citados – como Duarte, Kato e Barbosa (2001) e Duarte (2008) –, Aragão e Araújo (2017) discorrem sobre as diferenças entre o PB e o PE no que diz respeito aos mecanismos de retomada do tópico em posição do sujeito. Os dados foram obtidos a partir de inquéritos disponibilizados virtualmente por projetos de pesquisa da Universidade de Lisboa, de transcrições dos inquéritos do Projeto Norma Urbana Culta e de entrevistas concedidas a programas televisivos do Brasil e de Portugal na década de 2010. Os resultados assemelham-se muito com trabalhos feitos na vertente variacionista que serão apresentados posteriormente, pois uma das conclusões obtidas que merece destaque diz

respeito ao PB apresentar uma necessidade do preenchimento de sujeito, evitando-se formas em que o traço de pessoa é realizado apenas pela forma verbal.

Vê-se, portanto, uma diferenciação nos resultados encontrados entre o PB e o PE. Nesse sentido, um estudo sobre o tema, com foco nos nomes gerais, poderá trazer contribuições ao estudo dos itens que indeterminam o sujeito. A seguir, discutem-se os conceitos de impessoalização, indefinição e indeterminação do sujeito.

2.3 Impessoalização, indefinição e indeterminação do sujeito

Nesta seção, busca-se demonstrar uma distinção entre a impessoalização, a indefinição e a indeterminação do sujeito. Assim, as próximas seções contemplam uma discussão sobre a impessoalização, seguida da indefinição e, por fim, da indeterminação do sujeito. Os autores são apresentados em ordem cronológica, a fim de trazer os mais variados pontos de vista sobre o assunto.

2.3.1 Impessoalização

A impessoalização, a indefinição e a indeterminação têm sido, por muito tempo, termos de difícil conceituação entre os gramáticos. Milanez (1982), ao pesquisar a fala culta paulistana a partir de dados do projeto NURC, propõe uma distinção entre indeterminação, indefinição e impessoalização, mostrando como as noções costumam ser confundidas. A autora elucida que o conceito de indeterminação foi confundido com o de impessoalização durante muito tempo e que, só a partir de 1957, quando o Ministério da Educação e Cultura sugeriu a Nova Nomenclatura Gramatical Brasileira, os autores parecem ter passado a fazer a distinção entre indeterminação e impessoalização. Assim, o que se passou a ver nas gramáticas foram explicações, como as de Cunha e Cintra (2008), que afirmam:

Não deve ser confundido o SUJEITO INDETERMINADO, que existe, mas que não pode ou não se deseja identificar, como a inexistência de sujeito. Em ‘Chove’, ‘Anoitece’ ‘Faz frio’, interessa-nos o processo verbal em si, pois não o atribuímos a nenhum ser. Dize-se então que o verbo é IMPESSOAL; e o sujeito INEXISTENTE (CUNHA; CINTRA, 2008, p. 143).

Nota-se, portanto, que Cunha e Cintra (2008) tentam fazer uma distinção entre indeterminado e impessoal, mostrando que, no primeiro caso, o sujeito existe, mas não pode ou deseja ser identificado e que, no segundo, ele de fato não existe. Os casos de sujeito impessoal, ou, como também denominado por Cunha e Cintra (2008), de *oração sem sujeito*, dizem

respeito àqueles verbos ou expressões que denotam fenômenos da natureza, verbo *haver* na acepção de “existir”, verbos *haver*, *fazer* e *ir*, quando indicam tempo decorrido, ou o verbo *ser*, na indicação do tempo em geral.

Assim, no que diz respeito à impessoalização, Milanez (1982, p. 47) define-a como “a propriedade de certos verbos que, por sua própria natureza, não atribuem a seus SN-sujeitos nenhuma função referencial”. Tal definição é capaz de marcar a grande diferença existente entre a impessoalização e a indeterminação, uma vez que a primeira é um fenômeno que, ocorrendo ou não com sujeito lexical, sempre envolverá uma referência.

Mateus *et al.* (2003), ao discutirem a questão dos clíticos, afirmam que, em frases como (2) e (3), em que o sujeito denota uma entidade arbitrária, tem-se o que alguns autores denominam de impessoal, outros de indeterminado e alguns de *se-nominativo*. Vale lembrar que as gramáticas tradicionais do PB também diferenciam tais exemplos:

(2) Aluga-se casas

(3) Trabalha-se demais

Barbosa (2005, p. 369) define os verbos impessoais tendo em vista a tradição gramatical, isto é, como aqueles que “apenas se conjugam na terceira pessoa do singular”. Contudo, o autor discute e apresenta que os verbos podem se modificar dependendo do contexto, ou seja, um verbo como *parecer*, por exemplo, pode ser pessoal ou impessoal a depender da sentença. Sendo assim, o autor propõe um estudo descritivo da impessoalização nas gramáticas e demonstra que seu foco não é classificar um ou outro como impessoal ou pessoal, pois ele entende que “na sintaxe de todos se encontra sempre representada no sintagma verbal a função de sujeito” (BARBOSA, 2005, p. 375).

Negrão e Viotti (2008) fazem um recorte de uma investigação maior sobre como as línguas africanas podem ter influenciado nas diferenças existentes entre o português brasileiro e o português europeu. Elas ainda pontuam que

enquanto os trabalhos que discutem as origens do português tratam de questões léxico-morfológicas, nós trazemos para a discussão alguns dados de natureza sintático-semântica que mostram certas estratégias de impessoalização e de reordenação dos argumentos nas sentenças do português brasileiro (NEGRÃO e VIOTTI, 2008, p. 2).

É importante reforçar que as autoras relatam que essas características são as principais diferenças existentes entre as duas variedades da língua.

Na definição das pesquisadoras, “construções impessoais são, tradicionalmente, consideradas aquelas que não têm um sujeito com conteúdo semântico (NEGRÃO e VIOTTI,

2008, p. 5). Contudo, além de apontar formas de impessoalização já discutidas pelos gramáticos tradicionais, Negrão e Viotti (2008, p. 6) apresentam outras formas de impessoalização que não são tratadas com frequência na literatura tradicional, pois são apenas consideradas construções resultantes da queda do clítico *se*, tais como:

- (4) Tem que comprar um computador pra sala de projetos
- (5) Tem-se que comprar um computador pra sala de jogos
- (6) Com a mudança do Lattes, precisa verificar tudo, item por item.
- (7) Com a mudança do Lattes, precisa-se verificar tudo, item por item.

Assim, segundo a explicação das autoras, as sentenças (4) e (6) não seriam completamente impessoais, porque elas teriam um sujeito semântico indeterminado que não foi realizado foneticamente, mas realizado em (5) e (7) com o clítico *se*. Negrão e Viotti (2008) ainda propõem um contínuo de impessoalização, que vai do nível mais impessoal ao menos impessoal. Assim, o contínuo teria a seguinte sequência: sentenças transitivas-ativas, sentenças passivas verbais analíticas, sentenças passivas verbais sintéticas/sentenças de sujeito indeterminado, sentenças impessoais, construídas com verbos de alternância causativa, sentenças passivas adjetivas, sentenças estativas com sujeito inanimado e, por fim, sentenças impessoais.

Luft (2008) descreve ainda mais formas de ocorrência do fenômeno da impessoalização na língua portuguesa. Segundo o autor, os principais casos de impessoalidade são:

- a) expressão de fenômenos meteorológicos: *chove, chuvia*;
- b) com o verbo *ser*: *é dia/noite, é uma hora*;
- c) com o verbo *passar* + *de*: *passa de meia noite*;
- d) eventualmente, com o verbo *parecer* (que, aliás, implica *ser*: *parece primavera, parecia noite*; e outros verbos de ligação (aspectos de *ser*): *ficou noite, foi ficando tarde*;
- e) nas indicações de tempo sinalizadas por instrumentos (sinos, relógios, cometas etc.), com os verbos *tanger, tocar, dar, bater* (acidentalmente impessoais, pois podem ter sujeito – sino, relógio etc): *bateu meia-noite, deu 8 horas*;
- f) na indicação de tempo decorrido, duração, com os verbos *haver* e *fazer*: *há/faz dez anos (que) não estuda*;
- g) com o verbo *ir* + *para* ou *em*: *vai para dez anos que ele partiu; vai em cinco meses que ela voltou*;

- h) em expressões de existência, acontecimento ou realização, como verbo *haver*: *há pessoas de bem entre nós; haverá jogos;*
- i) o verbo *ser* na expressão da narrativa inicial era uma vez: *era uma vez três porquinhos;*
- j) verbo *ir* + *bem/mal*: *mal vai a quem fazer isso;*
- k) na indicação de espaço e distância, com os verbos *ser* e *ir*: *quanto é?; vai daqui até à praia?;*
- l) na expressão de sensações, com os verbos *dor*, *comichar*, *coçar*, *cheirar*: *onde lhe dói? Dói-me as costas; cheira bem aqui;*
- m) nas locuções *está bem* ou *está mal*: *está bem, vamos adiante; está mal, assim não pode ser;*
- n) o verbo *dar* (importar) na locução *pouco se lhe dá de*: *pouco se lhe (me, nos etc.) dá das críticas;*
- o) os verbos *basta* e *chegar* + *de*: *basta (ou chega) de reclamações.*

Como é possível notar, Luft (2008) detalha melhor os casos de impessoalização do que outros gramáticos, como Cunha e Cintra (2008), por exemplo, contudo, todos eles reforçam a ausência de sujeito nos casos de impessoalização.

Ao se tratar de estudos linguísticos descritivos contemporâneos, tem-se os trabalhos de Siewierska e Papastathi (2011) e de Gast e van der Auwera (2013), que discorrem sobre as construções denominadas *R-impersonals* ou, mais recentemente, de *human impersonal pronouns*. Contudo, essas denominações não fazem referência ao que consideramos aqui como sujeito impessoal e, sim, sujeito indeterminado, pois ainda que em tradução literal, elas se assemelhem à impessoalização, neste estudo, o que se considera indeterminado é o referente extralinguístico, sendo, dessa forma, próximo ao que os autores tratam de *R-impersonals* ou *human impersonal pronouns*. Gast e van der Auwera (2013) explicam que essas construções são feitas tipicamente a partir das palavras ‘pessoa’ ou ‘pessoas’, o que se assemelha aos nossos casos de indeterminação, uma vez que, quando dizemos “*se a pessoa comer, ela ficará forte*”, tem-se um sujeito indeterminado e não impessoal.

Polachini (2015) tenta compreender a controvérsia dos termos *impessoal* e verbo *unipessoal* em gramáticas de várias línguas do século XIX e como isso pode ter influenciado nas gramáticas portuguesas. Sendo assim, inicialmente a autora constata que essa controvérsia é anterior ao século XIX, pois Sanchez ([1587] 1976, p. 77 *apud* POLACHINI, 2015) entendia que verbo deveria ter tempo, pessoa e número e que, portanto, apenas verbos no infinitivo poderiam ser considerados impessoais. Além disso, para o mesmo autor, toda sentença deve possuir nome e verbo e, dessa forma, segundo Polachini (2015, p. 58), “seria falso dizer que os

verbos impessoais não têm sujeito, dado que, para Sánchez, o sujeito estaria implícito no verbo”.

Nas gramáticas do português, a autora constata que, diferentemente de gramáticas de outras línguas, houve uma variação de uso dos termos *impessoal* e *unipessoal* por portugueses e brasileiros. Gramáticos, como Morais Silva (1806), Coruja (1835), Ribeiro (1881) e Maciel [1887] 1902), todos citados por Polachini (2015), preferem o uso do termo *impessoal*; já Sotero dos Reis (1866) e Freire da Silva ([1871] 1875) optam pelo termo *unipessoal*. Com isso, “observa-se que todos os autores falam sobre a defectividade do verbo impessoal ou unipessoal em sua definição, ressaltando que este só tem terceira pessoa (alguns incluem o plural, outros não)” (POLACHINI, 2015, p. 60). Ainda segundo a autora:

Ribeiro (1881) e Maciel ([1887] 3 1902) aludem ainda a relação entre esse verbo e o sujeito, sendo este para Ribeiro impessoal e para Maciel, indeterminado. Dessa nomenclatura, pode-se depreender que Ribeiro considera a presença não aparente do sujeito, visto que o verbo o encerra em si. Por fim, do nome sujeito indeterminado conferido por Maciel infere-se que esse verbo não tem sujeito, sobretudo por conta desse comentário: “o critério da impessoalidade se estatue pela falta do sujeito” (POLACHINI, 2015, p. 60).

Nessa análise feita pela autora, é possível inferir uma diferença já apresentada nas gramáticas antigas entre a indeterminação e a impessoalização: na impessoalização, há ausência de sujeito, o que não ocorre com a indeterminação e com a indefinição, ainda que elas sejam confundidas.

Por fim, a autora conclui que existe, então, uma diferença no uso dos termos dependendo da linha que os autores seguem, já que, para aqueles que não consideram a defectividade do verbo para a definição, o mais interessante é analisar o sujeito. Isso reforça, ao nosso ver, a confusão causada pelos termos *indefinição* e *indeterminação*.

Widera (2019) propõe uma discussão sobre o uso de *ele* em construções impessoais no português europeu. A autora relata que o português europeu é considerado uma língua de sujeito nulo, contudo, na variedade não padrão, é possível observar o uso de *ele* em construções impessoais. Os exemplos dados por Widera (2019, p. 21) são:

(8) Ele choveu toda a noite!

(9) Ele há cada uma!

A autora argumenta que *ele* não se trataria de um pronome expletivo, mas de um elemento discursivo. O estudo ainda se encontra em estágio experimental, mas a pesquisadora

conclui que *ele* tem um valor expressivo para a frase e que o pronome se distingue dos sujeitos expletivos por não ser usado de forma uniforme e nem obrigatória.

Após os estudos apresentados, pode-se notar que, apesar de a maioria das gramáticas tradicionais apresentarem um grupo restrito de formas que podem tornar o sujeito impessoal, estudos mais recentes têm demonstrado a dificuldade de precisar o que é ou não impessoal, quando se considera o contexto, por exemplo. Sendo assim, para este trabalho, foram excluídas as formas consideradas impessoais pela tradição gramatical, tendo em vista que o nosso foco são as orações com sujeito indeterminado.

2.3.2 *Indefinição*

Milanez (1982) demonstra que a indeterminação é também, muitas vezes, confundida com a indefinição, ainda que sejam consideradas distintas pela análise sintática. Milanez (1982) mostra que a diferença está no fato de que, quando dizemos “Alguém bateu à porta”, o sujeito é determinado, ainda que indefinido. A indefinição pressupõe um conjunto fechado de elementos e remete à noção de quantificação. Dessa forma, Milanez (1982, p. 39) exemplifica que em (10) tem-se um exemplo de indeterminação e em (10a) de indefinição, tal como ilustrado abaixo:

(10) “Lá no Norte come-se fruta no café da manhã” (MILANEZ, 1982, p. 39)⁵.

(10a) “Lá no Norte, todos comem frutas no café da manhã” (MILANEZ, 1982, p. 39).

Baseando-se em Milanez (1982), Carvalho (2010) aponta alguns aspectos que servem para diferenciar a indeterminação da indefinição; a indeterminação restringe-se a seres humanos; refere-se a qualquer pessoa do discurso, ainda que sua marcação seja uma das três pessoas do discurso; possui um caráter generalizador como uma de suas condições primárias; considera irrelevante a quantificação; depende do contexto para ser interpretada e não identifica de modo exato as pessoas do discurso. Contrariamente à indeterminação, a indefinição refere-se a seres humanos e não humanos; não remete a outras pessoas do discurso, apenas à 3ª pessoa; toma a generalização como uma possibilidade; considera relevante a quantificação; pode ser

⁵ Gramáticas como a de Cunha e Cintra (2008) consideram que o *se*, no contexto mencionado, seria uma partícula apassivadora.

recuperada no nível frasal e é determinada, visto que uma das pessoas do discurso pode ser recuperada.

Teixeira (2014) busca, em seu trabalho, demonstrar que a indeterminação é um processo pragmático e não semântico, como muito se tem feito hoje. Sendo assim, a autora propõe uma distinção entre a indeterminação e a indefinição:

Assim sendo, neste trabalho, propõe-se que o termo **Indeterminação**, entendido como estratégia argumentativa do Falante, indica o conhecimento (e a intencionalidade de não revelar um referente) ou o não conhecimento de um de um referente por parte de um ou de ambos os interlocutores do discurso. À falta de designação e à presença de genericidade, dá-se o nome de **Indefinição**. (TEIXEIRA, 2014, p. 60).

Além disso, a autora ainda assume três pontos relevantes, isto é, que a indefinição é formulada no nível representacional (genericidade e papel semântico), que a indeterminação é formulada no nível interpessoal (identificabilidade e especificidade) e que ambas são codificadas por diferentes formas no nível morfossintático (estrutura e forma de indeterminação).

Onofre e Facundes (2021) objetivam explicar a noção de indefinição tendo em vista as várias abordagens linguísticas. Inicialmente, as autoras fazem questionamentos sobre o que seria a indefinição e o que seria a indeterminação. Para endossar a discussão, são apontadas gramáticas prescritivas, descritivas e interacionistas. Finalmente, Onofre e Facundes (2021) chegam ao ponto chave do estudo, isto é, a Teoria das Operações Enunciativas e Predicativas. Por meio dessa teoria, tem-se a seguinte noção de indefinição:

A marca da noção de indefinição é uma operação de construção enunciativa no ato do dizer que envolve os valores referenciais ali presentes em cotextos sempre enunciativos. As análises são pautadas na perspectiva da enunciação em que temos a preocupação em explicar o processo de produção e reconhecimento dos enunciados como marcas de indefinição na interação textual que envolve o sujeito enunciador e o sujeito coenunciador. Neste ato de construção de enunciação está presente o que define Culioli como Sit (Situação de enunciação) em relações que marcam articulação com as categorias de tempo e de espaço e o léxico, constituindo o sistema de relações que envolvem o conceito de noção, de domínio nocional e as operações predicativas e enunciativas que Culioli chamou de operação de *repérage*. (ONOFRE e FACUNDES, 2021, p. 59).

Ademais, ainda que as autoras tratem a questão da indefinição considerando os artigos definidos e indefinidos, a partir da definição apresentada por elas, tem-se que na indefinição é possível recuperar referências por meio de um contexto. É importante destacar que Milanez (1982) já apontava algo semelhante, quando pontuou que a indefinição pode ser recuperada até mesmo no nível frasal, diferentemente da indeterminação.

Oliveira e Ramos (2021), sob uma ótica variacionista, propõem um estudo sobre os pronomes de 1ª e 2ª pessoas na representação do sujeito. Para isso definem como *corpus* textos de natureza jornalística, que possuem a modalidade formal da língua e que foram retirados das revistas *Cláudia*, *Caras*, *Veja*, *Época* e *Isto é*.

A partir dessa coleta de dados, as autoras buscaram “refletir sobre as implicações das mudanças em curso no sistema pronominal brasileiro, tanto em relação ao preenchimento ou apagamento do sujeito na estrutura linguística, quanto em relação ao processo de referência (definição ou indefinição do agente)” (OLIVEIRA; RAMOS, 2021, p. 1).

Vale ressaltar que, nesse trabalho, as autoras utilizam indistintamente a indefinição e a indeterminação como sendo o mesmo processo. Sendo assim, em alguns momentos, elas optam pelo termo indeterminação e outras vezes pelo termo indefinição, reforçando a confusão no uso das noções em trabalhos até mesmo recentes. Salienta-se que as autoras pontuam que até mesmo gramáticos utilizam esses termos indistintamente.

Após esta apresentação das propriedades da indefinição e de como ela muitas vezes se mistura com a indeterminação até mesmo em trabalhos recentes, a próxima seção trata da indeterminação do sujeito e dos resultados de pesquisas que discorrem sobre o mesmo tema deste estudo. Vale lembrar que contextos de indefinição também não serão analisados neste trabalho e que aqui se entende a indefinição assim como em Milanez (1982), isto é, que, em casos de indefinição, sempre será possível recuperar um referente e que essa recuperação se dá até mesmo no nível frasal.

2.3.3 Indeterminação do sujeito

A indeterminação do sujeito, tema deste trabalho, foi tratada por vários estudos em diversas regiões do país e, comprovadamente, a partir dessas pesquisas, é possível observar que existem muito mais formas de se indeterminar o sujeito do que as gramáticas tradicionais atestam. Cunha e Cintra (2008) apontam que existem duas formas de se indeterminar o sujeito em português, sendo elas: verbo na terceira pessoa do plural e verbo na terceira pessoa do singular mais a partícula *se*, conforme exemplos (11) e (12).

(11) **Reputavam**⁶-no o maior comilão da cidade (CUNHA; CINTRA, 2008, p. 142).

(12) Ainda **se vivia** num mundo de certezas (CUNHA; CINTRA, 2008, p. 142).

⁶ Grifos da autora.

Neves (2000), entretanto, em uma perspectiva descritiva, apresenta várias outras maneiras de indeterminar o sujeito em português. A autora cita o pronome *eles*, verbos na terceira do plural, verbos na terceira pessoa do singular, verbos na terceira pessoa do singular mais o pronome *se*, o pronome *você*, o pronome *eu*, a primeira pessoa do plural, ou seja, *nós* e, ainda, a forma *a gente* e sintagmas nominais como *a pessoa*, *o cara* e *o pessoal*, estes últimos, segundo a autora, utilizados mais comumente na linguagem popular. Alguns exemplos citados por Neves (2000) são:

(13) Nessas horas **a gente** não pensa em nada, perde a cabeça (NEVES, 2000, p. 469).

(14) Eu gosto de você, mas **o pessoal** fala que você é meio bruta (NEVES, 2000, p. 470).

A autora destaca que a forma *eles* possui uma indeterminação parcial por se referir ao universo somente das terceiras pessoas e ainda que a forma *nós* também não indetermina totalmente porque sempre haverá, pelo menos, uma referência determinada, o *eu*. No caso de *a gente*, Neves (2000, p. 470) acrescenta que essa forma “sempre deixa indicado o envolvimento da primeira pessoa no conjunto”. Contudo, estudos sociolinguísticos que serão apresentados a seguir demonstram que, em alguns casos, a forma *a gente*, não inclui a primeira pessoa.

Autores como Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Oliveira (2006), Pontes (2008), Carvalho (2010), Souza (2014), Barbosa (2016) e Oliveira (2018) fizeram pesquisas, com dados de língua oral, buscando demonstrar os diversos recursos de indeterminação que a população brasileira tem utilizado. Milanez (1982) pesquisa a fala culta paulistana a partir de dados do projeto NURC. Os itens analisados pela autora são as expressões denominadas de itens lexicais que normalmente ocupam posição de sujeito (*a gente*, *você*, *eles*, *eu*, *nós*, *o/um indivíduo*, *o/um sujeito*, *a/uma pessoa*, *o/um cara* e *a/uma pessoa*), as constituídas por formas verbais despronominalizadas ((\emptyset + 3ª p. sing.), (\emptyset + 3ª p. pl.)), a construção com *se*, ou seja, (\emptyset + VERBO + *se*) e (\emptyset + infinitivo). A pesquisadora ainda reforça que formas, como *eu*, *você*, *indivíduo*, *cara*, *pessoa*, também foram consideradas formas de indeterminação devido à seguinte explicação:

o mecanismo de indeterminação funciona como uma projeção de um dos elementos do nível da determinação ao da indeterminação, sendo o envolvimento da 1ª pessoa, 2ª e 3ª meramente hipotético, ou seja, as mesmas seriam usadas para efeito de

ilustração de uma situação onde qualquer pessoa poderia se encontrar (MILANEZ, 1982, p. 31).

Ao tratar de sujeito, Milanez (1982) mostra que não está se referindo apenas a uma noção puramente sintática, ou seja, à posição que ele ocupa na frase, mas, sim, a uma referência que ele possui. Além disso, ela elucida que somente as formas pronominais *a gente*, *ocê*, *eu*, *eles* se prestam a indeterminar outros elementos da frase. Os sintagmas nominais *o sujeito*, *o indivíduo* e *uma pessoa* parecem esvaziar-se de seus efeitos de indeterminação em outras posições.

Menon (1994), a partir de dados do NURC/SP, realiza um estudo que serviu de referência para muitos trabalhos que vieram posteriormente. Com base na sociolinguística variacionista, selecionou 12 variantes para serem estudadas, sendo elas *a gente*, *eles*, *eu*, *formas nominais*, *nós*, *se*, *ocê*, *ocês*, *voz passiva sem agente*, *voz passiva sintética*, *verbo na terceira pessoa do singular* e *verbo na terceira pessoa do plural*. Vale ressaltar que a autora entende as formas nominais como aquelas constituídas de um artigo definido mais um substantivo, as quais servem como formas de generalização. Os exemplos encontrados pela autora em dados do projeto NURC/SP foram *o camarada*, *o homem*, *o indivíduo*, *a pessoa*, *o sujeito*, *a turma*, *o cara*, *o cidadão* e *o público*. No caso de *a gente*, Menon (1996) demonstra o processo de gramaticalização desse item e salienta que o mesmo processo pode ter acontecido também com as formas nominais. Sobre isso, a autora afirma que

essas expressões teriam se cristalizado e assumido um caráter pronominal, visto que não poderiam (principalmente no caso de *a gente*) mais sofrer expansão ou concordância ou receber adjetivação ou determinação sob perda do significado específico que assumem tais formas ao serem utilizadas para fazer referência a um sujeito indeterminado (MENON, 1996, p. 623).

Setti (1997) e Godoy (1999) fizeram uma análise sobre as formas de indeterminação do sujeito a partir de dados do VARSUL. A primeira trabalha com dados das capitais Curitiba, Florianópolis e Porto Alegre e a segunda com cidades as cidades Irati, Londrina e Pato Branco, todas localizadas no interior do Paraná. As variantes estudadas como indeterminadoras do sujeito são *a gente*, *eles*, *eu*, *formas nominais*, *nós*, *se*, *tu*, *ocês(s)*, *voz passiva sem agente*, *voz passiva sintética*, *verbo na 3ª pessoa do singular* e *verbo na 3ª pessoa do plural*.

Oliveira (2006) fez uma pesquisa com o intuito de observar as formas pronominais *nós* e *a gente* e o clítico *se* seguido de verbo no infinitivo no português brasileiro e no português europeu. A autora relata que, para o *corpus* do português brasileiro, foram coletadas 45 entrevistas, de uma hora cada, de fala espontânea com falantes de Uberlândia (MG), utilizando-se de um questionário guia. Já o *corpus* do português europeu foi retirado de entrevistas do

Castro Laboreiro (Outeiro, Perafita), do Baixo-minhotos-durienses-beirões (Vila Boa de buços, Granjal, Figueiró), do Baixo-Minho e Douro Litoral (Vila Praia de Ancora) e de regiões meridionais (Golpilhal, Moita do Martinho, Parreira, Alcochete, Lavre, Serpa, Sagres, Porches, Alte), todas publicadas pelo Instituto Camões, e entrevistas realizadas pelo projeto Português Fundamental com informantes do Porto, Faro, Lisboa, Braga, Évora, Aveiro, São Miguel, Guarda, Bragança, Portalegre, Vila real, Castelo Branco, Coimbra, Santarém, Beja, Viseu, Setúbal, Funchal, Terceira, Faial, Leiria, Viana do Castelo e Serpa.

Posteriormente, foi feita a apresentação de como foram coletados os dados e selecionaram-se os fatores externos e internos utilizados no trabalho. É válido ressaltar que, nas amostras do português europeu, não foram computados os fatores sociais, visto que o *corpus* não fornece informações desse tipo.

Dessa maneira, para a análise, foram elencados os seguintes fatores: variável dependente (*nós*, *a gente* e clítico *se*); variedades do português (português de Portugal e português do Brasil); tipo de sentença (finita e infinitiva); paralelismo (presença e ausência); presença/ausência de modalizadores e/ou orações performáticas; presença/ausência de preposições; número de argumentos (um ou dois ou mais argumentos); classe social (alta, média e baixa) e faixa etária (20 a 30 anos, 31 a 45 anos, acima de 45 anos). Vale ressaltar que, segundo a autora, a maioria dos fatores sociais e linguísticos desta pesquisa estão em Lopes (1988).

Assim, a autora conclui que *nós*, *a gente* e *se + infinitivo* são estratégias empregadas tanto no português brasileiro quanto no português europeu, mas que a frequência em que essas estratégias ocorrem são diferentes. Enquanto, no português brasileiro, há preferência por *a gente*, no português europeu, as estratégias *nós* e *a gente* são igualmente favorecidas.

No que se refere ao *se + infinitivo* (*Para se ter sucesso na vida, é preciso persistência*)⁷, era esperado que, no português brasileiro, houvesse um favorecimento desta estratégia, o que não aconteceu. Verificou-se “que embora seja necessário o *se* junto ao infinitivo no PB para marcar a indeterminação do sujeito, quando comparada a outras estratégias pronominais, essa estratégia é menos favorecida” (OLIVEIRA, 2006, p. 151).

A partir dos resultados mostrados, observa-se que as formas pronominais com as quais as gramáticas tradicionais não trabalham são as mais utilizadas pelos falantes quando se indetermina o sujeito, tanto no PB quanto no PE.

Ponte (2008) discorre sobre as formas de indeterminação no português oral do interior da Bahia. O trabalho, que trata de comunidades africanas, tem como objetivo compreender as

⁷ Exemplo retirado de Oliveira (2006, p. 54).

influências dessas línguas no português e, como comunidades afastadas, algumas delas formadas de antigos quilombos, utilizam o PPB (Português Popular Brasileiro). As comunidades pesquisadas são Helvécia, Rio de Contas, Sapé, Cinzendo e a cidade de Santo Antônio de Jesus.

No que se refere às variantes, a autora apresenta as seguintes formas: *a gente, nós, você, eles, verbo na terceira pessoa do plural sem sujeito explícito, verbo na terceira pessoa do singular sem sujeito explícito* e o *se*. Este último só foi analisado qualitativamente devido à pouca frequência. Os fatores linguísticos selecionados foram: realização fonética do sujeito indeterminado; desinência verbal; tipo de verbo; tipo de frase; nível de referencialidade; modo verbal; inclusão do falante e forma antecedente. Já as variantes sociais escolhidas foram: a localidade do falante; gênero; faixa etária; estada fora da comunidade e escolaridade. O trabalho pertence ao “Projeto Vertentes do Português Rural do Estado da Bahia, que tem se dedicado à constituição de um acervo da fala do Português Rural do Estado da Bahia, em meio digital, e à realização de análises linguísticas das diversas variedades da língua falada no interior do Estado” (PONTE, 2008, p. 60).

Após a análise dos dados, a pesquisadora aponta algumas conclusões relevantes. Segundo ela, a estratégia mais utilizada foi a forma *a gente*, seguida de *verbo na terceira pessoa do singular sem sujeito explícito*. Além disso, a variável linguística mais selecionada foi *forma antecedente*, mostrando, assim, um paralelismo formal. No que diz respeito à variável social, a *localidade do informante* foi a mais selecionada pelos itens pesquisados. Assim como em outras pesquisas, as mulheres se mostraram mais conservadoras do que os homens. Nesse sentido, o trabalho buscou compreender como a língua muda e como as comunidades interioranas têm se adequando a essas mudanças.

Carvalho (2010), ao pesquisar em Salvador as formas de indeterminação do sujeito no português oral, teve como base metodológica a Teoria Variacionista. As formas selecionadas no *corpus* são as previstas pela gramática, tais como *verbo na terceira pessoa do plural* ($\emptyset + V3PP$), *verbo na terceira pessoa do singular mais o pronome “se”* ($\emptyset + V3PS + SE$), *verbo no infinitivo pessoal* ($\emptyset + VINF$) e outras estratégias como *você, a gente, nós, eles, eu, voz passiva sem agente (VPSA), voz passiva sintética (VPASSINT)* e as *formas nominais* (*o cara, o indivíduo, o sujeito, nego, o pessoal, as pessoas, o povo, o público, o homem e gente*). Constatou-se que em Salvador as formas *você, a gente* e as *formas nominais* foram as mais utilizadas pelos falantes que compõem os *corpora*.

Souza (2014) pesquisa a indeterminação do sujeito em duas comunidades rurais do interior da Bahia (Matinha e Casinha). A primeira está localizada no município de Feira de Santana e a segunda, no município de Jeremoabo. A constituição do *corpus* se deu a partir de 24 entrevistas obtidas por meio do projeto denominado *A língua portuguesa falada no semi-árido Baiano*, da Universidade Estadual de Feira de Santana.

As variantes selecionadas para esse estudo foram *a gente, nós, você, eles, verbo na terceira pessoa do plural, verbo na terceira pessoa do singular, verbo + se e formas nominais*. Como se trata de um estudo variacionista, as variáveis extralinguísticas analisadas foram sexo/gênero; escolaridade; faixa etária e localidade. No que diz respeito às variáveis linguísticas, discutiu-se tipo de verbo; tempo/modo verbal; tipo de oração e paralelismo.

Os resultados obtidos mostraram favorecimento das formas nominais no conjunto das variantes. De acordo com Souza (2014, p. 139),

Tomando como base a frequência, nota-se que o pronome *a gente* (41%) e o $\emptyset+V3PS$ (28%) foram os recursos mais usados para indeterminar o sujeito nas duas comunidades. Por sua vez, as formas nominais (que representam 17% dos dados) foram um recurso bem produtivo, sendo a terceira variante mais usada. Registramos várias formas nominais que também foram encontradas em outras pesquisas, como *o cara, o camarada, o homem, a pessoa, o sujeito, a turma, o pessoal e nego* (esta última registrada apenas na pesquisa de Carvalho, 2010) bem como houve ocorrências da forma nominal *o cabra*, que foi registrada somente entre os informantes de Casinhas e que, até então, não havia sido encontrada em outros trabalhos sobre a indeterminação do sujeito.

Ao tratar dos resultados estatísticos, a autora argumenta que as variantes preferidas pelos falantes das comunidades Matinha e Casinhas são o pronome *a gente* (41%) e o *verbo na terceira pessoa do singular* (28%). As formas nominais, relevantes para este trabalho, foram favorecidas pelos verbos de ligação e intransitivos, pelos informantes do sexo masculino, pelos informantes de Casinhas, nos contextos em que essa estratégia ocorre isoladamente na sentença ou depois de forma diferente, entre os mais velhos, com verbos no futuro do subjuntivo, no presente do subjuntivo e no pretérito, conforme dados obtidos por Souza (2014).

Barbosa (2016) encontrou no falar ludovicense dez formas que indeterminam o sujeito. São elas *eu, tu, você, a gente, nós, eles, formas nominais, vocês, se e verbo na terceira pessoa do plural*. No entanto, o trabalho da autora não corrobora estudos, como o de Carvalho (2010), uma vez que a escolaridade mostrou ser um fator de relevância na pesquisa de Barbosa. Entretanto, salienta-se que a autora utilizou uma divisão diferenciada dos demais estudos realizados até o momento, já que seus informantes foram divididos em até o quinto ano, até o nono ano, ensino médio completo e ensino superior completo, enquanto, em Carvalho (2010) e em outros trabalhos que vieram posteriormente, como o de Oliveira (2018), a divisão se deu

apenas em ensino fundamental, ensino médio e ensino superior, estando eles completos ou não. Apesar dessa diferença, assim como na maioria dos trabalhos realizados sobre o tema, as formas nominais ocuparam a terceira posição em número de ocorrências, reforçando, mais uma vez, a tese de que essas formas ocupam uma posição estável independentemente da localidade da pesquisa.

Posio (2016), em seu texto sobre as estratégias de indeterminação no espanhol e no português europeu, propõe comparar estruturas como pronomes e formas verbais de segunda pessoa. Ele destaca que, nos estudos sobre a língua espanhola, têm-se focado muito sobre estudar a variação entre o *tu* indeterminado e o pronome indefinido-impessoal *uno*. Já no caso do português europeu, segundo o autor, parece haver menos alternativas para se fazer uso da indeterminação, uma vez que não há nada equivalente ao pronome *uno* e o *tu* não parece tão frequente se comparado ao espanhol. O português conta, contudo, com uma expressão que parece desempenhar função indeterminadora, sendo ela *a/uma pessoa*. Posio (2016) salienta, ainda, que a gramaticalização de sintagmas nominais é um fenômeno frequente no português, conforme pode ser observado em *a gente*.

Os dados⁸ coletados apresentam importantes resultados, tais como de que o *tu* indeterminado é muito mais frequente no espanhol do que o pronome *uno*, sendo este último usado, metade das vezes, por um único informante de 72 anos. Quando comparado ao espanhol, o português apresenta cerca de cinco vezes menos o *tu* como forma de indeterminação. Ressalta-se que todos os participantes do espanhol utilizam o *tu* como forma de indeterminação, diferentemente do português em que sete participantes não utilizam tal forma. Além disso, no espanhol, as formas *a pessoa* e *uma pessoa* são pouco frequentes em comparação ao *tu* indeterminador.

Tendo isso em vista, é possível concluir que a frequência de uso de construção de indeterminação é menor no *corpus* português do que em espanhol, o que indica que os participantes portugueses empregam outras construções em contextos em que os participantes espanhóis escolheriam o *tu* para indeterminar.

Além da diferença quantitativa existente, também há uma diferença qualitativa. Em espanhol, existem dois tipos proeminentes do pronome *tu*; ele pode ser utilizado para estabelecer uma referência genérica a qualquer pessoa em uma determinada situação ou ainda

⁸ Os dados do trabalho de Posio (2016) foram retirados do *Corpus de habla culta de Salamanca* e de gravações realizadas pelo autor na cidade do Porto entre maio e agosto de 2014 com a ajuda de estudantes de linguística.

para casos em que o falante não quer ser descoberto, como, por exemplo, ao contar experiências do passado. No caso do português, o uso do pronome só foi encontrado em situações genéricas.

Ao pontuar sobre as ocorrências de *a pessoa/uma pessoa* no *corpus* do português, Posio (2016) argumenta que os dados não permitem fazer uma distinção funcional entre *a pessoa* e *uma pessoa*, já que ambas as formas se prestam a leituras genéricas e direcionadas ao informante. Assim, a escolha de uma forma ou de outra parece ser uma questão de variação idioletal. Ainda conforme o pesquisador, o item *a pessoa/uma pessoa* demonstra indícios do processo de gramaticalização, pois apresenta uma ampliação nos contextos de uso, redução fonético/fonológica etc. Além disso, Posio (2016) propõe que o pouco uso da forma *a gente* pelos falantes do *corpus* pode estar ligado ao nível de escolarização dos participantes, que é superior completo ou incompleto.

Assim, de forma geral, Posio (2016) conclui que, enquanto os participantes da pesquisa feita em Salamanca tendem a usar a segunda pessoa do singular como estratégia referencial para falar sobre suas próprias experiências, em uma perspectiva geral, os participantes portugueses são mais propensos a usar a primeira pessoa do plural em contextos semelhantes. Além da diferença quantitativa, o uso de segunda pessoa do singular impessoal também é qualitativamente mais limitado no *corpus* português, uma vez que possui apenas usos genéricos e não ocorre em contextos narrativos, diferentemente do *corpus* espanhol. Além disso, as formas *a/uma pessoa* podem ser usadas em situações que não ocorrem no espanhol. De qualquer forma, pode-se concluir que a expressão de uma referência indeterminada individualizadora é mais frequente nos dados espanhóis do que nos dados do português.

Oliveira (2018) realizou um trabalho com dados da cidade de Bambuí, no estado de Minas Gerais e constatou que os nomes gerais, foco da pesquisa, foram selecionados em terceiro lugar como relevantes dentre as variantes estudadas. Tal resultado corrobora o que foi encontrado por Setti (1997) na cidade de Florianópolis, Carvalho (2010) em Salvador, Souza (2014) em comunidades no interior da Bahia e Barbosa (2016) no falar de São Luís-MA. Esse resultado indica como esses itens estão estáveis na língua e tendem a continuar, uma vez que, independentemente da região, eles parecem ser bastante usados quando se indetermina o sujeito. Ainda segundo Oliveira (2018), *povo* foi o item mais recorrente deste grupo. Além disso, na rodada realizada no programa GoldVarb X, que teve como objetivo observar a relevância dos nomes gerais no conjunto das variantes, cinco fatores foram selecionados como favorecedores desses nomes. São eles presença ou ausência de anáfora, grau de indeterminação, tempo verbal, tipo de oração e sexo. Apresenta-se, a seguir, um quadro com esses dados:

Quadro 2 – Resumo dos resultados obtidos em Oliveira (2018)

Fator	Divisão dos fatores selecionados	Número de ocorrências	Porcentagem (%)	Peso relativo⁹
Presença ou ausência de anáfora	Ausência	174/949	18,3	0,62
	Presença	17/408	4,2	0,25
Grau de indeterminação	Indeterminação total	59/426	13,8	0,54
	Parcial com referente implícito	112/673	16,6	0,53
	Parcial com referente explícito	20/258	7,8	0,37
Tempo verbal	Pretérito imperfeito	23/95	24,2	0,70
	Presente habitual	22/97	22,7	0,65
	Presente contínuo	63/359	17,5	0,62
	Futuro do pretérito	2/11	18,2	0,61
	Pretérito perfeito	23/113	20,4	0,60
	Presente do subjuntivo	3/21	14,3	0,52
	Futuro do subjuntivo	6/54	11,1	0,48
	Presente atemporal	37/424	8,7	0,38
	Perífrase de futuro	5/65	7,7	0,36
	Infinitivo	5/67	7,5	0,34
	Presente histórico	2/46	4,3	0,23

⁹ “Os pesos relativos são valores que vão de zero a um e que indicam matematicamente o peso com que um fator (linguístico ou extralinguístico) influencia o uso de uma variante, em relação a todos os fatores levados em conta na observação de um fenômeno de variação linguística. Quando o peso relativo de um fator é próximo de zero, significa que tal fator desfavorece o uso da variante. Quando o peso relativo é igual a 0,50, significa que ele não está correlacionado ao uso da variante – tal valor é, pois, o ponto denominado neutro. Finalmente, quanto mais próximo for de 1 (um), maior será o peso com que o fator favorece o uso da variante” (BELINE, 2003, p. 132).

Tipo de oração	Absoluta	42/222	18,9	0,58
	Subordinada	76/535	14,2	0,54
	Principal	67/465	14,4	0,50
	Coordenada	6/135	4,4	0,26
Sexo	Feminino	96/572	16,8	0,55
	Masculino	95/785	12,1	0,46

Fonte: Adaptado de Oliveira (2018).

A partir desse breve resumo, nota-se como a indeterminação do sujeito tem se mostrado um conjunto com muito mais variantes do que as formas apontadas pelas gramáticas. Além disso, buscou-se evidenciar que, neste estudo, entende-se que a indeterminação se dá quando não é possível recuperar o referente por meio do contexto ou de outros recursos disponíveis na fala. Dessa maneira, acredita-se que um estudo entre as variedades linguísticas do Brasil e de Portugal trará resultados interessantes sobre a realização do fenômeno da indeterminação do sujeito.

2.4 Nomes gerais

O estudo dos nomes gerais, para este trabalho, terá como base as discussões de Halliday e Hasan (1995 [1976]), Koch (2004), Amaral e Ramos (2014), Cappeau e Schnedecker (2014), Amaral (2017), Mihatsch (2017), Amaral e Mihatsch (2019) e Posio (2021). Todos esses tratam de nomes gerais com traço [+humano]. Trabalhos, como o de Fronek (1982), Mihatsch (2002), Mihatsch (2006) e Ramos (2013), discorrem sobre nomes gerais com traço [-humano] e, por isso, não fazem parte da bibliografia.

Halliday e Hasan (1995 [1976]) são os precursores sobre os estudos dos nomes gerais, ou, como denominado por eles, *general nouns*. Os autores explicitam que esses itens são um caso limítrofe entre um item lexical e um item gramatical e possuem traços mínimos de significado. Assim, Halliday e Hasan (1995 [1976]) utilizam-se dos seguintes exemplos para demonstrar a classe desses nomes: *people, person, man, woman, child, boy, girl* [humano], *creature* [não humano inanimado], *thing, object* [inanimado concreto], *stuff* [inanimado concreto contínuo], *business, affair, matter* [inanimado abstrato], *move* [ação], *place* [lugar], *question e idea* [fato].

Ao tratar da função coesiva dos nomes gerais, os autores ressaltam que, frequentemente, eles vêm acompanhados por *the*, sendo que o sintagma [*the* + nome geral] funcionaria como um item referencial anafórico. Essa junção de artigo mais um nome também foi citada por Menon (1994) ao definir o que seriam as formas nominais. Além disso, como eles estão na fronteira entre um item lexical e um item gramatical, do ponto de vista semântico, eles seriam membros superordenados de grandes conjuntos lexicais e, de um ponto de vista gramatical, a combinação desses itens com um determinante resultaria em algo semelhante a um item referencial.

Koch (2004), ao discutir sobre as diferentes formas de se retomar uma anáfora, salienta que os nomes genéricos também se enquadram nesse estudo, pois “*coisa, pessoa, negócio, criatura, indivíduo*” (KOCH, 2004, p. 250) podem desempenhar esse papel. A variedade regional ou social do falante também é muito importante para a seleção dos itens a serem usados, uma vez que *trem*, por exemplo, marcaria, provavelmente, o dialeto mineiro.

Segundo constatações da autora, o uso desses termos é muito comum na língua falada, mesmo quando se trata de falantes cultos. Koch (2004, p. 250) ainda explica que “na fala, em que planejamento e verbalização são quase simultâneos, a busca de um termo mais específico teria maior custo processual, de modo que se torna mais fácil recorrer a um termo imediatamente acessível”.

Amaral e Ramos (2014) discorrem sobre o comportamento morfossintático, semântico-textual e sociolinguístico de quatro nomes gerais (*coisa, negócio, trem, pessoa*) e salientam que eles podem vir precedidos ou não por determinantes. Nesses casos, eles serão elementos dêiticos ou anafóricos, e, portanto, poderão ser parafraseados por pronomes. Os autores destacam ainda que esses itens possuem outras funções e, ao citar Mihatsch (2006), elucidam que um nome geral é utilizado quando não há uma designação, naquele momento, acessível ao falante. Outra função é a de que eles podem substituir um referente que seja incômodo para aquele que fala.

Cappeau e Schnedecker (2014), ao analisarem formas como *gente* (fr. *gens*), *pessoa(s)* (fr. *persone (s)*) e *indivíduo(s)* (fr. *individu(s)*) na língua francesa, demonstram que *gens* não permite a individualização de componentes, além de mostrar uma tendência à pronominalização. Ainda segundo os autores, ao citarem Schnedecker (2012), um item como *gens* é homogêneo e heterogêneo; homogêneo por causa dos elementos humanos que compõem o grupo e heterogêneo no sentido de que esses elementos parecem pertencer a diferentes subcategorias. Algo semelhante parece ocorrer com o item *população*, no que diz respeito a ser

um conjunto homogêneo em seu sentido mais amplo, seja uma população constituída por animais, pessoas etc., e heterogêneo pela diversidade dos elementos que podem pertencer a este grupo dentro de subcategorias.

Amaral (2017), em um estudo contrastivo sobre o português e o espanhol, demonstra que existe certa diferença de aceitabilidade de alguns nomes gerais nessas línguas. Além disso, sua maior contribuição para esta tese é observar como o item *pessoa* possui uma situação diferente dos demais nomes explorados e como seu uso enquanto forma de indeterminação do sujeito é comum. Ainda segundo o autor, nomes gerais, como *ser humano* e *indivíduo*, são mais comuns em textos acadêmicos, e *sujeito* e *pessoa* são mais frequentes em registros ficcionais e orais. Tendo isso em vista, essa constatação pode ser corroborada no estudo de Oliveira (2018), visto que não foi encontrado, em seu *corpus*, a forma *indivíduo*, enquanto *ser humano* ocorreu raríssimas vezes.

Mihatsch (2017) apresenta um estudo sobre a gramaticalização dos nomes gerais, isto é, nomes que apresentam uma função tipicamente gramatical, com dados do alemão, espanhol, francês e português. A pesquisadora realiza um estudo de itens com traços [+humano] e salienta que existe um grupo que contém nomes bastante eruditos e outro que contempla itens mais usuais da língua, como o caso de *pessoa*. A autora conclui que *pessoa*, por exemplo, já mostra indícios de pronominalização. Além disso, seu trabalho demonstrou que, do ponto de vista lógico, os nomes gerais com traço [+humano] encontram-se abaixo de itens como *chose* (*coisa*). Nomes como *chose*, segundo a autora, não pertencem mais ao léxico e são utilizados sem nenhum tipo de referência autônoma.

Amaral e Mihatsch (2019) propõem um estudo sobre itens lexicais (*pessoa*, *pessoal* e *povo*) que estão se transformando em pronomes impessoais. Além disso, os pesquisadores ainda observam se o processo de gramaticalização desses nomes tem seguido caminhos conhecidos. A metodologia utilizada é composta por amostras de entrevistas sociolinguísticas coletadas em diferentes cidades do estado de Minas Gerais, amostras do Projeto NURC-RJ, com foco no PB falado padrão do Rio de Janeiro (NURCRJ), além do *Corpus* do Português, CORAL-ROM e C-ORAL-BRASIL. Os autores concluem que é possível distinguir dois grupos dentro desse conjunto de expressões. O primeiro deles diz respeito aos nomes coletivos *o pessoal* e *o povo* e o plural *as pessoas*; o segundo, ao item *a pessoa/uma pessoa*. Assim, o primeiro grupo dá indícios de que segue o mesmo caminho dos impessoais de terceira pessoa, exceto *as pessoas*. No caso de *a pessoa/uma pessoa*, segundo os autores, eles não são comumente usados em

contextos genéricos, ainda que em português seu uso seja mais aceitável do que em outras línguas.

Posio (2021) tem como objetivo esclarecer as propriedades semânticas, pragmáticas e morfossintáticas das construções de *pessoa* no português europeu e discutir seu processo de gramaticalização. O autor ainda demonstra que construções como *a pessoa* e *uma pessoa* são favorecidas porque no PE prefere-se evitar uma referência direta à pessoa e porque existe uma preferência por sujeitos expressos mesmo em contextos que permitem sujeitos nulos. Além disso, nesse processo de análise, tendo em vista a gramaticalização dessas construções, é proposta uma análise para avaliar até que ponto elas são semelhantes ao impessoal *homem* encontrado em outras línguas.

Após Posio (2021) apresentar uma breve discussão sobre o termo impessoal para a referência humana, ele passa a tratar de uma fonte de construções impessoais humanas pouco discutida, isto é, os nomes *homem* ou *pessoa*. Segundo o autor, na região conhecida como área de Carlos Magno, que compreende territórios em que se fala o alemão, o francês e línguas escandinavas, o *homem* impessoal está totalmente gramaticalizado, possuindo uso genéricos, tais como no exemplo dado pelo Posio (2021, p. 3):

(15) **On** ne vit qu'une fois. 'Só se vive uma vez.'; uso genérico, ou seja, não ancorado em nenhum ponto específico no tempo (POSIO, 2021, p. 3).

Ainda de acordo com o autor, apenas o francês manteve o uso genérico do pronome sujeito *on*, tanto na língua escrita quanto falada, diferentemente do que ocorreu em português e no espanhol. Contudo, segundo Posio (2021, p. 5 [Tradução nossa]), “Apesar da perda de *homem* como sujeito impessoal, o português moderno tem dois outros sintagmas nominais que denotam ‘humanos’, viz. *a gente* ‘o povo’ e *a/uma pessoa* ‘a/a pessoa’ que desenvolveram usos como dispositivos referenciais impessoais humanos”¹⁰.

Nesse sentido, após a discussão e análise dos dados, o autor conclui que as construções com o item *pessoa* parecem ser sintagmas nominais semigramaticalizados e que tal item não possui sinais fortes de uma gramaticalização completa como a de *homem* no francês.

Observa-se, portanto, que nomes gerais, assim como as demais formas de indeterminação do sujeito, são utilizadas pelos falantes quando eles não querem ou não podem nomear o referente. Assim, uma forte relação se dá entre os nomes gerais e as formas de

¹⁰ No original: “Despite the loss of *homem* as an impersonal subject, modern Portuguese has two other noun phrases denoting ‘humans’, viz. *a gente* ‘the people’ and *a/uma pessoa* ‘a/the person’ that have developed uses as human impersonal referential devices”.

indeterminação e, por isso, busca-se a comparação dessas variantes entre o PB e o PE, uma vez que o fenômeno da indeterminação possui resultados muito semelhantes no Brasil, o que não parece ocorrer quando comparado ao português europeu.

Assim, passa-se agora para a metodologia, seguida da análise e discussão dos dados.

3 METODOLOGIA

Este capítulo discorre sobre os processos metodológicos utilizados ao longo deste trabalho. Inicialmente, apresenta-se e se justifica a escolha das cidades. Posteriormente, são discutidos os critérios adotados para a seleção dos informantes, para a gravação, para o registro e transcrição dos dados, além da preparação para a análise estatística. Após essas informações, descrevem-se as variantes e as variáveis desta pesquisa. Finalmente, relata-se sobre a codificação utilizada no intuito de obter os resultados.

3.1 A escolha das cidades

Ao se buscar fazer um trabalho que tivesse um tema inédito, optou-se por selecionar duas cidades, inicialmente, que fossem comparáveis no que se diz respeito à estrutura, à história e à vida dos habitantes. Tendo isso em vista, foram selecionadas Ouro Preto, cidade do interior de Minas Gerais, e Coimbra, cidade da região central de Portugal, principalmente, porque se nota que elas possuem características socioculturais semelhantes.

Contudo, devido à pandemia da Covid-19, a coleta de dados em Ouro Preto foi inviabilizada e, então, optou-se por se utilizar dados coletados pela autora no ano de 2017 na cidade de Bambuí/MG em vez de Ouro Preto/MG. Como os dados foram coletados da mesma forma, foi constatado pela banca de qualificação que tal escolha não causaria diferença nos resultados. A coleta em Coimbra foi finalizada antes do período de isolamento social e, com isso, este estudo manteve uma comparação entre Brasil e Portugal.

3.2 História de Bambuí e de Coimbra

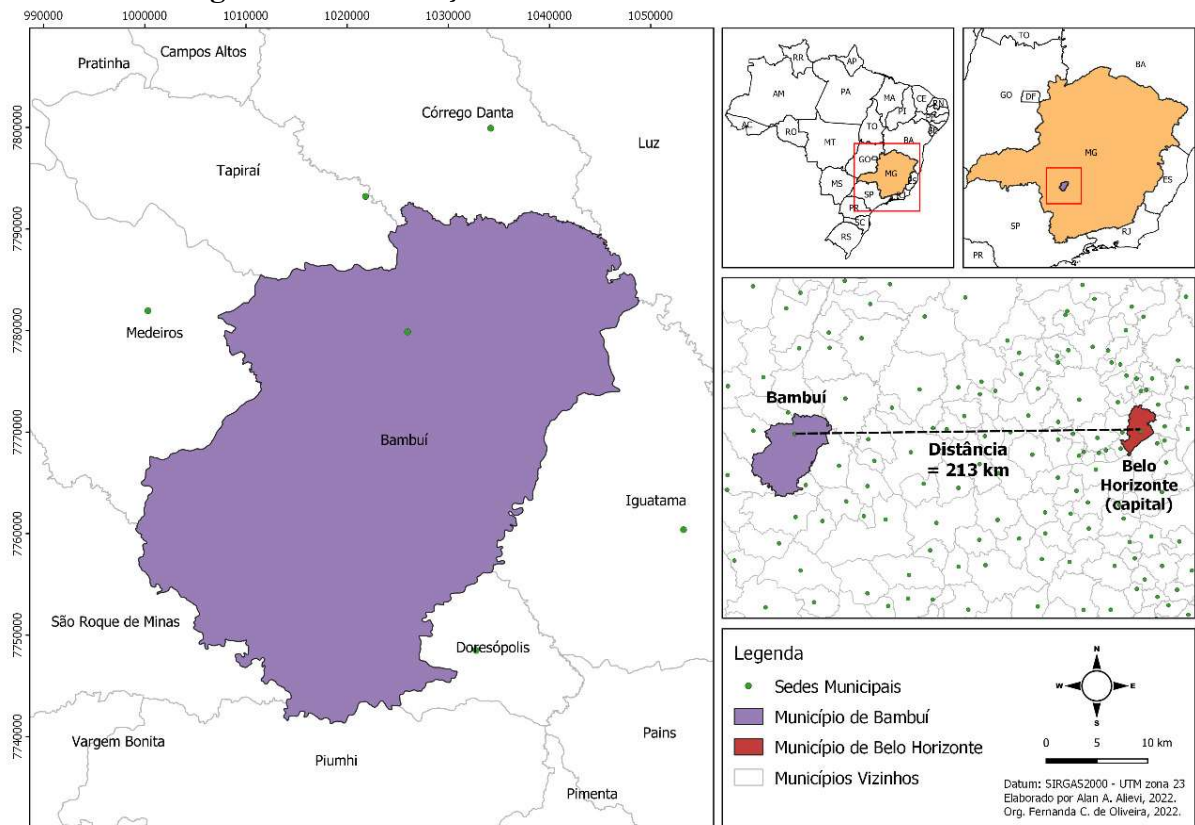
Nesta seção, é apresentado um pouco da história das cidades selecionadas, discorrendo-se sobre a localização e marcos importantes desses locais.

3.3 Bambuí

Bambuí, que, de acordo com dados do Instituto Brasileiro de Geografia e Estatística (IBGE) em 2021, tem uma população estimada em 23.964 habitantes, é uma cidade interiorana,

localizada no centro-oeste mineiro, a 270 km da capital Belo Horizonte, conforme mapa que pode ser visto na Figura 1:

Figura 1 – Localização de Bambuí no estado de Minas Gerais



Fonte: Elaborado pelo geógrafo Alan Alves Alievi a pedido da autora.

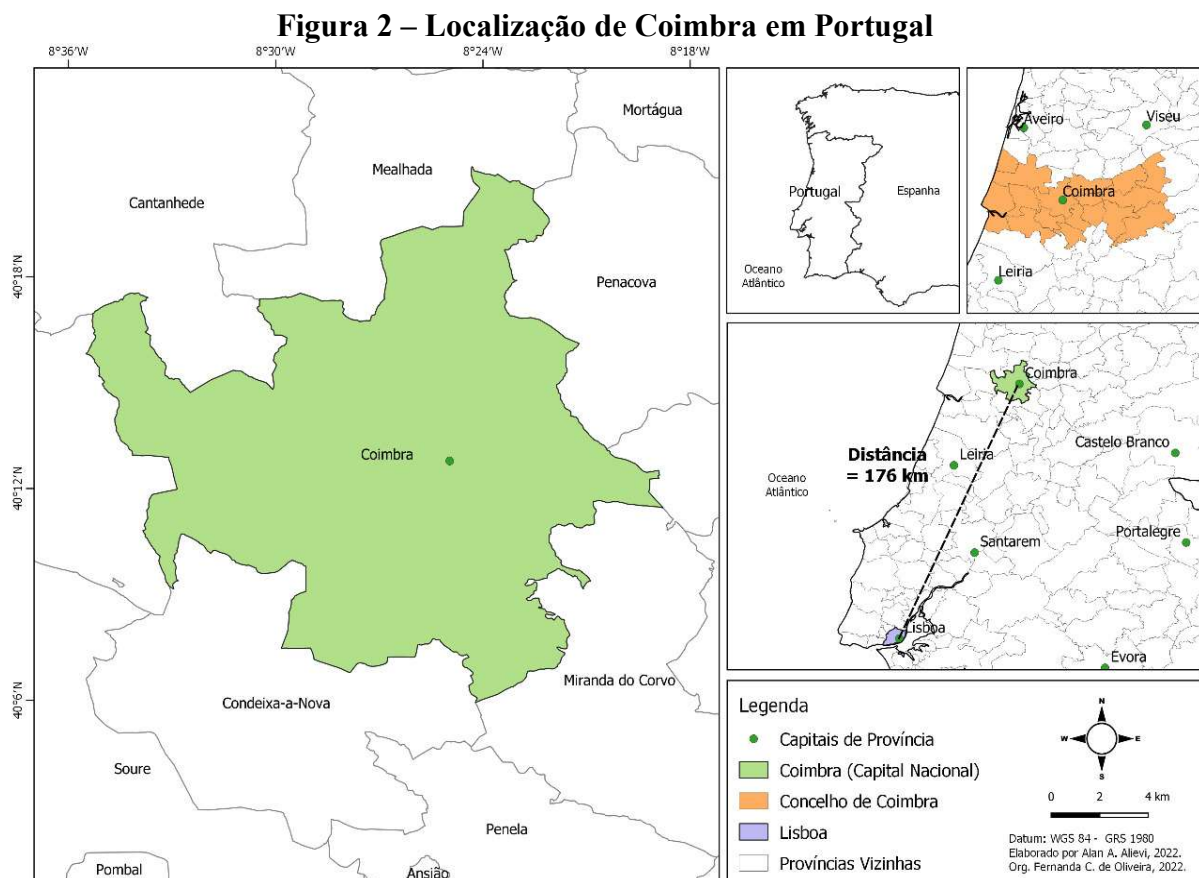
Lindiomar J. Silva publicou, em 2010, um livro intitulado *Bambuí nas Trilhas da Picada de Goiaz*, que foi considerado por uma lei municipal como a história oficial da cidade (Lei Municipal Nº 2178, de 07/11/2011). De acordo com o historiador, o fundador de Bambuí teria sido Inácio Correa Pamplona, um português, nascido na Ilha Terceira, no Bispado de Angra, no ano de 1731.

Ainda segundo Silva (2010, p. 115), “Bambuí talvez seja a cidade brasileira cujo nome tenha mais significados, dentre eles “rio das águas sujas”, “rio das borboletas” e “rio que corre na planície”. Contudo, de acordo com as pesquisas do autor, nenhum desses seria correto. O nome da cidade parece ter se derivado de uma tribo ou comunidade de nome Bambuí, a qual contava com aproximadamente dez mil pessoas, número este contabilizado em 1983, localizado em Camarões, na África Ocidental. Dessa forma, uma explicação para o nome tem relação com o fato de os negros quilombolas trazidos para a região terem dado o mesmo nome de sua tribo à cidade.

BambuÍ pertence à região da Serra da Canastra, local este com muitas cachoeiras e com uma produção de queijo artesanal conhecida em todo o país. Assim, por ser uma cidade antiga e de habitantes com hábitos bem campestres, acredita-se que uma pesquisa de caráter sociolinguístico trará resultados não só para o conhecimento do modo de falar da cidade, mas também contribuirá para uma comparação com a cidade de Coimbra, tendo em vista que os resultados de Bambuí assemelham-se aos demais resultados encontrados no Brasil, demonstrando uma estabilidade do fenômeno.

3.4 Coimbra

Coimbra, que, de acordo com o censo de 2021 realizado pelo Instituto Nacional de Estatística, tem 140.838 habitantes, é uma cidade portuguesa localizada na província da Beira Litoral, região central de Portugal, e se encontra a aproximadamente 200km de Lisboa, conforme Figura 2:



Fonte: Fonte: Elaborado pelo geógrafo Alan Alves Alievi a pedido da autora.

Costa (2011), em seu livro *Conhecer Coimbra*, conta com detalhes a história da cidade, fazendo uma descrição dos principais pontos turísticos, além de demonstrar a importância histórico-cultural de Coimbra para o país. De acordo com o autor, a posição geográfica privilegiada da cidade, que era conhecida como *Aeminium* no período romano, favoreceu a defesa de invasões durante toda história, o que auxiliou no controle de todo o território que a cercava. Um grande marco de Coimbra foi a presença dos povos germânicos que ocuparam todo o território europeu entre os séculos IV e VIII. Essa ocupação causou grandes transformações na arquitetura na cidade, no entanto, percebe-se ainda hoje muito da arquitetura romana que conseguiu sobreviver a todas essas mudanças.

Além da ocupação germânica, no ano 711, com as invasões muçumanas, a cidade passou, mais uma vez, por grandes mudanças, conforme relata Costa (2011):

com as invasões muçumanas de 711, a cidade seguiu o destino de todas as que se situavam a sul das montanhas das Astúrias e, durante mais de três séculos, foi islâmica, não obstante breves momentos de domínio das tropas cristãs, como aconteceu em 878. As marcas deixadas nas gentes e nos seus costumes foram profundas, mas, do campo artístico, os testemunhos dessa intensa colonização não chegaram até nós (COSTA, 2011, p. 16).

Avançando no tempo, tem-se a reconquista definitiva da cidade em 1064 pelas tropas de Fernando Magno e não há muitos relatos sobre o aspecto da cidade no período medieval. Assim, com o início dos reinados, o crescimento da cidade ocorreu de forma moderada e, somente a partir de 1537, Coimbra viveu grandes mudanças com a construção de grandes edifícios e a disseminação de colégios universitários. Ademais, nesse século, a população duplicou e foi possível observar a presença de pessoas em áreas ainda pouco habitadas na cidade. No que diz respeito aos séculos posteriores, a cidade continuou crescendo e se desenvolvendo de forma gradual até chegarmos à Coimbra de hoje, a qual é uma mistura de passado e presente.

Vale destacar a importância da Universidade de Coimbra para a cidade, uma vez que, desde sua fundação em 1290, mantém toda uma tradição, desde as festas e as vestimentas típicas dos estudantes, até sua arquitetura preservada. Assim, a universidade faz parte da vida dos moradores que já se acostumaram com a movimentação dos estudantes e percebem como a cidade fica diferente no período das férias.

Conclui-se que a cidade de Coimbra tem um grande valor histórico e cultural para todo o país devido à forte presença de eventos culturais que buscam a valorização das tradições.

Salienta-se que toda essa promoção de eventos e festas é feita juntamente com a universidade, o que a torna, de alguma forma, próxima da população.

Após a apresentação das cidades, descreve-se o passo a passo de como foi realizada a pesquisa.

3.5 A seleção dos informantes

A seleção dos informantes teve como base os preceitos apontados por Labov (2008 [1972]) como sendo essenciais para uma pesquisa sociolinguística, ou seja, os informantes foram escolhidos de acordo com os fatores extralinguísticos selecionados. Ademais, eles nasceram em suas respectivas cidades ou viveram nelas a maior parte da infância e adolescência.

Além do estudo de Tarallo (2003), os trabalhos de Carvalho (2010), Souza (2014), Oliveira (2018) e os demais apontados nesta pesquisa auxiliaram na seleção das variáveis extralinguísticas, visto que essas pesquisas contemplaram três delas: sexo, faixa etária e nível de escolaridade. Entretanto, é importante salientar que neste estudo o fator escolaridade foi descartado por duas principais motivações. A primeira delas está relacionada ao fato de que apenas em Barbosa (2016) o fator escolaridade foi observado como relevante, o que pode estar relacionado à seleção feita pela autora, diferente da dos demais trabalhos que estudaram o tema. A segunda motivação veio do fato de que não se acreditava na possibilidade de encontrar jovens, em Coimbra, com escolaridade inferior ao Ensino Secundário, equivalente ao Ensino Médio, considerando-se as condições oferecidas pelo país, conforme demonstra o “Atlas da Educação – Desempenho e potencial de sucesso e insucesso por concelho”, realizado entre 2012 e 2013 pela associação EPIS – Empresários pela Inclusão Social, em parceria com o Centro de Estudos de Sociologia da Universidade Nova (EPIS, 2012/2013). A pesquisa mostra que Coimbra tem uma das taxas de escolarização mais elevadas do país. Salienta-se que, apesar de os dados de Bambuí possuírem uma divisão por escolaridade, esse fator foi descartado na rodada para manter a amostra equiparada.

Assim, considerando que cada variável extralinguística contará com quatro informantes, sendo dois informantes portugueses e dois brasileiros, tem-se o Quadro 3, que apresenta a distribuição dos informantes por variáveis extralinguísticas:

Quadro 3 – Distribuição de informantes por variáveis extralinguísticas

Idade	Portugueses		Brasileiros	
	M	F	M	F
18-25 anos	4	4	4	4
30-45 anos	4	4	4	4
Acima de 50 anos	4	4	4	4
Total	12	12	12	12
TOTAL GERAL: 48				

Fonte: Dados da autora.

Feita a seleção e o levantamento da documentação necessária (vide APÊNDICES A e B), o projeto foi submetido ao Comitê de Ética da UFMG para avaliação e posterior início das gravações¹¹. Após a aprovação, buscou-se o contato com os informantes portugueses. No caso de Bambuí, as gravações já haviam sido realizadas em 2017¹² e, portanto, elas foram reutilizadas.

No caso de Bambuí, a coleta deu-se por meio de contatos com parentes e amigos que foram sugerindo nomes para serem entrevistados, uma vez que a pesquisadora é nascida na cidade. Por ser uma cidade interiorana e pelo conhecimento da pesquisadora com seus habitantes, não houve problemas na constituição do grupo de informantes, pois a população mostrou-se muito interessada em participar da pesquisa.

Tratando-se da cidade de Coimbra, ou seja, uma localização em outro país, a pesquisadora, primeiramente, fez contato com algumas pessoas conhecidas que residem na cidade para que a viagem pudesse ser marcada. É importante destacar que a pesquisadora residiu durante seis meses na cidade durante um intercâmbio feito em 2013/2014, o que facilitou, de certa forma, o contato inicial com os portugueses. Feito o contato, foi possível agendar a viagem para o período de 22/01/2020 a 09/02/2020. Em Coimbra, apesar do auxílio

¹¹ A pesquisa foi aprovada pelo Comitê de Ética em 17/03/2019. Número do projeto: CAAE 08126919.0.0000.5149 e Parecer nº 3.203.630.

¹² A pesquisa foi aprovada pelo Comitê de Ética em 28/09/2016. Número do projeto: CAAE 59734116.7.0000.5149 e Parecer nº 1.747.464.

de algumas pessoas, foi bastante difícil conseguir informantes. Muitas entrevistas foram obtidas na rua ou na universidade e, em alguns casos, os portugueses não se mostraram receptivos, considerando que a pesquisadora era estrangeira. As entrevistas foram realizadas também em asilos, associações de bairro e igrejas, dentre outros locais onde as pessoas se reuniam, o que contribuiu para que a pesquisadora conseguisse informantes. Todas as entrevistas foram realizadas no período de 28/01/2020 a 07/02/2020.

A seguir, descreve-se o processo de gravação e registro dos dados.

3.6 A gravação e o registro dos dados

Antes da realização das gravações, foram elaboradas a ficha do informante (APÊNDICE C), para que os dados dos informantes fossem coletados de forma organizada, e o roteiro de entrevistas (APÊNDICE D), que serviria como um norteador dos diálogos. Ressalta-se que o roteiro visou criar discussões que buscassem formas de indeterminar o sujeito e ainda considerou-se o que Meyerhoff, Schlee e Mackenzie (2015) discutem sobre a boa realização de uma entrevista, isto é, construir tópicos que façam com que o informante desenvolva o assunto sem necessitar de interrupções do entrevistador.

Após o final da coleta, foram obtidas 48 gravações, sendo 24 de Bambuí e 24 de Coimbra, as quais contabilizaram 27h48min30s de gravação e 275.023 palavras, distribuídas conforme Quadro 4:

Quadro 4 – Distribuição por tempo e número de palavras das gravações obtidas

Bambuí			Coimbra		
Identificação	Tempo	Nº de palavras	Identificação	Tempo	Nº de palavras
BAM02	00:46:53	8.112	COI10	00:16:35	2.289
BAM04	00:45:33	7.965	COI11	00:41:42	6.914
BAM05	00:25:31	4.721	COI12	00:23:51	4.209
BAM06	00:40:45	6.867	COI13	00:26:49	3.616
BAM07	00:44:00	8.211	COI14	00:13:50	2.096
BAM08	00:50:48	9.068	COI15	00:37:50	4.854
BAM09	00:45:34	7.204	COI16	00:21:18	3.272
BAM10	00:47:47	9.843	COI17	00:25:04	3.528
BAM11	00:30:31	5.459	COI18	00:27:44	3.138

BAM12	00:59:31	7.086	COI19	00:23:17	3.164
BAM13	00:30:25	5.692	COI20	00:18:23	3.078
BAM14	00:40:00	8.594	COI21	00:19:27	3.426
BAM16	00:50:31	7.822	COI22	00:23:53	3.769
BAM17	00:59:30	7.421	COI25 e 26	00:28:03	5.121
BAM18	00:37:30	6.986	COI28	00:14:15	2.304
BAM19	00:49:32	9.067	COI29	00:45:37	8.138
BAM21	00:39:48	6.325	COI30	00:37:41	4.920
BAM23	00:34:45	5.070	COI31	01:06:21	10.709
BAM24	00:26:09	5.135	COI32 e 33	00:42:45	5.277
BAM25	00:28:46	5.050	COI34	00:31:47	5.560
BAM26	00:45:53	7.327	COI35	00:25:35	4.367
BAM27	00:39:48	6.246	COI36	00:35:38	6.567
BAM28	00:20:39	4.571	COI38	00:26:10	5.407
BAM29	00:27:12	4.710	COI39	00:27:23	4.748
TOTAL	16:07:21	164.552	TOTAL	11:40:58	110.471

Fonte: Dados da autora.

3.7 A transcrição dos dados

Ao final das gravações, sejam as realizadas em Bambuí ou em Coimbra, iniciaram-se as transcrições, que duraram um período de quatro meses para cada uma das cidades e se basearam nos critérios estabelecidos por Amaral (2000), adaptados para os projetos sobre nomes gerais¹³, de acordo com o Quadro 5:

¹³ Informações completas em: www.lettras.ufmg.br/nomesgerais.

Quadro 5 – Normas adotadas para transcrição das entrevistas

SINAIS	OCORRÊNCIAS
()	Incompreensão de palavras ou segmentos
/	Truncamento
...	Qualquer pausa
((comentários))	Comentários descritivos do transcritor
“ ”	Discurso direto
[]	Supressão de diferentes segmentos sonoros
(...)	Corte na transcrição
Elevação/abaixamento das vogais pretônicas (<i>intão</i> < <i>então</i>) ¹⁴	
Vocalização da palatal (<i>trabaia</i> < <i>trabalha</i>)	
Permuta, apagamento ou inserção de diferentes segmentos (<i>ês</i> < <i>eles</i> ; <i>feiz</i> < <i>fez</i>)	

Fonte: Baseado em Amaral (2000).

Todas as transcrições foram feitas no programa *Microsoft Word*, a partir de um processo em que se ouvia a gravação e transcrevia-se o que foi dito, respeitando-se todas as normas estabelecidas acima. Depois de todas as transcrições realizadas, foi possível constatar a ocorrência, como mencionado, de 275.023 mil palavras no *corpus*.

3.8 A preparação dos dados para análise estatística

Realizadas as transcrições, as ocorrências foram coletadas para as futuras rodadas realizadas no Programa GoldVarb X. De acordo com Guy e Zilles (2007, p. 41), esse tipo de programa “utiliza um algoritmo baseado no procedimento de máxima verossimilhança para estimar os efeitos dos fatores”. Dessa maneira, primeiramente, converteram-se as transcrições em arquivo .txt para que os dados pudessem ser buscados pelo *software* AntConc. Para o que não foi possível ser buscado pelo programa, utilizou-se a própria ferramenta de busca do Word,

¹⁴ Exemplos retirados do *corpus* deste trabalho.

ou, em casos como o dos verbos sem sujeito explícito, fez-se a busca a partir da leitura das transcrições.

As formas indeterminadoras encontradas nas transcrições de Bambuí foram *a gente*, *você/ocê/cê*, *nomes gerais*, *nós/nóis/Ø+V1PP*, *eles/ês*, $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF$ e *se*. Já em Coimbra, encontraram-se as formas *a gente*, *você(s)/cê*, *nomes gerais*, *nós/ Ø+V1PP*, *eles*, $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF$, *se* e *tu*. Todos os dados coletados foram inseridos em uma planilha do Excel e foram analisados a partir dos seguintes fatores: sexo, faixa etária, presença ou ausência de locativo, presença ou ausência de anáfora, grau de indeterminação e mais ou menos coletivo. Salienta-se que fatores como tempo verbal, modo e tipo de oração foram excluídos dos grupos de fatores após a qualificação, pois foi visto que esses dados não traziam contribuições significativas para este estudo, uma vez que aqui se analisa a fala e se basear em determinadas categorias da gramática tradicional para fazer tais análises não se mostrou muito satisfatório.

As próximas seções tratarão das variantes e de cada variável separadamente, e, posteriormente, serão apresentadas as rodadas realizadas no *software* GoldVarb X.

3.9 As variantes

Nesta pesquisa tem-se como variável dependente a indeterminação do sujeito e as variantes são as dez formas distintas em que essa indeterminação se apresenta. As variantes selecionadas são *a gente*, *você(s)/ocê/cê*, *nomes gerais* (*cara*, *gente*, *fulano*, *homem*, *pessoa*, *peçoal*, *peçoas*, *povo*, *ser humano*), *nós/nóis/Ø+V1PP*, *eles/ês*, $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF^{15}$, *se* e *tu*. As formas foram selecionadas a partir de trabalhos anteriores, tal como o de Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010), Souza (2014), Oliveira (2018) e, ainda, só foram consideradas indeterminadas quando poderiam ser comutadas pelo *se*, conforme pontuado por Menon (1994) e exemplo (16) apresentado a seguir:

(16) nó a **pessoa** fala que [a]caba o vício mai[s] num [a]caba não
(BAM29F)

(16a) nó **se fala** que [a]caba o vício mai[s] num [a]caba não
(BAM29F)

15 No caso $\emptyset+VINF$, observou-se que, na maioria das vezes, a comutação por *se* não parece ser possível. No entanto, assim como em trabalhos já apresentados, acredita-se que isso não torna a forma menos indeterminadora que as demais, conforme poderá ser visualizado em exemplos apresentados posteriormente.

(16b) nó **fala-se** que [a]caba o vício mai[s] num [a]caba não
(BAM29F)

A seguir serão apresentadas as variantes, divididas de duas formas: os nomes gerais, foco deste trabalho, e as demais variantes indeterminadoras.

3.9.1 Os nomes gerais

As formas nominais já estudadas como maneiras de se indeterminar o sujeito por Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010) e Souza (2014) e tratadas por Oliveira (2018) e neste estudo como nomes gerais, ocorrerem, em um número relevante, no conjunto de variantes indeterminadoras nos trabalhos citados. Nota-se que há uma estabilidade no Brasil quanto ao número de ocorrência desses nomes gerais, uma vez que, nos trabalhos em que se trata do tema, eles ocupam, normalmente, a terceira posição, seguido apenas de formas como *a gente* e *você*.

Salienta-se que só foi considerada como ocorrência de nome geral aqueles itens que eram acompanhados por artigo definido ou aqueles sem determinantes. Itens acompanhados de artigo indefinido ou ocorrências como *as pessoas aqui*, isto é, com um locativo seguido de uma das formas de indeterminação pesquisadas neste estudo, foram descartados. Essa escolha justifica-se pelos fatos de que Menon (1994) já tratava da composição das formas nominais, ou seja, artigo definido + substantivo, e de que, mais recentemente, Amaral e Ramos (2014), ao tratar especificamente dos nomes gerais, salientam que eles podem vir precedidos ou não por determinantes. No entanto, apesar da constatação de Amaral e Ramos (2014) sobre determinantes, há que se considerar que este trabalho tem como foco a indeterminação do sujeito e percebeu-se que, em ocorrências como *pessoas aqui*, a indeterminação não ocorria, uma vez que o locativo junto ao nome consegue, de alguma forma, apontar sobre qual grupo se está falando.

Tendo isso em vista, esta pesquisa encontrou dez nomes gerais, sendo eles *cara*, *fulano*, *gente*, *homem*, *pessoa*, *pessoal*, *pessoas*, *povo* e *ser humano*. Algumas ocorrências desses nomes podem ser visualizadas nos exemplos abaixo:

(17) e quem recebe aí um salário de/de aposentadoria é... marido morre a isposa morre o **cara** vai recebe[r] só metade? eu acho uma sacanagem tamém (BAM11F)

(18) não porque muita gente vai na onda assim **fulano** tá vamo[s] tamém eu acho que mui/uns sabem o que tá fazeno otros não vai na/na mesmice (BAM07F)

(19) só veju o p[e]ssual aqui pela av[enida] acima... tudu jovens... **gent[e]** [es]tão (dretinho) ao bar aqui nesta rua tem ali o ambe...ah não... já não tem... (COI29F)

(20) num tô falano o que o po/o que o **home[m]** fala eu tô falano o que Deus de[i]xô (BAM05F)

(21) tem... a **peessoa** tem que/ que fica lá na portaria a dizer quem vai visitar (COI11F)

(22) Portugal é um país... entre política e futebol... se for preciso um jogo ímportamnte o **peessoal** não vai votar (COI18M)

(23) as **p[e]ssoas** ficam contentis purque houve alguém que lhes disse bom dia (COI34M)

(24) que[i]ma o filme do Brasil na minha opinião que num é legal não mais acho que tamém o **povo** tamém tem que acorda[r] né se tá ruim vamo[s] luta[r] p[r]a melhora[r] mais eu num sei se é o certo não (BAM07F)

(25) p[r]a joga[r] na internet ixatamente é/é/é isso aí até é chato né a/e/e/o tipo o **ser humano** num podia chega[r] a esse ponto não (BAM19M)

A partir dos exemplos acima, pode-se visualizar quando o nome geral ocorre com a presença do artigo definido, seja ele no plural como em (23) ou no singular como em (17), (20) (21), (22), (24) e (25), ou sem a presença de artigo, como em (18) e (19).

Passa-se agora para a apresentação das demais formas encontradas no *corpus*.

3.9.2 As demais formas de indeterminação encontradas

Outras formas encontradas neste trabalho para indeterminar o sujeito foram *a gente*, *você(s)/ocê/cê*, *nós/nóis/Ø+V1PP*, *eles/ês*, *Ø+V3PS*, *Ø+V3PP*, *Ø+VINFL*, *se* e *tu*. Essas formas foram selecionadas tendo em vista os trabalhos de Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Oliveira (2006), Carvalho (2010), Souza (2014), Oliveira (2018) e, também, a partir dos dados coletados nesta pesquisa.

Itens como *ocês* e *cês* e construções com voz passiva sintética e voz passiva sem agente foram pesquisadas, mas não foram encontradas. Abaixo, apresentam-se as variantes e os exemplos retirados de nosso *corpus*.

A gente

(26) ah... tem co/assim eu acho que quanto mais **a gente** vai aprendeno a ciência **a gente** vai de[i]xano de acredita[r] na religião porque é/é difícil cê acredita[r] nas duas coisa o mesmo tanto (BAM26M)

Você(s)/ocê/cê

(27) se **você** faz uma brincadeira ou dá uma piada... as p[e]ssoas ainda ficam assim a olhar com ar de disdém...(COI29F)

(28) já fora d'horas... purque eu acabei o ens/na altura acabei o sicundáriu... que **vocês** chamam o insino...(COI30M)

(29) não acredito que se ela tive[r] força de vontade porque tem pessoas que/que eu cunheço que já internô sete vezes se a pessoa num quise[r] ela nunca que vai sai[r] tem que te[r] vontade num adianta **ocê** i[r] lá e interna[r] e fica[r] lá não porque **ocê** sai mai[s] pior do que cê foi (BAM29F)

(30) tanto que muitas vezes **cê** vê os jovens lá concentrados nesses grandes (edifícios) cumerciais...(COI29F)

Nós/nóis

(31) às vezes **nós** temos porque esperar ... **nós** temos que ser nós mesmos ... não temos que esperar que as pessoas sejam como nós ... como um é cada como qual ... **nós** temos que ser assim (COI22F)

(32) ah eu acho que o Brasil afundô numa crise né crise não só econômica mais pulítica social de ética acho que **[es]tamo[s]** no fundo do poço (BAM06M)

(33) saúde paiz né isso aí **nóis** deseja (BAM05F)

Ês/eles

(34) porque os outro é assim **ês** fala nossa fulano tá com a casa ah o fulano tem um carrinho p[r]a anda[r] ah aí **ês** fala ah mais mais **ês** num vê né que que a gente faiz (BAM14M)

(35) mas se não houver bola das sete às oito **eles** dão esse/esse/esse concurso (COI11F)

Verbo na terceira pessoa do singular sem sujeito explícito (Ø+V3PS)

(36) intão acaba que falta um po[u]quinho de/de/de/de olha[r] mesmo **Øgeri** melhor essa questão (BAM02M)

Verbo na terceira pessoa do plural sem sujeito explícito (Ø+V3PP)

(37) dipois não/não **Øprocuram** empregos cuja r[e]muneração seja baixa portanto a indústria tem muito imprego... mas mal r[e]munerado...(COI34M)

Verbo no infinitivo sem sujeito explícito (Ø+VINFIN)

(38) tem muitas calçadas que são irregulares né por exemplo o idoso tem/tem dificuldade né... é... talvez[i]z mais atenção aos idosos em relação a isso... **cria[r]** também é... programas culturais num sei alguma coisa que possa né leva[r] os idosos a te[r] uma vida também fora de casa (BAM02M)

Se

(39) pode haver/pode-se... pode-se dizer que as ligações entre a cidad[e] e a universidad[e] não são boas (COI34M)

Tu

(40) tantu que **tu** não consegue sobreviver porque não há políticas suciais.... tantu não há nada que diga assim "olha essa senhora hoje... este senhor está nessas condições... mas tem aqui este apoio que não paga nada..." não... isto não existe (COI12F)

Nas próximas seções serão apresentadas e discutidas as variáveis extralinguísticas e linguísticas selecionadas para esta pesquisa.

3.10 Variáveis extralinguísticas

Para este trabalho, foram selecionadas as variáveis sexo e faixa etária, as quais, como aponta Mollica (2004), têm sido tratadas por cada vez mais estudos, mas sobre as quais ainda se tem muitos questionamentos a serem desenvolvidos.

Cada uma dessas variáveis será discutida nas próximas seções.

3.10.1 Sexo

A partir dos trabalhos que tratam da indeterminação já citados anteriormente, optou-se por testar a variável sexo, pois acredita-se que a seleção ou não de uma forma de indeterminação pode estar diretamente correlacionada a este fator.

Setti (1997), Godoy (1999), Carvalho (2010) e Souza (2014) constataram em seus trabalhos que o sexo masculino é favorecedor das formas nominais como meio de indeterminação. Além disso, em Carvalho (2010), tem-se uma informação relevante, isto é, a de que sexo foi o único fator extralinguístico selecionado, ou seja, não foi selecionado nem grau de escolaridade e nem faixa etária pelos itens que aqui denominamos de nomes gerais. No caso de Carvalho (2010), os homens demonstraram um favorecimento quanto ao uso dos nomes gerais e, no estudo de Oliveira (2018), concluiu-se que homens e mulheres utilizam as formas de indeterminação de maneira equiparada, não sendo possível observar, portanto, nenhum traço mais conservador por parte das mulheres ou mesmo algum tipo de estigma quando se considera tal fator.

Tendo tal diferença dos resultados em vista, a comparação proposta neste estudo poderá nos mostrar os caminhos que a indeterminação tem tomado tanto no PB quanto no PE.

3.10.2 Faixa etária

Refletindo-se sobre o que Bortoni-Ricardo (2014) apresenta a respeito dos papéis ocupados pelos membros das comunidades devido a sua faixa etária e baseando-se em trabalhos anteriores que trataram do mesmo tema, selecionaram-se para esta pesquisa três faixas etárias, sendo elas F1 – de 18 a 25 anos –; F2 – de 30 a 45 anos – e F3 – acima de 50 anos. Na primeira faixa, têm-se pessoas mais jovens, que estão ainda estudando ou prestes a se inserir no mercado de trabalho. A segunda faixa contempla aquelas que, possivelmente, já estão inseridas no mercado de trabalho e em fase adulta. A terceira faixa diz respeito às pessoas mais velhas e que viveram em tempos com costumes diferentes das demais faixas. Acredita-se, portanto, que, devido a esses papéis diferentes que cada um desses grupos exerce na sociedade, a variável faixa etária seja um fenômeno que ajude a avaliar casos de mudança em progresso, além de provocar resultados interessantes nos dados coletados.

Desse modo, espera-se que os mais velhos façam maior uso de formas mais conservadoras, isto é, aquelas que a gramática tradicional considera como indeterminadoras, e

os mais jovens utilizem mais as formas com nomes gerais, conforme já comprovado em estudos anteriores, como o de Oliveira (2018). Salienta-se que, diferentemente do fator *sexo*, o fator *faixa etária* não foi um fator extralinguístico selecionado pelos trabalhos sobre o tema, exceto no de Barbosa (2016), no qual esse fator extralinguístico demonstrou resultados significativos. Contudo, a autora utilizou critérios diferentes dos demais trabalhos sobre o tema. Espera-se, portanto, que esse fator, ainda que não selecionado na maioria dos estudos sobre o tema, traga resultados interessantes para a discussão do fenômeno.

3.11 Variáveis linguísticas

As variáveis linguísticas selecionadas basearam-se, principalmente, nas pesquisas feitas por Carvalho (2010) e Oliveira (2018), visto que esses são os estudos que mais se aproximaram de nosso trabalho. Assim, selecionaram-se quatro fatores que serão discutidos abaixo. São eles: a) presença ou ausência de anáfora; b) grau de indeterminação; c) presença ou ausência de locativo e d) mais ou menos coletivo.

3.11.1 Presença ou ausência de anáfora

Carvalho (2010) e Oliveira (2018) mostraram em seus trabalhos a relevância da variável presença ou ausência de anáfora para o estudo das formas nominais. Tendo isso em vista, optou-se por selecioná-la para esta pesquisa classificando-a como presença ou ausência de anáfora. O que se pretende aqui é avaliar se quando uma forma de indeterminação é utilizada, o informante vai mantê-la ou vai utilizar uma forma distinta, talvez mais ou menos indeterminada.

No que concerne aos critérios adotados para a delimitação de tal variável, consideraram-se aqueles adotados por Santana (2006) e Carvalho (2010), sendo eles:

- 1 – selecionou-se apenas um turno de fala, isto é, não deveria haver nenhuma interferência do entrevistador;
- 2 – o turno de fala não deveria ultrapassar o limite de dez linhas; e
- 3 – não deveria haver mudança do tema tratado e, caso acontecesse, a ocorrência seria considerada a primeira de uma nova série.

Apesar de ser uma análise semelhante à dos autores citados, optou-se, neste trabalho, por se considerar apenas três formas para classificar os dados, adotando os critérios de Oliveira (2018). Assim, a análise foi então elaborada da seguinte forma: a) retomada anafórica com o

mesmo elemento, ou seja, no caso das variantes que possuem pronomes, por exemplo, só seria considerada como retomada quando o informante repetisse uma forma idêntica à citada anteriormente; b) ausência de anáfora, isto é, quando o informante não retomasse com nenhum outro elemento, por fim; c) retomada anafórica com outro elemento, que ocorreu quando se usava duas formas distintas de indeterminação em um mesmo turno de fala.

Abaixo apresentam-se exemplos das três maneiras de forma antecedente encontradas:

a) Retomada anafórica com o mesmo elemento

(41) as **peessoas** num lutam em geral as **peessoas** num lutam pelos direitos elas lutam pelos seus interesses são po[u]cas as pessoas que lutam por direitos né? talvez o que faça as **peessoas** busca[r]/re/reivindica[r] alguma coisa é algum tipo de/de interesse né num sei tenho que pensa[r] milho[r] sobre isso (BAM02M)

No exemplo (41), vê-se que a variante *peessoas* é retomada na oração seguinte da mesma forma e, por isso, ela possui todos os critérios estabelecidos anteriormente para ser considerada como *retomada anafórica com o mesmo elemento*.

b) Ausência de anáfora

(42) e a nível pulíticu eu achu que tem ficadu assim porque **a gente** isquece essa part[e] é... já qui pulíticu é assim...(COI38F)

No exemplo (42), não existe nenhuma forma de retomada ao longo do contexto, tendo-se, portanto, no turno de fala selecionado, apenas uma aparição do item *a gente*. Este é, portanto, um exemplo que consideramos como *ausência de anáfora*.

c) Retomada anafórica com outro elemento

(43) até essas saídas na rua aí esse manifesto de rua aí acho que o **peessoal** agora não tão assim muito seno bajulado e seno assim engambelado acho que **eles** tão intendo o valor do voto sabe (BAM04M)

Os dados classificados como *retomada anafórica com outro elemento* referem-se aos casos em que é usada uma forma e, posteriormente, ela é retomada por outra forma da utilizada anteriormente. No exemplo (43), o autor utiliza o nome geral *peessoal* e depois faz a retomada

com o pronome *eles*, sendo, portanto, o exemplo classificado como retomada anafórica com outro elemento.

3.11.2 Grau de indeterminação

A variável *grau de indeterminação* foi escolhida para esta pesquisa tendo em vista que, em Carvalho (2010) e em Oliveira (2018), essa variável foi selecionada pelo *software* Goldvarb X como sendo relevante para a análise.

Siewierska e Papastathi (2011, p. 576), em uma pesquisa sobre a terceira pessoa do plural em inglês, apresentam cinco maneiras de se indeterminar a terceira pessoa do plural, sendo elas universal (“*Na Espanha, eles comem tarde*”), corporativa (“*Eles mudaram as leis fiscais no ano passado*”), vaga (“*Eles encontraram sua bicicleta na parte de trás de um celeiro*”), inferencial (“*Eles estão fritando batatas fritas aqui*”) e existencial específica (“*Eles estão batendo na porta*”)¹⁶. Tomando tal classificação e considerando os estudos de Santana (2006), Carvalho (2010) e Oliveira (2018), estabeleceram-se três classificações para este trabalho, sendo: a) indeterminação parcial com referente explícito; b) indeterminação parcial com referente implícito; c) indeterminação total.

A indeterminação parcial com referente explícito contempla aqueles casos em que é possível inferir, de alguma forma, o referente e em que, no texto, ocorre alguma marca explícita. Esses casos assemelham-se ao que Siewierska e Papastathi (2011) classificam como inferencial. A segunda, no entanto, diz respeito àqueles casos em que é possível inferir de quem se trata pelo contexto, mas não existe nenhuma marca pela qual seja possível identificar o referente, tal como o que Siewierska e Papastathi (2011) nomeiam de existencial específico. Já a indeterminação total diz respeito àqueles casos, tal como nomeado por Siewierska e Papastathi (2011), de universal, em que não é possível nenhum tipo de inferência e que, com isso, servem para se referir a qualquer pessoa em qualquer lugar do mundo. Mostram-se, assim, alguns exemplos retirados do *corpus*:

¹⁶ Exemplos traduzidos pela autora. No original: *In Spain, they eat late; they changed the tax laws last year; they've found his bike in the back of a barn; they've been frying chips here; they're knocking on the door.*

a) Indeterminação parcial com referente explícito

(44) porque desde de quando eu moro aqui assim toda vida morei mai[s] **a gente** num vê fala[r] em usina te deu lucro feiz isso feiz num vê cê num vê fala[r] uai (BAM14M)

Em (44), tem-se um dado que pode ser classificado como indeterminado com referente explícito, pois, a partir do contexto, é possível perceber que a informante está falando dos moradores da sua própria rua e dele mesmo, ainda que isso seja indeterminado, pois não é possível precisar, exatamente, quem são as pessoas.

b) Indeterminação parcial com referente implícito

(45) tem... a **pessoa** tem que/ que fica lá na portaria a dizer quem vai visitar... (COI11F)

O exemplo (45) é classificado como indeterminação parcial com referente implícito, pois o contexto envolve uma discussão sobre os asilos e como funciona a visita aos idosos que lá residem. Assim, a informante usa *pessoa* para se referir aos parentes, amigos ou conhecidos de idosos que vivem nesses locais. Como não é possível determinar exatamente quem são essas pessoas, esse tipo de ocorrência tem a classificação apresentada.

c) Indeterminação total

(46) mas parece que d'algum tempu pra cá **nota-se** muito mais organizações a fazerem aos fins de simana... (COI32 e 33F)

A indeterminação total ocorre quando se tem uma indeterminação por excelência, ou seja, não se pode recuperar o referente de forma nenhuma pelo contexto. O exemplo (46) é um desses casos, pois *nota-se* não faz referência a ninguém e, portanto, é um bom exemplo desse tipo de indeterminação.

3.11.3 Presença ou ausência de locativo

A seleção desta variável se deu com base na proposta de Oliveira (2018). Além disso, ao se transcrever as gravações, percebeu-se que, assim como observado por Oliveira (2018),

com certa frequência, os informantes indeterminavam o sujeito, mas, ao mesmo tempo, faziam usos de termos, como *aqui* e *ali*, ou mesmo alguma referência à cidade ou ao país por meio de lugares, ruas ou pelo próprio contexto abordado.

Tendo isso em vista, viu-se a necessidade de classificar os dados coletados a partir da presença de locativo ou da ausência de locativo. É importante lembrar que as entrevistas versaram sobre temas relacionados à fundação e ao modo de vida na cidade, acreditando-se, portanto, assim como Oliveira (2018), que esse fator seja relevante para a análise. Apresentam-se abaixo dois exemplos, um ilustrando a presença de locativo (47) e outro a ausência de locativo (48).

(47) não o ritmo de vida é otro que leva as pessoas tamém a agi de forma diferente né aqui **nós** andamos na rua e conhecemos grande parte das pessoas...(BAM02M)

(48) não porque muita gente vai na onda assim **fulano** tá vamo[s] tamém eu acho que mui/uns sabem o que tá fazeno otros não vai na/na mesmice (BAM07F)

O exemplo (47) foi classificado como *presença de locativo*, porque no contexto são feitas referências como “aqui nós andamos”, “conhecemos grande parte das pessoas”, tudo para demonstrar que está se referenciando pessoas daquela região. Diferentemente do que ocorre em (47), em (48) não é possível encontrar nenhuma dessas formas que servem para demarcar o lugar de referência e, portanto, na análise, esse item foi classificado como *ausência de locativo*.

3.11.4 *Mais ou menos coletivo*

A variável *mais ou menos coletivo* foi estudada somente por Oliveira (2018) e, apesar de não ter sido selecionada nos dados de Bambuí pelo *software* Goldvarb X, acredita-se que ela possa ser relevante em nossa análise, por possivelmente ser capaz de condicionar a realização do grupo de variantes.

Flaux e Velde (2000) definem coletivos como aqueles nomes constituídos de uma pluralidade interna, assim como *buquê*. Flaux (1999), ao tratar deste mesmo tema, explica que os nomes coletivos não são completamente independentes, visto que possuem uma relação lexical com os nomes correspondentes aos membros das coleções que denotam. Reforça ainda que eles são genéricos, pois abarcam todos aqueles que pertencem ao conjunto.

Tendo em vista os estudos das autoras apresentadas, bem como a proposta de análise de Oliveira (2018), definiriam-se as formas de indeterminação em *mais e menos coletiva*, considerando como coletivos aqueles itens que possuem uma pluralidade interna, tal como a

gente, pessoas, pessoal, povo, eles, nós etc. Já itens como *você, pessoa, cara, camarada, fulano* e outros foram considerados menos coletivos, visto que são genéricos, mas não possuem uma pluralidade interna, assim como relatam Flaux e Velde (2000).

Assim, apresentam-se exemplos de itens classificados como mais coletivo e menos coletivo.

a) Mais coletivo

(49) ah o **povo** ixagera sá cê anda lá nas bera dos corgo num tem uma fruta pros macaco cumê tadinho ês tá morreno de fome (BAM05F)

O exemplo (49) foi classificado para a análise dos dados como mais coletivo, tendo em vista sua referência a mais de uma pessoa.

b) Menos coletivo

(50) e quem recebe aí um salário de/de aposentadoria é... marido morre a isposa morre o **cara** vai recebe[r] só metade? eu acho uma sacanagem tamém (BAM11F)

No exemplo (49), nota-se que o item *povo*, diferentemente do exemplo (50), trata apenas de um indivíduo e, portanto, não possui pluralidade interna.

Assim, após delimitação das variáveis linguísticas e extralinguísticas apresentadas, passou-se para a codificação dos dados a serem analisados pelo programa GoldVarb X. Todo esse tratamento dos dados será apresentado na seção seguinte.

3.12 A codificação dos dados

Scherre e Naro (2014) apresentam uma discussão sobre a coleta, transcrição e codificação dos dados para uma pesquisa sociolinguística. No que diz respeito às codificações, definem que “codificar é transformar em código identificável pelos programas computacionais disponíveis tudo o que queremos que seja quantificado” (SCHERRE; NARO, 2014, p. 155). Tendo em vista tal conceito, fez-se, assim, a codificação dos dados (APÊNDICE E) e se utilizou o *software* GoldVarb X para a seguinte rodada binária: sintagmas compostos por nomes gerais *versus* demais formas de indeterminação do sujeito.

Após a codificação, a partir da utilização dos códigos criados e da formatação em arquivo em formato .txt, fez-se a primeira rodada dos dados. No caso de Bambuí, na rodada inicial, houve um *KnocKout* na variável linguística *presença ou ausência de anáfora*, isto é, 100% dos dados em que ocorreram *retomada anafórica com outro elemento* aconteceram nas demais formas de indeterminação. Nesse sentido, para se obter os pesos relativos, foi feita a junção da *retomada anafórica com o mesmo elemento* com a *retomada anafórica com outro elemento* e, então, o *KnocKout* foi resolvido. Assim, no caso de Bambuí, a análise do fator *presença ou ausência de anáfora* ocorrerá entre *ausência de anáfora* e *presença*, seja ela semelhante ou diferente.

Na rodada de Coimbra, também ocorreu um *KnocKout*. No entanto, nesse caso, houve 100% de ocorrência de *indeterminação parcial com referente explícito* nas demais formas de indeterminação. Sendo assim, optou-se pela junção dos dados de *indeterminação parcial com referente implícito e explícito*. Dessa forma, no grupo de fator *grau de indeterminação*, com os dados de Coimbra, a rodada ocorreu da seguinte forma: *indeterminação total versus indeterminação parcial com referente implícito ou explícito*.

É importante destacar que, para a equiparação dos dados, foram feitas rodadas idênticas das duas cidades, isto é, juntou-se, em Bambuí, o fator *grau de indeterminação* e, em Coimbra, o fator *ausência ou presença de anáfora*. Contudo, a diferença com os dados obtidos na rodada inicial não foi significativa e, por isso, optou-se pela primeira rodada já explicitada.

Ademais, após a retirada dos *KnocKouts*, realizaram-se as rodadas finais para a obtenção dos pesos relativos, as quais serão apresentadas e discutidas nas seções seguintes.

4 RESULTADOS E ANÁLISE DOS DADOS

Neste capítulo, serão apresentados dois quadros, um com os resultados obtidos em Bambuí e outro com os resultados obtidos em Coimbra, além de uma discussão sobre as variantes encontradas. Posteriormente, tem-se a apresentação das rodadas realizadas no Goldvarb X de Bambuí e de Coimbra, discutindo-se os fatores linguísticos e extralinguísticos que influenciaram ou não o fenômeno pesquisado. Finalmente, para que as conclusões sejam obtidas, será feita uma comparação entre os resultados das duas localidades.

4.1 Apresentação dos resultados e das variantes

No *corpus* analisado, foram encontradas 2.297 ocorrências distribuídas entre as formas de indeterminação. Em Bambuí, as variantes totalizaram 1357 ocorrência e, em Coimbra, 940. *A gente* foi a forma mais recorrente em Bambuí e *nós/desinência de 1ª pessoa do plural* foi a variante mais recorrente em Coimbra, conforme pode ser visualizado nas Tabelas 1 e 2:

Tabela 1 – Número de ocorrência e porcentagem de cada uma das variantes em Bambuí

Forma de indeterminação	Número de ocorrências	Porcentagem (%)
a gente	376	27,7
você/ocê/cê	325	23,0
nomes gerais	191	14,1
nós/nóis/desinência de 1ª pessoa do plural	134	9,9
eles/ês	122	9,0
Ø+V3PS	122	9,0
Ø+V3PP	39	2,9
Ø+VINF	36	2,7
Se	12	0,9
TOTAL	1357	100,0

Fonte: Dados da autora.

Tabela 2 – Número de ocorrência e porcentagem de cada uma das variantes em Coimbra

Forma de indeterminação	Número de ocorrências	Porcentagem (%)
nós/desinência de 1ª pessoa do plural	451	48%
SE	151	16,1%
Ø+V3PP	127	13,5%
nomes gerais	105	11,2%
a gente	36	3,8%
Ø+VINF	21	2,2%
Eles	19	2,0%
Ø+V3PS	13	1,4%
você(s)/cê	11	1,2%
Tu	6	0,6%
TOTAL	940	100%

Fonte: Dados da autora.

A partir das tabelas apresentadas, faz-se, a seguir, uma discussão das variantes encontradas.

4.1.1 A gente

A gente é a forma mais recorrente no *corpus* coletado em Bambuí, sendo encontradas 376 ocorrências desse item. Esse primeiro lugar também foi encontrado em Setti (1997), na cidade de Florianópolis, e em Oliveira (2006), Ponte (2008) e Souza (2014). Em Coimbra, no entanto, esse item aparece apenas 36 vezes. Vale salientar que a variante mais recorrente em Coimbra é *nós/desinência de 1ª. pessoa do plural*. Isso foi um resultado relevante, pois o *a gente* concorre amplamente no PB com a forma *nós*, conforme apresentado por Silva e Namiuti (2019) e, portanto, os resultados mostram que em Portugal utiliza-se mais a forma prevista pela gramática que indica primeira pessoa do plural quando se busca indeterminar o sujeito.

4.1.2 *Você(s)/ocê/cê*

Essas três formas ocupam o segundo lugar, em número de ocorrências, em Bambuí. Elas totalizam conjuntamente 325 dados. *Cê* foi a forma mais recorrente com 250 dados, seguida de *você* com um número bem inferior, apenas 53 ocorrências e, finalmente, com um número mais reduzido, está *ocê* com 22 dados. Essa forma de indeterminar ocupa o primeiro lugar no conjunto das variantes em Menon (1982), Rollemberg *et al.* (1991), Setti (1997), na cidade de Curitiba, além de Godoy (1999) e Carvalho (2010). Em Coimbra, essa é uma variante pouco produtiva, ocorrendo apenas 11 vezes. Além disso, essa variante ocorre no plural (*vocês*) em Coimbra por duas vezes, forma esta que não ocorre em Bambuí.

Vale destacar, conforme Nascimento *et al.* (2018), que a variante *você* tem comportamento diferente no PB e no PE. No PE, esse é um item que tende a ser evitado e que tem se mostrado mais produtivo em relações assimétricas e sem proximidade. Contudo, as autoras reforçam que esse quadro parece estar sofrendo alteração e os mais jovens têm utilizado com maior frequência esse item. No PB, no entanto, existe uma preferência pelo uso de *você* e a concorrência entre *tu* e *você* só existe de forma regionalizada devido a migrações internas para algumas regiões do país.

De forma geral, os resultados, no Brasil, demonstram um grande número de ocorrências desse item, o que não acontece em Coimbra. Ademais, um fato percebido pela autora no decorrer da coleta de dados é que, em geral, os portugueses se preocupam em manter uma formalidade durante as gravações.

4.1.3 *Nomes gerais*

Os nomes gerais, já estudados como formas de indeterminação por Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010) e Souza (2014), tratados por estes autores como formas nominais, possuem resultados semelhantes nesses trabalhos. Em Bambuí, conforme pode ser visualizado na Tabela 1, os nomes gerais ocupam a terceira colocação e, em Coimbra, a quarta. Esse é um indício de que os nomes gerais, ainda que não sejam os preferíveis nem pelos falantes de Coimbra nem pelos falantes de Bambuí, são, sem dúvidas, formas usuais no PB e no PE.

Em Bambuí encontraram-se oito nomes gerais em construções de indeterminação, sendo eles *povo*, *pessoas*, *pessoa*, *peçoal*, *cara*, *fulano*, *homem* e *ser humano*. A ordem de ocorrência desses itens é demonstrada na Tabela 3, a seguir.

Tabela 3 – Ocorrências dos nomes gerais no corpus de Bambuí

Nome Geral	Número de ocorrências	Porcentagem (%)
Povo	56	29,3
Pessoas ¹⁷	42	21,9
Pessoa	39	20,4
Peçoal	37	19,4
Cara	6	3,2
Fulano	5	2,6
Homem	3	1,6
ser humano	3	1,6
TOTAL	191	100,0

Fonte: Dados da autora.

Já em Coimbra, encontraram-se cinco nomes em construções de indeterminação, que são *pessoas*, *pessoa*, *peçoal*, *povo* e *gente* e que estão elencados na Tabela 4:

Tabela 4 – Ocorrências dos nomes gerais no corpus de Coimbra

Nome Geral	Número de ocorrências	Porcentagem (%)
Pessoas	85	81%
Pessoa	11	10,5%
Peçoal	4	3,8%

¹⁷ Salienta-se que os itens *pessoa* e *pessoas* não foram contabilizados de forma conjunta porque, além de aparecerem com graus de indeterminação diferentes nos contextos analisados, neste estudo, considera-se o fator mais ou menos coletivo como influenciador dos nomes gerais e, portanto, não seria possível analisar esses dados conjuntamente.

Povo	3	2,8%
Gente	2	1,9%
TOTAL	105	100,0

Fonte: Dados da autora.

É importante destacar que, apesar de os nomes gerais refletirem um resultado equiparado nas duas localidades analisadas e em outros trabalhos realizados, em Coimbra, observa-se que o nome geral utilizado de fato e que é responsável por praticamente todo o resultado, isto é, 85 das 105 ocorrências de nomes gerais, é *pessoas*. Itens como *cara*, *indivíduo* e *fulano* não foram encontrados. Nos resultados de Bambuí, vê-se que os informantes utilizam uma variedade maior de nomes gerais e de forma mais equiparada. Essa discussão será mais bem desenvolvida na comparação dos resultados.

4.1.4 Nós/nóis/-mos

A variante *nós/nóis* e a desinência modo temporal *-mos* foi a quarta forma de se indeterminar em Bambuí mais recorrente, com 134 ocorrências, distribuídas da seguinte forma: *nós* ocorre 68 vezes, seguido da desinência modo temporal *-mos* com 44 dados e, finalmente, *nóis* com apenas 22 ocorrências. Essa variante não ocupa posições muito altas nos trabalhos já citados, exceto no de Oliveira (2006), em que *nós* ocupa o segundo lugar no conjunto de suas variantes, sendo a pesquisa dessa autora voltada para uma comparação entre o PB e o PE. A autora conclui que *nós*, *a gente* e *se + infinitivo* são estratégias empregadas tanto no PB quanto no PE, mas que a frequência em que essas estratégias ocorrem são diferentes. Enquanto no PB há preferência por *a gente*, no PE as estratégias *nós* e *a gente* são igualmente favorecidas.

Em Coimbra, conforme já foi demonstrado, a forma *nós* e a desinência modo temporal *-mos* foi a variante mais recorrente, com 451 ocorrências, totalizando quase 50% dos dados obtidos nesta localidade. É importante destacar que a forma *nós* ocorre 259 vezes e a desinência modo temporal *-mos*, 192. Assim, pode-se concluir até o momento que as formas que designam primeira pessoa do plural – *nós* e *a gente* – são as formas de indeterminação preferíveis por portugueses e brasileiros.

4.1.5 *Eles/ês*

Eles/ês foi considerada uma maneira de indeterminar o sujeito por Milanez (1982), Rollemberg *et al.* (1991), Setti (1997), Godoy (1999), Ponte (2008) e Souza (2014). Neves (2000), ao tratar desse tema, destaca que a forma *eles* possui uma indeterminação parcial por se referir ao universo, somente, das terceiras pessoas, mas, ainda assim, considera a forma como indeterminadora. Além desses estudos, Souza (2007) também faz um estudo detalhado sobre tal pronome e conclui que *eles* é um recurso natural de indeterminação em Belo Horizonte. Em Bambuí, essa forma totalizou 122 dados. Salienta-se que Vitral e Ramos (2006) observaram que os homens favorecem a forma *ês*, semelhante aos dados encontrados em Bambuí. Em Coimbra, ela não se mostrou muito produtiva, ocorrendo, apenas, 21 vezes. A partir do estudo de Vitral e Ramos (2006), é possível levantar a hipótese de que os dados do PB reforçam uma cliticização que não se mostra tão avançada no PE, visto que no PE a terceira pessoa do plural sem o sujeito explícito foi muito recorrente, demonstrando usos diferentes dessas formas nas cidades pesquisadas.

4.1.6 *Verbo na terceira pessoa do singular sem sujeito explícito (Ø+V3PS)*

A variante $\emptyset+V3PS$ mostrou-se bastante produtiva nos trabalhos de Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Ponte (2008) e Souza (2014) e, dessa forma, foi selecionada para este trabalho. Nos estudos anteriores, essa forma ocupa, nas pesquisas no Brasil, em média, a segunda ou terceira posição em número de ocorrência das variantes selecionadas, suplantando, muitas vezes, os nomes gerais. Em Bambuí, coincidentemente, ela ocupou a quinta posição no conjunto das variantes (122 ocorrências), juntamente com a forma *eles/ês*. Em Coimbra, essa forma se mostrou ainda menos produtiva que em Bambuí, com apenas 13 ocorrências.

4.1.7 *Verbo na terceira pessoa do plural sem sujeito explícito (Ø+V3PP)*

A forma $\emptyset+V3PP$, apresentada pela gramática tradicional (CUNHA; CINTRA, 2008) como uma das maneiras de se indeterminar o sujeito em português, mostra-se pouco produtiva em nossos dados do português de Bambuí. Ela conta com apenas 39 ocorrências, reforçando a

ideia de que o português apresentado pela gramática tradicional e aquele utilizado na oralidade distanciam-se cada vez mais.

Nota-se que, mesmo em trabalhos mais antigos, como o de Milanez (1982), Menon (1994), Setti (1997) e Godoy (1999), essa forma já não era muito frequente nem mesmo em dados de pessoas com escolaridade de nível superior. Em estudos mais recentes, como o Ponte (2008), Carvalho (2010) e Souza (2014), essa pouca frequência da variante também é reforçada. Vê-se, portanto, como a língua, mesmo nas décadas de 80 e 90, já fazia uso de outras formas de indeterminação além daquelas previstas pelas gramáticas tradicionais. Vale lembrar que a forma *eles/ês* foi produtiva nos dados de Bambuí, o que parece nos demonstrar que talvez essa forma tenha entrado em desuso na língua, porque agora os brasileiros preenchem o sujeito em vez de suprimi-lo, como demonstrado nos trabalhos sobre indeterminação citados anteriormente que relatam um número considerável de ocorrência de *eles*.

No entanto, o português europeu parece ter seguido o caminho inverso no caso dessa variante, pois, como pode ser visto na Tabela 2, $\emptyset+V3PP$ foi a terceira variante mais recorrente em Coimbra, com 127 dados. Nota-se, assim, que os portugueses parecem preferir as formas canônicas de indeterminação, enquanto que os brasileiros têm preferido outros tipos de variantes. Isso é reforçado por Duarte, Kato e Barbosa (2001) quando estes afirmam que no PE ocorre o inverso do PB com o favorecimento da utilização do *se*.

4.1.8 Verbo no infinitivo sem sujeito explícito ($\emptyset+VINF$)

A variante $\emptyset+VINF$ foi considerada como indeterminadora do sujeito apenas nos trabalhos de Milanez (1982), Carvalho (2010) e Oliveira (2018). Contudo, optou-se aqui por selecioná-la, porque ela demonstrou que seu uso pelos falantes é muito próximo no PB e no PE. Dos trabalhos em que ela foi analisada, apenas em Carvalho (2010) $\emptyset+VINF$ se mostrou mais produtiva, contando com 171 ocorrências no *corpus* analisado. Em Milanez (1982), essa variante ocorre 30 vezes e, em nosso *corpus* de Bambuí, 36 vezes. Em Coimbra, foram encontradas 21 ocorrência dessa forma, o que parece demonstrar uma equiparidade desse item nos dados coletados em Bambuí e em Coimbra.

4.1.9 Se

A variante *se* pode ser considerada a forma de indeterminação por excelência, pois, nos contextos em que há sua ocorrência, dificilmente é possível recuperar o referente, conforme vê-se nos exemplos retirados do nosso *corpus* “ouviam-se mais o fado do que se ouve agora”; “é criada-se a falsa sensação de/de que tá tudo ruim”. Ademais, Cunha e Cintra (2008), assim como outros gramáticos, relatam que existem apenas duas formas de indeterminar o sujeito, com verbos na terceira pessoa do plural ou na terceira pessoa do singular mais o pronome *se*, ou seja, pela análise tradicional, as formas elencadas aqui sequer seriam consideradas como indeterminadas. No entanto, o que pode ser visto até o momento por todos os estudos que já discutem essa questão é que os falantes têm utilizado com muita frequência outras formas de indeterminação.

No Brasil, o *se*, também pesquisado por Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010), Souza (2014) e Oliveira (2018), mostrou-se mais produtivo nos trabalhos mais antigos, tais como o de Milanez (1982) e Menon (1994). Nos trabalhos mais recentes no Brasil, a forma não é muito utilizada.

Salienta-se que isso parece acontecer apenas no PB, pois em Coimbra a forma *se* tanto anterior quanto posterior ao verbo foi bastante recorrente, ocorrendo 151 vezes no *corpus*. Assim, é possível notar mais uma diferença entre o PB e o PE: enquanto o *se* encontra-se em segundo lugar em número de ocorrências no PE, no PB, ele ocupa a última. Esse resultado corrobora o que Duarte, Kato e Barbosa (2001) já haviam encontrado, de que no PB a forma *se* tende a ser suprimida, já que se prefere a utilização de formas pronominais, o que não ocorre no PE, que favorece o uso de *se*.

4.1.10 *Tu*

A variante *tu* não se mostrou produtiva em nossos dados, uma vez que em Bambuí não foram encontradas ocorrências e em Coimbra, apareceu apenas seis vezes, mesmo assim, quatro delas utilizadas por apenas uma informante. Dos trabalhos que pesquisaram o tema apenas Setti (1997) e Godoy (1999) encontraram esses itens em suas pesquisas. Contudo, vale lembrar que Setti (1997) utilizou um *corpus* coletado em Curitiba, Florianópolis e Porto Alegre e Godoy (1999) pesquisou o tema nas cidades de Irati, Londrina e Pato Branco, ou seja, todas localizadas no sul do Brasil, onde, de fato, essa variante tem se mostrado mais produtiva.

Franceschini e Loregian-Penkl (2015) pesquisaram o uso de *tu* e *você* em cinco cidades catarinenses (Florianópolis, Blumenau, Lages, Chapecó e Concórdia) e em quatro cidades

gaúchas (Porto Algre, Panambi, São Borja e Flores da Cunha). Costa (2016) pesquisou o mesmo tema em Cametá, no estado do Pará, assim como Vítório (2018), em Maceió, Alagoas. Todos eles concluíram que a variante *tu* é mais utilizada quando o falante tem mais intimidade com seu interlocutor, ou em casos em que há uma referência específica de quem se fala, além de a variante ser mais utilizada por mulheres. Sendo assim, ainda que essa variante tenha sido pouco recorrente em nosso *corpus*, pôde-se observar que, em nossos dados, também houve predominância do uso de *tu* por mulheres e, ainda, como os informantes não tinham intimidade com a entrevistadora, esse pronome pode ter sido evitado.

Nas próximas seções, são feitas as análises dos dados.

4.2 Análise dos fatores selecionados estatisticamente pelo programa GoldVarb X em Bambuí

Este trabalho, utilizando-se do *software* Goldvarb X, optou por fazer uma rodada binária contendo os nomes gerais de um lado e, de outro, as demais formas de indeterminação. Além disso, como o foco desta pesquisa são os nomes gerais, as explicações mais detalhadas terão como base esses itens. As demais formas indeterminadoras serão apresentadas como forma de comparação.

Os grupos de fatores selecionados estatisticamente como favorecedores dos nomes gerais em Bambuí são: a) presença ou ausência de anáfora; b) grau de indeterminação e c) sexo. Essa ordem foi a selecionada pelo programa e, dessa maneira, será a mesma em que os fatores serão discutidos detalhadamente. Os demais fatores que não foram selecionados pelo *software* serão analisados qualitativamente na seção 4.3.

4.2.1 Presença ou ausência anáfora

De acordo com a subseção 3.11.1, em que se discute a variável linguística *presença ou ausência de anáfora*, selecionaram-se três maneiras de se classificar os itens no *corpus* para este grupo: a) retomada anafórica com o mesmo elemento; b) ausência de anáfora e c) retomada anafórica com outro elemento. Entretanto, quando os dados foram rodados no *software* Goldvarb X, houve um *Knockout* nessa variável e, dessa forma, reagrupou-se a variável *retomada anafórica com o mesmo elemento* com *retomada anafórica com outro elemento*. De acordo com o que foi explicitado na seção 3.12, optou-se pelo reagrupamento, porque o número

de dados era relativamente alto. Feita essa junção, obtêm-se os resultados expostos na Tabela 5:

Tabela 5 – Influência do fator presença ou ausência de anáfora na indeterminação do sujeito por meio de nomes gerais em Bambuí

Presença ou ausência de anáfora	Número de ocorrências	Porcentagem (%)	Peso relativo
Ausência	174/949	18,3	0,62
Presença	17/408	4,2	0,25
Total	191/1357	14,1/100,0	

Fonte: Dados da autora.

De acordo com a Tabela 5, observa-se que, no caso dos nomes gerais, predominantemente, quando um item é utilizado, não há uma retomada com esse mesmo item. Tal resultado não era esperado, visto que, em Carvalho (2010), por exemplo, a presença é mais frequente do que a ausência, uma vez que, quando um item é usado, ele tende a ser repetido.

A partir desses dados, pode-se inferir que, quando o falante utiliza o nome geral, predominantemente, esse nome está em uma oração absoluta, não havendo, portanto, uma oração que retome a anterior. Isso parece ser, portanto, o maior influenciador do resultado. Alguns exemplos podem ser visualizados a seguir:

(51) é o/o **povo** tem muita crença ainda igual teve São Sebastião tinha bastante gente né a missa (BAM07F)

(52) intão eu acridito assim que a maioria das pessoas aqui são bem receptivas propícias a tá/tá ajudano gente que vem de fora eu acridito que sim o **pessoal** é legal (BAM26M)

Outro fator que parece ter proporcionado tal resultado possivelmente tenha a ver com o fato de, no caso dos nomes gerais, por haver mais ocorrências de uma indeterminação parcial com um referente implícito, como em (51) e (52), não ser necessária uma retomada desse sujeito, visto que o uso de tal forma já objetiva indeterminar, ainda que algo possa remeter a quem se está falando. Salienta-se que tanto o item *povo*, em (51), quanto o item *pessoal*, em (52), fazem remissão às pessoas de Bambuí.

4.2.2 Grau de indeterminação

A segunda variável selecionada como favorecedora dos nomes gerais foi *grau de indeterminação*. De acordo com a subseção 3.11.2, a *indeterminação parcial com referente explícito* se dá quando existe alguma palavra, expressa pelo informante, que torna possível retomar de quem se está falando. A *indeterminação parcial com referente implícito* ocorre, no entanto, quando se recupera de quem se está falando pelo contexto. A *total*, contrariamente às outras, não torna possível a recuperação. A Tabela 6 apresenta os resultados obtidos a partir do *corpus*:

Tabela 6 – Influência do fator grau de indeterminação na indeterminação do sujeito por meio de nomes gerais em Bambuí

Grau de indeterminação	Número de ocorrências	Porcentagem (%)	Peso relativo
Indeterminação total	59/426	13,8	0,51
Parcial com referente implícito	112/673	16,6	0,55
Parcial com referente explícito	20/258	7,8	0,36
Total	191/1357	14,1/100,0	

Fonte: Dados da autora.

A partir da Tabela 6, nota-se um leve favorecimento da indeterminação total e da parcial com referente implícito nos nomes gerais. Salienta-se que Carvalho (2010) encontrou que a indeterminação completa é favorecedora das formas nominais.

Observa-se, a partir dos dados, que formas indeterminadoras como *a gente, você e nós* parecem incluir mais o falante ou o interlocutor do que os nomes gerais, conforme exemplos (53) e (54) abaixo:

(53) eu acho que a gente cria né um/um laço com a comunidade onde **a gente** tá intão acho que isso é importante (BAM02M)

(54) intão quer dizer ela é ruim e boa... ela pode algumas pessoas influencia[r] mais num é tanto porque as pessoas que tem que influencia igual cê vê os presi/os presídio o **cara** vai preso ele sai da cadeia pio[r] (BAM08F)

Vê-se que, com a forma *a gente*, na maioria dos casos, sempre haverá a inclusão de quem se fala, como em (53), em que o informante descreve as relações que vão se criando quando se vive em determinada comunidade ou ainda os laços que ele mesmo criou. No caso dos nomes gerais, como em (54), ou mesmo aqueles que representem um grupo de pessoas, tal como *povo*, *pessoa(s)* ou *pessoal*, isso tende a ocorrer com frequência menor nos dados coletados. Posio (2021) relata, por exemplo, o processo de gramaticalização que esses itens estão sofrendo, o que possivelmente esteja contribuindo para o esvaziamento de sentido desses itens.

4.2.3 Sexo

De acordo com os dados apresentados na Tabela 7, as mulheres, diferentemente dos trabalhos de Carvalho (2010) e de Souza (2014), favoreceram o uso dos nomes gerais.

Tabela 7 – Influência do fator sexo na indeterminação do sujeito por meio de nomes gerais em Bambuí

Sexo	Número de ocorrências	Porcentagem (%)	Peso relativo
Feminino	96/572	16,8	0,55
Masculino	95/785	12,1	0,46
Total	191/1357	14,1/100,0	

Fonte: Dados da autora.

Além disso, esse favorecimento pode ser explicado pelo uso do nome geral *povo*. Este item, o mais recorrente dos nomes gerais, foi realizado 40 vezes por mulheres, sendo que seu total de ocorrência é 57. Nota-se, ainda, que a informante BAM07F foi responsável por realizar 10 das 40 realizações de *povo* (55). Nos demais itens, observou-se certa equiparação, exceto em *pessoal*, que, dos 38 dados, apenas seis foram realizados por falantes do sexo feminino (56). Dos três dados do item *homem*, os três foram realizados por mulheres (57) e, das três ocorrências de *ser humano*, as três foram realizadas por homens (58), conforme exemplificado a seguir:

(55) tinha muita compra de voto né ainda tem mais hoje o **povo** não é tão bobo né porque votava a troca de dez reais acho que o **povo** hoje já pensa antes de pega[r] o dinhe[i]ro (BAM07F)

(56) antes o **peçoal** ia pra istuda[r] e dificilmente voltava né (BAM06M)

(57) quanta coisa que o **homem** caminhô? quanta coisa que o **homem** descobriu? quantas coisas maravilhosas? e o próprio homem eles istá é... não istá dando valor naquilo que ele próprio con/construiu intão eu tenho a isperança de que haverá di/dias melhor acontecer (BAM12F)

(58) eu acho isso um absurdo ixtremo essa questão de preconceito uma coisa que não deveria te[r] nunca/nun/nunca em hipótese alguma em relação a nada/nada mais tá inraizado no ser humano... o **ser humano** é/é complicado né... meu/meu professor fala o **ser humano** é uma desgraça e eu acho que nunca vai acaba[r] (BAM13M)

A partir desses resultados, não seria correto afirmar que Bambuí contradiz trabalhos que mostram como as mulheres tendem a utilizar as formas mais tradicionais, pois, com esse pequeno favorecimento, não seria possível apontar uma conclusão tão assertiva. Salienta-se que a informante que realizou 10 das 40 ocorrências de *povo* tem 30 anos de idade, sendo possivelmente sua juventude a justificativa para o uso frequente desse nome geral. Acrescenta-se, ainda, que essa informante praticamente não utilizou outros nomes gerais.

Tem-se, como em Carvalho (2010), que o fator *sexo* foi o único selecionado para a realização de nomes gerais em Bambuí, ainda que nosso resultado não seja semelhante ao encontrado pelo autor. Ressalta-se que os pesos relativos encontrados por Carvalho (2010) foram muito próximos aos obtidos nesta pesquisa.

Um dado relevante sobre as demais formas de indeterminação quando comparadas com os nomes gerais se refere ao fato delas se mostrarem diferentes no que diz respeito ao fator *sexo*, uma vez que as demais formas foram mais utilizadas por homem do que por mulheres, em números absolutos. Contudo, de forma geral, observa-se que em Bambuí ocorreu uma distribuição bastante equiparada de homens e mulheres utilizando as demais formas de indeterminação do sujeito, bem como fazendo uso dos nomes gerais.

A seguir, apresentam-se os resultados obtidos por meio de uma análise qualitativa dos fatores não selecionados pelo GoldVarb X.

4.3 Análise qualitativa dos fatores que não foram selecionados pelo programa GoldVarb X em Bambuí

Nesta seção são apresentados e discutidos qualitativamente os fatores a) faixa etária; b) presença ou ausência de locativo e c) mais ou menos coletivo. Conforme explanado na seção

4.2, esses fatores não foram selecionados pelo GoldVarb X, contudo, acredita-se que eles possam trazer contribuições para esta e para futuras pesquisas sobre o tema, motivo pelo qual serão brevemente discutidos. Por outro lado, salienta-se que esses fatores tiveram uma distribuição bastante equiparada e, portanto, não será possível apresentar conclusões assertivas. Além disso, as porcentagens são dadas em relação ao conjunto total de dados devido ao objetivo principal da pesquisa, que é estudar os nomes gerais no conjunto das variantes indeterminadoras.

4.3.1 Faixa etária

O fator faixa etária também se mostrou muito equiparado quanto aos resultados obtidos, conforme Tabela 8:

Tabela 8 – Influência do fator faixa etária na indeterminação do sujeito por meio de nomes gerais em Bambuí

Faixa etária	Número de ocorrências	Porcentagem (%)
18-25 anos	39/271	14,4
30-45 anos	84/564	14,9
Acima de 50 anos	68/522	13,0
Total	191/1357	14,1/100,0

Fonte: Dados da autora.

O fator faixa etária, considerando os dados em porcentagem, mostra resultados que já eram esperados inicialmente, ainda que a diferença entre as porcentagens seja pequena, já que a expectativa era de que as faixas mais novas fariam maior uso dos nomes gerais. Conclui-se, portanto, que os nomes gerais possuem grande espaço nas formas que indeterminam o sujeito e que o fato de as faixas etárias mais jovens mostrarem maior uso desses nomes é um indício de que eles devem permanecer na língua.

4.3.2 Presença ou ausência de locativo

O fator presença ou ausência de locativo, com 14,7% e 13,3%, respectivamente, mostra, de acordo com a Tabela 9, uma confirmação de nossa hipótese inicial de que o falante faz uso de alguns mecanismos para marcar sobre onde se está falando.

Tabela 9 – Influência do fator presença ou ausência de locativo na indeterminação do sujeito por meio de nomes gerais em Bambuí

Presença ou ausência de locativo	Número de ocorrências	Porcentagem (%)
Presença	113/769	14,7
Ausência	78/588	13,3
Total	191/1357	14,1/100,0

Fonte: Dados da autora.

Esse maior uso da *presença de locativo* pode estar ligado ao fato de a entrevista tratar de perguntas sobre o Brasil e sobre a cidade do informante, o que faz com que o falante utilize locativos para fazer referência à localidade sobre a qual se trata, conforme exemplo (59). No caso do exemplo (60), não se tem o locativo explícito, no entanto, é possível recuperar, por meio do contexto, que se trata de pessoas de Bambuí:

(59) sai... por isso queês cortô o carnaval essas coisa assim... porque o **povo** num sabe diverti[r] **aqui** só sabe briga[r] (BAM25F)

(60) antes o **peçoal** ia pra istuda[r] e dificilmente voltava né (BAM06M)

No exemplo (59), a presença do locativo *aqui* reforça que a informante está falando de eventos da própria cidade. Já em (60), não consta nenhum locativo, contudo, pelo contexto da entrevista, é possível fazer tal recuperação. Assim, ambas as formas foram consideradas como *presença de locativo*.

Dessa forma, como os resultados são muito semelhantes, pode-se levantar duas hipóteses para os resultados obtidos. A primeira delas é que o roteiro criado tenha, de alguma forma, influenciado o uso desses locativos, ou, como segunda hipótese, que o falante, quando indetermina, tenta, em boa parte dos casos, deixar alguma marca do local sobre o qual se está falando. Isso talvez tenha se dado por causa dos laços que os falantes criam com determinada comunidade, assim como pontuado por um dos informantes no exemplo (61):

(61) eu acho que a gente cria né um/um laço com a comunidade onde **a gente** tá intão acho que isso é importante (BAM02M)

4.3.3 Mais ou menos coletivo

Conforme já apresentado anteriormente, consideraram-se como *mais coletivo* as formas indeterminadoras que possuíam uma pluralidade interna. Na Tabela 10, vê-se um maior uso das formas mais coletivas:

Tabela 10 – Influência do fator mais ou menos coletivo na indeterminação do sujeito por meio de nomes gerais em Bambuí

Mais ou menos coletivo	Número de ocorrências	Porcentagem (%)
Mais coletivo	135/948	14,2
Menos coletivo	56/409	13,7
Total	191/1357	14,1/100,0

Fonte: Dados da autora.

A partir dos dados apresentados, entre os nomes gerais, há uma preferência pelo uso de formas mais coletivas, enquanto, nas demais formas, ocorre o contrário. Isso pode ser justificado pelo nome geral mais recorrente no *corpus ser povo*, forma esta classificada como mais coletiva. Nesse sentido, esse fator apontou resultados muito equiparados, contudo, eles também podem servir para futuras comparações em outros trabalhos.

A seguir, apresenta-se a análise dos resultados obtidos com as gravações coletadas em Coimbra.

4.4 Análise dos fatores selecionados estatisticamente pelo programa GoldVarb X de Coimbra

Ao se tratar da rodada realizada com os dados de Coimbra, quatro fatores foram selecionados como favorecedores dos nomes gerais, sendo eles: a) presença ou ausência de locativo; b) faixa etária; c) grau de indeterminação e d) mais ou menos coletivo. A seguir, são apresentadas as discussões desses resultados.

4.4.1 Presença ou ausência de locativo

O fator presença ou ausência de locativo, conforme explicitado anteriormente, tem o objetivo de analisar se um falante, quando utiliza uma forma de indeterminação, busca marcar, de alguma maneira, sobre qual local se está falando. Dessa forma, no caso dos nomes gerais,

obteve-se que os informantes de Coimbra, quando utilizam um nome geral, não utilizam recursos de marcação para demonstrar sobre qual local se está falando (TABELA 11). Esse resultado pode servir como reforço para demonstrar o quanto os nomes gerais são capazes de indeterminar, além de evidenciar que os informantes sequer usam locativos para marcar o local de onde se está falando quando utilizam esses itens.

Tabela 11 – Influência do fator presença ou ausência de locativo na indeterminação do sujeito por meio de nomes gerais em Coimbra

Presença ou ausência de locativo	Número de ocorrências	Porcentagem (%)	Peso relativo
Presença	45/586	7,7	0,45
Ausência	60/354	16,9	0,58
Total	105/940	11,2/100,0	

Fonte: Dados da autora.

Para uma discussão mais completa dos resultados, apresenta-se, a seguir, o exemplo (62), em que se tem um caso de ausência de locativo, e em (63), de presença.

(62) e as **p[e]ssoas** andam muito de cara fechada... andam... já não são assim tão... (COI29F)

(63) sim ... Portugal é um país ... entre política e futebol ... se for preciso um jogo importante o **pessoal** não vai votar (COI18M)

A partir dos exemplos apresentados, percebe-se contextos semelhantes, em que se fala dos portugueses em geral, de como eles são, de como é a vida em Coimbra ou mesmo no país. O informante 29 faz um comentário de como as pessoas hoje estão mais fechadas, reservadas, que um pouco da alegria e das conversas na rua têm se perdido. Ele não faz uso de nada que marque, no período, de qual local são essas pessoas, podendo se tratar, portanto, de portugueses ou das pessoas do mundo.

No exemplo (63), o informante 18 fala também do comportamento das pessoas, de como entretenimentos, como futebol, têm sido vistos como mais importantes do que a política, que, do ponto de vista do informante em questão, é extremamente relevante para a vida em sociedade e que, portanto, deveria ser objeto de interesse de todos. Vale ressaltar que o entrevistado deixa claro que está tratando de portugueses; por isso, essa ocorrência foi considerada uma indeterminação parcial.

Tem-se, então, que o fator presença ou ausência de locativo, ainda pouco estudado, serve para reforçar a ideia de que a indeterminação pode ser dividida em graus, conforme se defende aqui, pois, de acordo com o que foi demonstrado nos exemplos, têm-se momentos em que os informantes marcam de onde estão falando, mas isso não invalida o fato de ser um sujeito indeterminado, uma vez que não é possível precisar-se sobre quem se está falando.

4.4.2 Faixa etária

O segundo fator selecionado em Coimbra foi faixa etária, conforme já apresentado anteriormente. Diferentemente dos resultados encontrados no Brasil, tal fator mostra-se relevante para os nomes gerais. Como é possível notar na Tabela 12, o resultado, previsível, foi de que os jovens usam mais os nomes gerais do que os informantes da segunda faixa, que, por sua vez, usam mais do que os mais velhos, que são os que menos utilizam nomes gerais. Salienta-se que, no estudo de Barbosa (2016), há uma parte dedicada à discussão do que ela considera como formas nominais e, apesar de a autora não ter obtido pesos relativos, sua análise demonstrou que, no falar ludovicense, considerando-se as porcentagens, os mais jovens utilizaram os nomes gerais (40,2%), seguidos da segunda faixa etária (31,5%) e posteriormente da terceira (28,3%). Nos demais trabalhos encontrados, esse fator não apresenta resultados significativos.

Tabela 12 – Influência do fator faixa etária na indeterminação do sujeito por meio de nomes gerais em Coimbra

Faixa etária	Número de ocorrências	Porcentagem (%)	Peso relativo
18-25 anos	41/242	16,9	0,63
30-45 anos	44/432	10,2	0,49
Acima de 50 anos	20/266	7,5	0,40
Total	105/940	11,2/100,0	

Fonte: Dados da autora.

Assim, no caso dos portugueses, é perceptível a influência do fator faixa etária no uso de nomes gerais. Sobre o uso mais recorrente dos itens pelos jovens, pode-se levantar duas hipóteses: a de que esses itens estão chegando agora no português europeu como forma de indeterminação ou a de que os mais velhos têm alguma resistência às formas inovadoras. Como os nomes gerais não demonstraram ser formas estigmatizadas em nenhum dos trabalhos

encontrados sobre o tema, talvez a primeira hipótese seja a mais coerente, contudo, como não foram encontrados estudos sobre o tema no português europeu, deixa-se aqui a possibilidade para futuras pesquisas.

4.4.3 Grau de indeterminação

Na seção 3.12, apontou-se que o fator *grau de indeterminação* sofreu uma junção para que pudesse se obter os pesos relativos, tendo em vista que houve um *Knockout*. Sendo assim, as formas de indeterminação parcial, sejam elas explícitas ou implícitas, foram agrupadas, ficando o fator subdividido apenas em indeterminação total ou parcial. Na Tabela 13, pode-se visualizar que, no caso de Coimbra, a indeterminação total favorece o uso de nomes gerais. Esse resultado é de extrema valia, pois reforça como os nomes gerais têm sido usados quando não se quer ou não se pode recuperar o referente, assim como as formas de indeterminação canônicas (verbo na terceira pessoa do plural sem sujeito explícito e verbo na terceira pessoa do singular + *se*).

Tabela 13 – Influência do fator grau de indeterminação na indeterminação do sujeito por meio de nomes gerais em Coimbra

Grau de indeterminação	Número de ocorrências	Porcentagem (%)	Peso relativo
Indeterminação total	58/349	16,6	0,65
Parcial com referente implícito	47/591	8,0	0,41
Total	105/940	11,2/100,0	

Fonte: Dados da autora.

Os exemplos (64), (65) e (66) ilustram o que foi classificado como indeterminação parcial com referente implícito, indeterminação parcial com referente explícito e indeterminação total, respectivamente:

(64) e **a gente** esteve portanto naquele chão ali em cima toda a gente tinha ali cordas que se estendia ali em cima... (COI11F)

(65) **nós** não temos aumentus a função pública foi p[e]nalizada ao máximu... porque (ninguém... era aos nossos descontus... né? e o aumentu dos nossos vencimentus...tantu e **nós** ficamus estagnados (COI12F)

(66) se a **pessoa** não for de acordo com as nossas expectativas... não tem que estar... (COI19F)

No exemplo (64), tem-se um caso de *a gente*, forma esta que normalmente inclui quem está falando e, portanto, na maioria dos nossos casos, não apresenta uma indeterminação total. Em (65), vê-se um caso de uma indeterminação parcial com referente explícito, porque a pessoa específica que está falando sobre os servidores públicos, isto é, um grupo restrito de pessoas. Salienta-se que, no conjunto de nomes gerais, não foram encontradas ocorrências com referência explícita. Por fim, no exemplo (66), é possível visualizar uma indeterminação total, ou seja, um caso de indeterminação por excelência, na qual não se consegue identificar nenhum tipo de referência sobre quem se está falando.

4.4.4 Mais ou menos coletivo

O último fator selecionado como favorecedor dos nomes gerais foi o *mais ou menos coletivo* e, como pode ser visualizado na Tabela 14, houve um favorecimento de itens mais coletivos no grupo de nomes gerais.

Tabela 14 – Influência do fator mais ou menos coletivo na indeterminação do sujeito por meio de nomes gerais em Coimbra

Mais ou menos coletivo	Número de ocorrências	Porcentagem (%)	Peso relativo
Mais coletivo	94/773	12,2	0,55
Menos coletivo	11/167	6,6	0,28
Total	105/940	11,2/100,0	

Fonte: Dados da autora.

Conforme apontado anteriormente, o fator *mais ou menos coletivo* não foi selecionado como favorecedor dos nomes gerais em Bambuí, contudo, em Coimbra, o inverso ocorreu. Além disso, salienta-se que, de acordo com a Tabela 4, o nome geral *pessoas* é responsável por 81% de todo o grupo de nomes gerais encontrados no *corpus*, o que, com certeza, fez com que esse resultado acontecesse.

Sendo assim, acredita-se que esse resultado dá-se pela preferência do item *pessoas* como forma de indeterminação do sujeito em Coimbra, seguido de uma pouca variedade dos demais nomes gerais encontrados. A seguir, apresentam-se os exemplos (67) e (68) para ilustrar o que se denominou de mais coletivo:

(67) *comu é uma cidad[e] qui não é muito grand[e]... as p[e]ssoas acabam pur não gastar tantu em certas coisas... transport[e]... (COI38F)*

(68) *as p[e]ssoas gostam de se (enganar) (COI39M)*

Em (67) e (68), temos exemplos de nomes gerais classificados como mais coletivos. Em (67), tem-se um comentário sobre como, em cidades pequenas, como Coimbra, os habitantes não gastam muito dinheiro com transporte, por exemplo, o que os permite ter uma melhor qualidade de vida. Em (68), o informante pontua que as pessoas, de um modo geral, têm dificuldade de aceitar certas verdades sobre política, situação do país etc. e, por isso, gostam de se enganar.

A próxima seção tratará dos fatores linguísticos que não foram selecionados como favorecedores dos nomes gerais, isto é, *sexo e presença e ausência de anáfora*.

4.5 Análise qualitativa dos fatores que não foram selecionados pelo programa GoldVarb X em Coimbra

Nesta seção são apresentados e discutidos qualitativamente os fatores a) sexo e b) presença ou ausência de anáfora, conforme foi feito com os dados de Bambuí. Salienta-se, novamente, que as porcentagens são dadas em relação ao conjunto total de dados devido ao objetivo principal da pesquisa, que é estudar os nomes gerais no conjunto das variantes indeterminadoras.

4.5.1 Sexo

A Tabela 15 demonstra um resultado equiparado no uso de nomes gerais por homens e mulheres em Coimbra. No Brasil, esse fator também não traz resultados significativos, demonstrando que tanto homens quanto mulheres utilizam essas formas para indeterminar o sujeito. Em número de ocorrências, tem-se que as mulheres utilizaram mais essas formas, mas, como o fator não apresenta pesos relativos, não é possível se falar em preferências de mulheres por esses itens.

Dessa maneira, os resultados demonstram que tanto homens quanto mulheres fazem uso desses nomes de maneira bastante equiparada e que, portanto, esse fator não influencia o fenômeno em questão.

Tabela 15 – Influência do fator sexo na indeterminação do sujeito por meio de nomes gerais em Coimbra

Sexo	Número de ocorrências	Porcentagem (%)
Masculino	46/427	10,8
Feminino	59/513	11,5
Total	105/940	11,2/100,0

Fonte: Dados da autora.

4.5.2 Presença ou ausência de anáfora

A variável presença ou ausência de anáfora também não foi selecionada como favorecedora dos nomes gerais em Coimbra. Contudo, como demonstra a Tabela 16, os falantes, em Coimbra, optam por retomar os nomes gerais em um contexto de indeterminação primeiramente com outros elementos, seguido da ausência de forma de retomada e, por fim, com retomada com o mesmo elemento.

Tabela 16 – Influência do fator presença ou ausência de anáfora na indeterminação do sujeito por meio de nomes gerais em Coimbra

Presença ou ausência de anáfora	Número de ocorrências	Porcentagem (%)
Ausência	88/762	11,5
Retomada anafórica com o mesmo elemento	13/157	8,3
Retomada anafórica com outro elemento	4/21	19,0
Total	105/940	11,2/100,0

Fonte: Dados da autora.

Os exemplos (69), (70) e (71) demonstram ocorrências desse fenômeno:

(69) e às vez[e]s se não for o **povu** a falar um bucadinhu... a coisa não anda muito pra frent[e]...mas não/não ligo muito pra pulítica não (COI38F)

(70) as **peessoas** tem muito o hábito de sair a noite cá em Coimbra... é ...umas mais que nas outras faculdades... mas cá em Coimbra as **peessoas** saem muito... e... (COI19F)

(71) **nós** temos um presidente que está a fazer muitas obras que há ... é conhecido por fazer muitas (...) mas as **peçoas** também não percebem (...) pronto ... não tenho muito a informar sobre... (COI18M)

Em (69) tem-se um exemplo de ausência de anáfora, isto é, o nome geral *povo* é usado e depois não é retomado. Já em (70), tem-se a retomada com o mesmo elemento, o item *peçoas*, que é utilizado e depois retomado. Por fim, tem-se a retomada anafórica com outro elemento, que é quando se usa uma forma de indeterminação, no caso de (71), a forma *nós*, e depois a retomada é feita com outro elemento, nesse caso, *peçoas*. Assim, mesmo com este fator não sendo selecionado, é perceptível uma preferência pela não retomada anafórica, ainda que ela não seja a maior em porcentagem, demonstrando, assim, que um nome geral parece não ser retomado quando é utilizado.

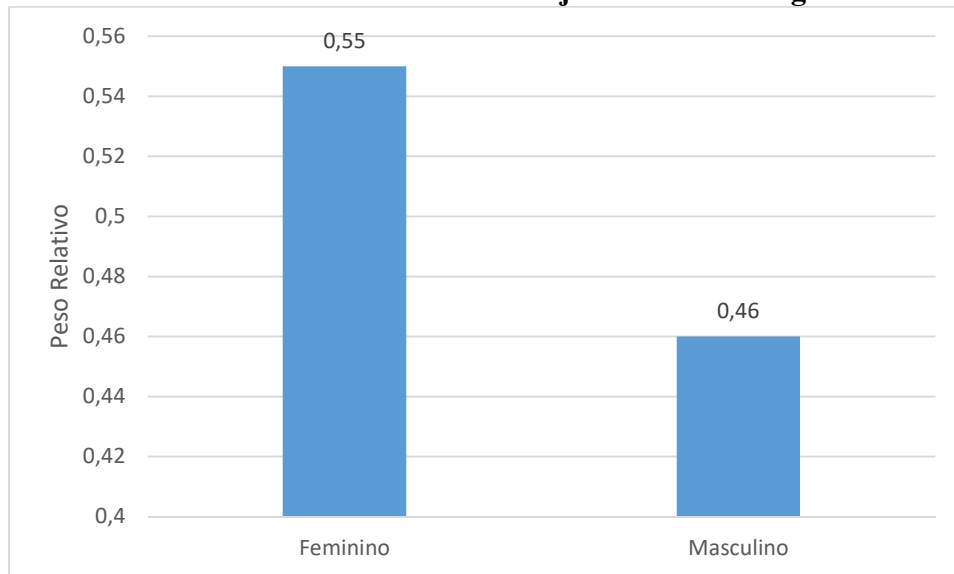
Na próxima seção, será apresentada uma comparação dos resultados de Bambuí e de Coimbra.

4.6 Comparação entre os resultados de Bambuí e de Coimbra

Nesta seção, conforme já foi indicado na 3.12, será feita uma comparação entre os resultados encontrados a partir das rodadas já apresentadas nas discussões anteriores. É importante lembrar que uma rodada equiparada foi feita, mas, como os resultados não apresentaram diferenças significativas dos encontrados inicialmente, optou-se por comparar as rodadas iniciais.

4.6.1 Sexo

O fator *sexo* foi selecionado como fator favorecedor dos nomes gerais em Bambuí, mas não em Coimbra. Em Bambuí, como pode ser visto no Gráfico 1, as mulheres utilizaram mais os nomes gerais como forma de indeterminação do sujeito do que os homens.

Gráfico 1 – Influência da variável sexo no conjunto dos nomes gerais em Bambuí

Fonte: Dados da autora.

Contudo, alguns pontos precisam ser lembrados, como o fato de que o nome geral *povo*, o mais recorrente nos dados de Bambuí, foi utilizado 40 vezes por pessoas do sexo feminino, sendo que o item teve 57 ocorrências em todo o *corpus*. Os demais itens demonstraram equiparação, exceto *pessoal*, que foi mais utilizado por homens. Além disso, esse resultado não corrobora o que foi encontrado por Carvalho (2010) e Souza (2014), mas, devido à proximidade dos resultados de homens e mulheres, não é possível afirmar que nossos resultados contradizem trabalhos anteriores, pois esse resultado pode ter se dado por causa da realização das repetidas vezes de um único nome geral.

Diferentemente de Bambuí, em Coimbra, *sexo* não foi uma variável selecionada, entretanto, a Tabela 15 demonstra que, em termos de porcentagem, tem-se um uso muito equiparado de nomes gerais por homens e mulheres. Sabe-se que não é possível comparar de forma idêntica as duas cidades, uma vez que em Bambuí obteve-se pesos relativos e em Coimbra não, mas pode-se afirmar a partir disso que não parece existir uma preferência ou rejeição pelo uso dos nomes gerais quando se analisa a variável *sexo*. Vale lembrar que os resultados de trabalhos apresentados ao longo deste estudo também chegaram a uma conclusão semelhante.

Assim, tanto no Brasil quanto em Coimbra, homens e mulheres utilizam os nomes gerais de maneira relativamente equiparada, ainda que em Coimbra este não tenha sido um fator selecionado como favorecedor desses nomes. Nesse sentido, nosso estudo consegue demonstrar que a variável *sexo* não é um fator que influencia diretamente o fenômeno em questão e que são formas usadas indistintamente por homens e mulheres. Isso pode ser justificado, porque

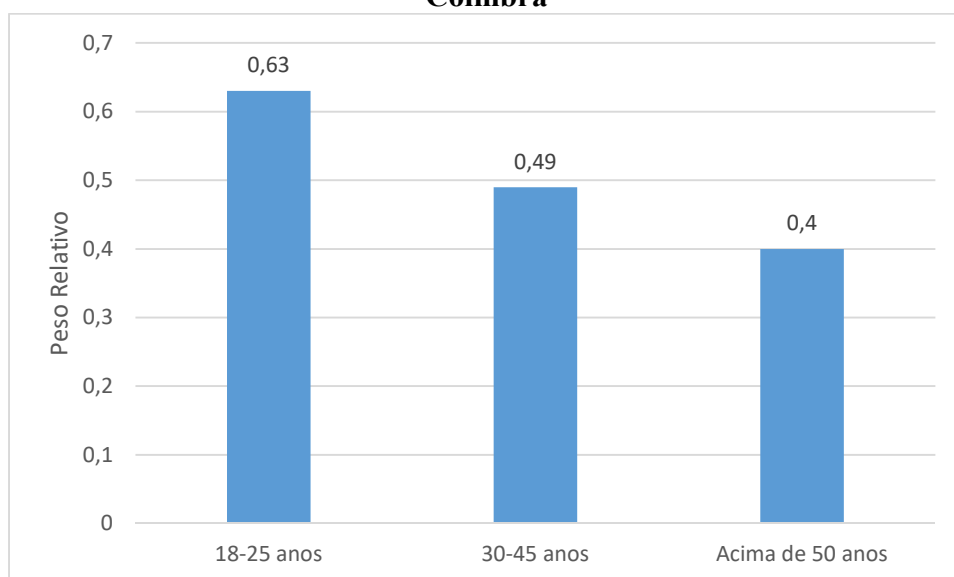
essas variantes não demonstram nenhum tipo de estigma e, conforme os estudos linguísticos, a exemplo de Freitag (2015), as diferenças de fenômenos linguísticos por homens e mulheres ao longo dos estudos da sociolinguística se dão por causa da distinção de papéis sociais entre eles, em que mulheres tendem a ser mais sensíveis a formas consideradas conservadoras. Contudo, como a indeterminação não é um fenômeno estigmatizado, esse já era inicialmente um resultado esperado.

4.6.2 Faixa etária

A faixa etária, neste estudo, conforme apresentado no Gráfico 2, foi dividida em três grupos: pessoas de 18 a 25 anos, pessoas de 30 a 45 anos e pessoas acima de 50 anos, com o intuito de contemplar gerações diferentes e, assim, poder trazer contribuições maiores ao fenômeno.

Essa variável foi selecionada como favorecedora dos nomes gerais em Coimbra, diferentemente do que ocorreu em Bambuí e em vários estudos sobre o tema no Brasil. De acordo com o Gráfico 2, é possível perceber uma preferência do uso dos nomes gerais em Coimbra pela faixa etária mais jovem, seguida da segunda faixa etária e, por fim, da última, que contempla os informantes mais velhos. Esse resultado é muito relevante, mas esperado, pois já se acreditava que os jovens fariam mais uso dessas formas.

Gráfico 2 – Influência da variável faixa etária no conjunto dos nomes gerais em Coimbra



Fonte: Dados da autora.

Em Bambuí, ainda que os pesos relativos não tenham sido obtidos, o resultado, em porcentagem, foi semelhante ao de Coimbra, com as faixas mais jovens utilizando mais os nomes gerais.

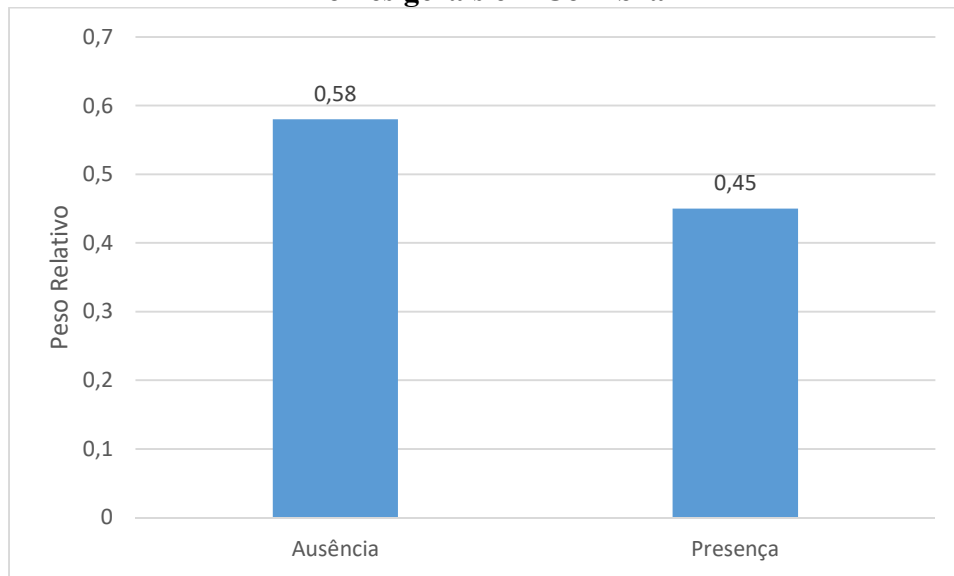
A partir da comparação das duas variáveis extralinguísticas selecionadas, pode-se observar que, tanto em Bambuí quanto em Coimbra, no caso de sexo, tem-se um resultado muito equiparado quanto à utilização dos nomes gerais por homens e mulheres. No que se refere à variável faixa etária, há uma preferência pelos mais jovens pelo uso dos nomes gerais.

Dessa maneira, a preferência dos mais jovens por utilizarem os nomes gerais demonstra um indício de que eles permanecerão na língua. Além disso, tanto em Bambuí quanto em Coimbra, as faixas mais velhas também utilizaram esses itens. Isso reforça o que Bortoni-Ricardo (2014) apresenta sobre os papéis ocupados pelos membros das comunidades tendo em vista a faixa etária e como esse fator é capaz de avaliar casos de mudança em progresso. Assim, como em Coimbra tem-se mais jovens utilizando os nomes gerais, podemos trazer à tona a possibilidade de uma permanência desses itens na língua e de uma possível mudança nas formas de indeterminação, com os nomes gerais ocupando cada vez mais espaço no conjunto das variantes. Essa hipótese, entretanto, deverá ser testada em pesquisas futuras.

4.6.3 Presença ou ausência de locativo

O fator presença e ausência de locativo foi selecionado apenas em Coimbra, conforme mostra o Gráfico 3, tendo como favorecedor a ausência de locativo no conjunto dos nomes gerais. De acordo com o que já foi apresentado, esse fator ocorre quando um período contém alguma forma de indeterminação do sujeito e o informante utiliza marcas como *aqui* ou alguma outra marcação que torne possível a localização do lugar do qual se está falando. Em Bambuí, esse foi um fator não selecionado, mas, assim como foi apresentado pela Tabela 9, tem-se que a presença (14,7%) ocorreu mais que a ausência (13%) quando comparadas as porcentagens.

Gráfico 3 – Influência da variável presença ou ausência de locativo no conjunto dos nomes gerais em Coimbra



Fonte: Dados da autora.

No Gráfico 3, pode-se ver uma diferença considerável desses contextos em que ocorre a presença ou ausência de locativo. Isso pode ter se dado porque a indeterminação total, outro grupo de fatores analisado, favorece a indeterminação por meio dos nomes gerais em Coimbra, o que não ocorre em Bamuí. Sendo assim, acredita-se que, pelo fato de os portugueses utilizarem os nomes gerais de forma a indeterminar totalmente sobre quem se está falando, eles não fariam qualquer referência ao lugar desses indivíduos. Uma outra hipótese possível é que o item *peessoas* foi o mais recorrente em Coimbra e, por ser um item mais comumente utilizado para se indeterminar o sujeito em português, esse resultado também tenha acontecido, conforme exemplo (72):

(72) as **peessoas** de uma maneira geral nunca estão satisfeitas (COI28M)

No exemplo retirado dos dados de Coimbra, o item *peessoas* é analisado como indeterminação total, tendo em vista que o informante está falando sobre as pessoas do mundo, de qualquer nacionalidade. Em (72), o informante explica que as pessoas de qualquer lugar nunca estão satisfeitas, independente do que se o ofereça a elas. Acredita-se, portanto, que o fato de a indeterminação total ter favorecido o uso dos nomes gerais tenha favorecido, também, a ausência de locativo.

Essa hipótese pode ser ainda mais assertiva quando se compara esse resultado com os dados de Bamuí, pois, ainda que este não tenha sido um fator selecionado pelo GoldVarb X na localidade, a indeterminação total não foi a mais recorrente nos dados, mas, sim, a

indeterminação parcial com referente implícito. Além disso, o nome geral mais recorrente foi *povo* e, no que diz respeito a esse item, somente uma ocorrência foi analisada como sendo totalmente indeterminada, uma vez que, normalmente, quando se utilizou tal item, fez-se referência aos brasileiros ou aos bambuienses, conforme exemplo (73).

(73) eu acho que hoje o **povo** deseja muito é dinhe[i]ro né (BAM07F)

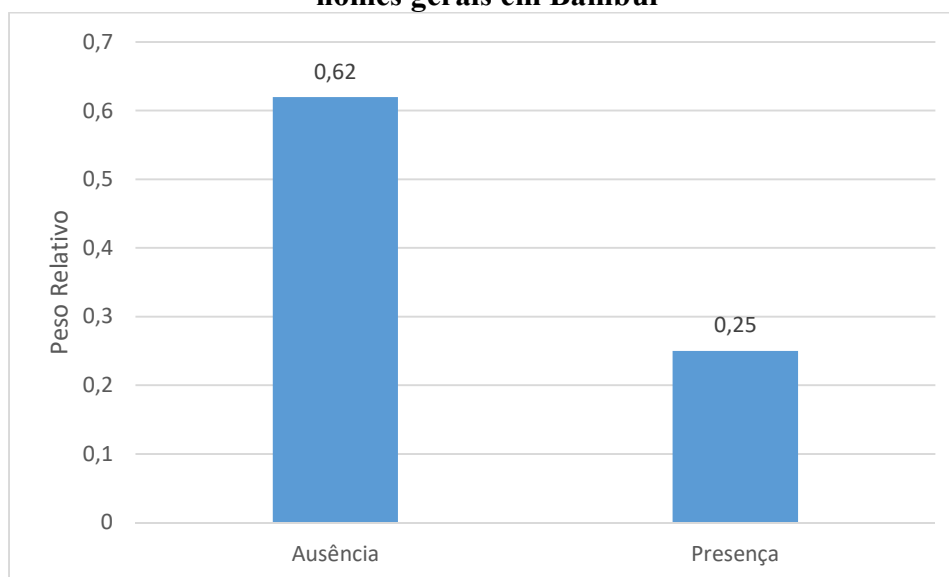
O exemplo (73) foi o único classificado como indeterminação total no conjunto de dados de *povo*, pois, somente nesse caso, fala-se de um povo sem marcar nenhuma localidade ou referência.

Nesse sentido, acredita-se ser possível estabelecer uma relação entre este fator e a variável grau de indeterminação, pois, nos dados classificados como indeterminação total, não há nenhum tipo de referência ao local, diferente do que ocorre nos dados analisados como indeterminação parcial.

4.6.4 Presença ou ausência de anáfora

O fator presença ou ausência de anáfora foi selecionado nos dados de Bambuí, contudo, na rodada inicial ocorreu um *Knockout* nessa variável. Dessa forma, foi necessário o reagrupamento dos resultados em que ocorreram *retomada anafórica com o mesmo elemento* com a *retomada anafórica com outro elemento*. Assim, a partir dessa junção, chegou-se aos dados apresentados no Gráfico 4, no qual se pode visualizar uma prevalência da ausência de anáfora em Bambuí, ou seja, quando um item ocorre, ele não tende a ser retomado.

Gráfico 4 – Influência da variável ausência ou presença de anáfora no conjunto dos nomes gerais em Bambuí



Fonte: Dados da autora.

O Gráfico 4 reflete uma diferença significativa entre a ausência e a presença de anáfora nos dados de Bambuí, o que demonstra que, quando um informante utiliza um nome geral, ele não tende a retomá-lo. Em Coimbra, no entanto, esse item não foi selecionado e não houve nenhum *Knockout* na rodada inicial. Assim, o que se obteve, em Coimbra, foi a *retomada anáfora com outro elemento* (19%), seguida da *ausência de anáfora* (11,5%) e, por fim, a *retomada anafórica com o mesmo elemento* (8,3%).

Sabe-se que não é possível uma análise equiparada entre as duas cidades nesse caso, até mesmo, porque o fator foi selecionado em uma e não em outra, mas é possível concluir que os bambuienses não fazem a retomada de um nome geral após o utilizarem e isso pode se dar por se tratar uma forma, ainda que inconsciente, de não se marcar sobre quem se está falando. No caso de Coimbra, a retomada tende a ser feita com outro elemento e talvez isso tenha se dado pelo fato de que, em Portugal, somente o item *pessoas* tenha ocorrido de forma significativa, demonstrando que outros nomes gerais são poucos usuais e, portanto, em contextos de retomada, outras formas apareçam, como no exemplo (74):

(74) **estamos** numa era global e se não for no global em nível internacional tem que ser no mínimo dos mínimos um global a nível do país e portanto as **pessoas** têm que porque hoje em dia num tem que ser como/como **nós** tínhamos esse sonho há vinte trinta anos é... (COI15M)

O exemplo (74) inicia o período com a 1ª pessoa do plural, retoma-a com o item *pessoas* e, por fim, retoma a forma *nós*, variando, dessa maneira, as formas de retomar. Vale lembrar

ainda que em Coimbra obteve-se apenas as porcentagens sobre a presença e que os resultados sobre ausência de anáfora também foram significativos. Assim, apesar de aqui estar se propondo uma análise desses itens, acredita-se que um grupo maior de dados, em que se obtenha pesos relativos, possa trazer resultados diferenciados.

Ademais, a hipótese aqui levantada a respeito dos resultados de Bambuí está relacionada ao fato de o informante não retomar um nome geral como forma de indeterminação exatamente por não querer chamar ainda mais a atenção do ouvinte para aquele do qual se está falando.

4.6.5 Grau de indeterminação

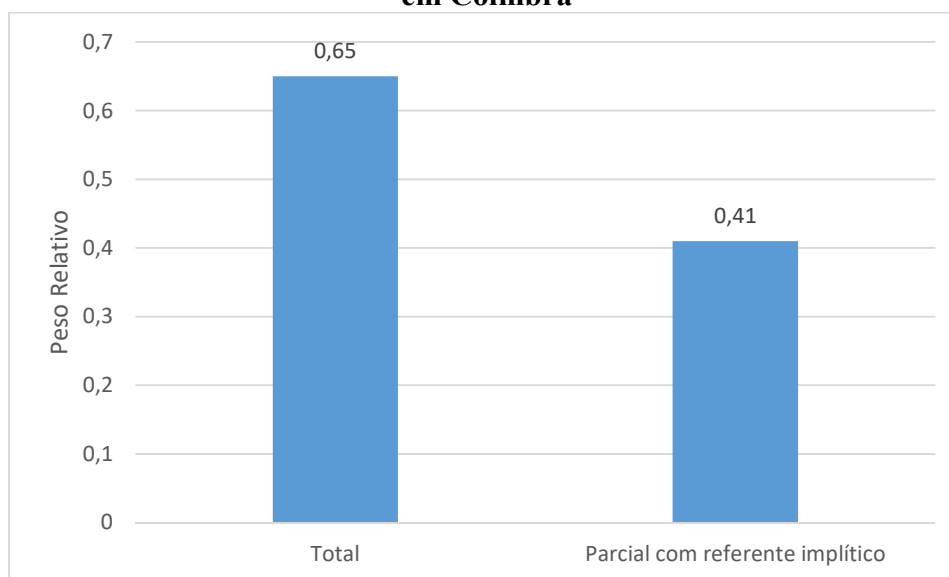
A variável *grau de indeterminação* foi a única selecionada como favorecedora dos nomes gerais nas duas localidades, conforme pode ser visualizado nas demais análises. Esse resultado demonstra como esse fator é relevante para as pesquisas com nomes gerais, tanto no Brasil quanto em Portugal.

É importante lembrar que, no caso de Coimbra, houve um *Knockout* nessa variável e, para que os pesos relativos fossem obtidos, foi necessária a junção da *indeterminação parcial com referente explícito* com a *indeterminação parcial com referente implícito*.

Conforme o Gráfico 5, em Coimbra, os nomes gerais foram favorecidos pela indeterminação total, ou seja, nos contextos em que ocorrem nomes gerais, não é encontrado nada que possa recuperar sobre quem se está falando. Já no caso de Bambuí, conforme Gráfico 6, o que se tem é um favorecimento de nomes gerais nos contextos de indeterminação parcial com referente implícito, seguidos da indeterminação total e, por fim, com referente explícito.

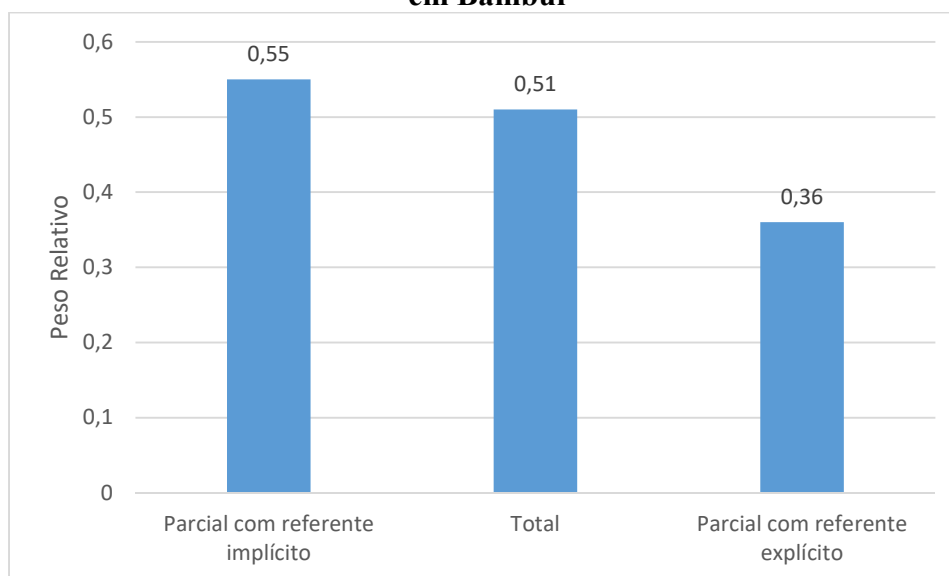
Tendo em vista que esse fator foi selecionado nas duas localidades, algumas hipóteses podem ser levantadas. A primeira delas é que esse resultado poderia indicar que a indeterminação, tanto no PB quanto no PE, pode ser analisada a partir de graus, isto é, do mais indeterminado ao menos indeterminado, mas que, ainda assim, seriam contextos de indeterminação e não de indefinição. Assim, a indeterminação com maior grau estaria relacionada, dentro dos nomes gerais, a contextos em que se fale sobre todas as pessoas do mundo, sem trazer elementos sobre o local de que se está falando ou sobre quem se está falando. Além disso, nesses casos, itens como *pessoa* ou *pessoas*, que demonstram um processo de gramaticalização mais avançado, conforme estudos anteriores, parecem ser os maiores responsáveis por essa indeterminação total.

Gráfico 5 – Influência da variável grau de indeterminação no conjunto dos nomes gerais em Coimbra



Fonte: Dados da autora.

Gráfico 6 – Influência da variável grau de indeterminação no conjunto dos nomes gerais em Bambuí



Fonte: Dados da autora.

Ao se tratar da indeterminação parcial com referente implícito em Bambuí, predominante na localidade, acredita-se que esse fato tenha se dado porque o item mais recorrente foi *povo* e, ao menos nos contextos encontrados em nosso estudo, ele demonstrou ser menos indeterminado que *pessoas*. Além disso, também foi possível constatar que a diferença entre os resultados de indeterminação parcial com referente implícito e total não parece ser tão significativa, o que pode estar ligado ao fato de que o segundo item mais recorrente em Bambuí foi *pessoas*, variante essa, que, conforme estudos já apresentados aqui,

possui uma capacidade de indeterminação grande, exatamente por estar em um processo avançado de gramaticalização.

Por fim, a indeterminação parcial com referente explícito não se mostrou significativa, em termos de pesos relativos, sendo ela bastante inferior nos dados obtidos, que pode ser visualizado no Gráfico 6. Em Coimbra, como não houve ocorrência de indeterminação parcial com referente explícito, ela sequer pôde ser analisada. Nas demais formas de indeterminação nessa mesma localidade, houve apenas quatro ocorrências em que foi possível a análise de indeterminação com referência explícita. Uma delas é demonstrada no exemplo (75) a seguir:

(75) porque ela faz lá surfe... e "ei... **a gent[e]** faz ondas assim... faz ondas assado"...
sábado e domingo e todos dias há onda (COI35M)

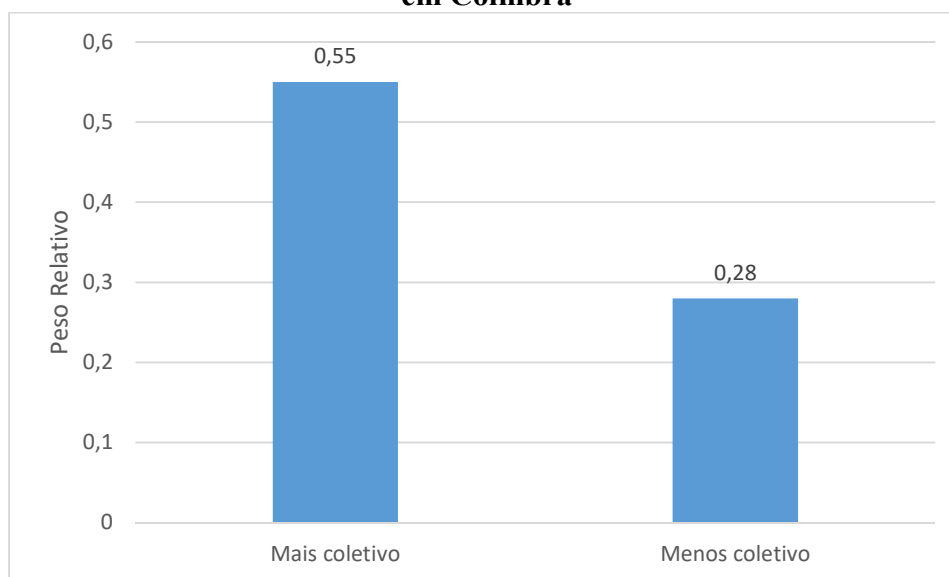
A forma *a gente*, de acordo com Neves (2000), foi classificada também como uma indeterminação parcial, porque normalmente ela inclui quem está falando, como no caso do exemplo em que o informante descreve o processo de surfar, incluindo a si próprio. Essas ocorrências não são comuns em contextos com nomes gerais, ao menos em nosso conjunto de dados, o que leva a mais uma hipótese: os nomes gerais são capazes de indeterminar em um grau até mesmo mais alto do que formas já discutidas de maneira mais aprofundada em estudos da área.

Sendo assim, pode-se concluir, a partir desse resultado, que o estudo da indeterminação por graus se mostrou satisfatória em nosso conjunto de dados e que nomes gerais têm mostrado uma capacidade de indeterminar até mesmo maior do que itens considerados como variantes indeterminadoras em estudos sociolinguísticos.

4.6.6 Mais ou menos coletivo

O fator mais ou menos coletivo foi selecionado apenas em Coimbra e favoreceu contextos de itens coletivos, como pode ser visto no Gráfico 7. Em Bambuí, os resultados, em porcentagem, foram muito equiparados, com uma leve preferência pelo mais coletivo, com 14,2% contra 13,7% para itens que não possuem o traço de +coletividade.

Gráfico 7 – Influência da variável mais ou menos coletivo no conjunto dos nomes gerais em Coimbra



Fonte: Dados da autora.

Observou-se, na primeira análise, que, tanto em Coimbra quanto em Bambuí, os itens mais recorrentes entre os nomes gerais – *povo*, seguido de *pessoas* em Bambuí, e o item *pessoas*, em Coimbra, que representou 81% dos dados – possuem o traço da coletividade, o que, com certeza, influenciou esse resultado. Além disso, esse fator parece demonstrar que nomes gerais que indicam a coletividade parecem ser preferidos por falantes brasileiros e portugueses nos contextos de indeterminação.

A partir de todo esse levantamento, é possível inferir, portanto, que os nomes gerais que possuem o traço +coletividade parecem favorecer contextos de indeterminação com nomes gerais em Coimbra, o que aparentemente também deve ocorrer em Bambuí.

Assim, após essa análise comparativa, a seguir serão feitas as considerações finais e a apresentação de uma síntese dos resultados obtidos a partir da proposta deste estudo.

5 CONSIDERAÇÕES FINAIS

Esta pesquisa teve como objetivo principal observar a relevância dos nomes gerais no conjunto das variantes indeterminadoras, verificando a influência de fatores linguísticos e extralinguísticos na realização do fenômeno. Além disso, foi também nosso objetivo propor uma comparação entre o PB e o PE, a partir de uma amostra de 48 gravações, coletadas no Brasil, na cidade de Bambuí, no estado de Minas Gerais, e em Coimbra, localizada em Portugal.

Inicialmente, por meio da bibliografia apresentada, foi possível observar, a partir de Milanez (1982), Menon (1994), Setti (1997), Godoy (1999), Carvalho (2010), Souza (2014) e outros, que existem, tanto no PB quanto no PE, muito mais maneiras de se indeterminar o sujeito do que as formas que as gramáticas tradicionais atestam, dentre as quais estão os nomes gerais.

No que se refere aos dados obtidos, em Bambuí, foi possível encontrar 191 ocorrências de nomes gerais em um conjunto de 1.357 dados. Os nomes gerais encontrados foram *cara*, *fulano*, *homem*, *pessoa*, *pessoal*, *pessoas*, *povo* e *ser humano*. No que diz respeito às demais formas de indeterminação, encontraram-se as seguintes: *a gente*, *ocê/ocê/cê*, *nós/nóis/desinência de 1ª pessoa do plural*, *eles/ês*, $\emptyset+V3PS$, $\emptyset+V3PP$, $\emptyset+VINF$ e *se*. Em Coimbra, foram obtidos 940 dados e 105 ocorrências de nomes gerais. Os nomes encontrados foram *pessoas*, *pessoa*, *pessoal* *povo* e *gente*. Foram também encontradas as variantes *nós/desinência de 1ª pessoa do plural*, *se*, $\emptyset+V3PP$, *a gente*, $\emptyset+VINF$, *eles*, $\emptyset+V3PS$, *ocê(s)/cê* e *tu*.

Os nomes gerais, foco desta pesquisa, ocuparam a terceira posição em Bambuí e, dessa forma, corroboraram o resultado encontrado por Setti (1997), na cidade de Florianópolis, Carvalho (2010), em Salvador, e Souza (2014), em comunidades no interior da Bahia. *Povo* foi o item mais recorrente deste grupo. Em Coimbra, os nomes gerais ocuparam a quarta posição com 105 ocorrências, resultado muito próximo ao encontrado no Brasil. Contudo, em Coimbra, houve uma menor variação de itens dentro desse grupo, sendo *pessoas* a variante mais recorrente.

A gente foi a forma mais recorrente no *corpus* coletado em Bambuí, sendo encontrada 376 vezes. Esse primeiro lugar em número de ocorrências entre as variantes também foi encontrado em Setti (1997), em pesquisa na cidade de Florianópolis, além de em Oliveira (2006), Ponte (2008) e Souza (2014). Em Coimbra, no entanto, esse item aparece apenas 36 vezes.

A variante *você(s)/ocê/cê* ocupou segundo lugar, em número de ocorrências em Bambuí. Essas três formas totalizam conjuntamente 325 dados. Em Coimbra, essa é uma variante pouco produtiva, ocorrendo apenas 11 vezes. Além disso, essa variante ocorre no plural (*vocês*) em Coimbra por duas vezes, forma esta que não ocorre em Bambuí. De forma geral, os resultados, no Brasil, demonstram um grande número de ocorrências desse item, o que não acontece em Portugal.

A forma *nós/nóis* e a desinência modo temporal *-mos* foi a quarta forma de se indeterminar mais recorrente em Bambuí, com 134 ocorrências. Essa variante não ocupa posições muito altas nos trabalhos já citados, exceto no de Oliveira (2006), em pesquisa voltada para uma comparação entre o PB e o PE, em que o *nós* ocupa o segundo lugar no conjunto de suas variantes. Em Coimbra, conforme já foi demonstrado, a forma *nós* e a desinência modo temporal *-mos* foi a variante mais recorrente, com 451 dados.

A forma *eles/ês*, em Bambuí, totalizou 122 dados; já em Coimbra, ela não se mostrou muito produtiva, ocorrendo apenas 21 vezes. Uma constatação encontrada neste trabalho que corrobora Vitral e Ramos (2006) diz respeito ao fato de homens favorecerem a forma *ês*, semelhante aos dados encontrados em Bambuí.

A variante $\emptyset+V3PS$ mostrou-se bastante produtiva em trabalhos anteriores, já que, em média, ela ocupa, nas pesquisas no Brasil, a segunda ou terceira posição em número de ocorrência das variantes selecionadas, suplantando, muitas vezes, os nomes gerais. Em Bambuí, coincidentemente, ela ocupou a quinta posição no conjunto das variantes (122 ocorrências), juntamente com a forma *eles/ês*. Em Coimbra, essa forma se mostrou ainda menos produtiva que em Bambuí, com apenas 13 ocorrências.

A forma $\emptyset+V3PP$, não muito recorrente em trabalhos antigos, como em Milanez (1982), Menon (1994), Setti (1997) e Godoy (1999) nem mesmo em trabalhos recentes, como Ponte (2008), Carvalho (2010) e Souza (2014), foi encontrada apenas 39 vezes nos dados de Bambuí. Vale lembrar que a variante *eles/ês* foi recorrente em Bambuí, o que pode ter desfavorecido o uso de $\emptyset+V3PP$. No entanto, o português europeu parece ter seguido o caminho inverso no caso dessa variante, tendo em vista que $\emptyset+V3PP$ foi a terceira variante mais recorrente em Coimbra, com 127 dados.

A variante $\emptyset+VINFINF$ foi considerada como indeterminadora do sujeito apenas nos trabalhos de Milanez (1982), Carvalho (2010) e Oliveira (2018). Contudo, optou-se aqui por selecioná-la, porque ela demonstrou ser usada com frequência semelhante em Bambuí e em

Coimbra. Em Bambuí, foram encontradas 36 ocorrências e, em Coimbra, 21, o que parece demonstrar uma equiparidade desse item nos dados coletados no Brasil e em Portugal.

O item *se* pode ser considerado a forma de indeterminação por excelência, pois, nos contextos em que ela aparece, dificilmente é possível recuperar de quem se trata. Tal variante se mostrou mais produtiva no Brasil em trabalhos mais antigos. Em Coimbra, no entanto, foram encontradas 151 ocorrências dessa variante no *corpus*, enquanto em Bambuí apenas 12. Assim, é possível notar mais uma diferença entre o PB e o PE, isto é, enquanto o item *se* encontra-se em segundo lugar em número de ocorrências do PE, no PB ele ocupa a última.

A variante *tu* não se mostrou produtiva em nossos dados, uma vez que em Bambuí não foram encontradas ocorrências e, em Coimbra, esta apareceu apenas seis vezes, sendo quatro delas utilizadas por apenas uma informante.

No que se refere aos fatores selecionados estatisticamente pelo programa GoldVarb X, que teve como objetivo observar a relevância estatística dos nomes gerais no conjunto das variantes, três fatores foram selecionados como favorecedores desses nomes em Bambuí. São eles presença ou ausência de anáfora, grau de indeterminação e sexo. Já em Coimbra, os fatores selecionados foram presença ou ausência de locativo, faixa etária, grau de indeterminação e mais ou menos coletivo.

No caso do fator *presença ou ausência de anáfora*, notou-se um favorecimento da ausência de retomada anafórica (0,62), que se supõe estar relacionado ao fato de que o informante não retoma um nome geral como forma de indeterminação exatamente por não querer chamar ainda mais a atenção do ouvinte para aquele sobre quem se está falando.

Ao se discutir o fator *grau de indeterminação*, foi possível constatar que os nomes gerais favorecem a indeterminação parcial com referente implícito (0,55), seguida de uma indeterminação total (0,51). Este resultado foi muito relevante para nosso estudo, visto que reforça como os nomes gerais são capazes de indeterminar. Salienta-se que este resultado pode ser pensado conjuntamente com o fator *presença ou ausência de anáfora*, pois, se não há retomada, é possível notar uma dificuldade ainda maior de se descobrir o referente.

O único fator extralinguístico selecionado como favorecedor dos nomes gerais em Bambuí, assim como em Carvalho (2010), foi o *sexo*. Notou-se um leve favorecimento no sexo feminino, embora, em números absolutos, isso seja representado por apenas um dado. Assim, tanto em Bambuí quanto em Coimbra, homens e mulheres utilizam os nomes gerais de maneira relativamente equiparada, ainda que em Coimbra este não tenha sido um fator selecionado como favorecedor desses nomes. Nesse sentido, nosso estudo consegue demonstrar que a variável *sexo*

não é um fator que influencia diretamente o fenômeno em questão e que, portanto, os nomes gerais são formas usadas indistintamente por homens e mulheres. Além disso, como o objeto de estudo não é estigmatizado, as diferenças de papéis sociais não parecem influenciar na escolha de uma ou outra variante, conforme relata Freitag (2015).

Os fatores linguísticos que não foram selecionados, ou seja, faixa etária, presença ou ausência de locativo e mais ou menos coletivo foram analisados qualitativamente. As conclusões mais relevantes são que, na maioria dos casos, quando o informante utiliza o nome geral, ele tende a usar formas que remetam ao lugar do qual está tratando. No que diz respeito à faixa etária, os informantes mais jovens (faixa etária 1 e 2), conforme esperado, fazem maior uso dos nomes gerais que os mais velhos. Isso pode ser um indício de que esses nomes tendem a continuar na língua.

O fator presença ou ausência de locativo, selecionado nos dados de Coimbra, favoreceu a ausência de locativo (0,45). Dessa forma, pôde-se observar que, quando os informantes utilizam um nome geral, eles não fazem uso de recursos de marcação para se demonstrar sobre qual local se fala, tal como apontado por Oliveira (2018). Esse resultado pode servir como reforço para demonstrar o quanto os nomes gerais são capazes de indeterminar, a ponto de os informantes sequer usarem locativos para marcar sobre onde se está falando quando utilizam esses itens. Além disso, a partir da análise desse fator extralinguístico, é possível reforçar a ideia de que a indeterminação pode ser dividida em graus, conforme se defende aqui, pois, de acordo com o que foi demonstrado nos exemplos, têm-se momentos em que os informantes marcam de onde se está falando, mas isso não invalida o fato de o sujeito ser indeterminado, uma vez que não é possível se precisar sobre quem se está falando.

O segundo fator selecionado em Coimbra foi faixa etária. Diferentemente dos resultados encontrados no Brasil (CARVALHO, 2010; OLIVEIRA, 2018), tal fator mostra-se relevante para os nomes gerais. Como é possível observar nos resultados obtidos, o resultado, previsível, foi de que mais jovens usam os nomes gerais, seguidos da segunda faixa e, por último, o uso de nomes gerais pelos mais velhos. Isso reforça o que Bortoni-Ricardo (2014) apresenta sobre os papéis ocupados pelos membros das comunidades devido a sua faixa etária e como esse fator é capaz de avaliar casos de mudança em progresso. Assim, como em Coimbra há mais jovens que utilizam os nomes gerais, podemos trazer à tona a possibilidade de uma permanência desses itens na língua e ainda de uma possível mudança nas formas de indeterminação nessa localidade, com os nomes gerais ocupando cada vez mais espaço no conjunto das variantes.

O único fator selecionado nas duas localidades foi *grau de indeterminação*, e em Coimbra a indeterminação total favorece o uso dos nomes gerais. Esse resultado é de extrema valia, pois reforça como os nomes gerais têm sido usados quando não se quer ou não se pode recuperar o referente, assim como já pontuaram Amaral e Ramos (2014), Amaral (2017), Mihatsch (2017) e Amaral e Mihatsch (2019).

O último fator selecionado como favorecedor dos nomes gerais em Coimbra foi o *mais ou menos coletivo* e, nesse caso, houve um favorecimento de itens mais coletivos no grupo de nomes gerais. Salienta-se que tal resultado é atribuído ao fato de que o nome geral *pessoas* foi responsável por 81% de todo o grupo de nomes gerais encontrados no *corpus*, o que, com certeza, fez com que esse resultado acontecesse.

Os fatores *sexo* e *presença ou ausência de anáfora* não foram selecionados nos dados de Coimbra, contudo, no caso de *sexo*, foi possível perceber que homens e mulheres utilizam de maneira semelhante os nomes gerais e, no caso do fator *presença ou ausência de anáfora*, percebeu-se que os falantes optam por retomar os nomes gerais em um contexto de indeterminação, primeiramente com outros elementos, seguido da ausência de forma de retomada e, por fim, de retomadas com o mesmo elemento.

A partir do resumo dos dados obtidos, pode-se perceber que este trabalho trouxe contribuições tanto para a pesquisa com nomes gerais, quanto para os estudos comparativos entre o PB e o PE. Dessa maneira, é possível reforçar como os nomes gerais são capazes de indeterminar no PB e ainda demonstrar como no PE esses itens se mostram produtivos atualmente. A frequência de ocorrências é semelhante nas duas localidades e a grande diferença entre os resultados se dá porque em Bambuí tem-se uma maior variedade de nomes gerais sendo utilizados, o que não ocorreu em Coimbra. Além disso, os fatores selecionados demonstraram que o *sexo* do informante não influencia diretamente no uso de nomes gerais, até mesmo em Bambuí, pois a diferença foi de apenas um dado.

Ademais, uma contribuição trazida por este estudo se refere ao fato de o único fator selecionado nas duas localidades ter sido o *grau de indeterminação*, o que reforça que, para se entender esse fenômeno, é necessário analisar caso a caso e não apenas separar a indefinição e a indeterminação a partir de diferenças pontuais. Sendo assim, este estudo se posiciona a favor de que a indeterminação pode ser dividida em graus e que a ocorrência de um locativo, por exemplo, não invalida o fato de um item ser indeterminado.

Vale salientar que a confusão entre a indeterminação e a indefinição foi tratada em vários trabalhos, mas aqui se propõe a divisão em graus (total e parcial), o que seria um caminho

para sanar as dúvidas quanto a essa classificação, pois muitas vezes os termos são até mesmo utilizados indistintamente.

A partir desses resultados, tem-se, então, um panorama das formas mais utilizadas em Bambuí e em Coimbra para se indeterminar o sujeito. No caso de Bambuí, é nítido que os resultados corroboram as pesquisas anteriores, mostrando, assim, que, independentemente da região do Brasil em que os falantes estejam, as formas utilizadas para indeterminar o sujeito são bastante equiparadas (MENON, 1994; SETTI, 1997; GODOY, 1999; PONTE, 2008; SOUZA, 2014). Bambuí não parece mostrar um comportamento diferente no que concerne aos fatores selecionados, conforme Oliveira (2018) já havia demonstrado; no entanto, o estudo de dados da cidade contribui não só para o conhecimento do modo de falar do município, mas também para comparações com outros trabalhos e para auxiliar na ampliação de produção de conhecimento voltado para o português falado em Minas Gerais.

Os resultados de Coimbra também contribuem para os estudos comparativos entre o PB e o PE, além de demonstrar que os nomes gerais ocorrem em quantidade semelhante nas duas localidades analisadas. De modo geral, a hipótese inicial foi parcialmente confirmada, já que se acreditava que brasileiros utilizariam mais nomes gerais que os portugueses; entretanto, em Coimbra os nomes gerais foram mais utilizados do que o esperado. Observa-se que em Bambuí a variedade de nomes gerais foi maior, enquanto em Coimbra o item *pessoas* foi responsável por grande parte das ocorrências desse grupo de variantes. Além disso, confirmou-se que as demais formas utilizadas por brasileiros e portugueses são diferentes, pois enquanto os segundos preferem formas como *nós*, *se* e *verbo na terceira pessoa do plural sem sujeito explícito*, os primeiros preferem variantes como *você* e *a gente*.

Durante a pesquisa, foram encontrados poucos trabalhos que estudassem o fenômeno da indeterminação em Portugal, principalmente na vertente da sociolinguística variacionista e, por isso, acredita-se que esse estudo possa contribuir, ainda que minimamente, sobre o fenômeno no PE.

Espera-se que este trabalho sirva de fonte para futuras pesquisas sobre o tema, visto que ainda se carece de estudos na área de indeterminação do sujeito. Em Minas Gerais, por exemplo, são escassas as pesquisas que versam sobre as diversas formas de indeterminação em dados de língua oral. No que concerne aos nomes gerais, percebe-se que são itens bastante recorrentes na língua e que ocupam posições altas em número de ocorrências em todas as pesquisas realizadas até o momento. Além disso, muito se tem pesquisado sobre esses nomes na última década, por meio de autores como Amaral e Ramos (2014), Amaral (2017), Mihatsch (2017),

Amaral e Mihatsch (2019). Eles observam, por meio de dados de língua oral e escrita, o comportamento desses itens das mais variadas formas e demonstram como eles têm se esvaziado de sentido e ocupado espaços ainda pouco analisados, como a de formação de pronomes.

Conclui-se, portanto, que Bambuí, ainda que seja uma cidade interioriana, distante da capital, tem comportamento semelhante na realização desse fenômeno quando comparada com outras localidades brasileiras. Por outro lado, Coimbra mostrou comportamento diverso no conjunto das variantes quando se considera a recorrência de uso de cada uma delas, entretanto, os nomes gerais, tanto em Bambuí quanto em Coimbra, obtiveram resultados semelhantes. Isso parece mostrar que o uso dos nomes gerais como recurso de indeterminação do sujeito possui variações relativamente estáveis em todos os lugares pesquisados até o momento. Futuras pesquisas poderão analisar outras variedades, como o português angolano ou moçambicano, sobre os quais até o momento não se encontram estudos.

REFERÊNCIAS

- ALKMIM, Tânia Maria. Sociolinguística. Parte I. In: MUSSALIM, Fernanda; BENTES, Ana Cristina. **Introdução à linguística: domínios e fronteiras**, v.1, 6. ed. São Paulo: Cortez, 2006. p. 21-47.
- AMARAL, Eduardo Tadeu Roque. A transcrição das fitas: abordagem preliminar. In: MEGALE, Heitor (org.). **Filologia Bandeirante: estudos**. v. 1. São Paulo: Humanitas FFLCH/USP, 2000. p. 195-208.
- AMARAL, Eduardo Tadeu Roque. Estudio contrastivo de nombres generales para humanos en español y en portugués. **Lingüística y literatura**, n. 72, p. 54-79, 2017.
- AMARAL, Eduardo Tadeu Roque; MIHATSCH, Wiltrud. Incipient impersonal pronouns in colloquial Brazilian Portuguese based on 'pessoa', 'pessoal' and 'povo'. **Linguistische Berichte, Sonderhefte** 26, p. 149-185, 2019.
- AMARAL, Eduardo Tadeu Roque; OLIVEIRA, Fernanda Carla. A influência de fatores semânticos no uso de nomes gerais para a indeterminação do Sujeito. **Acta Scientiarum (UEM)**, v. 41, p. e44977, 2019.
- AMARAL, Eduardo Tadeu Roque; RAMOS, Jânia. **Nomes gerais no português brasileiro**. Belo Horizonte: Faculdade de Letras da UFMG, 2014.
- ARAGÃO, Shelton de; ARAÚJO, Edivalda. O tópico retomado na posição de sujeito: diferenças entre português brasileiro e português europeu. **A Cor das Letras**. Revista Digital dos Programas de Pós-Graduação do Departamento de Letras e Artes da UEFS Feira de Santana, v. 18, n. 2, p. 192-215, maio-agosto 2017.
- BARBOSA, Alana Brito. **A indeterminação do sujeito no falar ludovicense**. 122f. Dissertação (Mestrado em Letras) - Setor de Ciências Humanas da Universidade Federal do Paraná, Curitiba, 2016.
- BARBOSA, Jorge Morais. Verbos impessoais? In: RIO-TOTO, Graça Maria; FIGUEIREDO, Olívia Maria; SILVA, Fátima (Orgs.) **Estudo em Homenagem ao Professor Doutor Mário Vilela**. Porto: Faculdade de Letras da Universidade do Porto, 2005, p. 369-376. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/9381/2/homenagemmmvilelavol1completo000065979.pdf#page=369>. Acesso em: 08 mar. 2022.
- BELINE, Ronald. A variação linguística. In: FIORIN, José Luiz. **Introdução à Linguística**, 2003.
- BORTONI-RICARDO, Stella Maris. **Manual de sociolinguística**. São Paulo: Contexto, 2014.
- CÂMARA MUNICIPAL DE BAMBUÍ. **Lei Municipal nº 2178**, de 07/11/2011. Publicada no quadro de publicações da Prefeitura, nos termos do art. 127 da Lei Orgânica do município de Bambuí.

CAPPEAU, Paul; SCHNEDECKER, Catherine. Gens, personne(s), individu(s): trois saisies de l'humain. In: NEVEU, F.; Blumenthal, P., HRIBA, L.; GERSTENBERG, A.; MEINSCHAEFER, J. (Orgs.) **4ième Congrès Mondial de Linguistique Française**, Jul 2014, BERLIM, Allemagne. Volume 8, 2014 (8), p. 3027-2040.

CARVALHO, Valter de. **Você, a gente et alia indeterminam o sujeito em Salvador**. 197f. Dissertação (Mestrado em Estudos da Linguagem) – Departamento de Ciências Humanas, Universidade do Estado da Bahia, Salvador, 2010.

CARVALHO, Gislaine Aparecida de. **A realização do sujeito na fala do português europeu**. 259. Tese (Doutorado em Linguística e Língua Portuguesa) – Programa de Pós-Graduação em Letras da Universidade Estadual Paulista – Júlio de mesquita Filho Campus de Araraquara, Araraquara, 2009.

COELHO, Izete Lehmkuhl *et al.* **Para conhecer sociolinguística**. São Paulo: Contexto, 2015.

COSTA, Raquel Maria da Silva. A alternância das formas pronominais *tu, você* e o(a) senhor(a) na função de sujeito no português falado em Cametá-Estado do Pará. **Revista de Letras**, Universidade Federal do Ceará, v. 2, n. 35, 2016, p. 64-76.

COSTA, Vítor. **Conhecer Coimbra**. Turismo de Coimbra, 2011.

CUNHA, Celso; CINTRA, Luís F. Lindley. **Nova gramática do português contemporâneo**. 5. ed. Rio de Janeiro: Lexicon, 2008.

DUARTE, Maria Eugênia Lamoglia. **O sujeito de referência indeterminada em sentenças infinitivas**. Revista do GEL, S. J. do Rio Preto, v. 5, n. 1, p. 9-30, 2008.

DUARTE, Maria Eugênia Lamoglia.; KATO, Mary Aizana.; BARBOSA, Pilar. Sujeitos indeterminados em PE e PB. In: **Boletim da ABRALIN**. Fortaleza: Maria Elias, 2001. v. 26. p. 1-15.

ECKERT, Penelope. Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. **Annual review of anthropology**, Palo Alto, n.41, p. 87-100, 2012.

EPIS – Empresários pela Inclusão Social. **Atlas da Educação** – Desempenho e potencial de sucesso e insucesso por concelho, 2012/2013. Disponível em: <http://www.epis.pt/mediadores/atlas-da-educacao/atlas>. Acesso em: 31 ago. 2018.

FERREIRA, Carlota. A geografia linguística no Brasil. **DELTA: Documentação e Estudos em Linguística Teórica e Aplicada**. v. 11, n. 2, 2019. Disponível em: <https://revistas.pucsp.br/index.php/delta/article/view/45189>. Acesso em: 20 mai. 2022.

FLAUX, Nelly. À propos des noms collectifs. **Revue de linguistique romane**. 63/251-252, 1999, p. 471-502.

FLAUX, Nelly; VAN DE VELDE, Danièle. **Les noms en français**: esquisse de classement. Ophrys, 2000, p. 56-61.

FRANCESCHINI, Lucelene Teresinha; LOREGIAN-Penkal, Loreni. A variável sexo/gênero e o uso de tu e você no Sul do Brasi. **SIGNUM: Estud. Ling.**, n.18/1, p. 182-205, 2015.

FREITAG, Raquel; MARTINS, Marco Antônio; TAVARES, Maria Alice. Bancos de dados sociolinguísticos do português brasileiro e os estudos de terceira onda: potencialidades e limitações. **Revista Alfa**, v. 56, n. 3, p. 917-944, 2012.

FREITAG, Raquel Meister Ko. (Re)Discutindo Sexo/Gênero na Sociolinguística. In: FREITAG, Raquel Meister Ko.; SEVERO, Cristine Gorski (Org). **Mulheres, Linguagem e Poder** – Estudos de Gênero na Sociolinguística Brasileira. São Paulo: Blucher, 2015, p. 17-74. Disponível em: <file:///C:/Users/Usu%C3%A1rio/Downloads/OpenAccess-Freitag-978-85-8039-121-3-0001.pdf>. Acesso em: 30 mai. 2022.

FRONEK, Josef. *Thing* as a function word. **Linguistics**, Haia v. 20, n. 9-10, 1982, p. 633-654.

GAST, Volker; AUWERA, Johan van der. Towards a distributional typology of human impersonal pronouns, based on data from European languages. In: BAKKER, Dik, HASPELMATH, Martin (Orgs.) **Languages Across Boundaries: Studies in Memory of Anna Siewierska**. Mouton: De Gruyter, 2013, p. 1-48.

GIVÓN, Talmy. **Syntax**. Revised edition. Philadelphia: Jonh Benjamins. (vol. 1 and 2), 2001.

GODOY, Maria Alice Maschio. **A indeterminação do sujeito no interior paranaense: uma abordagem sociolinguística**. 1999. 128 f. Dissertação (Mestrado em Letras e Linguística), Universidade Federal do Paraná, Paraná, 1999.

GUY, Gregory Riordan.; ZILLES, Ana. **Sociolinguística quantitativa: instrumental de análise**. São Paulo: Parábola, 2007.

HALLIDAY, Michael; HASAN, Ruqaya. **Cohesion in English**. 14. ed. London/New York: Longman, 1995 [1976].

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Bambuú**. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/mg/bambui.html>. Acesso em: 22 de mar. 2022.

INSTITUTO NACIONAL DE ESTATÍSTICA. **Coimbra**. Disponível em: https://censos.ine.pt/scripts/db_censos_2021.html. Acesso em: 20 de maio de 2022.

KOCH, Ingedore. Sobre a seleção do núcleo das formas nominais anafóricas na progressão referencial. In: NEGRI, L.; FOLTRAN, M. J.; OLIVEIRA, R. P. (Orgs.). **Sentido e significação: em torno da obra de Rodolfo Ilari**. São Paulo: Contexto, 2004. p. 244-262.

LABOV, William. **Padrões sociolinguísticos**. São Paulo: Parábola Editorial, 2008 [1972].

LAVANDERA, Beatriz R. **Variación y significado**. Y discurso. Ciudad Autónoma de Buenos Aires: Paidós, 2014.

LUFT, Celso Pedro. **Moderna Gramática Brasileira**. 2. ed. São Paulo: Editora Globo, 2008. Disponível em: <https://books.google.com.br/books?hl=pt-BR&lr=&id=tK5FQq3PoygC&oi=fnd&pg=PA15&dq=sujeito+impessoal&ots=B3KMspIMb>

Y&sig=iDlekKTzVe3GbVkm_KYBfX27_-
I&redir_esc=y#v=onepage&q=sujeito%20impessoal&f=false. Acesso em: 08 mar. 2022.

MATEUS, Maria Helena Mira *et al.* **Gramática da Língua Portuguesa**. 7. ed. Lisboa: Editora Caminho, 2003.

MENON, Odete Pereira da Silva. **Analyse sociolinguistique de l'indetermination du sujet dans le portugais parlé au Brésil, à partir des données du NURC/SP**. 1994. 397 f. Tese (Doutorado) – Département de Recherches Linguistiques, Université de Paris VII, Paris, 1994.

MENON, Odete Pereira da Silva. A gente: um processo de gramaticalização. *In: Estudos Linguístico*, XXV. **Anais do XLII Seminário GEL UNAERP**, Ribeirão Preto, 1996, p. 622-628.

MENON, Odete Pereira da Silva. *Estudos Linguísticos*, XXV. **Anais do XLIII Seminários GEL**, Ribeirão Preto, 1996, p. 622-628.

MEYERHOFF, Miriam; SCHLEEF, Erik; MACKENZIE, Laurel. **Doing sociolinguistics: a practical guide to data collection and analysis**. London/New York: Routledge, 2015.

MIHATSCH, Wiltrud. De plantas, animales y (otros) objetos: lexemas cultos genéricos entre léxico y gramática. *In: ALEXANDRE VEIGA; GONZÁLEZ PEREIRA, M; SOUTO GÓMEZ, Montserrat (eds.): Léxico y gramática*. Lugo: Tris Tram (Linguas e lingüística; 3), 2002, p. 237-248.

MIHATSCH, Wiltrud. **Kognitive grundlagen lexikalischer hierarchien**: untersucht am Beispiel des Französischen und Spanischen. Tübingen: Max Niemeyer, 2006.

MIHATSCH, Wiltrud. Les noms d'humains généraux aux limites de la grammaticalisation. **Syntaxe & Sémantique**, n° 18, p. 67-99, 2017.

MILANEZ, Wânia. **Recursos de indeterminação do sujeito**. 1982. 149 f. Dissertação (Mestrado) – Departamento de Linguística do Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1982.

MOLLICA, Maria Cecília. Relevância das variáveis não linguísticas. *In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). Introdução à Sociolinguística: o tratamento da variação*. 2. ed. São Paulo: Contexto, 2004, p. 27-31.

MORENO FERNÁNDEZ, Francisco. **Principios de sociolingüística y sociología del lenguaje**. 4. ed. corregida y actualizada. Barcelona: Ariel, 2009. 407 p.

NASCIMENTO et al. Sobre formas de tratamento no português europeu e brasileiro. **Diadorim**, Rio de Janeiro, vol. 20 – Especial, p. 245-262, 2018.

NEGRÃO, Esmeralda.V.; VIOTTI, Evani. Estratégias de impessoalização no português brasileiro. *In: Fiorin, J.L.; Petter, M. África no Brasil. A formação da língua portuguesa*. São Paulo: Contexto, 2008, p. 171-203. Disponível em: https://www.researchgate.net/publication/318860711_Impessoalizacao_no_portugues_brasileiro/link/5981ec38a6fdccb9101fb112/download. Acesso em: 07 mar. 2022.

NEVES, Maria Helena M. **Gramática de usos do português**. São Paulo: UNESP, 2000.

OLIVEIRA, Fernanda Carla de. **Nomes gerais e outras formas de indeterminação do sujeito no português oral de Bambuí-MG**. 569f. Dissertação (Programa de Pós-Graduação em Estudos Linguísticos) - Universidade Federal de Minas Gerais, 2018.

OLIVEIRA, Gaylha Wégila; RAMOS, Marta Anaísa Bezerra. Os pronomes de 1ª e 2ª pessoas na representação do sujeito: um olhar sob a ótica variacionista. **III Congresso Brasileiro sobre Letramento e Dificuldade de Aprendizagem**, 2021.

OLIVEIRA, Josane Moreira. A Sociolinguística Laboviana: festejando o cinquentenário e planejando o futuro. **Cadernos de Estudos Linguísticos**, Campinas, p. 481-501 - set./dez. 2016.

OLIVEIRA, Karine Rios de. **Nós, a gente e o clítico se como estratégias de indeterminação do sujeito no Português**. 173f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Uberlândia, 2006.

ONOFRE, Marília Blundi; FACUNDES, Leonildes Pessoa. A “noção da indefinição” nas abordagens linguísticas. **Traços De Linguagem - Revista De Estudos Linguísticos**, 5(1), jan-jun, 2021. Disponível em: <https://periodicos2.unemat.br/index.php/tracos/article/view/5995/4448>. Acesso em: 09 mar. 2022.

POLACHINI, Brua Soares. Verbos *impessoais* e *unipessoais* em gramáticas brasileiras oitocentistas do português. **Revista argentina de historiografia linguística**, VII, 1, 55-67, 2015. Disponível em: <file:///C:/Users/Usu%C3%A1rio/Downloads/Dialnet-VerbosImpessoaisEUnipessoaisEmGramaticasBrasileira-7425459.pdf>. Acesso em: 07 mar. 2022.

PONTE, Vanessa. **A indeterminação do sujeito no português rural do interior do estado da Bahia**. 2008. 125 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Letras e Linguística do Instituto de Letras, Universidade Federal da Bahia, Salvador, 2008.

PORTAL INE – Instituto Nacional de Estatística. **Statistics Portugal**. Disponível em: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_unid_territorial&menuBOUI=13707095&contexto=ut&selTab=tab3. Acesso em: 10 jan. de 2021.

POSIO, Pekka. You and we: Impersonal second person singular and other referential devices in Spanish sociolinguistic interviews. **Journal of Pragmatics**, p. 1-16, 2016.

POSIO, Pekka. *A pessoa* and *uma pessoa*: Grammaticalization and functions of a human impersonal referential device in European Portuguese. **Journal of Portuguese Linguistics**, p. 1-21, 2021.

RAMOS, Jânia M. O surgimento de um nome geral: a lexia *trem* no dialeto mineiro. In: RAMOS, Jânia M.; COELHO, Sueli M. **Português brasileiro dialetal: temas gramaticais**. Campinas: Mercado de Letras, 2013. p. 137-147.

ROLLEMBERG, Vera *et al.* Os pronomes pessoais sujeito e a indeterminação do sujeito na norma culta de Salvador. *In: Estudos linguísticos e literários*, v. 11, p. 53-74, 1991.

RUMEU, Márcia Cristina de Brito. Estratégias de indeterminação do argumento externo do verbo em jornais portugueses e brasileiros. **Veredas On line** – Atemática – 1/2011, p. 49-65 – PPG LINGUÍSTICA/UFJF – Juiz de Fora, 2011.

SALOMÃO, Ana Cristina Biondo. Variação e Mudança Linguística: Panorama e perspectivas da Sociolinguística Variacionista no Brasil. **Fórum Linguístico**, Florianópolis, v. 8, n. 2, p. 187-207, jul./dez. 2011.

SANTANA, Neila Maria Oliveira. **A indeterminação do sujeito no português rural do semi-árido baiano**. 2006. 153 f. Dissertação (Mestrado em Letras e Linguística) – Programa de Pós-Graduação em Letras e Linguística do Instituto de Letras, Universidade Federal da Bahia, Salvador, 2006.

SCHERRE, Marta Maria Pereira; NARO, Anthony Julius. Análise quantitativa e tópicos de interpretação do Varbrul. *In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). Introdução à Sociolinguística: o tratamento da variação*. 2. ed. São Paulo: Contexto, 2004.

SCHERRE, Marta Maria Pereira; YACOVENCO, Lilian Coutinho. A variação linguística e o papel dos fatores sociais: o gênero do falante em foco. **Revista da ABRALIN**, v. 10, n. 3, 31, dez. 2011. Disponível em: <https://revista.abralin.org/index.php/abralin/article/view/1088>. Acesso em: 20 abr. 2022.

SETTI, Adriane Cristina Ribas. **A indeterminação do sujeito nas três capitais do Sul do Brasil**. 1997. 116 f. Dissertação (Mestrado em Linguística). Universidade Federal do Paraná, Curitiba, 1997.

SIEWIERSKA, Anna; PAPASTATHI, Maria. Towards a typology of third person plural impersonals. **Linguistics**, v. 49, n. 3, 575-610, 2011.

SILVA, Josany Maria de Jesus; NAMIUTI, Critiane. ‘Nós’ e ‘a gente’: diferenças de uso nas variedades de português. **Acta Scientiarum**. Language and Culture, v.41, n. 1, jan-jun. 2019. Disponível em: periodicos.uem.br/ojs/index.php/ActaSciLangCult/article/view/43588/pdf. Acesso em: 30 mai. 2022.

SILVA, Lindiomar José. **Bambuí nas trilhas da picada de Goiaz, quilombolas e o povoamento do Campo Grande**. Contagem: Santa Clara Editora, 2010, 370 p.

SILVA, Vera Lúcia Paredes. Relevância das variáveis linguísticas. *In: MOLLICA, Maria Cecília; BRAGA, Maria Luiza (org.). Introdução à Sociolinguística: o tratamento da variação*. 2. ed. São Paulo: Contexto, 2004, p. 67-71.

SOUZA, Soliane Silva. **Formas de indeterminação do sujeito: Uma comparação entre as comunidades rurais Martinha e Casinhas (BA)**. 2014. 148f. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Estudos Linguísticos, Universidade Estadual de Feira de Santana, Feira de Santana, 2014.

TARALLO, Fernando. **A pesquisa sociolinguística**. 7. ed. São Paulo: Ática, 2003.

TAVARES SILVA, Cláudia Roberta. Distribuição e leitura referencial de sujeitos nulos e plenos em línguas pro-drop e não pro-drop: evidências da natureza semi-pro-drop do português brasileiro. *Leitura, [S. l.]*, v. 1, n. 47, p. 15–41, 2013.

TEIXEIRA, Maria Luiza de Sousa. **A indeterminação pragmática e semântica**. 2014.124f. Tese (Doutorado) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”, 2014.

UNIVERSIDADE FEDERAL DE GOIÁS. **Coimbra-Portugal**. Disponível em: <https://files.cercomp.ufg.br/weby/up/586/o/mapacoimbraa.gif>. Acesso em: 31 mai. 2022.

VITÓRIO, Elyne Giselle de Santana Lima Aguiar. A variação tu e você em Maceió, Alagoas. **Todas as Letras**, n. 2, v. 20, p. 85-89, mai/ago/2018.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. **Fundamentos empíricos para uma teoria da mudança linguística**. São Paulo: Parábola, 2006 [1975].

WIDERA, Carmen. O uso de ele em construções impessoais em português europeu – um estudo experimental. In: **XIX Fórum de Partilha Linguística**, n. 6, 2019, Lisboa. Resumos. Lisboa: Universidade de Lisboa, 2019, p. 21-22. Disponível em: http://jiclunl.fcsh.unl.pt/wp-content/uploads/sites/43/2019/07/XIVFPL_Livro-Resumos.pdf#page=25. Acesso em: 08 mar. 2022.

APÊNDICE A – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA INFORMANTES DE BAMBUÍ

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Projeto: Nomes gerais e outras formas de indeterminação do sujeito no português oral de Bambuí-MG

Prezado Senhor (a):

Este Termo de Consentimento pode conter palavras que você não entenda. Peça ao pesquisador que explique as palavras ou informações não compreendidas completamente.

1) Introdução

Você está sendo convidado(a) a participar de uma pesquisa que tem o objetivo de estudar a diversidade social e linguística em Bambuí. Posteriormente, os dados coletados serão comparados com os de outras localidades. Você foi selecionado porque reside em Bambuí, tendo morado a maior parte da sua vida nesta cidade, e sua participação não é obrigatória.

2) Procedimentos do Estudo

Para participar deste estudo, solicito a sua especial colaboração em participar das conversas que serão gravadas. Os dados coletados serão armazenados e utilizados para estudar fenômenos da língua.

3) Riscos e desconfortos

Consideramos que o risco possível para sua participação neste estudo é algum tipo de desconforto que possa ocorrer no momento das conversas gravadas. Esclarecemos que todos os procedimentos serão feitos conforme sua disponibilidade e se, eventualmente, esta pesquisa lhe causar qualquer tipo de insatisfação, o pesquisador compromete-se a reparar o fato, ou prover meios para a reparação.

4) Benefícios

Consideramos que essa pesquisa não trará benefícios diretos para você, mas sua contribuição servirá para ajudar na produção do conhecimento, que será exposto em trabalhos e eventos científicos.

5) Custos/Reembolso

Você não terá nenhum gasto com a sua participação no estudo e também não receberá pagamento por ela.

6) Caráter Confidencial dos Registros

A sua identidade será mantida em sigilo. Os resultados do estudo serão sempre apresentados como o perfil de um grupo e não de uma pessoa. Dessa forma, você não será identificado(a) quando o material de seu registro for utilizado, seja para propósitos de publicação científica ou educativa. As gravações comporão o *corpus* de uma dissertação de mestrado e todas as informações que possam vir a identificá-lo(a) serão mantidas em total sigilo. Seu nome, endereço e outras informações pessoais não serão divulgados e não constarão em nenhum trabalho científico que for decorrente da análise dos dados desta pesquisa.

No entanto, caso você autorize, esta gravação e/ou sua respectiva transcrição serão publicadas em obra científica ou em apresentação de trabalho em congresso. Você tem total liberdade para autorizar ou não:

- () autorizo a publicação da gravação e da transcrição;
 () autorizo a publicação somente da () gravação () transcrição;
 () não autorizo a publicação nem da gravação nem da transcrição.

7) Participação

Sua participação neste estudo é muito importante e voluntária. Você tem o direito de não querer participar ou de sair deste estudo a qualquer momento. Em caso de você decidir retirar-se do estudo, favor informar ao pesquisador.

8) Informações

Em caso de dúvida e eventuais esclarecimentos é possível entrar em contato com o Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, pelo telefone 3409-4592, por email coep@prpq.ufmg.br ou no seguinte endereço: Av. Antonio Carlos, 6627 – Unidade Administrativa II, sala 2005. CEP 31270-901 - Belo Horizonte, MG.

A pesquisadora responsável poderá fornecer qualquer esclarecimento sobre essa pesquisa, assim como tirar dúvidas, bastando contato no seguinte endereço e/ou telefone:

Nome da pesquisadora: Fernanda Carla de Oliveira

Endereço: Avenida Afonso Pena, 1456 – Boa Viagem – Belo Horizonte, 30130-005

Telefone: (31) 3567-3729/(31) 99531-2478

Email: fernanda.carladeoliveira@gmail.com

9) Declaração de Consentimento

Li ou alguém leu para mim as informações contidas neste documento antes de assinar este termo de consentimento. Declaro que toda a linguagem técnica utilizada na descrição deste estudo de pesquisa foi satisfatoriamente explicada e que recebi respostas para todas as minhas dúvidas. Confirmando também que recebi uma cópia deste Termo de Consentimento Livre e Esclarecido. Compreendo que sou livre para me retirar do estudo em qualquer momento, sem perda de benefícios ou qualquer outra penalidade.

Dou meu consentimento de livre e espontânea vontade para participar deste estudo.

Nome do participante (em letra de forma):

Assinatura do participante ou representante legal

Data: _____

Obrigado pela sua colaboração e por merecer sua confiança.

Nome (em letra de forma) e Assinatura do pesquisador

Data: _____

APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA INFORMANTES DE COIMBRA

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título do Projeto: NOMES GERAIS E OUTRAS FORMAS DE INDETERMINAÇÃO DO SUJEITO: UM ESTUDO VARIACIONISTA SOBRE O PORTUGUÊS ORAL DE OURO PRETO (BR) E DE COIMBRA (PT)

Prezado Senhor (a):

Este Termo de Consentimento pode conter palavras que você não entenda. Peça ao pesquisador que explique as palavras ou informações não compreendidas completamente.

1) Introdução

Você está sendo convidado(a) a participar de uma pesquisa que tem o objetivo de estudar a diversidade social e linguística em Coimbra. Posteriormente, os dados coletados serão comparados com os de outras localidades. Você foi selecionado porque reside em Coimbra, tendo morado a maior parte da sua vida nesta cidade, e sua participação não é obrigatória.

2) Procedimentos do Estudo

Para participar deste estudo, solicito a sua especial colaboração em participar das conversas que serão gravadas. Os dados coletados serão armazenados e utilizados para estudar fenômenos da língua.

3) Riscos e desconfortos

Consideramos que o risco possível para sua participação neste estudo é algum tipo de desconforto que possa ocorrer no momento das conversas gravadas. Esclarecemos que todos os procedimentos serão feitos conforme sua disponibilidade e se, eventualmente, esta pesquisa lhe causar qualquer tipo de insatisfação, o pesquisador compromete-se a reparar o fato, ou prover meios para a reparação.

4) Benefícios

Consideramos que essa pesquisa não trará benefícios diretos para você, mas sua contribuição servirá para ajudar na produção do conhecimento, que será exposto em trabalhos e eventos científicos.

5) Custos/Reembolso

Você não terá nenhum gasto com a sua participação no estudo e também não receberá pagamento por ela.

6) Caráter Confidencial dos Registros

A sua identidade será mantida em sigilo. Os resultados do estudo serão sempre apresentados como o perfil de um grupo e não de uma pessoa. Dessa forma, você não será identificado(a) quando o material de seu registro for utilizado, seja para propósitos de publicação científica ou educativa. As gravações comporão o *corpus* de uma tese de doutorado e todas as informações que possam vir a identificá-lo(a) serão mantidas em total sigilo. Seu nome, endereço e outras informações pessoais não serão divulgados e não constarão em nenhum trabalho científico que for decorrente da análise dos dados desta pesquisa.

No entanto, caso você autorize, esta gravação e/ou sua respectiva transcrição serão publicadas em obra científica ou em apresentação de trabalho em congresso. Você tem total liberdade para autorizar ou não:

- () autorizo a publicação da gravação e da transcrição;
- () autorizo a publicação somente da () gravação () transcrição;
- () não autorizo a publicação nem da gravação nem da transcrição.

7) Participação

Sua participação neste estudo é muito importante e voluntária. Você tem o direito de não querer participar ou de sair deste estudo a qualquer momento. Em caso de você decidir retirar-se do estudo, favor informar ao pesquisador.

8) Informações

Em caso de dúvida e eventuais esclarecimentos é possível entrar em contato com o Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais, pelo telefone 3409-4592, por email coep@prpq.ufmg.br ou no seguinte endereço: Av. Antonio Carlos, 6627 – Unidade Administrativa II, sala 2005. CEP 31270-901 - Belo Horizonte, MG – Brasil.

A pesquisadora responsável poderá fornecer qualquer esclarecimento sobre essa pesquisa, assim como tirar dúvidas, bastando contato no seguinte endereço e/ou telefone:

Nome da pesquisadora: Fernanda Carla de Oliveira

Endereço: Avenida Afonso Pena, 1456 – Boa Viagem – Belo Horizonte, 30130-005 – Brasil.

Telefone: (31) 3567-3729/(31) 99531-2478

Email: fernanda.carladeoliveira@gmail.com

9) Declaração de Consentimento

Li ou alguém leu para mim as informações contidas neste documento antes de assinar este termo de consentimento. Declaro que toda a linguagem técnica utilizada na descrição deste estudo de pesquisa foi satisfatoriamente explicada e que recebi respostas para todas as minhas dúvidas. Confirmando também que recebi uma cópia deste Termo de Consentimento Livre e Esclarecido. Compreendo que sou livre para me retirar do estudo em qualquer momento, sem perda de benefícios ou qualquer outra penalidade.

Dou meu consentimento de livre e espontânea vontade para participar deste estudo.

Nome do participante (em letra de forma):

Assinatura do participante ou representante legal

Data: _____

Obrigado pela sua colaboração e por merecer sua confiança.

Nome (em letra de forma) e Assinatura do pesquisador

Data: _____

APÊNDICE C – FICHA DO INFORMANTE

FICHA DO INFORMANTE

IDENTIFICAÇÃO DO ARQUIVO DE ÁUDIO: _____ **DATA:** _____
DURAÇÃO DO ÁUDIO: _____

NOME	
SEXO	
IDADE	
NATURALIDADE	
ENDEREÇO (Bairro e cidade)	
ESCOLARIDADE	
NATURALIDADE DOS PAIS	
PROFISSÃO	
ESTADO CIVIL	

1. Grau de cooperação durante a entrevista:

Alto Médio Baixo

2. Espontaneidade do informante:

Muita Média Pouca

3. Grau de conhecimento entre informante e pesquisadores:

Grande Médio Pequeno Nenhum

4. Houve participação de outras pessoas? Relacione-as.

OUTRAS OBSERVAÇÕES:

DADOS DO(S) ENTREVISTADOR(ES):

Nome do(s) entrevistador(es):

Dados para contato (telefone e e-mail):

APÊNDICE D – ROTEIRO DE ENTREVISTAS

ROTEIRO DE ENTREVISTAS

DADOS PESSOAIS E FAMÍLIA

1. Você trabalha/estuda? Trabalha com quê ou estuda o quê? Já trabalhou em uma profissão diferente? Por que mudou?
2. Você tem filhos (ou irmãos, sobrinhos etc.)? Como se chamam? Onde moram? O que fazem? Estudam? Trabalham?
3. Você é casado? Casou em qual cidade? Quando? Seu marido/sua esposa são de onde?
4. Tem muitos parentes nesta cidade? Quem são? O que fazem? São casados? E fora daqui?
5. Você conhece alguém que se mudou de Coimbra/Ouro Preto nos últimos anos? Por que essa pessoa foi embora? Você sente/sentiu muita falta?

INFÂNCIA

6. O que você sente ao falar daqueles tempos? Gostaria que voltassem?
7. Acha que eram melhores que hoje? Por quê? O que você acredita que as crianças de hoje querem/pensam?
8. Você seria capaz de lembrar o dia mais feliz de sua infância? Conte como foi. E o mais triste?

CIDADE E PAÍS

9. Você mora em Coimbra/Ouro Preto há quanto tempo? Mora com quem? Sua família é daqui também?
10. Você conhece um pouco da história da cidade? A idade, seu surgimento? Sabe de algum acontecimento que marcou a história de Coimbra/Ouro Preto?
11. Você gosta de Coimbra/Ouro Preto? Por quê?
12. Você acha que Coimbra/Ouro Preto é uma boa cidade para os jovens? Por quê?
13. E para os idosos? A cidade oferece boas condições?
14. Já pensou em sair daqui? Gostaria de morar em outra cidade?
15. O que espera que aconteça em Coimbra/Ouro Preto para melhorar a cidade (política, policiamento, indústrias etc.)?
16. Fale sobre as coisas que você gosta e não gosta em Coimbra/Ouro Preto. Quais são as principais festas/eventos daqui? Você participa?
17. O que você acha da Universidade de Coimbra/Universidade de Ouro Preto? É bom para a população?
18. O que você acha da situação política e econômica do país?

OPINIÕES

19. Para você quais as influências que os meios de comunicação exercem sobre a sociedade?
20. Você gosta de futebol? Torce para qual time? Gosta de outro esporte? Admira algum esportista? O que você está achando do desempenho da seleção?
21. Qual a sua opinião sobre os cidadãos de Coimbra/Ouro Preto?
22. Você acha que existe muito preconceito na cidade ou no país? Qual é sua opinião sobre o assunto?

EXPERIÊNCIAS PESSOAIS

23. Você viaja com frequência? Costuma ir sozinho ou acompanhado? Quais são suas companhias?
24. Você já passou por alguma situação de risco em que achou que não fosse sobreviver? Já presenciou algum acidente? Como foi?
25. Lembra de algum fato acontecido na cidade de muita repercussão? Como e com quem foi?
26. Você já se decepcionou com alguém que gostava muito? Qual foi o motivo? Hoje vocês já se entenderam? Por quê?
27. Você poderia me contar sobre um fato que marcou muito sua família?
28. Qual o dia mais marcante da sua vida? Fale um pouco sobre ele.

RELIGIÃO

29. Qual é a importância da religião na sua vida? Sua família é religiosa?
30. Você acha que Coimbra/Ouro Preto é uma cidade religiosa? Por quê?
31. Você acredita em milagres? Conhece algum caso milagroso?
32. Você acredita em vida após a morte? Na sua opinião, qual é o nosso destino depois da morte?

SOBRENATURAL

33. Em alguma ocasião você já sentiu a presença do sobrenatural? Como foi?
34. Já aconteceu de alguma vez você (ou pessoa que você conhece) dizer ou sonhar com algo e depois isso vir a acontecer realmente? Como foi?
35. Você acha possível alguém prever o futuro? Conhece alguém que passou por isso?

ASPIRAÇÕES

36. Se você pudesse realizar um desejo, qual seria? Qual você acredita ser o maior desejo de cada um?

APÊNDICE E – CODIFICAÇÃO DOS DADOS

G1: Variável dependente	I – Demais Formas de indeterminação G – Nomes gerais
G2: Sexo	m – Masculino f – Feminino
G3: Faixa Etária	1 – 18 a 25 anos 2 – 30 a 45 anos 3 – Acima de 50 anos
G4: Presença ou ausência de locativo	P – Presença A – Ausência
G5: Presença ou ausência de anáfora	X: Retomada anafórica semelhante Y: Ausência de anáfora Z: Retomada anafórica diferente
G6: Grau de indeterminação	U: Parcial com referente implícito V: Parcial com referente explícito T: Total
G7: Mais ou menos coletivo	8 – menos coletivo 9 – mais coletivo