



**UNIVERSIDADE FEDERAL DA BAHIA  
INSTITUTO DE LETRAS**

**PROGRAMA DE PÓS GRADUAÇÃO EM LÍNGUA E CULTURA**

Rua Barão de Jeremoabo, nº 147 - CEP: 40170-290 - Campus Universitário Ondina Salvador – BA Tel.:  
(71) 3283-6256 – Site: <http://www.ppglinc.letas.ufba.br> - E-mail: [pgletba@ufba.br](mailto:pgletba@ufba.br)

**GRACIELLE DE BARROS JESUS**

**A REALIZAÇÃO DO SUJEITO PRONOMINAL NO  
PORTUGUÊS DOS TONGAS DE SÃO TOMÉ E PRÍNCIPE**

Salvador

2022

**GRACIELLE DE BARROS JESUS**

**A REALIZAÇÃO DO SUJEITO PRONOMINAL NO  
PORTUGUÊS DOS TONGAS DE SÃO TOMÉ E PRÍNCIPE**

Tese apresentada ao Programa de Pós-Graduação em Língua e Cultura do Instituto de Letras da Universidade Federal da Bahia, como requisito parcial para a obtenção do grau de Doutora em Língua e Cultura.

Orientador: Prof. Dr. Dante Lucchesi

Coorientador: Prof. Dr. Danniell Carvalho

Salvador

2022

Jesus, Gracielle de Barros.

A realização do sujeito pronominal no português dos Tongas de São Tomé e Príncipe / Gracielle de Barros Jesus. - 2022.

210 f.: il.

Orientador: Prof. Dr. Dante Lucchesi.

Coorientador: Prof. Dr. Danniel Carvalho.

Tese (doutorado) - Universidade Federal da Bahia, Instituto de Letras, Salvador, 2022.

1. Sociolinguística - São Tomé, Ilha de (São Tomé e Príncipe). 2. Língua portuguesa - São Tomé, Ilha de (São Tomé e Príncipe) - Variação. 3. Língua portuguesa - Português falado - São Tomé, Ilha de (São Tomé e Príncipe). 4. Língua portuguesa - Aspectos sociais - São Tomé, Ilha de (São Tomé e Príncipe). 5. Mudanças linguísticas. I. Lucchesi, Dante. II. Universidade Federal da Bahia. Instituto de Letras. III. Título.

CDD - 469.79866994

CDU - 81'27(669.95)

**GRACIELLE DE BARROS JESUS**

**A REALIZAÇÃO DO SUJEITO PRONOMINAL NO  
PORTUGUÊS DOS TONGAS DE SÃO TOMÉ E PRÍNCIPE**

Tese apresentada ao Programa de Pós-Graduação em Língua e Cultura do Instituto de Letras da Universidade Federal da Bahia (PPGLinC/UFBA) como requisito parcial para a obtenção do grau de Doutora em Língua e Cultura.

Aprovada em 31 de maio de 2022.

**BANCA EXAMINADORA**

Professor Doutor Dante Eustacchio Ramaccioti Lucchesi – UFBA / UFF (orientador)

Professora Doutora Lanuza Lima – UFBA / IFBA

Professora Doutora Elisângela Mendes – UFBA / IFBA

Professor Doutor Carlos Felipe Pinto – UFBA

Professora Doutora Maria Cristina Figueiredo - UFBA

*Para (e por) Elis e Clarisse*

## AGRADECIMENTOS

A Danilo, o companheiro que honra diariamente o significado dessa palavra. Seu amor, apoio, incentivo e confiança guiaram meus caminhos em todos os momentos, sobretudo nos mais obscuros. Sem você, eu jamais ousaria a sequer sonhar com esta tese;

Às minhas meninas, Elis e Clarisse, que nasceram junto com esta tese. Todo o esforço do mundo vale a pena em troca da esperança de que vocês se orgulharão de mim um dia;

Aos meus pais e à minha irmã, Antonio, Kadja e Adriele, por terem sido a rede de apoio inabalável que toda mulher deveria ter;

Ao professor Dante Lucchesi, pela orientação cuidadosa, atenta e constante. Ser orientada por um dos maiores nomes da Linguística do Brasil me é causa de muito orgulho e gratidão;

Ao professor Danniell Carvalho, por ter feito considerações de suma importância durante o exame de qualificação e por, a partir de então, se tornar o meu coorientador. Tamanha generosidade será motivo da minha eterna gratidão.

A cada professor que fez parte da minha trajetória acadêmica nos últimos 5 anos;

Aos informantes, cujos dados linguísticos fornecidos são o coração dessa pesquisa;

Ao PPGLinC;

À CAPES, por financiar essa pesquisa.

Toda Causa tem seu Efeito; todo Efeito tem sua Causa; todas as coisas acontecem de acordo com a Lei. O Acaso é simplesmente um nome dado a uma Lei não conhecida; existem muitos planos de causalidade, mas nada escapa à Lei.

O Caibalion – Hermes Trimegisto

## RESUMO

As línguas formadas através do contato massivo são objeto de inúmeros estudos nos mais diversos âmbitos linguísticos, todos visando à compreensão da influência do contato e das vicissitudes socio-históricas que envolveram a formação dessas variedades para as suas características estruturais. Nessa perspectiva, o objetivo desta tese é realizar a descrição e a análise do funcionamento do Parâmetro do Sujeito Nulo no português dos Tongas, variedade de língua portuguesa falada pelos descendentes de africanos continentais contratados para trabalhar nas roças de cacau e café na ilha de São Tomé, no decorrer do século XX. A análise, realizada a partir do aporte teórico da Sociolinguística Paramétrica (TARALLO; KATO, 1989, 2007), da teoria de Princípios e Parâmetros (CHOMSKY, 1981) e da Sociolinguística Variacionista (WEINREICH; LABOV; HERZOG, 2006[1968]; LABOV, 2008[1972]), tem como *corpus* uma amostra de fala composta por 18 entrevistas realizadas na roça de Monte Café, estando os entrevistados estratificados em sexo, idade e escolaridade. A análise tomou por base estudos anteriores responsáveis por apresentar o feixe de propriedades linguísticas que, geralmente, estão associadas à mudança na marcação do Parâmetro do Sujeito Nulo, sobretudo os de Duarte (1995, 2012, 2018, 2019), Lucchesi (2009), Kapetula (2016), Oliveira (2016), dentre outros. Depois de identificadas, as ocorrências obtidas foram codificadas e submetidas ao programa de regras variáveis Goldvarb X, que forneceu os percentuais e pesos relativos para as variáveis selecionadas. Foi encontrado um total de 980 dados na amostra de fala analisada, sendo 50,2% de sujeitos expressos foneticamente e 49,8% de sujeitos nulos. Esses números são bastante significativos, no que diz respeito ao comportamento do fenômeno na comunidade de fala dos Tongas: embora seja inferior ao obtido em outras variedades de língua portuguesa, como o português brasileiro e o português afro-brasileiro, o alto índice de sujeitos expressos aponta para uma provável preferência pelo sujeito realizado, que pode se acentuar em alguns anos. Por outro lado, os resultados, muito próximos aos encontrados para o português moçambicano, mostram que, em comparação ao português brasileiro e ao português afro-brasileiro, ainda há grande influência do português europeu sobre as variedades africanas de língua portuguesa. As variáveis linguísticas selecionadas como estatisticamente relevantes foram Referência anterior ao sujeito, com a preferência pelos sujeitos foneticamente realizados em contextos de primeira menção (peso relativo 0.717), Caracterização semântica do sujeito, sendo o traço [+animado/ -humano] um favorecedor do sujeito expresso (peso relativo 0.597), Presença de constituinte à esquerda do verbo, cujos resultados apresentaram os contextos sem constituintes à esquerda do verbo como favorecedores do sujeito expresso (peso relativo 0.535), e Desinência número-pessoal do verbo, sendo o padrão [+pessoa/ +plural] favorecedor do sujeito expresso (peso relativo 0.648). A única variável social selecionada foi a Faixa etária, sendo os informantes mais jovens os que mais utilizam o sujeito expresso (peso relativo 0.602).

Palavras-chave: português dos Tongas, Parâmetro do Sujeito Nulo, Sociolinguística Variacionista, Sociolinguística Paramétrica, Contato entre línguas.

## ABSTRACT

Languages formed through massive contact are the subject of numerous studies in the most diverse linguistic areas, all aimed at understanding the influence of contact and the socio-historical vicissitudes that involved the formation of these varieties for their structural characteristics. In this perspective, the objective of this thesis is to carry out the description and analysis of the operation of the Null Subject Parameter in Tonga Portuguese, a variety of Portuguese language spoken by descendants of continental Africans hired to work in cocoa and coffee plantations on the island of São Tomé, during the twentieth century. The analysis, carried out from the theoretical contribution of Parametric Sociolinguistics (TARALLO; KATO, 1989, 2007), Principles and Parameters theory (CHOMSKY, 1981) and Variationist Sociolinguistics (WEINREICH; LABOV; HERZOG, 2006[1968]; LABOV, 2008[1972]), has as corpus a speech sample composed of 18 interviews carried out in the Monte Café farm, with the interviewees being stratified by sex, age and education. The analysis was based on previous studies responsible for presenting the bundle of linguistic properties that are generally associated with the change in the marking of the Null Subject Parameter, especially those by Duarte (1995, 2012, 2018, 2019), Lucchesi (2009), Kapetula (2016), Oliveira (2016), among others. After being identified, the occurrences obtained were coded and submitted to the Goldvarb X variable rules program, which provided the percentages and relative weights for the selected variables. A total of 980 data were found in the analyzed speech sample, with 50.2% of subjects expressed phonetically and 49.8% of null subjects. These numbers are quite significant with regard to the behavior of the phenomenon in the Tonga speech community: although it is lower than that obtained in other varieties of Portuguese language, such as Brazilian Portuguese and Afro-Brazilian Portuguese, the high rate of subjects expressed points to a probable preference for the realized subject, which may accentuate in a few years. On the other hand, the results, very close to those found for Mozambican Portuguese, show that, compared to Brazilian Portuguese and Afro-Brazilian Portuguese, there is still a great influence of European Portuguese on African varieties of Portuguese. The linguistic variables selected as statistically relevant were Previous reference to the subject, with a preference for phonetically performed subjects in contexts of first mention (relative weight 0.717), Semantic characterization of the subject, with the feature [+ animated/ -human] favoring the subject expressed (relative weight 0.597), Presence of constituent to the left of the verb, whose results showed the contexts without constituents to the left of the verb as favoring the expressed subject (relative weight 0.535), and Personal-number ending of the verb, with the pattern [+ person/ +plural] favoring the expressed subject (relative weight 0.648). The only social variable selected was the age group, with younger informants the ones who most used the express subject (relative weight 0.602).

Keywords: Tongan Portuguese, Null Subject Parameter, Variationist Sociolinguistics, Parametric Sociolinguistics, Contact between languages.

## LISTA DE ILUSTRAÇÕES

- Figura 1 Ilhas do Golfo da Guiné
- Figura 2 Engenho de açúcar em São Tomé século XVI
- Figura 3 Estrutura das roças
- Figura 4 Rua principal da roça de Monte Café
- Figura 5 Esquema de classificação das línguas em relação à realização do sujeito
- Gráfico 1 A realização do sujeito pronominal no português dos Tongas
- Figura 6 Hierarquia de referencialidade
- Quadro 1 Efeito prosódico dos constituintes que antecedem o verbo

## LISTA DE TABELAS

|           |   |
|-----------|---|
| Tabela 1  | População de São Tomé (XIX – XX)  |
| Tabela 2  | Línguas faladas em São Tomé e Príncipe em função da idade da população                |
| Tabela 3  | Tempo, modo e aspecto no crioulo havaiano   |
| Tabela 4  | Tempo, modo e aspecto no crioulo guineense  |
| Tabela 5  | Fatores demográficos para a formação de línguas em contexto de contato                |
| Tabela 6  | Os pronomes pessoais retos do latim   |
| Tabela 7  | Pronomes pessoais do caso reto no português europeu (MATEUS, 2003, p. 819)            |
| Tabela 8  | Quadro pronominal do português brasileiro   |
| Tabela 9  | O quadro pronominal do português afro-brasileiro                                      |
| Tabela 10 | Quadro pronominal do crioulo de São Tomé (MENDES, 2016, p. 153)                       |
| Tabela 11 | Quadro pronominal do português dos Tongas   |
| Tabela 12 | Frequência de sujeitos nulos em PE  |
| Tabela 13 | O paradigma verbal do PB no decorrer do tempo   |
| Tabela 14 | Uso do sujeito nulo de primeira pessoa em função da Faixa Etária                      |
| Tabela 15 | Uso de sujeito nulo de terceira pessoa em função da Faixa Etária                      |
| Tabela 16 | A realização do sujeito em função do tempo, do padrão sentencial e do traço semântico |
| Tabela 17 | Número de ocorrências em função da pessoa do discurso                                 |
| Tabela 18 | Número de ocorrências em função da desinência verbal                                  |
| Tabela 19 | Número de ocorrências em função do traço semântico                                    |
| Tabela 20 | Hipótese de resultado para a tarefa 1   |
| Tabela 21 | Resultados da primeira tarefa   |
| Tabela 22 | Resultados da segunda tarefa  |
| Tabela 23 | Resultados da terceira tarefa   |

- Tabela 24 Resultados da quarta tarefa
- Tabela 25 Resultados obtidos para a variável Morfologia verbal
- Tabela 26 Resultados obtidos para a variável Traço semântico do sujeito
- Tabela 27 Resultados obtidos para a variável Regularidade morfológica do verbo
- Tabela 28 Resultados obtidos para a variável Faixa etária
- Tabela 29 Resultados obtidos para a variável Tempo verbal
- Tabela 30 Expressão/ omissão do sujeito em algumas variedades do português
- Tabela 31 A realização do sujeito pronominal em função da variável Pessoa do discurso
- Tabela 32 A primeira pessoa do plural e a forma verbal
- Tabela 33 A primeira pessoa do singular e a forma verbal
- Tabela 34 Referência anterior ao sujeito
- Tabela 35 Sujeito de 3ª pessoa X feixe de traços semânticos (valor de aplicação: sujeito nulo)
- Tabela 36 O sujeito pronominal realizado e a variável Traço semântico do referente
- Tabela 37 Realização do sujeito pronominal em função da variável Faixa Etária
- Tabela 38 Realização do sujeito e a variável Presença de constituinte à esquerda do verbo
- Tabela 39 Uso do sujeito nulo X Pessoa do discurso e desinência verbal
- Tabela 40 Paradigmas verbais
- Tabela 41 O sujeito expresso e a variável Desinência número-pessoal
- Tabela 42 Cruzamento entre Sexo e Faixa Etária
- Tabela 43 Cruzamento entre Sexo e Escolaridade
- Tabela 44 O sujeito expresso e a variável Tipo de frase
- Tabela 45 Sujeito nulo e a variável Tipo sintático da oração no PB
- Tabela 46 Realização do sujeito pronominal em função da variável tipo de oração
- Tabela 47 Paradigma flexional do Kimbundu

## SUMÁRIO

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>INTRODUÇÃO</b>  | <b>16</b> |
| <b>2</b> | <b>O CONTEXTO SÓCIO-HISTÓRICO DE SÃO TOMÉ E AS LÍNGUAS FALADAS NO PAÍS</b>                                     | <b>20</b> |
| 2.1      | A SOCIO-HISTÓRIA DE SÃO TOMÉ E PRÍNCIPE  | 20        |
| 2.1.1    | <b>A primeira colonização de São Tomé: o ciclo do açúcar (XV-XVI)</b>  | <b>22</b> |
| 2.1.2    | <b>A segunda colonização da ilha de São Tomé: o ciclo do café e cacau e os trabalhadores contratados (XIX)</b> | <b>26</b> |
| 2.2      | O ESTABELECIMENTO E A ESTRUTURA DAS ROÇAS  | 33        |
| 2.2.1    | <b>A roça de Monte Café</b>  | <b>37</b> |
| 2.3      | VARIEDADES LINGUÍSTICAS DE SÃO TOMÉ  | 38        |
| 2.3.1    | <b>O forro</b>   | <b>39</b> |
| 2.3.2    | <b>O lung'le</b>   | <b>40</b> |
| 2.3.3    | <b>O angolar</b>   | <b>41</b> |
| 2.3.4    | <b>O kabuverdianu</b>  | <b>43</b> |
| 2.3.5    | <b>O português de São Tomé</b>   | <b>43</b> |
| 2.3.6    | <b>O português dos Tongas</b>  | <b>45</b> |
| <b>3</b> | <b>PRESSUPOSTOS TEÓRICOS</b>   | <b>49</b> |
| 3.1      | A SOCIOLINGUÍSTICA VARIACIONISTA   | 49        |
| 3.1.1    | <b>A variável linguística</b>  | <b>51</b> |
| 3.1.2    | <b>Fundamentos empíricos para uma teoria da mudança</b>  | <b>54</b> |
| 3.1.3    | <b>Comunidade de fala <i>versus</i> Comunidade de prática</b>  | <b>58</b> |
| 3.2      | O ESTUDO CIENTÍFICO DO CONTATO ENTRE LÍNGUAS MASSIVO E RADICAL   | 60        |
| 3.2.1    | <b>A Crioulística: conceitos fundamentais</b>  | <b>60</b> |
| 3.2.2    | <b>A Transmissão Linguística Irregular</b>   | <b>67</b> |
| <b>4</b> | <b>O SUJEITO PRONOMINAL NAS DIVERSAS VARIEDADES DA LÍNGUA PORTUGUESA</b>                                       | <b>74</b> |
| 4.1      | O CONCEITO DE PRONOME E DE PESSOA PRONOMINAL   | 74        |
| 4.2      | OS PRONOMES PESSOAIS DO CASO RETO: DO LATIM AO   |           |

|              |  |            |
|--------------|--|------------|
|              | PORTUGUÊS DOS TONGAS   | 77         |
| 4.3          | O CONCEITO DE SUJEITO  | 85         |
| 4.4          | O PARÂMETRO DO SUJEITO NULO  | 91         |
| <b>4.4.1</b> | <b>O Parâmetro do Sujeito Nulo em algumas variedades de língua portuguesa</b>            | <b>96</b>  |
| 4.4.1.1      | <i>O sujeito pronominal no português coloquial europeu (DUARTE, 1995)</i>                | 96         |
| 4.4.1.2      | <i>A perda do princípio “evite pronome” no português brasileiro (DUARTE, 1995; 2012)</i> | 99         |
| 4.4.1.2.1    | Refinando o tratamento da 3ª pessoa (DUARTE, 2012)                                       | 103        |
| 4.4.1.3      | <i>O parâmetro do sujeito nulo no português afro-brasileiro (LUCCHESI, 2009)</i>         | 106        |
| 4.4.1.4      | <i>Interpretação dos sujeitos nulos no português de Angola (KAPETULA, 2016)</i>          | 109        |
| 4.4.1.5      | <i>A expressão do sujeito no português de Moçambique (OLIVEIRA, 2016)</i>                | 116        |
| 4.4.1.6      | <i>Null subjects in (Romance-derived) creoles: routes of evolution (LIPSKI, 1999)</i>    | 121        |
| <b>5</b>     | <b>A METODOLOGIA</b>   | <b>128</b> |
| 5.1          | A CONSTITUIÇÃO DAS AMOSTRAS DE FALA  | 128        |
| <b>5.1.1</b> | <b>O vernáculo e o paradoxo do observador</b>  | <b>131</b> |
| <b>5.1.2</b> | <b>O perfil dos informantes</b>  | <b>133</b> |
| <b>5.1.3</b> | <b>A natureza das entrevistas</b>  | <b>137</b> |
| 5.2          | O ENVELOPE DA VARIAÇÃO   | 138        |
| 5.2.1        | <b>A variável dependente</b>   | <b>138</b> |
| 5.2.1.1      | <i>Critérios para inclusão e exclusão de dados</i>                                       | 139        |
| 5.2.1.2      | <i>As variáveis linguísticas</i>   | 141        |
| 5.2.1.2.1    | Pessoa pronominal  | 141        |
| 5.2.1.2.2    | Desinência número-pessoal do verbo   | 142        |
| 5.2.1.2.3    | Caracterização semântica do sujeito  | 143        |
| 5.2.1.2.4    | Tipo de frase  | 143        |
| 5.2.1.2.5    | Paralelismo formal   | 144        |

|           |  |     |
|-----------|--|-----|
| 5.2.1.2.6 | Localização em relação ao antecedente e as condições de referencialidade                           | 145 |
| 5.2.1.2.7 | Presença de constituinte à esquerda do verbo na oração   | 147 |
| 5.2.1.3   | <i>As variáveis sociais</i>  | 148 |
| 5.2.1.3.1 | Sexo/gênero  | 149 |
| 5.2.1.3.2 | Faixa etária   | 150 |
| 5.2.1.3.3 | Escolaridade   | 151 |
| 5.2.1.3.4 | Estada fora da comunidade  | 152 |
| 5.3       | PROCESSAMENTO QUANTITATIVO DOS DADOS   | 153 |
| 6         | <b>ANÁLISE VARIACIONISTA DA REALIZAÇÃO DO SUJEITO PRONOMINAL NO PORTUGUÊS DOS TONGAS</b>           | 156 |
| 6.1       | A REALIZAÇÃO/ NÃO REALIZAÇÃO DO SUJEITO PRONOMINAL NO PORTUGUÊS DOS TONGAS                         | 156 |
| 6.1.1     | <b>O Parâmetro do Sujeito Nulo e a variável Pessoa do sujeito: algumas considerações</b>           | 159 |
| 6.2       | AS VARIÁVEIS SELECIONADAS PELO GOLDVARB X  | 163 |
| 6.2.1     | <b>Referência anterior ao sujeito</b>  | 164 |
| 6.2.2     | <b>Caracterização semântica do sujeito</b>   | 167 |
| 6.2.3     | <b>Faixa etária</b>  | 169 |
| 6.2.4     | <b>Presença de constituinte à esquerda do verbo</b>  | 171 |
| 6.2.5     | <b>Desinência número-pessoal do verbo</b>  | 173 |
| 6.3       | A INTERAÇÃO ENTRE AS VARIÁVEIS SOCIAIS: CRUZAMENTO ENTRE SEXO X FAIXA ETÁRIA E SEXO X ESCOLARIDADE | 178 |
| 6.4       | VARIÁVEIS LINGUÍSTICAS NÃO SELECIONADAS  | 180 |
| 6.4.1     | <b>Tipo de frase</b>   | 181 |
| 6.4.2     | <b>Posição do sujeito em relação ao antecedente</b>  | 182 |
| 7         | <b>CONSIDERAÇÕES FINAIS</b>  | 185 |
|           | <b>REFERÊNCIAS</b>   | 191 |

## 1 INTRODUÇÃO

Sabe-se que a variação é um fenômeno inerente a todas as línguas humanas, e que, uma vez motivada por fatores internos e externos à estrutura linguística, consiste em um fenômeno condicionado, e não aleatório. Assim, propomos, nesta tese, o estudo da variação da realização do sujeito pronominal – ou do Parâmetro do Sujeito Nulo - no português dos Tongas, população descendente dos africanos continentais contratados para trabalhar nas grandes roças de café e cacau de São Tomé e Príncipe, na costa ocidental da África, entre os séculos XIX e XX. Nesta análise, pretendemos, além de descrever o *status* do fenômeno nessa variedade linguística, escrutinar quais fatores extra e intralinguísticos são relevantes para essa variação.

São muitos os estudos acerca do funcionamento do Parâmetro do Sujeito Nulo nas línguas humanas, sendo que esse parâmetro rege a possibilidade de alternância entre a presença e a ausência de um elemento na posição de sujeito em sentenças com verbos finitos. Muito já foi descoberto acerca desse parâmetro, sobretudo através de estudos prototípicos como o de Duarte (1995). Entretanto, investigações científicas são inesgotáveis, e, por isso, ainda há muito para descobrir sobre os fatores que licenciam a supressão do sujeito pronominal, sobretudo em línguas formadas em situações de contato massivo entre línguas tipologicamente diferentes. Nesse sentido, os estudos acerca do Parâmetro do Sujeito Nulo em variedades linguísticas faladas em comunidades rurais são particularmente interessantes para observar os fatores que atuam quando esse parâmetro se encontra em processo de mudança. No português dos Tongas, por exemplo, observa-se um processo similar ao que ocorreu no português brasileiro, com a inserção das formas gramaticalizadas *você* e *a gente* no quadro dos pronomes pessoais como um fator motivador da variabilidade da realização pronominal. Resta, então, avaliar quais outras questões estruturais e sociais se mostram importantes para a mudança do parâmetro, verificando se são as mesmas questões causadoras das mudanças em variedades urbanas da língua.

A análise aqui apresentada se insere no âmbito da Sociolinguística Paramétrica, reunindo os modelos teórico-metodológicos da Sociolinguística Quantitativa e da Teoria de Princípios e Parâmetros. O *corpus* utilizado é constituído por 18 entrevistas, realizadas em São Tomé e Príncipe, em 1992. Os informantes estão estratificados em sexo (9 homens e 9 mulheres) e divididos em três faixas etárias: faixa 1 (20-40 anos), faixa 2 (41-60 anos) e faixa 3 (60 anos ou mais). Essa investigação tem como foco a fala popular e, por esse motivo, os falantes entrevistados são pessoas com pouco ou nenhum acesso à escolarização formal.

Temos, então, como objetivo geral do trabalho, descrever e analisar, com base na Sociolinguística Variacionista e no modelo gerativista de Princípios e Parâmetros (CHOMSKY, 1981), os sujeitos pronominais de referência definida na variedade da língua portuguesa dos Tongas, falada na localidade de Monte Café, na ilha de São Tomé. Como objetivos específicos, temos:

a) Verificar se, assim como no português brasileiro, há um movimento de mudança em direção aos sujeitos preferencialmente realizados foneticamente;

b) Verificar se há, no português dos Tongas, a redução do paradigma flexional verbal, bastante reduzido nas variedades afro-brasileiras do português (LUCCHESI, 2009).

c) Traçar similaridades e/ou diferenças entre esses resultados e os encontrados em estudos realizados em outras variedades de língua portuguesa formadas em circunstâncias semelhantes às que ensejaram a formação do português dos Tongas.

A socio-história das línguas é fundamental para a compreensão da sua configuração estrutural, e essa relação fundamenta a hipótese de que as circunstâncias que envolveram a formação da variedade de língua portuguesa falada em Monte Café deixaram vestígios recuperáveis através da análise da estrutura linguística, sendo esses vestígios comparáveis aos verificados em outras variedades de línguas formadas através de contato massivo.

Com essa fundamentação e objetivos, esta tese está estruturada da seguinte forma:

Na seção 2, intitulada *O contexto socio-histórico de São Tomé e as línguas faladas no país*, apresentamos os dados demográficos e a socio-história de São Tomé e Príncipe. Em 2.1, apresentaremos os ciclos de colonização de São Tomé, comentando as circunstâncias que envolveram a formação da sociedade são tomense e a importância dos dois ciclos de colonização e povoamento para a história do país. Na subseção 2.2, apresentamos a concepção e a estrutura das roças, em torno das quais a população de São Tomé se estabeleceu. Apresentamos, também, a roça de Monte Café, onde vivem os informantes que forneceram os dados para essa pesquisa. Em 2.3, *Variedades linguísticas de São Tomé*, apresentamos os grupos étnicos que vivem em São Tomé e as línguas que são faladas na ilha.

A seção 3, *Pressupostos Teóricos*, traz as bases teóricas que sustentam as discussões estabelecidas aqui. Na subseção 3.1, intitulada Sociolinguística Variacionista, apresentaremos a gênese dessa vertente dos estudos linguísticos e os seus principais conceitos: variável linguística, variação estável, mudança em curso, tempo real, tempo aparente, comunidade de fala e comunidade de prática. Comentamos, também, os mecanismos através dos quais a mudança linguística acontece, formalizados através dos chamados *cinco problemas da mudança linguística*. Em 3.2, apresentamos as duas principais teorias que objetivam explicar a influência do contato massivo entre línguas para as novas variedades linguísticas que se formam: a Crioulística, em 3.2.1, e a Transmissão Linguística Irregular, em 3.2.2.

A seção 4, que tem como título *O sujeito pronominal nas diversas variedades da língua portuguesa*, apresenta as bases para a compreensão do funcionamento do sujeito pronominal nas línguas humanas. Trazemos, na subseção 4.1, os conceitos de pronome e de pessoa pronominal, buscando debater as imprecisões epistemológicas desses conceitos com o embasamento de um dos mais importantes trabalhos na área, *A natureza dos pronomes*, de Émile Benveniste (1956). Na subseção 4.2, apresentamos os quadros dos pronomes pessoais retos do latim, do português europeu, do português brasileiro, do português afro-brasileiro, do crioulo de São Tomé e do português dos Tongas, tendo sido este último elaborado por nós, a partir da observação dos dados. Em 4.3, discutiremos a noção de sujeito através

das concepções apresentadas por Rocha Lima (2011), Evanildo Bechara (2009), Ataliba de Castilho (2010) e Maria Eugênia Duarte (2007). Em 4.4, discutimos o Parâmetro do Sujeito Nulo, comentando sua gênese e aplicabilidade, bem como as características das línguas de marcação positiva e negativa para esse parâmetro. Por fim, em 4.4.1, apresentamos, resumidamente, os resultados encontrados por alguns estudos que focalizaram o Parâmetro do Sujeito Nulo em algumas variedades de língua portuguesa e em algumas línguas crioulas.

A quinta seção, *A metodologia*, busca apresentar os passos que foram seguidos para que se alcançasse os resultados aqui apresentados. Para tanto, discutimos, na subseção 5.1, os passos que devem reger a constituição das amostras de fala, apresentando, também, formas de mitigar o chamado *paradoxo do observador*, com o objetivo de alcançar o *vernáculo* dos informantes. Apresentamos, em 5.1.2, o perfil social dos informantes e os dados relativos às demais línguas que falam, à descendência de seus pais e ao período de estada fora da comunidade e, em 5.1.3, comentamos a natureza das entrevistas e o método de gravação e de transcrição. Na subseção 5.2, trazemos o envelope da variação, apresentando detalhadamente a variável dependente, a realização/não realização do sujeito pronominal, bem como os critérios seguidos para considerar ou desconsiderar dados. Além disso, apresentamos as variáveis linguísticas e sociais empregadas na análise. Por fim, em 5.3, discutimos o suporte quantitativo da análise, apresentando os passos seguidos para o processamento quantitativo dos dados.

Na seção 6, *Análise variacionista da realização do sujeito pronominal no português dos Tongas*, apresentamos os resultados obtidos na análise do Parâmetro do Sujeito Nulo no português tonga, comentando as frequências encontradas e as variáveis linguísticas e sociais selecionadas e descartadas pelo Goldvarb X.

Por fim, nas *Considerações Finais*, apresentamos um resumo das discussões realizadas no decorrer da tese, sintetizando os resultados alcançados e os possíveis próximos passos.

## **2 O CONTEXTO SOCIO-HISTÓRICO DE SÃO TOMÉ E AS LÍNGUAS FALADAS NO PAÍS**

Nesta seção, apresentaremos as circunstâncias que envolveram a consolidação da ilha de São Tomé como colônia portuguesa e, posteriormente, como país independente. Comentaremos, também, acerca dos fatos socio-históricos que deram forma às variedades linguísticas que se desenvolveram e são faladas no país. Desse modo, a seção será dividida em três subseções. Em 2.1, apresentaremos as duas fases do povoamento de São Tomé. Na subseção 2.2, discutiremos a formação e a estrutura das grandes roças de cacau e café, formadas durante a segunda colonização de São Tomé. Por fim, em 2.3, apresentaremos os aspectos gerais das línguas faladas em São Tomé: o forro, o lung'le, o angolár, o kabuverdianu, o português de São Tomé e o português dos Tongas.

### **2.1 A SOCIO-HISTÓRIA DE SÃO TOMÉ E PRÍNCIPE**

São Tomé e Príncipe, oficialmente denominado República Democrática de São Tomé e Príncipe, é um pequeno país insular localizado na Costa Ocidental Africana, mais especificamente na região do Golfo da Guiné. O país é formado por duas ilhas: a ilha de São Tomé, dividida em seis pequenos distritos, na qual se localiza a capital do país, São Tomé, e a ilha de Príncipe, formada por apenas um distrito e localizada ao norte (HAGEMEIJER, 2009).

Em relação à demografia, trata-se de um país com pouco mais de 197 mil habitantes (INE 2017), sendo importante comentar que o reduzido contingente populacional e as modestas dimensões geográficas não são empecilhos para a sua grande variedade linguística: nas palavras de Hagemeijer (2017), São Tomé é um “labirinto e laboratório de línguas”, característica que é uma marca das circunstâncias que envolveram o processo de povoamento e colonização da ilha.

Figura 1: Ilhas do Golfo da Guiné



Fonte: <https://www.naval.com.br/blog>. Acesso em 25 de setembro de 2020.

A sociedade são-tomense foi uma das primeiras insulares formadas através da miscigenação de povos de origens distintas. Os europeus, embora detentores de considerável homogeneidade étnica e cultural e de maior poder político, sempre foram demograficamente minoritários em relação à população africana, que, além de numericamente superior, foi oriunda de diferentes pontos da África, o que garantiu a esse primeiro contingente populacional significativa diversidade (HAGEMEIJER, 2009). Tinha-se, então, no início do povoamento da ilha, um quadro sociolinguístico extremamente propício ao surgimento de línguas pidgin e crioulas:

Devido às tarefas de povoamento e à ausência de uma atividade económica de envergadura, a sociedade de habitação caracterizava-se por um baixo número de povoadores, essencialmente de sexo masculino, que recorria a mão-de-obra africana, em maioria numérica, para as tarefas domésticas, rurais e obras. Por força das circunstâncias, a fase de habitação terá sido, desde logo, propícia à crioulação. Havia, por um lado, um contato mais intenso entre portugueses e escravos e, por outro, existia a necessidade iminente de comunicação que implicava uma aproximação, por parte dos escravos, ao código linguístico utilizado pelos povoadores portugueses (HAGEMEIJER, 2009, p. 2).

Por esses fatores, acredita-se que pode ter surgido, durante os primeiros anos de colonização, um pidgin utilizado para a comunicação emergencial entre os autóctones e a população europeia, bem como podem ter sido comuns outras estratégias de simplificação linguística com vistas a facilitar a comunicação, que certamente foram relevantes para as variedades linguísticas que se formaram a partir daí.

### **2.1.1 A primeira colonização de São Tomé: o ciclo do açúcar (XV-XVI)**

A ocupação de São Tomé se deu em duas fases distintas: a fase da habitação, compreendida entre 1485 até 1515, quando iniciou-se o cultivo de açúcar na ilha, e a fase de plantação, compreendida entre 1515 e os anos finais do século XVI, quando se observou o declínio da cultura açucareira (GARFIELD, 1992). Porém, segundo Hagemeyer (2009, p. 2), há fortes indícios de que a colonização da ilha se inicia, de forma efetiva, apenas em 1493, quando o fidalgo português Álvaro de Caminha recebe a capitania e para lá se desloca, juntamente com colonos brancos, que receberam terras através do regime de sesmarias, duas mil crianças judias que foram retiradas de seus pais, fugidos da Espanha para Portugal em 1492, alguns degredados e escravos da África Continental (MAGALHÃES, 2009). A partir daí, tem início um dos mais longos períodos de domínio colonial de que se tem notícia, uma vez que o pequeno país foi uma colônia portuguesa por quase quinhentos anos.

Inicialmente, o povoamento de São Tomé objetivava o estabelecimento e ascensão do empreendimento açucareiro, o que foi possível através da instalação de engenhos de açúcar principalmente na faixa litoral do país. Nesse período, os africanos escravizados eram trazidos principalmente do delta do rio Niger, mais especificamente da região entre o rio Real e o reino de Benin, e também dos reinos do Congo e do Ndongo (Angola), sendo que esses contingentes populacionais eram falantes, respectivamente, das línguas Edo, Kikongo e Kimbundo (SEIBERT, 2015, p 100).

As boas condições de plantio da ilha, tais como terrenos férteis, clima equatorial, bons cursos aquáticos e proximidade da costa marítima favoreceram o estabelecimento do plantio da cana-de-açúcar, que se efetivou a partir do ano de



noção que se tinha de inseparabilidade entre a cultura sacarina e a submissão dos povos africanos através da escravidão, uma vez que o sucesso de cada engenho estava atrelado à sujeição da população africana ao trabalho escravo, em condições sub-humanas (SEIBERT, 2015). Esses números, inclusive, possivelmente não condizem com a realidade, devido ao método de contabilização de mão de obra escrava da época:

Escravo era mercadoria susceptível de ser contabilizada. Uma peça era uma medida padrão correspondente a um escravo na pujança do vigor, com uma estatura de 1,75m. A peça das índias representa um negro de 15 a 25 anos. Um de 8 a 15 ou de 25 a 35 não forma uma peça inteira: são preciso três pra fazer duas peças. As crianças com menos de 8 anos e os adultos de 35 a 45 contam por meia peça (FERRONHA, 1989, p. 357 apud GUIMARÃES, 2010, p. 4).

Há, nessa fase do cultivo de cana-de-açúcar em São Tomé, uma mudança na rota do resgate de mão de obra escrava, o que teria importantes implicações para as variedades linguísticas que se formariam:

Ao contrário da fase de habitação, claramente ligada ao delta do Níger, a fase de plantação deve ser associada à predominância do resgate em zonas onde eram faladas línguas bantu, nomeadamente variedades do kikongo e do kimbundu, tipologicamente muito distintas do Edo do reino de Benim. Como veremos, esta diferença tipológica, que demarca fases distintas, permite reconstituir, em linhas gerais, a origem das línguas crioulas do Golfo da Guiné (HAGEMEIJER, 2009, p. 2).

Desde o início de sua história, São Tomé apresentou a miscigenação étnica, sobretudo luso-africana, como a mais preponderante característica do seu contingente populacional: cada povoador recebia, através de decreto régio, o direito a uma escrava capturada na Costa Africana, com vistas a salvaguardar o povoamento da ilha. Essas africanas escravizadas, bem como os filhos gerados por essas relações, receberam liberdade no ano de 1515. Logo depois, em 1517, também receberam alforria os primeiros escravizados, que chegaram com os colonos (HAGEMEIGER, 1999, 2009).

Assim, a mestiçagem, tanto linguística quanto étnica e cultural, permitiu a emergência de um contingente populacional com direitos e características próprias: os forros.

Nessa fase da formação da sociedade colonial, os mestiços rivalizaram com os brancos e, por outro lado, tentaram impedir a ascensão social dos negros forros. Formou-se uma elite local constituída por europeus residentes, mestiços e negros livres. Essas categorias traficaram escravos, detiveram escravos e terras, ocuparam cargos seculares e religiosos e obtiveram patentes militares (NASCIMENTO, 2008 apud SEIBERT, 2015, p.105).

Esse grupo de ex-escravos e descendentes formaram a Confraria de Nossa Senhora do Rosário dos Negros de São Tomé, instituição que obteve “privilégios” registrados em carta régia em 1526, prova da dinamicidade da população negra na sociedade são-tomense. Assim, o desejo de um espaço europeizado impunha à coroa portuguesa a necessidade cada vez maior de disputar o espaço da ilha, o que complexificou substancialmente os modelos administrativos na época (GUIMARÃES, 2010).

A monocultura sacarina seguiu sua expansão na ilha de São Tomé até aproximadamente 1580. Nessa fase, a prosperidade incontestável dos engenhos de açúcar exigia, no ponto de vista dos colonizadores, maior endurecimento do regime de trabalho e um maior contingente de mão de obra escravizada, fatores que motivaram fugas recorrentes, que eram favorecidas pela topografia difícil da ilha. Se tornaram comuns, também, os ataques às áreas de cultivo de cana-de-açúcar e de plantação para subsistência (GUIMARÃES, 2010). Esses fatores internos foram bastante significativos para o declínio da cultura de açúcar em São Tomé, embora o fato preponderante tenha sido a emergência do plantio de açúcar no Brasil:

Os engenhos brasileiros eram mais produtivos, e a qualidade do açúcar, muito superior à de São Tomé. Devido à umidade, o açúcar de São Tomé era pouco consistente e escuro, enquanto o brasileiro era firme, bem refinado e branco. [...] Além disso, no Brasil o clima era menos insalubre para os europeus e houve menos instabilidade política do que nas ilhas no Golfo da Guiné. Consequentemente, muitos fazendeiros se mudaram para o Brasil, onde reproduziram o sistema de plantation (SEIBERT, 2015, p. 107).

Assim, a agricultura açucareira, no decorrer do século XVII, deixou paulatinamente de existir, passando as atividades agrícolas a voltarem-se para a plantação de mantimentos para subsistência e abastecimento dos navios negreiros, sendo cultivados os mais diversos produtos, tais como farinha de mandioca, milho, feijão, banana, inhame, coco, azeite de dendê, bois, porcos, cabras e sal (SILVA,

2002). O tráfico de escravos, agora em menor escala, voltou-se para a região do Gabão, até cessar totalmente em 1850. A partir daí, a desproporção entre negros e brancos passou a ser cada maior: em 1758, de uma população total de 12.672 nas duas ilhas, havia 53 brancos (0,4%) e 8.880 escravos (70%) (SEIBERT, 2014).

Depois dessa transformação ocorrida em São Tomé e devido ao longo período de instabilidade política pelo qual Portugal passou, a ilha passa a ser “esquecida” pela coroa, passando a colônia a ser ponto de parada para o comércio de escravos trazidos para o Brasil. Assim, os forros, filhos de colonos portugueses com nativas da ilha e ex-escravos, passam a dominar a administração da região, situação que se seguiu praticamente inalterada pelos 200 anos seguintes, aproximadamente (GONÇALVES, 2016; SANTIAGO, AGOSTINHO, 2020).

Considerando a importância dos fatores socio-históricos para a formação de variedades de língua, o isolamento de São Tomé teve fortes implicações sociolinguísticas: a saída em massa da população portuguesa da ilha, motivada pelo declínio da cultura sacarina, contribuiu significativamente para a expansão de línguas crioulas, uma vez que a falta de referência da língua-alvo, o português, propiciou maior influência das línguas de substrato.

### **2.1.2 A segunda colonização da ilha de São Tomé: o ciclo do café e cacau e os trabalhadores contratados (XIX)**

A partir da segunda parte do século XIX, a ilha de São Tomé volta vivenciar a experiência de um grande ciclo econômico, com a introdução do plantio de cacau e café, influenciada pelo sucesso desse empreendimento no Brasil. Esse período, que coincide com a abolição da escravatura, ocorrida em 1869, e, em seguida, com a formalização das condições jurídicas da população recém-liberta, provoca importantes transformações na estrutura social e linguística da ilha, conforme explanaremos a seguir. A capital do arquipélago, que desde 1753 era a vila de Santo Antônio, no Príncipe, voltou a ser São Tomé em 1852, o que marca o início efetivo da chamada recolonização (NASCIMENTO, 2008). Sobre a população forra e seu *status* na sociedade colonial do século XIX, Seibert afirma que:

No início da recolonização, a maior parte das terras estava nas mãos dos forros, ou por título de posse ou por usufruto habitual. Ainda em 1872, 96 de 153 proprietários foram classificados como negros. Inicialmente, proprietários forros participaram com sucesso na produção das novas culturas de rendimento. Os mais abastecidos conseguiram mandar seus filhos para estudar em Portugal. No início do século XX, houve em Lisboa mais estudantes são-tomenses do que de todas as outras quatro colônias em conjunto. Muitos aderiram às associações pan-africanistas da época e opuseram-se à discriminação racial promovida pelo colonialismo europeu. Contudo, muitos proprietários forros foram sucessivamente expropriados e politicamente marginalizados pelos portugueses. Os forros perderam suas terras por venda, mas também por práticas fraudulentas e pela força. Os portugueses também se beneficiaram de títulos de posse defeituosos, da ausência de um cadastro atualizado e de disputas pela terra devido a fronteiras mal delineadas. Além disso, as autoridades coloniais não reconheceram seu direito de usufruto de terras pertencentes à coroa que foram vendidas a colonos brancos. No fim do século XIX, os proprietários portugueses possuíam 90% das terras e dominaram a economia do arquipélago (2015, p. 109).

Assim, apostando em um modelo análogo ao aplicado nas monoculturas de açúcar, os portugueses vão, gradativamente, se apossando das terras dos forros, expandindo cada vez mais as roças de café e cacau, inclusive por áreas até então inexploradas das ilhas. A população nativa, que ocupava cargos altos no clero e na administração colonial como um todo, passa a ser afastada desses cargos, tendo de si destituídos a cidadania e os direitos, situação que, com o passar dos anos, se tornaria insustentável e culminaria, no ano de 1953, em um grande e violento confronto entre a população nativa e as autoridades coloniais da ilha de São Tomé: O massacre de Batepá (SEIBERT, 2015).

A mão de obra escrava era, de forma clandestina, a empregada nas novas monoculturas de café e cacau – as chamadas roças. Essa mão de obra era importada, sobretudo, das colônias vizinhas, atividade que era possível mesmo após a abolição do tráfico negreiro. Entretanto, a lei aprovada em 1875, que prenunciava a abolição total da escravidão no império colonial para o ano seguinte, mudou o estado das coisas: a população escravizada, de posse da notícia, deixou as plantações e exigiu o fim imediato da escravidão – solicitação revoltosa que não pôde ser contida pelo então governador Gregório Ribeiro (SEIBERT, 2014, 2015).

Esse novo contingente livre, que posteriormente seria integrado à população crioula, se recusou a voltar às plantações de cacau e café, assim como os crioulos

nativos, que consideravam inferior e degradante o trabalho nas lavouras. Essa recusa fez com que os portugueses recorressem ao emprego de mão de obra não mais de africanos continentais escravizados, mas sim de trabalhadores contratados, regime instituído a partir do ano de 1875 e que durou até os anos 1950 (NASCIMENTO, 2001). Esse novo regime de trabalho consistia na contratação de trabalhadores oriundos de outras colônias portuguesas (Angola, Cabo Verde e Moçambique, principalmente) para trabalhar nas grandes roças são-tomenses.

É notável a importância da segunda colonização para as alterações sociolinguísticas que se observaram na ilha de São Tomé a partir de então. A introdução do contingente populacional dos contratados em circunstâncias análogas à escravidão, com vistas a “repor os contingentes de braços perdidos na crise braçal de 1875” (NASCIMENTO, 2001, p. 7), modificou significativamente a dinâmica da ilha, sobretudo no que diz respeito à demografia e ao então predomínio das línguas crioulas. A tabela a seguir, extraída de Gonçalves e Hagemeyer (2015), mostra as mudanças demográficas ocasionadas pela inserção do regime de contratos na ilha:

Tabela 1: População de São Tomé (XIX – XX)

| <b>Ano</b> | <b>Nativos</b> | <b>Portugueses</b> | <b>Serviçais (contratados)</b> | <b>Total</b> |
|------------|----------------|--------------------|--------------------------------|--------------|
| 1807       | 11636          | 121                | --                             | 11767        |
| 1827       | --             | --                 | --                             | 12713        |
| 1843       | 12568          | 185                | --                             | 12753        |
| 1860       | 10282          | 151                | --                             | 10433        |
| 1870       | 17568          | 449                | --                             | 18017        |
| 1875       | ?              | 741                | ?                              | 29441        |
| 1900       | 18128          | 1012               | 18033                          | 42103        |
| 1921       | 19196          | 998                | 38697                          | 59055        |
| 1940       | 31036          | 995                | 28459                          | 60490        |
| 1950       | 34947          | 1152               | 24060                          | 60159        |

Fonte: GONÇALVES, HAGEMEIJER, 2015, p. 89.

É importante salientar que a reposição da mão de obra, uma vez da recusa da população forra à submissão ao trabalho agrícola imposto pelos roceiros europeus, não foi a única motivação pela busca de trabalhadores fora da ilha de São Tomé:

A importação de serviçais não só ajudaria a transformar economicamente as ilhas como acabaria por escorar a hegemonia dos roceiros. [...] o afluxo de braços permitiria a tão desejada dissociação entre as roças e o tecido social no qual elas se implantaram. Na verdade, era mais fácil sujeitar aos padrões de vida e de trabalho das roças, indivíduos socialmente desenraizados e sem qualquer ligação ao universo social circundante. Mais importante, emergiriam, por via de fenómenos complexos e amiúde contraditórios próprios da vida em universos fechados, processos de identificação com a roça e com o roceiro e, concomitantemente, de afastamento ou mesmo antagonismo face aos ilhéus. Deste ponto de vista, algo da hegemonia dos roceiros escorou-se na importação de um segmento estranho às ilhas e, ao contrário de retrospectivos apriorismos ideológicos sobre a solidariedade racial, aos ilhéus. A distância social procurada pelos ilhéus que, também eles empregadores de serviçais e com padrões de vida cosmopolitas e pautados pela sociabilidade europeia, não podiam deixar de sentir genuinamente diferentes dos serviçais, contribuía para a animosidade recíproca entre uns e outros (NASCIMENTO, 2001, p. 8).

Assim, a partir da entrada em massa desse novo contingente populacional, advindo de diferentes regiões do contingente africano, passam a fazer parte da ilha de São Tomé, além dos crioulos autóctones e da língua portuguesa, as línguas dos serviçais, tais como o kimbundo, o umbundo e o crioulo caboverdiano (GONÇALVES, HAGEMEIJER, 2015). O contato dessas línguas com o português dos roceiros foi o responsável pelo surgimento do chamado português dos Tongas, conforme falaremos posteriormente.

Apesar de serem admitidos através de contratos formais, as disposições legais desse regime de contratação eram apenas teóricas. Os trabalhadores contratados eram, ao chegar em São Tomé, destituídos de sua vontade, fato que pode ser comprovado pela renovação automática dos contratos de trabalho, que deveriam ter duração de cinco anos (NASCIMENTO, 2013). O descumprimento dessa cláusula contratual, aliado ao fato de que parte do pagamento do salário dos serviçais só era feito quando esses eram repatriados – o que raramente acontecia –, gerou forte reação por parte de alguns países, notadamente Alemanha, Inglaterra e Estados Unidos, que viam nessa prática uma forma (mal) disfarçada de escravidão:

[...] dois principais aspectos da vida nas roças chamaram a atenção. Primeiro foi a extrema insalubridade das ilhas e as taxas de mortalidade alarmantes: Príncipe, que só foi seriamente aberto ao cultivo do cacau desde 1893, estava no auge de sua desastrosa epidemia de doença do sono, e as taxas de mortalidade estavam

chegando a 200 por mil - cinco vezes a média normal. O segundo objeto de crítica foi o fato de que nenhum trabalhador jamais foi repatriado. Um grupo de missionários escreveu uma carta para o *The Times*, de Londres, em 1909, afirmando categoricamente que nunca ninguém conheceu um serviçal que tivesse sido devolvido. Nessas circunstâncias, ser contratado para trabalhos nas ilhas foi amplamente considerado uma sentença de morte e foi justamente temida pelo continente. (BOUCHARD, 2017, p. 42)<sup>1</sup> (tradução nossa).

Nascimento (2013) ressalta, no entanto, que essa reação internacional só passou a existir a partir de 1904, justamente quando se iniciou a concorrência entre cartéis de cacau no mundo. Assim, a acusação de que Portugal produzia *slave cocoa* (cacau escravo) em sua colônia insular não tinha motivação humanitária, mas sim visava o estremecimento de São Tomé enquanto concorrente na produção de cacau, uma vez que as lavouras de café já estavam em declínio econômico nas ilhas. De todo modo, tais acusações forçaram o governo português a melhorar as condições de vida dos serviçais, sobretudo em relação à sua repatriação, que se tornou regular a partir de 1912 (SANTOS, 2014).

O modo de vida dos serviçais contratados foi profundamente relevante para as variedades linguísticas que se formaram e difundiram a partir da chegada desse contingente na ilha. Os trabalhadores viviam segregados tanto cultural quanto linguisticamente da população forra, uma vez que estes os consideravam inferiores devido à sua sujeição ao trabalho análogo ao escravo. Deve-se a este fato o pouco contato que os contratados tiveram com as línguas crioulas. Em contrapartida, a sua exposição ao português tornou-se maior gradativamente, já que o contingente europeu, responsável pelas plantações, tornou-se demograficamente superior em relação ao período anterior à segunda colonização.

Assim, para a comunicação no interior das roças, os trabalhadores contratados passam a utilizar o português como L2, com todas as modificações

---

<sup>1</sup> [...] two main aspects of life on the roças attracted attention. First was the extreme unhealthiness of the islands and the alarmingly high mortality rates; Príncipe, which had only seriously been opened up for cocoa growing since 1893, was in the throes of its disastrous sleeping sickness epidemic, and death rates were reaching 200 per thousand – five times the normal average. The second object of criticism was the fact that no laborers were ever repatriated. A group of missionaries wrote a letter to *The Times* of London in 1909, stating categorically that no one had ever met a serviçal who had returned. In these circumstances, to be contracted for work in the islands was widely held to be a death sentence and was justly feared by mainland Africans and Cape Verdeans alike.

estruturais que caracterizam as línguas adquiridas em contexto de contato, sendo a língua-alvo o português falado pelos portugueses e pelos colegas contratados, que eram das mais variadas origens étnicas (GONÇALVES, 2016; LOPES, BAXTER, 2011). Os descendentes dessa população passam a falar essa versão nativizada da língua portuguesa como língua materna, conhecida como português dos Tongas (cf. 2.3.6). Ainda sobre a fronteira entre o trabalho contratado dos serviçais e a escravidão, Mateus (2018) diz que:

Apesar dos relatos duvidosos, mesmo no século XX, a existência de escravatura nas ilhas de São Tomé e Príncipe afigura-se bastante credível. No entanto, pode afirmar-se que as condições de vida dos serviçais nas roças progrediram muito em relação ao período anterior à abolição da escravatura e foram melhorando mais ao longo do tempo. É claro que o bom tratamento dos serviçais por parte dos roceiros podia simplesmente advir do excesso de procura e do elevado preço de compra, fazendo do serviçal um bem precioso e indispensável. Os baixos salários e a impossibilidade de muitos não conseguirem regressar à terra de origem constituem factores que contribuem para fortalecer a tese da existência de trabalho escravo na colónia portuguesa (MATEUS, 2018, p. 209).

Conforme comentamos acima, os trabalhadores contratados eram tratados a partir de um *status* discriminatório na sociedade são-tomense, sendo que estes viviam separados até mesmo espacialmente da população nativa. De acordo com Seibert (2015), a hierarquia populacional de São Tomé, durante o ciclo do cacau e café, era liderada pelos colonos brancos, seguidos pela população crioula e, por fim, pelos africanos contratados. Tem-se, então, a população contratada como alvo de dupla marginalização: por parte dos colonizadores, que administravam as roças, e por parte dos nativo-forros, que viam nos serviçais contratados a representação do fracasso ocasionado pelo trabalho agrícola forçado, que remetia diretamente à tão repelida escravidão. Sabe-se, inclusive, que esse distanciamento entre a população nativa e os chamados “africanos do continente” era fomentada pelos próprios colonos, que temiam o surgimento de algum tipo de resistência em caso de associação entre esses dois grupos, conforme afirma Feio (2016):

A criação de distanciação e mesmo de rivalidade entre pessoas e “grupos” foi uma das armas usadas no e pelo próprio regime colonial. O convívio interétnico entre nativos e contratados era punido, sendo que eram os capatazes (moçambicanos, angolanos, cabo-verdianos) que estavam incumbidos de vigiar essas aproximações. As

animosidades eram tão fortes que depois da independência se registaram algumas perseguições feitas por forros e angolares a alguns ex-contratados (sobretudo aos moçambicanos e angolanos) (FEIO, 2016, p. 3).

Três momentos da história representaram a lenta evolução do *status* dos trabalhadores das roças na sociedade são tomense: a) o início da Primeira República Portuguesa (1910-1926), uma vez que fazia parte dos interesses republicanos a denúncia da incompetência monárquica na administração e distribuição de trabalho nas roças de São Tomé, bem como das irregularidades presentes em todas as colônias administradas pela monarquia portuguesa; b) a campanha realizada principalmente pelo governador-geral de São Tomé, Carlos Gorgulho (1945-1948), com vistas a acirrar a rivalidade entre a população nativa e os trabalhadores contratados, associando a aqueles a culpa pelas condições de trabalho precárias dos serviçais e os incitando a associar-se aos portugueses no massacre contra os forros, que culminou no Massacre de Batepá, em 1953; c) os movimentos africanos pela independência das colônias portuguesas, iniciados na década de 1960 (EYZAGUIRRE, 1993). Entretanto, a situação dos trabalhadores contratados só mudou de forma permanente com a independência do país, a cessão de direitos iguais e cidadania a todos os habitantes, em 12 de julho 1975.

Atualmente, São Tomé e Príncipe é dividido em sete distritos: Água Grande, Lobata, Lembá, Cauê, Cantagalo, Mé-Zochi e Pagué. No que diz respeito ao uso de línguas, o último censo (INE 2017) mostrou que o português é o idioma mais falado no país, sendo usado por 98,4% da população, enquanto, dentre as línguas nacionais, a mais falada é o forro, sendo utilizada por 36,2% dos habitantes de São Tomé. Não há, entretanto, nesse censo, informações sobre qual idioma o indivíduo considera principal, nos casos daqueles que afirmaram falar mais de um idioma. A tabela 2, adaptada de Bouchard (2017), detalha a distribuição do uso das línguas de São Tomé em função da idade dos habitantes.

Tabela 2: Línguas faladas em São Tomé e Príncipe em função da idade da população

| Línguas faladas (%) |           |       |         |         |              |
|---------------------|-----------|-------|---------|---------|--------------|
| Idade (anos)        | Português | Forro | Angolar | Lung'le | Caboverdiano |
| 5-9                 | 99,8      | 9,0   | 1,9     | 0,1     | 2,7          |
| 10-19               | 99,8      | 25,8  | 4,4     | 0,3     | 5,6          |
| 20-29               | 99,7      | 47,3  | 8,0     | 1,2     | 11,4         |
| 30-39               | 99,6      | 56,2  | 10,3    | 1,7     | 13,8         |
| 40-49               | 99,3      | 63,3  | 11,7    | 2,5     | 16,0         |
| 50 ou mais          | 97,8      | 73,3  | 13,8    | 2,7     | 14,2         |

Fonte: BOUCHARD, 2017, p. 78.

Conforme podemos ver através dos dados apresentados na tabela, há um significativo decréscimo da frequência de uso das línguas crioulas pelas gerações mais novas, ao passo que, nessas faixas etárias, o uso da língua portuguesa é mais frequente se comparado à frequência encontrada nas gerações mais velhas. Esse dado é um importante indício do *status* da língua portuguesa e das línguas crioulas na sociedade são-tomense, sendo que, enquanto o uso do português é estimulado, sobretudo em contextos mais formais e nos veículos de informação, as línguas crioulas tendem a ser utilizadas em contextos informais e com cada vez menos frequência.

O contexto social de São Tomé, nos dias atuais, ainda é significativamente complexo. Mais da metade (66,2%) da população são-tomense se encontra abaixo da linha de pobreza, embora o país seja rico em petróleo, uma das principais fontes de renda do país. A forte presença do turismo também é um dos fatores que mais movimentam a economia local, bem como a agricultura (BOUCHARD, 2017).

## 2.2 O ESTABELECIMENTO E A ESTRUTURA DAS ROÇAS

Durante a expansão do principal ciclo econômico de São Tomé e Príncipe – o do café e cacau, nos séculos XIX e XX, a economia das ilhas foi pautada na implantação de grandes estruturas agrícolas denominadas roças, palavra que significa ‘desbravar mato’. Esse modelo agrícola, baseado no sistema de *plantation*

e na expansão territorial, principalmente em relação à penetração na vegetação densa do arquipélago, encontra grande influência em estruturas semelhantes existentes em outros ambientes tropicais, principalmente os prósperos engenhos de açúcar brasileiros (BERTHET, 2012). Sobre as roças, Pape afirma que:

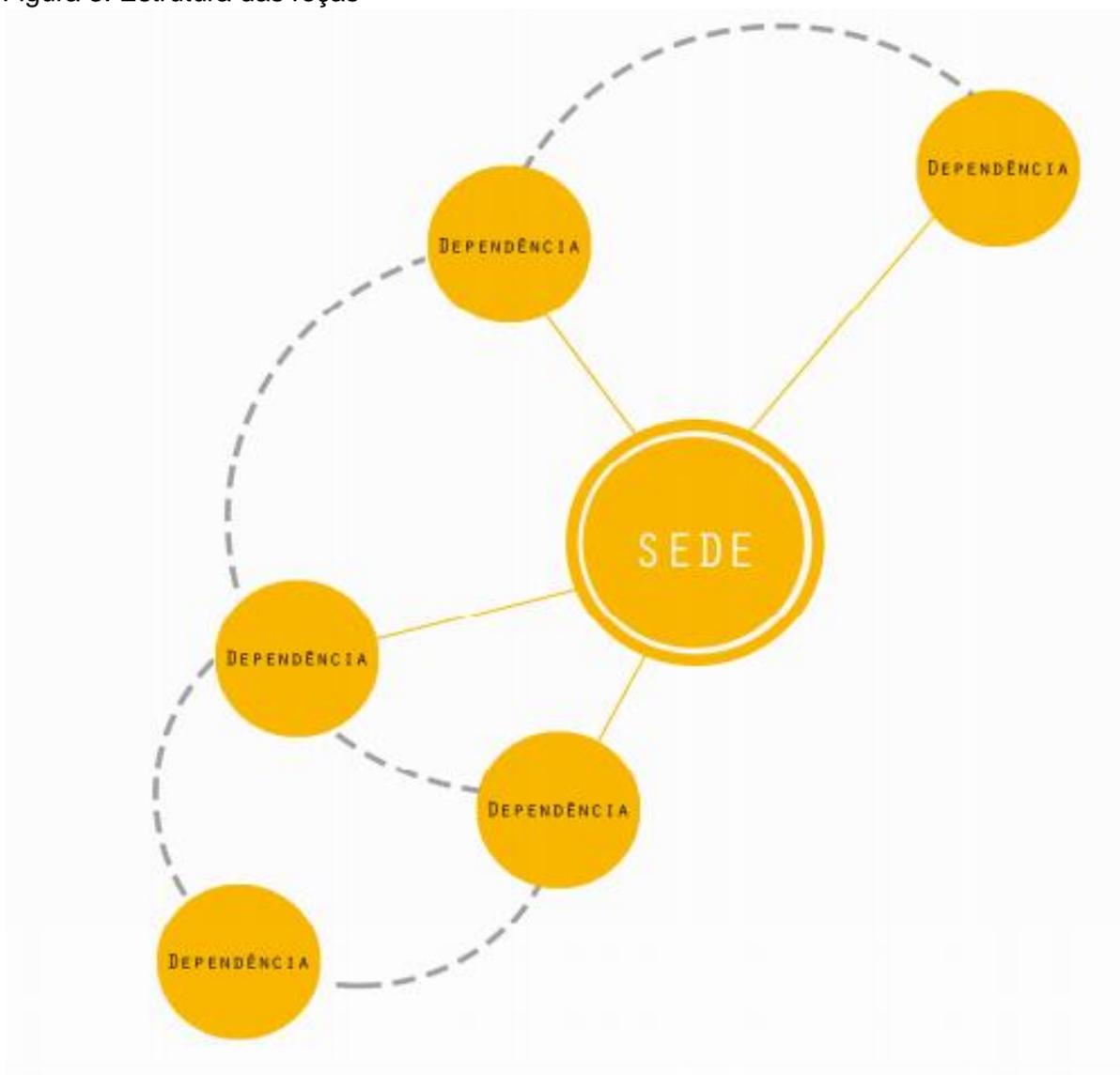
É em São Tomé e Príncipe que o termo “roça” ganha força, reclamando para si um carácter quase exclusivo, não só como afirmação identitária e específica do seu sistema agrário e social, mas também pela forma como a implantação desse sistema se amarrou à sua cultura e à sua própria história (2013, p. 23).

As roças são-tomenses eram as organizações estruturadoras do território das ilhas, uma vez que através delas se fixavam todas as estruturas necessárias para a existência de uma vida urbana, ainda que modesta, em São Tomé. Assim, configuradas como micro centros urbanos, as roças eram autossuficientes, sendo seu contato com o exterior necessário apenas para o escoamento da produção, que era dotada de alta complexidade e disciplina. O alto nível de isolamento dessas organizações agrícolas, característico de sua lógica produtiva, foi responsável pelo surgimento de muitas especificidades culturais e identitárias no interior das roças, que se refletem até hoje na população.

Quanto às características habitacionais e agroindustriais, essas estruturas contavam com terras de cultivo e por estruturas de suporte à produção, tais como elementos habitacionais, como casas dos administradores e dos trabalhadores, elementos industriais, como secadores e armazéns, elementos assistenciais, como saúde, educação e apoio social, e estruturas de manutenção, transformação e apoio alimentar. Essas estruturas, com o passar do tempo, evoluíram de acordo com a necessidade de aumento da produtividade (PAPE, 2013).

Assim, diferentemente dos extintos engenhos de açúcar, as roças passaram a ser complexas estruturas produtivas altamente industrializadas, formadas por uma sede e suas dependências, que se configuravam como redes interligadas entre si, contando com hierarquia e organizada gestão interna (figura 3).

Figura 3: Estrutura das roças



Fonte: VICENTE, 2016, p. 8.

A independência nacional de São Tomé, declarada em 12 de julho de 1975, teve como principal característica política a luta pela melhoria de vida dos trabalhadores das roças. Em 30 de setembro do mesmo ano, inicia-se o processo de nacionalização das roças, ou seja, o domínio das mesmas passa a pertencer ao estado são tomense, marca da chamada reforma agrária que ocorreu no país e da libertação simbólica do jugo colonial. O trecho a seguir, proferido durante o comício ocorrido nesta data pelo, a partir de então, presidente da República, Pinto da Costa, deixa claras as aspirações do povo são tomense em relação à posse das roças:

Fomos ou não fomos explorados durante centenas de anos pelo colonialismo fascista português? Somos ou não somos ainda explorados pelas companhias agrícolas estrangeiras existentes no nosso país? Temos ou não temos o direito de nacionalizar as terras, de tomar as terras que foram roubadas ao nosso povo? O que é que a Assembleia Geral do povo de S. Tomé e Príncipe decide... [sobre as roças]. Tomamos ou não tomamos essas roças? (Grito unânime do povo: Tomamos! Tomamos!) (BERTHET, 2016, p. 969).

As terras passam a pertencer, a partir daí, aos forros, que, em sua maioria, não apresentavam a experiência necessária para a boa administração das roças. Assim, a produção das roças tem sua lógica alterada, passando a monocultura de cacau a competir com a cultura de subsistência. Tal inabilidade no controle da produção e funcionamento das roças ocasionou o seu declínio, o que levou à pobreza das comunidades que dependiam diretamente dessa produção agrícola. A partir daí, grande parte da população que habitava as roças passa a dirigir-se para os centros urbanos (SEIBERT, 2015; BERTHET, 2016).

No pós-independência, em relação à demografia, observou-se uma diminuição drástica da população branca na ilha, devido ao processo de descolonização, bem como a repatriação de boa parte dos serviçais angolanos e moçambicanos, permanecendo parte significativa dos trabalhadores caboverdianos. (SEIBERT, 2014, 2015). A partir daí, viver nas roças passou a ter um significado diferente:

Nos anos noventa, morar numa roça continuou a ser associado a uma situação depreciativa, porquanto a roça era um lugar onde moravam os descendentes pobres de migrantes cabo-verdianos, sem energia elétrica e, por vezes sem água potável e sem serviços públicos básicos. Em termos de infraestruturas – moradia, água, estradas, e transporte –, as condições passaram a se degradar. Em muitas roças, por exemplo, o sistema de água passou a não fornecer mais a água potável necessária, obrigando muitos dos habitantes a percorrer vários quilômetros a pé para se abastecer na fonte mais próxima. No imaginário nacional, “pertencer à roça” era (e é ainda) associado a um conjunto de preconceitos, certezas sobre os hábitos e costumes dos moradores considerados isolados dos centros urbanos das duas ilhas ou em espaços de difícil (BERTHET, 2016, p. 980).

As políticas públicas implementadas em São Tomé, sobretudo após a reforma agrária, não contribuíram muito para a mudança da vida e para mitigar o preconceito

do qual a população que vive nas roças é vítima, situação que, entretanto, vem sendo, lentamente, modificada pelas gerações mais novas.

### **2.2.1 A roça de Monte Café**

A roça de Monte Café, na qual vivem os informantes que forneceram os dados para a realização deste trabalho, foi uma das primeiras fundadas no arquipélago de São Tomé e Príncipe, sendo que seus primeiros registros de existência datam de 1858. A roça, fundada por Manuel da Costa Pedreira, fica localizada na região de Mé-Zóchi, a 13 km da cidade de São Tomé e é considerada uma das mais importantes produtoras de cacau e café do arquipélago (MENDES, 2011).

Figura 4: Rua principal da roça de Monte Café



Fonte: LANDIM, 2014, p. 86.

Em relação à sua composição étnica, seus proprietários captaram, na ocasião da fundação da roça de Monte Café, trabalhadores da região sul de Angola. Esse contingente populacional, falante de línguas umbundo, foi maioria até meados do século XX, quando um grande número de trabalhadores moçambicanos e

caboverdianos foi trazido para trabalhar na roça (BAXTER, 2002). Desse modo, as línguas do tronco kimbundo e o crioulo caboverdiano passaram a coexistir com as versões de português faladas na roça (cf. 2.4).

### 2.3 VARIEDADES LINGUÍSTICAS DE SÃO TOMÉ

Como destacado nas seções anteriores, os primeiros contingentes populacionais trazidos para São Tomé para o trabalho agrícola nos engenhos de açúcar eram falantes de línguas de troncos distintos – Edo e línguas Bantu. Devido à necessidade de comunicação e mútua inteligibilidade entre os portugueses e a população escravizada e à situação de contato entre línguas, característica das circunstâncias de colonização, surge um código emergencial de comunicação, ou pidgin, cujo léxico advém do português, mas com a estrutura morfossintática qualitativamente distinta da do português.

Esse código emergencial, estruturalmente expandido e nativizado, originou o chamado protocrioulo do Golfo da Guiné, que, por sua vez, com o passar dos anos, devido principalmente à mobilidade da população por todo o Golfo da Guiné, motivada pela chegada dos colonizadores portugueses às outras ilhas africanas, deu origem às línguas crioulas faladas no Golfo da Guiné: o forro, o angular, o lung'le e o fa d'ambô<sup>2</sup>:

O grande número de características compartilhadas entre os quatro CGG (crioulos do Golfo da Guiné) deixa poucas dúvidas de que eles compartilham um ancestral comum, que naturalmente se desenvolveu no Forro, enquanto os outros crioulos são descendentes do protocrioulo em algum estágio particular de seu desenvolvimento. (HAGEMEIJER, 2000, p. 8, apud BOUCHARD, 2017, p. 32)<sup>3</sup> (tradução nossa).

Esses crioulos, embora apresentem uma origem comum, guardam peculiaridades decorrentes da diferenciação ocasionada pelo tempo e pelo espaço, sendo que essas especificidades foram suficientes para torná-los ininteligíveis entre si.

---

<sup>2</sup> O fa d'ambô é falado na guiné equatorial, mais especificamente nas ilhas de Ano Bom e Bioko.

<sup>3</sup> The large number of shared features between the four GGC leaves little doubt that they share a common ancestor, which naturally developed into [Forro], while the other creoles are offspring of the Proto-creole at some particular stage of its development.

Atualmente, em São Tomé e Príncipe, são falados três dos quatro crioulos mencionados acima: o forro e o angolar, em São Tomé, e o lung'le, no Príncipe, além do português, que é a língua oficial do país, do português dos Tongas e do kabuverdianu. Falaremos, na próxima seção, acerca do surgimento e do *status* atual de cada uma dessas línguas, com o objetivo de ilustrar o quadro sociolinguístico atual de São Tomé.

### **2.3.1 O forro**

Tendo sido registrado pela primeira vez no ano de 1627 pelo padre Alonso de Sandoval (ARAUJO, HAGEMEIJER, 2013), o forro, ou santome, é a variedade linguística mais falada em São Tomé depois do português, e está diretamente ligado ao primeiro grupo étnico genuinamente são-tomense. Essa relação é o que explica, inclusive, a perpetuação e abrangência dessa variedade linguística no país. A gênese da estrutura do santome está no contato entre as línguas do tronco edoide, faladas pelos primeiros habitantes africanos de São Tomé, e o português dos europeus (HAGEMEIJER, 2011).

Desse modo, o santome, que nasceu durante a fase de habitação da ilha de São Tomé, atingiu bastante prestígio na sociedade colonial por ser a língua dos primeiros “filhos da terra”, nomeadamente os filhos dos arranjos entre os primeiros habitantes portugueses e as mulheres africanas que para eles foram cedidas. Essa população mestiça, que nasceu livre e rapidamente ascendeu no bojo da sociedade são tomense, preservou a sua variedade linguística como símbolo de resistência e da sua diferenciação da população escravizada:

Nada parece mais natural do que associar esta comunidade de forros (escravos que recebiam a carta de alforria), com uma identidade própria, ao surgimento de uma língua crioula que mais tarde se difundiu para as roças, tornando-se a língua-alvo dos escravos recém-chegados (HAGEMEIJER, 1999, p. 78).

Atualmente, embora ainda seja grande o seu prestígio dentre as línguas autóctones, o santome é língua materna de cada vez menos são-tomenses, devido ao avanço da língua portuguesa pelo país e ao crescimento da urbanização de São Tomé. No interior do país, o uso do santomense está relacionado à preservação das tradições orais dos forros (BANDEIRA, 2017).

### 2.3.2 O lung'le

O lung'le ou principense, língua crioula que, assim como o angolár e o forro, detém status de língua nacional de São Tomé e Príncipe, é uma variedade linguística falada na ilha do Príncipe, cujas missões de povoamento se iniciaram no ano de 1500, sete anos de pois do início da colonização de São Tomé. Na ocasião do povoamento de Príncipe, o protocrioulo do Golfo da Guiné já havia se formado e foi levado de São Tomé para a ilha do Príncipe pela população escravizada, sendo que a diferença temporal na colonização das duas ilhas e o isolamento geográfico da ilha do Príncipe foram os principais fatores responsáveis pela diferenciação dessa variedade linguística em relação aos outros crioulos autóctones (BANDEIRA, 2017).

Além dos escravizados levados de São Tomé, a ilha do Príncipe recebeu, também, um contingente de africanos continentais trazidos de áreas de predomínio de línguas do grupo edoide, fator que explica algumas características estruturais do lung'le:

Por ter sido isolado desde muito cedo, conjectura-se que, do ponto de vista do léxico e da fonologia, o lung'le seria a língua do grupo que mais representa o léxico de origem edo e o que mais reteve aspectos fonológicos edoídes, exclusivos da área onde essas línguas eram faladas (HAGEMEIJER, 2009, p. 15-16. apud BANDEIRA, 2017, p. 129).

O uso do lung'le, nos dias atuais, ainda é restrito à ilha do Príncipe, sendo a língua crioula menos falada no país (INE 2017). O declínio no uso dessa variedade linguística se iniciou ainda no século XIX, quando a ilha do Príncipe foi acometida por uma epidemia de tripanossomíase, que provocou uma grande queda populacional da ilha, uma vez que apenas cerca de 300 pessoas sobreviveram. Essa queda demográfica proporcionou, devido à necessidade de repor a mão de obra perdida, a entrada de trabalhadores contratados e de suas variedades linguísticas no Príncipe. A entrada desse contingente populacional, que veio principalmente de Cabo Verde, levou também à queda da frequência de uso do lung'le. O uso reduzido dessa variedade linguística pode ser explicado, também, pela perspectiva da avaliação dos falantes:

Alguns jovens têm um conhecimento passivo do lung'le, mas não têm competência linguística para falar. Segundo Espírito Santo (em preparação), 57% dos alunos do ensino secundário afirmam ter um familiar falante de lung'le, mas apenas 18% alegam compreender a língua. Observou-se também que a língua não é transmitida intergeracionalmente. Um dos motivos apontados pelos próprios habitantes da Ilha do Príncipe é a relutância que os pais apresentavam em transmitir a língua, pois pensavam que seu aprendizado atrapalharia o do português. Em relação a isso, Pinho (2008) argumenta que falantes de línguas minoritárias podem ter atitudes positivas para com a sua língua, mas podem não querer transmiti-la por razões como a suposta interferência no aprendizado da língua majoritária, preconceitos com o 'sotaque' e sentimentos de inferioridade. Outrossim, os habitantes lembram que os pais e avós falavam em lung'le, porém crianças poderiam sofrer abusos físicos e psicológicos se não respondessem em português. Segundo Agostinho (2014), mesmo entre os falantes competentes em lung'le, a língua utilizada em seu cotidiano atualmente é predominantemente o português (AGOSTINHO et al, 2016, p. 594).

Apesar de alguns autores apontarem o lung'le como uma língua que corre risco de extinção, há, por parte do governo regional do Príncipe, o incentivo à difusão da cultura e, conseqüentemente, da língua principense, o que é posto em prática através da obrigatoriedade do ensino de lung'le nas escolas, da transmissão de um programa de rádio voltado para a discussão do lung'le na sociedade principense e de um encontro semanal que ocorre no centro cultural do Príncipe voltado para debates, que devem ser feitos em lung'le (AGOSTINHO, 2014).

### **2.3.3 O angolar**

As hipóteses que explicam a origem do grupo étnico angolar não são consensuais, sendo que Seibert afirma que existem três possibilidades explicativas para a chegada desse grupo à ilha. Segundo o autor, a primeira e mais antiga hipótese afirma que os angolares seriam descendentes de escravizados que vieram de Angola no século XVI. O navio no qual esses angolanos estavam teria naufragado na costa da ilha. A segunda hipótese supõe que os angolares seriam, na verdade, um grupo que já habitava a ilha antes da chegada dos portugueses, e a terceira hipótese afirma que os angolares seriam descendentes de uma população quilombola que teria fugido da costa para partes mais afastadas da ilha nos séculos XVI e XVII, formando, assim, uma comunidade isolada (SEIBERT, 2004, p.44).

Dentre as três hipóteses, a mais comumente aceita é a terceira, uma vez que existem fatores, sobretudo linguísticos, que corroboram essa explicação. A língua dos angolares, o angolar, apresenta muitas características que a aproximam de outros crioulos nascidos a partir do protocrioulo do Golfo da Guiné, sendo que o afastamento geográfico dos angolares não possibilita que essa aproximação linguística tenha acontecido totalmente através do contato. Assim, é possível supor que esse contingente tenha tido algum contato com o protocrioulo do Golfo da Guiné, tendo fugido e se isolado depois da sua formação.

Outra evidência de que os angolares são descendentes de quilombolas são os registros, feitos, inclusive, pelo próprio donatário da colônia, Álvares de Caminha, de recorrentes fugas de escravos para o interior da ilha, uma vez que a difícil topografia de São Tomé favorecia o sucesso dessas fugas (BANDEIRA, 2017). O estabelecimento desses quilombos, entretanto, não garantiu o isolamento demográfico e linguístico total dos angolares. Eram comuns saques e ataques dos angolares à cidade, além de sequestro de mulheres, sendo também recorrentes as tentativas de resgate desses negros fugidos - as chamadas "guerras do mato". Essas ações ocorreram até 1693, quando o capitão do mato Mateus Pires entra em acordo com o chefe dos quilombos:

Mateus Pires assina junto de um chefe do quilombo um “acordo de tréguas”, de modo a ver os seus interesses representados, bem como os dos restantes colonos, pretendo obter um aliado “nativo” que pudesse servir de intermediário social, político e económico entre os fugidos e os europeus. A aposta na colonização de modo indirecto foi uma prática muito usada em muitas outras ex-colónias. O chefe “angolar” recebe assim o título de capitão e inicia uma série de contactos regulares com a autoridade colonial: desempenhava funções “governativas e jurídicas”, dispondo de uma tropa composta por “tenentes e alferes” (FEIO, 2008, p. 25).

Desse modo, é possível afirmar que não foi total o isolamento étnico e linguístico dos angolares, já que essas ações possibilitaram o contato e, muito provavelmente, algum nível de troca linguística com os são-tomenses e até mesmo com variedades da língua portuguesa (NASCIMENTO; ROCHA, 2018, p. 57). Hoje, apenas cerca de 6% da população - 11.377 pessoas (INE 2017) é falante do crioulo angolar, sendo que esse contingente populacional se encontra em alguns pontos do litoral são-tomense, principalmente nos distritos de Lembá e Caué.

### **2.3.4 O kabuverdianu**

Estima-se que os primeiros trabalhadores contratados de Cabo Verde chegaram a São Tomé e Príncipe por volta de 1870, com o objetivo de trabalhar nas lavouras de cacau e café das ilhas (GONÇALVES, HAGEMEIJER, 2015). Grande número dos caboverdianos contratados foram enviados para o Príncipe e, apesar da resistência ao trabalho na lavoura, que, conforme comentamos anteriormente, tinha os mesmos moldes da escravidão, poucos caboverdianos foram repatriados, o que justifica a forte presença do crioulo kabuverdianu na ilha do Príncipe até hoje. Atualmente, essa língua, nascida na Alta Guiné, é um símbolo de orgulho e afirmação identitária dos descendentes dos contratados de Cabo Verde:

Com base em trabalho de campo realizado em 2018, foi possível notar que há inúmeras crianças que têm o kabuverdianu como primeira língua (aprendendo o português somente após a entrada na escola) e que a língua é amplamente usada para comunicação diária, sendo perceptível a ideia de que o uso do kabuverdianu serve como forma de manter a identidade caboverdiana. Muitos informantes afirmam que, mesmo tendo nascido no Príncipe, se consideram um pouco caboverdianos, já que essa é a nacionalidade de seus pais e eles falam a língua de Cabo Verde (ensinada pelos pais para manter a identidade), havendo, sempre que as condições financeiras permitem, relatos de visitas ao arquipélago para conhecer parentes que lá ficaram (AGOSTINHO, BANDEIRA e FREITAS, 2020, p. 71).

Embora seja falado somente na ilha do Príncipe, o kabuverdianu é a terceira língua mais falada no país, estando depois, respectivamente, do português e do forro (INE 2017). Desse modo, há, no Príncipe, mais falantes de kabuverdianu que de lung'le, sendo aquela a língua materna de muitos habitantes da ilha do Príncipe – vários deles, inclusive, monolíngues (AGOSTINHO, 2016).

### **2.3.5 O português de São Tomé**

A chegada da língua portuguesa à ilha de São Tomé acontece concomitantemente com o seu descobrimento e colonização, processo iniciado nos últimos anos do século XV. O fato é que, nas situações de sujeição que caracterizam a relação entre o colonizador e o colonizado, a língua da colônia representa um símbolo da suposta superioridade cultural do império, sendo as línguas autóctones vistas como inferiores. Entretanto, a diversidade linguística das

ilhas, atribuída aos diferentes contingentes populacionais que para lá se dirigiram na fase da habitação, e a falta de uma política de difusão da língua portuguesa, sobretudo devido ao não acesso da população autóctone ao ensino formal, impediram que, em primeiro momento, a língua portuguesa alcançasse estatuto de língua majoritária: sua referência, cada vez mais pontual, criou as circunstâncias adequadas para o surgimento do crioulo do Golfo da Guiné, conforme discutimos nas seções anteriores.

O *status* do português em São Tomé só se modificou com a chegada dos trabalhadores contratados para trabalhar nas lavouras de cacau e café, que passaram a adotar a língua portuguesa falada pelos responsáveis pelas plantações como L2. Posteriormente, com a independência, em 1975, a língua passa a ser oficial no país, sendo a norma portuguesa a ensinada nas escolas e utilizada nos veículos escritos de informação. É preciso ressaltar, no entanto, que a variedade de português de São Tomé não deixou de carregar as marcas estruturais das línguas crioulas faladas no país (HAGEMEIJER, 2009; GONÇALVES, HAGEMEIJER, 2015).

A língua portuguesa, em São Tomé, é vista como um meio de unificação do país, frente à grande diversidade linguística da população. Assim, a universalização da educação, que se tornou uma realidade após a independência, e o amplo acesso aos veículos de comunicação em massa são os principais fatores responsáveis pela difusão do português em São Tomé. A partir daí, o idioma passa a ser utilizado como L2 pelas gerações mais antigas, que passam a empregar o idioma sobretudo nos contextos mais formais, e como L1 pelas novas gerações, ao passo que os crioulos autóctones vão se tornando cada vez menos utilizados:

O português é a língua mais falada, uma tendência que atualmente parece estar a intensificar-se em detrimento das línguas crioulas autóctones. Se, no passado, o mundo português e o mundo crioulo se encontravam num contexto de diglossia clássica, favorável à manutenção dos crioulos, a independência alterou o rumo dos acontecimentos. A maior mobilidade social, à qual não é alheio o fenómeno da emigração, o acesso generalizado ao ensino e aos meios de comunicação na língua oficial, a ausência de políticas orientadas para as línguas crioulas, assim remetidas à informalidade e à oralidade, são fatores que têm desfavorecido cada vez mais as línguas minoritárias das ilhas. O censo de 2001 referente à língua falada por grupo etário mostra, por exemplo, que entre os jovens com

menos de 20 anos há uma quebra acentuada do número de falantes que alega falar o santome (HAGEMEIGER, 2009, p.19).

A compreensão das circunstâncias que envolveram a transmissão do português como L1 tendo como *input* uma variedade da língua utilizada como L2, sobretudo com o início do regime contratual, é fundamental para a explicação da emergência de outra variedade da língua portuguesa: o português dos Tongas, sobre a qual nos debruçamos neste trabalho.

### **2.3.6 O português dos Tongas**

Conforme comentamos anteriormente, a inserção do regime de contratos na ilha de São Tomé modificou significativamente a realidade sociolinguística da ilha. Os trabalhadores contratados, advindos de diferentes áreas da África Continental, levaram para São Tomé as línguas kimbundo, kikongo, umbundo e o kabuverdianu, sendo que essas línguas serviram de substrato para as mudanças ocorridas na língua portuguesa a partir daí. Dentre esses contingentes populacionais, tem-se aos angolanos como os mais representativos para as mudanças sociais e linguísticas que ocorreram a partir daí:

A presença dos angolanos nesse período foi marcante, porque, antes de 1909, de modo geral, os angolanos não foram repatriados, e quando o foram, os filhos nascidos em São Tomé foram obrigados a ficar na ilha, tornando-se propriedade das roças (BAXTER, 2019, p. 298).

Embora o forro estivesse presente no dia a dia das roças, Baxter elenca alguns motivos pelos quais essa língua não era utilizada em larga escala pelos trabalhadores contratados e nem pelos seus descendentes, ainda que seja possível que esse crioulo tenha exercido alguma influência sobre o português dos Tongas: a) a população local não costumava trabalhar nas roças de cacau e café, devido à proximidade entre o regime de contratos e a escravidão, o que produzia uma resistência intrínseca por parte da população forra; b) os administradores das roças não dominavam o santomense; c) os trabalhadores contratados ficavam confinados nas roças, o que dificultava o contato com o mundo externo e, conseqüentemente, com o forro (BAXTER, 2019).

Assim, os trabalhadores contratados foram levados a São Tomé e, nas grandes roças, passam a adotar, além de uma koiné baseada no umbundo, uma vez que essa língua era falada pelo contingente populacional mais numérico – os angolanos –, uma versão modificada do português, cujo *input* era o português L1 falado pelos administradores das roças e o português L2 falado pelos colegas (BAXTER, 2002). Nesse contexto, como é esperado em situações de contato massivo entre línguas tipologicamente diferentes, a língua que resultou apresenta um conjunto de mudanças estruturais que a afastam do português falado pelos administradores, e até mesmo da variante de língua portuguesa tida como oficial no país. (BAXTER, 2002; HAGEMEIJER, 2009). Os descendentes dos trabalhadores, denominados *Tongas*, foram os responsáveis por nativizar e propagar essa nova versão do português.

Há bastante divergência em relação ao estatuto do português dos Tongas na sociedade são tomense. Rougé (1992) afirma que a língua pode ser classificada como um crioulo, enquanto Baxter (2002), utiliza a noção de semicrioulização para defini-la. Porém, embora tenha sofrido alterações típicas de línguas formadas através do contato, o português dos Tongas não apresenta as reestruturações profundas que caracterizam as línguas crioulas. Esse fator, somado às especificidades sociodemográficas do funcionamento das roças, nos faz assumir a hipótese de que o português dos Tongas consiste não em uma língua crioula, mas sim em uma versão nativizada do português, que, inclusive, devido à pequena quantidade de falantes, estaria em vias de desaparecimento (BAXTER, 2002).

É unânime, em muitos estudos que abordam essa variedade linguística, a denominação de “português Tonga”, e não “língua Tonga” ou “crioulo Tonga” (LOPES; BAXTER, 2011; GONÇALVES; HAGEMEIJER, 2015; BAXTER; MELO, 2014, entre outros). Considerando que as línguas crioulas são definidas primeiro pela sua história social e por um conjunto de traços morfossintáticos, apontaremos, abaixo, alguns critérios que nos levam a assumir, aqui, essa perspectiva:

- a) Sistema TMA: Característica que consiste na presença de uma partícula antes do verbo para representar tempo-modo-aspecto, o sistema TMA está presente em todas as línguas crioulas conhecidas. A título de

exemplificação, trazemos as tabelas abaixo, adaptadas de Bickerton (1980, p. 14) e Couto (1996, p. 106), respectivamente:

Tabela 3: Tempo, modo e aspecto no crioulo havaiano

| <b>Tempo</b> | <b>Modo</b> | <b>Aspecto</b> | <b>Exemplo</b>  |
|--------------|-------------|----------------|-----------------|
| -            | -           | -              | Wok             |
| -            | -           | +              | Stei wok        |
| -            | +           | -              | Go wok          |
| -            | +           | +              | Go stei wok     |
| +            | -           | -              | Bin wok         |
| +            | -           | +              | Bin stei wok    |
| +            | +           | -              | Bin go wok      |
| +            | +           | +              | Bin go stei wok |

Fonte: BICKERTON, 1980, p. 14.

Tabela 4: Tempo, modo e aspecto no crioulo guineense

| <b>Tempo</b> | <b>Modo</b> | <b>Aspecto</b> | <b>Exemplo</b> |
|--------------|-------------|----------------|----------------|
| -            | -           | -              | I fuma         |
| +            | -           | -              | I fuma ba      |
| -            | +           | -              | I ba fuma      |
| -            | -           | +              | I ta fuma      |
| +            | +           | -              | I ba fuma ba   |
| -            | +           | +              | I ba ta fuma   |
| +            | -           | +              | I ta fuma ba   |
| +            | +           | +              | I bata fuma ba |

Fonte: COUTO, 1996, p. 106.

No português dos Tongas, essa característica gramatical não está presente. A língua dos Tongas apresenta várias evidências de processos de reestruturação do português, como é o caso do sistema de determinantes, do sistema pronominal, da

representação dos objetos dativos, da variação nas concordâncias - mas o sistema verbal TMA não é reestruturado para constituir um sistema com morfemas e diferentes da língua leixificadora, o português.

b) A história social de São Tomé: Desde o início do estabelecimento das roças, em São Tomé, um número significativo de serviçais, falantes principalmente de línguas bantu, conviviam com um número reduzido de administradores que falavam português. Contudo, a constante chegada de novas levas de contratados (ciclicamente, no século XX) fez que uma língua bantu continuasse como meio de comunicação intragrupo, o que pode ter impedido a formação de uma língua estruturalmente diferente, derivada do português.

Essas razões nos levam a adotar, para este trabalho, a perspectiva de que a língua dos Tongas se constitui em um dialeto do português, e não em uma língua tipologicamente diferente e totalmente reestruturada gramaticalmente. Assim, as características socio-históricas de São Tomé e as circunstâncias que envolvem o estabelecimento das roças nos levam a assumir uma outra hipótese explicativa para as características da língua Tonga, que será discutida posteriormente: a Transmissão Linguística Irregular.

### 3 PRESSUPOSTOS TEÓRICOS

Nessa seção, serão apresentadas as bases teóricas que fundamentaram essa tese. Em 3.1, apresentaremos os principais conceitos da Sociolinguística Variacionista, seus principais conceitos e objetivos. Discutiremos, em 3.1.1, a noção de variável linguística, comentando acerca das sua gênese no âmbito dos estudos labovianos e os impasses envolvidos nessa definição. Na subseção 3.1.2, apresentaremos os fundamentos empíricos que norteiam a mudança linguística, debatendo as noções de variação estável, mudança em curso e os cinco problemas da mudança linguística. Em 3.1.3, discutiremos as diferenças entre a comunidade de fala e a comunidade de prática.

Na subseção 3.2, apresentaremos as teorias que dão conta do contato entre línguas como evento responsável pela formação de novas variedades linguísticas. Também iremos comentar os conceitos de pidgin e crioulo, que fazem parte dos estudos da Crioulística (3.2.1), buscando compreender até que ponto o contato massivo influencia a estrutura e as tendências de mudança das línguas que se formam nessas condições. Trataremos, também, da teoria da Transmissão Linguística Irregular, com vistas a explicar a estrutura de línguas formadas em situação de contato sem que uma nova língua tipologicamente se forme (3.2.2).

#### 3.1 A SOCIOLINGUÍSTICA VARIACIONISTA

A Sociolinguística Variacionista ou Sociolinguística Quantitativa (LABOV, 1972; WEINREICH, LABOV e HERZOG, 1968) é a corrente dos estudos linguísticos que tem como principal objetivo analisar a correlação entre a variação sistemática das línguas e a estrutura das sociedades, explicando os mecanismos que norteiam o surgimento das diferentes variantes linguísticas, os fundamentos da mudança linguística e compreendendo a heterogeneidade e variabilidade dos usos linguísticos como inerentes à competência linguística dos falantes. Desse modo, o objeto de estudo da Sociolinguística Variacionista é o uso concreto da língua em uma determinada comunidade de fala, inserida e uma matriz social específica. Nas palavras de Labov (2008, p. 241) a Sociolinguística Variacionista tem como principal

objetivo “o estudo da língua em seu contexto social”, explicando a correlação entre fatores sociais e a variação linguística através de dados estatísticos. De acordo com Bright (1974, apud GONÇALVES, 2008, p. 33):

O termo sociolinguística é bastante recente. Da mesma forma que seus coirmãos mais antigos, ‘etnolinguística’ e ‘psicolinguística’, não é fácil defini-lo com precisão; na verdade, estes três termos tendem a se interseccionar parcialmente quanto ao seu objeto e, até certo ponto, a refletir diferenças nos interesses e abordagens dos pesquisadores mais do que no objeto de estudo propriamente dito. De fato, é correto afirmar que os estudos sociolinguísticos, da mesma forma dos abarcados sobre o nome ‘sociologia da linguagem’, tratam da relação entre língua e sociedade. Mas tal afirmação é excessivamente vaga. Se tentarmos ser mais exatos, observaremos que a sociolinguística difere de algumas preocupações anteriores com as relações língua-sociedade, pelo fato de que, seguindo novas perspectivas da própria linguística, considera tanto a língua quanto a sociedade como sendo uma estrutura e não uma coleção de itens. A tarefa da sociolinguística é, portanto, demonstrar covariação sistemática das variações linguística e social, e, talvez, até mesmo demonstrar uma relação causal em uma ou outra direção.

Um dos principais axiomas da Sociolinguística Variacionista é o da heterogeneidade ordenada, através do qual é possível sistematizar os fatores condicionantes da variação e da mudança linguística, o que pressupõe uma visão de língua como um fenômeno heterogêneo e sistematicamente variável. Assim, considerando que a funcionalidade da língua reside na sua capacidade de se adequar aos diferentes contextos comunicativos, a Sociolinguística Variacionista quebra a suposta proporcionalidade entre funcionalidade e homogeneidade e busca descortinar os condicionamentos sociais e linguísticos da variação e os mecanismos que possibilitam a implementação da mudança linguística. A correlação entre os fenômenos linguísticos e sociais é assim apresentada por Faraco (2005, p. 65):

A língua é uma instituição com uma autonomia que lhe é própria; deve-se, portanto, determinar as condições gerais da mudança de um ponto de vista puramente linguístico (...). Contudo, como a língua é uma instituição social, segue daí que a linguística é uma ciência social, e o único elemento variável ao qual se pode apelar para dar conta de uma mudança linguística é a mudança social (...).

Nessa perspectiva, a Sociolinguística Variacionista se constitui em uma importante vertente de análise das línguas, tanto no que diz respeito à explicação dos fatos linguísticos em uma perspectiva científica, quanto em relação ao

surgimento de políticas de ensino que busquem incluir em suas pautas a heterogeneidade inerente às línguas.

A estreia da Sociolinguística Variacionista é marcada pela publicação de três textos fundamentais: a dissertação de mestrado de William Labov, intitulada *The Social History of a Sound Change on the Island of Martha's Vineyard, Massachusetts*, de 1963, a sua tese de doutorado sobre a estratificação do inglês falado em Nova York, *The Social Stratification of English in New York City* (1966), ambos orientados por Uriel Weinreich, na Universidade de Colúmbia, e o ensaio de Weinreich, Labov e Herzog, *Empirical Foundations of the Language Change*, em 1968. Essa última obra, cuja principal proposta é a contestação do suposto caráter homogêneo da língua e a discussão dos fundamentos empíricos que norteiam a mudança linguística, inaugura uma linhagem de estudos que reconhecem a importância de se correlacionar fatores como idade, classe social, sexo, entre outros, na compreensão da variação e mudança. Esses fatores, arrolados, na perspectiva variacionista, sob o nome de variáveis extralinguísticas, regem, juntamente com as variáveis linguísticas, internas ao sistema linguístico, a concorrência entre duas ou mais formas linguísticas alternativas em uma comunidade de fala, processo que pode resultar na implementação da mudança linguística.

[...] a Sociolinguística Variacionista é mais apropriadamente descrita como o ramo da Linguística que estuda as características principais da linguagem em equilíbrio entre si - estrutura linguística e estrutura social; significado gramatical e significado social - aquelas propriedades da linguagem que requerem referência a fatores externos (sociais) e internos (sistêmicos) em sua explicação (tradução nossa) (TAGLIAMONTE, 2006, p. 5).

Assim, ao incorporar o componente social à análise do funcionamento do sistema linguístico, a Sociolinguística Variacionista, em consonância com a Antropologia e a Sociologia, volta seu olhar não ao falante de forma individual, mas sim à comunidade de fala como um todo.

### **3.1.1 A variável linguística**

Com o advento dos estudos sociolinguísticos, a partir da década de 1960, passa a receber maior atenção a correlação entre os usos linguísticos e as

características sociais dos falantes, passando-se, então, a reconhecer a relevância dos fatores extralinguísticos e da análise da comunidade de fala para o entendimento da variação e mudança linguísticas, uma vez que a sistematicidade da variação linguística permite o mapeamento dos fatores envolvidos na alternância de valores de determinado fenômeno variável. Em outras palavras, pode-se afirmar que, do ponto de vista da Sociolinguística Variacionista, a concorrência entre formas linguísticas diferentes, mas equivalentes, não acontece de forma acidental, pois a variação linguística é motivada por fatores internos e externos à estrutura linguística e completamente demonstráveis, uma vez que não são aleatórios. Desse modo, afirma-se que os mecanismos que norteiam a variação das línguas e os caminhos que fazem com que a variação culmine em mudança linguística são os pontos principais da análise Sociolinguística Variacionista.

O nascimento de formas linguísticas inovadoras é um evento comum a todas as línguas, e o período de variação consiste na competição entre essas formas e as antigas e já tradicionais na comunidade de fala. Entretanto, é válido comentar que nem todos os fenômenos linguísticos são passíveis de variação: existem aqueles cuja alteração implicaria no impedimento da plena comunicação, sendo esses, então, guiados por **regras categóricas**, em oposição àqueles que se encontram em movimento, estando sujeitos à variação e à mudança – as **regras variáveis**. No bojo dessa segunda possibilidade é que a Sociolinguística Variacionista desenvolve um de seus principais conceitos: o de variável linguística.

William Labov desenvolve a noção de variável linguística fundamentado principalmente na variação verificada no nível fonológico, baseado em seus estudos realizados em Martha's Vineyard e em Nova York. Desse modo, em sua sistematização conceitual, ele afirma que os valores que podem ser ocupados para uma variável linguística

oferecem meios alternativos de dizer “a mesma coisa”: ou seja, para cada enunciado em A existe um enunciado correspondente em B que oferece a mesma informação referencial (é sinônimo) e não pode ser diferenciado exceto em termos da significação global que marca o uso de B em contraste com A (WEINREICH, LABOV, HERZOG, 2006, p. 97).

Assim, nota-se que a aplicabilidade dessa noção inicial de variável linguística é perfeitamente viável aos fenômenos do âmbito fonológico, pois é um fato a observância da equivalência de significados mesmo quando há a alternância de fonemas em determinada palavra.

Nos anos 1970, entretanto, estudos sintáticos e morfossintáticos passam a tentar transplantar o conceito laboviano de variável linguística às suas análises, sendo que essa tentativa de ampliação na aplicabilidade do conceito foi responsável por uma importante crise epistemológica da teoria sociolinguística: não é tarefa simples aplicar o princípio da equivalência referencial a outros níveis de análise linguística (SANKOFF, 1982). As críticas mais duras às limitações do conceito foram feitas a partir da publicação do artigo de Beatriz Lavandera, *Where does the sociolinguistic variable stop?*, em 1978. No texto, a autora comenta a dificuldade de tratar sobre equivalência semântica em contextos de variação não fonológica, sobretudo devido à dificuldade de se delimitar as variantes de determinada variável sintática ou morfossintática.

Na realidade, a noção de equivalência semântica implicaria uma redução muito drástica da noção de significado referencial, se a sociolinguística insistisse em manter o princípio de que duas formas alternativas são variantes se representarem o mesmo significado no mesmo contexto de ocorrência. (CAMACHO, 2010, p. 151)

Ainda sobre os limites do conceito de equivalência semântica e a elasticidade da noção laboviana de variável linguística, Freitag (2009, p. 116) afirma:

Ampliando o escopo da definição de variável para além da fonologia, faz-se necessária uma teoria de gramática ampla, que envolva os componentes fonológico, morfológico, lexical, sintático, semântico e discursivo. As fronteiras não são discretas, e, em muitos casos, há sobreposição, originando um novo nível, como o morfossintático ou o semântico-discursivo.

Com vistas a tentar solucionar o impasse dos limites conceituais da variável linguística, Lavandera (1978) afirma que, em vez de adotar o conceito fonológico, adote-se a noção de equivalência funcional. Nessa perspectiva, a variável linguística passa a ser preenchida não por formas de mesmo significado, mas sim por estruturas que apresentam o mesmo valor de verdade, ou ainda, a mesma função, quando em contextos equivalentes (FREITAG, 2009). Esse conceito é o que

adotaremos aqui para compreender alternância na realização do sujeito nulo como uma variável linguística, mas é preciso comentar que ainda que haja equivalência funcional entre as variantes que compõem determinada variável, podem ser diferentes as valorações sociais que cada uma delas recebe. Assim, para esclarecer essas diferenças entre as valorações, é importante, nesse ponto, discutir a categorização laboviana das diferentes avaliações sociais dos fenômenos variáveis.

Os usos linguísticos estão atrelados às diferentes necessidades sociais e culturais dos falantes, de modo que existe, por parte deles, uma valoração social feita sobre cada um desses usos. Essa valoração, de acordo com Labov (1972), não é homogênea: ela segue uma gradação que se relaciona ao nível de consciência dos falantes sobre essas variáveis e às “pressões sociais que operam constantemente sobre a língua” (FREITAG et al, 2016, p. 139).

A classificação laboviana consiste na divisão das variáveis linguísticas em três categorias: os **estereótipos**, os **marcadores** e os **indicadores**. Labov define os estereótipos como formas linguísticas que estão sujeitas à extrema estigmatização e rejeição por parte dos falantes, enquanto os marcadores, embora reconhecidos pela comunidade linguística, revelam apenas estratificação estilística ou social, como nível de monitoramento e faixa etária, por exemplo. Os indicadores, por sua vez, apesar de regulares, não se encontram no nível de consciência dos falantes. É importante ressaltar que as variantes menos prestigiadas socialmente podem tanto vir a desaparecer quanto ser incorporadas na estrutura linguística de modo a alcançar prestígio social.

Para Labov, reconhecer a avaliação social e os diferentes *status* sociais das variáveis linguísticas é fundamental para que se possa prever, ainda que minimamente, a forma como uma possível mudança linguística acontecerá naquela comunidade. As formas linguísticas mais bem avaliadas pelos falantes tendem a ganhar espaço na comunidade de fala, tendendo a ser vitoriosas em caso de competição entre mais de uma variante. Qualquer variável linguística é passível de receber significado social, incluindo as que não são conscientemente controladas.

### **3.1.2 Fundamentos empíricos para uma teoria da mudança**

As línguas podem sofrer alterações tanto no tempo, quanto no espaço, sendo as segundas representadas tanto pela variação social quanto pela variação observada nas diferentes áreas geográficas. Entretanto, conforme comentado nas seções anteriores, nem todos os fenômenos linguísticos variáveis levam à mudança, e a compreensão dos mecanismos da mudança linguística se relaciona diretamente com esse entendimento:

Há uma diferença entre variáveis estáveis e variáveis instáveis. A variação estável é aquela em que não existe predominância de uma variante sobre outra, havendo um equilíbrio relativo que tende a se manter por algum tempo. Diferentemente, a variação é instável quando o processo pende para uma das variantes, em detrimento das demais. É possível, então, que o uso dessa variante favorecida tenda a se generalizar, fazendo com que as demais deixem de ser usadas. A mudança linguística é fruto da variação que ocorre em variáveis instáveis. Por isso é que, algumas vezes, elas são chamadas mudança em progresso (FIORIN, 2013, p. 146).

As mudanças pelas quais as línguas passam são fatos perceptíveis por todos que as utilizam, independentemente de sua capacidade de teorização científica acerca dessas mudanças. O fato é que, se observarmos a diacronia de qualquer língua humana, perceberemos que a homogeneidade e a estaticidade são inexistentes: todos os níveis linguísticos estão passíveis de mudança, e são essas mudanças que garantem a vitalidade e a funcionalidade das línguas nos diferentes contextos sociais. Através desse entendimento, a Sociolinguística Variacionista afirma a variabilidade e a mutabilidade como características inevitáveis das línguas.

Como a mudança é gradual, é necessário passar primeiro por um período de transição em que há variação, para em seguida ocorrer a mudança. Como a mudança e a variação estão estreitamente relacionadas, é muito difícil estudar uma sem estudar a outra (CHAGAS, 2007, p. 149).

A Sociolinguística Variacionista utiliza ferramentas estatísticas para avaliar se determinado fenômeno variável se encontra em **variação estável**, quando não há um movimento de mudança a favor de uma ou outra variante, como no caso, por exemplo, da variação **nós versus a gente** no português brasileiro, e para analisar os casos em que uma das variantes se sobrepõe às outras, como na transição **vossa mercê > você** – a chamada **mudança em curso** (LABOV, 1972). Assim, essa vertente dos estudos linguísticos, além de assumir a heterogeneidade e variabilidade

das línguas, compreende que, em alguns casos, essa variabilidade culmina em processos de mudança linguística, uma vez que a língua se atualiza, tanto em função da evolução da sociedade, quanto das mudanças das necessidades comunicativas dos falantes. Para Paiva (2016, p. 24):

Formas alternativas com o mesmo significado podem estar em concorrência e, gradualmente, uma delas pode ampliar seu uso e se instalar em toda uma comunidade de fala, acarretando o desaparecimento de sua rival. Ainda que essa caracterização seja um tanto simplificadora, ela traduz bem o princípio central de que toda mudança linguística implica variação que pode durar por longos períodos de tempo. Em outros termos, a mudança é a outra face da variação característica das línguas humanas.

A Sociolinguística sempre buscou ferramentas metodológicas específicas para analisar se determinado fenômeno se encontra em variação estável ou em mudança em curso. Nesse sentido, Weinreich, Labov e Herzog (2006[1968]) propõem os famosos cinco problemas da mudança linguística, que consistem em cinco questões norteadoras das mudanças linguísticas e que devem ser respondidas por teorias que objetivem compreender e explicar os mecanismos através dos quais a mudança linguística acontece (LUCCHESI, 2012). São eles:

1. O problema das restrições (*constraints problem*), que permite identificar os contextos linguísticos e extralinguísticos em que as formas concorrentes encontram resistência e/ou favorecimento em sua aplicação, o que comprova a inexistência da variação livre;
2. O problema da transição (*transition problem*), que busca verificar, em função do tempo, qual o percurso da nova forma, ou seja, busca explicar os estágios intermediários da transição da forma A para a B;
3. O problema da implementação (*actuation problem*), que analisa quais as origens da nova forma e quais contextos favorecem seu uso, explicando de forma global o processo da mudança propriamente dita, desde o início da difusão em um determinado grupo social até a sua generalização;
4. O problema do encaixamento (*embedding problem*), que visa verificar qual ou quais outros fenômenos do sistema linguístico sofrem alterações em decorrência da implementação da nova forma, uma vez que as mudanças estão encaixadas no

sistema (socio)linguístico. Em outras palavras, pode-se afirmar que o problema do encaixamento busca avaliar a correlação entre a estrutura linguística e social e os usos de determinada variável linguística; e

5. O problema da avaliação (*evaluation problem*), que analisa o enquadramento social da mudança, avaliando a consciência social sobre o processo de variação e mudança e permitindo que se faça uma previsão do futuro de determinada forma linguística na comunidade de fala.

Responder a esses questionamentos significa lançar luz ao processo de mudança linguística que, além de lento e gradual, é pouco perceptível durante o seu curso, uma vez que, nas palavras de Fiorin (2013, p. 57), “a sincronia faz abstração do movimento da língua no tempo; ela se apresenta como um complexo de sistema de valores puros, que se interdefinem”. Desse modo, se torna necessário desenvolver aparatos metodológicos capazes de analisar, sincronicamente, o movimento diacrônico das línguas, driblando a aparente abstração da sincronia frente à mudança linguística.

Para que se identifique e interprete as mudanças linguísticas, a Sociolinguística Variacionista desenvolveu duas abordagens possíveis: a análise em **tempo real** e em **tempo aparente**, sendo essas duas estratégias metodológicas que visam à descrição da evolução de um processo de mudança diacrônica. O estudo da mudança em tempo real, que analisa o comportamento das variantes na comunidade em diversas sincronias em um período determinando e real de tempo, consiste tanto na visitação à comunidade de fala, realizando o mesmo estudo em dois ou mais momentos diferentes, na mesma comunidade, quanto na análise de textos antigos que registrem a variação do fenômeno, comparando-os com dados recentes, sendo essa segunda estratégia a mais comumente utilizada nesse método de análise. Entretanto, ela encontra algumas barreiras em seu desenvolvimento, já que,

Ao recuar no tempo, o linguista conta apenas com o testemunho da escrita. [...] Mais problemático ainda é o fato de que forçosamente textos escritos só apresentam evidências positivas, ou seja, aquilo que era usado, mas nada podemos dizer acerca daquilo que, por razões diversas, até mesmo obra do acaso, não é atestado nos registros escritos (PAIVA, 2016, p. 24).

Considerando essas dificuldades, o modelo de análise mais recorrente nas pesquisas sociolinguísticas é o estudo da mudança em tempo aparente, utilizado nesta pesquisa. Esse método consiste em uma projeção sincrônica da mudança, que acontece no plano diacrônico, e é baseado no pressuposto de que o comportamento linguístico dos indivíduos se consolida por volta dos 15 anos de idade, o que permite que se afirme que a fala de um indivíduo de 30 anos, por exemplo, repete um modelo de língua de 15 anos atrás (LABOV, 1981).

O estudo em tempo aparente pode fornecer importantes indícios acerca do processo de mudança linguística, pois permite, através da estratificação dos informantes em diferentes faixas etárias e da análise quantitativa das variáveis distribuídas nessas diferentes faixas, afirmar se determinado fenômeno se encontra em variação estável ou em mudança em curso: se informantes mais jovens estão utilizando com menor frequência uma variável que é recorrente na fala dos mais velhos, é possível que esta esteja em vias de desaparecimento (NARO, 2003). É possível, também, através do estudo em tempo aparente, verificar o comportamento das variantes fazendo inferências sobre o seu desenvolvimento, difusão e implementação em detrimento das que eram tradicionais na comunidade no momento em que o processo está em curso.

### **3.1.3 Comunidade de fala *versus* Comunidade de prática**

O conceito de comunidade de fala é um dos pilares da Sociolinguística Variacionista. A importância desse conceito pode ser explicada através da evidente indissociabilidade entre língua e sociedade que caracteriza os estudos sociolinguísticos. Ao entender a língua como um produto social e cultural, é inevitável assumir a influência compartilhamento de normas linguísticas para a formação das variedades utilizadas por determinado grupo social. A necessidade de compreender como se dá essa influência é o que torna a comunidade de fala o objeto de estudo da Sociolinguística. Labov conceitua comunidade de fala da seguinte forma:

A comunidade de fala não é definida por nenhum acordo marcado quanto ao uso dos elementos da língua, mas, sobretudo, pela participação em um conjunto de normas compartilhadas. Essas podem ser observadas em tipos claros de comportamentos

avaliativos, e pela uniformidade de seus termos abstratos de variação, que são invariáveis com relação aos níveis particulares de uso (LABOV, 1972, p. 120-121).

A partir dessa abordagem, é possível inferir que inserir falantes em uma mesma comunidade de fala vai além de assumir que eles compartilham entre si determinadas formas linguísticas: é preciso compartilhar, também, o modo de avaliar essas formas. Desse modo, as fronteiras da comunidade de fala são definidas através do compartilhamento de normas e de atitudes linguísticas. Essa decisão não é puramente metodológica: assumir a comunidade de fala como um grupo de falantes que compartilham formas linguísticas atribuiria à língua uma homogeneidade que não lhe caracteriza. O significado social das formas linguísticas no interior da comunidade de fala é que é uniforme (VANIN, 2009).

A multiplicidade de papéis e a mobilidade social do indivíduo permite que ele pertença a diferentes comunidades de fala em diferentes situações, e esse é o ponto de tensão do conceito laboviano de comunidade de fala. Pautados nisso, alguns autores, tais como Mello (2001), Eckert (2012), dentre outros, destacam, para que se possa estudar a variação linguística com maior precisão, há a necessidade de delimitar subníveis de análise da comunidade, desviando o foco da comunidade de fala para a **comunidade de prática**.

Os limites da comunidade de prática são diferentes daqueles que delimitam a comunidade de fala: enquanto, para essa, a relevância está no compartilhamento de normas e atitudes linguísticas, para aquela o foco deve ser as redes sociais dos indivíduos, ou seja, as relações sociais com as quais eles estão engajados, seja no âmbito familiar, na igreja, no ambiente escolar ou no trabalho, por exemplo. Desse modo, compreende-se que o significado das formas linguísticas reside na interação intensa e no compartilhamento de práticas sociais entre os indivíduos. A dinamicidade, sobretudo linguística, é uma das características mais notáveis da comunidade de prática:

O repertório compartilhado em uma comunidade de prática é o resultado cumulativo de negociações internas. Nesse sentido, os repertórios mencionados são dinâmicos, isto é, modificam-se conforme o seu uso e as negociações dentro da comunidade. Com isso, os seus membros compartilham de práticas culturais diversas e, por esse motivo, as escolhas na língua estão em constante

negociação, e essas ocorrem implicitamente, no decorrer das trocas comunicativas. (MEYERHOFF, 2004 apud VANIN, 2009, p. 151).

Essa abordagem é bastante cara à Sociolinguística, uma vez que a comunidade de prática é caracterizada por uma interação constante entre os indivíduos, sendo essa interação o elemento gerador das práticas linguísticas comuns. A compreensão da forma como os falantes estão vinculados é fundamental para que se compreenda como as relações sociais podem ser relevantes para a construção de significado das formas linguísticas, o que dá importantes indícios sobre as tendências de mudança linguística na comunidade analisada.

### 3.2 O ESTUDO CIENTÍFICO DO CONTATO ENTRE LÍNGUAS MASSIVO E RADICAL

Uma das principais consequências sociolinguísticas das Grandes Navegações é a formação de variedades linguísticas advindas do contato entre as línguas levadas pelos colonizadores e aquelas encontradas por eles nos continentes, sobretudo o africano. As situações de contato entre essas distintas populações, além de possibilitar significativa miscigenação étnica, fez com que as variedades linguísticas que se formaram tivessem características muito peculiares, sendo necessário que a Linguística desenvolvesse aparatos teóricos específicos para lidar com elas.

Temos, então, nos últimos anos, um crescimento significativo das teorias que buscam explicar a formação e as características inerentes às variedades linguísticas formadas nessas situações de contato, abarcando tanto as situações em que o contato massivo possibilitou o nascimento de línguas novas, os chamados pidgins e crioulos, quanto os contextos em que o contato ocasionou significativa erosão e reestruturação gramatical, mas sem que houvesse a formação de uma nova língua. Estudos pretéritos já buscavam explicar as consequências da reunião de línguas diferentes, tendo a Sociolinguística papel decisivo no que diz respeito à introdução do contexto social para a melhor explicação dessas situações.

#### 3.2.1 A Crioulística: conceitos fundamentais

A Crioulística, ciência que pode ser definida como o “ramo da Linguística que se dedica ao estudo de pidgins e crioulos” (LUCCHESI, 2016, p. 73), volta a sua atenção para os crioulos de base portuguesa desde os primeiros estudos de Hugo Shuchardt, Marques de Barros, Leite de Vasconcelos e de Adolfo Coelho, nos primeiros anos da década de 1880. No Brasil, Silva Neto foi o precursor dos estudos das línguas crioulas, principalmente após a publicação de sua obra *Introdução ao Estudo da Língua Portuguesa no Brasil*, na década de 1950. Esses autores buscaram, em suas análises, investigar a forma como a interação entre línguas distintas pode resultar em um sistema linguístico com características estruturalmente diferentes tanto da língua-alvo, ou língua leixificadora, quanto das línguas de substrato, considerando sempre a relevância das vicissitudes históricas e sociais dos povos envolvidos no contexto. Nas palavras de Moraes (2018, p. 174), a Crioulística:

trouxe importantes contribuições à linguística geral, principalmente à linguística histórica, no sentido de que a questão da mudança linguística pode ser observada quase *in vitro* nesses crioulos, devido ao seu “processo de formação (pidginização e crioulição) e transformação (descrioulição)”. Por esse motivo, “a crioulística lança muita luz sobre o assunto pelo fato de, contrariamente ao que ocorre com as línguas não-crioulas, podermos dizer com relativa segurança quando surgiu determinada língua crioula”, além de a maioria dos crioulos ter uma “história relativamente curta”.

Assim sendo, afirmamos que a Crioulística tem significativa importância no âmbito dos estudos sobre a variação e a mudança linguísticas, uma vez que é a ciência que fornece bases para a compreensão das circunstâncias que favorecem a reestruturação gramatical intensa que caracteriza as línguas pidgins e crioulas.

As definições de língua pidgin e língua crioula não são unânimes, assim como também não o são os critérios que enquadram determinada língua em uma dessas duas categorias. Rougé afirma, em 2009, que

quando se aborda a questão dos crioulos, uma das maiores dificuldades é que não existe acordo entre os linguistas sobre a definição dessas línguas. Segundo os critérios considerados, encontramos definições extremamente diferentes, a tal ponto que algumas autorizam a classificação das línguas românicas como crioulas e outras restringem essa categoria a duas ou três línguas (p. 63).

Diversos autores (HOLM, 1988; ROBERTS, 2007; PARKVALL, 2012, entre outros) compreendem as línguas crioulas como resultado de um processo de nativização de um código emergencial, o pidgin, cuja gênese se daria em um contexto de contato entre falantes de línguas ininteligíveis entre si, fato que pode ser exemplificado pelas situações de contato linguístico verificadas no período das Grandes Navegações. O pidgin, necessário para a comunicação verbal nesse contexto, passa, então, a ser a língua materna da geração seguinte, sendo que essa nativização do código emergencial origina as línguas crioulas. Desse modo, no ponto de vista desses autores, ser precedida por um pidgin é um pré-requisito para que determinada língua seja considerada um crioulo, e as línguas geradas nessas circunstâncias apresentam características estruturais que as diferem daquelas formadas em contextos “tradicionais” de transmissão linguística (SANTOS, 2020).

Um dos principais modelos de explicação da origem das línguas crioulas como advindas de pidgins nativizados é a tese do Bioprograma de Bickerton (1981, 1984, 1988), baseada nas análises contrastivas da estrutura do pidgin e do crioulo havaiano. Essa teoria, que é inatista e inscrita na base teórica da Gramática Gerativa, afirma que é possível atribuir o nascimento de uma língua crioula tendo como base o *input* linguístico advindo da língua que surge como um código emergencial em situações de contato, mas que não é a língua materna dos falantes que a utilizam e que apresenta insuficiência na sua estrutura morfossintática, o que a impede de ser considerada uma língua plena – o pidgin. Ao ter contato com o pidgin, as crianças, acessando os mecanismos da Faculdade da Linguagem, preenchem as lacunas e transformam esse *input* restrito em uma língua plena.

Desse modo, para este trabalho, teremos como base a definição de crioulo de Bickerton (apud COUTO, [1984]1996, p. 173): “Um crioulo surge quando crianças adquirem um pidgin como sua língua nativa, e o pidgin é uma língua auxiliar que surge quando falantes de diversas línguas mutuamente ininteligíveis entram em contato estreito”. Essa concepção nos permite afirmar que o surgimento de línguas pidgin e crioulas está diretamente condicionado ao contato entre duas ou mais línguas de estruturas diferentes, sendo, então, realidades exclusivas de contextos multilíngues.

O caso africano, em particular, é proeminente no que diz respeito ao surgimento de línguas crioulas de base portuguesa. Muitos estudiosos, debruçando-se sobre os crioulos africanos, dedicaram-se a traçar critérios tanto linguísticos quanto estruturais que permitem classificar essas línguas. Um dos mais importantes estudos nesse aspecto é o de Honório Couto, *Introdução ao estudo das línguas crioulas e pidgins*, escrito em 1996, que, valendo-se, também, de alguns estudos anteriores sobre o tema, tais como o de Hymes (1971), traçou critérios estruturais e socio-históricos fundamentais para a definição de um crioulo. Como critérios estruturais, o autor aponta as seguintes características (COUTO, p. 34):

- 1) Número de fonemas menor do que o das línguas que entraram em sua formação.
- 2) Preferência pela estrutura silábica CV, em geral em vocábulos dissílabos.
- 3) Ausência quase total de morfologia derivacional e flexional.
- 4) As funções sintáticas são indicadas preferencialmente pela ordem, em geral SVO.
- 5) Léxico menos numeroso do que o das línguas de superstrato e substrato.

Em relação à socio-história dessas línguas, se baseando em estudos feitos por Chaudenson (1989), o autor indica as seguintes características (op. cit.):

- 1) Todos os crioulos conhecidos surgiram do contato de povos aloglotas que não se entendiam mutuamente, sendo que uma das línguas era sempre falada por um povo socioeconomicamente mais forte, em geral o colonizador. Em suma, ex-pidgins, isto é, os crioulos, surgem sempre em comunidades multilíngues ou, na melhor das hipóteses, bilíngues. O multilinguismo continua mesmo depois que a comunidade em questão se torna um país independente.
- 2) Em geral, os crioulos surgem em ilhas ou regiões isoladas, critério conhecido como insularidade. [...]
- 3) Uma outra característica dos crioulos é, para ele [Chaudenson], a exogeneidade das populações. [...]
- 4) Para Chaudenson, só são considerados crioulos as línguas que surgiram do processo de colonização da Ásia, África e América pela Europa.

Viveka (2015, p.107, tradução nossa) vai ao encontro de Hildo Couto ao listar fatores demográficos que contribuem para a formação de uma língua em contextos de contato:

Tabela 5: Fatores demográficos para a formação de línguas em contexto de contato

### **ESTÁGIO CARACTERÍSTICAS**

|           |   |
|-----------|---|
| Evento 0  | A colônia é fundada.  |
| Evento 1  | A população de substrato supera a população de superestrato.  |
| Evento 2  | Os nativos falantes da língua de substrato superam a população total de superestrato.               |
| Evento 3  | A importação de mão de obra é interrompida.   |
| Evento X* | A população de substrato de origem local ultrapassa a população de substrato de origem estrangeira. |

\* O Evento X pode ocorrer a qualquer momento entre o Evento 0 e até depois do Evento 3, mas normalmente ocorre entre os Eventos 1 e 3.

Fonte: VIVEKA, 2015, p. 107.

Assim, conforme mostraram os autores mencionados, acreditamos ser fundamental considerar, além dos fatores estruturais, as circunstâncias socio-históricas de contato entre línguas em contextos multilíngues para que se afirme que determinada língua pode, com segurança, ser classificada como uma língua crioula. Desse modo, a historicidade dessas línguas assume caráter crucial para a sua diferenciação em relação às demais, formadas em contextos que não envolvem colonização e contato linguístico massivo e abrupto.

Em relação aos reflexos linguísticos dos processos de crioulação, Bickerton (1981, 1984, 1988 apud Lucchesi, 2009, p. 112) elenca algumas tendências estruturais das línguas crioulas, todas elas motivadas, principalmente, pela não aquisição da morfologia flexional decorrente do pouco acesso às estruturas sintáticas da língua-alvo. São elas:

- (i) focalização do constituinte através do seu deslocamento para o início da oração; (ii) sistema de artigos baseado na distinção entre os níveis de referencialidade 'específico/não específico', marcados, respectivamente pela presença/ausência do artigo; (iii) sistema analítico, em que partículas independentes marcam as categorias verbais de tempo, modo e aspecto; (iv) orações relativas com cópia pronominal; (v) estruturas de negação dupla; (vi) uso de um mesmo verbo para indicar 'posse' e 'existência'; (vii) ausência de cópula em predicados que envolvem adjetivos; (viii) o uso de adjetivos que funcionam como verbos; (ix) ordem invariável para as orações

afirmativas, imperativas e interrogativas; (x) forma bimorfêmica analítica assumida pelas palavras interrogativas (constituintes WH); (xi) escassez de construções passivas.

Nesse ponto, é válido que afirmemos que efeitos da aquisição e transmissão linguística em situação de contato fazem parte de um *continuum* de desenvolvimento. Isso significa que o contato heterogêneo pode, também, gerar como resultado línguas que, apesar de apresentarem estruturas diferentes tanto das línguas leixificadoras quanto das línguas de substrato, não podem ser classificadas como crioulos, sendo, então, originadas a partir de um processo de reestruturação parcial (HOLM, 2004; LUCCHESI, 2006), devido à especificidade das circunstâncias sociais nas quais os falantes estão inseridos:

O grau de diferença entre as línguas crioulas estaria correlacionado estritamente à quantidade de *imput* da língua dominante, que por sua vez é controlada por fatores extralinguísticos, de forma que os crioulos resultantes poderiam estar mais ou menos aproximados da língua de superestrato (SANTOS, 2019, p. 48).

As línguas resultantes desses processos de reestruturação parcial tiveram suas características explicadas em relevantes estudos linguísticos. Hildo Couto, objetivando abarcar as hipóteses epistemológicas acerca dessas línguas, traça uma série de manifestações linguísticas que podem surgir como resultado da interferência linguística, mas sem que haja a formação de uma língua nova. Comentaremos muito brevemente quatro dessas manifestações, sendo essas as que julgamos mais pertinentes às discussões levantadas nessa seção (COUTO, p.82-106):

- a) Desdialeletalização: esse processo é observado quando há a perda dos traços dialetais em um contexto de imigração, como o observado na formação da cidade de Brasília, o que o autor também chama de “nivelamento dialetal”.
- b) Semicrioulo: Para alguns estudiosos, essa designação seria a mais adequada para definir o português popular brasileiro. O semicrioulo seria o processo que define línguas que apresentam tanto traços crioulos quanto não crioulos. É importante comentar que tais traços são observáveis nessas línguas pelo fato de elas terem iniciado e não concluído, por questões sociais, o processo de criouliização, e não por uma possível criouliização – descriouliização.

- c) Anticrioulo: Este conceito envolve uma situação linguística bastante singular e, ao mesmo tempo, análoga e oposta à da criouliização. Enquanto, no caso dos crioulos, há a tendência de os falantes das línguas de substrato – que, quase sempre, estão em situação de dominação política e cultural – absorverem o léxico da língua do dominador e conservarem a estrutura gramatical das suas línguas maternas – vindo desse contexto o termo “língua leixificadora” para definir as línguas de superestrato, nos anticrioulos o processo que ocorre é o oposto: por questões de resistência cultural, o povo dominado mantém o seu léxico de origem com vistas a manter a identidade do grupo, assimilando a gramática da língua dominante. Como exemplo, o autor menciona a língua geral brasileira, que apresentava léxico majoritariamente tupi e gramática próxima à do português.
- d) *Gastarbeiter-deutsch*: Esse termo define uma espécie de pidgin derivado do alemão, que recebeu uma grande massa de estrangeiros em seu processo de industrialização. Devido ao fato de a sociedade alemã ser, nesse período, extremamente xenofóbica, esses trabalhadores estrangeiros, que, em sua maioria tinham pouca ou nenhuma escolaridade, tinham pouco acesso aos alemães e ao alemão *standart*, o que fez com que os segregou geograficamente, situação que favoreceu o surgimento de uma interlíngua baseada no alemão e estruturalmente reduzida.

Línguas formadas através de processos como os exemplificados acima costumam apresentar em suas estruturas uma “retenção substancial da morfossintaxe das línguas-fonte” (SANTOS, 2019, p. 55), ou seja, apresentam muitas características estruturais herdadas da língua-alvo, o que as afastam das línguas crioulas. Holm menciona o português vernacular brasileiro (PVB) para exemplificar um possível caso de reestruturação parcial que não culmina na emergência de uma língua nova:

Muitas das mudanças a que o PVB se submeteu podem ser caracterizadas como reduções estruturais: morfologia reduzida na marcação de tempo e pessoa em verbos, de número em substantivos e outros elementos do SN, de caso nos pronomes. Algumas vezes, essas reduções significam a perda da complexidade sintática, como no caso de orações interrogativas, ou perda de itens funcionais, como pronomes reflexivos. Se compararmos as

diferenças entre o PVB e variedades do português europeu, encontraremos muitas semelhanças com mudanças ocorridas em outras variedades parcialmente reestruturadas. Entretanto, as mudanças nos traços listados não aconteceram randomicamente nas variedades parcialmente reestruturadas, ao contrário, aquelas tornam tais variedades mais semelhantes às línguas de substrato com as quais estiveram em contato. Este é, obviamente, o resultado que a linguística de contato nos levaria a esperar (1994, p. 169).

Desse modo, entendemos que é de fundamental importância compreender a influência do contato entre línguas não somente para a formação de variedades linguísticas novas – as línguas crioulas, mas também para a compreensão das circunstâncias em que o contato linguístico enseja alterações estruturais na língua-alvo, gerando, como resultado, uma variedade nova dessa língua. Na seção seguinte, apresentaremos o conceito que busca explicar as circunstâncias de formação dessas novas variedades da língua-alvo: a Transmissão Linguística Irregular.

### **3.2.2 A Transmissão Linguística Irregular**

As investigações acerca dos efeitos do contato massivo entre línguas tipologicamente diferentes tem sido tema recorrente dos estudos sociolinguísticos nos últimos anos. Autores como Lucchesi (1994, 1996, 2001, 2002, 2003, 2006), Baxter (1998, 2003), Baxter e Lucchesi (1993, 1997), Castro (2002), Holm (1987, 1988, 1992, 2004), Mattos e Silva (1988, 1998, 2004), dentre vários outros, têm buscado demonstrar que, além da formação de pidgins e crioulos, o contato entre línguas pode dar ensejo à formação de variedades linguísticas que, apesar da proximidade em relação à língua-alvo, apresentam características gramaticais muito singulares, típicas das línguas formadas nesse contexto, sobretudo em comunidades rurais isoladas, como é o caso das grandes roças de São Tomé.

As Grandes Navegações, iniciadas no final do século XV e encabeçadas, sobretudo, por Portugal, promoveram significativo contato entre os diferentes povos envolvidos nesse processo nos quatro continentes. Consequentemente, esse contato permitiu que essas populações vivenciassem contextos de intensas trocas culturais e linguísticas, que culminaram, em muitas situações, na formação de línguas estruturalmente diferenciadas das línguas dos colonizadores – os pidgins e crioulos (LUCCHESI, 2006). Entretanto, a formação de línguas pidgins e crioulas

não é a única possibilidade envolvida no que diz respeito às consequências do contato massivo entre línguas: as variedades linguísticas que se formam podem não constituir uma língua nova, mas sim uma versão reestruturada da língua-alvo.

A formação de variedades novas da língua-alvo ocorre através do processo de Transmissão Linguística Irregular (doravante TLI) (BAXTER, 1997; LUCCHESI, 2003, LUCCHESI; BAXTER, 2006). Esse processo caracteriza as línguas que se formaram a partir do contato das línguas dominantes, sobretudo o português, o espanhol e o inglês, com a população subalternizada nas situações de aculturação características da colonização. Nessas circunstâncias, a população aloglota precisa adquirir a língua do dominador de forma bastante adversa, abrupta e, quase sempre, violenta, o que se opõe à Transmissão Linguística Regular, onde os descendentes adquirem sua língua materna de forma natural.

Nas línguas formadas por processos de TLI, há a dificuldade, por parte dos descendentes, de acesso aos elementos gramaticais essenciais da língua-alvo, em decorrência da pobreza dos dados linguísticos oferecidos a eles pelos adultos, que falam a língua do dominante como L2. De acordo com Lucchesi e Baxter (2009, p. 102), essa pobreza de estímulo é decorrente:

- (i) da dificuldade de acesso dos falantes das outras línguas aos modelos da LA, sobretudo nas situações em que o grupo dominante é numericamente muito inferior aos falantes das outras línguas; (ii) do fato de os falantes das outras línguas serem, em sua grande maioria, adultos, o que faz com que esses falantes não possam contar com os dispositivos inatos que atuam naturalmente no processo de aquisição da língua materna; (iii) da ausência de uma ação normatizadora, ou seja, de uma norma ideal que oriente e restrinja o processo de aquisição da LA, já que esse processo tem seus objetivos circunscritos aos intentos comunicativos.

Desse modo, a língua formada nessas circunstâncias de sujeição características da relação colonizador/colonizado apresenta significativa erosão em sua estrutura gramatical, já que

mecanismos gramaticais mais abstratos e sem valor referencial são normalmente eliminados, do mesmo modo os falantes adultos teriam maior dificuldade em adquirir mecanismos gramaticais da LA que estivessem ausentes em sua língua nativa, bem como fixar os parâmetros da LA distintos dos da sua língua (BAXTER; LUCCHESI; RIBEIRO, 2006, p. 104).

Sendo assim, as línguas formadas a partir de processos de TLI tendem a apresentar especificidades que servirão como modelo de aquisição para as gerações seguintes – daí se justifica compreender como *irregular* esse modelo de *transmissão linguística*, em oposto à *transmissão geracional natural* (BAXTER; LUCCHESI; RIBEIRO, 2009). Desse modo, a língua resultante desse contato é uma nova variedade histórica da língua-alvo.

Através do conceito de Transmissão Linguística Irregular, pretende-se delinear os parâmetros socio-históricos e linguísticos que possam servir de base para uma visão sistemática dos processos históricos de mudanças induzidas pelo contato entre línguas que não resultam na formação de pidgins e crioulos típicos” (LUCCHESI, 2003, p. 272).

O português popular brasileiro é um exemplo clássico de língua formada através de um contexto de contato massivo que propiciou o nascimento não de uma língua crioula, mas sim de uma versão reestruturada da língua-alvo. Segundo Lucchesi e Baxter (2009, p. 189) as razões que impediram a formação de um crioulo brasileiro podem ser explicadas através dos fatores extralinguísticos a seguir:

- a) a proporção entre a população de origem africana e branca, que permitia um nível de acesso maior à língua-alvo do que o observado nas situações típicas de contato, considerando que o nível de crioulização é inversamente proporcional ao acesso dos falantes à língua-alvo;
- b) a ausência de vida social e familiar entre as populações de escravos, provocada pelas condições sub-humanas de exploração, pela alta taxa de mortalidade e pelos sucessivos deslocamentos, o que impedia ou dificultava muito a nativização de uma possível língua crioula, sendo que essa socialização é indispensável para a emergência da língua;
- c) o uso das línguas francas africanas como instrumento de interação dos escravos segregados e foragidos: ao invés de se comunicar através de uma possível língua crioula incipiente, os africanos escravizados privilegiavam o uso de línguas francas baseadas suas línguas nativas;

- d) o incentivo à proficiência do português;
- e) a maior integração social dos escravos urbanos, domésticos;
- f) a miscigenação racial.

As circunstâncias socio-históricas mencionadas acima apresentam extrema semelhança do observado na roça de Monte Café: essa roça, assim como as demais, formadas a partir do século XX, contaram com a mão de obra de trabalhadores que, apesar de contratados, viviam e trabalhavam em situações análogas à escravidão. Assim, o contexto sociolinguístico favoreceu a formação de uma versão reestruturada da língua portuguesa, já que esses trabalhadores contavam com dados instáveis da língua-alvo para seu aprendizado. Essa versão instável da língua portuguesa serviu como modelo de aquisição para as gerações subsequentes, conforme podemos verificar em Gonçalves (2012, p. 417):

[...] o português primeiramente adquirido como L2 se começa a afirmar como a L1 da população, nomeadamente das gerações mais novas. Embora a língua-alvo seja, a priori, o português europeu, e o processo de mudança linguística não seja suficiente para o surgimento de uma língua totalmente nova, como ocorreu por exemplo com a criouliização, a verdade é que o resultado desta mudança permitirá a emergência de uma variedade linguística com propriedades distintas das do português europeu. No dia a dia é produzido um português local, com possíveis influências do crioulo, decorrente da transição histórica de segunda a primeira língua e das situações de contacto linguístico acima descritas. Esta variedade do português, resultado de uma transmissão linguística irregular, desencadeada sobretudo através da oralidade, uma vez que o acesso à escola era bastante limitado, constitui o input a que as gerações mais jovens estão expostas, transformando-se, portanto, na língua-alvo no processo de aquisição da sua L1.

Variedades linguísticas formadas nessas circunstâncias podem fornecer dados extremamente significantes no que diz respeito aos mecanismos que permeiam a reformulação linguística em contextos de contato, bem como acerca da importância dos fatores socio-históricos na remarcação dos parâmetros da língua. Os reflexos linguísticos dessas remarcações paramétricas são amplos e podem ser recuperados através da análise da estrutura das línguas resultantes. Lucchesi (2006) elenca algumas características estruturais comuns às línguas formadas a partir do processo de TLI:

(i) não aquisição/incorporação, ou variação no uso, de morfologia flexional e palavras gramaticais; (ii) gramaticalização de itens lexicais para preencher as lacunas na estrutura linguística; (iii) transferência funcional que acomoda itens lexicais do superstrato em funções gramaticais da(s) língua(s) do substrato; e (iv) alteração dos valores dos parâmetros sintáticos em função de valores não marcados.

Além das mencionadas acima, uma importante característica estrutural das línguas formadas a partir de processos de TLI é a perda ou intensa variação da morfologia flexional, especificamente da concordância verbal e nominal. Na variedade linguística dos Tongas, essa é uma característica facilmente identificável, conforme veremos nos exemplos a seguir, retirados do *corpus* que compõe essa pesquisa:

(01) Ele transmite os palavra muito linda mesmo.

(02) [Eu] Nasceu aqui em Monte Café.

(03) Para uns ôtros que não aceitava, e fugiu.

(04) Elas permanece o tempo aqui e assim as melga também.

(05) Não dava conta das criança como é devido.

Nos casos típicos de criouliização, os reflexos linguísticos são outros, estruturalmente diferentes em comparação à TLI. Há, comumente, nesses casos, o uso de itens gramaticais da língua de substrato para suprir necessidades funcionais da variedade linguística que se forma, e a aplicação de estruturas lexicais desempenhando funções gramaticais, processos denominados, respectivamente, gramaticalização (06) e transferência funcional (07) (LUCCHESI, 2009, p. 115). Os exemplos a seguir ilustram esses fenômenos:

(06) **Jénti** m'pódi komfiá kù mulé.<sup>4</sup>

(Jénti = substantivo gente gramaticalizado, exercendo papel de índice de indeterminação do sujeito)

'Não se pode confiar nas mulheres.'

---

<sup>4</sup> Papiá kristang, crioulo de base portuguesa de Malaca, na Malásia. Exemplo extraído de Lucchesi (2014, p. 283).

(07) Eu dinheiro jon-pâ já-dá tinha.<sup>5</sup>

(Pâ = forma reduzida da preposição 'para', que passa a ser um morfema de dativo)

'Eu tinha dado o dinheiro para João'

Lucchesi (2019, p. 249) elenca, ainda, outros fatos linguísticos comumente encontrados em línguas crioulas, mas que não estão presentes em línguas formadas por processos de Transmissão Linguística Irregular:

(a) a realização obrigatória do pronome sujeito, mesmo quando a língua lexificadora é uma língua de sujeito nulo; (b) a eliminação da flexão nominal de número e a criação de um pluralizador nominal, com base no pronome de 3ª pessoa do plural, como ocorre no crioulo português de São Tomé, em que a expressão *inem mina* (lit. 'eles criança') significa 'as crianças'; (c) a eliminação da flexão nominal de gênero e a indicação do gênero dos seres animados por composição lexical, com construções do tipo um filho fêmea bonito, como acontece nos crioulos portugueses da costa ocidental da África; (d) a perda da preposição a/para, com a relação de dativo indicada pela serialização verbal, como em construções do tipo comprar chapéu dar mim que significam 'comprar chapéu para mim', que ocorrem nos crioulos portugueses do Golfo da Guiné ;(e) a gramaticalização do nome para 'corpo' ou 'cabeça' como pronome reflexivo.

O estudo acurado desses eventos linguísticos possibilita a compreensão das consequências linguísticas do contato massivo entre línguas, sendo que a dimensão e a intensidade dessas reestruturações permitem que se afirme se determinada variedade linguística foi formada a partir da TLI ou se se trata de uma língua crioula, estruturalmente nova. Apesar de não serem exclusividade das variedades formadas a partir do contato, as características que se observam nessas línguas, sobretudo no que diz respeito às alterações estruturais, são muito mais rápidas em comparação às mudanças pelas quais as línguas naturais passam. Nas palavras de Lucchesi,

A diferença [entre línguas crioulas e línguas 'naturais'] residiria, então, na quantidade e na intensidade das mudanças ocorridas num lapso tão curto de tempo no processo de transmissão linguística irregular vis-à-vis ao ritmo dessas mudanças no desenvolvimento, digamos, normal da língua. Em primeiro lugar, as inovações que surgem quotidianamente na interação verbal dos falantes de uma

---

<sup>5</sup> Crioulo português do Sri Lanka. Exemplo extraído de Smith (1977) apud Lucchesi (2009, p. 116).

língua fora da situação de contato têm que vencer a força das formas e estruturas fixadas na tradição do uso linguístico, para que uma mudança possa ocorrer. Com isso, mudanças desse tipo são muito limitadas, se comparadas ao que acontece no contato abrupto e radical entre línguas que constitui a gênese do processo de transmissão linguística irregular. A erosão da estrutura gramatical que costuma acontecer nesses casos produz uma enorme pressão estrutural, que abrirá caminho para uma legião de processos de gramaticalização e transferência funcional (2009, p. 118).

Desse modo, a história social do surgimento dessas línguas, suas especificidades estruturais e a velocidade com que as mudanças ocorrem durante a sua formação é o que permite afirmar que as línguas formadas através de contato massivo e abrupto precisam receber um olhar diferenciado por parte daqueles que pretendem compreender o funcionamento das línguas humanas. Buscamos, nesta tese, ao analisar uma variedade linguística cujas características estruturais e sociais nos permitem afirmar que se formou através de um processo de Transmissão Linguística Irregular, prestar significativa colaboração para esse entendimento.

## 4 O SUJEITO PRONOMINAL NAS DIVERSAS VARIEDADES DA LÍNGUA PORTUGUESA

Esta seção é dedicada aos estudos do sujeito pronominal nas diversas variedades do português, abrangendo também a forma como esse aspecto da gramática se manifesta no latim e no crioulo de São Tomé, de base lexical portuguesa. Para cumprir esse objetivo, a seção se estrutura da seguinte maneira: em 4.1 discutimos dois conceitos bastante relevantes para esta análise: o de pronome e de pessoa pronominal. A subseção 4.2 é destinada à descrição dos quadros pronominais do latim e de algumas variedades de língua portuguesa: o português europeu, o português brasileiro, o português afro-brasileiro, o crioulo de São Tomé, que tem como base leixificadora a língua portuguesa, e o português dos Tongas. A subseção 4.3 aborda as diferentes concepções de sujeito, debatendo definições apresentadas por autores de diferentes correntes teóricas. Por fim, a subseção 4.4 é dedicada ao Parâmetro do Sujeito Nulo, apresentando as características das línguas de marcação positiva para esse parâmetro e trabalhos que buscaram analisar e discutir o funcionamento desse parâmetro em diferentes variedades linguísticas.

### 4.1 O CONCEITO DE PRONOME E DE PESSOA PRONOMINAL

Os *pronomes* são identificados como uma das categorias gramaticais que fazem parte das situações de discurso desde as primeiras gramáticas descritivas do latim. A herança do entendimento latino de pronome se faz fortemente presente na tradição gramatical do português brasileiro, cujo quadro pronominal se baseia na flexão de casos do latim: os pronomes retos ocupam posição de sujeito, que equivale ao caso nominativo do latim, e os pronomes oblíquos, que têm função de complemento, equivalem aos casos acusativo e dativo.

Em relação à definição, o pronome é comumente entendido como a palavra que acompanha ou substitui o substantivo, explicação à qual Said Ali já ia de encontro desde o século passado:

Por muito sugestivo que seja o termo, não satisfaz, contudo, à ciência da linguagem definir o pronome como palavra supridora do nome substantivo. Nada autoriza a crer que o homem, ao designar pela primeira vez os seres por meio de nomes com que os distinguir uns dos outros, se lembrasse ao mesmo tempo de criar substitutos para esses nomes (1966, p. 92).

Desse modo, o entendimento das categorias pronominais vai além de uma caracterização simples, baseada em propriedades superficiais: muitos autores, como Benveniste (1956), Câmara Jr. (1970), Monteiro (1994), Castilho (2010), dentre outros, buscam captar a chamada natureza dos pronomes, sobretudo no que diz respeito às diferenças de comportamento entre as diferentes pessoas do discurso, na busca de uma definição ideal.

Uma das mais conhecidas discussões acerca da natureza dos pronomes é a de Émile Benveniste, em *A natureza dos pronomes* (1956). Segundo o autor, a universalidade das formas pronominais torna a importância do seu entendimento não somente uma questão de línguas, mas sim de linguagem, havendo, então, a necessidade de compreender os pronomes como uma classe composta por elementos bastante diferentes entre si. Essas diferenças se iniciam, segundo o autor, no agrupamento das três pessoas do discurso na categoria 'pessoa'. Assim, ao discutir a real função dos pronomes no discurso, Benveniste insere uma nova visão da categoria pessoa pronominal. De acordo com o autor, considera-se pessoas do discurso apenas aqueles que dele participam de forma efetiva. Desse modo, a noção de pessoa pronominal se aplicaria somente à primeiras e segundas pessoas, sendo esses atores intercambiáveis. A terceira pessoa, por sua vez, não apresenta essa particularidade, por se encontrar fora da situação comunicativa e por não necessariamente fazer referência a um ser humano, ao contrário das primeiras e segundas pessoas.

Por causa dessas especificidades da terceira pessoa, Benveniste afirma que essa seria, na verdade, a *não pessoa*. Esse conceito será fundamental posteriormente, para que possamos analisar as diferenças de comportamento dos pronomes de terceira pessoa em relação ao Parâmetro do Sujeito Nulo.

Na classe formal dos pronomes, os chamados de "terceira pessoa" são inteiramente diferentes de *eu* e *tu*, pela sua função e pela sua natureza. Como já se viu há muito tempo, as formas como *ele*, *o*, etc. só servem na qualidade de substitutos abreviativos: "Pierre est

malade; il a la fièvre [Pedro está doente; ele está com febre]"; substituem um ou outro dos elementos materiais do enunciado. Essa função, porém, não se prende somente aos pronomes; pode ser cumprida por elementos de outras classes; conforme o caso, em francês, por certos verbos: "cet enfant écrit maintenant mieux qu'il ne faisait l'année dernière [esta criança escreve melhor agora do que o fazia no ano passado]". É uma função de "representação" sintática que se estende assim a termos tomados às diferentes "partes do discurso", e que corresponde a uma necessidade de economia, substituindo um segmento do enunciado e até um enunciado inteiro, por um substituto mais maleável. Assim, não há nada de comum entre a função desses substitutos e a dos indicadores de pessoa (1956, p. 282).

Assim, a compreensão de pronome pessoal proposta por Benveniste é baseada na diferenciação entre o eixo *eu/tu* em contraposição à não pessoa, representada por elementos externos à situação comunicativa e que, por isso, não podem designar nada. Desse modo, só poderiam existir duas pessoas pronominais – aquela que fala e aquela que ouve -, cujos papéis podem se inverter durante a situação comunicativa.

Monteiro (1994), indo ao encontro das afirmações de Benveniste, ressalta, também, a imprecisão da compreensão dos pronomes como substitutos dos nomes, já que várias outras categorias gramaticais podem substituir o nome, e não apenas o pronome. O autor acrescenta, também, mais duas razões pelas quais a definição dos pronomes como meros substitutos dos nomes não é precisa: nem todos os pronomes funcionam como substitutos, e aqueles que substituem nem sempre aparecem substituindo nomes (Monteiro, 1994, p. 29).

Sobre a polêmica acerca da categorização ideal das pessoas do discurso, Monteiro afirma que

julgamos ser inviável a esta altura propor uma nova designação ou tentar uma sistematização mais adequada. É incoerente, sem dúvida, colocar na mesma classe vocábulos como *eu* ou *ele*, desde que ambos possuem funções e traços morfossintáticos inteiramente diversos. O dilema então se cria: ou se reorganiza tudo à luz de critérios mais lúcidos ou se busca aproveitar o que já está feito, chamando a atenção para as falhas de interpretação (1994, p. 32).

Castilho (2010), em sua *Nova Gramática do Português Brasileiro*, busca, através de uma abordagem descritiva, compreender e explicar o estatuto dos pronomes. O autor destaca, depois de apresentar a forma como alguns teóricos

ocidentais compreendiam a categoria pronominal, a necessidade de abordar os pronomes levando em consideração as suas propriedades semântico-discursivas, que se relacionam à dêixis, à anáfora e à catáfora, às suas propriedades gramaticais de pessoa, número, gênero e caso, e às suas propriedades funcionais e morfológicas. A marcação casuística, ressaltada por Castilho, é uma particularidade dos pronomes pessoais herdada do latim.

Outra propriedade das categorias pronominais destacada por Castilho diz respeito à proximidade ou adjacência, verificada nos contextos em que a forma pronominal acompanha o nome, e à substituição, propriedade mais marcante da categoria dos pronomes, a qual é responsável pela escolha do vocábulo pro + nome (em lugar do nome) (CASTILHO, 2010, p. 475).

As diferenças comportamentais entre as duas primeiras pessoas do discurso em comparação com a terceira também são abordadas por Castilho:

Aparentemente, as línguas naturais organizam seu quadro pronominal privilegiando a codificação (i) das pessoas do discurso; (ii) dos lugares ocupados por elas no espaço físico; e (iii) de seu tempo. Por estas propriedades, os pronomes são fundamentalmente dêiticos. Mas aí surge o pronome pessoal da terceira pessoa, que traz para dentro do quadro outra propriedade, a de retomar conteúdos já mencionados. Ao caráter dêitico, mostrativo dos pronomes, agrega-se outra propriedade, a da retomada de elementos do texto, ou anáfora. Enquanto a dêixis aponta para o lugar físico ocupado por esses participantes, a anáfora aponta para um lugar abstratizado, que é o lugar-no-texto. A terceira pessoa não acarretou aos pronomes uma perda de propriedades, e sim um acréscimo [...] (2010, p. 476).

De fato, definir adequadamente a categoria dos pronomes não é uma tarefa simples, e os debates propostos pelos autores acima são apenas alguns poucos exemplos das tentativas de traçar uma definição que abarque, de forma adequada, todas as suas propriedades. Não pretendemos, aqui, nos ater às discussões sobre definições de pronome ou de pessoa do discurso, visto que esse não é o objetivo deste trabalho. Em vez disso, pretendemos apresentar, de forma descritiva, a configuração dos quadros dos pronomes pessoais da língua portuguesa e de algumas das suas variedades.

#### 4.2 OS PRONOMES PESSOAIS DO CASO RETO: DO LATIM AO PORTUGUÊS DOS TONGAS

Sabe-se que, no latim clássico, existiam formas pronominais específicas para a primeira e segunda pessoas pronominais: o pronome *ego* dizia respeito à primeira pessoa do singular, enquanto o *nos* se referia à primeira pessoa do plural. A segunda pessoa do singular era representada pelo pronome *tu*, e a segunda pessoa do plural, pelo pronome *vos*. A tabela 6, adaptada de Almeida (1995), resume a distribuição dos pronomes pessoais retos do latim clássico:

Tabela 6: Os pronomes pessoais retos do latim

|                 | <b>Pessoa</b> | <b>Nominativo</b> |
|-----------------|---------------|-------------------|
| <b>Singular</b> | 1ª pessoa     | Ego               |
|                 | 2ª pessoa     | Tu                |
| <b>Plural</b>   | 1ª pessoa     | Nos               |
|                 | 2ª pessoa     | Vos               |

Fonte: ALMEIDA, 2000, p. 136.

Como é possível observar na tabela, não havia, no latim clássico, uma forma pronominal com a função de se referir ao mundo externo à enunciação, tarefa dos ditos pronomes de terceira pessoa. A flexão verbal especificava quem era o sujeito, ou, em casos específicos, essa ausência era suprida pelo uso de substantivos ou de pronomes demonstrativos, situação que se modificou na transição latim-português, quando houve a incorporação do demonstrativo *ille* no roll de pronomes pessoais retos. Sobre essa incorporação, Penkal (2006, p. 75) afirma que

cumprir a ressalva de que nas línguas românicas, em comparação ao latim, ao invés de uma ampliação, ocorreu, de fato, uma redução no sistema de formas para indicar a terceira pessoa, uma vez que, conforme elencado acima, havia em latim uma série de demonstrativos que eram utilizados para a terceira pessoa e apenas uma dessas formas – o demonstrativo *ille* – passou a ser o pronome de terceira pessoa do português, dando origem ao pronome sujeito *ele*, assim como a forma *illu[m]*, cuja evolução compreende os estágios: *illu[m]* > *ello* > *lo* > *o* deu origem ao pronome objeto *o* e ao artigo *o*.

A origem diferenciada dos pronomes de terceira pessoa, em comparação aos de primeira e segunda, motiva a diferenciação do comportamento entre a terceira pessoa e as duas primeiras, o que tem implicações no funcionamento das flexões

causais e na realização pronominal do sujeito, por exemplo (MENDES, 2016, p. 138). Em sua explanação acerca do quadro pronominal do português europeu, Mateus (2003) aborda esse comportamento dos pronomes de terceira pessoa, conforme podemos ver na tabela a seguir:

Tabela 7: Pronomes pessoais do caso reto no português europeu

| <b>Pessoas gramaticais</b>   | <b>Apenas valor dêitico</b> | <b>Valor dêitico/ correferencial</b> |
|------------------------------|-----------------------------|--------------------------------------|
| <b>1ª pessoa do singular</b> | Eu                          | --                                   |
| <b>2ª pessoa do singular</b> | Tu/você                     | --                                   |
| <b>3ª pessoa do singular</b> | --                          | Ele/ela                              |
| <b>1ª pessoa do plural</b>   | Nós                         | --                                   |
| <b>2ª pessoa do plural</b>   | Vós/vocês                   | --                                   |
| <b>3ª pessoa do plural</b>   | --                          | Eles/elas                            |

Fonte: MATEUS, 2003, p. 819.

Para explicar a configuração do quadro pronominal do português europeu, a autora traz, inicialmente, a distinção entre os conceitos de anáfora e de correferência: a relação anafórica pode ser compreendida como aquela em que são utilizadas expressões que não têm o que a autora chama de “referência virtual” em si mesmas, como é o caso dos pronomes, por exemplo. Já nos casos em que expressões nominais com referência virtual autônoma são utilizados e relacionados discursivamente à expressão anterior, temos exemplos de correferência. Como exemplo de anáfora, a autora apresenta a oração 08:

(08) A criança magoou-**se** no jardim.

Nesse caso, o pronome reflexivo *se* não apresenta referência autônoma, sendo, portanto, o seu valor referencial fixado no sintagma nominal sujeito *a criança*. Como correferência, por sua vez, Mateus apresenta o exemplo 09:

(09) O Pedro chegou de férias; **ele** vinha moreno e satisfeito.

Nesse exemplo, o pronome *ele* pode, ainda de acordo com a autora, assumir dois valores diferentes: pode remeter ao sintagma *o Pedro*, ou seja, assumir uma referência que faz parte da situação comunicativa, tendo, assim, valor dêitico, ou pode, ainda, remeter a uma entidade referida em outro momento do discurso, o que envolve fatores semânticos e textuais para além dos valores sintáticos. Segundo Mateus, há correferência quando há coincidência entre o pronome – ou a categoria vazia – e a expressão nominal que o antecede. Os contextos de primeira e segunda pessoa, desse modo, não permitem, semanticamente, a ocorrência de correferência, sendo essa uma realidade exclusiva dos pronomes de terceira pessoa.

Castilho (2010), na *Nova Gramática do Português Brasileiro*, também aborda as diferenças entre dêixis e anáfora:

[...] os dêiticos, não é que sejam empregados em substituição aos nomes, mas sim onde os nomes não possam ser usados. Só na medida em que um pronome possa colocar-se no lugar de um nome previamente usado será portador da anáfora, posto que a peculiaridade da anáfora é uma segunda menção da pessoa anteriormente nomeada. Assim, eles [os pronomes] se dividem em dêiticos e anafóricos, embora todos caiam sob a rubrica de pronomes, pois, apesar de suas respectivas diferenças, numa coisa coincidem: denominar no lugar do nome, já que se usam ou quando o nome não pode ser empregado, ou quando, dito uma vez, não pode voltar a ser repetido (CASTILHO, 2010, p. 473).

Em seguida, o autor apresenta o quadro pronominal do português brasileiro atual. Nessa variedade de língua portuguesa, vem ocorrendo uma intensa reorganização do quadro pronominal, o que, segundo Castilho, repercute em toda a estrutura da língua: na morfologia verbal, nos outros grupos pronominais, na concordância verbal e na estrutura da sentença. O quadro pronominal de Castilho é o apresentado a seguir, e foi construído a partir de dados obtidos no *corpus* do NURC:

Tabela 8: Quadro pronominal do português brasileiro

| <b>Pessoas gramaticais</b>   | <b>FORMAL</b>                 | <b>INFORMAL</b>  |
|------------------------------|-------------------------------|------------------|
| <b>1ª pessoa do singular</b> | Eu                            | Eu, a gente      |
| <b>2ª pessoa do singular</b> | Tu, você, o senhor, a senhora | Você, ocê, tu    |
| <b>3ª pessoa do singular</b> | Ele, ela                      | Ele/ei, ela      |
| <b>1ª pessoa do plural</b>   | Nós, a gente                  | A gente          |
| <b>2ª pessoa do plural</b>   | Vós, os senhores, as senhoras | Vocês, ocês, cês |
| <b>3ª pessoa do plural</b>   | Eles, elas                    | Eles/eis, elas   |

Fonte: CASTILHO, 2010, p. 477.

No quadro acima, observa-se a coocorrência entre as formas *nós* e *a gente* para a primeira pessoa do plural, sendo que, na língua formal, a forma *a gente* é utilizada com o verbo na terceira pessoa do singular, e a forma *nós* é realizada com o verbo em primeira pessoa do plural, segundo os dados de Omena (1978) apresentados por Castilho. Na língua não-padrão, por sua vez, tanto o *nós* quanto o *a gente* são utilizados com formas verbais em terceira pessoa do singular e na primeira pessoa do plural. O autor aponta que a opção por uma ou outra forma tem sido motivada por diferenças discursivas, como abordagem de temática central e emissão de opiniões e generalizações. A forma *a gente* também aparece, em contextos de menos formalidade, como primeira pessoa do singular.

No caso das segundas pessoas, vem ocorrendo, nos grandes centros, a substituição do pronome *tu* pelo *você*, o que é resultado da gramaticalização e posterior alteração fonológica da forma de tratamento *vossa mercê*. Nesse contexto, *o senhor/ a senhora* passam a ser utilizados como uma espécie de pronome pessoal utilizado para tratamento cerimonioso. Nas regiões onde o *tu* ainda é utilizado, o *você* surge como forma de tratamento com distanciamento.

Acerca da terceira pessoa, o que mais chama atenção no quadro apresentado por Castilho é a redução fonológica que se observa nos pronomes *ele/eles* em contextos informais, que passam a ser realizados como *ei/eis*, respectivamente. Para o autor, essa redução pode ser explicada por uma questão fonética: *ele/eles* são formas dissilábicas, o que contrasta com as formas monossilábicas *eu, tu* e *cê*.

Essa redução fonológica também é observada em outras variedades da língua portuguesa. No caso do português afro-brasileiro, variedade de língua portuguesa falada por moradores de comunidades<sup>6</sup> que descendem de escravizados africanos que permaneceram isolados no interior do país até meados do século XX (LUCCHESI, 2009, p. 75), o contato massivo entre línguas foi responsável por importantes reduções flexionais, sobretudo nos pronomes oblíquos, que não são objeto de análise dessa tese. As formas pronominais correspondentes às demais pessoas apresentam comportamento idêntico à do português brasileiro informal, conforme apresentado na tabela 9:

Tabela 9: O quadro pronominal do português afro-brasileiro

| <b>Pessoas gramaticais</b>   | <b>Sujeito</b>    |
|------------------------------|-------------------|
| <b>1ª pessoa do singular</b> | Eu                |
| <b>2ª pessoa do singular</b> | Tu, você, ocê, cê |
| <b>3ª pessoa do singular</b> | Ele, ela          |
| <b>1ª pessoa do plural</b>   | Nós, a gente      |
| <b>2ª pessoa do plural</b>   | Vocês             |
| <b>3ª pessoa do plural</b>   | Eles, elas        |

Fonte: LUCCHESI, 2009, p. 481

Os dados apresentados até então comprovam que, em situações de contato entre línguas tipologicamente diferentes, a morfologia flexional é fortemente afetada. No entanto, essas alterações não são uniformes: elas podem ocorrer em maior ou menor grau, a depender das características socio-históricas que ensejaram o contexto linguístico em questão. Para compreender mais detalhadamente esse panorama, apresentamos, a seguir, o quadro pronominal do crioulo de São Tomé:

<sup>6</sup> Com base em critérios socio-históricos e socioeconômicos, foram escolhidas para fazer parte do *corpus* do português afro-brasileiro do projeto Vertentes quatro comunidades do estado da Bahia: Helvécia, Cinzento, Rio de Contas e Sapé.

Tabela 10: Quadro pronominal do crioulo de São Tomé

| <b>Pessoas gramaticais</b>   | <b>Sujeito</b>   |
|------------------------------|------------------|
| <b>1ª pessoa do singular</b> | n ~ iN           |
| <b>2ª pessoa do singular</b> | Bo               |
| <b>3ª pessoa do singular</b> | e ~ ele          |
| <b>1ª pessoa do plural</b>   | Nõ               |
| <b>2ª pessoa do plural</b>   | Nãse             |
| <b>3ª pessoa do plural</b>   | inéN ~ néN ~ inê |

Fonte: FERRAZ, 1979, p. 62 apud MENDES, 2016, p. 156.

No caso das línguas crioulas, especificamente do crioulo são tomense, cujas reduções e reestruturações gramaticais foram mais intensas em relação a outros crioulos de base portuguesa, observa-se, na primeira pessoa do singular, um processo de redução que tem como resultado a cliticização do sujeito, que passa a ser reduzido a um travamento nasal (FERRAZ, 1979, p. 62-63 apud LUCCHESI, 2009, p. 479) (10). O pronome de terceira pessoa do singular também passa por um processo de redução fonológica, conforme exemplo (11), sendo que a forma tônica *ele* é exclusiva em contextos em que é necessário dar ênfase ao sujeito, conforme exemplo (12) (LUCCHESI, 2009, p. 479):

(10) N tabla.

(Eu trabalhei.)

(11) E-fla.

(Ele disse.)

(12) Ele mé as plumê.

(Ele mesmo é o primeiro.)

Conforme já comentado, a variedade linguística analisada aqui, o português dos Tongas, se formou em circunstâncias semelhantes às que ensejaram a formação de algumas outras variedades de língua portuguesa, o que nos leva a esperar, para essa variedade, a ocorrência de fenômenos linguísticos semelhantes, sobretudo em relação a processos de variação e mudança e de redução na morfologia flexional.

Desse modo, para melhor compreender as características formais dessa variedade e para compará-la estruturalmente às demais elencadas aqui, elaboramos, com base nas ocorrências encontradas no *corpus* dessa pesquisa, o quadro pronominal do português dos Tongas, listando os pronomes pessoais do caso reto empregados em cada uma das seis pessoas do discurso.

Tabela 11: Quadro pronominal do português dos Tongas

| <b>Pessoas gramaticais</b>   | <b>Sujeito</b>    |
|------------------------------|-------------------|
| <b>1ª pessoa do singular</b> | Eu                |
| <b>2ª pessoa do singular</b> | Tu, você, ocê, cê |
| <b>3ª pessoa do singular</b> | Ele, ela, ei      |
| <b>1ª pessoa do plural</b>   | Nós, a gente      |
| <b>2ª pessoa do plural</b>   | Vocês             |
| <b>3ª pessoa do plural</b>   | Eles, elas, eis   |

Fonte: Elaboração própria.

No quadro, é importante destacar que, no caso da segunda pessoa, há a ocorrência do pronome *você* e de suas formas reduzidas *ocê* e *cê*, em concorrência com o pronome canônico *tu*, estando esse, juntamente com a morfologia verbal canônica de segunda pessoa do singular, em acentuado desuso na norma linguística do português tonga. Assim, pode-se afirmar que o uso corrente do pronome *você* e de suas variedades, bem como a inserção do sintagma nominal *a gente* no paradigma de primeira pessoa do plural e do uso do *vocês* como pronome de segunda pessoa do plural são os pontos responsáveis pelo sincretismo da

morfologia verbal do português dos Tongas, assim como se observa em outras variedades de língua portuguesa.

Outro ponto a se destacar é a variação/redução das formas de terceira pessoa, conforme já havia sido apontado por Castilho para o português brasileiro informal. Essa redução, que é de natureza crioulizante, é um reflexo da erosão estrutural que caracteriza as línguas formadas através de contato massivo e abrupto.

Observar as mudanças ocorridas nos quadros pronominais de diferentes variedades de língua portuguesa permite que se compreenda o efeito do contato entre línguas, característica presente na formação da maioria dessas variedades, para a categoria dos pronomes. Essas alterações provocaram mudanças que afetaram outros fenômenos linguísticos, como a flexão dos verbos, o que tem implicações em toda a estrutura da língua.

#### 4.3 O CONCEITO DE SUJEITO

A sistematização das categorias linguísticas consiste em um importante reflexo da forma como determinada época compreende os fatos linguísticos. A categorização clássica, comumente encontrada nas gramáticas tradicionais e nos compêndios didáticos, é baseada em supostos atributos que, juntos, permitem que os elementos linguísticos sejam agrupados em uma mesma categoria gramatical.

O *sujeito* é uma categoria gramatical que suscita muitos debates, tanto no âmbito dos estudos descritivistas quanto na tradição gramatical prescritiva. No âmbito da gramática tradicional, é comum que a função *sujeito* seja identificada através da análise da organização oracional, o que envolve tanto a posição do elemento sujeito na sentença quanto a sua relação de concordância com o elemento verbal (CUNHA E CINTRA, 1997).

Desse modo, tanto nos livros didáticos quanto nas gramáticas descritivas e prescritivas é comum que se atribua ao sujeito três características gerais: 1) estabelecer relação de concordância com o verbo da sentença, 2) praticar a ação descrita pelo verbo, 3) ser o ser sobre o qual se declara alguma coisa. Entretanto, são muitos os casos em que o sujeito oracional não atende a todos esses atributos,

estando esses relacionados apenas aos casos ditos prototípicos. Esse impasse epistemológico, ao contrário do que se pode pensar, não tem relação com a categoria gramatical em si, mas sim com problemas atinentes à forma de compreendê-la.

Assim sendo, traremos, aqui, muito resumidamente, uma vez que a discussão das diferentes concepções de sujeito não faz parte do escopo desse trabalho, abordagens que consideramos prototípicas no que se diz respeito à discussão do conceito de sujeito: a de Rocha Lima (2011), a de Evanildo Bechara (2009), a de Ataliba de Castilho (2010) e a de Maria Eugênia Duarte (2007). Não buscamos, com isso, esgotar a discussão acerca das possibilidades epistemológicas da categoria, muito menos sugerir um novo conceito de sujeito. Nosso objetivo, aqui, é apresentar possibilidades de entendimento do conceito de sujeito, abordando perspectivas diferentes de compreensão dessa categoria gramatical. É importante destacar que, embora seja a abordagem tradicional a de maior alcance no que diz respeito ao estudo das prescrições gramaticais, a definição de sujeito não é, em muitos casos, consenso nem mesmo entre autores prescritivistas.

A *Gramática Normativa da Língua Portuguesa*, de Rocha Lima, inicia a sua análise do sujeito definindo-o através de seu caráter semântico: para o autor, o sujeito pode ser compreendido como “o ser sobre o qual se diz algo”, sendo o predicado, por sua vez, “aquilo que se diz sobre o sujeito”. Assim, pode ocupar o posto de sujeito tanto um sintagma nominal quanto apenas um substantivo ou elementos equivalentes, como os pronomes.

Sobre a omissão do sujeito, o autor aborda duas possibilidades: o caso dos sujeitos indeterminados, que ocorrem quando o falante não deseja ou não pode expressar o sujeito, e o caso das orações sem sujeito, que ocorre quando a ação verbal não é atribuída nenhum ser, o que, segundo o autor, é observado nos verbos que denotam fenômenos da natureza e nos verbos impessoais.

No que diz respeito à posição dos constituintes nas orações, Rocha Lima afirma que, na língua portuguesa, é comum que o sujeito apareça “no rosto da oração”, embora haja possibilidade de inversão dessa ordem. Porém, ele destaca que é preciso ter cautela ao realizar essas inversões, uma vez que verbos

intransitivos podem levar a interpretações incorretas em relação ao sujeito. Assim, o autor afirma que, para que se evite a confusão entre objeto direto e sujeito posposto, é necessário analisar a transitividade verbal e “procurar” o sujeito do verbo.

Em relação ao sujeito pronominal, Rocha Lima afirma que é comum, na língua portuguesa, a elipse do sujeito, uma vez que “são explícitas as nossas desinências verbais”. Em casos de ênfase, entretanto, é necessário que o sujeito esteja foneticamente expresso.

A gramática de Evanildo Bechara, intitulada *Moderna Gramática Brasileira*, aborda o sujeito atribuindo a essa categoria uma característica que vai ao encontro do que Rocha Lima afirma. O autor compreende essa categoria como

unidade ou sintagma nominal que estabelece uma relação predicativa com o núcleo verbal para constituir uma oração. É, na realidade, uma explicitação léxica do sujeito gramatical que o núcleo verbal da oração normalmente inclui como morfema número-pessoal (BECHARA, 2009, p. 338).

Desse modo, Bechara interpreta o sujeito como categoria fundamentada apenas na sintaxe, já que sua característica mais marcante é a concordância com o verbo da sentença – a exceção seria o sujeito indeterminado, que, para Bechara, apresenta caráter semântico. Ele destaca, também, que o sujeito, ao contrário do que a maioria das definições afirmam, não é um termo imprescindível à oração, o que pode ser comprovado através das orações sem sujeito - embora sua presença seja mais frequente na língua portuguesa em relação à não realização. Para Bechara, a não realização – ou não lexicalização, conforme o próprio autor nomeia, não deve ser tratada como elipse de sujeito, uma vez que a desinência número-pessoal do verbo representa, por si só, a presença do sujeito na oração.

Não se pode falar a rigor em elipse do sujeito, quando aparece apenas o núcleo verbal da oração (estudo, brincamos), já que ele aparece sempre presente na forma verbal flexionada no morfema que representa o sujeito gramatical (1ª, 2ª e 3ª pessoas do plural e do singular). Trata-se, pelo contrário, da sua expansão ou não, mediante o sujeito explícito (quando há, é claro, relação predicativa referida, mas do texto, para a transmissão efetiva e clara da mensagem) (BECHARA, 2009, p. 409).

Assim, Bechara compreende que a lexicalização do sujeito atende a critérios pragmáticos da comunicação.

Castilho, na *Nova Gramática do Português Brasileiro* (2010), destaca a dificuldade de se encontrar uma definição adequada de sujeito pronominal. Segundo o autor, a tradição gramatical simplifica demasiadamente a análise dessa categoria, que é multissistêmica e, por isso, deve ser abordada considerando todas as suas diferentes propriedades. Desse modo, Castilho aponta como primeira generalização sobre a categoria as suas possibilidades de preenchimento: a posição do sujeito pode ser ocupada por diferentes classes, tais como expressões nominais, como os substantivos, expressões pronominais ou por uma sentença. A posição de sujeito pode, também, ser “ocupada” por uma categoria vazia.

No que diz respeito à colocação do sujeito, o autor destaca a possibilidade de o sujeito preenchido ser anteposto ou posposto ao verbo. Ao analisar dados do NURC sobre a realização do sujeito e fazer um recorte em relação à sua ordem na sentença, Castilho destaca que 60% dos sujeitos preenchidos são antepostos ao verbo, tendo como condicionantes para esse número os seguintes fatores (CASTILHO, 2010, p. 291):

- (i) ser foneticamente “leve”, isto é, conter até 7 sílabas;
- (ii) ser sintaticamente construído com especificadores dispostos à esquerda do núcleo e por complementadores dispostos à direita do núcleo;
- (iii) ser semanticamente não específicos e agentivos;
- (iv) ser informacionalmente já conhecidos;
- (v) ser textualmente já mencionados.

Sobre os sujeitos pospostos ao verbo, há, também, certos fatores que atuam para o seu favorecimento:

- (i) Nas sentenças reduzidas de infinitivo, gerúndio e particípio;
- (ii) Nas orações intercaladas “construídas com verbos do tipo ‘dizer’, ‘sugerir’, ‘perguntar’, ‘responder’ e sinônimos, que arrematam enunciados em discurso direto” (CUNHA E CINTRA, 1985, p. 159);
- (iii) Nas sentenças interrogativas, exclamativas, imperativas e optativas;
- (iv) Nas condicionais em que se tenha omitido o *se*;
- (v) Nas sentenças de verbo intransitivo, como *aparecer*, *correr*, *andar*, *entrar*, *existir*, *faltar*, *ocorrer*, *sair*, *soar* etc.

Sobre o sujeito como categoria vazia, o autor destaca os estudos acerca do parâmetro *pro-drop* no âmbito do PE e do PB, destacando as diferenças de comportamento entre essas duas variedades no que diz respeito ao licenciamento do sujeito nulo. Castilho reconhece a correlação entre a omissão do sujeito e a riqueza da morfologia verbal das línguas, apontando, com base nos estudos de Duarte (1993), os contextos que favoreceriam a omissão fonética do sujeito pronominal. Os dados apresentados pelo autor através dos estudos de Mary Kato (1996), por sua vez, mostram que nem sempre há a correlação entre morfologia verbal pobre e preenchimento do sujeito, uma vez que, no *corpus* analisado por ela, o sujeito preenchido é altamente favorecido pela primeira pessoa do singular, pessoa cuja morfologia verbal se diferencia das demais em quase todos os tempos verbais.

O próximo tópico abordado por Castilho traz as propriedades discursivas do sujeito, através das quais se pode reconhecer o sujeito como aquele sobre o qual se declara algo, destacando, entretanto, a necessidade de se diferenciar as categorias tema discursivo e sujeito sintático.

A análise de Castilho se encerra com a discussão sobre as propriedades semânticas do sujeito: a agentividade, a animacidade/não-animacidade, a referencialidade/não referencialidade e a determinação/indeterminação. O agente, que, em muitos casos, coincide com o sujeito, é o constituinte responsável pela ação expressa pelo verbo. Entretanto, nem sempre essa coincidência ocorre, conforme aponta o autor no exemplo transcrito em 13:

(13) O assassino foi preso pelo **guarda**.

Em relação à animacidade e não-animacidade, o autor afirma que essa propriedade só pode ser atribuída a sujeitos de terceira pessoa, uma vez que os sujeitos de primeira e segunda pessoas, por remeterem, invariavelmente, às pessoas envolvidas no discurso, sempre apontam para entidades animadas. Essas características são, segundo Castilho, especificadas no discurso.

A referencialidade/não referencialidade é uma propriedade que se relaciona à definitude do sujeito. Desse modo, é referencial o sujeito que “destaca determinado

referente dentre o conjunto dos referentes possíveis”. Os sujeitos indeterminados, por sua vez, são aqueles cuja referência é genérica, característica que só pode ser avaliada contextualmente.

A análise de Castilho é uma das mais importantes no âmbito da compreensão da categoria do sujeito, uma vez que é uma das poucas de caráter descritivo-científico, baseada em linguística de *corpus*. Apesar de não se preocupar em apresentar definições fechadas para o sujeito e para as suas propriedades, Castilho levanta importantes discussões acerca da importância de se compreender essas propriedades antes de tentar traçar uma definição, devido à complexidade dessa categoria.

Por fim, traremos a última abordagem da noção de sujeito gramatical: a de Maria Eugênia Duarte em *Termos da Oração*. Assim como Bechara, a autora inicia a sua discussão questionando a suposta essencialidade do sujeito gramatical, uma vez que essa noção pode fazer com que se acredite que o termo essencial tem mais importância no contexto oracional que os demais termos. Sua crítica gira em torno da pouca ou nenhuma contribuição que a noção de essencialidade traz para a compreensão do funcionamento dessa categoria gramatical, uma vez que as suas relações no contexto da sentença em nada se relacionam com o seu nível hierárquico. Além disso, para Duarte, as orações sem sujeito também comprovam a insustentabilidade da noção de essencialidade.

Para a autora, é ideal que se considere não o sujeito e o predicado como termos essenciais, e sim o predador em torno do qual os demais termos se articulam. Nesse caso, o sujeito passa a ser compreendido como argumento externo, cuja existência está ancorada à seleção semântica realizada pelo predador. As orações sem sujeito, nesse caso, seriam, na realidade, aquelas cujo predador seleciona apenas argumentos internos.

Duarte afirma, ainda sobre a forma de se compreender o sujeito, que não tem relevância a sua classificação em simples ou composto, uma vez que a coordenação entre sintagmas não é uma exclusividade da função sujeito, podendo ser aplicada também aos objetos, por exemplo. Em relação ao sujeito oculto e ao sujeito indeterminado, só há, para a autora, sentido nessas classificações se elas se opuserem, respectivamente, às noções de sujeito expresso e sujeito determinado,

entendendo essas noções como necessárias pra que se compreenda a propriedade de omissão do sujeito nas diferentes línguas humanas.

De acordo com a autora, é fundamental que seja repensada a categorização tradicional sobretudo no ensino de língua portuguesa, já que

reconhecer e identificar os constituintes da sentença é importante para que o aluno compreenda, por exemplo, a concordância entre o verbo e o argumento externo e, sobretudo, porque existe dificuldade maior em realizar tal concordância quando o “sujeito” sintático se comporta como um argumento interno, estrutura típica dos verbos inacusativos [...] reconhecer e identificar é ainda mais importante para a boa utilização dos sinais de pontuação: o aluno entenderá melhor, por exemplo, que a vírgula não deve ser usada entre o predicador e os seus “argumentos” (DUARTE, 2007, p. 204).

Como vimos, são razoavelmente divergentes entre si as abordagens da noção de sujeito apresentadas acima, sendo claro que a definição tradicional de sujeito só atende aos casos prototípicos dessa categoria, estando os demais casos fora dos critérios pré-definidos. Por isso, muitas vezes, esses casos que fogem ao padrão são tratados através de acréscimos e estratégias epistemológicas que se afastam da noção inicial. Desse modo, é fundamental que a categoria *sujeito* seja compreendida de forma ampla, uma vez que se insere no plano sintático, no semântico e no discursivo.

#### 4.4 O PARÂMETRO DO SUJEITO NULO

Desde a gênese da Teoria Gerativa, entende-se a Gramática Universal como o estágio inicial do conhecimento linguístico humano, sendo composta por um conjunto de princípios, que são fenômenos invariáveis e comuns a todas as línguas humanas, e parâmetros, que consistem em propriedades subespecificadas e, por isso, passíveis de variação intersistêmica (KATO, 2002, p. 311). Assim, as razões que levam à variação e mudança na marcação desses parâmetros são muito relevantes para que se compreenda a forma como a variação interlinguística acontece, bem como para compreender as generalizações possíveis entre as diferentes línguas.

O protótipo da análise paramétrica é o Parâmetro do Sujeito Nulo ou o Parâmetro *pro-drop* (do inglês *pronoun drop*, ou “queda do pronome”) (CHOMSKY,

1981, 1986, PERLMUTTER, 1971, RIZZI, 1982, dentre outros), sendo esse, desde que foi formalmente proposto, no início dos anos 80, o parâmetro mais discutido da teoria de Princípios e Parâmetros. A partir daí, a sintaxe do sujeito, no bojo da Teoria Gerativa, ganha forma, uma vez que o sujeito é uma característica comum a todas as línguas, e a variabilidade em seu uso pode fornecer importantes dados acerca da aplicabilidade dos achados da Teoria de Princípios e Parâmetros às línguas humanas.

Assim sendo, entendemos como sujeito nulo, aqui, a forma de se interpretar os sujeitos referenciais, e, especificamente no caso desse trabalho, pronominais, sem realização fonética na posição tradicionalmente ocupada pelo sujeito na sentença. É importante ressaltar, entretanto, que não devemos fazer uma abordagem simplista desse parâmetro: não se trata simplesmente da marcação positiva ou negativa do sujeito pronominal, mas sim da análise das circunstâncias em que as línguas licenciam o uso do sujeito nulo e vice-versa (DUARTE; REIS, 2018, p. 175)

Hoje não fazemos apenas uma leitura paramétrica dos resultados das pesquisas variacionistas, pois já dispomos de uma teoria linguística suficientemente “adulta” e muito bem fundamentada. Tomamos o elenco de propriedades relacionadas a um determinado valor paramétrico como nosso ponto de partida, desde a formulação de hipóteses e o levantamento dos grupos de fatores até a interpretação dos resultados. Temos podido acompanhar a evolução da TP&P, desde sua formulação, passando por reformulações para acomodar as evidências empíricas vindas de diferentes sistemas linguísticos, que só contribuem para seu aperfeiçoamento.

Ao comparar o italiano com o inglês e o francês, Chomsky inicia a linhagem de estudos sobre o comportamento das línguas em relação ao licenciamento da não realização do sujeito pronominal. Inicialmente, o autor atribui como principal propriedade dos sistemas *pro-drop* a riqueza da morfologia verbal, assumindo, entretanto, a possibilidade da existência de sistemas mistos para esse parâmetro:

Chomsky aponta que línguas como o irlandês e hebraico, devido à natureza da flexão verbal das mesmas, poderiam não apresentar correlações absolutas a respeito dessas propriedades das línguas de sujeito nulo, mas sim uma “tendência” para uma ou outra possibilidade do sistema. (VERÍSSIMO, 2017, p. 79)

A discussão acerca da aplicabilidade do parâmetro do sujeito nulo e da existência de sistemas mistos ganhou ainda força em meados da década de 80, quando Huang publicou os resultados de sua análise desse parâmetro em línguas orientais. Apesar de não apresentarem flexão verbal, essas línguas admitem a ocorrência do sujeito nulo. Desse modo, para a autora, línguas de tópico, tais como o chinês e o japonês, licenciam e identificam sujeitos nulos não através da morfologia flexional, mas sim através do contexto comunicativo, uma vez que são orientadas para o discurso, e não para a sentença. Assim, línguas como o português europeu licenciam o sujeito nulo através de um controlador interno – a concordância verbal, enquanto nas línguas orientadas para o discurso a correferência identifica esses sujeitos (KATO; DUARTE, 2014).

Esse estudo, bem como vários outros realizados posteriormente, comprovam que a análise do Parâmetro do Sujeito Nulo não se resume apenas a uma questão de marcação ou não marcação, mas sim que envolve questões gramaticais, discursivas, sintáticas e semânticas mais complexas e que, portanto, precisavam ser mais bem sistematizadas.

Um estudo também de grande relevância acerca do Parâmetro do Sujeito Nulo é o de Nicolis (2008), realizado através da observação de nove línguas crioulas. O autor encontra, nessas línguas, algumas das propriedades anteriormente estabelecidas como pertencentes ao sujeito nulo, o que o leva a concluir que essas línguas seriam parcialmente positivas para o esse parâmetro (VERÍSSIMO, 2017).

Assim, estudos comprovam que o comportamento do Parâmetro do Sujeito Nulo está muito mais relacionado às restrições em maior ou menor grau ao licenciamento da marcação/não marcação do sujeito do que à simples admissibilidade de uma ou outra forma. Então, pretendemos, aqui, além de analisar a realização do sujeito no português dos Tongas, escrutinar os contextos sintáticos que favorecem um ou outro uso, com vistas a não reduzir a análise a uma perspectiva binária.

Com o intuito de interpretar o licenciamento do sujeito nulo como uma propriedade morfológica atinente a algumas línguas e motivada por um conjunto de fatores relacionados a essas línguas, alguns autores buscam listar quais

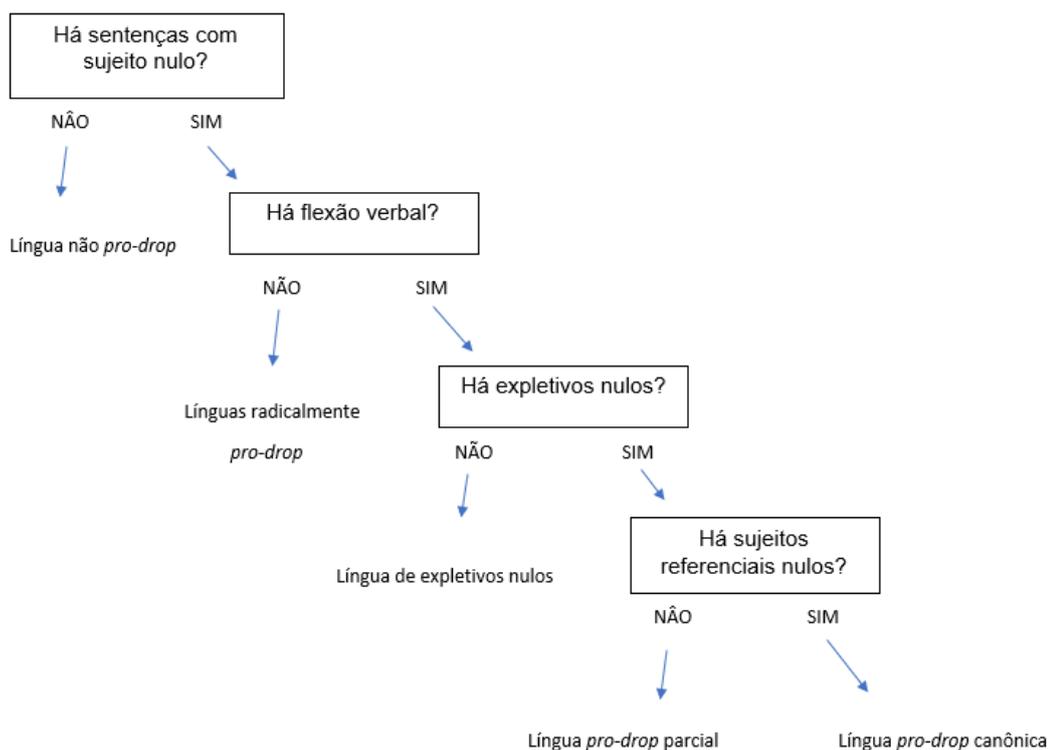
particularidades sintáticas permitem a omissão do sujeito pronominal, estando a maioria desses estudos relacionada à riqueza/pobreza flexional das línguas, ao modo de orientação (ao discurso ou à sentença) e às categorias funcionais do sujeito nulo, já que algumas línguas só licenciam o parâmetro *pro-drop* para sujeitos específicos, como o expletivo, por exemplo. Desse modo, Holberg e Roberts (2009), analisando um feixe de línguas indo-europeias distintas entre si, ampliam a concepção de sistemas mistos proposta por Chomsky, listando quatro tipologias linguísticas em relação ao licenciamento do sujeito nulo:

1. Línguas de sujeito nulo canônicas: São as línguas que licenciam o sujeito referencial não expresso/nulo, tais como o italiano, o português europeu, o grego, o turco e o basco. Nessas línguas, a pessoa do discurso pode ser retomada através da flexão verbal.
2. Línguas de sujeito nulo parcial: Nessas línguas, os sujeitos sem referência definida são, em geral, nulos, havendo restrições na aplicação do parâmetro *pro-drop* em sujeitos referenciais. Exemplos: o finlandês, o hebraico e o português brasileiro.
3. Línguas de sujeito nulo radicais: Línguas que licenciam a não expressão tanto o sujeito quanto do objeto, sem que haja flexão verbal indicativa da pessoa do discurso. Como exemplo, tem-se algumas línguas asiáticas, tais como o chinês, o japonês, o coreano, entre outras. É importante comentar, também, que as línguas de sujeito nulo radicais são um tipo diferente de línguas de sujeito nulo canônicas: enquanto, para essas, a omissão do sujeito é licenciada pela sintaxe, para aquelas esse licenciamento se dá através do discurso.
4. Línguas de sujeito nulo expletivo: São aquelas línguas em que um sujeito referencial completo não pode ser nulo, mas um sujeito expletivo pode. Exemplos: o holandês e o alemão.

Assim, enquanto línguas como o francês e o italiano licenciam o parâmetro *pro-drop* através da concordância verbal, línguas como o chinês e o japonês o fazem através dos níveis de referencialidade desses sujeitos (KATO; DUARTE, 2014), o que nos leva a concluir que são muitas as perspectivas de abordagem do

licenciamento no parâmetro *pro-drop*, sobretudo nas línguas românicas. Classificar essas línguas através de diferentes tipologias permite que se abarque as especificidades de cada uma delas. No intuito de sistematizar e hierarquizar essas características, Veríssimo (2017, p. 84) propõe o esquema a seguir:

Figura 5: Esquema de classificação das línguas em relação à realização do sujeito



Fonte: VERÍSSIMO, 2017, p. 84.

A compreensão do funcionamento do Parâmetro do Sujeito Nulo em sistemas compreendidos como mistos é especialmente relevante para o entendimento desse parâmetro em línguas que foram afetadas e estruturalmente modificadas pelo contato massivo entre línguas, como é o caso do português brasileiro e do português dos Tongas, uma vez que essas variedades de língua costumam apresentar características específicas em relação à marcação dos mais diversos parâmetros linguísticos. O tema ainda é objeto de muitos estudos no âmbito da Teoria Gerativa, e seu refinamento tem fornecido importantes contribuições para o entendimento do funcionamento das línguas naturais.

Na subseção seguinte, pretendemos apresentar, sucintamente, alguns dos mais importantes estudos que buscaram compreender o licenciamento do sujeito nulo nas línguas humanas e o poder de predição da Teoria de Princípios e Parâmetros no que diz respeito à sua aplicabilidade, apontando as múltiplas possibilidades que envolvem o entendimento desse parâmetro. Já que estamos lidando, aqui, com uma variedade africana de português, apresentaremos trabalhos realizados no âmbito do português europeu, português brasileiro, português afro-brasileiro e em algumas línguas crioulas.

#### **4.4.1 O Parâmetro do Sujeito Nulo em algumas variedades de língua portuguesa**

O funcionamento do sujeito nulo é um fenômeno recorrentemente analisado em diversas línguas e, no caso da língua portuguesa, há vários estudos do fenômeno em diferentes variedades. Desse modo, nessa subseção, abordaremos os aspectos centrais de algumas análises em relação à expressão do sujeito realizadas em algumas variedades da língua portuguesa: o português europeu, no português brasileiro, no português afro-brasileiro, no português angolano e no português moçambicano. Apresentaremos, também, os resultados encontrados em relação ao Parâmetro do Sujeito Nulo em algumas línguas crioulas que, embora não sejam de base portuguesa, se formaram em contextos socio-históricos semelhantes aos que ensejaram a formação do português dos Tongas.

Acreditamos que os resultados encontrados pelos autores desses trabalhos podem oferecer contribuições epistemológicas muito valiosas em relação à variação na realização do sujeito em línguas formadas em contextos de contato linguístico.

##### *4.4.1.1 O sujeito pronominal no português coloquial europeu (DUARTE, 1995)*

Esse é um estudo de fundamental importância para a compreensão da realização do sujeito em língua portuguesa, já que se trata da primeira análise desse parâmetro com dados reais de fala. No ponto de vista da autora, a adição de duas novas formas de tratamento de segunda pessoa (você e senhor) e o sincretismo de formas verbais que se manifesta através do uso da terceira pessoa do singular junto a essas formas não altera as seis formas distintivas que caracterizam a morfologia flexional dos verbos do português europeu (PE, doravante), o que permite que se

afirme que o PE é uma língua de sistema flexional funcionalmente rico. Assim, a autora teve como principal hipótese que a regularidade funcional do paradigma verbal licenciaria o uso de pronomes nulos em detrimento dos sujeitos plenos até mesmo em contextos de sincretismo flexional (exceto em casos em que não se fosse possível desfazer, semanticamente, a ambiguidade).

Com vistas a constituir sua amostra de fala, a Duarte analisou entrevistas realizadas em 1987, em Portugal. Os informantes que compuseram a amostra estavam divididos em duas faixas etárias (22-33 anos e 46-60 anos) e três níveis de escolaridade (básico, médio e superior), totalizando 30 informantes. Assim, a autora, aplicando uma leitura paramétrica aos resultados quantitativos obtidos, chega a cinco importantes conclusões, resumidas a seguir:

- I. O sujeito nulo é predominante em todas as pessoas do discurso e em todos os tipos de oração.

Tabela 12: Frequência de sujeitos nulos em PE

| <b>Pessoa</b>        | <b>Em oração coordenada</b> | <b>Sem oração coordenada</b> |
|----------------------|-----------------------------|------------------------------|
|                      | Sujeitos nulos (%)          | Sujeitos nulos (%)           |
| <b>1<sup>a</sup></b> | 60                          | 53                           |
| <b>2<sup>a</sup></b> | 73                          | 72                           |
| <b>3<sup>a</sup></b> | 73                          | 64                           |

Fonte: DUARTE, 1995, p. 8.

Os dados obtidos são muito significativos, sobretudo em relação à primeira pessoa, como é possível visualizar na tabela 12. Apesar de ter desinência exclusiva, tanto no singular quanto no plural, essa pessoa é a que apresenta os menores índices de sujeito nulo (ainda que sejam esses mais numerosos que os sujeitos foneticamente realizados). Esses dados dão indícios de que, na variedade analisada, a riqueza da flexão verbal pode não ser tão relevante para o preenchimento/não preenchimento da posição do sujeito. Sobre esse resultado, Duarte afirma que (p. 9):

A ocorrência equilibrada de sujeitos nulos e plenos de primeira pessoa não parece ser exclusiva do português europeu. Em estudo sobre o italiano oral, Ochs e Duranti (1979) já apontavam o fato de que o falante geralmente não se apresenta numa conversa usando apenas o mecanismo de concordância verbo-sujeito; na maioria dos casos um pronome pleno é utilizado.

II. O sincretismo flexional não impede o uso do sujeito nulo.

A análise feita por Duarte comprova, também, que não há o prejuízo no uso do sujeito nulo em contextos de sincretismo (ou, nas palavras da autora, de formas verbais homógrafas). Esse fato pode ser comprovado no exemplo a seguir, citado pela autora

...porque me parecia uma coisa detestável, um problema grande de consciência, se amanhã como magistrado viesse a verificar que tinha errado ou que tinha julgado mal, sobretudo se tivesse julgado mal em prejuízo de alguém... Hoje gostava de ser magistrado porque estou convencido que talvez evitasse mais injustiças como magistrado do que as que posso evitar como advogado (p. 11).

III. Não há diferença quantitativa significativa na frequência de uso entre orações coordenadas e não-coordenadas.

Outro ponto importante em relação ao encaixamento sintático do fenômeno diz respeito à frequência em orações coordenadas e não-coordenadas. Ao contrário do que se esperava, as diferenças de frequência entre um e outro tipo de oração é muito pequena, caindo sete e nove pontos na primeira e terceira pessoa, respectivamente, quando se excluem as orações coordenadas da rodada. Como era de se esperar, a frequência se mantém praticamente inalterada quando se faz uso da segunda pessoa.

IV. O traço semântico [-animado] favorece o uso do sujeito não realizado foneticamente.

O quarto ponto que julgamos importante em Duarte diz respeito ao traço semântico do sujeito: enquanto 69% dos sujeitos nulos apresentam o traço [+animado], 93% dos que apresentam o traço [-animado] são nulos. Esses últimos são, inclusive, preferenciais até mesmo nas posições que apresentam restrições para esse uso. As cinco ocorrências de traço [-animado] com sujeito pleno

aparecem, nas palavras da autora, em posições pouco ortodoxas, como em contextos com elementos intervenientes entre o sujeito e o referente, por exemplo.

- V. Não há diferença significativa de uso do sujeito nulo entre as diferentes faixas etárias e níveis de escolaridade.

O quinto e último ponto que consideramos relevante diz respeito a duas variáveis sociais importantes para o entendimento das tendências de mudança linguística: a faixa etária e a escolaridade. Na análise de Duarte, não há relevância dessas variáveis para o comportamento do sujeito nulo, o que nos fornece importantes indícios acerca do avanço do processo de apagamento do sujeito realizado no português europeu: acreditamos que as diferentes faixas etárias e os diferentes níveis de escolaridade tendem a apresentar comportamentos semelhantes quando se trata de uma situação de variação estável. Uma análise mais recente do fenômeno pode confirmar ou refutar essa hipótese.

#### 4.4.1.2 A perda do princípio “evite pronome” no português brasileiro (DUARTE, 1995; 2012)

A pesquisa realizada por Maria Eugênia Duarte em 1995, para a sua tese de doutorado, orientada pela professora Mary Kato, é, sem dúvida, o trabalho mais importante sobre o Parâmetro do Sujeito Nulo no âmbito do português brasileiro, tanto por seu pioneirismo quanto pela amplitude da análise, que ofereceu as bases para a compreensão do funcionamento do parâmetro *pro-drop* e dos principais fenômenos linguísticos correlacionados a esse parâmetro.

A hipótese que norteou esse trabalho foi a de que, em decorrência da perda do pronome *vós* e da diminuição de frequência de uso do pronome *tu*, que vem, paulatinamente, sendo substituído pelo pronome *você*, o paradigma flexional verbal do português brasileiro passou a apresentar um forte sincretismo, sobretudo no que diz respeito à segunda e terceira pessoas verbais, quadro que foi endossado pela gramaticalização do sintagma nominal *a gente*. A tabela a seguir, adaptada de Duarte (1995), descreve as mudanças sofridas pelo paradigma verbal do PB no decorrer do tempo:

Tabela 13: O paradigma verbal do PB no decorrer do tempo

| <b>Pessoa/número</b>         | <b>Pronome</b> | <b>Paradigma 1</b> | <b>Paradigma 2</b> | <b>Paradigma 3</b> |
|------------------------------|----------------|--------------------|--------------------|--------------------|
| <b>1ª pessoa do singular</b> | Eu             | Am o               | Am o               | Am o               |
| <b>2ª pessoa do singular</b> | Tu             | Am a s             | --                 | --                 |
|                              | Você           | Am a               | Am a               | Am a               |
| <b>3ª pessoa do singular</b> | Ele/Ela        | Am a               | Am a               | Am a               |
| <b>1ª pessoa do plural</b>   | Nós            | Am a mos           | Am a mos           | --                 |
|                              | A gente        | --                 | Am a               | Am a               |
| <b>2ª pessoa do plural</b>   | Vós            | Am a is            | --                 | --                 |
|                              | Vocês          | Am a m             | Am a m             | Am a m             |
| <b>3ª pessoa do plural</b>   | Eles/Elas      | Am a m             | Am a m             | Am a m             |

Fonte: DUARTE, 1995, p. 32.

Como podemos observar através da tabela, teríamos, então, no português brasileiro, um sistema verbal que evoluiu, com o passar do tempo, de seis para apenas três formas distintas, o que reduziria a frequência de uso do sujeito nulo, uma vez que um paradigma verbal sincrético necessita da diferenciação que seria realizada através do pronome sujeito expresso. Partindo dessa hipótese, a autora realiza uma análise que associa os modelos da Teoria de Princípios e Parâmetros à Sociolinguística Quantitativa, analisando uma amostra de fala que pertence ao projeto NURC e foi gravada no ano de 1992, sendo composta por doze informantes cariocas – seis do sexo masculino e seis do sexo feminino - com formação universitária, distribuídos em três faixas etárias: grupo 3 - 25-32 anos, grupo 2 - 45-53 anos e grupo 1 - 69-74 anos, gradação etária que caracteriza uma análise em tempo aparente.

Os possíveis fatores condicionantes internos à estrutura da língua considerados pela autora foram os seguintes:

1. Pessoa gramatical e desinência verbal;
2. Estrutura da oração e seu estatuto semântico no período;

3. A posição do referente do sujeito;
4. O traço semântico do referente do sujeito;
5. Duplo sujeito.

Os fatores extralinguísticos considerados como possivelmente relevantes foram:

1. Sexo;
2. Faixa etária;
3. Papel do indivíduo no grupo.

Dos 1.424 dados obtidos, Duarte obteve 29% de sujeitos nulos e 71% de sujeitos plenos. Apesar de muito significativos e condizentes com o esperado, esses números, isoladamente, dizem pouco sobre o estatuto do sujeito nulo e seus caminhos de mudança, o que só pode ser efetivamente analisado se considerarmos a forma como esse resultado está encaixado na estrutura da língua, sobretudo no que diz respeito à relação entre os resultados e a distribuição pelas diferentes faixas etárias. Portanto, comentaremos, a seguir, os resultados mais relevantes obtidos pela autora, com vistas a compreender o que há por trás desses percentuais.

Os resultados mostraram que a pessoa pronominal mais atingida pela mudança em direção aos sujeitos expressos foi a segunda, seguida, respectivamente, pela primeira e terceira, para as três faixas etárias, resultado que pode ser atribuído à inserção dos pronomes *você* e *a gente* no quadro pronominal do português brasileiro.

Em relação à primeira pessoa, Duarte obteve os resultados ilustrados na tabela a seguir:

Tabela 14: Uso do sujeito nulo de primeira pessoa em função da Faixa Etária

| <b>Pessoa</b>                | <b>Desinência</b> | <b>Grupo 1</b> | <b>Grupo 2</b> | <b>Grupo 3</b> | <b>Total</b> |
|------------------------------|-------------------|----------------|----------------|----------------|--------------|
|                              |                   | %              | %              | %              | %            |
| <b>1ª pessoa do singular</b> | -o                | 36             | 23             | 27             | 29           |
|                              | Zero              | 36             | 17             | 16             | 24           |
| <b>1ª pessoa do plural</b>   | -mos              | 29             | 29             | 0              | 28           |
|                              | Zero              | 0              | 0              | 5              | 9            |

Fonte: DUARTE, 1995, p. 52.

De posse desses resultados, pode-se fazer uma observação significativa: o grupo 1, composto pelos informantes mais velhos, não faz distinção entre presença e ausência de desinência verbal para aplicar o sujeito nulo à primeira pessoa do singular, característica que perde a força nas faixas etárias mais jovens. Esse dado pode ser interpretado como um indício para a implementação do sujeito nulo em função da presença/ausência da desinência distintiva, o que é uma forte característica das línguas *pro-drop*.

O caso da segunda pessoa é particularmente significativo: além da mais baixa frequência de uso do sujeito nulo, todas as ocorrências estão em contextos que favorecem a não realização do sujeito até mesmo em línguas não *pro*, como em orações encaixadas ou com possibilidade de retomada pragmática.

Conforme já esperado pela autora, os maiores índices de sujeito nulo estão na terceira pessoa, como mostra a tabela 15:

Tabela 15: Uso de sujeito nulo de terceira pessoa em função da Faixa Etária

|                              | <b>Grupo 1</b> | <b>Grupo 2</b> | <b>Grupo 3</b> | <b>Total</b> |
|------------------------------|----------------|----------------|----------------|--------------|
| <b>Pessoa</b>                | %              | %              | %              | %            |
| <b>3ª pessoa do singular</b> | 50             | 35             | 33             | 39           |
| <b>3ª pessoa do plural</b>   | 50             | 37             | 20             | 35           |

Fonte: DUARTE, 1995, p. 54.

Sobre os resultados relativos à terceira pessoa, a autora faz duas observações importantes:

1. A maior frequência de uso do sujeito nulo na terceira pessoa é um indício da eficiência de um referente externo reforçando esse sujeito não expresso foneticamente. Assim, um correferente externo, como um sintagma nominal, por exemplo, teria mais impacto para a não realização do sujeito do que a própria flexão verbal, nesses contextos;

2. Apesar da maior resistência, os sujeitos nulos de terceira pessoa também estão sofrendo queda em seu uso, o que pode ser comprovado pela forma como a frequência se distribui entre as faixas etárias.

Os resultados obtidos despertaram, em Duarte, a curiosidade de refinar o olhar para o comportamento dos sujeitos nulos de terceira pessoa. Na subseção seguinte, comentaremos a análise diacrônica realizada pela autora com vistas compreender melhor o comportamento dos sujeitos de terceira pessoa.

#### 4.4.1.2.1 Refinando o tratamento da 3ª pessoa (DUARTE, 2012)

Nessa segunda análise sobre o tema, Duarte, utilizando como amostra peças de teatro distribuídas em 7 períodos de tempo diferentes, compreendidos entre 1840 e 1990, busca compreender mais detidamente o comportamento diferenciado da realização do sujeito na terceira pessoa, no Português Brasileiro. Para tanto, a autora analisa dois aspectos específicos: o padrão sentencial, que diz respeito à

função do antecedente ao pronome de terceira pessoa e à sua função no período, e o traço semântico e referencialidade do antecedente do sujeito.

Em relação ao padrão sentencial, Duarte considerou 4 possibilidades:

Padrão 1: O antecedente se encontra no mesmo período e é o sujeito da oração precedente;

Padrão 2: O antecedente se encontra no período adjacente e tem a função de sujeito ou tópico estrutural;

Padrão 3: O antecedente é o sujeito de uma oração não adjacente no contexto precedente, havendo uma ou mais orações intervenientes;

Padrão 4: O antecedente está na oração precedente, mas tem uma função distinta da de sujeito.

Em relação ao traço semântico e à referencialidade do sujeito, foram consideradas as seguintes opções:

1. [+ humano] e [+ específico]
2. [+ humano] e [- específico]
3. [- humano] e [+ específico]
4. [- humano] e [- específico]

As rodadas foram realizadas em função da expressão plena do sujeito. A tabela a seguir apresenta os resultados obtidos em relação à frequência de realização do sujeito no decorrer dos diferentes períodos de tempo, em relação ao padrão sentencial e ao traço semântico:

Tabela 16: A realização do sujeito em função do tempo, do padrão sentencial e do traço semântico

| <b>Período de tempo</b>          | <b>Frequência</b> |
|----------------------------------|-------------------|
| <b>Período I (anos 1840)</b>     | 25%               |
| <b>Período II (anos 1880)</b>    | 32%               |
| <b>Período III (anos 1920)</b>   | 35%               |
| <b>Período IV (anos 1935)</b>    | 22%               |
| <b>Período V (anos 1955)</b>     | 37%               |
| <b>Período VI (anos 1975)</b>    | 50%               |
| <b>Período VII (anos 1990)</b>   | 58%               |
| <b>Padrão sentencial</b>         | <b>Frequência</b> |
| <b>Padrão 1</b>                  | 11%               |
| <b>Padrão 2</b>                  | 24%               |
| <b>Padrão 3</b>                  | 63%               |
| <b>Padrão 4</b>                  | 54%               |
| <b>Traço semântico</b>           | <b>Frequência</b> |
| <b>[+ humano] [+ específico]</b> | 43%               |
| <b>[+ humano] [- específico]</b> | 23%               |
| <b>[- humano] [+ específico]</b> | 12%               |
| <b>[- humano] [- específico]</b> | 4%                |

Fonte: DUARTE, 2012, p. 40.

Esse resultado, além de comprovar o crescimento do uso do sujeito realizado no decorrer do tempo, uma vez que a frequência passa de 25%, em meados dos anos 1800, a 58% nos anos de 1990, mostra que dois fatores de natureza estrutural e semântica são fundamentais para a realização do sujeito de terceira pessoa, sendo esses tão importantes quanto a avaliação morfossintática da presença/ausência de flexão verbal latente. Assim, a autora conclui que o traço [- específico] é o mais resistente ao uso do sujeito expresso, bem como contextos em que o referente se encontra espacialmente distante do sujeito.

Essas conclusões são muito importantes para que se compreenda o que, para além de questões relativas à morfologia dos verbos, permite que línguas estruturalmente diferentes licenciem o sujeito nulo. Assim, afirmamos que análises dessa natureza possibilitam a compreensão global do fenômeno que se encontra em variação e/ou processo de mudança, uma vez que, além de oferecerem as bases para que se compreendam quais fatores internos e externos à estrutura linguística estão sendo relevantes àquele quadro, permitem a compreensão do *status* da implementação da mudança no sistema linguístico, bem como o entendimento do funcionamento da mudança na marcação de um parâmetro linguístico.

#### 4.4.1.3 O parâmetro do sujeito nulo no português afro-brasileiro (LUCCHESI, 2009)

Chamando de português afro-brasileiro as variedades linguísticas faladas por moradores de três comunidades rurais do interior da Bahia – Helvécia, Cinzento, Barra/Bananal e Rio de Contas -, Dante Lucchesi, através do aporte teórico e metodológico da Sociolinguística Quantitativa, analisa o licenciamento do parâmetro *pro-drop* nessa variedade linguística, com vistas a verificar se comunidades linguísticas geograficamente isoladas e afetadas pelas peculiaridades causadas pelo contato massivo tendem a apresentar especificidades no que diz respeito ao sujeito pronominal.

O autor obteve 4.599 ocorrências em que o lugar do sujeito era ocupado por um pronome. Em linhas gerais, a análise apresentou resultados diferentes do que era esperado: tem-se, nessa variedade linguística, 27% (1.255 ocorrências) de uso do sujeito nulo, um percentual muito próximo ao obtido por Duarte (1995) em sua análise do fenômeno na norma culta. Entretanto, de acordo com o autor, essa proximidade entre o resultado da análise no português afro-brasileiro e o resultado da análise na norma culta brasileira pode ser explicado pela não exclusão, no *corpus*, de expressões fixas que favorecem o sujeito nulo e que foram desconsideradas por Duarte, tais como ‘não sei’, ‘sei lá’, entre outras. Ainda sobre esse resultado, o autor afirma que

Por outro lado, pode-se pensar, com base na visão das duas grandes vertentes evolutivas do PB proposta por Lucchesi (1994, 1998, 2001a, 2002b), que houve uma espécie de convergência na evolução gramatical dessas duas variedades linguísticas. Enquanto, na norma culta, a diminuição no nível de sujeito pronominal nulo derivou de

uma reorganização estrutural em função das substituições dos pronomes tu e nós por você e a gente, respectivamente; na norma vernácula, essa diminuição estaria ligada diretamente à queda de morfemas flexionais de pessoa e número do verbo, que não seria o resultado de uma reorganização da pauta dos pronomes pessoais sujeito, mas um reflexo direto do processo de transmissão linguística irregular que teria atuado mais intensamente na formação das variedades populares do PB (2009, p. 176-177).

Em relação ao encaixamento linguístico do fenômeno, comentaremos, a seguir, os resultados de algumas das variáveis que se mostraram estatisticamente relevantes para a análise de Lucchesi.

#### I) Pessoa do discurso

A pessoa do discurso que apresenta maior frequência de sujeito nulo é a terceira do singular, seguida pela primeira também do singular. Para Lucchesi, esse resultado pode ser explicado, no caso da terceira pessoa, pela possibilidade de retomada do referente do sujeito nulo e, no caso da primeira pessoa, pela presença do morfema flexional, que apresenta menos sincretismos em relação às demais pessoas. A tabela a seguir apresenta os resultados obtidos para todas as pessoas.

Tabela 17: Número de ocorrências em função da pessoa do discurso

| <b>Pessoa do discurso</b>    | <b>Nº de ocorrências/Total</b> | <b>Frequência</b> | <b>Peso relativo</b> |
|------------------------------|--------------------------------|-------------------|----------------------|
| <b>1ª pessoa do singular</b> | 659/2292                       | 29%               | .53                  |
| <b>2ª pessoa do singular</b> | 27/159                         | 17%               | .43                  |
| <b>3ª pessoa do singular</b> | 392/1167                       | 34%               | .55                  |
| <b>1ª pessoa do plural</b>   | 98/671                         | 15%               | .34                  |
| <b>2ª pessoa do plural</b>   | 6/35                           | 17%               | .43                  |
| <b>3ª pessoa do plural</b>   | 73/275                         | 27%               | .49                  |
| <b>Total</b>                 | 1255/4599                      | 27%               | --                   |

Fonte: LUCCHESI, 2009, p. 177.

#### II) Marca de pessoa e número na forma verbal

Baseando-se no princípio da saliência fônica, que consiste na maior propensão à marcação da pluralidade em formas verbais com maior distinção fônica, Lucchesi

analisa se há diferença significativa de uso do sujeito nulo de acordo com os morfemas flexionais. Os resultados obtidos estão descritos na tabela 18:

Tabela 18: Número de ocorrências em função da desinência verbal

| <b>Desinência verbal</b> | <b>Nº de ocorrências/Total</b> | <b>Frequência</b> | <b>Peso relativo</b> |
|--------------------------|--------------------------------|-------------------|----------------------|
| <b>-mos</b>              | 18/58                          | 31%               | .76                  |
| <b>-m</b>                | 20/63                          | 32%               | .57                  |
| <b>-o/ -el</b>           | 542/1844                       | 29%               | .51                  |
| <b>∅</b>                 | 675/2634                       | 26%               | .48                  |
| <b>Total</b>             | 1255/4599                      | 27%               | --                   |

Fonte: LUCCHESI, 2009, p. 178.

A análise dessa variável é particularmente interessante devido à possibilidade de avaliar a relação entre a realização do sujeito e a riqueza flexional dos verbos. Como mostram os dados da tabela, os morfemas mais salientes do paradigma verbal, *-mos* e *-m*, são os que apresentam maior frequência de sujeitos nulos (31 e 32%, respectivamente). Esse resultado, ainda que não muito expressivo, é um indício para a confirmação da correlação entre riqueza flexional e licenciamento *pro*.

### . III) Animacidade

Para essa variável, o autor considerou duas possibilidades: sujeitos com o traço [- animado] e sujeitos com o traço [+ animado]. A hipótese de que o traço [- animado] favorece o uso do pronome nulo se confirmou, conforme tabela a seguir:

Tabela 19: Número de ocorrências em função do traço semântico

| <b>Traço semântico</b> | <b>Nº de ocorrências/Total</b> | <b>Frequência</b> | <b>Peso relativo</b> |
|------------------------|--------------------------------|-------------------|----------------------|
| <b>[- animado]</b>     | 91/165                         | 63%               | .73                  |
| <b>[+ animado]</b>     | 374/1297                       | 29%               | .47                  |
| <b>Total</b>           | 465/1442                       | 32%               | --                   |

Fonte: LUCCHESI, 2009, p. 179.

Para analisar o encaixamento social do fenômeno, foram consideradas as variáveis sociais clássicas dos estudos sociolinguísticos de primeira onda: faixa

etária, sexo, escolaridade e estada fora da comunidade. Entretanto, nenhuma dessas variáveis mostraram resultados relevantes para a compreensão de como o fenômeno se encaixa na dinâmica social da comunidade. Para explicar esse resultado, o autor aponta as seguintes hipóteses:

1. A reação subjetiva dos falantes é uma possível influência, uma vez que o Parâmetro do Sujeito Nulo não sofre estigma social, ao contrário da concordância verbal, por exemplo. Assim, não existindo o estigma, não haveria impulsionamento em prol da implementação de uma ou outra forma;
2. A norma culta também não apresenta uso expressivo do sujeito nulo. Esse é um dado bastante relevante, considerando que essa é a norma que serve como modelo às classes menos prestigiadas, sobretudo quando essas passam a ter acesso mais amplo à escolarização formal.

#### *4.4.1.4 Interpretação dos sujeitos nulos no português de Angola (KAPETULA, 2016)*

A dissertação de mestrado de José Gueleka Kandjandja Kapetula, intitulada *Interpretação dos sujeitos nulos no português de Angola*, tem como principal objetivo verificar o funcionamento do Parâmetro do Sujeito Nulo nessa variedade de língua portuguesa, analisando se o português angolano se assemelha, em relação a esse parâmetro, ao português europeu, língua de sujeito nulo consistente, ou ao português brasileiro, que é considerada língua de sujeito nulo parcial. Para tanto, o autor, que escolheu como *corpus* apenas sujeitos de terceira pessoa do singular, selecionou dois grupos de informantes, sendo o primeiro composto por 35 adultos angolanos falantes de português como língua materna e o segundo composto por 21 estudantes portugueses falantes nativos de português europeu. Todos os informantes foram submetidos a um teste composto por quatro etapas, ou tarefas, nas palavras do autor:

- 1) Tarefa de aceitação de interpretações: Nessa tarefa, o autor teve como objetivo avaliar se há diferença na aceitação de sujeitos nulos e plenos de acordo com o caráter 'quantificado' ou 'não quantificado' do antecedente. Para tanto, os informantes foram apresentados a grupos de orações que deveriam ser avaliadas de acordo com a sua aceitabilidade. Essa tarefa oferecia quatro possibilidades (KAPETULA, 2016, p. 39):

i) Sujeito nulo encaixado com sujeito quantificado:

**Nenhum político** acha que **Ø** pode ser preso.

ii) Sujeito pronominal encaixado com antecedente sujeito quantificado:

**Todas as crianças** sabem que **elas** devem ir para a cama cedo.

iii) Sujeito nulo encaixado com antecedente sujeito não quantificado:

**Os alunos** acham que **Ø** vão ter uma boa nota.

iv) Sujeito pronominal encaixado com antecedente sujeito não quantificado:

**Os alunos** acham que **eles** vão ter boa nota.

2) Tarefa de leituras preferenciais: Os informantes devem apontar qual antecedente lhes parece mais aceitável como correferente de sujeitos nulos ou plenos em orações completivas e adverbiais, em contexto de ambiguidade. Desse modo, essa tarefa apresenta ao informante as seguintes possibilidades (KAPETULA, 2016, p. 40):

i) Sujeito nulo em completiva:

**O Rui** disse ao **Paulo** que **Ø** tinha ficado em primeiro lugar.

Pergunta: Quem tinha ficado em primeiro lugar?

Hipóteses de resposta: ( ) O Rui ( ) O Paulo

ii) Sujeito pronominal em completiva:

**O Pedro** perguntou ao **Zé** se **ele** tinha sido convidado.

Pergunta: Quem tinha sido convidado?

Hipóteses de resposta: ( ) O Zé ( ) O Pedro

iii) Sujeito nulo em adverbial:

**O avô** cumprimentou **o vizinho** quando **Ø** saiu de casa.

Pergunta: Quem é que saiu de casa?

Hipóteses de resposta: ( ) O Avô ( ) O vizinho

iv) Sujeito pronominal em adverbial:

**O Rui** fotografou o **Paulo** quando ele saiu de casa.

Pergunta: Quem é que saiu de casa?

Hipótese de resposta: ( ) O Paulo ( ) O Rui

3) Tarefa de interpretação: Nessa tarefa, buscou-se saber se, no caso de sujeitos nulos encaixados, há a aceitação de uma interpretação na qual o referente do sujeito nulo é idêntico a um sintagma presente em uma oração coordenada anterior. As perguntas apresentadas foram como as seguintes (KAPETULA, 2016, p. 41):

O Pedro disse que os pais estavam contentes e o Paulo disse que **estavam tristes**.

A expressão sublinhada refere-se:

- A. Aos pais do Pedro
- B. Aos pais do Paulo
- C. Aos pais de qualquer um deles

A Ana acha que o chefe é simpático e a Rita acha que é **antipático**.

A expressão sublinhada refere-se:

- A. Ao chefe de qualquer uma delas
- B. Ao chefe da Ana
- C. Ao chefe da Rita

4) Tarefa de juízos de gramaticalidade: Na quarta tarefa, os informantes fizeram a leitura de sujeitos nulos ou plenos em orações encaixadas, tendo como correferente o sujeito ou o objeto da oração matriz. Depois da leitura, os informantes avaliaram o nível de aceitabilidade de cada sentença. As possibilidades testadas foram:

i) Sujeito nulo encaixado correferente com objeto matriz:

**O Zé** perguntou à **Ana** se **Ø** estava grávida.

ii) Sujeito pronominal encaixado correferente com objeto matriz:

**O Rui** perguntou à **avó** se **ela** estava cansada.

iii) Sujeito pronominal encaixado correferente com sujeito matriz:

**A Ana** disse ao **Rui** que **ela** estava grávida.

iv) Sujeito nulo encaixado correferente com sujeito matriz:

**A avó** disse ao **neto** que **Ø** estava cansada.

De posse das respostas dadas pelos informantes aos testes, o autor comparou os dados obtidos através dos dois grupos para cada tarefa. Os dados obtidos foram os seguintes:

#### Tarefa 1

De acordo com o autor, os resultados esperados para essa tarefa são os descritos na tabela a seguir:

Tabela 20: Hipótese de resultado para a Tarefa 1

|                 |  | <b>PA</b> | <b>PE</b> |
|-----------------|--|-----------|-----------|
| <b>Tarefa 1</b> | Sujeito nulo com antecedente não quantificado    | -         | +         |
|                 | Sujeito nulo com antecedente quantificado        | -         | +         |
|                 | Sujeito expreso com antecedente não quantificado | Variável  | -         |
|                 | Sujeito expreso com antecedente quantificado     | Variável  | -         |

Fonte: Elaboração própria.

Os resultados obtidos por Kapetula estão dentro do esperado pelas suas hipóteses, conforme o disposto na tabela 21:

Tabela 21: Resultados da primeira tarefa

|   | PA     |    | PE    |    |
|---|--------|----|-------|----|
|   | Nº     | %  | Nº    | %  |
|   | 49/105 | 47 | 59/63 | 94 |
| <b>Tarefa 1</b>                                   |        |    |       |    |
| Sujeito nulo com antecedente não quantificado     | 64/140 | 46 | 67/84 | 80 |
| Sujeito expresso com antecedente não quantificado | 56/105 | 53 | 20/63 | 32 |
| Sujeito expresso com antecedente quantificado     | 58/140 | 41 | 19/84 | 23 |

Fonte: KAPETULA, 2016, p. 45.

Sobre esse resultado, o autor afirma haver importantes similaridades entre o comportamento do grupo de angolanos e o observado nas línguas de sujeito nulo consistente, nas quais, na maioria dos casos, o correferente do sujeito nulo, quando presente nas orações encaixadas, é interpretado como sendo o sujeito da oração matriz, o que não costuma ocorrer quando o sujeito da oração encaixada é expresso foneticamente. Apesar dessa tendência, os resultados apresentaram variação, ora se aproximando do comportamento observado no grupo de controle – o grupo dos falantes de português europeu – ora se afastando. Esse segundo grupo, por sua vez, apresenta características que confirmam a classificação do português europeu como língua de sujeito nulo, conforme estudos realizados anteriormente sobre essa variedade linguística, tais como o de Duarte: os sujeitos nulos são interpretados como ligados aos sujeitos das orações matrizes, o que quase não acontece em relação aos sujeitos expressos.

## Tarefa 2

Esse teste teve como objetivo avaliar a interpretação dos sujeitos nulos e expressos em contextos de ambiguidade em orações completivas e em orações adverbiais. Era esperado, tendo como base os resultados obtidos em trabalhos anteriores mencionados pelo autor, que o sujeito nulo retomasse com maior

frequência o sujeito matriz, enquanto o sujeito expresso retomaria o antecedente objeto. A tabela 22 descreve os resultados obtidos pelo autor:

Tabela 22: Resultados da segunda tarefa

|                 | PA      |    | PE      |    |
|-----------------|---------|----|---------|----|
|                 | Nº      | %  | Nº      | %  |
|                 | 146/210 | 70 | 85/126  | 67 |
|                 | 113/210 | 54 | 104/126 | 83 |
| <b>Tarefa 2</b> | 113/210 | 54 | 31/126  | 25 |
|                 | 99/210  | 47 | 35/126  | 28 |

Fonte: KAPETULA, 2016, p. 50.

Os resultados acima mostraram que o português angolano não apresenta um comportamento bem delineado para esse contexto, exceto nos casos das completivas com sujeito nulo, que retoma preferencialmente o antecedente sujeito da oração matriz. Nos demais casos, não há preferência demarcada. O caso do português europeu, por sua vez, é diferente: os pronomes nulos tendem claramente a retomar o sujeito da oração principal, enquanto os pronomes expressos comumente retomam o objeto. Esse resultado também está de acordo com o esperado pelo autor, uma vez que línguas de sujeito nulo costumam ser sensíveis à função sintática do correferente.

### Tarefa 3

Nessa tarefa, são testadas as leituras no que diz respeito à relação entre o sujeito nulo em oração encaixada e o DP antecedente. O autor buscou testar se o português angolano faz, nesses contextos, uma leitura estrita (*strict*), uma leitura consistente (*sloppy*) ou uma leitura mista. Os resultados obtidos foram os seguintes:

Tabela 23: Resultados da terceira tarefa

|                 |                    | PA     |      | PE     |    |
|-----------------|--------------------|--------|------|--------|----|
|                 |                    | Nº     | %    | Nº     | %  |
| <b>Tarefa 3</b> | <i>Strict</i>      | 83/210 | 39,5 | 81/126 | 64 |
|                 | <i>Sloppy</i>      | 62/210 | 29,5 | 12/126 | 10 |
|                 | Qualquer uma delas | 65/210 | 31   | 33/126 | 26 |

Fonte: KAPETULA, 2016, p. 55.

Os resultados obtidos para o português angolano apresentam bastante variação, embora haja uma leve preferência pelos contextos em que o referente do sujeito nulo da oração encaixada é interpretado como sendo o sujeito da oração matriz. O português europeu, por sua vez, interpreta o correferente desses sujeitos preferencialmente como o sujeito da oração principal, o que é compatível com as línguas de sujeito nulo consistente.

#### Tarefa 4

Nessa tarefa, os informantes avaliavam a aceitabilidade de sentenças com sujeito nulo e expresso encaixadas, tendo como correferente o sujeito ou o objeto presente na oração principal. Os resultados foram os seguintes:

Tabela 24: Resultados da quarta tarefa

|                 |                                    | PA    | PE    |
|-----------------|------------------------------------|-------|-------|
|                 |                                    | Média | Média |
| <b>Tarefa 4</b> | Nulo – correferente sujeito        | 3,3   | 3,5   |
|                 | Nulo – correferente objeto         | 3,1   | 3,0   |
|                 | Expresso – correferente<br>sujeito | 3,0   | 3,1   |
|                 | Expresso – correferente<br>objeto  | 3,0   | 3,75  |

Fonte: KAPETULA, 2016, p. 56.

Esses resultados consistem em uma média das notas atribuídas pelos participantes do teste a cada uma das sentenças apresentadas, sendo 1 as

sentenças inaceitáveis e 4 as sentenças totalmente aceitáveis. Como vimos, há certa coesão entre o comportamento dos dois grupos, sendo a maioria das sentenças consideradas aceitáveis.

Os resultados obtidos pelo autor mostram que, embora haja certa similaridade entre o comportamento dos falantes de português angolano e de português europeu, não é possível afirmar, com esses dados, que o português de Angola é uma língua de sujeito nulo consistente, uma vez que há constante variação nas referências linguísticas dos falantes. Desse modo, Kapetula destaca a necessidade de ampliar a investigação para outros contextos de fala, avaliando, também, o perfil sociolinguístico dos informantes e as condições de aplicação dos testes. De todo modo, o trabalho de Kapetula é de suma importância para o entendimento do funcionamento do Parâmetro do Sujeito Nulo nas diferentes variedades do português, uma vez que é um dos poucos que investiga o fenômeno do ponto de vista da avaliação, que é fundamental para a compreensão das tendências de mudança linguística.

#### 4.4.1.5 A expressão do sujeito no português de Moçambique (OLIVEIRA, 2016)

A dissertação de mestrado de Victor Oliveira, que investiga a variação na realização do sujeito no português de Moçambique, visa localizar o português moçambicano em um possível *continuum* afro-brasileiro, o que seria possível devido às similaridades observadas entre essa variedade e o português brasileiro. Assim, o autor busca avaliar se há, no português moçambicano, assim como há no português brasileiro, um movimento em direção aos sujeitos preferencialmente expressos.

O *corpus* da pesquisa, composto por 18 entrevistas concedidas por falantes do português moçambicano como L2, foi coletado no ano de 2007 em Maputo e estratificado de acordo com as variáveis sociais sexo, faixa etária e escolaridade. Ao contrário da maioria das pesquisas realizadas no âmbito do Parâmetro do Sujeito Nulo, o autor optou por inserir no *corpus* da análise, além dos sujeitos pronominais, as construções cujo sujeito é um sintagma nominal, como em (14):

(14) **Aquele homenzinho coitadinho** ficou na miséria.

Oliveira justifica sua escolha metodológica afirmando que, nesses contextos, o falante poderia ter optado por uma categoria vazia, uma vez que os referentes estão presentes no discurso. Se o informante opta por realizar foneticamente esse sintagma nominal sujeito, ele deve, então, ser contabilizado. Desse modo, o autor identificou as ocorrências de sujeito nulo e sujeito expresso, submetendo-as, depois de codificadas, ao programa estatístico GoldVarb X.

As variáveis linguísticas consideradas no estudo de Oliveira foram as seguintes:

- (i) realização fonológica do sujeito;
- (ii) categoria do sujeito;
- (iii) presença de elemento preposto ao verbo;
- (iv) tempo verbal;
- (v) regularidade morfológica do verbo;
- (vi) forma simples ou perifrástica do verbo;
- (vii) marca indicativa de número e pessoa;/
- (viii) concordância verbal;
- (ix) animacidade do sujeito;
- (x) (in)determinação referencial do sujeito;
- (xi) estatuto da oração quanto à sua (in)dependência sintática.

As variáveis sociais consideradas foram:

- (i) gênero;
- (ii) faixa etária;
- (iii) escolaridade;
- (iv) a forma como o português foi adquirido (em casa/ na escola).

As variáveis que foram consideradas pelo GoldVarb X estatisticamente relevantes para a ocorrência do sujeito nulo foram:

- (i) Morfologia verbal (marca número-pessoal);
- (ii) Traço [ $\pm$  animado];
- (iii) Regularidade morfológica do verbo (regular X irregular);
- (iv) Faixa etária;

(v) Tempo verbal.

A seguir, comentaremos, brevemente, os resultados encontrados por Oliveira para cada uma das variáveis selecionadas.

O autor obteve 58,4% de ocorrências de sujeito nulo, em contraposição a 41,6% de sujeitos plenos, números que colocam o português moçambicano em posição intermediária em relação ao português europeu e o português brasileiro no que diz respeito ao Parâmetro do Sujeito Nulo.

Em relação à morfologia verbal, primeira variável selecionada, os resultados obtidos pelo autor foram os seguintes:

Tabela 25: Resultados obtidos para a variável Morfologia verbal

| <b>Índice número-pessoal do verbo</b> | <b>Nº</b> | <b>%</b> | <b>Peso relativo</b> |
|---------------------------------------|-----------|----------|----------------------|
| <b>1ª pessoa do singular</b>          | 577/1053  | 54,8     | 0,48                 |
| <b>2ª pessoa do singular</b>          | 21/23     | 91,3     | 0,9                  |
| <b>3ª pessoa do singular</b>          | 425/789   | 53,9     | 0,41                 |
| <b>1ª pessoa do plural</b>            | 276/392   | 70,4     | 0,64                 |
| <b>3ª pessoa do plural</b>            | 215/336   | 49,5     | 0,56                 |
| <b>Total</b>                          | 2593      | --       | --                   |

Fonte: OLIVEIRA, 2016, p. 55.

Segundo o autor e de acordo com os dados obtidos por ele, o paradigma verbal do português moçambicano é semelhante ao do português europeu, inclusive no que diz respeito à ausência da flexão verbal de segunda pessoa do plural (vós). Entretanto, algumas diferenças foram observadas, tais como o uso do *você* com morfema de segunda pessoa do singular, por exemplo. A desinência *-mos*, de primeira pessoa do plural, é a mais numerosa e a que apresenta maior homogeneidade entre os informantes. Os dados também mostram que o sujeito nulo é fortemente favorecido pela segunda pessoa do singular, mas o pequeno número

de ocorrências de segunda pessoa impede que se tire conclusões mais aprofundadas sobre esse número.

A segunda variável selecionada foi Traço semântico do sujeito

Tabela 26: Resultados obtidos para a variável Traço semântico do sujeito

| <b>Traço do sujeito</b> | <b>Nº</b> | <b>%</b> | <b>Peso relativo</b> |
|-------------------------|-----------|----------|----------------------|
| <b>+ animado</b>        | 1349/2375 | 56,8     | 0,47                 |
| <b>- animado</b>        | 165/218   | 75,7     | 0,75                 |
| <b>Total</b>            | 2593      | --       | --                   |

Fonte: OLIVEIRA, 2016, p. 57.

No português moçambicano, o sujeito nulo é significativamente comum quando o referente de terceira pessoa é inanimado, característica também observada nas línguas de sujeito nulo canônicas. No caso específico do português europeu, por exemplo, os sujeitos nulos que têm um referente com traço [- animado] são mais numerosos até mesmo em contextos que desfavorecem o sujeito nulo.

A terceira variável selecionada foi a Regularidade morfológica do verbo:

Tabela 27: Resultados obtidos para a variável Regularidade morfológica do verbo

| <b>Tipo de verbo</b> | <b>Nº</b> | <b>%</b> | <b>Peso relativo</b> |
|----------------------|-----------|----------|----------------------|
| <b>Irregular</b>     | 963/1563  | 61,6     | 0,53                 |
| <b>Regular</b>       | 551/1030  | 53,5     | 0,44                 |
| <b>Total</b>         | 2593      | --       | --                   |

Fonte: OLIVEIRA, 2016, p. 58.

Apesar de essa variável ter apresentado resultados quantitativos relativamente próximos entre as suas variantes, o autor atribui a preferência de uso do sujeito nulo entre verbos irregulares a dois fatores: a saliência fônica, que se relaciona à diferença entre as formas verbais das seis pessoas gramaticais e a

identificação do sujeito, e o tipo de sujeito, uma vez que alguns verbos irregulares, como ser e estar, costumam aparecer em contextos que favorecem o sujeito nulo.

O próximo fator selecionado foi a variável social Faixa etária:

Tabela 28: Resultados obtidos para a variável Faixa etária

| <b>Idade</b>              | <b>Nº</b> | <b>%</b> | <b>Peso relativo</b> |
|---------------------------|-----------|----------|----------------------|
| <b>Até 35 anos</b>        | 303/532   | 57       | 0,47                 |
| <b>Entre 36 e 59 anos</b> | 840/1349  | 62,3     | 0,55                 |
| <b>Acima de 60 anos</b>   | 371/712   | 52,1     | 0,43                 |
| <b>Total</b>              | 2593      | --       | --                   |

Fonte: OLIVEIRA, 2016, p. 59.

Segundo o autor, não era esperado que o fator Faixa etária fosse selecionado, já que esse fator não se mostra estatisticamente relevante para o uso do sujeito nulo no português europeu. De todo modo, a faixa etária intermediária é a que mais utiliza o pronome sujeito não realizado devido ao fato de que essa faixa costuma estar mais susceptível à pressão pelo uso da norma, embora seja importante comentar que a realização do sujeito é um fenômeno linguístico que não está no nível de consciência dos falantes, não sendo, desse modo, alvo de estigmatização.

O quinto e último fator selecionado como relevante para a análise do sujeito nulo no português moçambicano foi o Tempo verbal, cujos resultados estão descritos na tabela 29:

Tabela 29: Resultados obtidos para a variável Tempo verbal

| <b>Tempo verbal</b>                    | <b>Nº</b> | <b>%</b> | <b>Peso relativo</b> |
|--|-----------|----------|----------------------|
| <b>Presente</b>                        | 1003/1734 | 57,8     | 0,47                 |
| <b>Pretérito perfeito</b>              | 294/463   | 63,5     | 0,59                 |
| <b>Pretérito imperfeito</b>            | 166/294   | 56,5     | 0,5                  |
| <b>Futuro simples</b>                  | 4/7       | 57,1     | 0,43                 |
| <b>Subjuntivo/ Futuro do pretérito</b> | 47/95     | 49,5     | 0,45                 |
| <b>Total</b>                           | 2593      | --       | --                   |

Fonte: OLIVEIRA, 2016, p. 60.

Conforme explicitado pelos dados da tabela, quase todos os tempos verbais favorecem o sujeito nulo, sendo o subjuntivo/ futuro do pretérito a única exceção. De acordo com Oliveira, esse resultado pode ser explicado pela pouca diferenciação entre as pessoas gramaticais no modo subjuntivo, o que faz com que seja necessária a realização fonética do sujeito para identificá-lo.

O trabalho de Oliveira é de suma relevância para a compreensão do funcionamento do Parâmetro do Sujeito Nulo em um contexto de aquisição de L2, sobretudo para compreender se há a aproximação entre o português moçambicano e a língua-alvo – o português europeu – ou se há a aproximação entre o português moçambicano e outras variedades de língua portuguesa, como o português brasileiro, por exemplo. Esses dados se mostram muito relevantes para a compreensão desse mesmo parâmetro no português dos Tongas, já que as circunstâncias de formação das duas variedades linguísticas são muito similares.

#### 4.4.1.6 *Null subjects in (Romance-derived) creoles: routes of evolution (LIPSKI, 1999)*

O estudo de John Lipski, um dos raros sobre o funcionamento do parâmetro *pro-drop* em línguas crioulas, é de fundamental importância para que se compreenda o funcionamento desse parâmetro em línguas nascidas a partir de situações de contato massivo. Ao comentar a ausência de flexão número-pessoal na

maioria das línguas crioulas, o autor afirma, associando o licenciamento do parâmetro *pro-drop* à riqueza flexional das línguas, que a maioria dos crioulos não licenciam sujeitos nulos referenciais, incluindo aqueles que derivam de línguas leixificadoras de sujeito nulo, tais como o espanhol e o português, pois não possuem flexão verbal de pessoa e número. A partir daí, o autor passa a apresentar e comentar características sintático-morfológicas dos crioulos que serviram de base para os seus estudos, as quais comentaremos brevemente a seguir.

a. Crioulo Filipino Espanhol (chabacano)

O crioulo filipino, ou chabacano, falado majoritariamente na cidade de Zamboagna, nas Filipinas, apresenta todas as características sintáticas das línguas crioulas no que diz respeito à flexão verbal, o que, entretanto, não impede o licenciamento do parâmetro *pro-drop* em alguns casos. O curioso é perceber que esse licenciamento não é observável nos padrões de licenciamento *pro-drop* das línguas de origem ibero-românica, e nem mesmo das línguas filipinas que serviram de substrato para a formação do chabacano. Os exemplos a seguir, listados pelo autor, mostram quais situações foram inalteradas na transição espanhol – chabacano, sendo, nesses casos, o comportamento do sujeito nulo, especificamente o sujeito nulo expletivo, uma herança da língua leixificadora.

(15) Noay pa hente na mundo.

(Ainda não havia pessoas no mundo.)

(16) Tiene muslim ke sabe chabacano?

(Tem muçulmano que sabe chabacano?)

(17) Tiene sabe, tiene no sabe.

(Tem que sabe, tem que não sabe.)

Contextos de encadearamento lógico-semântico também são favoráveis, no chabacano, ao uso do sujeito nulo, desde que o primeiro verbo tenha um sujeito explícito:

(18) Mama talya na bentana ta espera konmigo.

(Mamãe estava na janela [e ela estava] esperando por mim.)

(19) Ya lyama el rey kon el baw ya pregunta konele porké ele ta karga su kasa.

(O rei chamou a tartaruga [e ele] perguntou por que ele levou sua casa.)

(20) Kosa le ya ase ya sake su korta-pluma ya empesa pone aguhero na buli del mana olya.

(O que ele fez, [ele] tirou o canivete [e ele] começou a colocar buracos no fundo dos potes.

(21) Mana Hapón ya tene kambyo de korasón ya manda kanamon sale.

(Os japoneses tiveram uma mudança de coração e nos ordenaram a sair.)

Com esses exemplos, Lipski busca mostrar que os poucos contextos de licenciamento do sujeito nulo no chabacano são muito específicos e facilmente justificáveis. Além de serem permitidos apenas em casos de sujeitos expletivos, o uso do sujeito nulo está relacionado a contextos em que o discurso possibilita a retomada desse sujeito. Nas palavras do autor, comparando o crioulo filipino com o português e o espanhol,

Em espanhol e português, os falantes aceitam sem hesitação toda e qualquer frase contendo sujeitos pronominais nulos e, de fato, preferem pronomes nulos a pronomes expressos na maioria dos casos. No crioulo filipino espanhol, sentenças com sujeitos pronominais nulos nunca são aceitas isoladamente, pois a falta de AGR nessas línguas impossibilita a identificação do sujeito em orações simples. Por outro lado, quando apresentados exemplos nos quais os pronomes subjetivos foram omitidos seguindo um contexto bem definido que permite a identificação, a maioria dos falantes de crioulo filipino reconhece o uso de pronomes nulos como aceitável. Assim, esses exemplos não representam erros de desempenho, fenômenos de hesitação ou lapsos momentâneos e devem ser contabilizados dentro de modelos sintáticos de comportamento do pronome sujeito<sup>7</sup> (tradução nossa) (p. 6).

---

<sup>7</sup> In Spanish and Portuguese, speakers unhesitatingly accept any and all sentences containing null subject pronouns, and in fact prefer null pronouns to overt pronouns in most cases. In PCS, sentences with null subject pronouns are never accepted in isolation, since the lack of AGR in these languages makes identification of the subject impossible in simple clauses. On the other hand, when presented with examples like those in (3), in which subject pronouns have been omitted following a well-defined context which permits identification, most PCS speakers acknowledge the use of null subject pronouns as acceptable. Thus, examples like (3) do not represent performance errors, hesitation phenomena or momentary lapses, and must be accounted within syntactic models of subject pronoun behavior.

Outro ponto de fundamental importância na análise do crioulo filipino espanhol diz respeito aos contextos que licenciam o sujeito nulo referencial, sendo que esses contextos são, sintaticamente, muito específicos e diferentes do que é visto no filipino e do espanhol, sendo que somente essa última licencia o sujeito nulo. Assim, conforme afirma o autor, o comportamento do sujeito nulo no chabacano é uma simbiose entre o que é verificável em algumas variedades de segunda língua do espanhol e o cruzamento de algumas línguas filipinas. Essa tendência não é, para os estudos de línguas em contato, uma novidade: as línguas que nascem nessas situações tendem a apresentar características tanto da língua lexicadora quanto das línguas de substrato.

b. O crioulo de Maurício

O caso do crioulo de Maurício, falado na ilha africana de mesmo nome, é bastante diferente do caso do crioulo filipino em relação à gênese da língua. O crioulo de Maurício tem como língua lexicadora o francês, que é reconhecidamente uma língua negativa para o parâmetro *pro-drop*, apresentando características que não podem ser explicadas através da sintaxe das línguas de substrato (o malgaxe e algumas línguas Bantu), tais como, por exemplo, o sujeito nulo referencial licenciado apenas na presença de partículas tempo-modo-aspecto (TMA) pré-verbais nas respostas a perguntas, como podemos ver a seguir:

(22) To papa ki pe fer?

(O que seu pai está fazendo?)

Pe lave.

(Está lavando.)

Segundo o autor, em contextos como o mencionado no exemplo acima, a categoria vazia é governada pela partícula TMA, e o sujeito é retomável pragmaticamente, fatores que licenciam a não expressão do pronome sujeito. Desse modo, Lipski comenta a importância de se compreender o funcionamento do crioulo Maurício na análise do Parâmetro do Sujeito Nulo devido a essa peculiaridade no licenciamento do uso, o que, segundo ele, não é verificado em outros crioulos franceses do Oceano Índico. É importante comentar, então, que os dados trazidos

pelo autor mostram que a possibilidade de suprimir o sujeito no crioulo Maurício é, além de uma novidade, uma característica restrita apenas a alguns contextos discursivos, e de baixa frequência. Para ilustrar essa afirmação, o autor utiliza o inglês, língua reconhecidamente negativa para o parâmetro *pro*, como exemplo:

Sujeitos nulos limitados em inglês fornecem uma comparação útil com os dados do CM. No inglês coloquial (norte-americano), não é incomum ver mensagens postadas, telegramas e notas curtas, como "voltarei em uma hora", "ligarei na chegada", "já comi", [...]. Na maioria dos casos, o sujeito implícito é 'eu' ou 'nós', etc. Quando tais sentenças aparecem, por exemplo, em placas afixadas em portas, não é possível falar em 'vinculação do discurso', mas sim em pressupostos compartilhados pelo escritor e pelo leitor quanto à identificação do sujeito nulo. Uma característica fundamental dessas 'sentenças sem sujeito' em inglês é o fato de que são mais frequentemente escritas do que faladas, e que o sujeito compreendido é quase sempre {eu} ou {nós} (tradução nossa) (p. 15).

Assim, Lipski afirma acreditar que o crioulo Maurício não necessariamente desenvolveu uma sintaxe diferenciada para o sujeito pronominal, mas sim lançou mão de uma característica também encontrada no inglês, francês e outras línguas *pro-drop*: a possibilidade de omitir o sujeito em respostas simples e em contextos que permitam a retomada inequívoca da referência desse sujeito. Assim, apesar do seu relativo afastamento da sua língua leixificadora – o francês, o crioulo mauriciano não apresenta exemplos significativos que o permitam classificar como uma língua negativa para o parâmetro *pro*.

#### c. Sujeitos nulos em crioulos afro-hispânicos

Nessa seção, o autor trata de dois crioulos de base espanhola: o papiamento e o palenquero, afirmando que ambas apresentam, no que diz respeito ao licenciamento *pro-drop*, características semelhantes às do crioulo Maurício: enquanto no espanhol e no português são raras as construções que não permitem o sujeito nulo, o papiamento e o palenquero só licenciam o uso do sujeito não realizado quando o contexto permite a identificação inequívoca desse sujeito.

#### d. Um caso semelhante ao de Monte Café: Sujeitos nulos em um pidgin chinês de base espanhola

Apesar de se tratar reconhecidamente de um pidgin e não de um crioulo, o caso da língua dos culies precisa ser comentado, já que se assemelha bastante à língua dos Tongas. Os culies são trabalhadores contratados chineses levados para trabalhar nas plantações e usinas de açúcar em Cuba e no Peru, em situação análoga à escravidão, na segunda metade do século XIX. Assim, o aprendizado do espanhol por parte dos culies se deu de forma semelhante à aquisição dos trabalhadores contratados para as roças de Monte Café, uma vez que o modelo de língua disponível para esse aprendizado era o espanhol falado por alguns africanos que também trabalhavam nessas plantações – embora a convivência entre esses dois grupos, na maioria dos casos, fosse conflituosa. Ainda em relação à socio-história dessa população, é possível, segundo Lipski, que o crioulo português de Macau – de onde a maioria da população contratada vinha – e que outros pidgins e crioulos já presentes em Cuba tenham exercido influência na formação do pidgin sino-hispânico dos culies.

No que diz respeito ao licenciamento do sujeito nulo, há bastante diferença entre o pidgin dos culies, o espanhol e o cantonês – idioma nativo da maioria da população chinesa contratada. Embora o espanhol e o cantonês licenciem o uso do sujeito nulo, o fazem de formas bastante diferentes, sendo que não há, no cantonês, a relação entre preenchimento do sujeito e flexão verbal, estando a identificação do sujeito vinculada a antecedentes discursivos. O exemplo abaixo ilustra essa configuração:

(23) Tú, Malena, jabla mucho; no tlabaja, nenhum jase na; to día sentá la sillón.

(Você, Magdalena, fala demais; [você] não trabalha, [você] não faz nada, o dia todo [você] está sentada em uma cadeira.)

Os crioulos e o pidgin analisados por Lipski fornecem dados bastante interessantes em relação ao funcionamento do sujeito nulo nas línguas crioulas. Como pudemos observar nos exemplos supramencionados, o licenciamento do sujeito nulo está associado a contextos comunicativos bastante específicos e pouco vinculados a questões gramaticais, e sim ligadas à possibilidade de retomar o sujeito através do discurso, estando tais características presentes até mesmo em línguas crioulas que têm como lexificadoras línguas positivas para o parâmetro *pro-drop*.

Com isso, podemos afirmar que línguas formadas em contextos de contato massivo tendem a apresentar uma dinâmica que as afasta das línguas-alvo, apresentando características sintáticas particulares e bastante interessantes aos estudos linguísticos.

A discussão teórica e o feixe de pesquisas apresentadas nessa seção oferecem os fundamentos necessários para a análise que se segue. Acreditamos que os conceitos discutidos e os resultados obtidos através do estudo do Parâmetro do Sujeito Nulo nas mais diversas variedades da língua portuguesa, bem como a observação das características em relação a esse parâmetro em línguas crioulas oferecerão importantes meios de comparação para os resultados obtidos na análise do português dos Tongas.

## 5 A METODOLOGIA

Os procedimentos metodológicos são um dos primeiros pontos que devem ser definidos na realização de uma pesquisa científica, uma vez que a adequação desses procedimentos aos objetivos da investigação é o que garante ao pesquisador a fidedignidade dos seus resultados e de sua interpretação. No caso específico dos dados obtidos através de entrevistas sociolinguísticas, há desafios inerentes ao trabalho do pesquisador que só podem ser superados através de uma metodologia rígida que permita a observação sistemática dos fatos linguísticos, em consonância com o aporte teórico que conduz a pesquisa.

Desse modo, descreveremos, nesta seção, os procedimentos metodológicos adotados com vistas a alcançar os objetivos e confirmar ou refutar as hipóteses aqui levantadas. Na subseção 5.1, descreveremos o modo de constituição das amostras de fala das pesquisas sociolinguísticas, discutiremos os conceitos de *vernáculo* e *paradoxo do observador*, apontando possíveis soluções para mitigá-lo, e apresentaremos o perfil dos informantes cujas entrevistas serviram como *corpus* para este trabalho, com a caracterização das entrevistas. Em 5.2, apresentaremos o envelope da variação, listando as variantes que compõem a variável dependente aqui analisada e as variáveis linguísticas e extralinguísticas em torno das quais as nossas hipóteses foram construídas. Por último, em 5.3, serão descritos os passos que foram seguidos para a análise quantitativa dos resultados obtidos.

### 5.1 A CONSTITUIÇÃO DAS AMOSTRAS DE FALA

As amostras de fala são a essência da pesquisa sociolinguística, uma vez que a sua análise é o que possibilita o levantamento dos dados estatísticos necessários para a comprovação ou refutação das hipóteses sobre o fenômeno linguístico em questão. Desse modo, para que se construa uma amostra adequada ao que a análise propõe, é necessário seguir regras específicas de coleta de dados. Uma amostra de fala sociolinguística precisa ser estratificada de acordo com os dados sociais dos informantes, que são divididos em células sociais de acordo o que é

preconizado pela metodologia clássica da Sociolinguística de Primeira Onda (LABOV, 2008[1972]).

Sobre o número de informantes que devem compor uma amostra de fala sociolinguística, é importante afirmar que a necessidade de dispor de uma amostra numericamente significativa não implica dizer que é necessário um número muito grande de informantes, pois os usos linguísticos apresentam comportamento diferenciado em comparação a outros fatos sociais passíveis de análise estatística, o que garante uma certa homogeneidade do comportamento linguístico de uma comunidade de fala:

As pesquisas sociolinguísticas têm mostrado que não há necessidade de amostras tão grandes como as usadas em outras pesquisas de natureza social (de intenções de voto, por exemplo) para se analisar fenômenos variáveis, uma vez que o uso linguístico é mais homogêneo do que o comportamento humano acerca de outros fatos, em virtude de não estar tão sujeito à manipulação consciente (com a ressalva de que no caso dos estereótipos possa haver algum grau de manipulação consciente). Assim, por exemplo, numa dada comunidade, resultados obtidos a partir da análise da fala de cerca de 60 informantes serão apenas ratificados em amostras maiores, ou seja, a partir de certo número de informantes os resultados tornam-se redundantes (COELHO et al, 2015, p. 114).

Sendo assim, uma vez definida a comunidade de fala a ser analisada, os informantes que fornecerão os dados linguísticos para a pesquisa são escolhidos randomicamente, respeitando critérios sociais pré-definidos. Em seguida, procede-se, finalmente, à coleta dos dados.

O *corpus* do português dos Tongas, amostra de fala vernácula que compõe essa pesquisa, foi obtida por meio do acervo do Projeto Vertentes do Português Popular da Bahia<sup>8</sup> e é composta por entrevistas sociolinguísticas concedidas por pessoas nascidas na comunidade de Monte Café, na ilha de São Tomé. Os informantes, que apresentam pouca ou nenhuma escolaridade, foram estratificados em função das características sociodemográficas comuns aos estudos sociolinguísticos: *idade, sexo, escolaridade e estada fora da comunidade*. Foram consideradas três faixas etárias: a faixa I, composta por pessoas de 20 a 40 anos, a faixa II, formada por informantes de 41 a 60 anos, e a faixa III, com informantes

---

<sup>8</sup> <http://www.vertentes.ufba.br/>

acima de 60 anos. Em cada faixa etária há 3 homens e 3 mulheres, perfazendo 6 informantes por faixa e 18 informantes no total. A estratificação em relação à variável social *escolaridade* não foi possível, uma vez que quase todos os informantes da faixa III são analfabetos.

Todos os informantes são, em algum nível, falantes de uma variedade linguística que eles denominam *kimbundo*, mas, de acordo com Baxter (2002), é provável que essa referência mencionada seja, na realidade, o *umbundo*, já que eles afirmam que seus familiares vieram de regiões da África que falavam *umbundu* e não *kimbundu*, como Beguela, por exemplo.

Como a maior parte da documentação dessa roça foi destruída no processo de independência, imagina-se que o uso do termo *kimbundu* pode significar a presença de um bom número de *kimbundus* em algum período da história, ou pode indicar a generalização do nome pelos portugueses, ou ainda pode estar associado a algum prestígio do termo no passado. O fato é que havia uma certa preferência pela contratação de *umbundus* pelo fato de serem agricultores e que a fala africana dos informantes que compõem esse corpus se aproxima muito mais do *umbundu* do que do *kimbundu* (BAXTER, 2002, apud SANTANA, 2019, p. 63).

Desse modo, a história da roça de Monte Café e a descendência dos informantes foi de extrema importância para traçar seu perfil sociolinguístico, uma vez que precisamos considerar esses fatores para então construir e embasar as nossas hipóteses de trabalho. Durante séculos, diversas línguas tipologicamente diferentes, tais como variedades diferentes de português (como L1 e como L2) e línguas africanas, principalmente de Moçambique e o crioulo caboverdiano, estiveram presentes no contexto de Monte Café e ensejaram as características do português dos Tongas, o que faz com que seja necessário avaliar possíveis situações de bilinguismo e especificidades aquisicionais.

Constituir uma amostra de fala é um grande desafio, uma vez que o pesquisador que se propõe a essa tarefa se depara com barreiras comuns, sobretudo devido à necessidade de captação da fala natural: como obter a fala espontânea sendo o pesquisador um elemento estranho à comunidade, o que faz com que o informante altere seu comportamento? Desse modo, comentaremos, na seção a seguir, acerca dessas dificuldades, apresentando os aparatos metodológicos que permitem a sua mitigação.

### 5.1.1 O vernáculo e o paradoxo do observador

William Labov, ao inaugurar a linhagem de estudos no âmbito da Sociolinguística Variacionista, afirma que a forma adequada de se proceder em uma pesquisa sociolinguística é através da observação da língua em situações de interação social. Nessas situações, o falante tende a monitorar minimamente a sua fala, o que permite que o observador tenha acesso ao **vernáculo**: a língua natural, com o mínimo de monitoramento, comumente utilizada em situações informais de comunicação entre pessoas que se conhecem e mantém uma relação de pouca formalidade.

O vernáculo costuma ser mais facilmente acessado quando os falantes comentam assuntos que envolvem fortes emoções, uma vez que, nessas situações, aquele que narra os fatos passa a prestar mais atenção ao fato narrado em si do que à forma como está narrando. Desse modo, as entrevistas sociolinguísticas devem ser guiadas de forma a conduzir a emergência desses temas. Entretanto, nesse momento, surge um dilema: se interessa à sociolinguística somente a fala vernácula, como é possível aos falantes a manifestação desse vernáculo estando eles sendo observados e gravados?

O dilema causado pela inevitável interação entre o 'objeto a ser medido' e o 'instrumento de medição' não é uma realidade exclusiva dos estudos sociolinguísticos. O físico alemão Werner Heisenberg, por exemplo, ao formular o chamado Princípio da Incerteza de Heisenberg, ressalta que, devido à interferência do medidor e do instrumento de medição, há um limite para a observação dos sistemas quânticos, o que é uma característica própria da natureza da observação de fenômenos naturais:

Há quem afirme que não se pode localizar exatamente um elétron, porque ele não se encontra em um lugar determinado. (...) Existe um limite para os nossos poderes de observação e para o mínimo de perturbação que acompanha o nosso ato de observação, um limite inerente à natureza das coisas e que nunca pode ser vencido pelo aperfeiçoamento da técnica e da habilidade do observador (CARUSO; OGURI, 2006, p. 468).

Essas formulações, embora pertencentes à Física Quântica, descrevem perfeitamente o que a sociolinguística convencionou chamar de **paradoxo do**

**observador** (LABOV, 1972), que consiste na dificuldade de conciliar a presença do pesquisador – um elemento estranho à comunidade, e de seu gravador, e a coleta de dados adequados para a análise linguística, o que se torna muito difícil perante o natural estranhamento imposto pela situação da entrevista.

O objetivo da pesquisa linguística na comunidade deve ser descobrir como as pessoas falam quando não estão sendo sistematicamente observadas – no entanto, só podemos obter tais dados por meio da observação sistemática. O problema, evidentemente, não é insolúvel: ou achamos maneiras de suplementar as entrevistas formais com outros dados, ou mudamos a estrutura da situação da entrevista de um jeito ou de outro (LABOV, 2008[1972], p. 244).

Todas as atividades de pesquisa que envolvem a gravação da imagem e/ou da fala dos envolvidos precisam de consentimento, e essa exigência ética, apesar de absolutamente indispensável e justificada, implica em alguns problemas para o pesquisador que investiga fenômenos linguísticos: se o informante sabe que a sua forma de falar está sob análise, ele certamente aumentará o seu nível de monitoramento linguístico, o que conseqüentemente o afastará do vernáculo. Desse modo, cabe aos pesquisadores o desenvolvimento de soluções metodológicas capazes de superar a falta de naturalidade do contexto da entrevista sociolinguística.

Com o passar dos anos, a metodologia sociolinguística desenvolveu algumas técnicas que visam a mitigação do paradoxo do observador. A seguir, listaremos alguns métodos apontados por Labov (2008[1972], p. 246) com vistas a diminuir esse efeito:

1) Entrevistas com intervalos ou pausas: Com esse método, o pesquisador pode pausar as entrevistas e manter diálogos informais com o entrevistado, sendo que esses diálogos também serão gravados e farão parte do *corpus*, mas sem que o informante saiba.

2) Emoções fortes: Essa técnica visa envolver o entrevistado com a sua narrativa, a ponto de fazê-lo diminuir o monitoramento da sua fala devido à emoção causada pelos fatos narrados. Segundo Labov, narrativas relacionadas a risco de vida são as mais eficazes para essa técnica.

3) Conversas rápidas e anônimas: A partir de conversas rápidas sistemáticas realizadas em locais públicos, pode-se obter os dados de fala necessários, sem que seja necessário realizar entrevistas tradicionais.

4) Observações assistemáticas: É a realização de gravações em locais públicos, tais como trens, restaurantes e lanchonetes, com vistas a captar a fala natural das pessoas que compõem aquela comunidade.

5) Meios de comunicação: Nessa técnica, os dados são obtidos através de transmissões de meios de comunicação em massa, como rádio e televisão. Segundo Labov, é preferível que esses dados sejam obtidos através de falantes que estão sob impacto de forte emoção, como os entrevistados imediatamente após um desastre ou acidente, por exemplo.

Desse modo, nas entrevistas realizadas em Monte Café, os entrevistadores optaram por dirigir as perguntas para temas como a vida em família e em comunidade, o histórico das roças e a relação dos moradores com o lugar onde vivem, o que faz com que o informante se prenda à importância daquilo que está narrando e monitore minimamente a sua fala.

### **5.1.2 O perfil dos informantes**

Os dados dos informantes cujas entrevistas compõem a amostra utilizada nessa análise estão descritos na tabela a seguir. Todos os informantes possuem pouca ou nenhuma escolaridade, uma vez que nos interessa ter acesso a uma versão da língua que tenha sofrido pouca ou nenhuma influência da normatização imposta pela escolarização formal.

|   | Iniciais | Idade | Escolaridade      | Vê TV?       | Línguas que fala        |                              |              |              |              | Naturalidade | País   | Outras informações                               | Períodos fora de Monte Café |
|---|----------|-------|-------------------|--------------|-------------------------|------------------------------|--------------|--------------|--------------|--------------|--|--|-----------------------------|
|   |          |       |                   |              | Português               | “Kimbundu”                   | Caboverdiano | Forro        | “Moçambique” |              |  |  |                             |
| 1 | L. V.    | 30    | 6ª série primária | Às vezes     | +                       | +                            | -            | -            | -            | M.C          | Mãe: filha de pais tonga de avós angolanos (os dois lados). padrasto: tonga descendente de angolanos pai: Caboverdiano (Lúcia não morou com ele) | frequenta Igreja Evangélica (pastor brasileiro!) | -                           |
| 2 | J. F.    | 22    | 6ª série primária | Sim          | + dominante             | +                            | -            | - entende    | -            | M.C.         | pai tonga de M.C., de pais angolanos; mãe forra  | foi criado pela avó materna                      |                             |
| 3 | J. N.    | 29    | 7ª série primária | Sim (novela) | +                       | + (entende mais do que fala) | -            | -            | -            | M.C.         | pai e mãe tongas de M.C. ; avós (dois lados) angolanos; avó paterna de Benguela  | -  | -                           |
| 4 | F. J.    | 30    | 6ª série primária | Sim          | + (a partir dos 7 anos) | +                            | -            | -            | -            | M.C.         | pais tongas de M.C.; avós todos tongas de M.C.   | -  | fez tropa: 2,5 anos         |
| 5 | M. A.    | 29    | 4ª série          | Sim          | +                       | +                            | - entende    | -            | -            | M.C.         | filha de de Dona Isabel Nascimento pai: tonga de descendência angolana   | -  | -                           |
| 6 | F. A.    | 38    | 7ª série          | Sim          | +                       | +                            | -            | + (um pouco) | -            | M.C.         | pais tongas de M.C., avós angolanos; avós tongas de M.C. de descendência angolana  |  |                             |

|    |       |    |  |               |                         |             |              |                      |   |      |   |  |
|----|-------|----|--|---------------|-------------------------|-------------|--------------|----------------------|---|------|---|--|
| 7  | M. L. | 53 | Analfabeta   | Não           | +                       | + dominante | -            | -                    | - | M.C. | pais tongas de M.C.; avós angolanos (M+H os dois lados)   | -                                      |
| 8  | J. C. | 57 | Analfabeto   | Sim           | + a partir dos 7 anos   | + dominante | -            | -                    | - | M.C. | pai angolano; mão tonga de M.C., de pais angolanos  | -                                      |
| 9  | M. F. | 44 | 3ª série (escola para adultos depois da Independência) | Às vezes      | +                       | +           | + (um pouco) | -                    | - | M.C. | pais tongas de M.C. ; avós (dois lados) angolanos de Lobito   | -                                      |
| 10 | A.    | 54 | Analfabeto   | Sim (futebol) | +                       | + dominante | -            | -                    | - | M.C. | pais tongas de M.C. ; avós (dois lados) angolanos   | -                                      |
| 11 | I. S. | 49 | Analfabeta   | Não           | +                       | + dominante | + (um pouco) | -                    | - | M.C. | Pais angolanos  | -                                      |
| 12 | A. F. | 50 | 4ª série   | Sim           | + (a partir dos 7 anos) | + dominante | -            | - entende            | - | M.C. | pais tongas de M.C., avós angolanos (os dois lados)   | viveu quase 3 anos na ilha do Príncipe |
| 13 | R. S. | 89 | Analfabeta   | Não           | +                       | + dominante | -            | -                    | - | M.C. | Pais angolanos  | Passou 5 anos em Angola, já adulta     |
| 14 | P. F. | 68 | Analfabeto   | Não           | +                       | + dominante | -            | - (entende um pouco) | - | M.C. | pais tongas de M.C.; avós angolanos (M+H os dois lados)   | -                                      |
| 15 | I. N. | 67 | Analfabeta   | Assiste pouco | +                       | +           | - entende    | -                    | - | M.C. | pai tonga de desc moçambicana; avós paternos moçambicanos. mãe: tonga de desc. caboverdeana, mas falante de | -                                      |

|    |       |    |            |                          |   |             |   |   |   |      |  |  |   |
|----|-------|----|------------|--------------------------|---|-------------|---|---|---|------|--|--|---|
|    |       |    |            |                          |   |             |   |   |   |      | Kimbundu; -avós maternos caboverdeanos   |  |   |
| 16 | D. O. | 66 | Analfabeto | Sim                      | + | +           | - | - | - | M.C. | pai angolano de Huambo, mãe tonga de M.C.; avós (dois lados) todos de Angola         |  | - |
| 17 | C. C. | 91 | Analfabeta | Não                      | + | + dominante | - | - | - | M.C. | Pai e mãe angolanos  |  | - |
| 18 | A. J. | 70 | Analfabeto | Sim<br>(mas não entende) | + | + dominante | - | - | - | M.C. | pais tongas de M.C., de descendência angolana; avós (= H+M dos dois lados) angolanos |  |   |

As características socioculturais dos informantes são semelhantes, principalmente no que diz respeito à dificuldade de acesso ao mundo exterior à roça de Monte Café, uma vez que a cidade mais próxima fica a uma distância de aproximadamente 4 quilômetros. Além disso, o acesso à mídia também é bastante limitado, já que a única TV existente na sede da roça só é ligada à noite e apresenta programação bastante restrita (SANTANA, 2019, p. 63).

### 5.1.3 A natureza das entrevistas

As entrevistas que compõem o *corpus* deste trabalho foram colhidas entre os meses de janeiro e março de 1998, na roça de Monte Café, na ilha de São Tomé. Todas as entrevistas têm duração média de 60 minutos e foram conduzidas por membros do projeto Vertentes. A maioria dos entrevistados concederam as entrevistas em suas casas, o que contribui positivamente para a quebra da formalidade que a situação pode impor.

Conforme discutido em 5.1.1, as entrevistas sociolinguísticas têm como principal objetivo o acesso à fala espontânea do entrevistado, ou seja, o seu vernáculo. Desse modo, todas as entrevistas que compõem o *corpus* dessa pesquisa seguem uma estrutura que vai ao encontro das orientações labovianas no que diz respeito à tentativa de mitigar o paradoxo do observador e obter o vernáculo do informante. Assim, os temas relacionados à vida social na roça e à história de vida dos informantes, sobretudo em relação ao remoto período de trabalho nas lavouras, se mostraram bastante produtivos, bem como questões relacionadas aos costumes dos habitantes da roça e questões culturais da região. Sobre o eixo temático das entrevistas sociolinguísticas, Lucchesi (2009, p. 160) afirma que

Não obstante a produtividade de alguns temas, o leque de temas numa entrevista de tipo sociolinguístico é potencialmente ilimitado, não havendo em princípio tema ruim. O fator decisivo é a sensibilidade do pesquisador em capturar o estado psicológico do informante, tendo como apoio uma boa avaliação da realidade cultural da comunidade.

Desse modo, uma entrevista sociolinguística, além de fonte inesgotável de informações acerca dos dados de fala de determinada comunidade, se constitui, também, em uma importante fonte de acesso à sua história, uma vez que a

variedade de temas abordados durante a conversa podem fornecer dados históricos até então desconhecidos sobre a localidade em questão.

As entrevistas gravadas em Monte Café foram armazenadas em meio digital, com boa qualidade sonora, apesar das eventuais interrupções de indivíduos que, por vezes, se encontravam no ambiente. Posteriormente, essas entrevistas passaram por um processo de transcrição grafemática, cujos critérios estão de acordo com a Chave de Transcrição do Projeto Vertentes. Depois de gravadas, as entrevistas foram transcritas e as ocorrências de realização e não-realização do sujeito pronominal e foram identificadas, seguindo a cota de cinquenta ocorrências por entrevista. Em seguida, essas ocorrências foram codificadas de acordo com uma chave de codificação predefinida, que contém todas as variáveis linguísticas e sociais consideradas potencialmente relevantes, conforme demonstraremos na próxima seção.

## 5.2 O ENVELOPE DA VARIAÇÃO

Conforme já debatido nas seções anteriores, é sabido que os falantes têm à sua disposição várias formas diferentes de dizer a mesma coisa com o mesmo significado referencial, dentro do mesmo contexto. Essas diferentes formas – as variantes linguísticas -, juntas, formam o fenômeno variável. Dessa forma, nessa seção, apresentaremos a nossa **variável dependente** – a realização/não realização do sujeito pronominal, apresentando as variantes que compõem essa variável. Apresentaremos, também, as variáveis linguísticas e extralinguísticas que supostamente afetam a variável em estudo e são denominadas **variáveis independentes** ou **explanatórias**. Definido o fenômeno variável em foco, serão descritos os critérios que guiaram tanto o levantamento quanto a exclusão de ocorrências na amostra linguística analisada.

### 5.2.1 A variável dependente

A variável dependente – A realização/não realização pronominal do sujeito – é binária, ou seja, pode assumir duas possibilidades: a ausência de pronome sujeito foneticamente realizado ou a presença do pronome sujeito, compreendendo as ocorrências de todas as pessoas pronominais, tanto do singular quanto do plural.

Esse é um fenômeno bastante recorrente na língua, o que garante um número bastante expressivo de dados.

Foram consideradas ocorrências válidas sentenças com verbos em sua forma finita, precedidos por sujeitos pronominais realizados foneticamente ou não. Dessa forma, as ocorrências, grosso modo, apresentam o seguinte padrão:

- a) Sujeito pronominal realizado + verbo
  - (24) Mas **eu** não vejo isso como nada.
  - (25) **Ele** acompanhou o irmão dele.
  - (26) Conforme **eles** tavam pondo o dedo.
  
- b) Zero + verbo
  - (27) Sim, sim,  $\emptyset$  faço sim sem nenhum problema.
  - (28) Com seis ano,  $\emptyset$  começô trabaiare.
  - (29)  $\emptyset$  Creio que  $\emptyset$  nasceu,  $\emptyset$  nasceu aqui.

#### 5.2.1.1 Critérios para inclusão e exclusão de dados

a. Em sequências oracionais em que apenas o primeiro sujeito pronominal é foneticamente realizado, consideramos os zeros subsequentes como tendo o primeiro sujeito pronominal como referente.

- (30) **Eu** fui a escola com a idade de sete anos,  $\emptyset$  permaneci na escola quando  $\emptyset$  atingiu os doze ano  $\emptyset$  já tinha concluído, doze não, onze anos  $\emptyset$  tinha concluído a quarta série,  $\emptyset$  acho por impossibilidade dos pais  $\emptyset$  tive que parar.

b. Em contextos de repetição da sentença causada pela interrupção por participantes circunstanciais, consideraremos os dados como ocorrências diferentes, desde que as duas ocorrências apareçam em orações completas (31). Caso a repetição seja apenas do pronome, consideramos apenas o primeiro dado como uma ocorrência válida (32):

- (31) INF: **Ela** tava dizendo...  
CIRC: Boa tarde.  
INF: **Ela** tava dizendo...

(32) INF A quantidade **eu** não... (ocorrência desconsiderada)

CIRC: É vinte e quatro?

INF: A quantidade **eu** não lembro.

c. Sujeitos de frases consideradas expressões cristalizadas<sup>9</sup> foram desconsiderados:

(33) **Ø** Sei lá!

d. Ocorrências com verbos em formas não finitas não foram consideradas:

(34) **Eu** ir lá? Jamais.

e. Ocorrências com verbo implícito, ainda que seja possível recuperá-lo através do contexto, foram desconsideradas:

(35) **Ela** foi comigo... **Ele Ø** com o pai.

f. Ocorrências cuja audição não é completamente clara e, por isso, de acordo com o que orienta a chave de transcrição, estão grafadas entre colchetes, foram descartadas:

(36) [**Nós**] gostava muito daquela região.

g. Ocorrências com pronomes clivados ou pseudoclivados foram descartadas, uma vez que, nesses casos, de acordo com Duarte (1995), o pronome pleno é categórico:

(37) Foi **ele** que fez isso.

h. As ocorrências com verbos no imperativo foram descartadas, uma vez que favorecem, de forma categórica, a omissão do sujeito:

(38) **Ø** Vá pegar a cadeira!

Muitos são os fatores internos à estrutura linguística e externos a ela que podem influenciar a realização ou não do pronome sujeito. Descreveremos, a seguir,

---

<sup>9</sup> Entendemos, aqui, expressões cristalizadas como uma sequência de palavras fixa e consagrada por uma comunidade linguística, cujo sentido se dá através do contexto do diálogo e da idiomaticidade e não da relação entre as palavras que a compõem (CARNEIRO, 2016, p. 36).

as variáveis intralinguísticas e extralinguísticas em torno das quais giram as hipóteses levantadas.

### 5.2.1.2 As variáveis linguísticas

#### 5.2.1.2.1 Pessoa pronominal

Os estudos conduzidos por Duarte (1993, 1995) sobre o português brasileiro demonstraram que, na amostra sincrônica analisada, havia um movimento em direção à preferência pela realização do sujeito pronominal liderada, respectivamente, pela segunda e primeira pessoa pronominal, tendo a terceira pessoa apresentado um percentual mais elevado de sujeitos pronominais nulos. O estudo de Soares da Silva (2006), indo ao encontro ao de Duarte, apresenta sujeitos realizados com maior frequência para as primeiras e segundas pessoas, tendo as terceiras as mais altas de sujeito nulo, resultados que podem ser explicados através dos papéis das pessoas pronominais nas situações comunicativas.

Desse modo, através dessa variável linguística, pretende-se verificar a frequência de supressão dos sujeitos expressos em função da forma pronominal desses sujeitos, analisando se há, no português dos Tongas, comportamento semelhante ao observado no português brasileiro. As formas pronominais que compõem essa variável são as seguintes:

- a) Primeira pessoa do singular (eu/ Ø)  
(39) Mas **eu** não tem, meu parto bom já... não... Ø fico em casa. Ø Tem que corrê hospital.
- b) Segunda pessoa do singular (você/ Ø)  
(40) Qual trabalho que **você** andava trabaiá lá? Ø Trabaiava com ele?
- c) Terceira pessoa do singular (ele/ela/ Ø)  
(41) **Ela** também veio contratada e Ø voltô pra lá.
- d) Primeira pessoa do plural (a gente/ Ø)  
(42) Aí que **a gente** começô prendere, mas depois Ø não foi mais.
- e) Primeira pessoa do plural (nós/ Ø)

(43) **Nós** era três mulhé.

(44)  $\emptyset$  Não temos não.

f) Segunda pessoa do plural (vocês/  $\emptyset$ )

(45) **Vocês** falam que aqui pertence, mas  $\emptyset$  não sabem.

g) Terceira pessoa do plural (eles/elas/  $\emptyset$ )

(46) Então que **eles** vivia aqui. Depois de vas paraíso,  $\emptyset$  saíram para roça Amparo segundo...

#### 5.2.1.2.2 Desinência número-pessoal do verbo

Essa é uma variável extremamente importante, já que nos permite analisar a riqueza flexional dos verbos no português dos Tongas e a sua relação com o licenciamento do sujeito nulo. As hipóteses afirmam que um paradigma mais simplificado tende a favorecer a realização do sujeito pronominal, conforme afirma Chomsky (1981) ao explicar o parâmetro do sujeito nulo. Duarte (1995) afirma que o enfraquecimento do paradigma verbal, no caso do português brasileiro, motivado principalmente pelas mudanças ocorridas no quadro pronominal, seria o responsável pelo aumento da frequência de sujeitos realizados, uma vez que o sincretismo das formas verbais impede que a identificação do pronome sujeito seja feita através do paradigma verbal.

As desinências verbais que fizeram parte da análise foram as seguintes:

a. Morfema zero (- pessoa, - plural e formas homógrafas de 1ª e 3ª pessoa)

(47) Ele **matava** ni mato.

(48) Eu **disse** não.

b. Marcas de primeira pessoa do singular (+ pessoa, - plural)

(49) Eu **fui** pro escritório.

(50)  $\emptyset$  **Tenho** lá até aqui uma tia mais.

c. Marcas regulares de primeira pessoa do plural

(51) Putuguês é aqui, prunque nós **criamos**...

(52)  $\emptyset$  **Brincamos** muito bem mesmo.

d. Marcas de terceira pessoa do plural

(53) Eles **foram** viver noutra roça.

(54) **Ø São** de Angola.

#### 5.2.1.2.3 Caracterização semântica do sujeito

Os estudos mais importantes acerca da realização pronominal do sujeito têm apontado a caracterização semântica como uma das variáveis mais relevantes para o comportamento do fenômeno. O estudo realizado por Cyrino, Duarte e Kato (2000), por exemplo, comprova que, em línguas que licenciam o uso de variantes nulas e não nulas, a caracterização semântica do sujeito tem extrema importância para o uso de uma ou outra variante. De acordo com as autoras, os sujeitos com o traço semântico [+humano], representados pelas primeiras e segundas pessoas, apresentam as maiores frequências de realização fonética, enquanto os de terceira pessoa, por sua vez, se localizariam em um ponto mais baixo da escala de referencialidade, sendo foneticamente realizados com menor frequência.

Desse modo, ao analisar o traço semântico do referente do sujeito pronominal, buscamos verificar se também há, no português dos Tongas, a preferência por sujeitos realizados em contextos de referentes mais animados. Os níveis de referencialidade analisados foram os seguintes:

a. [+ animado/ + humano]

(55) Ahã, sim, **ela** fala o kimbundo. (referência: a mãe do informante)

b. [+ animado / – humano]

(56) Sim, sim, sim, **Ø** ambundam mais onde tem cabra. (referência: abelhas)

c. [- animado/ + concreto]

(57) **Ø** Cegou ele. (referência: uma faca)

d. [- animado/ + abstrato]

(58) Por acaso **ela** era boa. (referência: a criação de antigamente)

#### 5.2.1.2.4 Tipo de frase

O estudo de Duarte (1995) demonstrou que as orações interrogativas favorecem a não realização do pronome sujeito pronominal. Desse modo, motivados por esse resultado, pretendemos, nessa variável, verificar se o padrão sentencial também influencia a frequência de expressão fonética do sujeito pronominal no português dos Tongas. Estendemos essa análise para além do padrão interrogativo, conforme exemplificaremos a seguir:

a. Afirmativa

(59) **Eu** gosto muito pôco.

b. Negativa

(60) **Eles** não aceita isso.

c. Interrogativa

(61) Qual trabalho que **você** andava trabaiá lá fora?

d. Resposta a pergunta direta do tipo *yes/no question*

(62) DOC: O senhor costuma assistir TV?

INF: Ø Assisto.

e. Resposta a pergunta direta do tipo *yes/no question* com sim/não

(63) DOC.: Ahã, branco não gostava de forro não?

INF.: Não, Ø també nã entrava ni ni quintale não.

f. resposta a uma pergunta QU- (*Wh- Questions*)

(64) DOC: O que você fazia mais aqui?

INF: Ø Brincava muito.

g. Resposta epistêmica

(65) **Eu** acho que as coisas mudô muito.

#### 5.2.1.2.5 Paralelismo formal

O estudo realizado por Scherre e Naro (1993) apresenta o conceito de paralelismo formal, que consiste na tendência de o falante repetir as escolhas linguísticas feitas no nível discursivo, no nível sintagmático e no nível da palavra. Para os autores, quando o falante emprega uma determinada forma em seu

discurso, a tendência é que ele mantenha essa mesma forma, desde que não haja mudança no referente.

Desse modo, pretendemos verificar se esse princípio também se aplica à realização/não realização do sujeito pronominal, ou seja, analisaremos se há tendência de manutenção do zero (não realização) quando o sujeito da ocorrência imediatamente anterior também é nulo, e se há a tendência da manutenção do pronome realizado quando o da sentença anterior também é realizado. Lucchesi (2009) obteve, em sua análise sobre o português brasileiro, a confirmação do funcionamento do mecanismo do paralelismo formal, uma vez que o sujeito não realizado na oração anterior favoreceu a manutenção do sujeito nulo na oração subsequente. Há, também, o favorecimento da não realização do sujeito quando o correferente é um sintagma nominal. Os correferentes possíveis são:

a. Sintagma nominal

(66) **Meu pai** é de São Tomé. **Nasceu** em Monte Café.

b. Pronome

(67) **Eu** fiquei na cidade cinco anos só, depois **vi** embora pra roça.

c. Nulo

(68) **Tenho** lá até uma tia mas **eu** não sê vida dela.

d. Primeira menção (ocorrências que não estão ligadas a uma ocorrência anterior)

(69) DOC: Em Monte Café também. E começou trabalha quando?

INF: **Eu** comecei a trabalhar em mil novecentos e sessenta, mas na cidade capital.

#### 5.2.1.2.6 Localização em relação ao antecedente e as condições de referencialidade

O estudo de Duarte (1993) sobre o português europeu analisou a relação entre a função sintática da oração na qual o sujeito se encontra e a frequência de realização do sujeito pronominal. Os dados obtidos pela autora mostraram que, nas orações relativas, os sujeitos foneticamente realizados superam a frequência dos sujeitos nulos (70 e 61%, respectivamente). As orações sem encadeamento

sintático, por sua vez, favoreceram o uso do sujeito nulo, o que comprova a provável realização entre o tipo oracional e o fenômeno aqui analisado. Esse resultado nos motivou a incluir no nosso rol de variáveis estruturais o tipo de oração.

A correferencialidade é outro fenômeno que poderá ser analisado, paralelamente, através dos resultados obtidos nessa variável estrutural. Estudos importantes acerca do comportamento do sujeito pronominal associam a posição do sujeito em relação ao seu antecedente à frequência de realização e não realização do sujeito. Duarte (2003) afirma que um sistema com flexões verbais sincréticas só licencia o sujeito nulo em contextos com referentes sintaticamente acessíveis. Sobre essa questão, Paredes (2003, p. 105) afirma que "quanto mais estreita a conexão entre um referente/sujeito e sua menção prévia, menor a necessidade de explicitá-lo, seja por um pronome ou nome (no caso de terceira pessoa)", sendo essa relação uma realidade tanto da língua escrita quanto da língua falada. Assim sendo, antecedentes distantes tendem a favorecer o uso do sujeito pronominal. Essa hipótese poderá ser avaliada através da análise de orações cujo correferente esteja distante.

As variantes que compõem essa variável são as seguintes:

a. Sem encadeamento sintático

(70) DOC: E... as pessoas comem, a cobra aqui?

INF: Bom... **eu** nunca vi, assi... pessoa a comer

b. Primeira oração assindética

(71) **Eu** fui pa hospital, **eu** tinha que fazer corretagem.

(72) Eu toco viola, **eu** sou músico da igreja.

c. Segunda oração assindética

(73) Ele tinha uma mota, **ele** pegô na mota,...

d. Terceira oração assindética em diante

(74) **Ele** tinha uma mota, ele pegô na mota, **Ø** pôs na mota, ...

e. Primeira coordenada sindética

(75) **Eu** tive a tremer pensando que... meu pai, porque **eu** criei com padrasto que é meu pai...

f. Segunda coordenada sindética

(76) Eu saí cedo porque não gosto de me atrasar, mas mesmo assim não  $\emptyset$  consegui chegar a tempo.

g. Terceira coordenada sindética em diante

(77) Eu saí cedo porque não gosto de me atrasar, mas mesmo assim não consegui chegar a tempo  $\emptyset$  pois andei devagar

h. Oração principal posposta à subordinada em que está o antecedente

(78) Quando **eu** saí,  $\emptyset$  ingressei na igreja.

i. Oração adverbial com antecedente na oração principal

(79) **Ela** estava aqui quando  $\emptyset$  viu você.

j. Oração completiva com antecedente na oração principal

(80) **Eu** lembro que **eu** brincava muito ali.

k. Oração relativa com antecedente na oração principal

(81) Eu tinha ido ver **minha madrinha**, que  $\emptyset$  mora ali em cima.

l. Antecedente na fala do documentador

(82) DOC: Como é que foi que **eles** morreram?

INF:  $\emptyset$  Morreram muito súbito, muito súbito.

#### 5.2.1.2.7 Presença de constituinte à esquerda do verbo na oração

Analisamos, aqui, se há relação entre a presença ou ausência de constituintes anteriores ao verbo e se há relevância no comportamento dos sujeitos pronominais em relação aos tipos de constituintes, quando presentes.

a. ausência de constituinte

(83)  $\emptyset$  Mudei de ideia agora.

b. presença de constituinte adverbial

(84) **Na escola** a gente ia sempre.

c. presença de tópico

(85) **Ver filme**, eu gosto bastante.

d. presença de SN / SPrep com valor adverbial

(86) **Naquela casa** nós moramos por muito tempo.

e. presença de clítico

(87) Ela **se** machucou cortando cana.

f. presença de *não/nem/e não*

(88) A gente **nem** disse isso.

g. presença de palavra QU-

(89) **Onde** eles vão com tanta pressa?

### 5.2.1.3 *As variáveis sociais*

As variáveis sociais constituem a base dos estudos sociolinguísticos, uma vez que é a análise desses fatores que nos permite compreender a variação e as mudanças inerentes às línguas humanas como condicionadas não apenas por fatores internos ao sistema linguístico, mas também, e principalmente, por fatores que vão além da língua enquanto sistema. Por se tratar de produto e meio de interação social, não se deve analisar a língua de forma dissociada das questões sociais que a influenciam, pois tais questões são intrínsecas às línguas humanas e, portanto, fundamentais para a compreensão de seu funcionamento.

No âmbito dos estudos sociolinguísticos, a dimensão social da variação linguística é acessada através da estratificação dos informantes em categorias sociais específicas (LABOV, 2008[1972]). Nessa análise, foram consideradas três das variáveis sociais tradicionalmente analisadas nos estudos da Sociolinguística e nas análises acerca dos condicionamentos sociais da variação na realização do sujeito pronominal, bem como uma quarta que foi estabelecida no programa de pesquisa do Projeto Vertentes: Sexo/ gênero, Faixa etária, Escolaridade e Estado fora da comunidade. As hipóteses acerca de cada uma dessas variáveis serão apresentadas a seguir.

### 5.2.1.3.1 Sexo/ gênero

Apesar de ser uma categoria social quase unânime nos estudos sociolinguísticos, há uma longa discussão sobre a adequação no entendimento dos conceitos de sexo e gênero e de sua aplicabilidade à análise dos comportamentos linguísticos, uma vez que, enquanto o sexo se relaciona às características biológicas dos falantes, o gênero estaria ligado à sua identidade social, o que se aproxima mais dos interesses da Sociolinguística. Entendendo a impossibilidade de se prolongar nessa discussão neste trabalho, seguiremos, daqui em diante, a recomendação de Paiva (2004), passando a adotar, nas nossas categorizações, o termo sexo/gênero, por entender que a categoria sexo, no nosso estudo, abarca também o comportamento social de cada falante, o que, no nosso entendimento, inclui a noção de gênero.

Os estudos tradicionais da Sociolinguística de Primeira Onda, sobretudo os realizados nas décadas de 1960 e 1970 (LABOV, 1966, [2008]1972, entre outros), apontam como tendência do comportamento linguístico feminino o uso de variáveis linguísticas mais conservadoras e com maior prestígio social, mas essa tendência não pode ser avaliada de forma isolada. Os papéis sociais de homens e mulheres mudaram significativamente com o passar dos anos, de modo que a Sociolinguística, enquanto ciência que busca correlacionar a variação linguística às transformações que ocorrem na estrutura das sociedades, precisa acompanhar tais mudanças. Sobre isso, Freitag (2015, p. 18) afirma que

se a Sociolinguística tem como premissa, em tendência ampla, o estudo da relação entre língua e sociedade, precisa considerar que a sociedade muda; se a sociedade muda, as explicações do modelo teórico-metodológico deveriam, também, mudar: a explicação de as mulheres preferirem as formas padrão ou não estigmatizadas, por conta de seu papel como mães e educadoras, talvez fosse válida e pertinente nos anos de 1960; hoje, não se pode dizer que é este o papel das mulheres na sociedade.

Para ilustrar a correlação entre os diferentes comportamentos femininos e masculinos e a organização social da comunidade de fala analisada, Mollica e Braga (2003, p. 35) apresentam o estudo de Haeri (1987) sobre a língua árabe. Enquanto, nas comunidades de fala ocidentais, há indícios de maior conservadorismo linguístico entre falantes do sexo/gênero feminino, na variedade de língua árabe

analisada pelo autor há maior frequência de uso das formas mais prestigiadas entre os homens. As autoras apontam, também, que o cruzamento entre a variável sexo/gênero e outras variáveis sociais, como classe social e escolaridade, por exemplo, também deve ser considerado, uma vez que os papéis sociais de homens e mulheres se reorganizam em função de sua posição social, além de haver uma tendência apresentada pelas mulheres de menor resistência à ação normativa da escolarização tradicional.

Todas essas generalizações nos levam a afirmar que a variável sexo/gênero deve ser avaliada não apenas considerando eventuais papéis sociais que são esperados de uma mulher ou de um homem em determinada comunidade de fala devido à neutralização desses papéis nas sociedades atuais, sobretudo dentre os grupos mais jovens. Essas mudanças despertam a necessidade de se avaliar a variável sexo/gênero considerando sempre a estrutura social daquela comunidade de fala, principalmente no que diz respeito a sua estrutura social e às possíveis neutralizações decorrentes do cruzamento com outras variáveis sociais.

Outro ponto passível de atenção e cautela diz respeito às características do fenômeno que está sendo analisado aqui: a realização do sujeito pronominal. Testaremos a hipótese mais tradicional acerca da variável sexo/gênero – se existe, por parte das mulheres, maior tendência de uso da forma mais conservadora, o sujeito não realizado foneticamente, mas, considerando, além dos pontos discutidos acima, o fato de que a variável que está sendo analisada nesse trabalho não apresenta variantes socialmente estigmatizadas, espera-se que haja a neutralização dos resultados em relação à variável social sexo/gênero.

#### 5.2.1.3.2 Faixa etária

A faixa etária é uma das variáveis sociais mais tradicionais dos estudos no âmbito da Sociolinguística Variacionista, uma vez que ela permite que se perceba a dinâmica do fenômeno analisado na comunidade, inclusive no que diz respeito à estabilidade das variáveis envolvidas e às tendências de mudança, embora, assim como no caso da variável Sexo, estejam embricadas na variável Faixa etária algumas outras questões sociais, tais como redes sociais e escolaridade, por exemplo (FREITAG, 2005). Nas palavras de Freitag (2011, p 46),

A idade é uma das três supercategorias sociais nas sociedades industrializadas modernas, junto com a classe e o sexo, e seu atributo social é a correlação primária com a mudança linguística. Intuitivamente, percebemos a influência da idade nos processos de variação e mudança linguística: uso de uma expressão “fora de moda”, gírias desatualizadas, enfim, percebemos que o tempo passou e ainda guardamos traços daquela época em nosso repertório linguístico.

. A hipótese mais recorrente sobre a variável Idade é a de que as diferentes faixas etárias refletem diferentes estágios da mudança linguística na comunidade sob análise. Assim, a Sociolinguística Variacionista postula que as mudanças podem ser apreendidas no seu curso de implementação através do que se denominou análise em *tempo aparente*. Desse modo, o estudo da mudança em tempo aparente está baseado no pressuposto de que diferenças linguísticas entre gerações podem espelhar mudanças diacrônicas, ou seja, a diacronia através da qual as mudanças acontecem pode ser visualizada no recorte sincrônico do tempo aparente.

São várias as sugestões acerca dos recortes adequados para a análise da variável faixa etária nos estudos linguísticos. Nessa pesquisa, os informantes analisados estão estratificados em três faixas etárias: Faixa 1 - 20 a 40 anos, Faixa 2 - 41 a 60 anos e Faixa 3 - 60 anos ou mais. Nossa hipótese em relação à variável faixa etária é a de que, assim como no estudo realizado por Duarte (1995), os informantes mais jovens utilizem o sujeito pronominal foneticamente preenchido com maior frequência, uma vez que essa é a variante inovadora – embora a própria autora ressalte que essa variável, assim como a escolaridade, não apresentou resultados muito relevantes estatisticamente.

#### 5.2.1.3.3 Escolaridade

A variável social Escolaridade também é bastante relevante nas análises no âmbito da Sociolinguística Quantitativa, já que ela está diretamente relacionada à resistência às tendências de mudança linguística. Conforme afirmam Mollica e Braga (2003, p. 53), “a escola atua como preservadora de formas de prestígio. Veículo de familiarização com a literatura nacional, a escola incute gostos, normas, padrões estéticos e morais em face da conformidade de dizer e de escrever”. No entanto, nem todos os fenômenos linguísticos estão sob o controle normatizador da escola:

alguns deles, sobretudo aqueles que não envolvem estigmatização social, costumam não sofrer esse controle.

Por um lado, a escola controla, evita e pune, com veemência, o uso de formas com supressão e/ou troca de líquidas, como *framengo* e *pobrema*, e os fenômenos sintáticos com resquílios dos casos latinos nos pronomes, como *dá pra mim sair*. Por outro lado, é conivente com formas redundantes, do tipo *há anos atrás* (MOLLICA E BRAGA, 2003, p. 53).

Estudos que têm como objeto variedades populares de língua precisam dedicar atenção especial à variável Escolaridade, uma vez que sua análise, nesse contexto, é bastante complexa. Nesta pesquisa, os informantes são estratificados em analfabetos – pessoas com nenhuma escolaridade - e semianalfabetos – pessoas que tiveram, no máximo, quatro anos de acesso à escolarização, sendo que, na prática, há pouca diferença entre esses dois grupos em relação ao acesso à escolarização formal e, conseqüentemente, ao comportamento linguístico. No caso específico do fenômeno aqui analisado, a realização pronominal do sujeito, essa diferença é ainda menos sensível, uma vez que esse fenômeno é um dos que não costumam sofrer as sanções normativas da escola. Acreditamos, assim, que toda e qualquer mudança que possa ocorrer nesse fenômeno não esteja no nível de consciência dos falantes, embora exista a possibilidade de que aqueles que tiveram algum contato, ainda que mínimo, com a escolarização formal, apresentem tendência a utilizar a variante mais conservadora – o pronome nulo.

#### 5.2.1.3.4 Estada fora da comunidade

A variável Estada fora da comunidade avalia o quanto o mundo externo à comunidade de fala analisada influencia o comportamento linguístico dos falantes, no caso daqueles que se ausentaram por mais de seis meses do local onde vivem. Entretanto, o comportamento dessa variável se diferencia no contexto de Monte Café, uma vez que a maioria dos informantes nasceram e permaneceram por toda a vida no interior da roça, que funciona como uma comunidade fechada e isolada.

No *corpus* que compõe essa análise, dos 18 informantes analisados, apenas três estiveram fora de Monte Café por um período prolongado, o que desperta a necessidade de analisar de forma individual o comportamento linguístico desses

falantes, de forma a buscar similaridades ou diferenças na sua fala em relação aos demais.

### 5.3 PROCESSAMENTO QUANTITATIVO DOS DADOS

O suporte quantitativo é uma das principais ferramentas utilizadas pela Sociolinguística Variacionista para obter resultados confiáveis em relação ao entendimento da variação e mudança linguísticas. No estudo aqui apresentado, a ferramenta estatística escolhida foi o pacote de programas Varbrul (SANKOFF, 1988; PINTZUK, 1988), mais especificamente a sua versão mais recente, o Goldvarb X (SANKOFF, TAGLIAMONTE, SMITH, 2005), cuja metodologia de análise consiste em um método multivariado para a formulação de regras variáveis que visa, principalmente, “separar, quantificar e testar a significância dos efeitos de fatores contextuais em uma variável linguística” (GUY; ZILLES, 2007, p. 33-34). A análise multivariada é a mais adequada ao modelo de análise aqui desenvolvido, pois através dela é possível analisar, de forma simultânea, a influência e a significância de múltiplas variáveis independentes, sendo esse resultado apresentado através de pesos relativos. É possível, também, através desse método, avaliar se há influência entre as variáveis independentes (GUY, ZILLES, 2007).

O uso do método multivariado na análise linguística justifica-se pela natureza desse tipo de problema/fenômeno. A análise de fenômenos linguísticos revela a atuação concomitante de vários fatores de ordem linguística e extralinguística. Dessa forma, uma análise univariada, que trate separadamente essas influências, não comporta a complexidade e ação em cadeia dos fatores sobre o fenômeno em questão. A adoção de uma análise multivariada, por sua vez, é “mensurar” a influência mútua de fatores e “dará resultados mais precisos, porque, ao mesmo tempo em que computa o efeito de uma variável independente, ela controla explicitamente o efeito de todas as outras variáveis independentes conhecidas” (GUY, ZILLES, 2007, p. 34).

O Goldvarb X é capaz de processar um grande volume de dados linguísticos, o que o torna muito adequado a análises com grande quantidade de dados. Para proceder à análise, o programa foi utilizado de acordo com os passos explicitados a seguir.

Primeiramente, foi preciso gerar o arquivo de dados, no qual foi registrada a codificação de todos os dados. Nessa etapa, é fundamental estar atento à

linguagem exigida pelo programa, já que só serão lidas e quantificadas as linhas de códigos iniciadas por um parêntese aberto. Depois de exportar todas as linhas de codificação para a tela inicial do programa, aplicamos o comando *Generate Factor Specifications*, que tem como função apresentar todos os grupos de fatores (*Factor Specifications*) através de seus respectivos códigos pré-definidos na chave de codificação. Eventuais erros na codificação foram identificados e corrigidos nessa etapa. Depois de realizadas as correções, aplicamos o comando *Check Tokens* para que o programa lesse se a condição básica para a análise estava sendo atendida – a existência de, no mínimo, dois fatores por grupo.

O passo seguinte foi a criação do arquivo de condições através do comando *Tokens > No recode*. Nesse momento, foram realizadas as alterações de acordo com os objetivos da pesquisa, como amálgamas, eliminação de fatores, entre outros. Aplicamos, então, o comando *Cells > Load cells to memory* para criar o arquivo de células, através do qual foram gerados os resultados em valores percentuais. Nos casos em que alguma variante apresentou resultado categórico (0 ou 100%, os chamados *knockouts*), foi preciso realizar adaptações no arquivo de condições de modo e eliminar esses fatores, já que a análise aqui realizada considera apenas casos em que existe variação. O novo arquivo de condições gerou um novo arquivo de células sem *knockouts*, que foi submetido à análise através do comando *Cells > Binominal up and down*.

Esse é o ponto de maior importância da rodada dos dados. Com o comando *Cells > Binominal up and down*, o programa apresentou os dados que fundamentam esta análise, mostrando qual é a melhor rodada (*Best stepping up run*). Os grupos de fatores tidos como mais relevantes e seus respectivos pesos relativos foram retirados da melhor rodada, e esses resultados foram interpretados da seguinte forma, considerando uma escala de pesos relativos que vai de 0 a 1:

- Fatores que apresentaram peso relativo abaixo de 0,5 foram considerados como desfavorecedores da não realização pronominal do sujeito;
- Fatores com peso relativo igual a 0,5 foram considerados neutros em relação ao sujeito pronominal nulo;

- Fatores que apresentaram peso relativo acima de 0,5 foram interpretados como favorecedores para a aplicação da regra.

A interpretação adequada dos dados apresentados pelo Goldvarb X é o que fundamenta os resultados obtidos nessa análise, considerando que os números apresentados só ganham significado linguístico depois de devidamente interpretados. Segundo Mollica e Braga, as limitações do método estatístico aplicado à Sociolinguística Variacionista

são as do próprio linguista, a quem cabe a responsabilidade de descobrir quais são os fatores relevantes, de levantar e codificar os dados empíricos corretamente, e, sobretudo, de interpretar os resultados numéricos dentro de uma visão teórica da língua. O progresso da ciência linguística não está nos números em si, mas no que a análise dos números pode trazer para nosso entendimento das línguas humanas (2003, p. 25).

Desse modo, o entendimento do linguista e a pertinência das hipóteses levantadas para cada fenômeno linguístico são fundamentais para a interpretação dos dados estatísticos obtidos. O fato é que a presença ou ausência de determinadas variantes não é suficiente para a descrição de determinado fenômeno linguístico no bojo da comunidade de fala: é fundamental que se conheça as frequências relativas dessas variantes, sabendo interpretá-las corretamente (CHAMBERS, 2003).

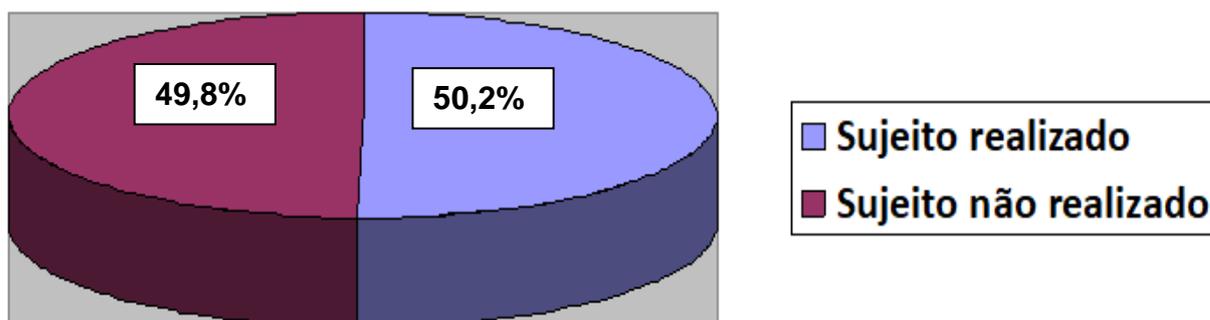
## 6 ANÁLISE VARIACIONISTA DA REALIZAÇÃO DO SUJEITO PRONOMINAL NO PORTUGUÊS DOS TONGAS

Apresentaremos, nessa seção, os resultados obtidos através da análise do fenômeno aqui discutido: a realização do sujeito pronominal no português dos Tongas. Além da frequência geral, apresentaremos a discussão acerca dos resultados apresentados tanto pelas variáveis selecionadas como estatisticamente relevantes pelo programa, apresentando seus respectivos pesos relativos, quanto pelas as variáveis não selecionadas, confrontando os dados obtidos com as nossas hipóteses e debatendo as possíveis motivações para a configuração em que o fenômeno se apresenta.

### 6.1 A REALIZAÇÃO/ NÃO REALIZAÇÃO DO SUJEITO PRONOMINAL NO PORTUGUÊS DOS TONGAS

Os resultados fornecidos pelo Goldvarb X mostram que, do total das ocorrências obtidas através da análise das 18 entrevistas (980 ocorrências), 50,2% dos dados encontrados (492 ocorrências) foram de sujeito foneticamente realizado, enquanto 49,8% dos dados (488 ocorrências) apresentaram sujeito não realizado foneticamente. O gráfico abaixo ilustra esse resultado:

Gráfico 1: A realização do sujeito pronominal no português dos Tongas



Fonte: Elaboração própria.

Os números obtidos, que se aproximam da neutralidade, se afastam, por um lado, dos encontrados por Duarte (1993, 1995) para o português europeu e para o português brasileiro, e se aproximam, por outro, dos resultados obtidos na análise

do parâmetro do sujeito nulo no português moçambicano, conforme ilustraremos na tabela a seguir, que apresenta R como sujeito realizado e N como sujeito não realizado:

Tabela 30: Expressão/ omissão do sujeito em algumas variedades do português

|          | Português dos Tongas | Português europeu (DUARTE, 1995) | Português brasileiro (DUARTE, 1995) | Português afro-brasileiro (LUCCHESI, 2009) | Português moçambicano (OLIVEIRA, 2016) |
|----------|----------------------|----------------------------------|-------------------------------------|--|--|
| <b>R</b> | 50,2%                | 36,9%                            | 71%                                 | 73%  | 41,6%                                  |
| <b>N</b> | 49,8%                | 63,1%                            | 29%                                 | 27%  | 58,4%                                  |

Fonte: Elaboração própria.

A tabela 30 apresenta dados importantes para que se compreendam os mecanismos de funcionamento do parâmetro do sujeito nulo em diferentes variedades da Língua Portuguesa, e é, sobretudo, a socio-história das línguas analisadas e as variáveis linguísticas envolvidas que fornecerão as bases para a interpretação dos resultados, conforme demonstraremos nas seções seguintes.

Os dados relativos ao português europeu foram obtidos através do estudo de Duarte (1995), que investigou, através de um *corpus* de entrevistas sociolinguísticas realizadas nos moldes propostos por Labov, os fatores linguísticos e sociais envolvidos na frequência de realização/ não realização do sujeito nulo na fala do português de Portugal. A amostra foi composta por 30 informantes distribuídos em três níveis de escolaridade – básico, médio ou superior -, e a autora obteve 2013 ocorrências de sujeito pronominal. Duarte observou que, na sua amostra, o sujeito nulo foi preferência em todas as pessoas pronominais, sendo as únicas exceções as orações relativas, as interrogativas diretas, as interrogativas *-qu*, e os sujeitos com referentes com traço [-animado]. Duarte afirma que é possível que a mudança rumo aos sujeitos expressos tenha se iniciado a partir desses pontos mais vulneráveis do sistema linguístico.

Em relação ao português brasileiro, a frequência apresentada, obtida por Duarte (1995) através da análise da fala culta, mostra um distanciamento em relação ao português dos Tongas, uma vez que, nessa variedade, 71% dos sujeitos

pronominais são foneticamente realizados, número bastante superior à variedade tonga. Essa frequência elevada é atribuída, de acordo com Duarte, ao sincretismo flexional ocasionado pela substituição quase total do pronome *tu* pelo pronome *você*, potencializado pela gramaticalização do sintagma nominal *a gente*, que passa a concorrer com o pronome canônico *nós*. Tendo em vista que esse sincretismo flexional também é uma realidade no paradigma verbal do português dos Tongas, temos o desafio de buscar compreender o que motiva a diferença entre as frequências encontradas para essas duas variedades linguísticas.

O resultado obtido por Lucchesi (2009) para o português afro-brasileiro, embora se aproxime dos números encontrados por Duarte para o português brasileiro culto, pode ser explicado por fatores diferentes:

Enquanto, na norma culta, a diminuição no nível de sujeito pronominal nulo derivou de uma reorganização estrutural em função das substituições dos pronomes *tu* e *nós* por *você* e *a gente*, respectivamente; na norma vernácula, essa diminuição estaria ligada diretamente à queda de morfemas flexionais de pessoa e número do verbo, que não seria o resultado de uma reorganização da pauta dos pronomes pessoais sujeito, mas um reflexo direto do processo de transmissão linguística irregular que teria atuado mais intensamente na formação das variedades populares do PB. (176-177).

Desse modo, pode-se concluir, com relação à comparação entre o português brasileiro e o português afro-brasileiro, que somente os dados numéricos não são suficientes para a compreensão dos mecanismos de variação das línguas. O fato é que o entendimento da história das línguas e da complexidade das variáveis envolvidas na variação é que darão as bases para a interpretação dos resultados numéricos obtidos. Desse modo, processos de variação e mudança linguísticas engatilhados por motivações diferentes podem convergir para uma mesma configuração paramétrica. É possível, então, que, assim como no português afro-brasileiro, o português dos Tongas apresente resultados que possam ser explicados como motivados mais pelas questões socio-históricas que ensejaram a formação dessa variedade linguística do que por questões puramente linguísticas/estruturais.

Por fim, a variedade linguística cujos resultados mais se aproximam do português Tonga é também uma variedade africana: o português de Moçambique. Nessa variedade, 58,4% das orações analisadas apresentam sujeito foneticamente

realizado, o que, inicialmente, pode indicar que há, para esse parâmetro, configuração semelhante em relação à variedade Tonga. Por outro lado, vemos que o português dos Tongas se encontra mais próxima do português brasileiro em comparação ao português moçambicano, já que, enquanto aquele exibe mais efeitos do contato entre línguas em comparação ao português de Moçambique, este apresenta um contingente populacional com mais acesso à educação, sofrendo, também, maior influência do Português Europeu. Daremos atenção especial a essa variedade na análise dos resultados, acreditando que similaridades no comportamento das variáveis linguísticas e sociais entre o português moçambicano e o português dos Tongas podem indicar trajetórias semelhantes entre essas variedades linguísticas, sendo esses processos motivados, para além das questões estruturais, pela socio-história dessas línguas.

### **6.1.1 O Parâmetro do Sujeito Nulo e a variável Pessoa do Sujeito: algumas considerações**

Uma das variáveis incluídas no nosso escopo de análise foi a Pessoa do sujeito, cuja observação teve como objetivo verificar como se comportam as diferentes pessoas em relação à realização ou não realização do sujeito pronominal. Embora essa variável não tenha sido selecionada pelo Goldvarb X como estatisticamente relevante, apresentaremos, aqui, a frequência de sujeitos expressos para cada pessoa pronominal, uma vez que esse dado é bastante relevante para a compreensão dos resultados que serão analisados posteriormente.

Tabela 31: A realização do sujeito pronominal em função da variável Pessoa do discurso

| <b>Pessoa do discurso</b>    | <b>Ocorrências/total</b> | <b>Percentual</b> |
|------------------------------|--------------------------|-------------------|
| <b>1ª pessoa do plural</b>   | 36/48                    | 75%               |
| <b>2ª pessoa do singular</b> | 12/21                    | 57,1%             |
| <b>1ª pessoa do singular</b> | 286/512                  | 55,8%             |
| <b>3ª pessoa do singular</b> | 152/328                  | 46,3%             |
| <b>3ª pessoa do plural</b>   | 16/70                    | 22,9%             |
| <b>Total</b>                 | 492/980                  | 50,2%             |

A tabela mostra que o sujeito preenchido é mais frequente com os pronomes de primeira pessoa do plural, com frequência de 75%. Esse dado contraria a hipótese inicial de Duarte (1995) de que os sujeitos nulos mais frequentes são os que acompanham formas verbais mais marcadas, como é o caso os verbos de primeira pessoa do plural. Nossa análise foge a esse padrão, e as hipóteses para justificar essa configuração são as seguintes:

1. Incluímos, no âmbito dos sujeitos de primeira pessoa do plural, o pronome *a gente*, em concorrência com o pronome canônico *nós*. O pronome *a gente* é tradicionalmente utilizado com formas verbais não marcadas e sincréticas, que são as correspondentes à terceira pessoa do singular. O sincretismo, nesse caso, motivaria a maior frequência de preenchimento do sujeito.
2. O sujeito canônico *nós* aparece em grande parte das ocorrências desacompanhado de sua forma verbal canônica (a desinência *-mos*), conforme apresentado na tabela X abaixo, o que é bastante comum em variedades populares de língua, utilizadas por pessoas menos escolarizadas. Desse modo, do ponto de vista gramatical, o contexto verbal sincrético em que esses pronomes foram preferencialmente aplicados justificaria a maior frequência da forma pronominal expressa.

Tabela 32: A primeira pessoa do plural e a forma verbal

| Forma verbal                |   | Nós          |      | A gente      |    |
|-----------------------------|---|--------------|------|--------------|----|
|                             |   | Ocorr./Total | %    | Ocorr./Total | %  |
| [- pessoa/ - plural] (zero) | R | 18/24        | 75   | 8/10         | 80 |
|                             | N | 6/24         | 25   | 2/10         | 20 |
| [+ pessoa/ + plural] (-mos) | R | 10/14        | 71,4 | --           | -- |
|                             | N | 4/14         | 28,6 | --           | -- |

Na tabela, vemos que, das 48 ocorrências de sujeito de primeira pessoa do plural, 34 foram utilizadas com formas verbais não marcadas. Das 38 ocorrências com a forma canônica *nós*, 24 foram utilizadas com formas verbais não marcadas, o que totaliza 63,2% das ocorrências com esse perfil. Esses números mostram que,

na variedade de português dos Tongas, está em inegável avanço o processo de sincretismo da morfologia verbal, o que pode estar favorecendo o alto número de pronomes expressos de primeira pessoa do plural. Duarte, ao encontrar resultados semelhantes para a primeira pessoa no português europeu, afirma que

A ocorrência equilibrada de sujeitos nulos e plenos de primeira pessoa não parece ser exclusiva do português europeu. Em estudo sobre o italiano oral, Ochs & Duranti (1979) já apontavam o fato de que o falante geralmente não se apresenta numa conversa usando apenas o mecanismo de concordância verbo-sujeito (que viria a ser popularizado como característico das línguas pro-drop); na maioria dos casos, um pronome pleno é utilizado (p. 9).

Depois dos pronomes de primeira pessoa do plural estão os sujeitos de segunda pessoa do singular, representados, aqui, apenas pelo pronome *você*, uma vez que não obtivemos ocorrências do pronome canônico *tu*. A alta frequência de uso da forma expressa desse pronome – 57,1% - também pode ser explicada, em primeira análise, pela sua alta frequência de uso com formas pronominais sincréticas: o pronome *você* compartilha formas verbais com praticamente todas as demais pessoas pronominais, sendo que todas as 21 ocorrências de *você* foram realizadas com formas verbais de traço [- pessoa, -plural] (90) (91) (92). Entretanto, os demais resultados indicam que não parece haver, no português dos Tongas, forte relação entre sujeito exposto e sincretismo da morfologia verbal, o que nos leva a acreditar não ser esse um fator preponderante para os resultados encontrados.

(90) Qual trabalho que você andava trabaiá lá fora?

(91) Você vai embora, vai acompanha sua família.

(92) você vai imbora ni sua terá. Vai ni pé di sua família qui tinha ficare aqui.

O pronome *eu*, por sua vez, embora seja canonicamente utilizado com formas verbais não sincréticas, e, por isso, tenda a ser preferencialmente nulo, apresentou frequência de 54,9% de realização fonética. Para compreender esse resultado, apresentaremos, abaixo, a tabela que mostra o uso da primeira pessoa do singular em função da desinência número-pessoal dos verbos com os quais essa forma pronominal foi utilizada:

Tabela 33: A primeira pessoa do singular e a forma verbal

| Forma verbal                  |   | 1ª pessoa do singular |    |
|-------------------------------|---|-----------------------|----|
|                               |   | Ocorr./Total          | %  |
| [- pessoa/ - plural] (zero)   | R | 165/312               | 53 |
|                               | N | 147/312               | 47 |
| [+ pessoa/ - plural] (-o/-ei) | R | 116/200               | 58 |
|                               | N | 84/200                | 42 |

Conforme aponta a tabela, das 512 ocorrências de primeira pessoa do singular como sujeito, 312 ocorrências (60,9%) foram utilizadas com formas verbais sincréticas. No entanto, as formas verbais específicas de primeira pessoa do singular [+ pessoa, - plural] também favoreceram o uso do sujeito expreso. A leitura desse resultado nos leva a afirmar que, no português dos Tongas, a relação entre realização pronominal do sujeito de primeira pessoa do singular e a riqueza da morfologia verbal podem não ser os fatores preponderantes para as altas frequências de sujeitos foneticamente realizados, mas sim outros fatores linguísticos, discursivos e sociais que serão explicitados e debatidos nas subseções seguintes.

O fato é que a relação entre a realização do sujeito e o paradigma flexional dos verbos pode ter sido uma importante porta de entrada para a mudança rumo à preferência pelos sujeitos foneticamente realizados, mas, posteriormente, outros fatores passaram a atuar de forma que o pronome expreso passou a ser preferencial para quase todas as pessoas verbais, independentemente da forma verbal que as acompanha. Ao passo que o sujeito preferencialmente realizado avança para todas as primeiras e segundas pessoas do discurso, o uso de verbos com desinências marcadas se torna cada vez menos frequente, o que certamente contribui para a aceleração desse avanço.

Voltando à tabela que apresenta a distribuição das pessoas do sujeito em relação à sua realização fonética, vemos que as terceiras pessoas apresentam um comportamento que vai de encontro ao das demais pessoas. Tanto a terceira pessoa do singular quanto a terceira pessoa do plural são as únicas pessoas do

discurso que desfavorecem o uso do sujeito expesso, apresentando frequência de 46,3% e 22,9%, respectivamente. O resultado é um forte indicativo de que, para a terceira pessoa, existem fatores discursivos mais relevantes que a flexão verbal para a realização fonética do sujeito, o que justifica a resistência da terceira pessoa ao sujeito realizado foneticamente. Sobre isso, e em relação ao português brasileiro, Duarte (1995) afirma que

Um predomínio muito significativo de sujeitos nulos de terceira pessoa (já atestado em outros trabalhos, entre os quais, Duarte (1993). Negrão & Müller (1994), Kato et alii (no prelo) sobre os de primeira e segunda confirmaria a ineficiência de um mecanismo de sujeito nulo centrado na flexão apenas e a resistência maior da terceira pessoa, que conta com o reforço da referência externa de um SN (DUARTE, 1995, p. 43-44).

Buscaremos, nos próximos passos, confirmar ou refutar as nossas hipóteses em relação às frequências de realização e não realização do sujeito encontradas no português dos Tongas para compreender o que há por trás dos números apresentados, qual a sua relação com outras variedades do português e quais são as variáveis linguísticas e sociais que contribuem para essa variação. Dessa forma, iremos, a seguir, analisar minuciosamente todos os fatores selecionados como estatisticamente relevantes pelo GoldVarb X. Em seguida, comentaremos, também, alguns pontos que julgamos relevantes acerca dos resultados apresentados pelas variáveis linguísticas não selecionadas: Tipo de frase e Localização em relação ao antecedente.

## 6.2 AS VARIÁVEIS SELECIONADAS PELO GOLDVARB X

Nessa subseção, apresentaremos os pesos relativos e percentuais encontrados para cada variável considerada relevante pelo o GoldVarb X. Nossas rodadas foram realizadas em função da variável R, ou seja, do preenchimento do sujeito pronominal como valor de aplicação da regra variável. As variáveis selecionadas como relevantes para o fenômeno foram as seguintes, na ordem em que aparecem:

1. Referência anterior ao sujeito;
2. Caracterização semântica do sujeito;
3. Faixa etária;

4. Presença de constituinte à esquerda da oração;
5. Desinência número-pessoal do verbo.

Pretendemos, sempre que julgarmos pertinente, comparar os nossos resultados com os encontrados em outros trabalhos sobre o tema mencionados anteriormente, como o de Duarte (1993, 1995, 2019), Lucchesi (2009), Oliveira (2016), dentre outros. Desse modo, poderemos avaliar possíveis rumos que a variedade de Língua Portuguesa dos Tongas pode estar tomando, apontando, assim, generalizações sobre os mecanismos que regem uma eventual mudança nesse sistema linguístico.

### **6.2.1 Referência anterior ao sujeito**

Essa variável, bastante recorrente nos estudos sociolinguísticos sobre o parâmetro *pro drop*, sistematiza o conceito de paralelismo formal, que consiste na tendência de manutenção de uma única marca em uma sequência de ocorrências. No caso da realização do sujeito, acredita-se que uma ocorrência de sujeito foneticamente realizado favorece a manutenção do sujeito expresso na oração imediatamente posterior, enquanto o sujeito não realizado é favorecido por uma ocorrência anterior de sujeito nulo.

As línguas que licenciam o sujeito nulo o fazem quando há, sintática ou discursivamente, um referente bem definido, o que, de acordo com a nossa hipótese, faz com que a frequência de sujeitos foneticamente realizados em orações iniciais/primeira menção seja maior, uma vez que não costuma haver, nessas ocorrências, um correferente próximo que licencie o sujeito nulo. Por outro lado, no caso dos sujeitos de terceira pessoa, é possível contar com a correferência forte de um sintagma nominal, o que faz com que, nesses contextos, o sujeito nulo seja preferencialmente utilizado (DUARTE, 1995).

Os resultados e discussões acerca da variável Referência anterior ao sujeito estão dispostos a seguir:

Tabela 34: Referência anterior ao sujeito

| Referência anterior ao<br>sujeito | Ocorrências/total | Percentual | Peso relativo |
|-----------------------------------|-------------------|------------|---------------|
| <b>Primeira menção</b>            | 241/344           | 70,1%      | 0.717         |
| <b>Pronome</b>                    | 156/270           | 57,8%      | 0.540         |
| <b>Sintagma nominal</b>           | 51/128            | 39,8%      | 0.457         |
| <b>Nulo</b>                       | 44/238            | 18,5%      | 0.193         |
| <b>Total</b>                      | 492/980           | 50,2%      | --            |

Significância: 0.004

Os dados acima confirmam em absoluto as hipóteses para essa variável. Em contextos em que o sujeito é mencionado pela primeira vez naquele turno de fala, há um favorecimento expressivo de uso do sujeito foneticamente realizado, sendo que, nesses casos, o sujeito realizado foneticamente totalizou 70,1% das ocorrências, apresentando peso relativo de 0.717. Esse resultado pode ser explicado, conforme comentamos acima, pela necessidade discursiva, em contextos iniciais, de primeira menção, de se especificar o sujeito para que não haja ambiguidades. Os exemplos a seguir, retirados do *corpus* aqui analisado, ilustram esse contexto:

(93) **Eu** colhia café, dipois, fazia desinfecção no coiso, nos cacueiros, nas cafieros com a máquina pulverizadora.

(94) Sim, sim, sim, tinha por acaso **nós** tinha o fardamento próprio.

(95) É, é, porque aqui memo essa história de viola aqui **eles** toca nota metade.

(96) É, **eu** trabalhava na roça mais ainda no stava diretamente integrado assim alistado.

Quando a ocorrência anterior é preenchida por um pronome expreso, há um favorecimento da continuidade do uso desse pronome foneticamente realizado, o que confirma o mecanismo do paralelismo formal. Nesses contextos, 57,8% dos sujeitos realizados foram precedidos por pronomes sujeitos também plenos, com o peso relativo de 0.540. As orações listadas abaixo exemplificam essas ocorrências:

(97) Se na altura **eu** tinha já quatorze anos **eu** fazia quando **eu** bem pretendia.

(98) O sacrifício que **eu** fazia no estudo, a noite hoje **eu** não consinto.

(99) A evolução **nós** temo tomamos a independência, **nós** tamos na democracia.

(100) Hoje quanto mais possibilidade **você** tiver pôr o seu filho a estudar memo que fôr até décima terceira não sê quê, ou pa um colégio ou um curso dentro da tua possibilidade **ocê** pode muito bem fazer.

O próximo resultado, relativo exclusivamente ao comportamento dos sujeitos de terceira pessoa, confirma a incidência do mecanismo de correferência externa ao discurso. Quando o correferente do sujeito é um sintagma nominal, há um desfavorecimento de uso do sujeito foneticamente realizado, que apresenta, nesse contexto, frequência de 39,8% e peso relativo de 0.457.

(101) Mas o ga, mais já o gato já é difícil cobra matá. É difícil a cobra matar porque não consegue atinge na pele.

(102) Meus filhos nasceram aqui, não falam kimbundo mas compreendem ainda.

(103) Meu tio fala com ele em quimbundo e explica, fala e escreve e dá pra ele ler.

Esse número confirma completamente a hipótese levantada para essa variável, uma vez que o sujeito exposto foi desfavorecido pela presença de um sintagma nominal como referente, já que a presença desse sintagma, que funciona como um correferente forte, dispensa o uso do sujeito foneticamente realizado. Uma proposta de pesquisa interessante seria a retomada dos sujeitos de terceira pessoa que apresentem sintagmas nominais como correferentes, com vistas a investigar os condicionamentos psicolinguísticos que motivam a supressão do sujeito em contextos específicos de retomada de SN.

O último fator que integra essa variável é o sujeito nulo como antecedente. As ocorrências com essa configuração seguem o padrão dos exemplos a seguir:

(104) Nasceram em monte café. Moraram aqui a vida toda.

(105) Falamos muito bem português. Conversamos tudo.

Como vimos, esse padrão sentencial também contribui para a confirmação da hipótese para essa variável, uma vez que, nesses contextos, o sujeito nulo foi retomado por um sujeito pleno em apenas 18,5% das ocorrências, com peso relativo de 0.193. Desse modo, podemos afirmar que, no português dos Tongas, o mecanismo do paralelismo formal é uma realidade, podendo ser corroborado pelos pesos relativos obtidos.

### 6.2.2 Caracterização semântica do sujeito

Diversos estudos acerca do funcionamento do Parâmetro do Sujeito Nulo têm demonstrado que, nos contextos em que o referente do sujeito apresenta o traço semântico [- animado], o sujeito preenchido foneticamente é preferencialmente utilizado, uma vez que esse traço é exclusivo das terceiras pessoas e, por isso, o referente pode ser retomado com facilidade através do discurso. O estudo de Duarte (2019), que compara o PE e o PB em relação ao funcionamento dessa e de algumas outras propriedades no âmbito do sujeito nulo, obteve resultados que confirmam esse mecanismo nas duas variedades linguísticas. Os resultados encontrados pela autora estão descritos na tabela a seguir:

Tabela 35: Sujeito de 3ª pessoa X feixe de traços semânticos (valor de aplicação: sujeito nulo)

| Traço                 | Português europeu |      |       | Português brasileiro |      |       |
|-----------------------|-------------------|------|-------|----------------------|------|-------|
|                       | Ocorr./Total      | %    | P.R.  | Ocorr./Total         | %    | P.R.  |
| <b>[- ani/ - esp]</b> | 12/12             | 100  | --    | 7/12                 | 58   | 0.863 |
| <b>[- ani/ + esp]</b> | 137/142           | 96,5 | 0.942 | 73/173               | 42   | 0.692 |
| <b>[+ ani/ - esp]</b> | 191/246           | 78   | 0.562 | 62/191               | 32,5 | 0.555 |
| <b>[+ ani/ + esp]</b> | 308/559           | 55   | 0.307 | 189/803              | 23,5 | 0.437 |

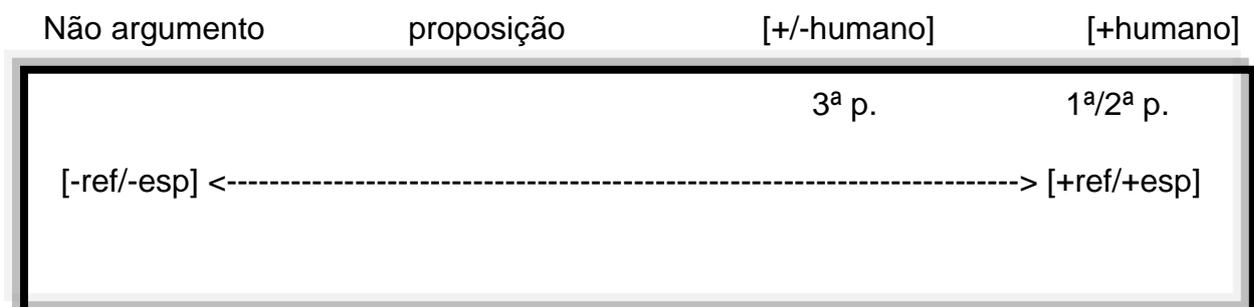
Fonte: DUARTE, 2019, p. 113

Como demonstram os dados da tabela, o traço semântico [– animado] é fortemente favorecedor do uso do sujeito nulo. No português europeu, esse traço, associado ao traço [- específico], favorece categoricamente o uso do sujeito nulo. De todos os contextos apresentados, o único que desfavorece, nas duas variedades analisadas, o uso do sujeito nulo, é o [+ animado/ + específico], que, na maioria das

ocorrências, representa correferentes humanos. Esse caráter dos sujeitos de terceira pessoa é, de acordo com Duarte, o motivo pelo qual esses sujeitos apresentam maior resistência à transição de sujeito nulo para sujeito pleno, o que não acontece com os sujeitos de primeira pessoa, que apresentam, invariavelmente, referentes humanos e específicos.

Essas diferenças de comportamento entre as pessoas do discurso levaram Cyrino, Duarte e Kato (2000) a propor uma sistematização representada por uma escala hierárquica, na qual estariam dispostos os argumentos verbais – tanto sujeitos nominais e pronominais quanto objetos – de acordo com o traço semântico dos seus referentes. De acordo com as autoras, o mecanismo de referencialidade está diretamente ligado à opção por variantes nulas ou não-nulas, sendo que quanto mais referencial for o sujeito, maior será a frequência de uso do sujeito expreso. A escala proposta é a seguinte:

Figura 6: Hierarquia de referencialidade



Fonte: CYRINO, DUARTE, KATO, 2000, p. 59.

Na escala, estão distribuídos no ponto hierárquico mais alto os sujeitos de primeira e segunda pessoa, seguidos pelos pronomes de terceira pessoa e pelos sujeitos proposicionais/oracionais. No ponto mais baixo da hierarquia se encontram os sujeitos expletivos, que apresentam o traço [-referencial/-específico].

Na análise aqui realizada, os resultados obtidos para o traço semântico do referente foram os seguintes:

Tabela 36: O sujeito pronominal realizado e a variável Traço semântico do referente

| Traço semântico         | Ocorrências/total | Percentual | Peso relativo |
|-------------------------|-------------------|------------|---------------|
| + animado/ - humano     | 54/94             | 57,4%      | 0.597         |
| + animado/ + humano     | 436/852           | 51,2%      | 0.519         |
| - animado/ + específico | 2/34              | 5,9%       | 0.049         |
| <b>Total</b>            | 492/980           | 50,2%      | --            |

Significância: 0.004

Nossa análise demonstrou que, no português dos Tongas, os traços [+humano] (106) e [-humano] (107) não apresentam diferenças significativas de frequência, estando, respectivamente, com 0.519 e 0.597 de peso relativo. O resultado obtido para esses traços vai, parcialmente, de encontro ao esperado, já que se esperava que o traço [+ humano] fosse mais favorecedor do sujeito expresso em comparação ao traço [- humano], conforme disposto na hierarquia da referencialidade de Cyrino, Kato e Duarte. Entretanto, para os sujeitos com o traço [-animado/ + específico] (108), o resultado foi exatamente o esperado: o uso do sujeito expresso foi totalmente desfavorecido, com peso relativo de 0.049, o que confirma a hipótese de que sujeitos menos referenciais tendem a ser representados por sujeitos nulos - em línguas *pro-drop* canônicas, inclusive, sujeitos com esses traços sequer são representados por pronomes pessoais (DUARTE, 1995).

(106) Eu parei, tendo em conta dificuldade da vida, e não continuei estudo.

(107) Mas cobra já é diferente, porque ela pica conforme eu expliquei aqui...

(108) O muro caiu, mas já tava perto de cair tinha muito tempo.

Para essa variável, eram previstos, também, sujeitos com referentes com o traço semântico [- animado/ - específico], mas não encontramos, no *corpus*, nenhuma ocorrência com esse perfil.

### 6.2.3 Faixa etária

A variável Faixa etária faz parte do escopo de análise da grande maioria dos trabalhos realizados no âmbito da Sociolinguística Variacionista. A distribuição do funcionamento das variantes linguísticas entre as diferentes faixas etárias é o que

permite que se avalie o andamento de um possível movimento de mudança linguística, uma vez que variantes utilizadas com menor frequência por faixas etárias mais jovens e com maior frequência entre os mais idosos apontam uma possível implementação de mudança rumo à extinção dessas variantes.

Para a análise variacionista da realização do sujeito no português dos Tongas, aplicamos a metodologia do tempo aparente (LABOV, 1972). Esperávamos, inicialmente, que a variante considerada conservadora – a não-realização fonética do sujeito pronominal - fosse utilizada com maior frequência pelos falantes mais idosos, enquanto o sujeito pronominal realizado, variante inovadora, fosse mais frequente dentre os mais jovens. É preciso comentar, entretanto, que o fator Faixa etária não costuma se mostrar relevante para esse fenômeno: na análise da norma não culta do Rio de Janeiro feita por Duarte (2003), por exemplo, a frequência de uso do sujeito expesso é superior a 75% para todas as faixas etárias, o que comprova que o sujeito expesso, para essa norma linguística, é indiscutivelmente preferencial, ao menos na fala, estando o processo de implementação da forma expressa em vias de finalização. Em Lucchesi (2009), essa variável não foi selecionada como estatisticamente relevante, e na análise de Oliveira (2016), a variável Faixa etária foi selecionada como relevante, sendo a faixa intermediária – de 36 a 59 anos – a que apresentou maior frequência de uso do sujeito nulo (peso relativo 0.55), seguida, respectivamente, pelos informantes mais idosos e pelos mais jovens, com peso relativo de 0.43 e 0.47, respectivamente.

Os resultados apresentados pela variável Faixa etária na análise da realização do sujeito no português dos Tongas estão dispostos na tabela a seguir:

Tabela 37: Realização do sujeito pronominal em função da variável Faixa etária

| <b>Faixa etária</b>              | <b>Ocorrências/total</b> | <b>Percentual</b> | <b>Peso relativo</b> |
|----------------------------------|--------------------------|-------------------|----------------------|
| <b>Faixa 1 (20 a 40 anos)</b>    | 197/345                  | 57,1%             | 0.602                |
| <b>Faixa 3 (60 anos ou mais)</b> | 177/315                  | 56,2%             | 0.530                |
| <b>Faixa 2 (41 a 60 anos)</b>    | 118/320                  | 36,9%             | 0.362                |
| <b>Total</b>                     | 492/980                  | 50,2%             | --                   |

Significância: 0.004

De acordo com os resultados apresentados, a faixa etária intermediária – a faixa 2 - é a que apresenta menor frequência de uso do sujeito foneticamente realizado, com percentual de 36,9% e peso relativo de 0.362, resultado que se aproxima ao encontrado por Oliveira para o português moçambicano. As faixas 1 e 3 favorecem o uso do sujeito expresso, com pesos relativos de 0.602 e 0.530, respectivamente.

Esse resultado, embora possa, inicialmente, parecer distante do esperado – as faixas etárias mais jovens à frente do processo de inserção da variante expressa no escopo da variedade linguística dos Tongas, e os mais velhos como o grupo que menos usa o sujeito expresso -, pode ser explicado através do comportamento social esperado para as diferentes faixas etárias: os falantes da faixa intermediária, comumente mais susceptíveis às pressões sociais pelo uso da norma por estar envolvidos em atividades onde se espera esse uso, como no trabalho, por exemplo, usam com maior frequência, ainda que inconscientemente, a variante mais próxima do padrão. Por outro lado, a quase neutralidade entre as frequências de uso das faixas 1 e 3 é um possível indicativo de que a mudança rumo à preferência pela forma expressa do sujeito pronominal está significativamente avançada na comunidade de fala analisada.

#### **6.2.4 Presença de constituinte à esquerda do verbo**

A variável Presença de constituinte à esquerda do verbo busca analisar se a presença ou ausência de elementos nessa posição influenciam na realização ou na não-realização do sujeito pronominal. Duarte (1995) avaliou, em sua análise, a influência da presença de clíticos, advérbios e da negação nessa posição. Para a autora, a presença desses elementos que, na oralidade e em sentenças de sujeito nulo, se cliticizam ao verbo, acabam, por uma questão de prosódia, por construir um contexto de intercambialidade com o sujeito expresso, sobretudo em orações principais com correferente nas encaixadas. Esse mecanismo já tinha sido identificado por ela em relação à preferência do sujeito nulo em orações com verbo no pretérito perfeito e por Callou & Silva (1995) em relação ao uso do artigo antes de pronomes possessivos. Para ilustrar a sua hipótese, a autora apresenta os exemplos a seguir:

(109) Por exemplo. nós éramos meninas: **não** éramos casadas. (p. 74)

(110) **Não** posso falar muito de horários porque eu sou super... Pontualidade não é o meu forte. (p. 74)

O paradigma abaixo, também elaborado pela autora, é apresentado por ela para comprovar o efeito prosódico existente entre um verbo com o sujeito exposto e esse mesmo verbo precedido por um advérbio, negação ou clítico:

Quadro 1: Efeito prosódico dos constituintes que antecedem o verbo

|          |             |           |            |
|----------|-------------|-----------|------------|
| Eu acho  | Não acho    | Ele disse | Só disse   |
| Eu quero | Só quero    | Ele fica  | Não fica   |
| Eu creio | Ainda creio | Ele foi   | Sempre foi |

Fonte: DUARTE, 1995, p. 74.

Dessa forma, ancorados a essas reflexões de Duarte, apresentamos os resultados obtidos para essa variável na análise do português dos Tongas, comentando, em seguida, cada um dos resultados encontrados.

Tabela 38: Realização do sujeito e a variável Presença de constituinte à esquerda do verbo

| <b>Tipo de constituinte</b>     | <b>Ocorrências/total</b> | <b>Percentual</b> | <b>Peso relativo</b> |
|---------------------------------|--------------------------|-------------------|----------------------|
| <b>Ausência de constituinte</b> | 399/769                  | 51,9%             | 0.535                |
| <b>Clítico</b>                  | 4/7                      | 57,1%             | 0.485                |
| <b>Constituinte adverbial</b>   | 19/43                    | 44,2%             | 0.439                |
| <b>Presença de não/nem</b>      | 68/142                   | 47,9%             | 0.406                |
| <b>Presença de palavra -QU</b>  | 2/19                     | 10,5%             | 0.106                |
| <b>Total</b>                    | 492/980                  | 50,2%             | --                   |

Significância: 0.004

Os nossos resultados confirmam por completo a hipótese de Duarte, o que, por sua vez, comprova a influência de elementos fonológicos, como a prosódia, para a configuração dos fenômenos morfossintáticos. A ausência de constituintes à esquerda do verbo (111) é a única configuração oracional que favorece a realização do sujeito exposto, com peso relativo de 0.535 e frequência de 51,9%. Por outro

lado, a presença de constituinte adverbial (112), de não ou nem (113), de palavras - QU (114) e de clíticos (115) desfavorecem o pronome foneticamente realizado, com pesos relativos de, respectivamente, 0.439, 0.406, 0.106 e 0.485. Os exemplos elencados abaixo ilustram cada uma dessas construções:

(111) Porque ele tinha muita paciência de aprender essa língua.

(112) **Já** sabe como porco é onguo.

(113) Eles nasceram aqui mesmo em Monte Café. **Não** vieram de Angola.

(114) Permaneci na escola, **quando** atingiu os doze ano já tinha concluído, doze não, onze anos tinha concluído a quarta série.

(115) Ele viu, assim que **me** disse.

#### 6.2.5 DESINÊNCIA NÚMERO-PESSOAL DO VERBO

Em torno da desinência verbal giram muitas hipóteses acerca do licenciamento do sujeito nulo. Estudos no âmbito do português brasileiro, sobretudo do português popular, vêm demonstrando que o forte sincretismo da morfologia verbal ocasionou a necessidade de se identificar os sujeitos através da sua expressão fonética, o que desencadeou uma mudança rumo aos sujeitos preferencialmente realizados. No caso do português brasileiro, a extensão dessa mudança atingiu todas as formas e pessoas verbais, e no português afro-brasileiro o fenômeno já é atestado até mesmo nas formas verbais mais marcadas, conforme apresentado por Lucchesi (2009):

Tabela 39: Uso do sujeito nulo X Pessoa do discurso e desinência verbal

| <b>Pessoa do discurso/ Desinência verbal</b> | <b>Ocorr./Total</b> | <b>%</b> |    |
|--|---------------------|----------|----|
| <b>1ª pessoa do singular</b>                 | -o/ -ei             | 542/1844 | 29 |
|  | ∅                   | 117/448  | 26 |
| <b>1ª pessoa do plural</b>                   | -mos                | 18/58    | 31 |
|  | ∅                   | 80/613   | 13 |
| <b>3ª pessoa do plural</b>                   | -m                  | 20/56    | 36 |
|  | ∅                   | 53/219   | 24 |

Fonte: LUCCHESI, 2009, p. 179.

Assim, buscamos analisar essa variável para verificar se a morfologia verbal é, no português da Tongas, realmente relevante para a realização do sujeito pronominal. Para tanto, agrupamos as desinências verbais de acordo com as possíveis combinações apresentadas por Galves (1990), que associou o traço pessoa ao traço plural. Desse modo, temos o seguinte esquema para os paradigmas verbais:

Tabela 40: Paradigmas verbais

| <b>Pessoa/número</b>             | <b>Aplicação</b>  |
|----------------------------------|---|
| <b>- pessoa, - plural (zero)</b> | Até 4 pessoas do discurso:<br>Primeira pessoa do singular no imperfeito do indicativo:<br>(116) Eu gostava também de falar essa língua.<br>Segunda pessoa do singular (você):<br>(117) Você viu que foi grande o movimento aqui.<br>Terceira pessoa do singular:<br>(118) Ele sabe que aqui é muita gente.<br>A gente:<br>(119) Aqui a gente gosta muito de ir nessas festas. |
| <b>- pessoa, + plural</b>        | Duas pessoas do discurso:<br>Terceira pessoa do plural<br>(120) Eles vieram morar aqui quando eram crianças.<br>Segunda pessoa do singular (vocês)<br>(121) Vocês viram o carro passar?   |
| <b>+ pessoa, - plural</b>        | Primeira pessoa do singular:<br>(122) Eu comprei muitos alimentos.<br>(123) Eu acho que sim, mas não tenho muita certeza.   |
| <b>+ pessoa, + plural</b>        | Primeira pessoa do plural:<br>(124) Nós viemos morar aqui já faz muito tempo.   |

Fonte: OLIVEIRA, 2020, p. 64.

O quadro flexional acima, baseado em dados do português brasileiro, mostra que essa variedade de língua apresenta paradigma que diferencia apenas quatro pessoas verbais, apresentando muitos sincretismos. Buscamos, desse modo, verificar se, no português dos Tongas, as formas mais sincréticas se apresentam como contextos favorecedores do sujeito expresso, e se as formas exclusivas são favorecedoras do sujeito nulo.

Os resultados da variável Desinência número-pessoal do verbo estão dispostos na tabela a seguir:

Tabela 41: O sujeito expresso e a variável Desinência número-pessoal

| <b>Desinência número-pessoal</b>   | <b>Ocorrências/total</b> | <b>Percentual</b> | <b>Peso relativo</b> |
|------------------------------------|--------------------------|-------------------|----------------------|
| <b>+ pessoa/ + plural (-mos)</b>   | 9/14                     | 64,3%             | 0.648                |
| <b>+ pessoa/ - plural (-o/-ei)</b> | 119/205                  | 58%               | 0.533                |
| <b>- pessoa/ - plural (-o)</b>     | 353/703                  | 50,2%             | 0.513                |
| <b>- pessoa/ + plural (-m)</b>     | 11/58                    | 19%               | 0.221                |
| <b>Total</b>                       | 492/980                  | 50,2%             | --                   |

Significância: 0.004

A tabela apresenta resultados bastante interessantes. Das quatro possibilidades de morfologia verbal, três favorecem o sujeito foneticamente realizado, o que confirma que o fenômeno está em estágio avançado de implementação na estrutura da língua. Porém, é fundamental que se comente os resultados obtidos para cada uma das variantes para a melhor compreensão das especificidades desse quadro de variação.

A forma verbal que mais favorece a ocorrência do sujeito foneticamente realizado é a que apresenta desinência número-pessoal [+ pessoa, + plural], representada pelos verbos de primeira pessoa do plural. Se analisarmos a hipótese inicial relativa à relação entre o sujeito nulo e o sincretismo do paradigma verbal, veremos que esse resultado não está de acordo com o esperado, uma vez que as formas verbais de morfologia mais latente dispensariam a necessidade expressão do sujeito devido à possibilidade de retomada através da flexão verbal.

Desse modo, atribuímos a alta frequência de sujeitos expressos para essa forma verbal a duas possibilidades: primeiramente, ao fato de que, das 9 ocorrências de sujeito realizado, 5 são da mesma informante, uma mulher da faixa etária mais baixa. Desse modo, é possível que esse resultado expresse mais um comportamento linguístico individual da informante do que uma tendência geral. Outra possibilidade seria, conforme já afirmamos algumas vezes, devido à uniformização da preferência pelo sujeito expresso para todas as formas verbais – tanto as marcadas quanto as não marcadas. De todo modo, é necessário que se obtenha mais dados de [+ pessoa, + plural] para confirmar ou refutar essa

possibilidade, uma vez que o baixo número de ocorrências pode ter motivado o resultado obtido.

A segunda forma verbal que mais favorece o sujeito expresso é a de traço [+ pessoa, - plural], representada pelos sujeitos de primeira pessoa do singular. Temos, então, novamente, verbos de desinência latente favorecendo o uso do sujeito expresso, o que é mais um indício que aponta para a uniformidade da queda de frequência do sujeito nulo à revelia das formas verbais com as quais esses sujeitos aparecem. É possível, também, que o informante escolha essa variante para dar ênfase às suas ações ou à sua participação nas ações. Para essa variante, a frequência de uso do sujeito expresso foi de 58 pontos percentuais e peso relativo de 0.533.

A variável seguinte é a que representa os verbos com traço [- pessoa/ - plural], sendo esse o traço mais sincrético, uma vez que verbos com esse traço podem ser utilizados com até quatro pessoas pronominais. Apesar disso, a frequência de uso entre o sujeito nulo e o sujeito expresso, nesse contexto, é bastante equilibrada, sendo os sujeitos expressos levemente mais favorecidos, com frequência de 50,2% e peso relativo de 0.513. Isso aponta, mais uma vez, para a baixa relevância da morfologia verbal para a frequência de realização ou não realização do sujeito.

Por último, comentaremos os resultados encontrados para os verbos com desinência [- pessoa/ + plural]. Essa desinência inclui duas possibilidades de pessoas: a segunda pessoa do plural e a terceira pessoa do plural. Entretanto, assim como o observado por Oliveira (2016) no português moçambicano, não houve, no nosso *corpus*, nenhuma ocorrência de segunda pessoa do singular. Isso pode ser atribuído à forma como as entrevistas foram guiadas: grande parte das perguntas têm narrativas como respostas, o que desfavorece o uso de segundas pessoas. Outra possibilidade de interpretação desse dado diz respeito à forma como o falante se dirige ao entrevistador: o uso do pronome você/vocês é categórico nessa variedade linguística, não sendo identificada nenhuma ocorrência de vós (em alguns casos, dois entrevistadores conduziam as entrevistas). Desse modo, apesar

do traço [- pessoa], os sujeitos analisados aqui são apenas de terceira pessoa do plural.

Como demonstra a tabela, esse traço desfavorece fortemente o uso do sujeito expresso, que aparece com frequência de 19% e peso relativo de 0.221. Conforme já afirmamos, o português dos Tongas não apresenta diferença significativa de frequência entre as diferentes desinências número-pessoais dos verbos, o que nos leva a crer que esse resultado não é motivado pela relação entre sujeito nulo e forma verbal mais marcada, mas sim devido ao comportamento dos sujeitos de terceira pessoa que, por causa do seu conjunto de traços semânticos descritos por Cyrino, Duarte e Kato e já apresentados anteriormente e ao seu caráter anafórico, são um contexto de resistência ao sujeito expresso.

### 6.3 A INTERAÇÃO ENTRE AS VARIÁVEIS SOCIAIS: CRUZAMENTO ENTRE SEXO X FAIXA ETÁRIA E SEXO X ESCOLARIDADE

Em análises variacionistas, há grande expectativa acerca do comportamento das variáveis sociais para a configuração do fenômeno em variação, uma vez que o entendimento da influência dessas variáveis para o comportamento dos diferentes fenômenos é a razão de ser da Sociolinguística Variacionista. Entretanto, estudos sobre o sujeito nulo em diversas variedades linguísticas têm demonstrado que, para a realização do sujeito pronominal, é pouca ou nenhuma a influência dos fatores sociais isolados, padrão que se repetiu na análise da variedade linguística dos Tongas. No nosso estudo, apenas a variável social Faixa etária foi selecionada como estatisticamente relevante pelo GoldVarb X, tendo as demais variáveis sociais (Sexo, Escolaridade e Estada fora da comunidade) apresentado resultados percentuais próximos da neutralidade.

Uma possibilidade de análise do comportamento das variáveis sociais é o cruzamento entre essas variáveis, com o objetivo de avaliar se a interação entre elas causa efeitos que não são previsíveis através dos seus resultados isolados. Desse modo, para verificar se há, efetivamente, efeito conjunto de diferentes variáveis sobre a variável dependente, resolvemos realizar alguns cruzamentos entre as variáveis sociais analisadas nesse estudo. Para tanto, fizemos uso do comando *Cross Tabulation* no Goldvarb X, para que fossem feitos os cruzamentos Sexo x

Faixa etária, Sexo x Escolaridade e Sexo X Estada fora da comunidade. Após realizar os *Cross Tab*, criamos, no arquivo de condições, uma variável nova para cada cruzamento, atribuindo uma variante para cada célula gerada. Dessa forma, foi possível avaliar estatisticamente a relevância desses cruzamentos, gerando percentuais e pesos relativos para cada um deles.

O cruzamento realizado entre Sexo e Estada fora da comunidade não foi selecionado pelo programa como estatisticamente relevante. Os cruzamentos Sexo x Faixa etária e Sexo X Escolaridade foram selecionados, e os resultados serão apresentados nas tabelas a seguir:

Tabela 42: Cruzamento entre Sexo e Faixa etária

| <b>Sexo X Faixa etária</b> | <b>Ocorrências/total</b> | <b>Percentual</b> | <b>Peso relativo</b> |
|----------------------------|--------------------------|-------------------|----------------------|
| <b>Mulheres da faixa 1</b> | 110/175                  | 62,9%             | 0.692                |
| <b>Mulheres da faixa 2</b> | 42/148                   | 28,4%             | 0.286                |
| <b>Mulheres da faixa 3</b> | 91/159                   | 57,2%             | 0.530                |
| <b>Homens da faixa 1</b>   | 87/170                   | 51,2%             | 0.535                |
| <b>Homens da faixa 2</b>   | 76/172                   | 44,2%             | 0.415                |
| <b>Homens da faixa 3</b>   | 86/156                   | 55,1%             | 0.517                |
| <b>Total</b>               | 492/980                  | 50,2%             | --                   |

Significância: 0.007

Os resultados apresentados pela tabela 42 mostram que, em relação ao movimento e mudança rumo a sujeitos preferencialmente realizados na variedade dos Tongas, o efeito da variável Sexo não é relevante mesmo quando ela é analisada em conjunto com a variável Faixa etária. Os grupos que mais favorecem o uso do sujeito expresso são, respectivamente, o das mulheres da faixa 1, com peso relativo de 0.692, e o dos homens da faixa 1, com 0.535, o que, portanto, confirma a homogeneidade do comportamento linguístico entre homens e mulheres. Em outras palavras, mesmo quando as variáveis são cruzadas, o fator Faixa etária segue sendo preponderante.

A tabela a seguir apresenta os resultados do cruzamento entre as variáveis Sexo e Escolaridade.

Tabela 43: Cruzamento entre Sexo e Escolaridade

| <b>Sexo X Escolaridade</b>      | <b>Ocorr./total</b> | <b>Percentual</b> | <b>Peso relativo</b> |
|---------------------------------|---------------------|-------------------|----------------------|
| <b>Mulheres semianalfabetas</b> | 127/225             | 56,4%             | 0.632                |
| <b>Mulheres analfabetas</b>     | 116/257             | 45,1%             | 0.416                |
| <b>Homens semianalfabetos</b>   | 110/232             | 47,4%             | 0.419                |
| <b>Homens analfabetos</b>       | 139/266             | 52,3%             | 0.539                |
| <b>Total</b>                    | 492/980             | 50,2%             | --                   |

Significância: 0.040

Pode-se apreender dos resultados apresentados acima que, embora o fator Escolaridade não seja, de forma isolada, considerado relevante, ele passa a sê-lo quando associado ao sexo dos informantes. As mulheres semianalfabetas favorecem o uso do sujeito expesso, com peso relativo de 0.632, enquanto as mulheres analfabetas desfavorecem seu uso, com peso relativo 0.416. No caso dos homens, o oposto é o que ocorre: aqueles que não têm escolaridade favorecem o uso do sujeito expesso, com peso relativo de 0.539, e os que apresentam pouca escolarização o desfavorecem, com peso relativo de 0.419. Esse resultado é bastante interessante para que possamos entender o quanto são particulares os mecanismos que norteiam o funcionamento das variáveis sociais em determinada comunidade de fala: as hipóteses acerca das categorias sociais pré-definidas necessitam de dados empíricos para sua confirmação, o que só é possível com a análise dos dados.

#### 6.4 VARIÁVEIS LINGUÍSTICAS NÃO SELECIONADAS

Nessa seção, iremos comentar os resultados obtidos para as duas variáveis linguísticas que não foram selecionadas como relevantes pelo Goldvarb X, mas que, ainda assim, apresentaram resultados interessantes: as variáveis Tipo de frase e Posição do Sujeito. Sobre variáveis não selecionadas, Guy afirma que mesmo que um grupo de fatores “não seja significativo, mesmo assim não seria aconselhável abandoná-lo, sem antes entender porque teve tanta importância em outros lugares” (GUY, ZILLES, 2007, p. 185). O fato é que, para o enriquecimento de qualquer pesquisa sociolinguística, se faz necessário compreender, de forma aprofundada, o funcionamento de todas as variáveis linguísticas e sociais envolvidas, fornecendo

respostas para questões que podem ter sido levantadas no decorrer do estudo e, principalmente, compreendendo de forma autônoma os resultados encontrados. Desse modo, apresentaremos e comentaremos, a seguir, os resultados das duas variáveis linguísticas não selecionadas.

#### 6.4.1 Tipo de frase

A inclusão da variável Tipo de frase no escopo da análise do sujeito nulo no português dos Tongas teve como motivação os resultados encontrados por Duarte (1995) no português europeu. De acordo com os dados obtidos pela autora, o sujeito nulo é preferencialmente utilizado nas orações interrogativas, com 77% de sujeitos nulos em contraposição aos 67% encontrados nas orações declarativas.

A nossa análise foi feita com base em entrevistas pautadas em diálogos em turnos de fala, formato que dificulta a incidência de orações interrogativas, uma vez que perguntas não são recorrentes nesse contexto. Desse modo, a grande maioria das orações foram declarativas/afirmativas, sendo essas seguidas pelas orações negativas. O menor percentual de sujeitos expressos foi observado nas orações negativas, conforme apresentamos na tabela a seguir:

Tabela 44: O sujeito expresso e a variável Tipo de frase

| <b>Tipo de frase</b> | <b>Ocorrências/total</b> | <b>Percentual</b> |
|----------------------|--------------------------|-------------------|
| <b>Afirmativa</b>    | 416/821                  | 50,7%             |
| <b>Interrogativa</b> | 7/14                     | 50%               |
| <b>Negativa</b>      | 69/145                   | 47,6%             |
| <b>Total</b>         | 492/980                  | 50,2%             |

Os resultados acima não confirmam os achados de Duarte para o português europeu no que diz respeito à maior frequência de sujeitos nulos em orações interrogativas, o que pode ser motivado pela baixa frequência desse tipo de oração no *corpus*. Sobre esses resultados, gostaríamos de chamar atenção para as orações negativas: nesse contexto, há um desfavorecimento, ainda que discreto, do sujeito expresso. Considerando que as orações negativas apresentam a partícula *não* antes do verbo, há a possibilidade de estar acontecendo a influência dessa partícula para um maior favorecimento do sujeito nulo, uma vez que, conforme já dissemos,

determinadas partículas pré-verbais são, para os falantes, intercambiáveis com o sujeito expreso. Os exemplos abaixo ilustram esse contexto:

(125) Ele morreu com trinta e três anos. Já **não** tinha direito, mas ele foi duas vezes pa Cabo-Verde.

(126) **Não** quero por exemplo apredir com machim ou com veneno porque se eu matar senhor com machim, eu vou pa preso.

#### 6.4.2 Posição do sujeito em relação ao antecedente

Duarte (1995) inseriu a variável Tipo sintático da oração para verificar se há, de fato, correlação entre a estrutura das orações e os índices de realização ou apagamento do sujeito pronominal. A autora, para proceder à análise, excluiu do escopo das possibilidades as orações coordenadas pois, segundo ela, orações sequenciadas com sujeito de mesma referência tendem a apagar o sujeito até mesmo em línguas que não são positivas para o parâmetro *pro drop*. Entretanto, para verificar se essa tendência se manteria no português dos Tongas, optamos por não excluir da nossa análise as orações com essa característica, considerando que, nessa variedade de língua, os sujeitos plenos são preferidos na maioria dos contextos. Os resultados obtidos por Duarte foram os dispostos na tabela a seguir:

Tabela 45: Sujeito nulo e a variável Tipo sintático da oração no PB

| Tipo sintático da oração | Frequência |
|--------------------------|------------|
| Oração independente      | 37%        |
| Adjunta posposta         | 30%        |
| Raiz posposta            | 26%        |
| Completiva               | 24%        |
| Raiz anteposta           | 23%        |
| Adjunta anteposta        | 15%        |
| Relativa                 | 7%         |

Fonte: DUARTE, 1995, p. 59.

Os resultados de Duarte demonstraram que, embora todos os contextos apresentem baixa frequência de sujeitos pronominais nulos, as orações relativas se mostram um contexto de extrema resistência, com frequência de apenas 7% de

sujeitos não realizados. Esses e os demais resultados se opõem aos obtidos pela autora em sua análise do Português Europeu (1993), que apresentou como resultado o favorecimento do sujeito nulo em todos os contextos sintáticos, embora, nessa variedade, as orações relativas também tenham apresentado, com sujeitos de primeira e terceira pessoa, um baixo índice de sujeitos nulos (30 e 39%, respectivamente).

Para a autora, a preferência por sujeitos plenos nas relativas pode ser motivada por uma questão estrutural, uma vez que, nesses contextos, a expressão do sujeito possibilitaria a marcação da função sintática do pronome relativo *que*, evitando ambigüações entre a função sujeito e a função objeto. Na nossa análise, apesar do baixo índice de orações relativas, foi possível confirmar essa hipótese, conforme apresentado na tabela 46:

Tabela 46: Realização do sujeito pronominal em função da variável Tipo de oração

| <b>Posição do sujeito</b>          | <b>Ocorrências/total</b> | <b>Percentual</b> |
|------------------------------------|--------------------------|-------------------|
| <b>Em oração relativa</b>          | 7/9                      | 77,8%             |
| <b>Em oração adverbial</b>         | 6/8                      | 75%               |
| <b>Sem encadeamento</b>            | 533/915                  | 58,2%             |
| <b>1ª coordenada sindética</b>     | 15/65                    | 23%               |
| <b>Principal posposta</b>          | 29/52                    | 50%               |
| <b>Em oração completiva</b>        | 4/8                      | 50%               |
| <b>2ª coordenada sindética</b>     | 2/6                      | 33,3%             |
| <b>No discurso do documentador</b> | 23/72                    | 29,2%             |
| <b>1ª oração assindética</b>       | 15/77                    | 22,1%             |
| <b>2ª oração assindética</b>       | 5/29                     | 17,2%             |
| <b>3ª oração assindética</b>       | 3/24                     | 12,5%             |
| <b>Total</b>                       | 492/980                  | 50,2%             |

Outro resultado digno de nota diz respeito ao comportamento das orações coordenadas, sejam elas sindéticas ou assindéticas. Conforme afirmado por Duarte,

as orações com essa configuração favorecem a não expressão do sujeito mesmo em línguas que não são *pro drop*. Nossos resultados confirmaram esse mecanismo, embora não de forma categórica: as frequências de sujeito foneticamente realizado vão diminuindo de acordo com a distância dessas orações coordenadas em relação ao referente. Desse modo, as primeiras coordenadas tendem a apresentar maior frequência de sujeitos expressos, e as terceiras coordenadas apresentam frequência menor de sujeitos foneticamente realizados. Os exemplos apresentados a seguir ilustram essa configuração:

(127) O pessoal mandou trazer ele porque ele tinha passado esse ano de servir tropa já, mas morreu com trinta e três anos e já não tinha mais direito, já não tinha direito.

(128) Tendo em conta eu morava aqui, estudava na cidade, dormia na cidade, lá pra mea noite é que ia pra cama.

Vimos, na análise do comportamento do Parâmetro do Sujeito Nulo no português dos Tongas, que são muitos os fatores linguísticos envolvidos, enquanto, por outro lado, não há grande influência das variáveis sociais para a configuração do fenômeno, assim como observado por Duarte em suas análises no Português Europeu e no Português Brasileiro.

A pouca relevância de fatores sociais, bem como a distribuição não linear do fenômeno entre as diferentes faixas etárias, é um forte indicativo do avanço da mudança naquela comunidade de fala, o que nos leva a crer que, no caso do português dos Tongas, é avançada e estável a implementação do uso do sujeito foneticamente expresso. Esse quadro, no entanto, não é indicativo da não influência dos fatores que envolveram a formação dessa variedade linguística para a sua atual configuração, mas sim da estabilidade e distribuição equilibrada da mudança na estrutura da língua e da sociedade dos Tongas.

## 7 CONSIDERAÇÕES FINAIS

A língua é a parte efetiva da linguagem, sendo o seu funcionamento condicionado a questões sociais e históricas que dizem respeito a cada comunidade de fala. Desse modo, as línguas se tornam objetos complexos e fascinantes, e investigar sua história é fundamental para compreender o que motiva a sua variabilidade e as suas tendências de mudança.

Esse trabalho teve como principal objetivo a análise do funcionamento do Parâmetro do Sujeito Nulo no português dos Tongas, variedade linguística falada pelos descendentes dos africanos continentais contratados para o trabalho nas grandes roças de cacau e café da ilha de São Tomé, buscando verificar o *status* do fenômeno nessa comunidade e analisar as possíveis tendências de mudança, bem como os fatores linguísticos e sociais que se mostram relevantes para a variação na marcação desse parâmetro.

São várias as perspectivas através das quais se pode investigar o funcionamento de uma língua. No âmbito dessa tese, realizamos o estudo do Parâmetro do Sujeito Nulo através da perspectiva da Sociolinguística Paramétrica, por acreditar que a mudança sintática pode ser mais bem observada se associada à mudança na marcação dos parâmetros que compõem a Gramática Universal. Desse modo, além da descrição do fenômeno, pudemos levantar o conjunto de fatores linguísticos e sociais associados às diferentes marcações paramétricas do fenômeno. Assim, para compreender e explicar o que motiva a variação nesse parâmetro, lançamos mão do que prescreve a Sociolinguística Paramétrica, que busca explicar a variação linguística como sendo motivada pelas diferentes valorações atribuídas aos diferentes parâmetros linguísticos.

Depois de traçar, na seção 2, um histórico da população tonga, explanando a forma como aconteceram os dois ciclos de colonização da ilha, apresentamos o conceito e a estrutura das roças, unidades em torno das quais girou a economia e a vida nos primeiros anos de existência de São Tomé. Em seguida, apresentamos a roça de Monte Café, onde vivem os Tongas, e a composição étnica de São Tomé,

comentando acerca das variedades linguísticas que são lá faladas: o forro, o lung'le, o angolar, o kabuverdianu, o português de São Tomé e o português dos Tongas. Esses dados foram de extrema importância para que pudéssemos compreender e explicar a relação entre a socio-história de São Tomé e a sua configuração sociolinguística.

Na terceira seção, tratamos dos pressupostos teóricos nos quais a análise se ancorou. Apresentamos os preceitos da Sociolinguística Variacionista e os seus conceitos fundamentais, como o de variável linguística, variação estável, mudança em curso e comunidade de fala. Além disso, apresentamos o modelo de Princípios e Parâmetros os fundamentos da associação entre modelos teóricos denominada Sociolinguística Paramétrica. Na subseção seguinte, apresentamos as principais teorias que visam escrutinar o efeito do contato massivo entre línguas tipologicamente diferentes para as variedades linguísticas que se formam: a Crioulística e a Transmissão Linguística Irregular.

Na quarta seção, apresentamos a metodologia seguida para a realização da pesquisa: comentamos acerca dos passos seguidos para proceder à constituição da nossa amostra de fala, as estratégias seguidas para mitigar o paradoxo do observador, com vistas a obter o vernáculo dos informantes, bem como o perfil social de cada um deles. Nossa análise foi realizada a partir do estudo de entrevistas fornecidas por 18 informantes, todos com pouca ou nenhuma escolaridade e estratificados em três faixas etárias. Foi possível, desse modo, apresentar o método de obtenção dos dados e identificar quais ocorrências deveriam ser descartadas da análise, bem como comentar as nossas hipóteses para cada variável linguística e social. Por fim, demonstramos, também, os passos para o processamento quantitativo dos dados obtidos.

Na seção 5, fizemos um panorama do sujeito pronominal em algumas variedades da língua portuguesa. Apresentamos as noções de pronome e de pessoa pronominal, trazendo o quadro pronominal do latim, do português europeu, do português brasileiro, do português afro-brasileiro, do português de São Tomé, do crioulo de São Tomé e do português dos Tongas. Em seguida, propusemos uma discussão acerca das várias concepções de sujeito, partindo, em seguida, para o debate sobre a definição do Parâmetro do Sujeito Nulo, mostrando as diferentes tipologias linguísticas em relação ao licenciamento do sujeito nulo. Por fim, com

vistas a apresentar um breve histórico de pesquisas anteriores acerca do funcionamento do parâmetro *pro drop*, apresentamos os resultados obtidos por alguns autores em relação ao funcionamento desse parâmetro em algumas variedades linguísticas formadas a partir do contato entre línguas.

Em nossa análise, tivemos o objetivo de responder a algumas questões essenciais:

- Está havendo, no português dos Tongas, a diminuição da frequência no uso do sujeito nulo?
- Pode-se traçar semelhanças entre o funcionamento do parâmetro *pro drop* no português Tongas e em outras variedades de língua portuguesa?
- Quais são os fatores sociais e estruturais que atuam no funcionamento desse parâmetro?
- Pode-se considerar o português dos Tongas uma língua de sujeito nulo parcial, assim como é o português brasileiro?

Sabemos que sistemas linguísticos que se encontram em processo de mudança podem se diferenciar não apenas na marcação ou não marcação de determinado parâmetro, mas também na frequência em que as propriedades relacionadas a esse parâmetro aparecem (TARALLO, KATO, 1989). Desse modo, resolvemos verificar se os fatores que atuam em prol da mudança na marcação do sujeito nulo em outras variedades linguísticas também atuam para motivar uma possível mudança desse parâmetro no português dos Tongas.

Em nossa análise, obtivemos 50,2% de sujeitos realizados e 49,8% de sujeitos nulos, números que se distanciam daqueles encontrados em outras variedades do português, como o português brasileiro, o português afro-brasileiro e o português europeu. A diferença na frequência encontrada para o português dos Tongas em relação a outras variedades linguísticas pode ser atribuída ao nível de isolamento dos habitantes da roça que, em sua maioria, sequer saíram da localidade. Resultado semelhante foi encontrado para o português moçambicano, o que indica que ambas as comunidades linguísticas estão passando, em relação ao Parâmetro do Sujeito Nulo, por processos semelhantes: nas duas variedades, a frequência de sujeitos foneticamente realizados parece estar gradativamente

aumentando, o que é um forte indício de que a preferência por sujeitos nulos, herança da gramática do português europeu, está perdendo força. É possível inferir, desse modo, que, em relação ao Parâmetro do Sujeito Nulo, o português dos Tongas está seguindo os mesmos passos que seguiu o português brasileiro no passado.

A variável Pessoa do sujeito apresentou resultados interessantes no que diz respeito à relação entre a possibilidade identificação do sujeito através da morfologia verbal e o licenciamento do sujeito nulo. Ao cruzar as variáveis Pessoa do sujeito e Forma verbal, entendemos que, embora o sincretismo da morfologia dos verbos e a inserção dos pronomes *a gente* e *você* no paradigma pronominal do português dos tongas possa ter sido uma importante porta de entrada para o aumento da frequência dos sujeitos foneticamente realizados, há a preferência de uso dos sujeitos plenos para quase todas as pessoas pronominais, com a exceção das terceiras pessoas, que apresentam referentes fortes, comumente sintagmas nominais.

É importante destacar, também, que todos os informantes são falantes de kimbundu, língua que apresenta altos níveis de sincretismo flexional, como se pode perceber na tabela 47. Desse modo, é possível que a preferência por sujeitos foneticamente realizados seja, também, uma influência do kimbundu.

Tabela 47: Paradigma flexional do Kimbundu

|            |             |
|------------|-------------|
| Eme ngidya | Eu como     |
| Eye udya   | Tu come     |
| Mwene udya | Ele come    |
| Etu tudya  | Nós comemos |

Fonte: Oliveira, 2019, p. 44.

As variáveis selecionadas pelo Goldvarb X dão um importante panorama do *status* do Parâmetro do Sujeito Nulo no português dos Tongas. A primeira delas, Referência anterior ao sujeito, apresentou resultados que confirmaram o mecanismo do paralelismo formal: os sujeitos em primeira menção são realizados com maior frequência, e os pronomes realizados são favorecidos por uma ocorrência pronominal anterior. Os resultados nos mostraram, também, que os sujeitos que têm sintagmas nominais como correferentes apresentam menores frequências de

sujeitos pronominais realizados, uma importante confirmação da força da correferência externa no caso das terceiras pessoas.

A variável seguinte, Caracterização semântica do sujeito, apresenta um panorama bastante claro no que diz respeito à relação entre o traço semântico [-animado] e a realização do sujeito pronominal: esse traço é um forte desfavorecedor do sujeito pronominal expresso, enquanto o traço [+animado] favorece o uso do sujeito realizado.

Embora a variável social Faixa etária tenha sido selecionada como estatisticamente relevante pelo Goldvarb X, ela não apresentou a gradação típica dos quadros de mudança em curso. Esse dado, no entanto, se mostra bastante condizente com o que foi encontrado até aqui: os pronomes plenos são mais utilizados pelos informantes mais novos, enquanto os pronomes nulos, por sua vez, são mais frequentes entre a faixa etária intermediária. Esses dados nos mostram que o fenômeno se encontra em variação estável, estando distribuído de forma equilibrada no contexto da comunidade de fala dos Tongas.

A variável seguinte selecionada pelo programa foi a Presença de constituinte à esquerda do verbo. Nos cenários em que não há constituintes, tais como advérbios, palavras -QU ou clíticos à esquerda do verbo, houve o favorecimento do sujeito realizado. Por outro lado, a presença de constituintes favorece o sujeito nulo, resultado condizente ao encontrado por Duarte (1995) para essa variável.

A última variável selecionada foi a Desinência número-pessoal do verbo, cujos resultados foram mais uma comprovação de que o fenômeno está distribuído de forma equilibrada entre as diferentes formas verbais e que o sincretismo verbal não é o fator decisivo para o comportamento do fenômeno no português dos Tongas, sendo o sujeito expresso a preferência até mesmo nos contextos em que a morfologia verbal permite a identificação do sujeito.

O português dos Tongas, assim como o português brasileiro, apresenta um forte sincretismo das formas verbais, constando no seu quadro pronominal a inserção dos pronomes *você* e *a gente*. Esses dados, juntamente com a extensa análise da influência das variáveis linguísticas e sociais apresentadas neste trabalho, nos leva a crer que o português dos Tongas é uma língua de sujeito nulo parcial, embora essa classificação pareça abarcar línguas com comportamentos

diferentes em relação a esse parâmetro, o que revela a complexidade da tarefa de buscar uma classificação exata. Sobre essa dificuldade, Veríssimo afirma que

Na medida em que mais línguas de origens diversas foram sendo analisadas [...] os dados começaram a revelar que o parâmetro do sujeito está relacionado a casos que vão além de uma simples marcação binária sobre a admissibilidade ou não de sujeitos nulos. Ou seja, o comportamento das línguas não se limitava ao licenciamento ou não de sujeitos nulos em contextos bem especificados, mas abarcava diferentes possibilidades que envolviam restrições em maior ou menor grau à expressão fonológica do sujeito. Tornou-se evidente que as propriedades demonstradas pelo sujeito nulo são determinadas por fatores gramaticais mais complexos, que não podem ser captados por uma simples binariedade nas opções (Veríssimo, 2017, p. 80).

Se a classificação das línguas se mostra uma tarefa complexa, enquadrar línguas formadas através de contato massivo é ainda mais desafiador, uma vez que o contexto sociocultural das comunidades de fala e a influência das línguas envolvidas no contato são fatores preponderantes. Dessa forma, compreender o funcionamento das línguas fruto do contato massivo é dar um importante passo rumo à compreensão aprofundada do funcionamento das línguas humanas em contextos multilíngues.

Esperamos que nosso trabalho possa ser somado à longa e importante tradição de estudos acerca do funcionamento do Parâmetro do Sujeito Nulo, podendo, também, despertar a curiosidade e motivar novas pesquisas acerca das variedades africanas da Língua Portuguesa, na busca de compreender a variação dos parâmetros sintáticos como motivadores dos processos de mudança linguística.

## REFERÊNCIAS

- AGOSTINHO, Ana Livia dos Santos. **Fonologia e método pedagógico do lung'le. 2014.** Tese (Doutorado em Letras) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2014.
- AGOSTINHO, Ana Livia dos Santos; LIMA, Manuele Bandeira de Andrade; ARAUJO, Gabriel Antunes. O Lung'le na educação escolar de São Tomé e Príncipe. **Trabalhos em linguística aplicada**, v. 55, p. 591-618, 2016.
- AGOSTINHO, Ana Livia; BANDEIRA, Manuele; FREITAS, Shirley. Línguas crioulas e o papel da língua portuguesa em São Tomé e Príncipe. **Lusofonias em debate**, p. 63-86, 2020.
- ALI, Manuel Said. **Gramática histórica da língua portuguesa.** São Paulo: Melhoramentos, 1966.
- ALMEIDA, Napoleão Mendes de. **Gramática Latina.** 26. ed. São Paulo: Saraiva, 1995.
- ARAUJO, Gabriel Antunes; HAGEMEIJER, Tjerk. **Dicionário livre santome-português.** Hedra, 2013.
- BANDEIRA, Manuele. **Reconstrução fonológica e lexical do protocrioulo do Golfo da Guiné.** 2017. Tese (Doutorado) - Universidade de São Paulo, São Paulo, 2017.
- BAXTER, Alan. A re-estruturação do português numa situação de contato africana. **Trabalho apresentado ao 3º Congresso Internacional da Associação Brasileira de Linguística**, Rio de Janeiro, 2003.
- BAXTER, Alan. O português dos Tongas de São Tomé. *In*: OLIVEIRA, Márcia; ARAUJO, Gabriel (Orgs.). **O português na África Atlântica: Angola, Cabo Verde, Guiné Bissau, São Tomé e Príncipe.** 2ª ed. São Paulo: FFLCH/USP, 2019.
- BAXTER, Alan. O português vernáculo do Brasil - Morfossintaxe. *In*: PERL, Matthias; SCHWEGLER, Armin (Orgs.). **América negra: panorâmica actual de los estudios lingüísticos sobre variedades hispanas, portuguesas y criollas.** Frankfurt am Main: Vervuert; Madrid: Iberoamericana, 1998. p. 97-134.
- BAXTER, Alan. 'Semicreolization'?—The restructured Portuguese of the Tongas of São Tomé, a consequence of L1 acquisition in a special contact situation. **Journal of Portuguese Linguistics**, v. 1, n. 1, 2002.
- BAXTER, Alan; LUCCHESI, Dante. A relevância dos processos de pidginização e crioulição na formação da língua portuguesa no Brasil. **Estudos Linguísticos e Literários**, Salvador, n.19, p.65-83, 1997.

BAXTER, Alan; LUCCHESI, Dante. Processos de descrioulização no sistema verbal de um dialeto rural brasileiro. **Papia - Revista de Crioulos de Base Ibérica**, Brasília, Universidade de Brasília, v.2, n.2, p. 59-71, 1993.

BAXTER, Alan; MELLO, Camila Ferreira; ALMEIDA SANTANA, Natali Gomes. A construção de objeto duplo e as influências do substrato no português afro-brasileiro (e africano). **PAPIA-Revista Brasileira de Estudos do Contato Linguístico**, v. 24, n. 2, p. 283-306, 2014.

BECHARA, Evanildo. **Moderna gramática brasileira**. Rio de Janeiro: Lucerna, 2009.

BENVENISTE, Émile. A natureza dos pronomes. In: \_\_\_\_\_ . **Problemas de linguística geral I**. 4. ed. Campinas: Pontes, 1956.

BERTHET, Marina Annie. Reflections on roças in São Tomé and Príncipe. **Estudos Históricos (Rio de Janeiro)**, v. 25, n. 50, p. 331-351, 2012.

BERTHET, Marina. São Tomé e Príncipe: reflexões sobre alguns aspectos de sua história agrícola no pós-independência. **Estudos Ibero-Americanos**, v. 42, n. 3, p. 961-986, 2016.

BICKERTON, Derek. Creole languages and the Bioprogram. *In*: NEWMeyer, Frederick (Org.). **Linguistics: the Cambridge survey**. Cambridge: Cambridge University Press, 1988. p. 268-284.

BICKERTON, Derek. Creolization, linguistic universals, natural semantax and the brain. **Issues in English creoles: Papers from the 1975 Hawaii conference**, p. 1-18, 1980.

BICKERTON, Derek. *Roots of language*. Ann Arbor: Karoma, 1981.

BICKERTON, Derek. The Language Bioprogram Hypothesis. **Behavioural and Brain Sciences**, Cambridge, n.7, 1984, p.173-203.

BOUCHARD, Marie-Eve. **Linguistic variation and change in the Portuguese of São Tomé**. 2017. Tese (Doutorado) - New York University, New York, 2017.

BRIGHT, William. As dimensões da sociolinguística. *In*: FONSECA, Maria Stella; NEVES, Moema (Orgs.). **Sociolinguística**. Rio de Janeiro: Eldorado, 1974.

CAMACHO, Roberto Gomes. Uma reflexão crítica sobre a teoria sociolinguística. **DELTA: Documentação de Estudos em Linguística Teórica e Aplicada**, v. 26, n. 1, p. 141-162, 2010.

CÂMARA JR, Joaquim Mattoso. **Estrutura da língua portuguesa**. Petrópolis: Vozes, 1970.

CARNEIRO, Amanda dos Santos. **Descrição e classificação das expressões cristalizadas com ser e estar do português do Brasil**. 2016. 158 f. Dissertação (Mestrado em Linguística) – Universidade Federal de São Carlos, São Carlos, 2016.

CARUSO, Francisco; OGURI, Vitor. **Física Moderna: origens clássicas e fundamentos quânticos**. Elsevier, 2006.

CASTILHO, Ataliba de. **Nova gramática do português brasileiro**. São Paulo: Contexto, 2010.

CASTRO, Yeda Pessoa de. **A língua mina-jeje no Brasil: um falar africano em Ouro Preto do século XVIII**. Belo Horizonte: Secretaria de Estado da Cultura, 2002.

CHAGAS, Paulo. A mudança linguística. *In*: FIORIN, José Luiz (org.). **Introdução à Linguística: Objetos teóricos**. 5. Ed. São Paulo: Contexto, 2007.

CHAUDENSON, Robert. Créolisation linguistique, créolisation culturelle. **Études créoles**, v. 12, n. 1, p. 53-73, 1989.

CHOMSKY, Noam. **Knowledge of language: its origin, nature and use**. New York: Praeger, 1986.

CHOMSKY, Noam. **Lectures on government and binding**. Dordrecht: Foris, 1981.

CHOMSKY, Noam. Principles and Parameters in syntactic theory. *In*: HORNSTEIN, Norbert; LIGHTFOOT, David (orgs.), **Explanation in Linguistics: the logical problem of language acquisition**. Londres: p. 32-75, 1981.

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

COUTO, Hildo Honório. **Introdução ao estudo das línguas crioulas e pidgins**. Editora UnB, 1996.

CUNHA, Celso; CINTRA, Lindley. Língua e Sociedade: variação e conservação linguística. *In*: \_\_\_\_\_. **Nova gramática do português contemporâneo**, v. 1, 1985.

CUNHA, Celso; CINTRA, Lindley. Pronomes. *In*: \_\_\_\_\_. **Nova Gramática do Português Contemporâneo**. Rio de Janeiro: v. 6, p. 289-382, 1997.

CYRINO, Sonia. **O objeto nulo no português do Brasil: um estudo sintático-diacrônico**. 1994. Tese (Doutorado) – Instituto de Estudos da Linguagem, UNICAMP, Campinas, 1994.

CYRINO, Sonia; DUARTE, Maria Eugênia; KATO, Mary. Visible subjects and invisible clitics in Brazilian Portuguese. *In*: KATO, Mary; NEGRÃO, Esmeralda Vailati. **Brazilian Portuguese and the null subject parameter**. Frankfurt: Vervuert-Iberoamericana, p. 55-73, 2000.

DUARTE, Maria Eugênia. Do pronome nulo ao pronome pleno: a trajetória do sujeito nulo no português do Brasil. In: ROBERTS, Ian; KATO, Mary (Orgs.). **Português brasileiro: uma viagem diacrônica**. Campinas: UNICAMP, p. 107-128, 1993.

DUARTE, Maria Eugênia. A evolução na representação do sujeito pronominal em dois tempos. In: PAIVA, Maria da Conceição; DUARTE, Maria Eugênia. **Mudança linguística em tempo real**. Rio de Janeiro: Contra Capa/Faperj, p. 115-128, 2003.

DUARTE, Maria Eugênia. A perda do princípio “evite pronome” no português brasileiro. 1995. Tese (Doutorado) – Instituto de Estudos da Linguagem, UNICAMP, Campinas, 1995.

DUARTE, Maria Eugênia. O sujeito nulo referencial no português brasileiro e no português europeu. In: GALVES, Charlotte, KATO, MARY. **Português Brasileiro: uma segunda viagem diacrônica**. Campinas: Ed. Unicamp, p. 93-126, 2019.

DUARTE, Maria Eugênia. On the embedding of a syntactic change. **Language Variation in Europe, Papers from ICLaVE**, v. 2, p. 145-155, 2004.

DUARTE, Maria Eugênia. Sociolinguística “Paramétrica”. **Sociolinguística, sociolinguísticas**. São Paulo: Contexto, p. 33-44, 2016.

DUARTE, Maria Eugenia. Termos da oração. **Ensino de gramática: descrição e uso**. São Paulo: Contexto, p. 186-204, 2007.

DUARTE, Maria Eugênia.; REIS, Eduardo Patrick. Revisitando o sujeito pronominal vinte anos depois. **ReVEL**, v. 16, n. 30, p. 173-197, 2018.

DUARTE, Maria Eugenia; MOURÃO, Gabriela; SANTOS, Heitor. Os sujeitos de terceira pessoa: revisitando Duarte 1993. In. DUARTE, Maria Eugênia. **Os sujeitos nas peças de teatro**. Parábola, 2012.

DUARTE, Pape; ANDRADE, Rodrigo. **As Roças de São Tomé e Príncipe**. Lisboa: Tinta da China, 2013.

ECKERT, Penelope. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. **Annual review of Anthropology**, v. 41, p. 87-100, 2012.

EYZAGUIRE, Pablo. Plantations, state farms and smallholders: cocoa production in São Tomé. **Cocoa and Development Conference**, 1993.

FARACO, Carlos Alberto. **Linguística histórica: uma introdução ao estudo da história das línguas**. Parábola, 2005.

FEIO, Joana Areosa. Cabo-verdianos e São-tomenses de ascendência cabo-verdiana em São Tomé e Príncipe na atualidade: uma abordagem etnográfica. **Diáspora Cabo-Verdiana: Temas Em Debate**, p. 200-226, 2016.

FEIO, Joana. **De étnicos a “étnicos”**: Uma abordagem aos “Angolares” de São Tomé e Príncipe. 2008. 108 f. Dissertação (Mestrado em Antropologia) - Instituto Superior de Ciências do Trabalho e Empresa, Lisboa, 2008.

FERRAZ, Luiz Ivens. The Creole of São Tomé. **African Studies**, v. 37, n. 1, p. 3-68, 1978.

FERRONHA, António Luís Alves. O comércio português de escravos nos séculos XV e XVI. **Portugal no mundo**, v. 1, p. 315-324, 1989.

FINGER, Ingrid. Aquisição de segunda língua: abrangência e limitações do modelo gerativista. **Revista da ABRALIN**, v. 2, n. 2, p. 31-57, 2003.

FIORIN, José Luiz (Org.). **Linguística? O que é isso?** São Paulo: Contexto, 2013.

FREITAG, Raquel Meister Ko. Discutindo sexo/gênero na sociolinguística. **Mulheres, linguagem e poder—Estudos de gênero na sociolinguística brasileira**. São Paulo: Blucher, p. 17-74, 2015.

FREITAG, Raquel Meister Ko. Idade: uma variável sociolinguística complexa. **Línguas & Letras**, v. 6, n. 11, p. 105-121, 2005.

FREITAG, Raquel Meister Ko. O "social" da sociolinguística: o controle de fatores sociais. **Revista Diadorim**, v. 8, n. 1, 2011.

FREITAG, Raquel Meister; SANTANA, Cristiane Conceição; ANDRADE, Thais Regina Conceição; SOUSA, Valéria Santos. Avaliação e variação linguística: estereótipos, marcadores e indicadores em uma comunidade escolar. *In*: FREITAG, **Sociolinguística e Política Linguística: Olhares Contemporâneos**. São Paulo: Blucher, p. 139-160, 2016.

FREITAG, Raquel. Problemas teórico-metodológicos para o estudo da variação linguística nos níveis gramaticais mais altos. **Matraga-Revista do Programa de Pós-Graduação em Letras da UERJ**, v. 16, n. 24, 2009.

GARFIELD, Robert. **A history of São Tomé island: 1470-1655: the key to Guinea**. San Francisco: Mellen Research University Press, 1992.

GILLIGAN, Gary Martin. 1987. **A Cross-linguistic Approach to the Pro-Drop Parameter**. 1987. Tese de Doutorado. Los Angeles: University of South California, 1987.

GONÇALVES, Clézio Roberto. **Uma abordagem sociolinguística dos usos das formas você, ocê e cê no português**. 2008. 349 f. Tese (Doutorado) –Programa de Pós-graduação em Linguística, Universidade de São Paulo, São Paulo, 2008.

GONÇALVES, Rita. **Construções ditransitivas no português de São Tomé**. 2016. Tese (Doutorado em Linguística) - Universidade de Lisboa, Lisboa, 2016.

GONÇALVES, Rita. Mudança linguística e variação no português de São Tomé. **Actas do Colóquio Internacional São Tomé e Príncipe numa perspectiva interdisciplinar, diacrónica e sincrónica**, p. 413-430, 2012.

GONÇALVES, Rita; HAGEMEIJER, Tjerk. O português num contexto multilingue: o caso de São Tomé e Príncipe. **Revista Científica da Universidade Eduardo Mondlane: Série Letras e Ciências Sociais**, v. 1, p. 87-107, 2015.

GUASTI, Maria Teresa. **Language acquisition: The growth of grammar**. MIT press, 2002.

GUIMARÃES, Cecília Silva. **Produtora de açúcar e armazém de escravos: mercados e política na Ilha de São Tomé**. 2010. 134 f. Dissertação (Mestrado em História) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2010.

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

HAGEMEIJER, Tjerk. 2000. **Serial verbs in São-Tomense**. Dissertação (Mestrado). Lisboa: FLUL, 2000.

HAGEMEIJER, Tjerk. As ilhas de Babel: A criouliização no Golfo da Guiné. **Revista Camões**, v. 6, p. 74-88, 1999.

HAGEMEIJER, Tjerk. As línguas de S. Tomé e Príncipe. **Revista de crioulos de base lexical portuguesa e espanhola**, v. 1, p. 1-27, 2009.

HAGEMEIJER, Tjerk. Initial vowel agglutination in the Gulf of Guinea creoles. *In*: ABOH, Enoch; SMITH, Norval (Orgs.). **Complex processes in new languages**. Amsterdam/Filadélfia: John Benjamins Publishing Company, p. 29-50, 2009.

HAGEMEIJER, Tjerk. São Tomé e Príncipe: labirinto e laboratório de línguas. *In*: SEIBERT, Gerhard. **Arquipélagos Atlânticos Crioulos: Cabo Verde e S. Tomé e Príncipe numa perspectiva comparada**, p. 41-70, 2017.

HAGEMEIJER, Tjerk. The Gulf of Guinea creoles: Genetic and typological relations. **Journal of Pidgin and Creole Languages**, v. 26, n. 1, p. 111-154, 2011.

HENRIQUES, Isabel Castro. **O ciclo do açúcar em São Tomé nos séculos XV e XVI**. Lisboa: Publicações Alfa, 1989.

HOLM, John. Creole influence on Popular Brazilian Portuguese. *In*: GILBERT, Gleen. (Org.). **Pidgin and Creole languages**. Essays in Memory of John E. Reinecke. Honolulu: University of Hawaii Press, 1987. p.406-429.

HOLM, John. **Languages in contact: the partial restructuring of vernaculars**. Cambridge: Cambridge University Press, 2004.

HOLM, John. A semicrioulização do português vernáculo do Brasil: evidência de contacto nas expressões idiomáticas. **Papia, Revista de Crioulos de Base Ibérica**, Brasília: Universidade de Brasília/Thesaurus, v. 3, n. 2, 1994.

HOLM, John. **Pidgins and Creoles**. Cambridge: Cambridge University Press, 1988.

HOLM, John. Popular Brazilian Portuguese; a semi-creole. *In: D'ANDRADE, Ernesto; KIHM, Alain (Orgs.). Actas do Colóquio sobre Crioulos de Base Lexical Portuguesa*. Lisboa: Colibri, 1992. p.37-66.

HOLMBERG, Anders; ROBERTS, Ian. Introduction: Parameters in Minimalist Theory 1. *In: Diachronic and Comparative Syntax*. Routledge, p. 473-534, 2018.

HYAMS, Nina. **Language acquisition and the theory of parameters**. Springer Science & Business Media, 2012.

HYMES, Dell. **Pidginization and creolization of languages**. CUP Archive, 1971.

INSTITUTO NACIONAL DE ESTATÍSTICA. INE: São Tomé e Príncipe em números. 2017. Disponível em: <<http://www.ine.st/docs/2017/Censos/2017/index.html>>. Acesso em: 07 de ago. de 2020.

JAEGGLI, Osvaldo; SAFIR, Kenneth . The null subject parameter and parametric theory. *In: The null subject parameter*. Springer, Dordrecht, p. 1-44, 1989.

JAEGGLI, Osvaldo; SAFIR, Kenneth. The null subject parameter and parametric theory. *In: \_\_\_\_\_*. **The null subject parameter**. Springer, Dordrecht, 1989. p. 1-44.

KAPETULA, José Gueleka Kandjandja. **Interpretação dos sujeitos nulos no Português de Angola**. 2016. 89 f. Dissertação (Mestrado em Ciências da Linguagem) – Universidade Nova de Lisboa, Lisboa, 2016.

KATO, Mary. A evolução da noção de parâmetros. **DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada**, v. 18, p. 309-337, 2002.

KATO, Mary. Português brasileiro falado: aquisição em contexto de mudança linguística. *In: DUARTE, Inês; LEIRIA, Isabel (orgs.), Atas do Congresso Internacional sobre Português*. Lisboa: Colibri, p. 211-237, 1996.

KATO, Mary; DUARTE, Maria Eugênia. Restrições na distribuição de sujeitos nulos no Português Brasileiro. **Veredas-Revista de Estudos Linguísticos**, v. 18, n. 1, 2014.

KENEDY, Eduardo. **Curso básico de linguística gerativa**. São Paulo: Contexto, 2016.

LABOV, William. What can be learned about change in progress from synchrony descriptions. *In*: SANKOFF, David; CEDERGREN, Henrietta (Orgs.). **Variation Omnibus**. Carbondale; Edmonton: Linguistic Research, p.177-199, 1981.

LABOV, William. **Sociolinguistic Patterns**. Philadelphia: University of Pennsylvania Press, 1972. [*Padrões Sociolinguísticos*. Trad.: Marcos Bagno; Marta Scherre e Caroline Cardoso. São Paulo: Parábola, 2008.]

LANDIM, Edmilsa. **Turismo Rural como factor de desenvolvimento para São Tomé**: Proposta das actividades turísticas para Roça de Monte Café e Suas Dependências. Trabalho de Conclusão de Curso, 2014.

LANGACKER, Ronald. A universalidade no plano da linguagem. *In*: \_\_\_\_\_. **A linguagem e sua estrutura**. Petrópolis: Vozes, 1977.

LAVANDERA, Beatriz. Where does the sociolinguistic variable stop?. **Language in society**, v. 7, n. 2, p. 171-182, 1978.

LENNEBERG, Eric. The biological foundations of language. **Hospital Practice**, v. 2, n. 12, p. 59-67, 1967.

LIMA, Carlos Henrique da Rocha. **Gramática normativa da língua portuguesa**. 49. ed. Rio de Janeiro: José Olympio, 2011.

LIPSKI, John. Null subjects in (Romance-derived) creoles: routes of evolution. **ANNUAL MEETING, SOCIETY FOR PIDGIN AND CREOLE LINGUISTICS**. Los Angeles: 1999.

LUCCHESI, Dante. A Teoria da Variação Linguística: um balanço crítico. **Estudos Linguísticos**, São Paulo, v. 41, n. 2, p. 793-805, 2012.

LUCCHESI, Dante. As duas grandes vertentes da história sociolinguística do Brasil. **DELTA**, São Paulo, v.17, n.1, p.97-130, 2001.

LUCCHESI, Dante. Norma linguística e realidade social. *In*: BAGNO, Marcos (Org.). **Linguística da norma**. São Paulo: Loyola, 2002.

LUCCHESI, Dante. O conceito de transmissão linguística irregular e o processo de formação do português do Brasil. *In*: RONCARATI, Cláudia; ABRAÇADO, Jussara. (Orgs.). **Português brasileiro: contato linguístico, heterogeneidade e história**. Rio de Janeiro: 7Letras, 2003. p.272-284.

LUCCHESI, Dante. Parâmetros sociolinguísticos do português brasileiro. **Revista da ABRALIN**, v.5, n.1 e 2, p. 83-112, 2006.

LUCCHESI, Dante. Variação e norma: elementos para uma caracterização sociolinguística do português do Brasil. **Revista Internacional de Língua Portuguesa**, Lisboa, n.12, p.17-28, 1994.

LUCCHESI, Dante. Variação, mudança e norma: a questão brasileira. In: CARDOSO, Suzana (Org.). **Diversidade linguística e ensino**. Salvador: EDUFBA, 1996. p.69-80.

LUCCHESI, Dante. **Sistema, mudança e linguagem**: um percurso na história da linguística moderna. Parábola, 2004.

LUCCHESI, Dante. Crioulística. In: MOLLICA, Maria Cecília; JUNIOR, Celso Ferrarezi (Ed.). **Sociolinguística, sociolinguísticas**: uma introdução. Editora Contexto, 2016.

MAGALHÃES, Joaquim Romero. O açúcar nas ilhas portuguesas do Atlântico séculos XV e XVI. **Varia História**, v. 25, p. 151-175, 2009.

MATTOS E SILVA, Rosa Virgínia. (Org.). **Sete estudos sobre o português kamayurá**. Salvador: UFBA; PROED, 1988.

MATTOS E SILVA, Rosa Virgínia. **Ensaio para uma socio-história do português brasileiro**. São Paulo: Parábola, 2004.

MATTOS e SILVA, Rosa Virgínia. Ideias para a história do português brasileiro: fragmentos para uma composição posterior. In: CASTILHO, Ataliba de (Org.). **Para a história do português brasileiro**. v.1: Primeiras ideias. São Paulo: Humanitas, 1998. p.21-52.

MELLO, Heloísa. Perfil sociolinguístico de uma comunidade bilingue da zona rural de Goiás. **Revista Linguagem e Ensino**, v. 4, n. 2, p. 61-92, 2001.

MENDES, Amaro Gaudêncio. **Roça Monte Café**: preservação e valorização da presença de um passado histórico de S. Tomé e Príncipe. 2011. Dissertação (Mestrado) - Universidade de Évora, 2011.

MENDES, Elisângela dos Passos. **A flexão de caso pronominal no continuum do português popular da Bahia**. 2016. 270 f. Tese (Doutorado em Língua e Cultura) – Universidade Federal da Bahia, Salvador, 2016.

MEYERHOFF, Miriam. Communities of practice. In: CHAMBERS, Jack; SCHILLING, Natalie (Orgs.). **Handbook of variation and change**. Oxford: Blackwell, 2004. p. 526-548.

MOINO, Ruth. **Passivas no discurso oral e escrito**. 1987. Dissertação (Mestrado) – PUC-SP, São Paulo, 1987.

MOLLICA, Maria Cecília; JUNIOR, Celso Ferrarezi (Orgs.). **Sociolinguística, sociolinguísticas**: uma introdução. Editora Contexto, 2016.

MONTEIRO, José Lemos. **Pronomes pessoais**: subsídios para uma gramática do português do Brasil. Fortaleza: Edições UFC, 1994.

MORAES, Jorge Viana de. Hugo Schuchardt, Silva Neto, Weinreich e Labov: a relação entre alguns scholars e suas contribuições para os Estudos Crioulísticos. **PAPIA: Revista Brasileira de Estudos do Contato Linguístico**, v. 27, 2018, p. 167-228.

NARO, Anthony. O dinamismo das línguas. *In*: MOLLICA, Cecília; BRAGA, Maria Luiza (orgs.). **Introdução à Sociolinguística: o tratamento da variação**. São Paulo: Contexto, p. 43-51, 2003.

NASCIMENTO, Augusto. As fronteiras da nação e das raças em São Tomé e Príncipe: São-tomenses, Europeus e Angolas nos primeiros decênios de Novecentos. **Varia Historia**, v. 29, p. 721-743, 2013.

NASCIMENTO, Augusto. **Atlas da lusofonia: São Tomé e Príncipe**. Lisboa: Prefácio, 2008.

NASCIMENTO, Augusto. **Mutações sociais e políticas em São Tomé e Príncipe nos séculos XIX-XX: uma síntese interpretativa**. Comunicação apresentada no Centro Cultural Português de São Tomé e Príncipe. São Tomé e Príncipe, 2001.

NICOLIS, Marco; BIBERAUER, Theresa. The null subject parameter and correlating properties. **The limits of syntactic variation**, v. 132, 2008.

NUNES, Jairo. **O famigerado se: uma análise sincrônica e diacrônica das construções com se apassivador e indeterminador**. 1990. Dissertação (Mestrado) – UNICAMP, Campinas, 1990.

OLIVEIRA, Gabryella Fraga de. **Distribuição de sujeito nulo e pleno no português falado e escrito por universitários guineenses e são-tomenses da UNILAB-Ceará: interferências do português brasileiro?**. 2019. Dissertação de Mestrado. Universidade Federal de Pernambuco.

OLIVEIRA, Kércia Rosário Fiuza. **Sujeito nulo ou sujeito preenchido?** uma análise da variedade falada pela comunidade quilombola Rio das Rãs–Bahia. Dissertação (Mestrado). Repositório Digital de Teses e Dissertações do PPGLin-UESB, v. 8, p. 104 p.-104 p., 2020.

OLIVEIRA, Victor. **A expressão do sujeito no português de Moçambique**. Dissertação (Mestrado). Unicamp, 2016.

OMENA, Nelize. **Pronome pessoal de terceira pessoa: suas formas variantes em função acusativa**. 1978. Dissertação (Mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1978.

PAGOTTO, Emilio. **A posição dos clíticos em português: um estudo diacrônico**. 1992. Dissertação (Mestrado) – UNICAMP, Campinas, 1992.

PAIVA, Maria da Conceição. A variável gênero/sexo. *In*: MOLLICA, Maria Cecília; BRAGA, Maria Luíza (Orgs.). **Introdução à Sociolinguística**. São Paulo: Contexto, 2004, p. 33-42.

PAIVA, Maria da Conceição. Mudança em tempo real e em tempo aparente. *In*: MOLLICA, Maria Cecília; FERRAREZI, Celso (Orgs.). **Sociolinguística, sociolinguísticas: uma introdução**. São Paulo: Contexto, p. 23-32, 2016.

PARKVALL, Mikael. **Da África para o Atlântico**. Editora da UNICAMP, 2012.

PENKAL, Loremi Loregian. Pronomes pessoais: conceituação versus uso. **Analecta**, v. 7, n. 1, p. 71-83, 2006.

PERLMUTTER, David. **Deep and surface structure constraints in syntax**. Tese de Doutorado. Massachusetts Institute of Technology, 1968.

PINKER, Steven, **O instinto da linguagem: como a mente cria a linguagem**. São Paulo: Martins Fontes, 2002.

PINTZUK, Susan. **Programas Varbrul**. Rio de Janeiro-RJ, UFRJ, 1988.

RAMOS, Jânia. “Sociolinguística Paramétrica” ou “Variação Paramétrica”? *In*: DA HORA, Dermeval.; CHRISTIANO, Elisabeth. **Estudos linguísticos: realidade brasileira**. João Pessoa: Ideia, p. 83-94, 1999.

RAMOS, Jânia. **Marcação de caso e mudança sintática no português do Brasil: uma abordagem gerativa e variacionista**. 1992. Tese (Doutorado) – Instituto de Estudos da Linguagem, UNICAMP, Campinas, 1992.

RAPOSO, Eduardo. **Teoria da gramática: A faculdade da linguagem**. Lisboa: Ed. Caminho, 1992.

RIZZI, Luigi. **Issues in Italian syntax**. Dordrecht: Foris, 1982.

ROBERTS, Ian. **Diachronic syntax**. Oxford University Press, 2007.

ROCHA, Fabiane; NASCIMENTO, Teixeira. São Tomé e Príncipe: aspectos históricos, econômico-sociais e linguísticos. *In*: BRANDÃO, Sílvia (org.). **Dois variedades africanas do Português: variáveis fonético-fonológicas e morfossintáticas**. São Paulo: Blucher, p. 43-74, 2018.

ROUGÉ, Jean-Louis. A inexistência de crioulo no Brasil. *In*: FIORIN, José Luiz; PETTER, Margarida Maria Taddoni (Orgs.). **África no Brasil: a formação da língua portuguesa**. Editora Contexto, 2009, p. 63-73.

ROUGÉ, Jean-Louis. Les langues des Tonga. **Actas do colóquio sobre crioulos de base lexical portuguesa**. Lisboa: Colibri, p. 171-176, 1992.

SANKOFF, David. Sociolinguistic method and linguistic theory. *In*: COHEN, L. *et al* (Orgs.). **Logic, Methodology and Philosophy of Science VI**. Amsterdam: North-Holland, p. 677-689, 1982.

SANKOFF, David. Variable rules. *In*: AMMON, Ulrich; DITTMAR, Norbert; MATTHEIR, Klaus (Orgs.). **Sociolinguistics: an international handbook of the science of language and society**. Berlin; New York: Walter de Gruyter, 1988, p. 984-998.

SANKOFF, David; TAGLIAMONTE, Sali; SMITH, Eric. **Goldvarb X: a variable rule application for Macintosh and Windows**. Department of Linguistics, University of Toronto, 2005.

SANTANA, Natali Gomes de Almeida. **As construções dativas no português de duas comunidades bilíngues de São Tomé (África)**. 2019. 110 f. Dissertação (Mestrado em Língua e Cultura) – Universidade Federal da Bahia, 2019.

SANTIAGO, Ana Maria; AGOSTINHO, Ana Livia. Situação linguística do português em São Tomé e Príncipe. **A cor das Letras**, v. 21, n. 1, p. 39-61, 2020.

SANTOS, Maciel Morais. Um episódio do slave cocoa: o relatório Nightingale de 1906. *In*: RIBEIRO, Jorge Martins; SILVA, Francisco Ribeiro (Orgs.). **Estudos em homenagem a Luis Antonio de Oliveira Ramos**. Porto: Faculdade de Letras da Universidade do Porto, p. 965-983, 2014.

SCHERRE, Maria Marta Pereira; NARO, Anthony Julius. Duas dimensões do paralelismo formal na concordância verbal no português popular do Brasil. **DELTA: Documentação e Estudos em Linguística Teórica e Aplicada**, v. 9, n. 1, 1993.

SEIBERT, Gerhard. Colonialismo em São Tomé e Príncipe: hierarquização, classificação e segregação da vida social. **Anuário Antropológico**, v. 40, n. 2, p. 99-120, 2015.

SEIBERT, Gerhard. Crioulização em Cabo Verde e São Tomé e Príncipe: divergências históricas e identitárias. **Afro-Ásia**, p. 41-70, 2014.

SEIBERT, Gerhard. Os angolares da Ilha de São Tomé: náufragos, autóctones ou quilombolas?. **Textos de História**, v. 12, n. 1-2, p. 43-64, 2004.

SILVA LOPES, Norma; BAXTER, Alan Norman. A concordância verbal variável no português dos Tongas. **PAPIA-Revista Brasileira de Estudos do Contato Linguístico**, v. 21, n. 1, p. 39-50, 2011.

SILVA, Daniel. O tráfico de São Tomé e Príncipe, 1799 a 1811: para o estudo de rotas negreiras subsidiárias ao comércio transatlântico de escravos. **Estudos de História: Revista do Programa de Pós-Graduação em História**, v. 9, n. 2, p. 35-51, 2002.

SILVA, Humberto Soares da. Sobre o alcance da sociolinguística no estudo da mudança paramétrica: uma perspectiva interlinguística. **Working Papers em Linguística**, v. 14, n. 2, p. 47-65, 2013.

SILVA, Vera Lúcia Paredes. Motivações funcionais no uso do sujeito pronominal: uma análise em tempo real. *In*: PAIVA, Maria da Conceição; DUARTE, Maria Eugênia. **Mudança linguística em tempo real**. Rio de Janeiro: Contra Capa, 2003, p. 97-114.

SILVA, Wellington Santos da. **A Língua Geral de Mina e o Ciclo do Ouro**: um capítulo da história dos contatos no Brasil. 2020. 310 f. Tese (Doutorado) - Universidade de São Paulo, 2020.

SMITH, Ian. Siri Lanka Creole Portuguese phonology. **International Journal of Davidian Linguistics**, n.7, p.247-406. 1977.

SNOW, Catherine. Relevance of the notion of a critical period to language acquisition. **Sensitive periods in development: Interdisciplinary perspectives**, p. 183-209, 1987.

TAGLIAMONTE, SALI. **Analysing sociolinguistic variation**. New York: Cambridge University Press, 2006.

TARALLO, Fernando. **A pesquisa sociolinguística**. São Paulo, Ática, 1994.

TARALLO, Fernando. Por uma Sociolinguística Românica “Paramétrica”: Fonologia e Sintaxe. **Ensaio de Linguística**, UFMG, v. 13, p. 51-84, 1987.

TARALLO, Fernando. Por uma sociolinguística românica “paramétrica”: fonologia e sintaxe. **Cadernos de Linguística e Teoria da Literatura**, n. 13, p. 53-85, 1985.

TARALLO, Fernando; KATO, Mary A. Harmonia trans-sistêmica: variação intra- e inter-linguística. **Revista Diadorim**, Campinas, p. 315-353, 1989.

VANIN, Aline Aver. Considerações relevantes sobre definições de comunidade de fala. **Language and Culture**, v. 31, n. 2, p. 147-153, 2009.

VELUPILLAI, Viveka. **Pidgins, creoles and mixed languages**: An introduction. John Benjamins Publishing Company, 2015.

VERÍSSIMO, Victor. A evolução do conceito de parâmetro do sujeito nulo. **Entrepalavras**, v. 7, n. 1, p. 76-90, 2017.

VICENTE, Marta Gonçalves. (Re) viver as roças: reabilitação participada das roças de São Tomé e Príncipe. **VIII Seminário Internacional de Investigación en Urbanismo, Barcelona-Balneário Camboriú**, 2016.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. **Empirical foundations for a theory of language change**. University of Texas Press, 1968 (2006).

XAVIER, Gildete. **Português brasileiro como segunda língua**: um estudo sobre o sujeito nulo. 2006. 213 f. Tese (Doutorado em Linguística) - Instituto de Estudos da Linguagem da Universidade Estadual de Campinas, Capinas, 2006.