



**Universidade do Estado do Rio de Janeiro**

Centro de Educação e Humanidades

Instituto de Letras

Elizabeth Aparecida Hautz

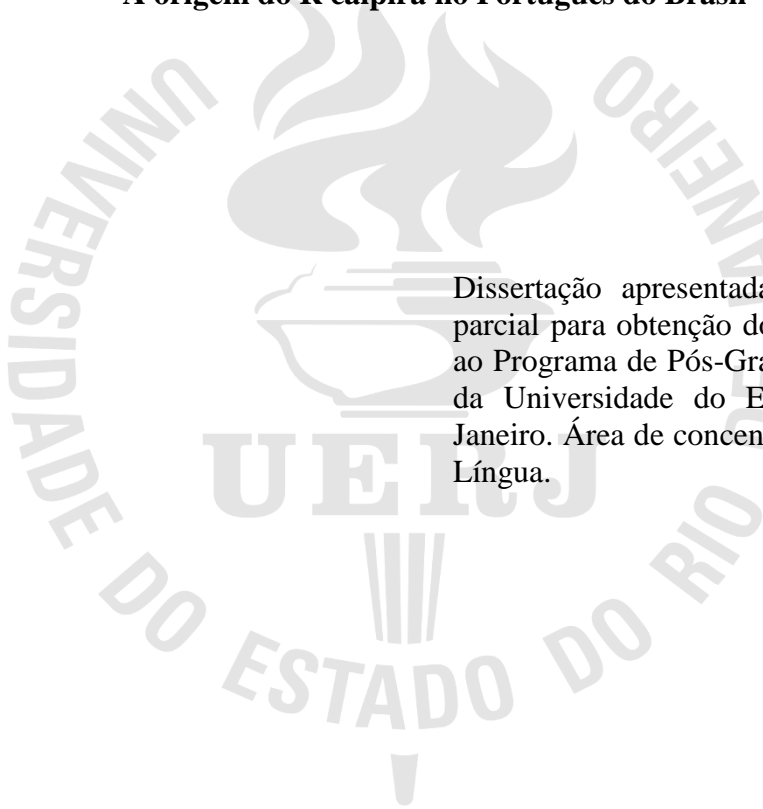
**A origem do R caipira no Português do Brasil**

Rio de Janeiro

2021

Elizabeth Aparecida Hautz

**A origem do R caipira no Português do Brasil**



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Letras da Universidade do Estado do Rio de Janeiro. Área de concentração: Estudos de Língua.

Orientadora: Prof<sup>ª</sup> Dra. Denise Salim Santos

Rio de Janeiro

2021

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/BIBLIOTECA CEH/B

H382 Hautz, Elizabeth Aparecida.  
A origem do R caipira no Português do Brasil / Elizabeth Aparecida  
Hautz. – 2021.  
94 f.: il.

Orientadora: Denise Salim Santos.  
Dissertação (mestrado) – Universidade do Estado do Rio de Janeiro,  
Instituto de Letras.

1. Língua portuguesa – Dialeto - Brasil - Teses. 2. Língua portuguesa -  
Brasil – Teses. 3. Língua portuguesa – Brasil - História – Teses. 4.  
Linguística – Teses. I. Santos, Denise Salim. II. Universidade do Estado do  
Rio de Janeiro. Instituto de Letras. III. Título.

CDU 806.90-087(81)

Bibliotecária: Mirna Lindenbaum. CRB7 4916

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta  
dissertação, desde que citada a fonte.

---

Assinatura

---

Data

Elizabeth Aparecida Hautz

**A origem do R caipira no Português do Brasil**

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Letras da Universidade do Estado do Rio de Janeiro. Área de concentração: Estudos de Língua.

Aprovada em 28 de outubro de 2021.

Banca Examinadora:

---

Prof<sup>a</sup> Dra. Denise Salim Santos (Orientadora)  
Instituto de Letras – UERJ

---

Prof<sup>a</sup> Dra. Maria Teresa Gonçalves Pereira  
Instituto de Letras – UERJ

---

Prof<sup>a</sup> Dra. Lucia Deborah Ramos de Araujo  
Colégio Pedro II

Rio de Janeiro

2021

## DEDICATÓRIA

Sou caipira, pira, pora, nossa... Senhora de Aparecida, que ilumina a mina escura e funda, o  
trem da minha vida  
(ROMARIA, *Renato Teixeira*)

## AGRADECIMENTOS

Aos meus saudosos caipiras mais amados: meus pais José Carlos Hautz (Sô Nei) e Ivone Maria de Souza Hautz (dona Ivone); aos caipiras da minha querida Guaxupé e a todos os caipiras desse mundão véio sem *portera*.



Sinto-me uma pessoa abençoada porque tenho tantas pessoas a quem deveria agradecer, que o espaço aqui não seria suficiente.

Gostaria de citar o trio parada dura sem o qual essa empreitada não teria sido concluída com êxito: professora Denise Salim Santos, minha orientadora sempre tão sábia e gentil e que me aceitou como orientada sem titubear; professora Maria Teresa Gonçalves Pereira, que lutou por mim como uma leoa quando tudo me levava a desistir; professora Tania Maria Nunes de Lima Camara por toda sua diplomacia e apoio que renderam tão bons frutos.

Ao querido professor Luiz Fernando DiasPita, pelo apoio na publicação dos meus artigos e por sempre acreditar em mim.

Às professoras Maria Teresa Gonçalves Pereira, Tania Maria Nunes de Lima Camara Lucia Deborah Araujo, por aceitarem participar desta banca e trazer contribuições à pesquisa realizada.

Aos colegas da Pós-Graduação em Letras da UERJ que estiveram o tempo todo me incentivando, em especial Cláudio Corrêa (amizade incondicional) e Helena Langoni (a temperança das horas difíceis) e aos colegas da USP, Rita de Cássia Varoli, Fábio Seiji Takaki e Fernanda Ferreira Dória.

Aos meus meninos, meus queridos filhos, João Paulo e Gustavo (vocês são a luz dos meus olhos!) e a Robson, obrigada! Assim como ao meu irmão Darlei, ao meu irmão posição Gilmar Dias Marinho e aos compadres Lúcio e Sueli Freitas.

Meu eterno agradecimento àquela que me iniciou nas primeiras letras, minha professora primária, Dona Azélia Lopes Correia, que deixou esse mundo tão cedo e a todos os outros queridos professores com quem tive o prazer desde então.

Por último, mas não menos importante, meu agradecimento ao bondoso Deus e a Nossa Senhora Aparecida a quem fui consagrada na pia de batismo.

## RESUMO

HAUTZ, Elizabeth Aparecida. **A origem do R caipira no Português do Brasil**. 2021. 94 f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

Esta dissertação objetiva fazer uma abordagem sobre o fenômeno linguístico referente ao R retroflexo, vulgarmente conhecido como R caipira e algumas hipóteses quanto a sua origem. O R caipira ou retroflexo é considerado por muitos a principal marca distintiva do que chamamos de dialeto caipira que, segundo os estudos de Amaral (1920), corresponde à variação linguística presente no interior de São Paulo, notadamente na região do Alto Tietê. Alguns autores, inclusive, apontam que o R caipira seria uma modalidade específica do R retroflexo, exclusiva da variedade linguística encontrada no interior do Brasil e marca distintiva do Português do Brasil. Conforme outros autores como Ribeiro (2016), tal fenômeno se estende por todo sul de Minas e de Mato Grosso, Norte do Paraná e interior de São Paulo. Há algumas teorias acerca da origem do R retroflexo no território brasileiro, entretanto, nenhuma das hipóteses encontra-se totalmente comprovada. Considerando que tal objeto de pesquisa está intimamente relacionado às questões de evolução do próprio Português do Brasil, uma releitura do histórico e da evolução da língua portuguesa trazida pelos colonizadores, bem como das línguas que com ela coexistiram durante a formação do Brasil (indígenas e africanas), se faz necessária. Assim, com base não apenas, mas especialmente nos estudos de Faraco (2016), Noll (2008), Naro & Scherre (2007), revisitamos essa evolução e apresentaremos algumas hipóteses existentes, até o momento, que tentam elucidar a origem do R caipira. Trata-se de um fenômeno de significativa relevância que se mantém vivo, contrariando os augúrios que previam sua extinção.

Palavras-chave: Português do Brasil. Dialeto caipira. R caipira.



## ABSTRACT

HAUTZ, Elizabeth Aparecida. **The origin of R caipira in Brazilian Portuguese**. 2021. 94 f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

This dissertation aims to approach the linguistic phenomenon related to the retroflex R, commonly known as R caipira and some hypotheses as to its origin. The *caipira* R or retroflex is considered by many to be the main distinguishing mark of what we call the *caipira* dialect which, according to studies by Amaral (1920), corresponds to the linguistic variation present in the interior of São Paulo, notably in the Alto Tietê region. Some authors even point out that the caipira R would be a specific modality of the retroflex R, exclusive to the linguistic variety, found in the interior of Brazil and a distinctive mark of Brazilian Portuguese. According to other authors such as Ribeiro (2016), this phenomenon extends throughout the south of Minas and Mato Grosso, North of Paraná and interior of São Paulo. There are some theories about the origin of retroflex R in the Brazilian territory, however, none of the hypotheses find itself fully proven. Considering that such object of research is closely related to issues of evolution of Brazilian Portuguese itself, a rereading of the history and evolution of the Portuguese language brought by the colonizers, as well as the languages that coexisted with it during the formation of Brazil (indigenous and Africans), if necessary. Thus, based not only but especially on studies by Faraco (2016), Noll (2008), Naro & Scherre (2007), we revisit this evolution as well as present some existing hypotheses so far that try to elucidate the origin of R caipira. It is a phenomenon of significant relevance that is still alive, contrary to the auspices that predicted its extinction.

Keywords: Brazilian portuguese. “Caipira” dialect. “caipira R”.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Mapa linguístico apresentando as variantes de “tangerina” (tangerina, mexerica, laranja cravo, tanja, poncã, mimosa, vergamota/bergamota) registradas nas 25 capitais brasileiras.....	20
Figura 2 - Mapa apresentando a configuração da província romana de Hispania em meados do século III d.C.....	25
Figura 3 - Mapa apresentando as línguas da Península ibérica por volta de 1300 d.C.....	26
Figura 4 - Mapa da Presença do retroflexo em determinadas áreas do Brasil.....	48
Figura 5 - Mapa da distribuição dos povos Tupi e Tapuia no litoral do Brasil, às vésperas do colonialismo do século XVI.....	55
Figura 6 - Mapa da escravidão nas Américas (locais de origem dos principais fluxos do tráfico) .....	71
Gráfico 1 - Taxa de urbanização brasileira no período compreendido entre 1940-2010.....	36
Gráfico 2 - Proporção da população residente, por situação do domicílio (1950-2010).....	36
Quadro 1 - Quadro de símbolos fonéticos consonantais relevantes para a transcrição do português.....	46
Quadro 2 - Quadro das consoantes conforme Cagliari (1981) .....	49
Quadro 3 - Ramificações do tronco linguístico Tupi. ....	58
Quadro 4 - Ramificações do tronco linguístico Macro-Jê.....	66
Quadro 5 - Números de escravos entrados no Brasil.....	70
Quadro 6 - Quadro fonética da consonância simples do Quimbundo segundo ponto de articulação na boca.....	74
Quadro 7 - Censos da distribuição étnica da população brasileira entre os séculos XVI e XIX.....	76
Quadro 8 - Evolução da população de brasileiros nativos entre os séculos XVI e XIX.....	84

## SUMÁRIO

	<b>INTRODUÇÃO</b> .....	10
1	<b>SOCIOLINGUÍSTICA E DIALETOLOGIA</b> .....	13
1.1.	<b>Dialetos no Brasil: eis uma questão</b> .....	16
1.2.	<b>Estudos dialetais no Brasil</b> .....	18
2	<b>HISTÓRIA, FORMAÇÃO E EVOLUÇÃO DO PORTUGUÊS DO BRASIL</b> .....	22
2.1.	<b>Formação histórica do Português Europeu</b> .....	23
2.2.	<b>Formação histórica do Português do Brasil</b> .....	30
3	<b>O DIALETO CAIPIRA</b> .....	38
3.1.	<b>Amadeu Amaral e o livro <i>O Dialeto Caipira</i></b> .....	40
3.2.	<b>Características fonéticas presentes em <i>O Dialeto Caipira</i></b> .....	43
4	<b>OS RÓTICOS E O RETROFLEXO</b> .....	45
5	<b>A ORIGEM DO R CAIPIRA</b> .....	51
5.1.	<b>Hipótese de natureza externa: produto do contato entre línguas</b> .....	52
5.1.1.	<b><u>A diversidade das línguas indígenas e a Língua Geral</u></b> .....	52
5.1.2	<b><u>Características linguístico-estruturais do Tupi Guarani</u></b> .....	60
5.1.3	<b><u>A Língua Geral Paulista</u></b> .....	61
5.1.4	<b><u>O Tronco Macro-Jê</u></b> .....	64
5.1.5	<b><u>As Línguas Africanas</u></b> .....	68
5.2.	<b>Hipótese de natureza interna</b> .....	77
6	<b>CRIOLIZAÇÃO E DERIVA</b> .....	81
	<b>CONSIDERAÇÕES FINAIS</b> .....	86
	<b>REFERÊNCIAS</b> .....	89

## INTRODUÇÃO

O dialeto caipira corresponde ao falar ou variação linguística observados sobretudo no interior de São Paulo, sul de Minas e de Mato Grosso e norte do Paraná, assim como em regiões de Goiás e de Mato Grosso bem como em áreas interioranas do Rio de Janeiro e do Espírito Santo. Amadeu Amaral dedicou-se ao estudo dessa variedade, conforme podemos verificar logo na introdução de sua obra *O Dialeto Caipira*, quando faz menção à existência de um dialeto peculiar encontrado no território da antiga província de São Paulo, o qual era falado pela maioria da população daquela região, estendendo-se, inclusive, ao falar da própria minoria culta.

Dentre as várias características do dialeto caipira levantadas por Amaral, o autor apresenta-nos uma descrição detalhada da realização de um rótico<sup>1</sup>, ao qual chama de R caipira, descrito como um R inter e pós-vocálico bastante peculiar e que, ao ouvido, se assemelhava bastante ao R inglês pós-vocálico.

Associar o R caipira exclusivamente ao falar do interior do Estado de São Paulo, entretanto, é um equívoco geográfico e histórico, pois embora o R retroflexo seja uma marca do falar do “matuto” paulista, estudos linguísticos identificaram que tal variante seguia as rotas dos bandeirantes paulistas em busca de ouro, encontrada em cidades de Minas Gerais, Mato Grosso, Mato Grosso do Sul, Paraná de oeste de Santa Catarina e do Rio Grande do Sul, e até mesmo em regiões interioranas do Rio de Janeiro e do Espírito Santo.

Ainda que o R caipira pareça seguir a trilha dos bandeirantes, tal fato serve mais como explicação para sua disseminação e não necessariamente com sua origem. De fato, existem diferentes hipóteses que se empenham em explicar a origem desse fenômeno linguístico, como veremos a seguir.

O R caipira, sendo resultado de uma mudança interna da língua ou fruto da influência dos povos que aqui se encontraram com o português, teria sua origem africana ou indígena? Existe uma explicação para a existência do R retroflexo na língua portuguesa, uma explicação linguística para a presença do R caipira no falar de algumas regiões do Brasil? Essas são perguntas para as quais procuramos respostas nesta pesquisa.

---

<sup>1</sup> Fonema da classe dos vibrantes, “Rótico” significa um fonema que é representado em línguas que usam o alfabeto latino pela letra “r”. É um termo genérico cujo sentido varia de língua para língua.

Assim como nas demais controvérsias referentes às marcas do Português do Brasil (PB), também a questão da origem do R caipira apresenta confrontos entre os defensores de tendência estruturalista, que tendem a explicar os fenômenos de variação linguísticas como consequência de uma evolução interna e natural da língua, e os adeptos de vertentes que privilegiam um olhar sociolinguístico ao explicar tais fenômenos.

Dessa forma, o objetivo geral do trabalho é revisitar a história da Língua Portuguesa, buscando identificar, com base na literatura especializada, notadamente, a produzida por Silva Neto (1950), Melo (1975), Elia (1979), Basso & Gonçalves (2014) a possível origem do R retroflexo. Temos então como objetivo específico trazer à baila as teorias e as hipóteses defendidas e propostas por representantes das diferentes vertentes, como Borba (1908), Amaral (1920), Rodrigues (1986), Head (1986), Dietrich (2010), Mendonça (2012), dentre outros, que tentam explicar a origem do R caipira bem como apresentar as lacunas identificadas por aqueles que refutam tais explicações.

Nossa proposta de trabalho assim se apresenta: primeiramente, discutiremos alguns pontos centrais acerca da conceituação teórica referente aos domínios da Sociolinguística e da Dialectologia que representam significativa parte da fundamentação teórica e metodológica dessa pesquisa, com base nos trabalhos de Benveniste (2005), Barthes (1988) e Coseriu (1982).

Em seguida, no capítulo 2, nos dedicamos à história, à formação e à evolução do Português do Brasil (doravante PB), visto que a marca linguística cuja origem tentamos analisar é tida como presente apenas no domínio dessa variante, formada a partir do Português Europeu (PE), conforme os estudos de Teyssier (1997), Ilari & Basso (2006), Noll (2008) Faraco (2016).

No capítulo 3 é feita a apresentação do que vem a ser o “dialeto caipira”, suas principais características, não apenas linguísticas e estruturais, mas também sociológicas e históricas, bem como no capítulo 4, apresentamos uma abordagem sobre os róticos, baseada principalmente nas ideias de Lehiste (1962), Lindau (1985), Calou, Moraes & Leite (1998) e Brandão (2007). Acreditamos que tal análise seja importante dado que o R retroflexo parece apresentar-se como a mais significativa marca distintiva atribuída ao dialeto caipira.

O capítulo 5 trata das principais hipóteses e teorias formuladas acerca da origem do fenômeno linguístico em comento sob o viés sociolinguístico e sob o viés estruturalista. Nele tratamos da influência das línguas indígenas e africanas como elemento desencadeador das

mudanças ocorridas no Português do Brasil. Também apresentamos a hipótese de variação de origem interna defendida principalmente por Head (1987).

No capítulo 6 analisamos a questão extremamente controversa acerca das teorias da diferenciação do PB como resultado de um processo de crioulização ou de deriva espontânea, bem como a hipótese que tenta abarcar a duas anteriores, isto é, a origem do R caipira como possível resultado de um processo de crioulização seguido de descrioulização. Este capítulo é fundamentado pelos estudos de Faraco (2016), Basso & Gonçalves (2014), Naro & Scherre (2007).

Com a revisão bibliográfica que nos propusemos realizar, espera-se reiterar a necessidade de estudos adicionais sobre tema tão abrangente e instigante, dada a importância dos estudos dialetológicos para o entendimento das circunstâncias e variáveis envolvidas no surgimento do R retroflexo como marca exclusiva do Português do Brasil.

## 1 SOCIOLINGÜÍSTICA E DIALETOLOGIA

Dada a indissociabilidade entre os fenômenos da linguagem e o contexto social, acreditamos que nosso estudo, predominantemente dialetológico, deve ser subsidiado pelas contribuições da Sociolinguística no campo da Dialetologia visto que, dentre os objetivos de estudo da Sociolinguística, temos a análise e a sistematização das variantes linguísticas usadas por uma mesma comunidade de fala.

Por muito tempo, os estudos linguísticos se concentraram em abordagens metodológicas formais de descrição da língua, enfatizando as relações internas dos fenômenos linguísticos e negligenciando os fatores extralinguísticos. Saussure, considerado o pai da Linguística moderna, apesar de considerar a língua como um fato social, imprime aos estudos linguísticos do século XX o caráter eminentemente estruturalista, do qual exclui toda consideração de natureza sócio-histórica e cultural na observação, descrição, análise e interpretação dos fenômenos linguísticos. O linguista genebrino instaura uma oposição entre Linguística Interna e Linguística Externa, ao afirmar que o estudo das características externas do linguístico, embora apreciável, não é indispensável ao conhecimento do organismo linguístico interno.

Com o tempo, essa cisão de perspectivas de estudos linguísticos, gerada pela rigidez paradigmática de Saussure, passou a ser contestada por diversas outras correntes linguísticas, como o Funcionalismo, a Sociolinguística, a Análise do Discurso, entre outras formas de abordagem. Estudiosos da língua passaram a considerar a história das línguas como inseparável da história da cultura e da sociedade.

Émile Benveniste destaca o caráter social do fenômeno linguístico, dado que, segundo essa visão, a cultura é essencialmente linguagem, o que impõe ao cientista da língua a adoção de uma nova objetividade, conforme destacado por Roland Barthes:

(...) Ele (Benveniste) não hesita em notar o nascimento de uma nova objetividade, imposta ao cientista pela natureza simbólica dos fenômenos culturais: longe de abandonar a língua no limiar da sociedade, como se ela não fosse mais do que um utensílio, afirma com esperança que é “a sociedade que começa a reconhecer-se como língua (BARTHES, 1988, p. 180).

Em meados da década de 1960, observa-se um acentuado crescimento dos estudos que procuram abordar a diversidade de relações entre língua e sociedade: os estudiosos da língua, passaram a perceber que muito das questões a respeito de mudanças de caráter idiomático não

poderia ser elucidado sem uma perspectiva mais abrangente que considerasse informações sobre os fatores sociais que afetam a fala.

A partir de 1964, a Sociolinguística surge como ciência num período marcado por pesquisas e estudos dedicados à relação entre linguagem e sociedade capitaneados por estudiosos expoentes como William Bright, William Labov, Dell Hymes, John Fischer, entre outros. Para a Sociolinguística, não se trata de um total abandono dos aspectos estruturais da língua ou descuido de uma análise linguística estruturada e criteriosa, mas de priorizar grupos humanos, ao invés da gramática por si só.

Denomina-se Teoria da Variação a análise Sociolinguística que se dedica ao estudo das variantes linguísticas numa dada comunidade; esse modelo teórico-metodológico estuda como se conduzem duas ou mais variantes que coexistem em uma dada língua até o momento em que, se assim o for, uma forma se sobrepõe e substitui a outra. A Sociolinguística considera especialmente como objeto de estudo essa variação, encarando-a como um princípio geral e universal e, portanto, passível de descrição e análise científica.

Podemos dizer que a Sociolinguística se constitui como uma ciência que traz consigo um posicionamento teórico proveniente tanto da Antropologia Linguística, quanto da Dialetoлогия. Os estudos da variação linguística têm-se desenvolvido bastante no Brasil, dada a sua heterogeneidade linguística e cultural; assistimos a um crescente número de significativos trabalhos de viés sociolinguístico presentes no mundo acadêmico.

O eixo teórico central do nosso trabalho se alicerça na Dialetoлогия, segmento da Linguística amparado por longa tradição e metodologia sólida que trata da identificação, descrição, interpretação e análise de fenômenos linguísticos num dado espaço geográfico, em especial os fenômenos relativos às diferenças entre os falares regionais, a origem dessas diferenciações, o fracionamento de uma língua, a padronização e a variação da língua em regiões de contato multilíngue.

A Dialetoлогия é uma disciplina que estuda as diferenças regionais de uma língua, ou dialetos, procurando descobrir e descrever as suas características, tentando “identificar áreas mais ou menos coesas, assim como determinar os fatores que levaram à sua formação” (FERREIRA *et al.*, 1996, p.480). Para esses autores, a descrição comparada das “formas de falar” em que uma língua se diversifica no espaço leva a que essa disciplina possa ser encarada como o “estudo cartográfico dos dialetos”, obedecendo para tanto a critérios rigorosos de análise e de representação das variantes geográficas.



Preocupa-se a Dialetoologia em identificar e descrever os diferentes usos de uma determinada língua, considerando a distribuição diatópica, sua natureza sociocultural e temporal. Surgida como ciência, em meados do século XIX, vem a materializar-se com a publicação do *Atlas linguistique de la France*, de Jules Gilliéron e Edmond Edmont, entre 1902 e 1910.

Segundo Moreira, amparado por Rector (1975), os dialetos passaram a objeto de estudo com uma frequência crescente e sob duas abordagens distintas: por um lado, elucidar os fenômenos do presente por meio da investigação do passado e das formas antigas do falar; por outro lado, numa segunda forma de abordagem, integrando-se a geografia às pesquisas dialetológicas pela instrumentalização cartográfica, o que permite desvelar os vínculos intrínsecos entre a vida social e cultural do homem em seu espaço natural. Assim, percebemos que a Dialetoologia estabelece uma relação dialógica com a etnografia, com a linguística histórica, bem como com a antropologia e outros campos interdisciplinares do saber.

A Dialetoologia tradicional ou monodimensional privilegia a relação diatópica, por meio do estudo e da documentação dos dialetos locais mais antigos. Historicamente, a Dialetoologia era considerada monodimensional, ou areal, pois considerava somente uma dimensão linguística da realidade, ou seja, o que importava era a localidade, deixando de lado aspectos como idade, sexo/gênero, escolaridade. Por sua vez, a Sociolinguística considerava os fatores extralinguísticos como sexo, escolaridade, idade, classe social em apenas uma comunidade de fala. Assim, a Dialetoologia analisava os aspectos linguísticos de maneira horizontal (preocupação com as diferentes áreas linguísticas), enquanto a Sociolinguística procedia a uma investigação de maneira vertical (preocupação com os fatores extralinguísticos).

As mudanças que ocorrem em uma língua, entretanto, não dependem simplesmente de características próprias do espaço geográfico limitante do fenômeno linguístico em análise, daí a importância também da Dialetoologia pluridimensional, pois, conforme Coseriu (1982, p.12):

A Dialetoologia tem como centro de interesse estudar as unidades sintópicas e sobretudo a diversidade diatópica, enquanto caberia à Sociolinguística o estudo das unidades sinstráticas e a diversidade diastrática, ficando com a Estilística a unidades sinfásicas e a diversidade diafásica.

Embora predominante a perspectiva diatópica nos estudos dialetológicos, faz-se necessária a busca por abordagens outras que adotem novas metodologias que ampliem o

escopo monodimensional da Dialetologia para uma atuação pluridimensional. Daí a importância do diálogo com outras áreas correlatas do saber tais como a Sociolinguística, a Antropologia, a Historiografia, dentre outras, pois a eleição de uma ou outra das possibilidades de abordagem é determinada pelo propósito do trabalho e pelas características de seu escopo de análise.

### 1.1. Dialetos no Brasil: eis uma questão

Por muito tempo defendeu-se de forma ferrenha a unidade da língua portuguesa no Brasil, como se a língua aqui falada apresentasse uma uniformidade surpreendente, mas, conforme advertem vários estudiosos da área, esse mito seria mais uma das marcas que compõem o imaginário e a mitologia do preconceito linguístico em nosso país, onde existe uma longa tradição de estudos filológicos e gramaticais baseada, durante muito tempo, no conceito distorcido da unidade linguística inamovível no Brasil.

Conforme bem pontuado por Faraco em sua obra *História Sociopolítica da Língua Portuguesa*:

O Brasil é, até agora, a única sociedade extra europeia em que a língua portuguesa se tornou a L1 da maioria absoluta da população. (...) Essa hegemonia não significa, claro, homogeneidade. Há, no Brasil, grande variedade regional (assinalada principalmente por pronúncias, prosódias e léxicos diferentes) (FARACO, 2016, p. 136).

Como enfatizado por Henriques (2019, p. 109), “a tese da unidade ou da diversidade da língua portuguesa no Brasil (teve?) muitos defensores”. Transcrevemos a seguir, nos apropriando da seleção feita por Henriques, a opinião de alguns especialistas no assunto:

É preciso ter na devida conta que *unidade* não é *igualdade*; no tecido linguístico brasileiro há, decerto, gradações de cores. Minucioso estudo de campo determinaria, com segurança, várias áreas. O que é certo, porém, é que o conjunto de falares brasileiros se coaduna com o princípio da *unidade na diversidade* e da *diversidade na unidade* (SILVA NETO, 1970, p. 632).

O que se conclui é que a língua comum aos povos brasileiro e português é a mesma (embora com traço coletivos nacionais), e que essa língua comum apresenta, no Brasil (aliás, igualmente em Portugal), notável unidade, de norte a sul e de leste a oeste, quer nas camadas cultas, quer nas populares (ELIA, 1979, p. 13)

É possível que, em fase do crescente entrecruzar-se entre campo e cidade, em algum futuro, a unidade, homogeneidade do Português do Brasil venha a ser a dominante,

mas, de fato, hoje, o que se documenta é uma intrincada e sutil diversidade dialetal (NEVES, 2004, p. 71)

O fato mais básico que salta aos olhos é muito simples: o português do Brasil sempre foi o português. No imenso território que atualmente ocupa como língua materna da grande nação brasileira, não se estabeleceu nenhum traço estrutural estranho à estrutura original com a qual a língua portuguesa aportou na América (NARO, 2004, p. 179)

Compartilhamos da opinião de Orlandi (1998), como também da opinião de Henriques (2019), ao ressaltar que, em vez de considerar uma oposição estrita entre unidade e diversidade, concebamos essa relação como uma relação necessária e dinâmica.

Vale acrescentar que a definição de dialeto passou, e ainda passa, por discussões e revisões conceituais. Considera-se dialeto a diferenciação regional de uma língua que pode ser fruto das características intrínsecas do processo original de colonização de uma dada região como também efeito de intercâmbios comunicativos e migracionais, além de outros fatores políticos, econômicos, educacionais etc.

É bastante complexo para a Dialectologia estabelecer uma acurada delimitação entre língua e dialeto. Muitos autores não identificam essa diferença ou a veem simplesmente como um caso de poder político, social ou econômico ou, nas palavras de Coseriu (1982: 11), tal diferenciação se resume na verdade a uma questão de *status* histórico. Para o autor, e para muitos outros que coadunam com a mesma visão, o termo dialeto se aplicaria a uma língua menor incluída em uma língua maior, ou seja, uma língua histórica (ou idioma oficial), que lhe deu origem.

Podem ser consideradas dialetos as variedades regionais que, em seu desenvolvimento histórico, se encontram em espaços geográficos distintos e não estão solidamente reguladas como as variedades padrão. De acordo com essa perspectiva, o dialeto seria uma variedade linguística regional ou social, suficientemente discernível, sem diferenciação semântica, mas apenas estilística, daí muitos estudiosos referirem-se a esse fenômeno como variedades linguísticas.

Dubois (1993) distingue dialeto e falar com base na extensão geográfica do alcance de cada um. Para o autor, o falar, por oposição ao dialeto, é um sistema de signos e regras combinatórias definido por um quadro geográfico restrito (uma aldeia, um grupo étnico específico etc. ). Por sua vez, o dialeto seria considerado relativamente coeso sobre uma área extensa e delimitada por critérios linguísticos da Dialectologia e da Geolinguística. Ainda segundo Henriques (2019, p. 129), “se a palavra dialeto não tivesse tido um dia a conotação

pejorativa que os povos (e os estudiosos) lhe deram, não teria sido necessário recorrer-se a uma palavra sem estigmas como falar, para dizer quase a mesma coisa”.

Diante do exposto, salientamos que, neste trabalho, consideramos os termos dialeto, linguajar, variedade linguística e falar como sinônimos.

## 1.2 Estudos dialetais no Brasil

Os estudos dialetais no Brasil remontam a 1826, quando o Visconde de Pedra Branca escreveu um artigo contrastivo entre o Português do Brasil e o Português de Portugal integrante da obra francesa *Introduction à l'Atlas Ethnographique du Globe*. Desde então, a Dialetoologia no Brasil presenciou um grande avanço e desenvolvimento.

Tradicionalmente, os estudos dialetológicos no Brasil são divididos em quatro fases ou períodos. A primeira fase ou fase lexical, compreende o período que vai de 1826 a 1920 e é marcada pelos trabalhos voltados ao estudo do léxico e de suas peculiaridades. Além da publicação de várias gramáticas, houve a elaboração de inúmeros dicionários, glossários e vocabulários regionais, entre os quais merecem destaque:

- *Dicionário da Língua Portuguesa*, de Luís Maria Silva Pinto, publicado em 1832;
- *Popularium Sulriograndense* e o *Dialeto Nacional*, de Apolinário Porto-Alegre, de 1872;
- *O Idioma Hodierno de Portugal Comparado com o do Brasil*, de Jorge Paranhos da Silva, de 1879 e que se destaca por ser o primeiro estudo de conteúdo gramatical publicado no Brasil;
- *O Tupi na Geografia Nacional*, de Theodoro Sampaio, de 1901;
- *Glossário Paraense*, de 1905 e o *Vocabulário Popular* (contendo regionalismos mineiros), de 1912;
- *A Língua Portuguesa no Brasil*, de Virgílio de Lemos, publicado em 1916.

A segunda fase, a fase dialetológica, se estende de 1920 a 1952 se inicia em 1920, com a publicação de *O Dialeto Caipira*, de Amadeu Amaral, considerado o primeiro estudo a destacar não apenas o léxico, mas também questões fonéticas, morfológicas e sintáticas. Essa fase é marcada pela publicação de estudos gramaticais de caráter geral, com o objetivo de descrever e analisar o Português do Brasil.

Também neste período há um volume considerável de trabalhos voltados para as variações linguísticas específicas de algumas regiões e para a contribuição africana, tais como:

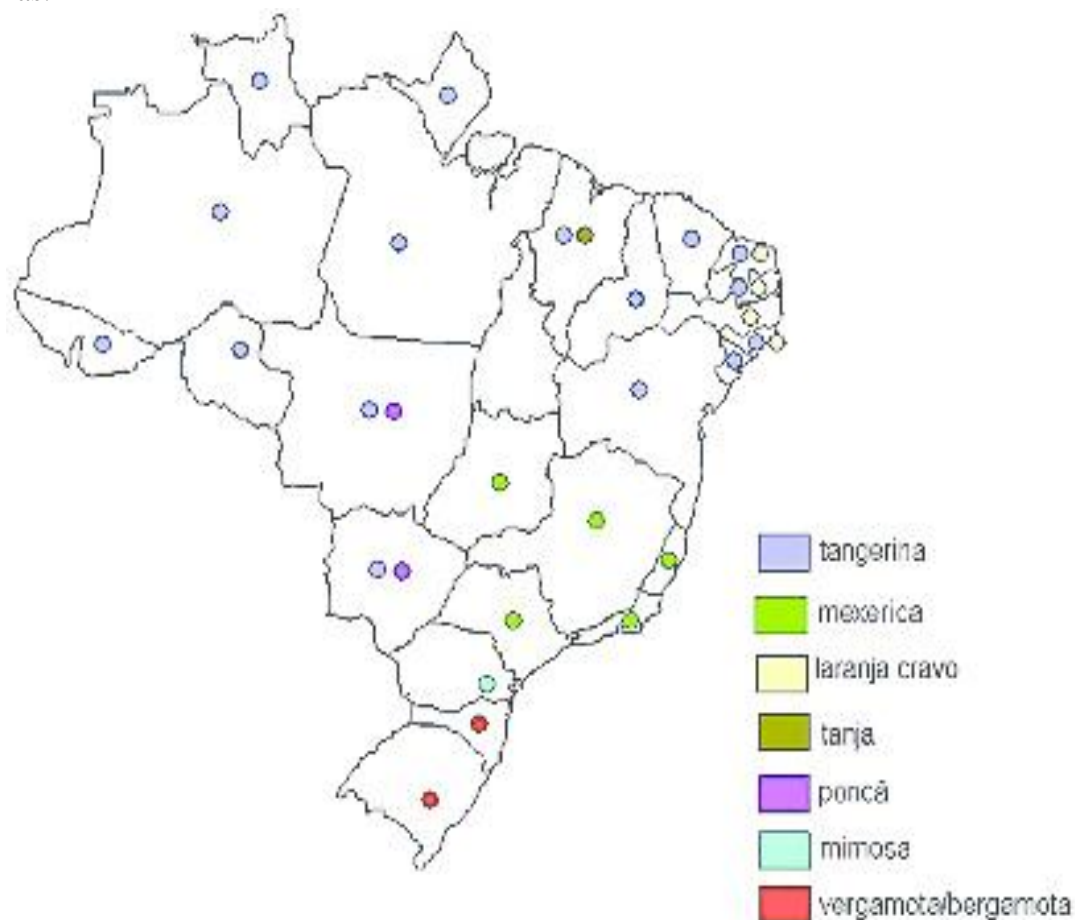
- 1922: *O Linguajar Carioca*, de Antenor Nascentes;
- 1926: *A Influência do Tupi no Português*, de Jacques Raimundo;
- 1933: *O Elemento AfroNegro na Língua Portuguesa*, de Jacques Raimundo e *A Influência Africana no Português do Brasil*, de Renato Mendonça;
- 1938: *O Falar Mineiro*, de José Aparecido Teixeira;
- 1944: *Estudos de Dialectologia Portuguesa: a linguagem de Goiás*, também de José Aparecido Teixeira;
- 1951: *A Linguagem Popular da Bahia*, de Edison Carneiro.

Além das obras citadas e outras similares, essa fase também foi profícua quanto à produção de glossários, vocabulários e léxicos regionais ou temáticos tais como o *Vocabulário Gaúcho*, de Roque Callage (1928); o *Vocabulário Pernambucano*, de Pereira da Costa (1937); o *Vocabulário Amazonense*, de Alfredo da Maia (1939), entre outros

A fase geolinguística inicia-se formalmente em 1952, com um ato do governo brasileiro, o Decreto n. 30.643, de 20 de março de 1952, que determinava a elaboração do *Atlas Linguístico do Brasil* pela Comissão de Filologia da Casa de Rui Barbosa e se estende até a criação do projeto Atlas Linguístico do Brasil (ALiB). Esse período foi marcado pela publicação de vários atlas mono ou bidimensionais, como os da Bahia, de Minas Gerais, da Paraíba, do Sergipe e do Paraná, que precederam a criação do projeto ALiB, ocorrida em 1996.

O que vem a ser um mapa linguístico? Recorremos à esclarecedora definição dada por Celso Cunha (1968, p. 53): “um atlas é um grande vocabulário dialetal que visa apresentar, de maneira científica e viva, sobre cada carta, as variantes de uma palavra ou de uma pequena frase em território mais ou menos vasto”. Para fins de ilustração e para melhor compreensão, reproduzimos abaixo um mapa linguístico apresentando as variantes de “tangerina” (tangerina, mexerica, laranja cravo, tanja, poncã, mimosa, vergamota/bergamota) registradas nas 25 capitais brasileiras.

Figura 1 – Mapa linguístico apresentando as variantes de “tangerina” (tangerina, mexerica, laranja cravo, tanja, poncã, mimosa, vergamota/bergamota) registradas nas 25 capitais brasileiras.



Fonte: ROMANO; AGUILERA, 2009.

A fase geolinguística contou com os trabalhos de nomes de peso como Serafim da Silva Neto, Celso Cunha, Nelson Rossi e Antenor Nascentes. Conforme enfatizado por HENRIQUES (2019), os trabalhos de Antenor Nascentes foram de suma importância nesse período, destacando-se a reedição em 1953 do seu *Linguajar Carioca*. Também merece destaque a obra *Guia para Estudos Dialectológicos*, de 1957, de Serafim da Silva Neto, lembrando que esse autor sempre ressaltou a importância da criação de “uma mentalidade dialectológica”, necessária para a criação de um ambiente favorável às pesquisas de campo. (1957, p. 9). Pioneiro na realização dos estudos linguístico-geográficos de um falar brasileiro, Nelson Rossi publicou o *Atlas Prévio dos Falares Baianos* (APFB), em 1963.

Com relação a Celso Cunha, segundo Henriques (2019), além de suas importantes obras relativas ao estudo do Português, foi também um dos protagonistas no processo de desenvolvimento do Projeto de Estudos da Norma Urbana Culta (NURC) que, por sua vez,

deu origem ao Projeto Gramática do Português Falado (PGPF) e o Projeto História do Português do Brasil (PHPB).

Ainda nessa fase foram publicados diversos glossários, vocabulários e léxicos regionais ou temáticos, como: *Novo Dicionário da Gíria Brasileira* de Manuel Viotti (1956), *Locuções Tradicionais no Brasil* de Câmara Cascudo (1970), *Dicionário de Termos e Expressões Populares* de Tomé Cabral (1973), *Dicionário Contrastivo Luso-Brasileiro* de Mauro Villar (1989), entre outros.

Finalmente, temos a chamada fase cartográfica nacional, de 1996 em diante. Essa fase se iniciou com a formação do comitê nacional para coordenar o Projeto ALiB. Sob o ponto de vista metodológico, os estudos deixam de ser apenas diatópicos e passam a analisar também outras variáveis em consonância com os atuais princípios da Dialectologia Pluridimensional. Especificamente a Dialectologia Pluridimensional e Relacional, que tem como precursores Radtke e Thun (1996), analisa as variedades linguísticas, os dialetos e as variedades em contato, a partir de diferentes dimensões e parâmetros (pluridimensional), promovendo relações entre elas (relacional). Vários projetos espalhados pelo país passam a mapear diversas marcas do PB em diferentes regiões com o objetivo de aprimorar a confecção de mapas linguísticos abrangentes.

Um mapa linguístico nos informará hoje uma fotografia sincrônica da realidade linguística de um determinado grupo ou espaço geográfico. Tão importante quanto essa visão sincrônica é o subsídio de uma análise em termos diacrônicos para um melhor entendimento dos fenômenos linguísticos; assim, acreditamos de extrema importância traçar, em linhas gerais, a evolução do Português do Brasil, como faremos no capítulo seguinte.

## 2 HISTÓRIA, FORMAÇÃO E EVOLUÇÃO DO PORTUGUÊS DO BRASIL

Compreender um dado idioma é também compreender sua história, visto que nenhuma língua é estática (com exceção das linguagens de programação – e mesmo estas são constantemente revistas e aprimoradas), ao contrário, a língua está sempre em movimento e compreender essa dinâmica envolve a apropriação de conhecimentos das diversas áreas do saber. Torna-se necessário que se faça um percurso multidisciplinar combinando quadros conceituais oriundos da Sociologia, Antropologia, dos Estudos Históricos sobre a Língua Portuguesa, da Sociolinguística e seus eventuais desdobramentos.

Como afirma Faraco (2019, p. 16), “as sociedades têm história. É no interior dessas totalidades que as línguas existem. Elas são, por consequência, também heterogêneas, contraditórias, simultaneamente integradas, fragmentadas e em constante devir.”. Segundo o autor, uma língua é “uma construção imaginária em que se mesclam fatos linguísticos com fatores históricos, políticos, sociais e culturais, ou seja, falantes de diferentes variedades linguísticas se reconhecem como falantes de uma mesma língua mesmo que haja poucas semelhanças léxico-gramaticais entre essas variedades, mas para Faraco:

(...) O contrário também ocorre, ou seja, variedades que, por critérios estritamente linguísticos, poderiam ser consideradas constitutivas de uma mesma língua, são assumidas por seus falantes como línguas diferentes por razões históricas, políticas e socioculturais. (...) Uma língua é, então, um conjunto de variedades (e só assim pode ser definida) distribuídas no espaço geográfico e social no eixo do tempo, conjunto que os falantes, por razões históricas, políticas e socioculturais, idealizam como uma realidade uma onde não há, efetivamente, unidade (FARACO, 2019, p. 35).

Ao estudar a história de uma língua há de se ter em mente que ela não nos fornece todas as respostas sobre a sua origem e constituição, porém não podemos negar que tanto a análise sincrônica quanto a diacrônica são instrumentos metodológicos essenciais para a explicação dos diversos fenômenos que tomam lugar num dado idioma.

Os estudos de Linguística Histórica ainda seguem, em linhas gerais, o duplo corte proposto por Ferdinand de Saussure (1857-1913) para contar a história de uma língua, ou seja, de um lado, a distinção que leva em conta o fator temporal (estudos sincrônicos *versus* estudos diacrônicos) e, por outro, a distinção que considera o próprio sistema da língua e seu meio externo (linguística interna e linguística externa).



Estudos posteriores levaram, progressivamente, a uma flexibilização dos cortes propostos por Saussure, enfatizando a necessidade do encaixe das mudanças linguísticas de ordem sistêmica interna na matriz mais ampla que abarca também suas condições externas bem como a consideração tanto dos aspectos sincrônicos quanto diacrônicos.

Estudar a variação linguística necessariamente passa por estudar não apenas a (s) língua (s), mas as culturas a ela (s) subjacentes, pois, segundo Faraco:

(...) a realidade cultural, tal como a realidade linguística, é caracterizada por pluralidade e heterogeneidade, mesmo que atrelada a um determinado estado-nação. Se o português brasileiro, por exemplo, é de fato um conjunto de variedades linguísticas, assim também temos de entender a expressão singular e genérica “cultura brasileira”: ela recobre um conjunto de culturas diferentes, de modos de vida diferentes, de diferentes respostas que os grupos e classes sociais dão às suas respectivas condições existenciais no espaço do estado-nação Brasil (FARACO, 2019, p. 35).

Acreditamos que o estudo da história do Português do Brasil (PB) é incompleto se não levamos em conta a análise dos principais fatos que tiveram impacto na formação do Português Europeu (PE) visto que, se a modalidade europeia não pode ser considerada como o principal componente na formação da variação brasileira (como muitos advogam), é inegável que em alguma medida nosso idioma recebeu considerável herança linguística dessa matriz.

## 2.1. Formação Histórica do Português Europeu

A língua mãe do Português Europeu se originou no noroeste da Península Ibérica numa região que corresponde atualmente ao norte de Portugal e à Galiza (integrante da Comunidade Autónoma da Espanha).

Ao tentarmos traçar a história de uma língua, o uso de algum tipo de periodização se faz necessário apesar da relativa arbitrariedade que toda periodização intrinsecamente possui. No caso da história do português, de acordo com os critérios utilizados por autores diversos, encontramos diferentes divisões cronológicas, sendo que alguns deles adicionam o período pré-literário, pré-histórico ou proto-histórico (geralmente desde a ocupação da Península Ibérica pelos romanos até meados do século XI) como parte integrante dessa cronologia. Exemplificando: a periodização que se apresenta como uma das mais utilizadas, conforme

podemos comprovar em Basso & Gonçalves (2014); Faraco (2016); Teyssier (1997) abrangeriam, com uma certa flexibilidade, as seguintes fases:

Primeiramente, temos o período denominado pré-literário ou proto-histórico que, resumidamente, corresponde à chegada do latim à Península Ibérica até o período imediatamente anterior aos primeiros textos em galego-português.

Os romanos chegaram à região que atualmente corresponde à Península Ibérica, por volta de 218 a.C., durante a Segunda Guerra Púnica contra o império cartaginês. Os cartagineses ocupavam parte da península, entretanto, os romanos os derrotaram na Espanha e iniciaram a ocupação do território. As conquistas se processaram de forma lenta e, até a pacificação completa da província romana da *Hispania*, foram necessárias várias campanhas até a completa estabilização durante o governo de Augusto, em 27 d. C.

Assim, conforme salientado por Basso & Gonçalves (2014), houve a gradual romanização das províncias como resultado desse laborioso processo de conquista romana onde temos que, nas províncias nas quais os romanos chegaram primeiro, os dialetos românicos desenvolvidos se mostraram mais conservadores do que os observados nas províncias cuja conquista romana teria sido mais tardia. Outro fator que contribuiu para a diferenciação do grau de romanização das províncias teria sido a distância desses territórios em relação a Roma bem como a dificuldade de acesso a algumas regiões.

Ainda com relação a essa diferenciação, ressaltamos a influência das diferentes línguas locais sobre a língua dos dominadores, conforme mencionado por Haury:

Assim, o latim levado para a Península Ibérica, por exemplo, em 197 a.C. mais ou menos, deve ter sido mais arcaico que o levado para a Dácia em 107 d.C. (...) Ao conjunto dos falares diversos dos povos vencidos e conquistados, cuja língua se infiltrou na do povo vencedor, dá-se o nome de *substrato linguístico*. (...) Na Lusitânia pré-romana foram os celtas o elemento de maior valor linguístico para a estruturação do português (HAURY, 2008, p. 27-28).

À época do imperador Diocleciano (244-311 d.C.), era assim a configuração de províncias nessa região:

Figura 2 - Mapa apresentando a configuração da província romana de Hispania em meados do século III d.C



Fonte: BASSO; GONÇALVES, 2014, p. 105

Ao longo dos séculos V, VI e VII, os visigodos invadiram a Península Ibérica, dominando grande parte da mesma, entretanto o latim permaneceu como a língua de cultura da península e os dialetos germânicos visigóticos deixaram suas marcas principalmente no léxico das línguas românicas que ali iriam se desenvolver e, conforme salientado por Basso & Gonçalves, com relação à constituição do galego-português:

Um dos reinos germânicos conquistados pelos visigodos ao longo desses séculos foi o do suevos, que constituíram um domínio cujas fronteiras eram quase coincidentes com as da anterior província da Gallaecia. A permanência deles ali, até a conquista dos visigodos, apenas em 574, permitiu um relativo isolamento linguístico que acabou por propiciar certos desenvolvimentos no latim vulgar dessa região, que depois viriam a diferenciar o romance falado ali, o galego-português – um estágio muito antigo do português, quando este ainda se confundia com o galego –, dos outros dialetos românicos da península. Esse isolamento teria grande influência no desenvolvimento do galego-português, pois algumas dessas mudanças, como as quedas de *l* e *n* intervocálicos (*luna* > *lua*, *soles* > *sóis*, por exemplo) e a mudança dos grupos *cl*, *pl* e *fl* para *ch* (como em *pluvia* > *chuva*, *clavis* > *chave*, por exemplo) já se manifestavam no período romano (por influência dos substratos linguísticos da região) (BASSO; GONÇALVES, 2014, p. 106).

O domínio visigodo mantém-se até o início do século VIII, quando os árabes muçulmanos invadem a península em 711 d.C. e acabam por dominar grande parte do território.

A invasão da península pelos muçulmanos e o processo de retomada da mesma pelos cristãos representam um momento crucial na história das línguas românicas da Península Ibérica, pois com a resistência dos reinos cristãos ao norte, o movimento da Reconquista precisou retomar todo o centro-sul num processo que levou sete séculos.

A chamada Reconquista, ou seja, a retomada lenta e gradual dos territórios da Península Ibérica pelos cristãos, se consolida apenas em 1492, com a conquista de Granada (último reduto mouro na península, região que hoje corresponde ao extremo sul da Espanha).

A Reconquista leva paulatinamente os dialetos românicos para o sul da península, definindo as fronteiras das línguas portuguesa, espanhola e catalã. Assim, pode-se dizer que, por volta do século XIII, o desenho político da península já apontava para a distribuição atual das línguas românicas, como pode ser observado no mapa a seguir.

Figura 3 – Mapa apresentando as línguas da Península ibérica por volta de 1300 d.C.



Fonte: <https://www.slideshare.net/amandagbento/do-latim-implcita-africanizao-do-portugus>

Além dos movimentos linguísticos descritos, é importante ressaltar que a Reconquista foi um período extremamente importante para o estabelecimento das unidades políticas da península, como o estado monárquico português. Sob o comando de Dom Afonso Henriques, Portugal passa a ser independente da Galiza e reconhecido pelo papa em 1173. Dom Afonso III continua a expansão em direção ao sul e, em 1250, conquista o Algarve, fixando as fronteiras atuais de Portugal.

A seguir, temos o que chamamos de período arcaico, que vai do nascimento da Língua Portuguesa, considerando como critério de início as datas prováveis dos textos mais antigos escritos em português, na verdade ainda galego-português, ou seja, fins do século XII e início do século XIII até o início das grandes navegações portuguesas, em torno de 1415 (data da tomada de Ceuta, no norte da África, pelos portugueses. Esse período apresenta uma fase pré-literária até finais do século XII e uma fase literária ou trovadoresca. Muitas mudanças foram identificadas no período, tendo-se como fonte textos reconhecidos e estudados como parte dele, tais como *Carta da Fundação da Igreja de Lardosa* (882), cuja escrita apresenta-se como uma mistura de latim vulgar e galego-português, e textos escritos em português arcaico propriamente dito, entre fins do século XII e início do século XIII (a *Notícia do Torto*, a *Demanda do Santo Graal*, a *Notícia de Fiadores*, o *Testamento de Afonso II*).

Durante esse período, algumas inovações fonéticas específicas do noroeste da Península Ibérica acabam por transformar os dialetos do romance ali falados bastante diferentes com relação aos dialetos do sul moçárabe e em relação aos dialetos do centro-norte da península (leonês e castelhano). Essas diferenciações começam a ocorrer durante a passagem do latim vulgar para o romance e terão grande influência nos desenvolvimentos posteriores do português.

Como dito anteriormente, não existe consenso absoluto entre as diversas periodizações propostas por diferentes autores, conforme informado por ILARI & BASSO (2006). Para Leite de Vasconcelos, esse período, por ele denominado de Português arcaico iria de 1216 até 1385-1412; para Serafim da Silva Neto, é o correspondente ao Português Trovadoresco e se estenderia de 1216 à 1420; Pillar Vázquez Cuesta o denomina de Galego-Português compreendendo o período de 1216 até 1385/1420 enquanto Luís Felipe Lindley-Cintra prefere nomear esse período como Português antigo. Para Haury:

Quando Portugal se separou da Galiza, no século XII, era o galego-português o idioma falado em toda a região da Galiza e da nascente nação portuguesa e, por três séculos ainda, foi o veículo da produção poética trovadoresca em toda a Península Ibérica. (...) A educação dos reis e membros da nobreza dos reinos em formação

passou a ser feita na Galiza, que se tornou também o centro irradiador da produção poética trovadoresca. (...) Além disso, esse foi, durante o período da Reconquista, o idioma de comunicação com as populações mouras e moçárabes (...) a língua trazida do Norte e adotada durante três séculos pela nação portuguesa sofreu gradativamente uma significativa transformação, acelerada ainda pela decisiva influência dos dialetos moçárabes e pelos fatos políticos do prestígio de Lisboa, até adquirir, no século XIV feição distinta que a caracterizou como língua portuguesa (HAUY, 2008, p. 42).

O galego-português é a língua da primitiva poesia lírica da Península Ibérica e foi conservada basicamente em três compilações: o *Cancioneiro da Ajuda* (a única compilação organizada ao tempo dos Trovadores, copiado em fins do século XIII ou princípios do século XIV); o *Cancioneiro da Vaticana* e o *Cancioneiro da Biblioteca Nacional de Lisboa* (ambos copiados na Itália, ao que tudo indica, nos primeiros anos do século XVI). Tais compilações são escritas numa língua que tem por base os falares da Galícia e do Norte de Portugal.

Assim como o castelhano, o português originou-se de uma variedade falada no norte (o galego-português medieval) levado ao sul durante a Reconquista. Cabe ressaltar que, quanto à norma, o português moderno diverge do castelhano, pois vai buscá-la não no norte, mas na região centro-sul, onde se localiza Lisboa; entretanto, durante todo o período compreendido entre o começo do século XIII e meados do século XIV, a língua comum é esse galego-português nascido no Norte.

Por volta de 1415, temos o início do chamado período do português clássico que vai até a publicação do poema épico *Os Lusíadas*, de Luís Vaz de Camões, datado de 1572. Cabe ressaltar que nesse período relativamente curto, ocorreram várias inovações e consolidações importantes na Língua Portuguesa, aproximando-a bastante da língua que falamos atualmente. Foi uma época muito relevante por abranger a fase da expansão das grandes navegações que levou Portugal à conquista de diversos territórios do além-mar.

O período também se mostrou fundamental em termos linguísticos, pois houve a consolidação de uma língua literária de forte expressão. A Universidade de Coimbra, fundada em 1290, é transferida para Lisboa por algumas décadas e exerce um papel importante na produção e disseminação das letras e da gramática, da dialética e da retórica. Há o florescimento da prosa historiográfica, tendo como expoente, Fernão Lopes. Assim afirmam Basso & Gonçalves:

(...) Todo esse movimento culminaria com o estabelecimento do português clássico, especialmente em virtude da produção renascentista, no qual a língua viria a se consolidar e retomar vários aspectos do latim, indo na contramão de várias mudanças naturais do romance que ocorreram no período anterior do português arcaico. (...) vemos o centro de cultura de Portugal ser transportado do norte

linguisticamente conservador para o sul, no eixo Coimbra-Lisboa. O resultado disso é uma mudança no modo de desenvolvimento da língua do sul, que se baseia nos clássicos para desenvolver uma língua literária comum (como uma *koiné*), supradialetal, que exercerá uma forte influência nos períodos posteriores. A ruptura com o português arcaico e com a variedade setentrional, o galego-português, se dá com a independência de Portugal e com o deslocamento da corte para o sul, além do declínio natural na atividade literária trovadoresca (...) ocorrendo assim uma mudança geral no panorama linguístico-literário em Portugal (BASSO; GONÇALVES, 2014, p. 138).

Algumas das discussões sobre as diferenças encontradas entre o PE e o PB se baseiam muito na relação entre essa passagem do português arcaico para o português clássico, que ocorreu justamente à época da expansão marítima portuguesa. Quando chega ao Brasil, a variedade de português que o escritor Caminha traz consigo teoricamente é o português clássico. Entretanto, em 1500 havia, na verdade, um processo ainda de consolidação linguística na qual coexistiam falantes tanto do português arcaico quanto do clássico. Assim, seria precipitado afirmar que os portugueses das primeiras ondas migratórias falassem, em sua maioria, o português clássico.

Por fim, temos o que chamamos de português moderno, que tem início em 1572 e segue seu curso, de modo diverso, tanto em Portugal, quanto em suas ex-colônias (Brasil, África e Ásia) até os dias de hoje. Remonta ao século XVI, quando a língua se uniformiza, adquirindo as características do português atual. Em 1536, o padre Fernão de Oliveira publicou a primeira gramática de Língua Portuguesa, a "*Grammatica de Lingoagem Portuguesa*" com um estilo que se baseava no conceito clássico de gramática, entendida como "arte de falar e escrever corretamente".

Com o aparecimento das primeiras gramáticas que definem a morfologia e a sintaxe, a língua entra na sua fase moderna: em *Os Lusíadas*, de Luis de Camões (1572), o português já é, tanto na estrutura da frase quanto na morfologia, muito próximo do atual. A partir daí, a língua sofrerá mudanças menores como na fase em que Portugal foi governado pelo trono espanhol (1580-1640) e o português incorpora algumas palavras castelhanas; ou ainda, a crescente influência francesa no século XVIII (sentida principalmente em Portugal). Fenômenos semelhantes ocorrem nos séculos XIX e XX quando o vocabulário português recebe novas contribuições de origem greco-latina e inglesa para designar os avanços tecnológicos da época como: automóvel, televisão, *check-up* e *software*. Tais condições fizeram com que o português da metrópole se afastasse paulatinamente daquele falado nas suas ex-colônias.

## 2.2. Formação histórica do Português do Brasil

Ao analisarmos as fases pelas quais o Português do Brasil passou, novamente nos deparamos com a problemática questão da periodização. Diversos autores propuseram diferentes periodizações, baseadas nas mais diversas estratégias, cada qual apresentando vantagens e desvantagens.

Cientes da intrínseca arbitrariedade que toda periodização comporta, adotaremos aquela proposta por Noll (2008, p. 269 -275):

- 1) Fase inicial: compreendendo o período de 1500 a 1550 que corresponde ao traslado da Língua Portuguesa para o Brasil.
- 2) Fase formativa (1550 a 1700): já compreendendo as múltiplas interações e contatos linguísticos entre as matrizes portuguesa, indígena e africana e na qual há o surgimento e fixação de algumas das características marcantes do Português Brasileiro.
- 3) Fase diferenciadora (1700 a 1800): período em que se acentuam as diferenças entre o português europeu e sua variante brasileira. Cabe lembrar que é também nessa fase que há o florescimento da região das Minas Gerais.
- 4) Fase do desenvolvimento da escrita e do ensino: período de 1800 a 1950, quando tem-se o desenvolvimento da escrita e do ensino com a implantação de políticas de ensino no Brasil (criação das primeiras universidades) e publicação de documentos diretamente em território nacional, fruto das demandas decorrentes da transferência da corte lisboeta para o Rio de Janeiro.
- 5) Fase de nivelamento: desde os anos de 1950 à atualidade, marcada pela urbanização e industrialização do país e uma certa homogeneização linguística, fruto do nivelamento acarretado pelo avanço dos meios de comunicação de massa, principalmente o rádio e a televisão.

Essa periodização proposta pelo romanista alemão apresenta uma estruturação mais analítica e pormenorizada e, de acordo com Basso & Gonçalves:

(...) A periodização de Noll nos permite olhar com mais cuidado para as mudanças sociais pelas quais a sociedade brasileira passou e de que modo o PB acompanhou tais mudanças. (...) o que Noll apresenta é mais uma análise da formação do PB e de sua fixação e expansão no Brasil e menos uma descrição estrutural; uma razão para



tanto é que, apesar de já sabermos vários fatores estruturais envolvidos na formação do PB, ainda há muita pesquisa a ser feita para podermos falar com segurança quando as maiores mudanças ocorrerem e quais foram suas fases. Além disso há outra razão que tem a ver com a ideia de que a formação de uma língua segue de perto a formação da sociedade (...) (BASSO; GONÇALVES, 2014, p. 223).

A propósito dessa demarcação, a primeira fase de evolução do PB se inicia em abril de 1500, quando o navegante português, Pedro Álvares Cabral, chega à costa do território brasileiro, de que tomou posse em nome do então monarca D. Manuel. A colonização portuguesa, entretanto, efetivamente apenas terá início em 1532, com a criação e distribuição de quinze capitanias hereditárias.

No momento do “descobrimento”, o Português Europeu passava pela sua fase “clássica”. Nesses anos iniciais, pode-se dizer que os primeiros contatos linguísticos entre os portugueses e as populações autóctones se resumiram a interações pontuais de troca nas quais utilizava-se algum *pidgin*, língua híbrida, rudimentar resultante do contato entre falantes de idiomas diferentes e também funcionando como eventual estratégia de uso dos *lingoas*. Os *lingoas* eram degredados portugueses propositalmente deixados entre os indígenas a fim de aprender seu idioma para posteriormente auxiliar os colonizadores nos negócios de interesse da Coroa. Acrescente-se que o interesse da Coroa pelo recém-descoberto território era diminuto e se resumia basicamente a exploração do pau-brasil. Assim, os portugueses pouco adentraram o território, e seus estabelecimentos se localizaram quase que exclusivamente no litoral.

Nessa fase, a influência das línguas autóctones no Português Europeu se resumia praticamente à adição de itens lexicais, palavras nativas que se referiam principalmente à flora e à fauna locais e que aparecem documentadas nas narrativas de alguns viajantes estrangeiros que por aqui passaram como Hans Staden (c. 1525 – c. 1579) e Jean de Léry (c. 1536 – c.1613). Além disso, há um guia de conversação escrito por Lery, mas em tupi-francês: *Les Singularités de la France Antarctique*. Assim, quanto ao português da fase inicial (1500-1550), este apenas teve seu léxico enriquecido por muitas palavras de origem indígena mencionadas, inclusive, nas primeiras gramáticas portuguesas de Fernão de Oliveira (1536) e de João de Barros (1540).

A fase seguinte, denominada de fase formativa, se estende de 1550 a 1700. Nesse período, Portugal se volta para o Brasil como uma alternativa às perdas que vinha enfrentando no contexto das suas colônias na Ásia. Por essa razão, Portugal inicia a ocupação agrícola da colônia a partir da segunda metade do século XVI: “Nesse processo, uma economia de coleta (baseada no corte do pau-brasil e na exploração do trabalho indígena) foi transformada numa

economia de produção açucareira intensiva baseada no trabalho escravo” (FARACO, 2016, p. 59).

Além disso, a fim de viabilizar a cultura da cana, a Coroa Portuguesa passou não só a se utilizar da mão de obra africana escravizada, nos mesmos moldes feitos anteriormente nos empreendimentos açucareiros realizados nos Açores e na Madeira, como também passou a controlar o tráfico negreiro praticamente sem concorrência, fornecendo escravos não apenas para o Brasil, mas também para as colônias açucareiras nas Antilhas, controladas por espanhóis, franceses, holandeses e ingleses. Diferentemente do que ocorrera em suas possessões na Ásia e na África, no caso do Brasil, a presença portuguesa não se limitou ao estabelecimento de entrepostos comerciais, mas se desenhou como uma colônia de exploração e povoamento.

Por volta de 1700, a população do Brasil era bem maior e havia diversos núcleos de povoamento embora quase que exclusivamente localizados na faixa litorânea. Apesar da ocupação predominantemente costeira, o interior já havia sido explorado pelos bandeirantes paulistas em busca de metais preciosos e de índios. Em suas inúmeras entradas e bandeiras, os paulistas ajudaram a desenhar a configuração atual do Brasil, notadamente no que diz respeito à interiorização da presença e do domínio portugueses e à expansão das fronteiras.

Nesse período temos a formação da língua geral, uma espécie de *koiné*<sup>2</sup>, alicerçada na família linguística Tupi-Guarani e sistematizada pelos jesuítas que chegaram ao Brasil a partir de 1549. Também chamada de língua brasílica ou língua da terra ou ainda “a língua mais usada na costa do Brasil”, tal espécie de língua franca foi de grande utilidade para os colonizadores portugueses e para a cristianização dos indígenas.

Os jesuítas escreveram diversas obras bilíngues (português-língua geral) para auxiliar na evangelização dos nativos. Neste contexto, se destacam as obras do Padre José de Anchieta, a “*Arte de Grammatica da Lingva Mais Vsada na Costa do Brasil*”, de 1595 e a obra de António de Araújo, “*Catecismo. Na Lingoa Brasilica, no qual se contem a svmma da doctrina christã...*”, de 1618. As consequências linguísticas da expansão portuguesa em terras brasileiras foram diversas das observadas nos entrepostos asiáticos e africanos.

Com a descoberta do ouro na região das *Minas Geraes*, inicia-se a ocupação do interior e uma urbanização mais dinâmica; também nesse período, há a entrada de um grande contingente de portugueses provenientes de diversas regiões de Portugal e a metrópole

---

<sup>2</sup> Em termos linguísticos, a etimologia da palavra se funda no grego antigo e se refere a uma espécie de língua comum simplificada que resulta da convergência de dois ou mais dialetos ou línguas da mesma família. (nota da autora)

começa a apresentar políticas linguísticas para a difusão e consolidação do português no Brasil e para a eliminação das línguas gerais.

Segue-se então a chamada fase diferenciadora. Neste período, encontram-se diversos documentos que já assinalam as diferenças entre o Português do Brasil e o Português Europeu, conforme Basso & Gonçalves:

Devido principalmente ao ciclo do ouro, o Brasil passa a ser o destino de muitos portugueses e outros europeus, que vinham para cá em busca de riqueza rápida. As constantes idas e vindas de portugueses e brasileiros – estes, uma vez ricos, iam à metrópole para desfrutar de sua repentina riqueza (...) permitiram um maior contraste entre as variedades faladas dos dois lados do Atlântico, pois os dois portugueses eram ouvidos aqui e lá com maior regularidade (BASSO; GONÇALVES, 2014, p. 231).

Segundo Noll (2008), como exemplo de um texto que evidencia traços considerados típicos do Português do Brasil falado na província das Minas, temos a peça de teatro *O periquito do ar*, escrita por volta de 1800, por Rodrigues Maia. A fala do personagem principal (Dom Periquito, um brasileiro estereotipado) apresenta significativos marcadores como a escrita das vogais que são alçadas, principalmente <e> para <i>, como *mi diga* (me diga), *sinhorinha* (senhorinha), *mitêlo* (metê-lo), *di* (de); o uso exagerado de diminutivos, a maior ocorrência de próclise em comparação à modalidade europeia (*mi deixe, le diga, mi consterna*); dentre outros.

Ainda segundo o autor, outras características marcantes do Português Brasileiro, e ainda atestadas em graus diferentes no Brasil atual, são encontradas nas normas educacionais do *Seminário Episcopal de Nossa Senhora da Grasa de Olinda* (Pernambuco), e do *Recolhimento de Nossa Senhora da Glória do Lugar de Boa-Vista*, de Recife (Pernambuco), ambas escritas em 1798 pelo então bispo de Pernambuco. Dentre os vários “vícios de linguagem” praticados pelos alunos e apontados pelo bispo, encontram-se a monotongação de [ej] para [e] (como “janeru” em vez de janeiro), a queda do /l/ final (Portugal como “Portugá” [portu’ga]), a falta de concordância no sintagma nominal para a formação do plural (“os menino”, “muitas flor”), dentre outros.

Importante ressaltar que há, por parte da Coroa, a promoção da Língua Portuguesa em substituição às línguas gerais sendo que, dentre as medidas tomadas, com o fim de alcançar esse objetivo, a mais importante foi o *Diretório dos Índios*”, em 1757, promulgado pelo Marquês de Pombal, do que destacamos:

Para desterrar este perniciosíssimo abuso, será hum dos principaes cuidados dos Directores, estabelecer nas suas respectivas Povoações o uso da Língua Portuguesa, não consentindo por modo algum, que os Meninos, e Meninas, que pertencerem ás Escólas, e todos aquelles Indios, que forem capazes de instrucção nesta materia, usem da Lingua própria das suas Naçoens, ou da chamada Geral; mas unicamente da Portuguesa, na forma, que Sua Magestade tem recomendado em repetidas ordens, que até agora se não observáão com total ruina Espiritual, e Temporal do Estado.<sup>3</sup>

Diante do exposto, constata-se que o português falado no Brasil já se encontrava próximo da língua falada contemporaneamente.

Temos, então, a fase que vai de 1800 a 1950 nomeada por Noll como fase de desenvolvimento da escrita e do ensino. Em teoria, corresponde a um período marcado por vários acontecimentos de relevante importância: a vinda da família real portuguesa e de sua corte (1808), a Proclamação da Independência (1822), a abolição da escravatura e o estabelecimento da república (respectivamente em 1888 e 1889), a entrada de milhões de imigrantes europeus como massa assalariada, a industrialização e a rápida urbanização de um país até então eminentemente rural e o conseqüente desenvolvimento da escrita e do ensino.

Do ponto de vista linguístico-estrutural, conforme ressaltado por Noll (2008), certos fenômenos encontrados a partir de 1800 parecem ser exclusivos do Português do Brasil, como os exemplificados abaixo:

- 1) a queda do /r/ final principalmente nos verbos (“cantá”, “fala”, “busca”),
- 2) a monotongação de [ej] em [e] (“manteiga”, “bera”, “dexa”, em vez de manteiga, beira, deixa),
- 3) aféreses: “tá”, “tô”, “cê”, “ocê”,
- 4) epêntese de [i] antes de /s/ final: “meis” (mês), “treis” (três);
- 5) africacão de /t/ e /d/ antes de /i/ (como na pronúncia para “tio” e para “dia”),
- 6) a pronúncia “chiantre” de /s/ e /z/ em fins de palavras e de sílaba (característico do Rio de Janeiro, de Belém e de várias cidades litorâneas).

Ainda duas características importantes do PB sempre mencionadas por pesquisadores estudiosos do tema são o uso de formas como “mi”, “lhi”, “filiz” (em vez de me, lhe, feliz) e a pronúncia /w/ para o <l> final de sílaba como em “caniu” (canil), Brasiu (Brasil), “abriu” (abril).

Interessante ressaltar que, de acordo com Noll, por volta dessa época, já se documenta a existência do chamado “erre caipira” (o <r> retroflexo) especialmente no interior de São Paulo, mas também no Paraná e em regiões de Minas Gerais e Mato Grosso.

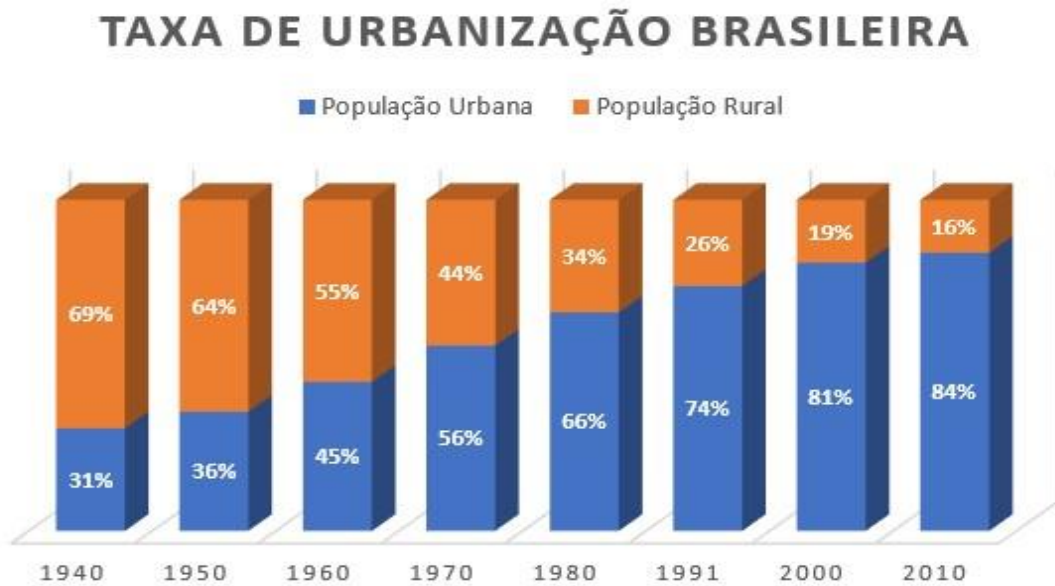
<sup>3</sup> Trecho do Diretório dos Índios, conforme fac-símile anexo em Almeida (1997).

A vinda da corte portuguesa para fixação no Rio de Janeiro contribuiu para a fixação do Português do Brasil como o conhecemos hoje visto que, de 1808 em diante, o país passou a contar com a autorização do uso da prensa tipográfica, com políticas educacionais e com uma urbanização cada vez mais intensa com a abertura dos portos. A entrada de escravos africanos beira os 2 milhões (IBGE, 2000, p.223) e estes aprendiam o português como segunda língua, de forma fragmentada e deficiente, na maioria dos casos, o que certamente causou impactos no desenvolvimento do Português Brasileiro. A imigração massiva de trabalhadores europeus também alterou o quadro sociolinguístico original visto que a maioria deles era formada de europeus de matriz não portuguesa (espanhóis, italianos, alemães, sírio-libaneses etc.).

Durante a República Velha, sob a influência dos ideais positivistas, a educação torna-se laica e gratuita e, durante o regime do Estado Novo (1937-1945). Uma das consequências linguísticas mais impactantes, principalmente durante a Segunda Guerra Mundial, é a obrigatoriedade do ensino do idioma português em todo território nacional, inclusive nas escolas das colônias de imigrantes, quando passa a ser proibido o ensino de línguas estrangeiras e há o fechamento de escolas mantidas por comunidades de imigrantes (principalmente alemães, italianos e japoneses) bem como a proibição da publicação de imprensa em outro idioma que não o português. Como é possível perceber, nesse período em especial, foi bastante expressivo o papel das políticas educacionais na padronização, disseminação e homogeneização da Língua Portuguesa falada/usada no Brasil.

Finalmente, se inicia a chamada fase de nivelamento, que corresponderia a meados de 1950 até os dias atuais. De 1900 em diante, o Brasil passou por um impressionante processo de urbanização. Para se ter uma ideia, por volta de 1940, a população rural correspondia à 69% da população do país, ou seja, uma proporção superior à 2/3 (dois terços) da população total; entretanto, em apenas três décadas, o quadro praticamente se inverte e em 1970 a população urbana passa a totalizar 56% da população nacional, conforme observarmos no quadro abaixo:

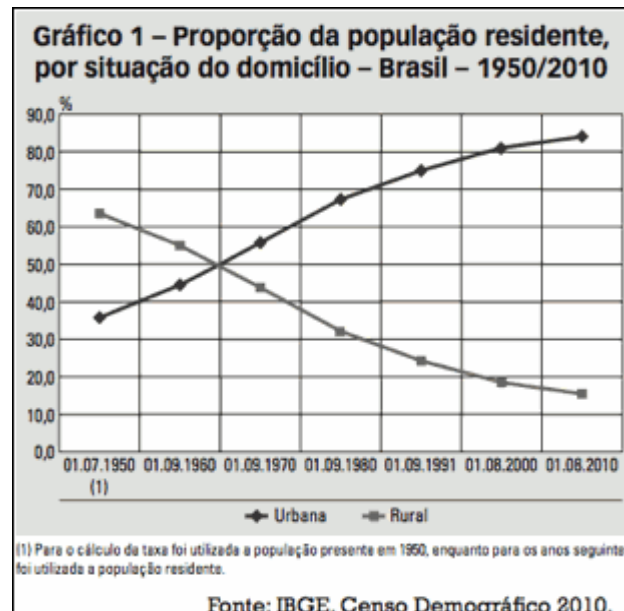
Gráfico 1 - Taxa de urbanização brasileira no período compreendido entre 1940-2010



Fonte: <https://www.proenem.com.br/enem/geografia/urbanizacao-brasileira/>

No próximo gráfico, nota-se que, por volta de 1950, ocorre a grande virada, quando então a população urbana ultrapassa rapidamente a população rural:

Gráfico 2 – Proporção da população residente, por situação do domicílio (1950-2010)



Esse rápido processo de urbanização tem vários impactos na linguagem: aumento das mudanças linguísticas devido a maior interação entre as populações anteriormente isoladas; aos diversos movimentos migratórios internos (deslocamentos populacionais

majoritariamente de nordestinos para a região norte, centro-oeste e, principalmente para a região sudeste. Há ainda o intenso fluxo de populações do interior dos estados do Sudeste e do Sul para as respectivas capitais. A exposição midiática e o papel dos meios de comunicação de massa, principalmente o rádio e a televisão, são fatores determinantes no processo de nivelamento de uma norma urbana ou, segundo Stella Maris Bortoni-Ricardo (2004) uma norma *rurbana*.

Deveria haver uma fase intermediária entre as fases 4 e 5 propostas por Noll e pelos autores citados, pois creditamos que a fase 4 abrange um período bastante longo (1800 a 1950), no qual ocorreram diferentes fatos linguisticamente relevantes: incremento da influência do Português Europeu com a chegada da corte; intensos contatos linguísticos entre imigrantes europeus e populações africanas, principalmente no interior do país com a abolição da escravatura e a substituição do braço escravo pela mão de obra assalariada (principalmente no período entre 1880 a 1920).

A intensa urbanização que toma impulso em meados do século XIX faz com que os falantes de comunidades localizadas no interior do país se desloquem para os grandes centros urbanos; em grandes capitais, como São Paulo, por exemplo, assistimos a uma babel de linguajares e sotaques que sinalizam a origem dos seus falantes. Dentre esses, encontra-se aquele no qual se situa o objeto do nosso estudo, ou seja, o R caipira (R retroflexo), marca distintiva do chamado dialeto caipira, variação que analisaremos mais detidamente a seguir.

Fez-se aqui um longo percurso para chegar a uma peculiaridade de nosso Português do Brasil e o que se pôde observar é que uma língua está sempre exposta à influência de diversas outras e tanto o Português Europeu quanto o Português do Brasil não são exceções: estiveram e estão expostos a esse fenômeno. Desse contato surgem acréscimos, trocas ou variações. Seria, então, o retroflexo um desses casos?

### 3 O DIALETO CAIPIRA

O dialeto caipira é um dialeto do Português do Brasil falado no interior do estado de São Paulo, leste e sul do Mato Grosso do Sul, sul e sudoeste de Minas Gerais, sul de Goiás, norte do Paraná e zonas rurais do sul do Rio de Janeiro.

Os estudos relativos a essa variedade do Português do Brasil têm aumentado nas últimas décadas, entretanto, pode-se dizer que o pioneiro na delimitação e caracterização da modalidade foi Amadeu Amaral em *O Dialeto Caipira*, de 1920.

Tivemos, até cerca de vinte e cinco anos atrás, um dialeto bem pronunciado, no território da antiga província de S. Paulo. É de todos sabido que o nosso falar caipira – bastante característico para ser notado pelos mais desprevenidos como um sistema distinto e inconfundível – dominava em absoluto a grande maioria da população e estendia a sua influência à própria minoria culta. (AMARAL, 1920, p. 01)

Uma possível explicação do que vem a significar o termo caipira se mostra necessária visto que, muito da conceituação popularmente disseminada se ampara no senso comum e em definições de perspectivas predominantemente preconceituosas:

Como jagunço, tabaréu e sinônimos, o termo caipira se fixou no uso comum brasileiro no século XIX, mas não tem acepção simples em português. Seu designado, o tipo ou indivíduo a que refere também é figura complexa e fértil não apenas como expressão literária ou personagem da história e da cultura. Vive ou sobrevive ainda e atua – contrariamente à previsão de Euclides – na sociedade brasileira contemporânea como um componente básico de sua constituição. E como assunto literário ou de estudos tem sido aproveitado em obras diversas, como as de Afonso Arinos, de Valdomiro Silveira, de Cornélio Pires, de Monteiro Lobato, de Alberto Faria, para lembrar apenas os mais próximos de Amadeu (CESCHIN, 1999, p. 45)

Dentre as várias definições do vocábulo encontradas, no dicionário Michaelis (2021) podemos observar que o termo caipira se refere a:

- 1 Pessoa que nasceu ou mora na roça ou em ambientes rurais e que comumente trabalha em serviços de lavoura de subsistência no Sudeste ou Centro-Oeste brasileiros, em especial no interior de São Paulo; araruama, cafumango, cariazal, jeca, mano-juca, mixuango, muxuango, pé-duro, pé no chão, pioca, piracuara, piraguara, piraquara, roceiro, saquarema (...)
- 2 Indivíduo simples, ingênuo, tímido, de pouca ou nenhuma instrução e hábitos rudes, em geral habitante do campo; capa-bode, catrumano, groteiro, jeca, matuto, mcorongo, queijeiro, tabaréu (...)

Ou ainda, funcionando como adjetivo:



1 Que vive no interior, fora dos centros urbanos; que vive no campo ou na roça; caboclo, capiau, jeca-tatu, matuto, roceiro, sertanejo, sitiano.

2 Que é rude, de pouca instrução, afastado do convívio social ou que leva uma vida de hábitos e modos rústicos.

3 Próprio de ou típico de caipira.

4 FIG Que é tímido, acanhado, envergonhado.

O termo é frequentemente utilizado de forma pejorativa, entretanto, cabe salientar que o vocábulo caipira continua de origem controversa, sugerida por Cunha (1986, p. 137) como proveniente da mistura das palavras *caipora* e *curupira*, de origem tupi.

Darcy Ribeiro, em *O povo brasileiro: a formação e o sentido do Brasil*, dedica um capítulo ao chamado Brasil Caipira, como observamos no excerto abaixo em que apresenta as circunstâncias da propagação, por assim dizer, da cultura caipira:

Esgotado o impulso criador dos bandeirantes que se fizeram mineiros, toda a economia da vasta população do Centro Sul entra em estagnação. Mergulha numa cultura de pobreza, reencarnando formas de vida arcaica dos velhos paulistas que se mantinham em latência, prontas a ressurgir com uma crise do sistema produtivo. A população se dispersa e se sedentariza, esforçando-se por atingir níveis mínimos de satisfação de suas necessidades. O equilíbrio é alcançado numa variante da cultura brasileira rústica, que se cristaliza como área cultural caipira. É um novo modo de vida que se difunde paulatinamente a partir das antigas áreas de mineração e dos núcleos ancilares de produção artesanal e de mantimentos que a supriam de manufaturas, de animais de serviço e outros bens. Acaba por esparramar-se, falando afinal a língua portuguesa, por toda a área florestal e campos naturais do Centro-Sul do país, desde São Paulo, Espírito Santo e estado do Rio de Janeiro, na costa, até Minas Gerais e Mato Grosso, estendendo-se ainda sobre áreas vizinhas do Paraná. Desse modo, a antiga área de correrias dos paulistas velhos na preia de índios e na busca de ouro se transforma numa vasta região de cultura caipira, ocupada por uma população extremamente dispersa e desarticulada. Em essência, exaurido o surto minerador e rompida a trama mercantil que ele dinamizava, a paulistânia se "feudaliza", abandonada ao desleixo da existência caipira (RIBEIRO, 1995, p. 381-382).

Dentre as referências literárias sobre a importância da cultura caipira nos estudos das Ciências Sociais acerca da Formação Cultural do Brasil, é evidente o destaque da obra de Antonio Candido:

Nos lugares onde se estabeleceu, desde o início, um estrato social dominante e ligado intimamente à tradição do reino (como foi o caso da Bahia e de Pernambuco, devido principalmente a fatores de ordem político-administrativa), a cultura portuguesa pôde desde logo transplantar-se. A consequência foi o desenvolvimento por assim dizer de dois planos culturais, o português e o sincrético (este, acentuado logo depois na faixa litorânea pela contribuição africana), estabelecendo-se uma distância apreciável entre a camada dominante e a camada dominada, tanto no plano estrutural quanto no da cultura (CANDIDO, 1956, p. 160).

Os indivíduos pertencentes a essa cultura compartilhavam, como era de se esperar, não somente hábitos e costumes semelhantes como também um característico linguajar ao qual convencionou-se chamar de dialeto caipira.

### 3.1. Amadeu Amaral e o livro *O Dialeto Caipira*

Amadeu Amaral é considerado por muitos como um dos principais estudiosos da linguagem no Brasil, bem como um dos precursores da Dialetologia brasileira graças a publicação de *O Dialeto Caipira*, em 1920.

O esforço e a importância do seu legado são inegáveis visto que, apesar das limitações existentes à época da sua pesquisa, Amaral pautou-se por princípios rigorosos de investigação do ponto de vista metodológico, esforçando-se ao máximo para conferir confiabilidade ao seu trabalho descritivo da língua. Realizou descrições fonológicas essenciais à caracterização do dialeto caipira, numa época em que não havia instrumentos nem tecnologia adequados, como os gravadores de áudio, por exemplo.

É uma obra de referência na história da Dialetologia brasileira e um marco na história da linguística no Brasil, considerando que um dos objetivos do autor foi descrever o falar caipira em seus diferentes aspectos: fonético, lexical, morfológico e sintático e retratar de forma abrangente um falar regional brasileiro. Conforme salientado por Laura do Carmo:

Até 1920, ano em que *O dialeto caipira* veio a público, os estudos dialetais no Brasil eram basicamente no campo do léxico, resultando na elaboração de dicionários, vocabulários e léxicos, de brasileirismos e de regionalismos. A publicação do trabalho de Amadeu, seguida, em 1922, de *O linguajar carioca*, de Antenor Nascentes, é marco de uma nova abordagem no estudo dialetal no Brasil. Os trabalhos de natureza lexi-cográfica continuam, mas os estudos dialetais, a partir de então, passam a incluir também os aspectos fonéticos, morfológicos e sintáticos. Apesar de não se realizar um trabalho de campo sistemático – como o que vai ser proposto a partir da década de 1950 –, havia a observação direta do “acontecer” da língua, orientações quanto à metodologia de abordagem e de anotações, bem como um interesse especial pelos diferentes aspectos da realidade do falante (CARMO, 2008, p. 375-376).

Em linhas gerais, Amadeu Amaral defende a tese de que o chamado dialeto caipira estava bem sedimentado até o século XIX no território da antiga província de São Paulo, sendo “bastante característico para ser notado pelos mais desprevenidos como um sistema distinto e inconfundível”, e que dominava em absoluto a fala da maioria da população e

estendia a sua influência até outros grupos, inclusive aos mais cultos, a minoria da população. Já àquela época, era possível perceber que o estabelecimento de padrões cultos no ensino estava diluindo as características típicas desse dialeto caipira, fato bem observado por Amaral (1920, p.1).

De acordo com Duarte (1976), conforme ressaltado por Zilio-Passerini (2020), o interesse inicial de Amaral era produzir um “Cancioneiro Popular” visto que já em 1916, o autor iniciou a coleta de material para sua pesquisa redigindo uma circular, remetida a todos os agentes do jornal *O Estado de São Paulo* distribuídos pelo território nacional, na qual solicitava o envio de quadrinhas, modinhas, poesias e demais elementos referentes aos usos e costumes populares, como verificado no excerto abaixo que reproduz a circular de Amaral:

Pretendo organizar um “Cancioneiro Popular” de São Paulo, a exemplo do que já se tem feito em todos outros países do Brasil e aqui mesmo no Brasil, em diversos estados. Serei, pois, muito grato a V.S. se me quiser prestar o favor de, com pequeno trabalho, colher, ou fazer colher por pessoa idônea, alguns versos desses que correm no meio do povo, entre adultos ou crianças, sejam em quadras, sejam em qualquer outra disposição.

No caso de V.S. estar disposto a prestar-me este obséquio, rogo-lhe observar as seguintes condições:

a. colher todos versos, por imperfeitos que sejam, indistintamente, visto que os mais imperfeitos e destituídos de sentido podem ter um grande interesse de folclore. b. escrever os versos sem introduzir neles a menor modificação. c. declarar ao pé a localidade onde foram apanhados.

Deixo aqui, desde já, os meus cordiais agradecimentos a V.S. na esperança de renová-los na própria obra que trato de organizar. (DUARTE, 1976, p. 34)

Com relação à estrutura da obra, esta se divide em cinco partes além da introdução: fonética, lexicologia, morfologia, sintaxe e vocabulário. Na introdução, Amadeu Amaral procede à contextualização do seu objeto de estudo: o dialeto corrente na antiga província de São Paulo até meados do século XIX (ZILIO-PASSERINI, 2020). Também há a menção ao trabalho de outros pesquisadores, linguistas e literatos tais como: Adolfo Coelho, Said Ali, Teodoro Sampaio, Afrânio Peixoto, Monteiro Lobato, dentre outros. As quatro primeiras partes seguem o modelo das gramáticas tradicionais, enquanto a quinta parte (Vocabulário) é a mais extensa de todo o livro e com esmerado detalhamento etimológico, como se observa no fragmento abaixo:

(...)

AFINCÁ(R), v. t. - embeber, cravar (qualquer objeto delgado e longo): "Afinquei o pau no chão". "Não afinque prego na parede". "O marvado afincô a faca no ôtro". | É port., como fincar, mas com acepções diversas.

AFITO, s. m. - mau olhado. | Apesar de nunca termos ouvido este voc., e só o haveremos encontrado num escrito ("A Superst. Paulistana", eng. E. Krug),

resolvemos registá-lo, por ser dos mais curiosos. É palavra antiga na língua com a significação de indigestão, diarréia ("Novo Dic."). Em cast. existe ahito, q., - o que padece de. indigestão ou embaraço gástrico. Comparando-se isto com o sentido que dão à palavra os caipiras. segundo o citado escritor, e com a expressão "deitar o fito", equivalente a "deitar mau olhado". que se encontra em Gil V., pode deduzir-se que a significação primitiva do voc. port. e cast. deve ter sido, mais extensamente, a de - indisposição causada por mau olhado, quebranto (AMARAL, 1920, p. 38).

O pioneiro trabalho realizado por Amaral, agora centenário, continua como um marco referencial e ponto de apoio de boa parte dos estudos relacionados ao dialeto caipira. A caracterização concebida evidencia o empenho e o cuidado do autor na descrição dessa variação do Português do Brasil.

Segundo Castro (2006), desde a obra de Amadeu Amaral, podemos identificar alguns traços característicos da fonética caipira. Estabelecendo diferença em relação ao português lusitano, Amaral (1982: 45) destaca como característica prosódica do dialeto caipira “o frasear lento, plano e igual”, associado à maior duração das vogais e à ocorrência de mais pausas na realização do grupo de palavras.

No que diz respeito à fonética, apresenta a caracterização desse aspecto da fala caipira de forma didática e detalhada. No tópico inicial da seção sob a etiqueta “Generalidades”, o autor traz algumas características gerais do linguajar caipira, e enfatiza o ritmo e musicalidade da linguagem, com seu tom lento, plano e igual sem variedade de inflexões, os quais considera essencialmente diferentes do português europeu. Ao analisar tal característica da prosódia caipira, Amaral salienta que a longa duração dos tempos das vogais, até por consequência da lentidão da fala, acaba por tornar a pronúncia vagarosa, cantada e com um excessivo estiramento das vogais.

Outra característica marcante, segundo o autor, que diferencia a pronúncia caipira da portuguesa, diz respeito à quase inexistência da elisão das vogais átonas tão comum na boca dos lusitanos o que faz com que sua fala soe muito mais vigorosa e rápida, ao abreviar os vocábulos, “comendo-lhe” as vogais e pela absorção das átonas finais nas vogais que se lhes seguem:, p'dáçu, c'rôa, 'sp'rança, tiátru, d'hoj'em diante, um'august'assemblêia; enquanto o caipira paulista, por apoiar-se mais demoradamente nas vogais, não praticasse em tão larga escala essas mutações e elisões, pronunciando as vogais de forma clara.

Assim, pode dizer-se que no dialeto não aparecem vogais ensurdecidas: todas soam distintamente, salvo os casos de queda ou de sinalefa. E, ainda, é a partir de uma prosódia peculiar ao falar do PB que Amaral (1920, p. 04-95) justifica a oscilação quanto à colocação pronominal:

(...) Seria, aliás, muito interessante um estudo acurado das feições especiais da prosódia caipira, com o objetivo de discriminar a parte que lhe toca na evolução dos diferentes departamentos do dialeto. Chegar-se-ia de certo a descobertas muito curiosas, até no domínio dos fatos sintáticos. A diferenciação relativa à colocação dos pronomes oblíquos, no Brasil, deve explicar-se, em parte, pelo ritmo da fala e pelo alongamento das vogais (3). Esses pronomes, no português europeu, se antepõem ou pospõem a outras palavras, que os atraem, incorporando-os. Prosodicamente, não têm existência autônoma: são sons ou grupos de sons, destinados a adicionarem-se aos vocábulos acentuados, segundo leis naturais inconscientemente obedecidas (ênclise, próclise). Passando para o Brasil, a língua teve que submeter-se a outro ritmo, determinado por condições fisiológicas e psicológicas diversas: era o suficiente para quebrar a continuidade das leis de atração que agiam em Portugal. O alongamento das vogais, dando maior amplitude aos pronomes na pronúncia, tornando mais sensível a sua individualidade, veio acentuar, de certo, aquele efeito (AMARAL, 1920, p. 4-95).

No que diz respeito especificamente ao objetivo imediato do nosso estudo, interessados, *a priori*, a seção de fonologia que Amadeu Amaral se ocupou em descrever o singular R retroflexo encontrado no linguajar caipira.

### 3.2. Características fonéticas presentes em *O Dialeto Caipira*

Com relação aos fonemas, Amaral defende a tese de que os fonemas do dialeto são em geral os mesmos do português, com a ressalva de desconsiderarem-se “ligeiras variantes fisiológicas, que sempre existem entre povos diversos e até entre frações de um mesmo povo”; mas observa algumas peculiaridades tais como:

- 1) O /s/ pós-vocálico tem sempre o mesmo valor, é uma linguodental ciciante, não se notando jamais as outras modalidades conhecidas entre portugueses e mesmo entre brasileiros de outras regiões (como o s “chiante”, tipicamente carioca).
- 2) A explosiva gutural gh tem uma tonalidade especial, sobretudo antes dos semiditongos cuja prepositiva é u, casos em que frequentemente se vocaliza: áu-ua = água, léu-ua = légua).
- 3) Apresenta ch e j palatais explosivos, característica que o autor considera como uma característica que ainda se conserva nos falares portugueses em certas regiões de Portugal.

- 4) A consonância palatal molhada lh não existe no dialeto, fenômeno que o autor salienta ser comum à maioria dos dialetos portugueses da África e Ásia, e como em vários dialetos castelhanos da América.

Em relação aos segmentos vocálicos, entre as peculiaridades apontadas por Amaral, temos:

- a realização de [e] e [o] átonos finais (est[e], pov[o]), não se verificando no dialeto caipira o alçamento das médias observado, nesse contexto, em outras regiões do país;
- o alçamento e nasalização do /e/ pretônico inicial, como ocorre em [ɪ]xame, [ɪ]ξεμπλο, [ɪ]leição; c) realização do ditongo nasal de bom, tom, som como [ãw].

No que se refere às consoantes, destacamos:

- a realização africada das palatais /Σ / ε /Z/ - [tΣ]ave para chave, [dZ]ente para gente - , pronúncia também registrada “entre o povo em certas regiões de Portugal”, conforme apontado por Amaral (1982: 48);
- a alternância entre /b/ e /v/, dando lugar a formas sincréticas como [b]assora / [v]assora; [b]espa / [v]espa; [b]amo / [v]amo.
- a rotacização do “L”, que é a permutação, em fim de sílaba, da aproximante lateral [l] pelo fonema /r/ (por exemplo, mil > mir, enxoval > enxovar, etc);
- a iotização do “LH” [ʎ] (<Falhou> [fa'jo]; <Mulher> [mu'jɛ]; <Alho> ['aju]; <Velho> [ve'ju]; <Olhei> [o'jej], etc.);
- a apócope da consoante /r/ na terminação dos verbos no infinitivo (<Brincar> [brĩ'ka]; <Olhar> [o'ja]; <Comer> [ko'me]; <Chorar>: [fo'ra]; etc.);
- a transformação de proparoxítonas em paroxítonas: a apócope ou síncope em palavras proparoxítonas e a aférese em muitas palavras (por exemplo, árvore .arvre, etc);
- o R caipira: o fonema /r/, em fim de sílaba ou em posição intervocálica, assume as características formas aproximante alveolar [ɹ], retroflexo [ɻ].

Nessa seção do trabalho de Amaral, o autor chama a atenção para o traço objeto do nosso estudo, o R caipira, de que trataremos com mais detalhamento no próximo capítulo.

#### 4. OS RÓTICOS E O RETROFLEXO

O estudo e classificação dos róticos apresenta considerável complexidade devido ao amplo espectro de suas realizações fonéticas. Segundo Natvig, em seu artigo *Rhotic underspecification: Deriving variability and arbitrariness through phonological representations*:<sup>4</sup>

As propriedades fonéticas dos róticos existem dentro de um espectro de gestos articulatórios inter-relacionados, mas sem uma característica unificadora comum que os distingue de outros sons. Portanto, não é possível classificá-los universalmente com base em características de superfície e, ao mesmo tempo, excluir sons que não demonstram comportamento fonológico semelhante ao rótico (...) Na verdade, Chabot afirma que “pode não haver maneira de unir róticos por meio de modelos representacionais; isto é, eles não podem ser entendidos fora do papel que desempenham dentro de um sistema” (2019:3), levando à conclusão de que “se os róticos estão em uma relação fonética/fonologia arbitrária, tal relação deve, em princípio, ser possível para toda fonologia” (Chabot 2019:18). A questão, então, é se o grau de arbitrariedade que os róticos exibem é atestado para outras classes fonêmicas, ou se há motivações estruturais para essa relação arbitrária entre fonética e fonologia (NATVIG, 2020, p. 1) (Tradução livre da autora).<sup>5</sup>

Os sons de /r/ - ou róticos (forma aportuguesada do inglês “rhotics”) não podem ser identificados por meio de características articulatórias comuns, mas por outros fatores, como sinal ortográfico utilizado ou, mais frequentemente, pela posição que ocupam nas estruturas silábicas de diferentes línguas (cf. Ladefoged & Maddieson, 1996, apud Lindau, 1985).

Com relação à ocorrência do R caipira em posição intervocálica (a[©]a[©]a) e pós-vocálica (ca[©]ta) – som identificado como “linguopalatal e guturalizado”, a articulação é minuciosamente descrita: o R retroflexo, como também é conhecido, é, talvez, o traço mais marcante do que hoje se identifica como uma pronúncia “caipira”; ou seja, como podemos observar, o R caipira, em especial, o retroflexo, se apresenta como uma marca distintiva do dialeto descrito por Amaral. Para o estudioso, o R inter e pós-vocálico (arara, carta) possui um

<sup>4</sup> Subespecificação Rótica: Derivando a variabilidade e arbitrariedade por meio de representações fonológicas. *Glossa: a journal of general linguistics* 5(1): 48. 1–28. DOI: <https://doi.org/10.5334/gjgl.1172>.

<sup>5</sup> The phonetic properties of rhotics exist within a spectrum of interrelated articulatory gestures, but without one common unifying characteristic that distinguishes them from other sounds. Therefore, it is not possible to universally classify them based on surface features while also excluding sounds that do not demonstrate rhotic-like phonological behavior (...) In fact, Chabot states that “there may be no way of uniting rhotics via representational models; that is, they cannot be understood outside of the role they play within a system” (2019:3), leading to the conclusion that “if rhotics are in an arbitrary phonetics/phonology relationship, such a relationship must in principle be possible for all phonology” (Chabot 2019:18). The question, then, is whether the degree of arbitrariness that rhotics exhibit is attested for other phonemic classes, or whether there are structural motivations for this arbitrary relationship between phonetics and phonology.

valor peculiar: é linguopalatal e guturalizado. Autores posteriores criticaram o termo “guturalizado”, demasiado vago, em termos fonéticos. Apesar disso, não podem negar o notável empenho de Amaral em tentar descrever a forma de produção do fonema:

Na sua prolação, em vez de projetar a ponta contra a arcada dentária superior, movimento este que produz a modalidade portuguesa, a língua leva os bordos laterais mais ou menos até os pequenos molares da arcada superior e vira a extremidade para cima, sem tocá-la na abóbada palatal. Não há quase nenhuma vibração tremulante. Para o ouvido, este R caipira assemelha-se bastante ao r inglês post-vocálico. É, muito provavelmente, o mesmo r brando dos autóctones. Estes não possuíam o rr forte ou vibrante, sendo de notar que com o modo de produção acima descrito é impossível obter a vibração desse último fonema (AMARAL, 1920, p. 5).

Embora a ocorrência do retroflexo seja mencionada em vários trabalhos sobre os róticos no Brasil, esse som, até o momento, ainda não teve uma descrição fonético-acústica específica que o relacionasse com o contexto vocálico adjacente e com a posição na palavra, tampouco temos uma representação fonológica abrangente para o retroflexo do Português do Brasil

Para uma melhor contextualização e entendimento do fonema em comento, apresentamos abaixo o quadro de símbolos fonéticos consonantais relevantes para a transcrição do português, conforme Cristóvão Silva (2009, p. 37):

Quadro 1 – Quadro de símbolos fonéticos consonantais relevantes para a transcrição do português

Articulação		Bilabial	Labiodental	Dental ou Alveolar	Alveopalatal	Palatal	Velar	Glotal
Maneira	Lugar							
Oclusiva	desv voz	p b		t d			k g	
Africada	desv voz				tʃ dʒ			
Fricativa	desv voz		f v	s z	ʃ ʒ		X Y	h ɦ
Nasal	voz	m		n		ɲ ỹ		
Tepe	voz			ʀ				
Vibrante	voz			ʁ				
Retroflexa	voz			ɻ				
Lateral	voz			l ɭ		ʎ ɸ		

**Tabela:** Símbolos fonéticos consonantais relevantes para transcrição do português



Uma das primeiras tentativas de caracterização acústica dos sons do /r/ retroflexo foi feita por Lehiste (1962). Por meio de dados obtidos de informantes naturais de Estados do meio-oeste norte-americano, a autora pôde observar o comportamento do /r/ nas posições inicial, medial e final.

Com referência aos sons de /r/ que precedem sufixos derivacionais, percebeu-se que quando o /r/ torna-se intervocálico pela presença do sufixo – *er*, como observado em *bearer*, *borer*, *dearer*, parece assumir as características de um /r/ inicial. A autora sugere que os róticos podem participar de processos fônicos com condicionamento gramatical.

Tanto Lindau (1985) quanto Silva (2002) observaram fenômenos parecidos no comportamento dos róticos em línguas diferentes, em épocas diferentes, e tais estudos parecem dialogar progressivamente.

O estudo dos róticos no Português do Brasil não é muito recente. Callou *et al* (1996) já haviam observado uma grande variação para esses segmentos no nosso idioma, reportando um considerável número de realizações fonéticas frente à variação encontrada para as outras consoantes. Os róticos apresentaram a seguinte distribuição variacional: tepe (32%), apagamento do segmento (26%), fricativa velar (21%), fricativa laríngea (aspirada) (18%), retroflexo (1%) e vibrante alveolar (1%).

Os movimentos articulatórios bem como a resposta acústica que caracterizam os róticos têm sido objeto de diversos estudos que verificaram, em especial, as propriedades comuns e os aspectos que permitem discriminar tais segmentos (Silva, 1996; Ferraz, 2005; Fraga, 2008; Clemente, 2009; Ladegofed, 2010; Seara et.al, 2011).

Acerca dos róticos do Português do Brasil, tratamos do segmento retroflexo – aquele produzido com a elevação e o encurvamento da ponta da língua em direção ao palato duro. Esse movimento específico da língua pode gerar alterações no espectro acústico, em especial, na região do terceiro formante (F3) – o arredondamento –, que apresenta um acentuado movimento de abaixamento durante a produção do segmento (Ferraz, 2005; Fraga, 2008; Clemente, 2009). Além dessas características, o segmento retroflexo é caracterizado pela presença de energia espectral, qualidade semelhante às vogais, indicando a ausência de uma constrição acentuada na cavidade oral (Clemente, 2009).

Tais características acústicas e articulatórias referentes à produção do segmento retroflexo assim como de outros segmentos que constituem a classe dos róticos têm sido estudadas também para os diferentes falares do Português do Brasil. Conforme Brandão, a

ocorrência do R retroflexo é mais abrangente do que se supunha, não se limitando ao interior da região centro sul do país:

As primeiras menções ao R retroflexo e a seu âmbito de ocorrência datam de inícios do século XX, quando Amadeu Amaral o caracterizou como típico do dialeto caipira, denominação com que cunhou o falar “do território da antiga província de São Paulo” (1976: 45) e que, até fins do século XIX, teria dominado “em absoluto a grande maioria da população”, estendendo sua influência “à própria minoria culta” (1976:45). Silva Neto (1963: 194), dentre outros filólogos, também dela tratou, indicando o território em que incide – “três Estados: sul de São Paulo, sul de Mato Grosso, e norte do Paraná” – bem como a sua possível motivação sócio-histórica: “Em todo o caso, o ponto de irradiação parece ter sido São Paulo e o povoamento bandeirante”. No entanto, como afirma Brandão (2007: 268), “parece importante retomar a questão de sua distribuição geográfica, pois, como testemunham os atlas linguísticos regionais já publicados e diversos outros estudos, seu alcance espacial parece ser mais amplo do que se supunha”. O levantamento levado a cabo pela autora com base em diversos estudos sobre o R demonstra que a variante retroflexa está registrada de Norte a Sul, em falares das mais diferentes regiões do país (BRANDÃO, 2007, p. 268).

O mapa seguinte, apresentado por Brandão (2007), nos mostra a presença do R retroflexo em determinadas áreas do território brasileiro:

Figura 4 - Mapa da Presença do retroflexo em determinadas áreas do Brasil



(\*Presença do retroflexo em determinadas áreas)

Fonte: BRANDÃO, 2007, p. 265-283.

Assinalam-se os Estados em que se conhecem registros de retroflexo, sem levar em consideração as suas áreas específicas de ocorrência ou seu maior ou menor índice de frequência.

Cagliari (1981, 29-34), em sua tese de Livre Docente, intitulada *Elementos de fonética do português brasileiro*, faz um detalhado levantamento das possíveis ocorrências do R no Português do Brasil<sup>6</sup>. Do quadro fonético formulado dessas consoantes temos:

Quadro 2 - Quadro das consoantes conforme Cagliari (1981)

Modos de Articulação	Lugares de articulação							
	Dental	Alveodental	Alveolar	Palato alveolar	Velar	Uvular	Faringal	Glotal
Fricativa								
Oclusiva								
Fricativa		[ɣ] [ɣ]			[x] [χ]	[χ][κ]	[h] [ʕ]	[h][ɦ]
Vibrante	[r] [r]		[r] [r]			[ʀ]		
Tepe	[r][r]		[r] [r]					
Flepe			[ɾ] [ɾ]					
Retroflexa			[ɻ] [ɻ]	[ɻ] [ɻ]				
Constritiva			[ɻ][r]	[ɻ]	[ɻ]	[κ]		

Quadro 1: Quadro das consoantes - (Cagliari (1981, p. 39)

Quadro das consoantes conforme Cagliari (1981)

Fonte: [http://www.revel.inf.br/files/artigos/revel\\_9\\_as\\_realizacoes\\_dos\\_roticos.pdf](http://www.revel.inf.br/files/artigos/revel_9_as_realizacoes_dos_roticos.pdf)

Dentre as inúmeras peculiaridades do dialeto caipira levantadas por Amaral, o autor apresenta-nos a descrição detalhada da realização de um rótico, ao qual chama de R caipira,

<sup>6</sup> Do quadro das consoantes formulado por Cagliari (1981, p. 39), são extraídas para este artigo apenas as unidades sonoras relativas às articulações fonéticas do fonema /r/ e do arquifonema /R/, conforme ilustrado no Quadro 1.

descrito como um R inter e pós-vocálico bastante peculiar e que, ao ouvido, se assemelhava bastante ao R inglês pós-vocálico.

Para o fenômeno em questão, ou seja, para as várias possibilidades de realização de fonema, não há, até o momento, teorias satisfatórias que expliquem com exatidão todos os porquês de suas ocorrências. Especificamente ao caráter retroflexo do /ɾ/, já foram atribuídas diversas possibilidades de origem e diferentes realizações na mesma área ou no mesmo contexto. Algumas hipóteses serão apresentadas a seguir.

## 5. A ORIGEM DO R CAIPIRA

A origem do R caipira, em especial, na sua modalidade retroflexa apresenta considerável controvérsia. Apesar do crescimento de pesquisas na área dialetológica no país, ainda há muito a ser explorado, principalmente no que diz respeito às pesquisas de cunho mais técnico-fonológico, visto haver uma predominância de trabalhos mais voltados para a análise da variação linguística com ênfase nos aspectos socioculturais o que, sem dúvida também é de extrema importância. Além dos estudos dialetológicos serem ainda recentes, a insuficiência de registros e fontes é grande, visto que boa parte das línguas autóctones faladas nos períodos colonial e pré-colonial foram extintas ou transmitidas somente de forma oral.

Graças ao trabalho de religiosos (notadamente jesuítas) e de alguns naturalistas estrangeiros, algo da cultura e língua indígenas pôde ser preservado para a posteridade, mas o resultado desses esforços, infelizmente, representa uma pequena parte de toda a diversidade linguística e cultural que havia no Brasil nos primeiros séculos após o seu “descobrimento”.

Dentre as teorias existentes acerca da origem do R caipira, algumas contam com um reduzido número de adeptos e a quantidade de trabalhos acerca das mesmas é pontual. Uma hipótese por contato com outra língua, mas que não conseguiu angariar muitos adeptos, se refere à possível influência norte-americana, devido ao influxo de imigrantes oriundos do sul dos Estados Unidos, depois da guerra civil que terminou em meados do século XIX. Essa hipótese, ao nosso ver, e em concordância com Head, é a menos plausível:

Quanto à possibilidade de influência da língua inglesa nessa particularidade da pronúncia caipira, mais uma vez as características das variedades da língua de presumível origem desfavorecem hipóteses de evolução determinada pelo contato entre as referidas formas de linguagem. É que a grande maioria dos imigrantes americanos que vieram ao Brasil depois da guerra civil americana (1861-65) era do sul dos Estados Unidos – justamente de regiões linguísticas não caracterizadas pela pronúncia retroflexa do “r” da língua inglesa. Aliás, a localidade onde os imigrantes americanos para o Estado de São Paulo se concentraram principalmente – atualmente o município de Americana – não se caracteriza pela pronúncia típica do “R caipira”. Além disso, nas outras regiões para onde imigraram elevados números de americanos na mesma época (por exemplo, no Estado do Pará), não se encontra a pronúncia típica do “R retroflexo” (HEAD, 1987, p. 12).

As hipóteses mais robustas que nos propusemos a analisar decorrem de duas perspectivas linguísticas que se referem a fatores de natureza distintas: um de âmbito interno e outro de âmbito externo. A perspectiva externa se apoia na tese da origem da variação em comento como resultado da influência do contato do Português Europeu com as línguas

indígenas e africanas. A perspectiva interna concebe o surgimento do R caipira como resultado de um processo interno da própria língua. As duas diferentes perspectivas são detalhadas a seguir.

### **5.1 Hipótese de natureza externa: produto do contato entre línguas**

Dentre as hipóteses vinculadas às variações e mudanças linguísticas como fruto de um fenômeno externo, destacamos, no caso do Brasil, a intensa situação de bilinguismo ou mesmo multilinguismo encontrada na sociedade colonial. O colonizador português falava diferentes dialetos, pois a colônia recebeu portugueses oriundos de diversas regiões de Portugal. Somado a isso, o território era ocupado por indígenas das mais diversas etnias e falantes de centenas de dialetos diferentes. Como se esse cenário já não fosse suficientemente babélico, em meados do século XVI, inicia-se a importação de milhões de africanos, também falantes de línguas diferentes. Assim, na tentativa de pelo menos encontrar pistas que expliquem a provável origem do nosso singular R caipira, faz-se necessário analisar a influência dos indígenas e dos africanos na formação do português falado no Brasil.

#### **5.1.1 A diversidade das línguas indígenas e a Língua Geral**

No século XV, quando da chegada dos primeiros europeus, fontes estimam que houvesse entre mil e duas mil línguas faladas nas Américas sendo que, de acordo com os escritos de Cardim (1980[1625], p.101-6), apenas no território que hoje corresponde ao Brasil, por volta de mil línguas diferentes.

Estudiosos de períodos posteriores afirmam que se trata de um quantitativo superestimado. Na verdade, não havia uma “babel nos trópicos” como muitos apregoavam, visto que as variações não eram suficientemente expressivas a ponto de criar situações de ininteligibilidade entre as diversas variedades dialetais. As fontes para a recuperação do saber sobre as línguas do Brasil quinhentista são escassas. Relativos ao século XVI, existem apenas dois documentos remanescentes: a *Arte de gramática da língua mais usada na costa do Brasil*

(*Arte de Grammatica da Lingva Mais Vsada na Costa do Brasil/ 1565*), de Anchieta e o *Vocabulário na língua brasílica* (séc. XVI), constituído por um dicionário português-tupi, compilado pelos missionários jesuítas, mas apenas publicado em 1938. No contexto dos primeiros anos da colonização, proliferam relatos de viagens nos quais encontram-se majoritariamente descrições da fauna, flora, geografia e dos povos aborígenes do Novo Mundo, mas menções sobre suas línguas estão diluídas e esparsas.

Ao longo de três séculos, os missionários católicos foram enviados ao Novo Mundo americano com a dupla função de catequizar e ensinar leitura e escrita, produzindo para os índios e para os colonos cartilhas, catecismos, hagiografias, e para auxiliar os próprios religiosos, as gramáticas, sermões e confessionários. A produção de gramáticas da América espanhola foi muito superior à da América portuguesa, sendo que a cidade do México possuía imprensa própria desde 1539. Diz-nos Altman:

Ao contrário, o Brasil só teria sua própria casa impressora no início do século XIX e, somente no século XX, uma cátedra para o estudo do tupi antigo. Assim, não é de se estranhar que, das eventuais centenas de línguas originalmente existentes em território português nos séculos XVI e XVII, só tenhamos tido a descrição gramatical de duas delas: do tupi (= tupinambá), elaborada pelos jesuítas José de Anchieta e Luís Figueira (1575-1643), e do kiriri, elaborada pelo também jesuíta Luís Vicencio Mamiani (1652-1730) (Anchieta, 1990 [1595]; Figueira, 1880 [1687] e Mamiani 1877 [1699]), todas escritas em português e publicadas, pela primeira vez, em Portugal (ALTMAN, 2003, p. 63).

No contexto colonial americano, as chamadas línguas gerais foram as línguas indígenas inicialmente escolhidas pela administração colonial e pela Igreja como língua franca de contato entre as etnias existentes. Essas línguas autóctones foram “reduzidas”, serviram de substrato para elaboração dos primeiros dicionários e gramáticas confeccionados pelos missionários religiosos: caso do *náhuatl* (língua falada por parte dos nativos do que hoje se conhece por México), do quéchua (uma das línguas autóctones da região andina), do guarani (falado principalmente na região oeste do estado brasileiro do Paraná e no atual Paraguai) e o tupi (variedade que predominava na costa brasileira nos séculos XVI e XVII).

De acordo com os dados etnográficos disponíveis, os aldeamentos tupinambás compunham-se de uma população bastante elevada para a época e se estendiam desde onde hoje se situa o atual estado do Pará até o Rio de Janeiro. A exceção era a divisa entre o Ceará e o Maranhão, a região da foz do rio Paraíba, a região limítrofe entre o sul da Bahia e o norte do Espírito Santo, visto que, nessas regiões, predominavam grupos étnicos pertencentes ao Tronco Macro-Jê.

O tronco linguístico Tupi é constituído por dez famílias com suposta origem comum: Arikém (AR), Awetí (AW), Jurúna (JU), Mawé (MA), Mondé (MO), Munduruku (MU), Puruborá (PU), Ramaráma (RA), Tuparí (TU) e Tupi-Guarani (TG).

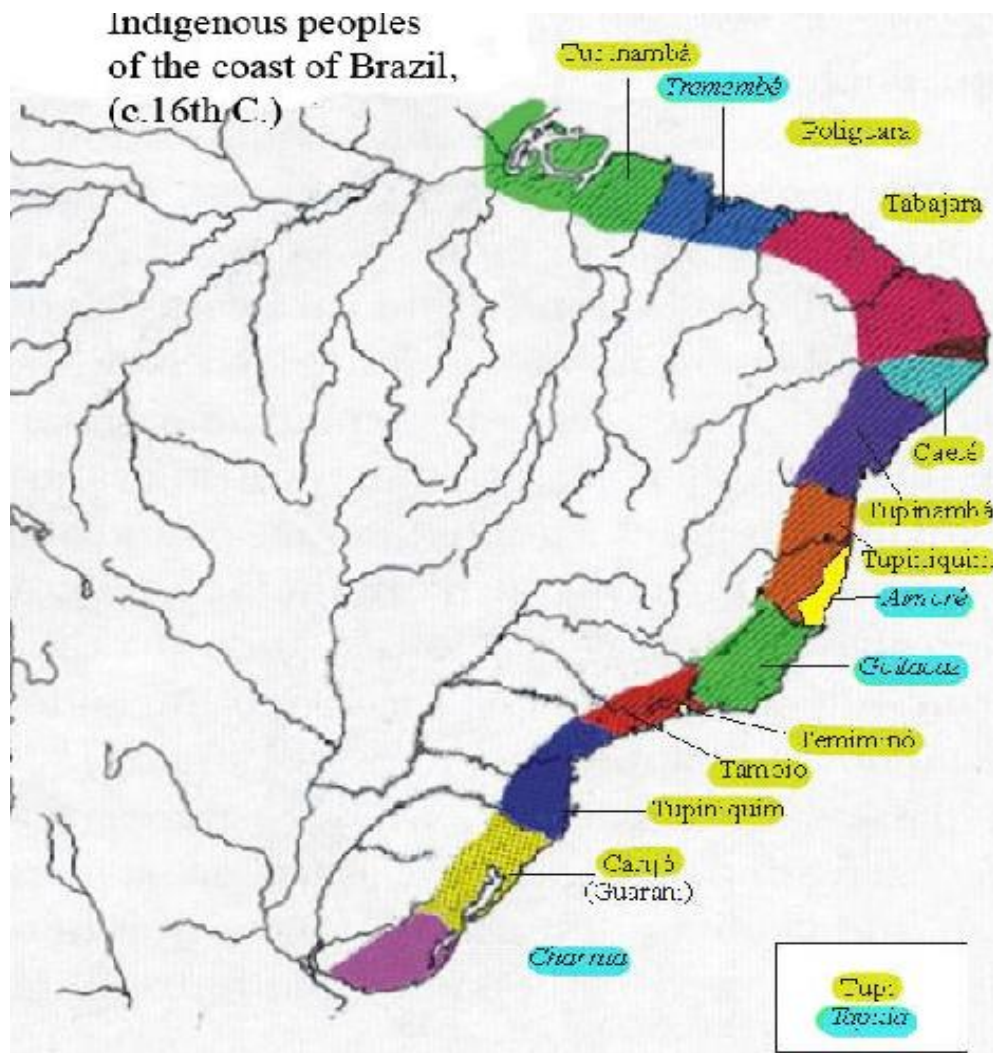
A língua Tupi-Guarani é um grupo linguístico que se divide em oito subgrupos, e um deles pertence aos Tupinambás, sendo suas características mais Tupi do que Guarani. Segundo Métraux (1948, p. 95):

Tupinambá. – Esse nome é aplicado aqui a todos os Índios, falantes de um dialeto Tupi-Guarani, que, no século XVI, foram os senhores da costa do Brasil, desde a foz do Rio Amazonas, até Cananéia, no sul do estado de São Paulo. Embora linguística e culturalmente relacionados de maneira muito próxima, esses índios eram divididos em muitas tribos, que empreendiam guerras desumanas umas contra as outras. À maior parte desses grupos foram dados nomes diferentes pelos colonizadores Portugueses e Franceses, mas o termo Tupinambá foi aplicado a tribos de regiões largamente separadas, como Rio de Janeiro, Bahia e Maranhão. Por serem essas as tribos melhor conhecidas, nós vamos, por conveniência, aplicar a todas elas o termo Tupinambá (MÉTRAUX, 1948, p. 95).

Nos séculos iniciais da colonização portuguesa no Brasil, a língua dos índios Tupinambás (pertencente ao tronco linguístico Tupi-Guarani) era falada em grande parte do litoral brasileiro, conforme se verifica no mapa a seguir:



Figura 5 – Mapa da distribuição dos povos Tupi e Tapuia no litoral do Brasil, às vésperas do colonialismo do século XVI



Fonte: <https://br.pinterest.com/pin/415808978073757218/>. Acesso em: 14 fev. 2021.

De acordo com Rodrigues (2001), o Tupi é a língua presente na maior parte do território brasileiro, com registros de ocorrência no Amapá, parte do Pará; no Paraná, Santa Catarina e Rio Grande do Sul; no litoral atlântico e em Rondônia.

Em meados do século XVI, essa língua Tupi passou a ser aprendida pelos exploradores e colonos portugueses. Paulatinamente, o uso da língua, chamada de *Brasílica*, intensificou-se e generalizou-se a ponto de ser falada por quase toda a população do período colonial brasileiro.

Adicionalmente, os jesuítas a incorporaram como instrumento de catequização das populações indígenas sob sua influência. Como dito anteriormente, o padre jesuíta José de Anchieta, inclusive, elaborou uma gramática (*Arte de Gramática da língua mais usada na*

*Costa do Brasil*), publicada em 1595. Em 1618, seguiu-se a publicação do primeiro Catecismo na Língua Brasílica e, em 1621, um manuscrito com o dicionário dos jesuítas, um vocabulário na Língua Brasílica. Altman relata:

José de Anchieta chegou à Bahia em 1553, indo quase imediatamente para São Vicente e para o Planalto do Piratininga, onde permaneceu por onze anos, até 1565. Foi nesse período que aprendeu a língua indígena, o tupi de São Vicente (ou tupiniquim), língua em que provavelmente escreveu a primeira versão da gramática encomendada pelo Pe. Luís da Grã (1523-1609) para o Colégio da Bahia, para onde, de fato, foi levada por volta de 1560 (Rosa, 1998 [1995], p. 276; Rodrigues, 1998, p. 73). Esta variante, o tupiniquim, falada na região de São Paulo e São Vicente, é que teria dado origem à chamada Língua Geral Paulista (LGP), usada na expansão bandeirante do sul e sudoeste do país: São Paulo, Minas Gerais, sul de Goiás, Mato Grosso, norte do Paraná, de meados do século XVII a meados do século XVIII (Rodrigues, 1994). Como tal e à exceção do dicionário de von Martius (1683), não lhe restam outros documentos (ALTMAN, 2003, p. 68).

Em meados do século XVII, essa língua, já bastante modificada pelo uso cotidiano por índios catequizados e não índios, passou a ser conhecida por Língua Geral. Havia dois ramos da Língua Geral no Brasil-colônia: a paulista e a amazônica; entretanto foi a primeira que marcou profundamente o vocabulário popular do português brasileiro, ainda de uso corrente, e que produz a falsa impressão de que a língua dos índios era apenas o tupi.

Com exceção do litoral, logo após o início da colonização, essas duas línguas de origem indígena, chamadas de línguas gerais, exerceram papéis de grande importância não apenas em termos demográficos, mas em termos das relações que estabeleceram, visto que os portugueses faziam uso delas nos espaços extraoficiais, utilizando-as em suas relações sociais.

Assim, houve a formação da Língua Geral Paulista (LGP), surgida em contexto de intensa miscigenação entre homens portugueses e mulheres indígenas, incentivados pela prática cultural dessas tribos, que consistia em dar a um estranho, como esposa, uma jovem da tribo:

A instituição social que possibilitou a formação do povo brasileiro foi o cunhadismo, velho uso indígena de incorporar estranhos à comunidade. Consistia em lhes dar uma moça índia como esposa. Assim que ele a assumisse, estabelecia, automaticamente, mil laços que o aparentavam com todos os membros do grupo. (RIBEIRO, 2004, p. 81).

Na verdade, a Língua Geral Paulista se originou da língua dos índios Tupi da Capitânia de São Vicente e do alto rio Tietê, e era ligeiramente diferente da língua dos Tupinambás e, no século XVII, era falada pelos bandeirantes paulistas. Por meio das Entradas

e Bandeiras paulistas, a Língua Geral Paulista penetrou em áreas jamais alcançadas pelos índios Tupi-Guarani, interiorizando-se e não mais se restringindo ao litoral brasileiro.

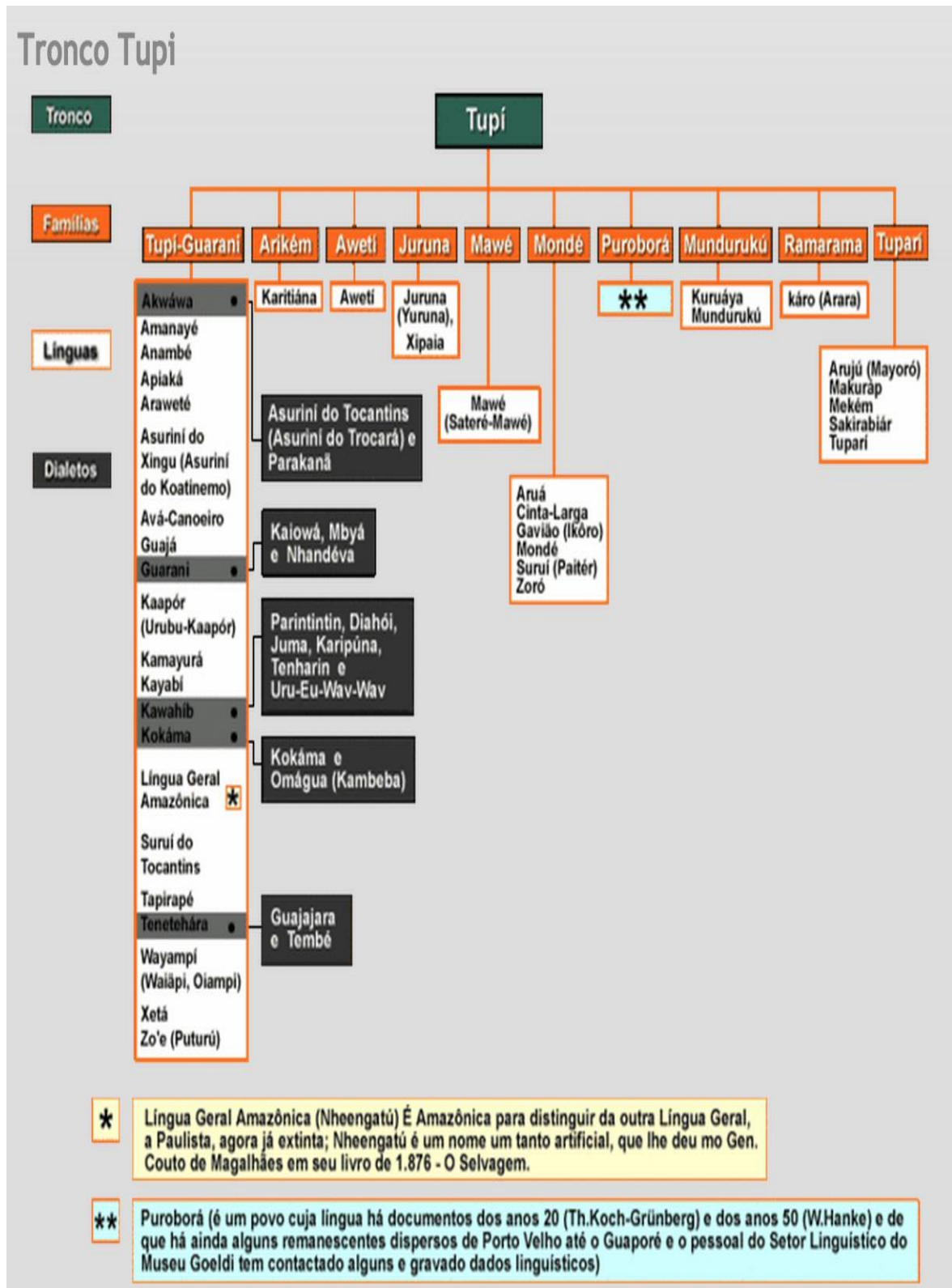
Existe uma certa divergência quanto a algumas denominações dadas aos dialetos e tribos causada por certas inconsistências em registros antigos que não deixavam explícita a existência de não apenas uma língua geral mas de “línguas gerais”. Assim, cabe um esclarecimento, mesmo que superficial, sobre a diversidade linguística aqui encontrada.

O que conhecemos por “tupi” na tradição brasileira a partir do século XIX corresponde a uma realidade linguística complexa. De acordo com Rodrigues (1996) e Dietrich (2010), as línguas gerais da época colonial - a língua brasílica, a Língua Geral Paulista (LGP) e a Língua Geral Amazônica (LGA) - se baseiam no tupinambá, extinto desde a primeira metade do século XVIII, uma das línguas da grande família linguística Tupi-Guarani.

Os estudos comparativos propõem a existência de aproximadamente 42 famílias linguísticas genéticas, das quais dez compõem o tronco Tupi, enquanto entre doze a quinze constituem o tronco Macro-Jê.

O Tronco Tupi tem grande importância pelo fato de ser um dos maiores agrupamentos linguísticos da América do Sul. De acordo com os estudos comparativos desenvolvidos por Rodrigues (1985, 2013), o tronco Tupi possui dez famílias linguísticas: Arikém, Awetí, Jurúna, Mawé, Mondé, Puruborá, Munduruku, Ramaráma, Tupari, e Tupi-Guarani. Essas línguas podem ser subdivididas em dois ramos: as do ramo ocidental e as do ramo oriental. Uma visão geral do tronco tupi com suas ramificações pode ser conferida no quadro abaixo:

Quadro 3 - Ramificações do tronco linguístico Tupi



Fonte: DICIONÁRIO ilustrado Tupi-Guarani. Disponível em: <https://www.dicionariotupiguarani.com.br/voce-sabia-que-nao-existe-a-lingua-Tupi-Guarani/>. Acesso em: 15 fev. 2021.

Conforme Rodrigues, enquanto a família Tupi-Guarani vem sendo documentada desde o século XVI, a documentação sobre as línguas das outras famílias se deu somente a partir do século XVIII. Examinamos no tópico seguinte aspectos internos à família Tupi-Guarani.

As línguas da família Tupi-Guarani encontram-se bastante distribuídas pela América do Sul, apresentando línguas situadas desde a Guiana até a Argentina, por um lado, e da Bolívia ao litoral do Brasil, por outro.

De acordo com Rodrigues (1984-1985), as línguas Tupi-Guarani se subdividem em, pelo menos, oito subgrupos cuja subclassificação toma como referência propriedades essencialmente fonológicas e morfossintáticas que essas línguas compartilham entre si. Ao todo, propõem-se oito ramificações, conforme se vê pela classificação (Dietrich, 2010) a seguir:

- Grupo Guarani Meridional.
- Guarani da Região Guaporé-Mamoré-Paraguai-Paraná.
- Grupo Tupi da Costa Brasileira.
- Grupo Asurini-Tenetebara-Tapirapé.
- Grupo do Xingu-Tocantins-Gurupi.
- Grupo do Norte de Mato Grosso e de Rondônia.
- Grupo do Alto Xingu.
- Grupo Amazônico Setentrional.

Dos grupos citados acima, merecem destaque:

- Grupo Guarani Meridional: nesse grupo, dentre outros, encontra-se o chamado Guarani clássico ou guarani antigo, cuja língua foi documentada pelos missionários jesuítas nas Reduções nos séculos XVII e XVIII.

- Grupo Tupi da Costa Brasileira: nesse grupo, além de outros dialetos de menor significância numérica, encontram-se:

a) o Tupinambá (a língua dos índios da costa brasileira nos séculos XVI e XVII) empregado pelos jesuítas nos primórdios da colonização e tornou-se a língua brasílica. Disseminou-se, pelas contínuas migrações dos tupinambás, nas regiões entre Santa Catarina e Bahia, estendeu-se posteriormente ao Maranhão e entrou na Amazônia no século XVII, dando origem à Língua Geral Amazônica (LGA), conhecida atualmente por Nheengatu.

b) o Tupiniquim: língua historicamente conhecida, mas não documentada, atualmente extinta.

c) o Potiguara: língua também historicamente conhecida, mas não documentada, atualmente extinta.

### 5.1.2 Características linguístico-estruturais do Tupi-Guarani

De acordo com Barbosa (1956), o Tupi identifica um conjunto de 31 fonemas, dos quais doze são vogais (orais a, e, i, o, u, y, que também podem ser nasais) três semivogais e dezesseis consoantes.: Todas as vogais podem ser também nasais e a nasalidade tanto vocálica como consonântica é um dos fatores principais na fonologia Tupi-Guarani, de acordo com Dietrich (2010).

Do ponto de vista morfológico, Dietrich (2010) observa que as línguas desse tronco apresentam elementos de todos os tipos: isolante, aglutinante, flexivo e incorporante. Sem nos aprofundarmos nesse tipo de análise, por sinal bastante complexa e que fugiria ao tema central proposto, cabe destacar uma peculiaridade relacionada ao caráter aglutinante da maioria das línguas tupis-guaranis.

Estudos de reconstrução do Tupi antigo, amparados nas línguas remanescentes como o nheengatu e o guarani permitem caracterizar tal ramo como pertencente ao tipo aglutinante pois, de acordo com Dietrich (2010), as línguas Tupis-Guaranis possuem grande número de sufixos e prefixos nominais e verbais. Assim, as línguas desse grupo possuem a faculdade de permitir aos falantes construir sintagmas complexos combinando vários sufixos numa ordem bem estabelecida.

(...) os nomes podem apresentar sufixos (...) de grau, como diminutivos, aumentativos, intensivos, de aspecto, de quantificação, de negação, de diversos tipos de nominalização e, no tupinambá, até de caso sintático. Os verbos podem ter prefixos de pessoa, de modo, de voz, de nominalização e sufixos de tempo, de aspecto e de nominalização (DIETRICH, 2010, p. 19).

Com relação às consoantes, em especial ao fonema /r/, Barbosa (1956) assim como outros autores, menciona que “O r é sempre brando, mesmo no princípio da palavra: roy: frio”.

Embora muitos autores tenham mencionado, como o próprio Barbosa e Amaral, o caráter “gutural” do tupi, não nos deram maiores detalhes sobre o que entendiam por gutural.

Amaral um pouco mais específico descreveu o R caipira, relatando, dentre outras características, seu som “gutural”.

Bastante conhecida é a declaração feita por Pero de Magalhães de Gândavo em sua obra *Historia da prouincia sa[n]cta Cruz, a qui vulgarme[n]te chamamos Brasil*:

A língua de que usam toda pela costa é uma [...]. Carece de três letras, convém a saber, não se acha nela f, nem l, nem R, cousa digna de espanto, porque assi não têm Fé, nem Lei, nem Rei: e desta maneira vivem desordenadamente sem terem além disto conta, nem peso, nem medida (GÂNDAVO, 1576, cap. 10, fl. 33 v.).

A maioria dos que defendem como indígena a origem do R caipira, bem como do fenômeno de alternância das vogais *l* e *r*, costumam se referir ao desconhecimento desses fonemas pelas populações aborígenes que aqui se encontravam; entretanto, a ocorrência do R retroflexo, concentra-se na região linguisticamente conhecida como pertencente ao dialeto caipira (considerável área do território centro-sul do país) e não espalhado por todo o território nacional como eram os dialetos Tupis-Guaranis, o que nos faz desconfiar da fragilidade de uma correspondência biunívoca. Assim, fazem-se necessárias algumas considerações acerca da Língua Geral Paulista, doravante (LGP), ao que tudo indica, mãe do dialeto caipira.

### 5.1.3 A Língua Geral Paulista

Nos primeiros anos da colonização, foram muito frequentes as relações entre homens portugueses e mulheres indígenas dada a escassez de mulheres brancas e também devido ao estímulo desse tipo de união inicialmente favorecido por líderes indígenas com a intenção de firmar alianças com os estrangeiros.

Em pouco tempo, o número de mestiços (filhos de pais portugueses e mães indígenas) começou a crescer. Essa prole mestiça, na maioria dos casos, crescia, com alguma variedade indígena como língua materna e aprendendo o português (em diferentes níveis) como segunda língua. Houve o predomínio do bilinguismo entre os habitantes da colônia. No primeiro século da colonização, o fenômeno foi mais intenso e profundo no planalto de Piratininga, em São Paulo.

(...) Aí o número de mestiços, chamados então de mamelucos, cresceu tanto usando a língua tupi (ligeiramente modificada pelo convívio com a portuguesa) que esta foi difundindo-se e passou a ser chamada *língua geral*, hoje mais especificada como *língua geral paulista* (LGP). Os mamelucos passaram a integrar as expedições ou bandeiras dirigidas mais para o interior de São Paulo, Minas Gerais, Goiás, Mato Grosso, levando consigo o uso das duas línguas, português e língua geral (...) A LGP, derivada do tupi, expandindo assim (...) vigorou da segunda metade do século XVI até a segunda metade do século XIX. Seu declínio e o consequente exclusivismo do português, nessas regiões, deveram-se principalmente à progressiva extinção dos indígenas tupi e à crescente introdução, ou imigração, naquelas regiões, tanto de escravos africanos e de mestiços de outras áreas do Brasil quanto, em meados do século XIX de novos colonos europeus (DIETRICH, 2010, p. 47-38).

A respeito do predomínio da língua Tupi na Capitania de São Vicente no século XVII, existe o seguinte relato do jesuíta Fernão Cardim:

Em toda esta provincia ha muitas e varias nações de diferentes linguas, porém uma é a principal, que comprehende algumas dez nações de Índios. Estes vivem na costa do mar, e em uma grande corda do sertão, porém são todos estes de uma só língua, ainda que em algumas palavras discrepam, e esta é a que entendem os Portuguezes; é fácil, elegante, e suave, e copioza; a dificuldade d'ella está em ter muitas composições; porém dos Portuguezes, quazi todos os que vêm do Reino e estão cá de assento e comunicam com os Indios a sabem em breve tempo, e os filhos dos Portuguezes cá nascidos a sabem melhor que o Portuguez, assim homens como mulheres, principalmente na Capitania de São Vicente (CARDIM, 1894, p. 205).

Diferentemente da LGA, a LGP não foi muito documentada. Segundo Leite (2013), sobre a LGP, temos somente três documentos de registro: um dicionário de verbos, do século XVIII, conhecido por sua publicação no *Glossaria Linguarum de Martius* (1863); um vocabulário, ao que tudo indica também do século XVIII, e uma lista de palavras colhidas em comunidades de mestiços de índios bororo e negros no início do século XIX. Para Rodrigues (1996), o *Dicionário de Verbos* constitui, hoje, a principal fonte sobre a LGP.

Segundo alguns autores, como Melo (1975), Noll (2008) e Rodrigues (2010), embora o dialeto caipira tenha recebido bastante influência da LGP, essa herança se restringiu ao léxico. Para Melo (1975), “o tupi terá sido uma das línguas das Bandeiras (...). Assim se compreende como, em regiões onde jamais habitou uma tribo tupi, a nomenclatura geográfica ateste a presença de considerável soma de topônimos oriundos da língua indígena (42-49)”, entretanto, o mesmo autor ressalta:

(...) Depois o português reagiu e (...) rápido desterrou a língua indígena, que hoje é falada apenas por reduzidas populações localizadas na Amazônia. (...) Porém dessa longa convivência com o tupi, o português não poderia sair incólume. (...) Qual a extensão e intensidade da influência tupi no português? (...) imensa foi a contribuição do tupi ao vocabulário da língua portuguesa americana. (...) Há, porém, um campo em que a hesitação é muito maior e as conclusões a que se tem chegado



se tornam passíveis de sérias críticas. Refiro-me às influências apontadas na fonética, na morfologia e na sintaxe do português (MELO, 1975, p. 42-49).

Para Noll (apud DIETRICH, 2010), autores como Teodoro Sampaio (1987 [1901]), Clóvis Monteiro (1926; 1959: 75-136), Francisco da Silveira Bueno (1973) e Antônio Geraldo da Cunha (1999: 18), dentre outros, defenderam um possível substrato tupi para explicar certas particularidades da pronúncia brasileira; entretanto, segundo o mesmo autor, “(...) Para os linguistas modernos (...), é evidente que os filólogos citados careciam de conhecimentos tanto das regras morfofonemáticas como da morfossintaxe das línguas desta família. (...)”, e ainda:

(...) Não há influência tupi nem na fonética, nem na morfologia do português brasileiro. (...) as particularidades discutidas que distinguem o Português do Brasil do português europeu se explicam por evoluções internas, não por influência da língua brasileira. Esta só se reflete no léxico e nos nomes (toponímia, hidronímia, coronímia e antroponímia) (Noll in DIETRICH, 2010, p. 85-86).

Um dos mais respeitados especialistas em línguas indígenas no Brasil, Aryon Dall’Igna Rodrigues, compartilha de opinião semelhante em relação às influências das línguas tupi no Português do Brasil, embora se posicione de forma mais moderada, admitindo a dificuldade na identificação das modificações sofridas pela língua portuguesa falada no Brasil por influência de uma das línguas indígenas:

Por um lado, o português do Brasil é um mosaico de variedades fonético-fonológicas, morfofonemáticas, estilísticas e lexicais, ainda carente de documentação, análise e descrição. Por outro lado, o estado atual dessas variedades já é o produto de reajustes e reorganizações associados à migrações e à ampliação e intensificação do ensino escolar, nem espaço de tempo de perto de cinco séculos. (...) poucas associações podem ser feitas entre fenômenos das línguas tupi e tupinambá e fenômenos correspondentes em variedades regionais do português do Brasil. Por exemplo: no âmbito da fonética (...) mais complicada, entretanto, é a situação com respeito à consoante lateral alveolar *l*, a qual foi alterada no fim de sílaba, mas conservada no início de sílaba: *alta* [‘alta] > [‘arta] (...) (RODRIGUES in DIETRICH, 2010, p. 44-45).

Com relação às várias realizações do R, Rodrigues acrescenta:

Dificuldade adicional ocorre quando as línguas influenciadoras divergem, como no caso da consoante vibrante alveolar *r*. Tanto o tupinambá como o tupi tinham o som [r], “vibrante” (rótico) simples ou *flap*, como o do português *caro* [‘karnu], mas não tinham o som *r*, vibrante (rótico) múltiplo sonoro, como em *carro* [‘karu], que pode ter sido substituído pelo vibrante simples sonoro *r* (*caro* [‘karu], pelo fricativo velar sonoro [ʀ] ([‘kaʀu]), pelo fricativo velar surdo [x] ([‘kaxu]) ou pelo fricativo glotal surdo [h] ([‘kahu]), variantes que ocorrem efetivamente em diversas variedades do português do Brasil. Em posição final, o tupinambá tinha o “vibrante” simples (p.

ex., *ajúr*, ‘eu vim’, *ere ár*, ‘você caiu’), mas o tupi e a LGP tinham “zero” (ausência de som) posição (*ajú* ‘eu vim’, *ere ´a* ‘você caiu’), logo, é possível que a situação do tupi e da LGP se correlacione com a de algumas variedades do português nas áreas em que este conviveu com aquelas línguas (p. ex., *parar* [pa’rar] > [pa’ra]). Entretanto, é claro que a ausência do *r* final nas regiões em que se falava o tupinambá não se pode explicar pelo contato com essa língua (Rodrigues in DIETRICH, 2010, p. 44-45).

Ainda em relação à ocorrência desse R peculiar na região correspondente à zona na qual floresceu a Língua Geral Paulista, salienta Melo:

Tal *r* é característico do Norte de São Paulo e Sul de Minas, e nele se transforma sistematicamente o -l final de sílaba. Quem já viajou por aquelas bandas sabe que, basta transpor-se a Mantiqueira, aparecem meninos vendendo “paster de carne”, com o seu errezinho particular. Esta consonância aproxima-se bastante do *r* final de sílaba americana, que figura por exemplo em *form*, porém é mais intenso. Está para o *r* inglês como o nosso *r* inicial está para o *j* espanhol ou *ch* alemão (MELO, 1975, p. 106).

Diante do que temos visto até o momento, mesmo aqueles estudiosos que consideram o contato entre línguas como a possível causa das variedades encontradas no Português do Brasil, no que se refere às variedades fonéticas, a incerteza é ainda maior, dada a escassa quantidade de registros descritivos detalhados sobre as particularidades sonoras/fonológicas/fonéticas das línguas indígenas faladas à época da colonização. Outra dificuldade, como já mencionada por Rodrigues, diz respeito ao aspecto diacrônico das análises, da amplitude temporal e das transformações prévias ocorridas.

Uma hipótese recentemente levantada tenta vincular a origem do R caipira ao contato com as línguas indígenas do tronco Macro-Jê.

#### 5.1.4 O Tronco Macro-Jê

Embora boa parte do vulgarmente conhecido dialeto caipira seja creditada como fruto da influência do Tupi-Guarani, estudiosos têm apontado que algumas peculiaridades, inclusive fonológicas, desse dialeto são fruto da influência de outro tronco linguístico dos autóctones brasileiros que habitavam as regiões interiores do território e que falavam línguas derivadas não do Tupi-Guarani, mas do tronco Macro-Jê.

De acordo com os estudos linguísticos atuais, em especial a classificação feita por Rodrigues (1986), as línguas indígenas brasileiras pertencem a dois grandes trocos: o Tupi-

Guarani e o Macro-Jê, além de 19 famílias linguísticas e famílias de uma única língua, também denominadas “línguas isoladas”.

O tronco linguístico Macro-Jê abrange doze famílias e tem uma peculiaridade hipotética, devido ao seu descobrimento recente e poucas pesquisas relacionadas ao mesmo. De acordo com Boswood (1973) Lévi-Strauss e Nimuendajú, os grupos Jê ocupam a metade leste do planalto brasileiro.

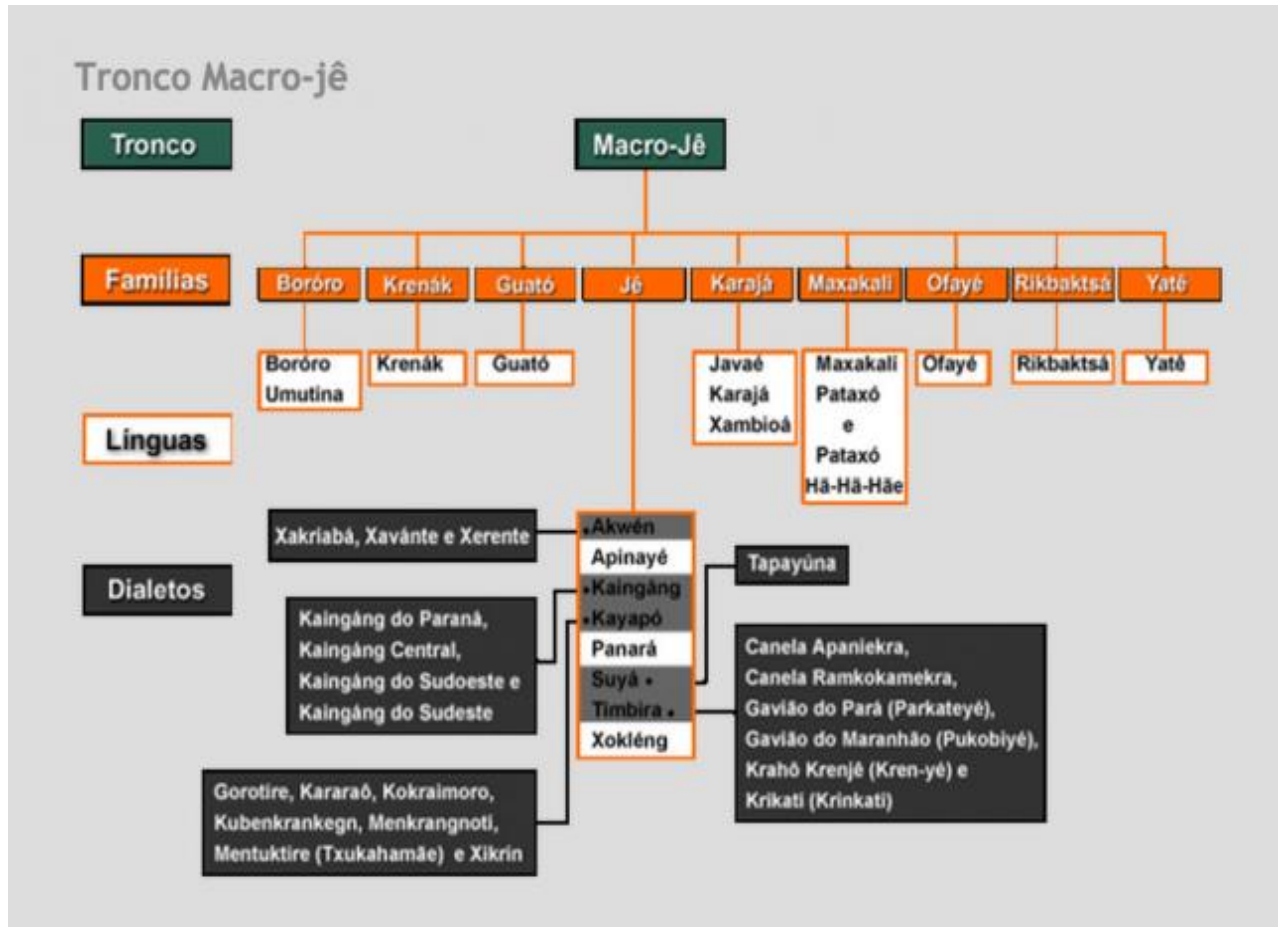
Segundo Rodrigues (1999), o Macro-Jê apresenta línguas distribuídas desde o Maranhão até o Rio Grande do Sul, no oeste e no leste da Amazônia, em Mato Grosso e em Mato Grosso do Sul, ao norte do rio Amazonas, nos estados do Amazonas, Roraima, Pará e Amapá, contando ainda com algumas línguas ao sul daquele rio, ao longo de seu afluente Xingu, nos estados do Pará e Mato Grosso. São pouco claras as evidências que existem até o momento para o reconhecimento de um tronco linguístico Macro-Jê, sendo que a família linguística Jê se apresenta como o constituinte maior e compreende línguas faladas especialmente nas regiões do cerrado, do sul do Maranhão e Pará, partes de Goiás e Mato Grosso, até os campos meridionais de São Paulo, Paraná, Santa Catarina e Rio Grande do Sul.

Cabe ressaltar que poucas línguas indígenas no Brasil foram estudadas extensamente e o conhecimento sobre elas se encontra ainda em construção e em constante revisão. Além disso, recentemente, boa parte do acervo de estudos linguísticos produzido por Nimuendajú foi destruída durante o incêndio do Museu Nacional do Rio de Janeiro em 2018. Infelizmente.

Duarte (2016), citando Rodrigues (1999), ressalta que “o termo Macro-Jê foi proposto inicialmente por Mason (1950: 287). Este termo cobria uma quantidade de línguas que estavam correlacionadas com a família linguística Jê. (p. 18)”. Salaria ainda que o termo foi criado por Martius (1867) com a finalidade de eliminar os termos “Tapuya” e “Tapuya-Jê” utilizados por outros autores no passado. Além disso, não há consenso entre os linguistas sobre quantas famílias existem de fato nesse tronco linguístico, pois não há na literatura informações convergentes sobre o número exato de línguas que constituem o tronco Macro-Jê. Assume-se, em geral, que existam de doze a dezesseis famílias.

As línguas que até momento conhecemos da família Jê se subdividem nos seguintes grupos:

Quadro 4 - Ramificações do tronco linguístico Macro-Jê



Fonte: DICIONÁRIO ilustrado Tupi-Guarani. Disponível em: <https://www.dicionariotupiguarani.com.br/dicionario/macro-je/>. Acesso em: 15 fev. 2021

Um grupo de destaque para o nosso estudo é o *Kaingang*, o grupo linguístico mais diferenciado dentro da família Jê e que se distribui nos estados do centro sul, em especial, São Paulo e Paraná.

O povo indígena Kaingang é o terceiro em número de população no Brasil, com 37.470 indivíduos. Seus integrantes vivem nos três estados da região Sul do país e no oeste de São Paulo. São falantes da língua que dá nome à etnia e identificados geneticamente à família Jê e ao tronco linguístico Macro-Jê. O contato com a sociedade não índia foi intensificado na segunda metade do século XIX e início do século XX quando vigorava a política de pacificação dos povos indígenas em território nacional. Esse contato gerou consequências drásticas como a expropriação territorial, os conflitos internos, as doenças e mortes.

De acordo com Wilmar D'angelis, importante pesquisador do Kaingang, a maioria das línguas e povos da família Jê vive bem mais ao norte: os Xavante (Mato Grosso), os Parakatéye (Pará), os Mebengokre, conhecidos como Kayapó (Pará e Mato Grosso), os

Xerente, os Krahô, os Apinayé (Tocantins), os Apaniekrá, os Pukobyé, os Krinkati (Maranhão) e alguns outros.

A língua Kaingang, por muito tempo, foi considerada de difícil aprendizagem, tratada como língua “travada”, o que significa que apresenta aspectos estruturais e, principalmente fonéticos, bastante diferentes das línguas românicas. A dificuldade de aprender a língua Kaingang é um dos motivos que explica o fracasso dos próprios jesuítas no trabalho religioso com esse povo. Telêmaco Borba conviveu por muito tempo com os Kaingang dos campos centrais do Paraná. O estranhamento que a língua falada pelos Kaingang causava em muitos exploradores e cronistas pode ser notado no seguinte fragmento, retirado da obra *Actualidade Indígena*:

Falam um idioma muito gutural, nada parecido ao Guarany; não pronunciam o L, o R forte e o Z. Das palavras que conheço, do idioma deles, só duas: Pirá, peixe, e Piraju, são da língua Guarany e têm a mesma significação em uma e outra; e kêfé, faca, que é parecida com Kicé, faca, do Guarany (BORBA, 1908, p. 19).

Um dos trabalhos linguísticos de maior significância em relação à língua Kaingang foi desenvolvido pela linguista alemã Úrsula Wieseemann, que iniciou seus estudos sobre esta língua, em 1958, no Posto Indígena Rio das Cobras, no município de Nova Laranjeiras, estado do Paraná. Segundo Wieseemann, existe a hipótese de que depois de se separarem de outros grupos Jê, do Brasil central, os Kaingang começaram sua migração para o sul. Uma segunda separação teria ocorrido entre o grupo de São Paulo e os demais, continuando estes a migrar em direção sul.

Após cruzarem o rio Paranapanema, os índios teriam se dividido em várias hordas, sendo inexistente o contato entre os grupos que atravessaram o rio Iguazu e aquelas hordas que se estabeleceram ao norte deste rio. Um dos grupos teria passado vários rios em direção sudeste, deixando de ter contato com os demais, ao passo que os grupos Central e Sudoeste mantiveram contato esporádico depois da separação (WIESEMANN, 1978, p. 216). Após anos de estudos, Wieseemann elaborou um dicionário kaingang-português-kaingang, o qual apresenta uma ideia geral de alguns aspectos dessa língua.

No oeste do que hoje é o estado de São Paulo, no período colonial, viviam indígenas das etnias Oti, Ofalê e Kaingang que falavam uma língua do ramo Macro-Jê. Os Otis e os Ofalês eram vulgarmente chamados também de Xavantes. Infelizmente, em meados do século XIX, estavam já praticamente extintos, entretanto, graças ao trabalho dos indigenistas Curt

Nimuendajú e Telêmaco Borba, foi possível o registro, mesmo que escasso, de traços da cultura e da língua desses povos.

Borba (1908), recolheu várias palavras do grupo Oti, em 1878, e identificou algumas peculiaridades do idioma falado por tal etnia. Os Otis, assim como outros povos pertencentes ao grupo Jê, possuíam sons que os grupos de língua tupi não tinham, como o /r/ forte, que Borba identificou como bastante gutural: o /r/ retroflexo, posteriormente descrito detalhadamente por Amadeu Amaral, conforme tratado anteriormente, em *O Dialeto Caipira*.

A observação de Borba é de extrema relevância, pois evidencia uma das possíveis origens do peculiar /r/ paulista, também chamado de R caipira, visto que essa marca fonética foi identificada nas línguas faladas pelos grupos Oti, Ofalê e Kaingang:

Ha sido opinião quasi geral dos escriptores que têm tratado das línguas dos nossos selvagens, que nellas faltam varias letras do nosso alphabeto, e entre elas o *r* forte, o *l* e o *z*; eu também até poucos annos, segui essa opinião; mas, em 1878, tendo tido ocasião de tratar com alguns selvagens da nação *Chavante*, que demoram nos Campos-novos da Província de S.Paulo, comarca de Botucatú, fiquei convencido de que aquella falta de letras não era tão geral como até aqui se acreditava, e que pelo menos na língua *Chavante* existem o *l*, o *r* forte e um *z* como o *th* inglez: é verdade que esta língua discorda completamente da língua geral de nossos selvagens; os *Chavantes*, até no physico, nada se parecem com as outras nações de indios que conheço.

Para facilitar algum estudo ethnographico, que porventura alguém queira emprender, junto a esta *notícia*, mais este pequeno vocabulario da língua *Caiguá* (que é a mesma *Guarani*, com pouca differença) e *Chavante*. Nesta, o *r* forte é muito gutural, o *J* sôa como no hespanhol e o *th* como no inglez (BORBA, 1908, p. 72).

Depreende-se do excerto de Borba, embora boa parte da Língua Geral Paulista tenha sofrido influência dos dialetos de origem Tupi-Guarani, que o processo de infiltração dessa língua no interior remoto do país levou a situações de contato e influências recíprocas com outros dialetos indígenas locais. Muitas pesquisas ainda são necessárias, porém não podemos descartar a probabilidade de que o R retroflexo seja mais uma herança do tronco Jê do que do tronco tupi, o que explicaria sua predominância nas regiões não litorâneas e mais ao centro sul do território brasileiro.

#### 5.1.5 As Línguas Africanas

Por mais de três séculos, a sociedade brasileira bem como sua economia viveram quase exclusivamente da mão de obra escrava, quase totalmente de origem africana (nas “peças” importadas ou nas suas descendentes). Para o Padre Antônio Vieira, o Brasil teria seu corpo na América e sua alma na África, dada o grande impacto da presença africana na formação do povo brasileiro.

As estatísticas sobre o fluxo histórico de negros para o Brasil, apesar da carência de fontes, refletem que a influência do elemento africano foi elevada, praticamente nunca interrompido e com picos significativos durante a expansão econômica do Brasil colonial. Segundo o historiador Laurentino Gomes:

O Brasil foi o maior território escravista do hemisfério ocidental por quase três séculos e meio. Recebeu, sozinho, quase 5 milhões de africanos cativos, 40% do total de 12,5 milhões embarcados para a América. Como resultado, é atualmente o segundo país de maior população negra ou de origem africana do mundo (GOMES, 2019, p. 24).

A escassez de fontes fidedignas impede a construção de números precisos, entretanto a aproximação das estimativas converge em torno da cifra citada por Laurentino (2019), conforme citado por Mendonça (2012), em levantamento com base em estatísticas aduaneiras, em documentos subsistentes nos Arquivos do Itamaraty, que apresenta a seguinte relação no quadro a seguir:

Quadro 5 - Números de escravos entrados no Brasil

<b>Números de escravos entrados no Brasil (Avaliação baseada em estatísticas aduaneiras subsistentes)</b>				
<i>Período</i>	<i>Regiões</i>	<i>Entradas anuais</i>	<i>Total anual</i>	<i>Total da importação</i>
Século XVI	Todo o Brasil	?	?	30.000
	Brasil holandês	3.000		
Século XVII	Brasil português	5.000	8.000	800.000
	Pará	600		
	Recife	5.000		
	Bahia	8.000		
Século XVIII	Rio	12.000		
			25.000	2.500.000
Século XIX (até 1830)	Todo o Brasil	?	50.000	1.500.000
<b>Durante o tráfico</b>				<b>4.830.000<sup>7</sup></b>

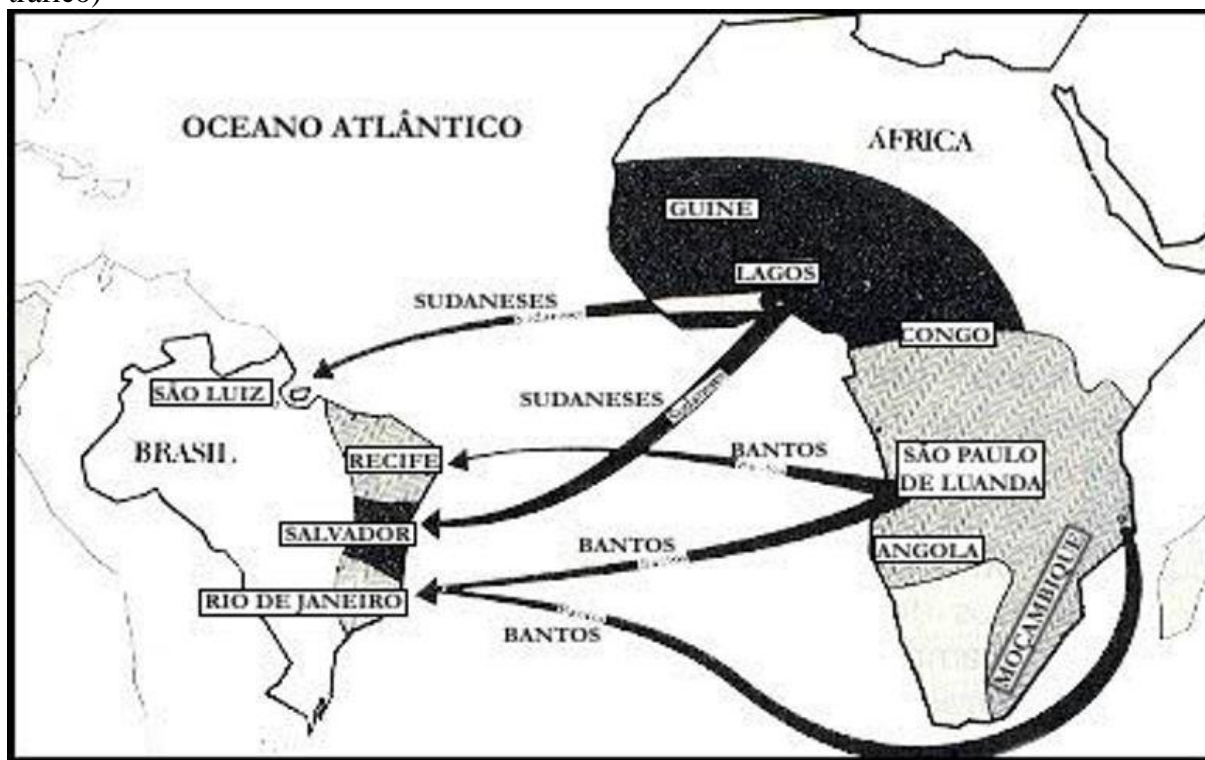
Fonte: tabela adaptada de MENDONÇA, 2012.

Com relação à procedência dessa massa de escravizados, os escravos trazidos para o Brasil, vieram de duas regiões do continente africano. Os *bantos*, importante grupo étnico das regiões Centro-Sul e Nordeste da África, regiões conquistadas pelos portugueses ainda no século XV, grosso modo compreendendo a região de Moçambique, Congo e Angola e os *sudaneses*, oriundos da Nigéria, Guiné e Costa do Ouro, predominavam na costa Centro-oeste do continente africano, conforme visualizados no mapa a seguir:

<sup>7</sup> A última coluna, conforme metodologia adotada pelo autor, corresponde ao número anual multiplicado pelo número de anos do período, no caso, por 100 (século).



Figura 6 - Mapa da escravidão nas Américas (locais de origem dos principais fluxos do tráfico)



Fonte: Portal Geledés. Disponível em <https://www.geledes.org.br/escravidao-nas-americas/>. Acesso em: 12 fev. 2020.

Dentre os pioneiros nos estudos relativos à influência linguística da matriz africana no português falado no Brasil, destacam-se Jacques Raimundo (com a publicação da obra *O elemento afro-negro na língua portuguesa*, em 1933) e Renato Mendonça, autor de *A influência africana no português do Brasil*, também de 1933. Tais autores trouxeram à baila a até então menosprezada importância da nossa herança africana em termos linguísticos.

De acordo com Yeda Pessoa de Castro, no prefácio da reedição da obra de Mendonça (2012), com base em pesquisas mais recentes, o mapa linguístico do continente africano, onde são faladas mais de 2.000 línguas, engloba em linhas gerais quatro grupos ou famílias etnolinguísticas, assim distribuídos:

- Afroasiática (hamíto-semítica) são as línguas da África do Norte.
- Khoisan (línguas de clique) cujos falantes concentram-se no deserto de Kalahari.
- Nilo-Saariana, são as línguas nilóticas do sul do Sudão e do Saara.
- Níger-Congo, a maior família linguística, abrange dois grandes grupos: o banto localizado abaixo da linha do equador, e o que é a oeste africano, ao longo da costa atlântica, do Senegal à Nigéria, na região do Golfo do Benin, com línguas tradicionalmente denominadas sudanesas. Entre elas, as do grupo linguístico gbe ou

ewe-fon, conhecidas no Brasil por minas ou jejes, e o iorubá falado na Nigéria Ocidental e no vizinho Benim, chamada de nagô.

A família nigero-congolesa é, sem dúvida, a mais significativa entre todas para a nossa História. A entrada de bantos ocorreu desde o início do tráfico para o Brasil no século XVI, com um fluxo ininterrupto até o século XIX, e foram dirigidos para todos os núcleos da colônia em formação que demandavam mão de obra escravizada.

Cumprе salientar que banto e iorubá são designações contemporâneas, no período colonial, os bantos era identificados por congos, angolas, benguelas, dentre outros, de acordo com a sua procedência africana, enquanto os iorubás, no Brasil, ainda são tradicionalmente conhecidos como nagôs.

Mendonça procura sistematizar os fatos da linguagem popular brasileira que lhe parecem resultado direto do contato de línguas africanas com o Português Europeu antigo, a partir da análise das transformações e adaptações resultantes dessa dinâmica de contato, ressaltando a interferência do quimbundo e do quicongo, duas línguas muito assemelhadas dentro do grupo banto, conforme a classificação linguística de Malcolm Guthrie em *The classification of the Banthu languages* (1948, Oxford).

A família negro-africana, na denominação de Mendonça,<sup>8</sup> ou Niger-Congo, de acordo com classificações mais recentes, como dissemos, congrega as línguas de matriz banto e as de matriz sudanesa e apresenta certos traços peculiares. Línguas sudanesas e bantu coincidem nas características gerais, o que supõe um ancestral linguístico comum.

Um dos traços distintivos das línguas da grande família negro-africana (banto-sudanês), especialmente no caso das banto, é a divisão do vocabulário em um número determinado de *classes*, que se distinguem entre si por *afixos* próprios a cada uma. A classe, por sua vez, forma uma espécie de sistematização das palavras por grupos.

Todavia o critério é diferente e na constituição da classe observam-se não só os prefixos (elemento material), mas também o sentido da palavra (elemento espiritual). Deste modo há uma classe constituída por seres humanos, outra pelos nomes de plantas, uma terceira pelos nomes abstratos, uma quarta pelos nomes de líquidos, e sucessivamente. (...) o afixo classificador de cada termo tem uma importância tão grande que se repete no correr da frase em todas as palavras que com ele se relacionam (MENDONÇA, 2012, p. 65-66).

A língua bantu é classificada como do tipo aglutinante, ou seja, apresenta uma tendência à aglutinação acentuada, pois costuma apresentar quantidade de termos enclíticos

---

<sup>8</sup> A exemplificação utilizada é a exposta por Mendonça (2012).

que, no uso diário, podem ser separados do radical pelos falantes. Para fins de elucidação, vejamos um exemplo fornecido pelo autor.

Na frase “nosso formoso homem aparece, nós o amamos”, a palavra “homem” equivale ao termo “*umuntu*” e cada vocábulo a ela ligado deve, em concordância com regras fixas, ter um prefixo que lembre o começo de *umuntu*; *mu* ou *u*, ou *w* ou *m*:

<b>umuntu</b>	<b>wetu</b>	<b>omuchle</b>	<b>uyabonakala</b>	<b>simtanda</b>
homem	nosso	bonito	aparece	amamos

A mesma frase no plural muda de prefixo: “homens” = *abantu*:

<b>abantu</b>	<b>betu</b>	<b>abachie</b>	<b>bayabonakala</b>	<b>sibatanda</b>
homens	nossos	bonitos	aparecem	amamos

No que se refere à influência de línguas africanas no Português do Brasil, Mendonça faz coro com João Ribeiro que, em 1888, em seu *Diccionario Grammatical* afirma que as alterações produzidas na linguagem brasileira por influência das línguas faladas pelos escravos introduzidos no Brasil, ao contrário do que se pensava e que era defendido por muitos intelectuais eugenistas da época, foram bem mais profundas tanto no léxico quanto no sistema gramatical da Língua Portuguesa. Entre essas alterações, mencionava a redução das formas verbais e a simplificação das flexões de plural presentes na fala popular.

Com relação ao aspecto que mais nos interessa nesses estudos, ou seja, peculiaridades fonéticas/fonológicas atribuídas às influências das línguas africanas no Português do Brasil, Castro (2001/2005) enfatiza que, quanto à tentativa para justificar a diferença entre a pronúncia da Língua Portuguesa no Brasil e em Portugal, sem dispor dos recursos de dados mais elaborados, Mendonça não passou de uma explicação muito simplificada, a de que “a pronúncia portuguesa possui mínima complexidade das sílabas, gerada pelo enfraquecimento das vogais e o acúmulo de articulações”.

Em outros termos, isso implica dizer que a proximidade relativa entre a estrutura fonológica do português antigo e das línguas do grupo banto, entre elas, o quimbundo ressaltado por Mendonça, potencializou a continuidade do tipo prosódico de base vocálica do português antigo na modalidade brasileira, afastando-a do Português de Portugal, cuja

tendência é a de eliminar vogais átonas, criando grupos consonânticos impronunciáveis pelos brasileiros (Yeda Pessoa de Castro (2001).

Em relação aos fenômenos fonéticos/fonológicos, Mendonça (1933) salienta algumas marcas importantes do quimbundo:

- o quimbundo tem as vogais *a, e, i, o, u*, ao lado das semivogais *y* e *w*.
- em quimbundo não existem ditongos e as combinações vocálicas *au, ai, eu, oi, ou* são dissilábicas e assim devem ser emitidas.
- as palavras do quimbundo, assim como as das línguas bantu, terminam sempre por vogal.
- as consoantes são, em maioria absoluta, simples: *b, f, j, p, h, k, l, m, n, ñ, r, s, t, v, x, z*.

Ainda, no que se refere às consoantes, de acordo com o ponto de articulação ou a maneira como são proferidas, podem ser classificadas de acordo com o seguinte quadro fonético:

Quadro 6 - Quadro fonética da consonância simples do Quimbundo segundo ponto de articulação na boca

		Velares	Palatais	Dentais	Labiais	
					Labiodentais	Bilabiais
Oclusivas		quê		tê		pê
		guê		dê		bê
Constritivas	Fricativas			sê zê	fê	
				chê jê	vê	
	Vibrantes			lê rê (fraco)		
	Nasais		nhê	nê		mê

Fonte: MENDONÇA, 2012, p.70.

Especificamente visando o objeto do nosso estudo, Mendonça salienta que o *r* é sempre brando e nunca forte ou desacompanhado de *i* como em *caricunda*. Ao tratar da influência africana no português, menciona a apócope que aparece em *l* e *r finais*:

general	<i>generá</i>
cafezal	<i>cafezá</i>
mel	<i>mé</i>
esquecer	<i>esquecê</i>
Artur	<i>Artú</i>

Ainda com relação ao *r*, observa a ocorrência de rotacismo. Segundo o autor, a inexistência do *r* nas línguas bantu deu origem à substituição do *r* forte português pela linguodental *l* ou o seu abrandamento em *r* fraco, como em:

rapaz	<i>lapassi</i>
carro	<i>calo</i>
era	<i>ela</i>
claro	<i>calado</i>
fora	<i>folá</i>

Para Aragão (2010), o rotacismo do *l* pelo *r* ocorre atualmente e com considerável frequência entre os falantes das zonas rurais brasileiras, exemplificando com a existência desse fenômeno na sentença “eu num gosto de farsidade”, donde temos “falsidade>farsidade”. Há também a presença de metátese (os sons trocam de posições um com o outro dentro de uma palavra) do “r” pós-vocálico no Português do Brasil como herança das línguas africanas , no exemplo “porque > praquê”; note-se o deslocamento do “r” após a vogal “o” para antes desse mesmo fonema. O mesmo fenômeno ocorre em alguns falares do norte de Portugal e na língua galega, conforme exemplo pautado na Real Academia Galega: “perguntar > preguntar”.<sup>9</sup>

Os censos, sobretudo a partir do século XVII, revelam que a população de africanos e de seus descendentes era numericamente superior ao número de europeus durante três séculos consecutivos:

<sup>9</sup> Conforme exposto no artigo Africanidades na fonética do Português Brasileiro. DOI: DOI: 10.18468/letras.2017v7n1.p269-287. Disponível em: < file:///C:/Users/lizha/Downloads/3005-12721-1-PB.pdf> acesso em: 28.05.2021.

Quadro 7 - Censos da distribuição étnica da população brasileira entre os séculos XVI e XIX

Tipos	1538 – 1600	1601- 1700	1701 - 1800	1801 - 1850	1851 - 1890
Africanos	20%	30%	20%	12%	2%
Negros brasileiros	—	20%	21%	19%	13%
Mulatos	—	10%	19%	34%	42%
Branco brasileiros	—	5%	10%	17%	24%
Europeus	30%	25%	22%	14%	17%
Índios integrados	50%	10%	8%	4%	2%

Fonte: Dados organizados por MUSSA, 1990 apud MATTOS; SILVA, 2004, p. 18.

Uma simples observação do quadro acima atesta que a influência africana na formação da população e conseqüentemente na formação da cultura brasileira não pode ter sido irrelevante conforme afirmaram alguns estudiosos da Dialectologia nacional notadamente entre as décadas de 1920-1950. Sílvia Elia (1979), a exemplo de Serafim da Silva Neto (1963), diminui a importância da influência africana no português do Brasil e avança a hipótese de criouliização<sup>10</sup>.

Percebe-se a expressiva ascendente miscigenação confirmada pelo percentual de mulatos na população nos anos finais do século XIX. Gladstone Chaves de Melo, em *A língua do Brasil*, publicada em 1946, faz referência ao expressivo número de mulatos na população e a influência africana no português falado no Brasil:

Ainda não se estudou convenientemente a importância e a influência do mulato na civilização brasileira (...)  
Quer-me parecer que a influência na língua até certo ponto correu paradas com a miscigenação racial e com a importância social do negro na formação histórica do Brasil, e que ao africano muito mais que ao índio se deve a tendência à simplificação das flexões e certas deturpações fonéticas (MELO, 1975, p. 75).

Ainda quanto à questão fonética vinculada ao contato da língua portuguesa falada no Brasil com as línguas de matriz africana, Melo defende que a influência africana embora mais profunda que a do tupi, foi menos extensa e teria atingido mais intensamente a fonética e a morfologia da língua, “E isso tem razões históricas. O negro viveu longamente em contato com o branco. Falou-lhe o idioma, como o diabo ajudou, naturalmente deturpando-o,

<sup>10</sup> Em linguística, criouliização corresponde ao processo de formação de crioulos (idioma que surge quando da adoção de uma língua estrangeira por uma comunidade, que é a mescla com o seu próprio idioma, criando um léxico e uma gramática distintos dos originais).

desfigurando-o, alterando-lhe fonemas, modificando-lhe articulações, simplificando-lhe a morfologia (...)” Melo (1975, p. 74). No entanto, o autor adverte que se deveria evitar exagerar a influência do negro no nosso “linguajar plebeu”, numa atitude bastante equivocada da mentalidade racista da sua época:

(...) Cumpre assinalar que muitos fatos pretensamente africanos são portugueses e cumpre sobretudo lembrar o acentuado instinto de imitação do negro do negro e do mulato, que, a par da ascensão social que lhes vem trazendo, lhes vai não raro determinando o nivelamento linguístico com os brancos de origem portuguesa (MELO, 1975, p. 75).

Melo tece críticas e contesta as suposições fonéticas, morfológicas, sintáticas levantadas por Mendonça em sua obra sobre a influência africana no Português do Brasil, entretanto, não faz nenhuma menção quanto a alguma marca linguística que se associasse ao surgimento do R caipira. Mas, ainda que Melo não faça maior menção ao R caipira, levanta a hipótese de que esse dialeto seria fruto de um crioulo que recebera influência não apenas indígena, mas também e, principalmente, africana. Assim, afirma o autor (1975, p. 77) que “Por razões de ordem histórica, sou levado a supor que se constituiu no planalto central paulistano um dialeto crioulo de tipo tupi-quimbundo, o qual, intensamente lusitanizado posteriormente, deu o *dialeto caipira*, que Amadeu Amaral tão bem estudou”.

## 5.2 Hipótese de natureza interna

Um dos principais defensores da teoria que se baseia no surgimento do R caipira como resultado de um processo interno é Brian Franklin. Head (1987, p. 13):

Baseiam-se em fatores externos as hipóteses referentes à origem do “R caipira” consideradas até agora no presente estudo. Por não ser convincente nenhuma das explicações examinadas anteriormente, torna-se necessário buscar informações que possam servir para fundamentar outra hipótese, mais plausível. Uma vez que as hipóteses refutadas se referem à fatores externos, podem ser especialmente relevantes dados de outra natureza – principalmente sobre características internas do português (HEAD, 1987, p. 13).

Head estrutura sua teoria de origem do R caipira como fruto de um processo interno, processo este de variação e mudança comum entre /r/ e /l/ que, em determinados contextos originara essa pronúncia típica do R caipira. Assim, no paradigma ou quadro geral dos

fonemas, /r/ e /l/ ocupam lugares muito próximos, compondo um pequeno conjunto ou subparadigma: o das líquidas (sonantes não nasais) apicais (ou “anteriores”), como também apresentam propriedades fonéticas em comum em termos sintagmáticos:

(...) Por um lado, /r/ e /l/ são os únicos fonemas que ocorrem como segundo membro de grupo de consoantes em posição inicial de sílaba; por outro, fazem parte de um pequeno conjunto de consoantes que são passíveis de ocorrência em posição final de sílaba interna e em posição final de palavra. Em formas de linguagem rurais e populares, /r/ e /l/ também acusam semelhanças entre si, quer por algumas das propriedades já mencionadas, quer por outras de natureza sincrônica ou diacrônica. Conforme a variedade, são os únicos fonemas a sofrerem, em determinados contextos, “vocalização” (no sentido de passarem a vogais assilábicas), a participarem (junto com /s/, ou não) de “queda” (eliminação) em final de palavra, e a manifestarem alternância entre si como segundo membro de grupo de consoantes e/ou em posição final de sílaba (HEAD, 1987, p. 15).

A perspectiva teórica de origem do R caipira por meio de um processo interno à própria língua se basearia em dois pilares básicos:

- (1) A relação entre /r/ e /l/ como membros exclusivos de uma mesma classe, pelas propriedades fônicas em comum ou pela participação em um conjunto de vários processos de alternância e condições de ocorrência;
- (2) a descrição detalhada das características fonéticas das diversas realizações do /r/ no dialeto caipira, segundo Head (1987, p. 15):

Abre-se uma perspectiva para explicar, sem recorrer a outros elementos, a origem da pronúncia típica do “R caipira”: em termos gerais, seria a participação de /r/, junto com /l/, de algum processo de alternância e evolução com esse resultado dentro da respectiva variedade do português do Brasil (HEAD, 1987, p. 15).

Cabe ressaltar que Head deixa claro sua posição terminantemente contrária às hipóteses sobre supostas causas externas e ratifica a importância dos dados sobre as estruturas e os processos internos da língua que sofre a mudança. A fim de trazer fielmente a posição do estudioso sobre o tema, citamos:

(...) Considero inaceitáveis as referidas hipóteses sobre supostas causas externas. Por outro lado, são sempre relevantes os dados sobre as estruturas e os processos internos da língua que acusa mudança. Por isso, considero precipitada toda e qualquer hipótese referente a condições externas (inclusive contato com outras línguas) antes que a mudança ou evolução em questão seja devidamente analisada à luz das propriedades e processos internos da respectiva língua. Aliás, a separação entre as condições linguísticas externas e as propriedades internas, sem considerar as relações entre ambos os contextos, na elaboração de hipóteses de origem (especialmente sobre características fonológicas ou gramaticais), aliada a uma



tendência a dar preferência a explicações em termos de fatores externos (linguísticos ou até extralinguísticos), sem fazer análise prévia e exaustiva dos processos e estruturas do contexto linguístico da mudança em questão – como se observa frequentemente em trabalhos sobre a língua portuguesa no Brasil – represente uma visão parcial e desatualizada (HEAD, 1987, p. 28).

Head salienta que, ao examinar casos de hipóteses que atribuem determinadas características do Português do Brasil à influência indígena ou africana, Melo (1971) refuta-os, baseado na análise de dados referentes a dialetos do Português Europeu ou em evoluções encontradas em outras línguas neolatinas ou no próprio latim.

Apoiado em Martinet (1955), Head considera que as possibilidades de influência de contatos com outras línguas são determinadas pelas estruturas e pelos processos já existentes na língua que vem a sofrer a evolução. Como exemplo, ele cita o caso da incorporação do fonema [ʒ] (fricativa palatal vozeada) pelo inglês pelo contato com o francês, durante um período de intenso contato entre as duas línguas (depois da invasão normanda); foi a natureza do sistema fonológico do próprio inglês que proporcionou as condições necessárias e propícias à “importação” dessa variante.

Há ainda uma lacuna na proposta teórica de Head sobre uma origem do R caipira a partir de um processo interno ao próprio idioma português: o R retroflexo é amplamente documentado no Brasil, mas não em Portugal e o próprio Head menciona não ter encontrado referências a variantes parecidas na linguagem popular de Portugal e a necessidade de mais pesquisas em torno do tema a fim de confirmar a viabilidade dessa perspectiva teórica:

O presente trabalho menciona uma possível relação entre o processo de evolução das líquidas, na língua portuguesa, que teria dado origem à pronúncia típica do “r” caipira (...). A noção de haver alguma tendência geral nesse sentido deverá ser examinada à luz de outros casos de evolução, quer na língua portuguesa, quer nas outras línguas românicas. (...) Quanto ao domínio geográfico da pronúncia típica do “R caipira”, haverá interesse em procurar determinar por que o “R retroflexo” e outras variantes afins teriam surgido no Brasil, mas não em Portugal (HEAD, 1987, p. 23).

Em resumo, Head propõe a hipótese na qual atribui uma possível origem do “R caipira” às propriedades comuns existentes entre o “R caipira” e a variante velarizada da lateral /l/ visto que, as variantes velarizadas (ou recuadas e rebaixadas) de /l/ e /r/ ocorrem tipicamente em posição final de sílaba.

Como se percebe, Head descarta *a priori* que a origem do R caipira seja fruto de influências de línguas em contato, assumindo uma posição convergente aos teóricos que consideram as variações encontradas no PB em relação ao PE como fruto de um processo de

deriva linguística como conceituado por Sapir (1949/1921), para o qual os processos de variação que uma língua sofre são controlados por um mecanismo da própria língua.

Ressaltamos que, tanto as hipóteses estruturalistas quanto as sociolinguísticas relativas à origem do R caipira são controversas assim como o são as causas que deram origem ao Português do Brasil, cujas variações com relação ao Português Europeu ainda se encontram sob análise: a variante brasileira ora vista como fruto de um processo de deriva linguística, ora como fruto de uma criouliização seguida de uma descriouliização, conforme veremos no capítulo seguinte.

## 6. CRIOLIZAÇÃO E DERIVA

Outra questão que permeia os estudos do Português do Brasil é a possibilidade de ele ser produto de uma língua crioula. A hipótese de que um processo de crioulização seria a explicação para a existência das diferenças encontradas entre o PE e o PB tem sido levantada há bom tempo, com um defensor radical, na pessoa do filólogo português Adolfo Coelho, na segunda metade do século XIX.

Estudiosos do Brasil e do exterior dividem-se entre dois extremos: a hipótese de que o PB seria uma nova língua, produto de um processo de crioulização e a hipótese de que o PB seria apenas uma variação do PE e que essas variações ocorreriam simplesmente por um processo natural de deriva.

Ao adentrarmos a discussão, primeiramente, faz-se necessária a delimitação de alguns conceitos essenciais para a compreensão dessa dinâmica. São eles os conceitos de *pidgin*, crioulo, língua de superstrato, língua de substrato, transmissão linguística irregular e deriva.

Um *pidgin* é uma língua simplificada, composta de elementos de duas ou mais línguas e utilizada como forma de comunicação entre comunidades linguísticas diferentes, servindo apenas como segunda língua para fins limitados e específicos, como o contato comercial. Assim, segundo a abordagem tradicional em estudos crioulos, um *pidgin* de base portuguesa tem a maioria do seu vocabulário derivado do português e a estrutura gramatical pode ser independente do vocabulário.

*Pidgin* e crioulo são conceitos conexos: para a maioria dos estudiosos, todo crioulo já foi um *pidgin*. O início de seu ciclo é um *pidgin*, uma língua que surge pela necessidade de transações comerciais. Visto que seu uso se restringe a uma esfera específica da sociedade, sua estrutura é bem simples e rudimentar. Esta língua reduzida, que surge do contato de dois grupos que não têm uma língua comum, começa a tomar forma, sendo o léxico provido pela língua de maior prestígio, o superstrato, e a gramática provida pela(s) língua(s) de menor prestígio, chamada(s) substrato.

Com a expansão marítima europeia no século XV, as línguas de maior prestígio, conforme atesta a história, são as línguas europeias dos colonizadores, enquanto as de menor prestígio, por outro lado, são as dos grupos autóctones sob dominação. Com o tempo, os *pidgins* passam a fazer parte de várias esferas da sociedade e, como decorrência, as crianças são expostas ao *pidgin* e o adquirem como idioma materno. Tal processo tem sido chamado

nativização ou criouliização (Holm, 2000, p.7). E é nessa condição que o *pidgin* passa a ser um crioulo.

A diferença básica entre um *pidgin* e um crioulo é o fato de os crioulos terem falantes nativos, e *pidgins*, não. Se um crioulo permanece em contato com sua língua lexificadora (o superestrato), a tendência é que passe pelo processo de descriouliização. A descriouliização, então, é o processo de perda das “características crioulas” e a aproximação a características da língua lexificadora. Para Naro e Scherre (2007, p. 26):

Um *pidgin* ou crioulo de “base portuguesa” tem a maioria do seu vocabulário derivado do português. A estrutura gramatical pode ser independente do vocabulário. De fato, muitos estudiosos são da opinião de que os crioulos de bases europeias (o francês do Haiti, o português de Cabo Verde etc.) possuem uma gramática africana. Quando estamos dizendo que um certo *pidgin* ou crioulo é de “base x”, estamos falando apenas do léxico e não da gramática. Um *pidgin* ou crioulo de base portuguesa terá necessariamente palavras portuguesas, mas suas estruturas gramaticais poderão ser portuguesas, africanas, ameríndias, ou de qualquer outra origem. Esse uso se deve ao fato de que é muito mais fácil identificar as origens do vocabulário do que as fontes da gramática (NARO; SCHERRE, 2007, p. 26).

Se o acesso dos falantes das outras línguas (e, principalmente, dos seus descendentes) aos modelos da língua alvo aumenta com a continuidade da situação social que originou o contato, e considerando-se que esses modelos tendem a gozar de um maior prestígio na estrutura da comunidade de fala, pode-se pensar que tais modelos tendem a suplantar os processos de transferência de estruturas das outras línguas e/ou de reestruturação original da gramática durante a nativização/criouliização. O resultado desse processo, então, pode não ser a formação de um sistema linguístico distinto da língua alvo (um *pidgin* ou um crioulo), mas uma nova variedade dessa língua alvo que não deixa de apresentar certas características decorrentes do processo de transmissão linguística irregular que se deu com a socialização/nativização da língua alvo entre os segmentos de falantes das outras línguas e seus descendentes.

Segundo Naro e Scherre (2007, p. 137), a “transmissão linguística irregular” é aquela que se dá “entre adultos e/ou com base em fala não suscetível de uma análise ordenada, talvez por ser caótica, ou por ser em quantidade insuficiente, ou ainda por outras razões”.

Para Sapir (1949/1921), os processos de variação que uma língua sofre são controlados por um mecanismo da própria língua, o que ele chama de deriva linguística, que consiste no fato de a língua mover-se através do tempo, em um curso que é próprio dela e que vai gerar determinadas transformações sintáticas e morfológicas.

Tendo em mente as conceituações expostas anteriormente, passaremos ao estudo das diferentes hipóteses assumidas com relação ao processo de variação do PB com relação ao PE.

Segundo os adeptos dessa corrente teórica, todas as diferenças entre o PB e o PE podem ser explicadas por características estruturais da própria língua portuguesa e não por influência externa da estrutura de outras línguas.

Naro & Scherre (2007), com base em pesquisas que tinham por foco a análises dialetais do PE, defendem a tese de que a língua portuguesa falada em Portugal antes da colonização do Brasil, se tomada em sua diversidade dialetal e não na sua forma padrão, “continha as sementes da variação que mais tarde germinaram e produziram uma vasta floresta no Brasil” e rejeitam veemente as hipóteses de pidginização/crioulização.

Assim, diferenças observadas entre PB e PE seriam simplesmente continuações de mudanças já iniciadas na fase de diferenciação do próprio latim vulgar que deu origem ao português. Simplesmente o português carrega em si diversas possibilidades de mudanças estruturais potenciais; a diferença é que algumas delas se desenvolveram no Brasil e outras não, e se daria o mesmo para Portugal.

Diante dessa premissa:

(...) Teorias que se situam no grupo internalista têm de responder ao motivo populacional, afinal é inegável que, na formação do português brasileiro, a maioria dos participantes eram aloglotas (pessoas que não tinham o português como língua materna), e se não houve um crioulo de base, qual a importância ou papel dos aloglotas na formação do PB? (BASSO; GONÇALVES, 2014, p. 288-289).

E mais, se tanto PB quanto PE estão destinados a sofrer essa “mudança natural” em um *continuum* de tempo, por que as mudanças que ocorreram foram diferentes no Brasil e em Portugal?

Por sua vez, os adeptos das teorias externalistas, que consideram a possibilidade de origem crioula, argumentam que o convívio entre uma minoria de falantes de português com populações que falavam outras línguas, principalmente os africanos, no Brasil fez emergir uma ou mais línguas crioulas de base portuguesa. Devido à influência cada vez maior da Língua Portuguesa, esses crioulos foram se descrioulizando e se aproximando cada vez mais do Português Europeu. Entretanto, uma descrioulização completa não é possível. Assim, essa reaproximação do PB ao PE não foi total, mantendo-se as peculiaridades de cada um deles.

Interessante ressaltar a teoria de Guy (1981), na qual o autor se apoia em argumentos externos (sociais) e internos (estrutura da língua). Pelo fato de a sociedade brasileira possuir um alto grau de mestiçagem desde os primeiros tempos da colonização, o processo de descrioulização foi mais rápido e o ciclo pidgin – crioulo- descrioulização teria acontecido diversas vezes e de forma acelerada.

Tal teoria foi duramente criticada, principalmente por Tarallo (2006), conforme citado por Basso & Gonçalves (2014, p. 293) “Segundo Tarallo, a ideia de que o PB estaria se descrioulizando é problemática porque com o passar do tempo a distância entre as estruturas do PB e do PE só aumenta.” e, ainda “sem essa aproximação entre o PB e o PE, a ideia de descrioulização fica bastante enfraquecida” e, por conseguinte, a própria ideia de criouliização.

Tanto a ideia de criouliização como de uma semicriouliização, esta inclusive formulada em outros termos há várias décadas por Silva Neto (1950, p.48), perderam força a partir das críticas de Tarallo.

Noll (2008) rejeita tanto a hipótese de criouliização prévia quanto a da existência de um semicrioulo no Brasil. O principal ponto de apoio de sua teoria se assenta no fato de que a população formada por brasileiros nativos foi a que mais cresceu ao longo da história e, como outros pesquisadores, redireciona o foco das influências das populações de origem portuguesa, indígena e africana para as influências da população de brasileiros nativos.

Conforme se observa na tabela apresentada por BASSO & GONÇALVES, o brasileiro nativo desempenhou um papel importante na formação da nossa língua pois, de 1600 em diante, a população de nativos só fez crescer; alcançando o percentual de 50% do total no século XVIII e 79% em fins do século XIX.

Quadro 8 - Evolução da população de brasileiros nativos entre os séculos XVI e XIX

População	1538-1600	1601-1700	1701-1800	1801-1850	1851-1890
Africanos	20%	30%	20%	12%	2%
<b>Brasileiros nativos</b>	<b>—</b>	<b>35%</b>	<b>50%</b>	<b>70%</b>	<b>79%</b>
Europeus	30%	25%	22%	14%	17%
Índios integrados	50%	10%	8%	4%	2%

Fonte: Adaptado de BASSO; GONÇALVES, 2014.

Assim, com relação a esse aspecto da história sociodemográfica da população brasileira, Noll (2008, p. 190-191) argumenta:

O fator de integração dos negros crioulos nascidos no país com a língua materna portuguesa diferencia o Brasil, de maneira essencial, das zonas de fala genuinamente crioula. Já Nina Rodrigues afirmava: “o negro *novo* era obrigado a aprender o português para falar com os senhores brancos, com os mestiços e os negros crioulos” (1977: 123). A pequena consideração que se dá ao papel dos negros crioulos conduz a um erro fundamental na interpretação das estatísticas da população pela Crioulística. Uma fórmula associada à população do Brasil segundo o princípio *branco* [= elemento português] vs. *Negro* [= fator potencial de criouliização] não é praticável do ponto de vista linguístico. Na estatística colonial da população, os negros crioulos deveriam ser, de fato, computados, com respeito ao seu significado para o desenvolvimento linguístico, juntamente com o contingente branco (português) e separado do africano (NOLL, 2008, p. 190-191).

Noll se apoia na dinâmica populacional acima apresentada e no fato de que estudos recentes sobre línguas crioulas consideram que, para a formação de uma língua crioula baseada em uma língua A num dado país, seja necessário que não mais que 20% da população tenha a língua A como materna.

Voltando às variações sobretudo fonológicas e morfológicas encontradas entre o PB e o PE, tanto Gladstone Chaves de Mello (1975), Noll (2008) e Naro & Scherre (2007) argumentam que são exemplos de deriva linguística com ocorrências similares em dialetos portugueses ou no português arcaico.

Mais especificamente, com relação ao objeto do nosso estudo, os fenômenos relacionados à realização do R como a neutralização de /r/ e /l/ (como na pronúncia caipira de “sol” – “sor”), os autores citados colocam o fenômeno no rol daqueles atestados em diferentes períodos na história da Língua Portuguesa, enfatizando a inutilidade de recorrer a influências externas para explicá-los.

## CONSIDERAÇÕES FINAIS

Como dito anteriormente, o /r/ retroflexo, nosso R caipira, aparece como a principal marca distintiva do chamado dialeto caipira, especialmente a modalidade encontrada no interior de São Paulo, sudoeste de Minas Gerais e norte do Paraná, além de partes do Mato Grosso e Goiás. Alguns estudiosos chegam a insinuar que o R caipira é invenção brasileira, além do principal traço distintivo entre o Português do Brasil e o Português Europeu.

Cabe ressaltar que os róticos estão presentes em aproximadamente 75% das línguas do mundo e destas, 18% possuem mais de um som de /r/, como é o caso do Português do Brasil (cf. Maddieson, 1984, *apud* Lindau, 1985).

Até o momento, a origem do R retroflexo ou caipira permanece desconhecida embora haja várias suposições e hipóteses ainda não totalmente confirmadas; a maioria aponta a herança indígena como possível fonte desse fenômeno linguístico, entretanto, autores divergem quanto aos ramos linguísticos de que o R caipira poderia ter-se derivado: se do tronco Tupi-Guarani ou do Macro-Jê.

O Tupi-Guarani foi o substrato da Língua Geral falada na colônia desde os primeiros tempos da colonização até meados do século XVIII e essa era a língua franca utilizada tanto pelos indígenas de diferentes etnias quanto pelos colonos portugueses e pelos exploradores paulistas que lideravam as bandeiras que adentraram os rincões do interior do centro sul brasileiro.

Embora boa parte do vocabulário popular do Português do Brasil tenha sofrido grande influência do Tupi-Guarani, o substrato da Língua geral dominante, estudos recentes convergem para a hipótese de que, no caso em especial do R retroflexo, a origem de tal fenômeno teria sido a herança fonética dos dialetos pertencentes ao tronco Macro-Jê falados pelos índios dessa etnia que habitavam as regiões não litorâneas onde predominavam os indígenas de língua tupi. Tal hipótese se assenta no fato de que, segundo alguns indigenistas e filólogos, o R retroflexo gutural se apresenta como um som característico de dialetos Macro-Jê como observado nas línguas faladas pelas etnias Oti, Kaingang, Ofalê, dentre outras.

Dado que o próprio tronco Macro-Jê ainda não se encontra totalmente estruturado e é fonte de uma reconstrução artificial, tal possibilidade não deve ser descartada, mas cuidados adicionais devem ser considerados antes de se eleger essa vertente em detrimento daquela que advoga tal fenômeno como fruto da nossa herança Tupi.



Os estudos descritivos das diversas modalidades linguísticas existentes entre as línguas indígenas brasileiras são essenciais para subsidiar não somente o estudo da origem do R caipira como também de diversas outras peculiaridades que se apresentam no Português do Brasil. Acreditamos que a valorização da nossa herança indígena através de políticas afirmativas e estudos correlatos podem contribuir bastante para o desenvolvimento das pesquisas sociolinguísticas e históricas que tratam da formação, evolução e diferenciação do português falado no Brasil.

Teóricos pertencentes a uma outra vertente, também externalista, traçam uma provável origem do R retroflexo como resultado do contato com as línguas africanas faladas pelos escravos que aqui chegavam. Essa hipótese foi levantada por Mendonça (1948) que atribuiu à influência de línguas africanas diversas propriedades do dialeto caipira.

Apesar de existirem estudos que apresentam as influências das línguas de matrizes africanas na formação do Português do Brasil e do dialeto caipira (AMARAL, 1920), essa influência parece se concentrar especialmente em aspectos relacionados ao léxico e à sintaxe ou até mesmo a um processo de crioulização do português.

Como dissemos anteriormente com relação ao estudo das línguas indígenas, o avanço dos estudos relativos às influências africanas na formação do nosso português também é essencial para subsidiar os estudos concernentes a origem do R caipira e a estruturação de tais estudos passa necessariamente por políticas e iniciativas que fomentem a valorização de nossa negligenciada herança africana.

Desde meados da década de 80 do século passado, estudiosos estruturalistas como Head, apontam para uma origem interna dos fenômenos de variação linguística, sobretudo, os fonéticos/fonológicos e, para essa vertente, o R caipira seria resultado de um processo interno inerente à própria língua portuguesa, ou seja, uma deriva. Estudos adicionais que tratem do complexo fenômeno da deriva linguística serão muito bem-vindos e, com certeza, irão auxiliar na busca por respostas relativas ao nosso objeto de estudo.

Crioulização? Deriva? Ou um mix disso tudo? Como fica o nosso enigma relativo à origem do R caipira?

Somos fruto de um caldeirão cultural. Como boa parte das línguas indígenas brasileiras ainda foram pouco estudadas em profundidade e muitas etnias foram extintas ou desaparecem sem haver registros dos respectivos dialetos, a maioria das hipóteses carece de comprovação e a origem do R caipira permanece ainda um enigma a ser desvendado pelas pesquisas e estudos futuros.

Com relação à influência de matriz africana, estudos comparativos em países africanos de colonização portuguesa podem representar um valioso subsídio para auxiliar a explicar variações típicas do PE em solo estrangeiro.

Estudos referentes à teoria internalista, mais especificamente aqueles que advogam pela deriva linguística, também apresentam um futuro promissor dado o avanço tecnológico a que temos assistido nas áreas de estudos fonético/fonológicos e correlatos.

Até o momento, o que podemos afirmar é que, contrariando a previsão de muitos dialetólogos e filólogos do passado, como o próprio Amadeu Amaral, o R caipira assim como o dialeto ao qual pertence não desapareceu; muito pelo contrário, o dialeto caipira e sua principal marca, o R retroflexo ou caipira, estão “firmes e fortes” fazendo parte do falar de milhões de brasileiros nos diversos rincões do Brasil.

Assim, há muito trabalho a se fazer para desvendarmos e entendermos as peculiaridades presentes no PB, suas origens e fatores de influência. Tanto a corrente de viés interno e estruturalista carece do subsídio de novas pesquisas e *inputs* como a corrente de viés externo de estudos adicionais sobre a historiografia do PB, dos dialetos presentes em solo português europeu, das línguas indígenas nativas e das línguas africanas que aqui chegaram; pois o processo de investigação da história de uma língua e de sua formação é extremamente complexo, dinâmica e se apresenta em um eterno devir.

## REFERÊNCIAS

ALKMIN, T. M. Sociolinguística. *In*: MUSSALIN, F.; BENTES, A. C. (org.). **Introdução à linguística**. São Paulo: Cortez, 2004.

ALMEIDA, M. M. S. **Aspectos fonológicos do português falado na baixada cuiabana: traços de língua antiga preservados no Brasil**. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2001.

ALMEIDA, R. H. **O Diretório dos índios: um projeto de “civilização” no Brasil do século XVIII**. Brasília: Ed. Universidade de Brasília, 1997. (Com fac-símile do Diretório dos índios em apêndice).

ALTMAN, M. C. A descrição das línguas ‘exóticas’ e a tarefa de escrever a história da linguística. **Revista da ABRALIN**, v. 10, n. 3, 31 dez. 2011.

AMARAL, A. **O dialeto caipira**. 3. ed. São Paulo: Hucitec; SCET; CEC, 1976 [1920].

ARAGÃO, M. S. S. Africanismos no português do Brasil. **Revista de Letras**, v. 30, n. 1/4, p. 7-16, jan. 2010/dez. 2011. Disponível em: [http://www.revistadeletras.ufc.br/Revista%20de%20Letras%20Vol.30%20-%201.4%20-%20jan.%202012%20.%20dez.%202011/rl30art01\\_Africanismos\\_no\\_portugues\\_d\\_o\\_Brasil.pdf](http://www.revistadeletras.ufc.br/Revista%20de%20Letras%20Vol.30%20-%201.4%20-%20jan.%202012%20.%20dez.%202011/rl30art01_Africanismos_no_portugues_d_o_Brasil.pdf). Acesso em: 12 fev. 2021.

BARBOSA, Pe. A. L. **Curso de Tupi Antigo: gramática, exercícios, textos**. Rio de Janeiro: Livraria São José, 1956.

BARTHES, R. **O rumor da língua**. São Paulo: Brasiliense, 1988.

BASSO, R. M.; GONÇALVES, R. T. **História concisa da língua portuguesa**. Petrópolis: Vozes, 2014.

BENVENISTE, E. **Problemas de linguística geral I**. Campinas, SP: Pontes, 2005.

BORBA, T. **Actualidade indígena no Paraná**. Curitiba: Imprensa Paranaense, 1908.

BORTONI-RICARDO, S. M. **Educação em língua materna: a sociolinguística na sala de aula**. São Paulo: Parábola Editorial.2004.

BOSWOOD, J. 1973. **Evidências para a inclusão do Aripaktsá no filo Macro-Jê**. Brasília: Summer Institute of Linguistics, [20--]. (Série Linguística, 1).

BRANDÃO, S. F. **A geografia linguística no Brasil**. São Paulo: Ática, 1991.

BRANDÃO, S. F. Nas trilhas do –R retroflexo. **Signum**: estudos da linguagem, Londrina, v. 10, n. 2, p. 265-283, 2007.

CAGLIARI, L. C. **Elementos de fonética do português brasileiro**. Tese (Livre Docente) - Universidade Estadual de Campinas, Campinas, SP, 1981.

CALLOU, D.; MORAES, J.; LEITE, Y. A realização das consoantes pós vocálicas no português do Brasil. *In*: CASTILHO, A. T. (org.) **Gramática do português falado**. Campinas: Editora da Unicamp, 1998.

CALLOU, D.; MORAES, J. A.; LEITE, Y. Variação e diferenciação dialetal: a pronúncia do /r/ no português do Brasil. *In*: KOCH, I. G. V. (org.). **Gramática do português falado**. v. VI: Desenvolvimentos. Campinas: Ed. da UNICAMP/FAPESP, 1996.

CANDIDO, A. Possíveis raízes indígenas de uma dança popular (1956); “Caipiradas” (1980). **Revista USP**, Dossiê 100 Anos de Antonio Candido. São Paulo, n. 118, p. 139–172, jul./set. 2018.

CARDIM, F. Principio e origem dos indios do Brazil e de seus costumes, adoração e ceremonias. **Revista do Instituto Histórico e Geográfico Brasileiro**, Rio de Janeiro, t. 57, pt. 1, p. 185-212, 1894.

CARMO, L. A voz do caipira em Amadeu Amaral. *In*: CARMO, Laura do; LIMA, Ivana Stolze (org.). **História social da língua nacional**. Rio de Janeiro: Casa de Rui Barbosa, 2008. v. 1.

CASTRO, V. **A resistência de traços do dialeto caipira**: estudo com base em Atlas linguísticos regionais brasileiros. Tese (Doutorado) - Curso de Linguística, Instituto de Estudos da Linguagem, UNICAMP, Campinas, SP, 2006.

CASTRO, V. S. O 'r-caipira' no Estado de São Paulo: estudo com base em dados do Atlas Linguístico do Brasil (ALIB). *In*: GRUPO DE ESTUDOS LINGUÍSTICOS DO ESTADO DE SÃO PAULO – GEL, 57., 2009, São Paulo. **Anais...** São Paulo: GEL, 2009.

CESCHIN, O. H. L. A respeito de Amadeu Amaral e d'ó dialeto caipira. **Língua e Literatura**, São Paulo, n. 25, p. 41-80, 1999.

COSERIU, E. **O homem e sua linguagem**: estudos de teoria e metodologia linguística. Trad. Carlos Alberto da Fonseca e Mário Ferreira. Rio de Janeiro: Presença; São Paulo: Universidade de São Paulo, 1982. cap. 4, p. 79-116.

CRISTÓFARO SILVA, T. **Fonética e fonologia do português**: roteiro de estudos e guia de exercícios. 9. ed. São Paulo: Editora Contexto, 2007.

CRISTÓFARO SILVA, T. *et al.* **Fonética acústica**: os sons do português brasileiro. 1. ed. São Paulo: Editora Contexto, 2019. v. 1, 272 p.

D'ANGELIS, W. **A língua Kaingang**. Disponível em: [http://www.portalkaingang.org/lgua\\_kaingang.pdf](http://www.portalkaingang.org/lgua_kaingang.pdf). Acesso em: 16 fev. 2021

DICIONÁRIO Michaelis. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/caipira>. Acesso em: 28 maio 2021.

DUARTE, F. B. Diversidade linguística no Brasil: a situação das línguas ameríndias. **Caletroscópio**, Belo Horizonte, v. 4, n. esp., 2016.

DUBOIS, J. *et al.* **Dicionário de linguística**. São Paulo: Cultrix, 1978.

ELIA, S. **A unidade linguística do Brasil**: condicionamentos geoeconômicos. Rio de Janeiro: Padrão, 1979.

FARACO, C. A. **História sociopolítica da língua portuguesa**. São Paulo: Parábola Editorial, 2016.

FERREIRA, M. B. *et al.* Variação Linguística: perspectiva dialectológica. *In*: FARIA, I. H. *et al.* (org.). **Introdução à linguística geral e portuguesa**. 2. ed. Lisboa: Caminho, 1996. p. 477-502.

GÂNDAVO, P. M. **Regras que ensinam a maneira de escrever e orthographia da lingua Portuguesa, com hum Dialogo que a diante se segue em defensam da mesma lingua**. Lisboa: Antônio Gonçalves, 1574.

GÂNDAVO, P. M. **Historia da prouincia sa[n]cta Cruz a qui vulgarme[n]te chamamos Brasil**. Lisboa: Antônio Gonçalves, 1576.

GOMES, L. **Escravidão**: do primeiro leilão de cativos em Portugal até a morte de Zumbi dos Palmares. Rio de Janeiro: Globo Livros, 2019. v. 1.

HAUY, A. B. Origem e formação da língua portuguesa. *In*: SPINA, Segismundo (org.). **História da língua portuguesa**. São Paulo: Ateliê Editorial, 2008. p. 22-33.

HEAD, B. F. Propriedades fonéticas e generalidade de processos fonológicos: o caso do R caipira. **Cadernos de Estudos Linguísticos**, Campinas, SP, v. 13, p. 5–39, 2012. DOI: 10.20396/cel.v13i0.8636786. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/cel/article/view/8636786>. Acesso em: 7 fev. 2021.

HENRIQUES, C. C. **Geo-História do português**: estudos sobre a história e a geografia do português na perspectiva brasileira. Rio de Janeiro: Gramma, 2019.

HOLM, J. **An introduction to pidgins and creoles**. Cambridge: Cambridge University Press, 2000.

ILARI, R.; BASSO, R. **O português da gente**: a língua que estudamos a língua que falamos. São Paulo: Contexto, 2006.

LEHISTE, I. **Acoustical characteristics of selected English consonants**. The Hague: Mouton, 1962. p. 51-115.

LEITE, F. R. **A língua geral paulista e o “Vocabulário elementar da língua brasílica”**. Dissertação (Mestrado) - UNICAMP, Campinas, SP, 2013.

LINDAU, M. The story of /r/, in Victoria Fromkin (org.). *In*: FROMKIN, Victoria. **Phonetic Linguistics**: essays in honor of Peter Ladefoged. New York: Academic Press, 1985. p. 157-168.

MARTINET, A. **Économie des changements phonétiques**. Berne: A. Francke, 1955.

MELO, G. C. **A língua do Brasil**. 3. ed. Rio de Janeiro: Fundação Getúlio Vargas, 1975.

MENDONÇA, R. **A influência africana no português do Brasil**. Brasília: FUNAG, 2012.

MÉTRAUX, A. The Tupinamba. *In*: STWEARD, J. (org.). **Handbook of south american indians**. Washington: Government printing office, 1984. p. 95-139.

MOREIRA, F. A. **O dialeto caipira de Amadeu Amaral e suas reminiscências na linguagem de regiões rurais da Zona da Mata mineira**. Tese (Doutorado) – Univesidade Federal Fluminense, Niterói, 2019.

NARO, A. J.; SCHERRE, M. M. P. **Origem do português brasileiro**. São Paulo: Parábola, 2007.

NEVES, M. H. M. **O Português são dois...** Novas fronteiras, velhos problemas. São Paulo: Parábola, 2004.

NIMUENDAJÚ, C. **Textos indigenistas**. São Paulo: Loyola, 1982.

NOLL, V. **O português brasileiro: formação e contrastes**. São Paulo: Globo, 2008.

NOLL, V; DIETRICH, W. (org.) **O português e o tupi no Brasil**. São Paulo: Contexto, 2010. p. 183-209.

RADTKE, E.; THUN, H. **Neue Wege der romanischen Geolinguistik: Akten des Symposiums zur empirischen Dialektologie**. Heidelberg/Mainz 21.-24.10. 1991, Kiel, Westensee, 1996.

RECTOR, M. **A linguagem da juventude**. Petrópolis: Vozes, 1975.

RIBEIRO, D. **Os índios e a civilização**. São Paulo: Companhia das Letras, 2004.

RIBEIRO, D. O Brasil caipira. *In*: RIBEIRO, D. **O povo brasileiro: a formação e o sentido do Brasil**. São Paulo: Companhia das Letras, 1995.

RODRIGUES, A. N. **O dialeto caipira na região de Piracicaba**. São Paulo: Ática, 1974

RODRIGUES, A. D. **Línguas brasileiras – para o conhecimento das línguas indígenas**. São Paulo: Edições Loyola, 1986.

RODRIGUES, A. D. Tupí, tupinambá, línguas gerais e português do Brasil. *In*: NOLL, V.; DIETRICH, W. (org.). *O português e o tupi no Brasil*. São Paulo: Contexto, 2013. p. 27-47. **Revista Brasileira de Linguística Antropológica**, v. 2, n. 1, p. 167-168.

ROMANO, V. P.; AGUILERA, V. A. A distribuição diatópica das variantes para tangerina: um estudo geo-sociolinguístico. *In*: ISQUERDO, A. N.; ALTINO, F. C.; AGUILERA, V.A. (org.). **Atlas Linguístico do Brasil**: descrevendo a língua, formando jovens pesquisadores. Londrina: Ed. da UEL, 2009. p. 148-157.

SAPIR, E. **Language** –an introduction to the study of speech. New York: Harcourt, Brace & World, [1949] [1921].

SILVA, A. H. P. **As fronteiras entre fonética e fonologia e a alofonia dos róticos iniciais em PB**: dados de dois informantes do sul do país. Tese (Doutorado) - UNICAMP, Campinas, SP, 2002.

SILVA NETO, S. **Introdução ao estudo da língua portuguesa no Brasil**. Rio de Janeiro: Presença, 1950.

SILVA NETO, S. Um traço de pronúncia caipira. *In*: SILVA NETO, S. **Língua, cultura e civilização**; estudos de filologia portuguesa. Rio de Janeiro: Acadêmica, 1960.

SOUZA, A. C. M. **Os parceiros do Rio Bonito**: estudo sobre o caipira paulista e a transformação dos seus meios de vida. [S.l.: s.n.], 2017.

TEYSSIER, P. **História da língua portuguesa**. Tradução de Celso Cunha. 2. ed. São Paulo: Martins, 2014.

TOLEDO NETO, S. A.; ALMEIDA, M. M. S. Variedade do Português do Brasil na trilha das bandeiras paulistas. *In*: NOLL, V.; DIETRICH, W. (org.). **O português e o tupi no Brasil**. São Paulo: Contexto, 2016.

WIESEMANN, U. G. **Kaingang-Português Dicionário Bilingue**. Curitiba: Editora Evangélica Esperança, 2002.

ZILIO-PASSERINI, T. Tradição e inovação em “O Dialeto Caipira”, de Amadeu Amaral. **Verbum**, São Paulo, v. 9, n. 2, set. 2020. Disponível em: <https://revistas.pucsp.br/index.php/verbum/article/view/50405/pdf>. Acesso em: 01 fev. 2021.