UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS
DEPARTAMENTO DE LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM SEMIÓTICA E LINGUÍSTICA GERAL

GUSTAVO MICAEL GOMES MARTINS

Aqui se está fazendo muy grande nação
Estar + gerúndio *versus* estar + infinitivo preposicionado no português brasileiro
oitocentista

Versão Corrigida

São Paulo, SP

2019

GUSTAVO MICAEL GOMES MARTINS

AQUI SE ESTÁ FAZENDO MUY GRANDE NAÇÃO

Estar + gerúndio *versus* estar + infinitivo preposicionado no português brasileiro oitocentista

Versão Corrigida

Dissertação apresentada como requisito final à obtenção de grau de Mestre em Semiótica e Linguística Geral, no curso de Pós-Graduação em Semiótica e Linguística Geral, na Universidade de São Paulo.

Orientador: Thomas Daniel Finbow

SÃO PAULO

2019

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação
Serviço de Biblioteca e Documentação
Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo

M379a Martins, Gustavo Micael Gomes
Aqui se está fazendo muy grande nação: estar + gerúndio versus estar + infinitivo preposicionado no português brasileiro oitocentista / Gustavo Micael Gomes Martins ; orientador Thomas Daniel Finbow. - São Paulo, 2019.
141 f.

Dissertação (Mestrado)- Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. Departamento de Linguística. Área de concentração: Semiótica e Linguística Geral.

1. Gerúndio. 2. Infinitivo preposicionado. 3. Português brasileiro. 4. Brasil Império. I. Finbow, Thomas Daniel, orient. II. Título.

fflch

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, LETRAS E CIÊNCIAS HUMANAS

ENTREGA DO EXEMPLAR CORRIGIDO DA DISSERTAÇÃO/TESE

Termo de Ciência e Concordância do (a) orientador (a)

Nome do (a) aluno (a): GUSTAVO MICHAEL GOMES MARTINS

Data da defesa: 10 / 06 / 2019

Nome do Prof. (a) orientador (a): THOMAS DANIEL FINBOW

Nos termos da legislação vigente, declaro **ESTAR CIENTE** do conteúdo deste **EXEMPLAR CORRIGIDO** elaborado em atenção às sugestões dos membros da comissão Julgadora na sessão de defesa do trabalho, manifestando-me **plenamente favorável** ao seu encaminhamento e publicação no **Portal Digital de Teses da USP**.

São Paulo, 8 / 08 / 2019

(Assinatura do (a) orientador (a))

AGRADECIMENTOS

Pobres criaturas, tão frágeis e temerosas ante um mundo de feras, são os humanos, sem garras ou presas para matar, sem placas ou couraças que as possam proteger. Mas dos deuses roubaram o fogo e a palavra e assim definiram seu caminho, aberto não por ossos e sangue, mas pelo verbo, a razão e a cooperação. *Ecce homo*, tão fraco por si, tão poderoso em união. Assim, de mãos dadas e corações unidos, seguem em direção ao seu lugar no mar de estrelas.

Aqui, porém, o objetivo era um tanto menos sublime. Mas ainda que o fim seja menor, o mecanismo central não se altera. Se essas páginas hoje chegam a você é pela bondade e generosidade de muitos.

Daqueles que ouviram os lamentos.

Daqueles que apontaram as pedras no caminho.

Daqueles que, sabendo ou não, foram a luz.

A todos esses, ofereço toda a gratidão de um ser humano consciente de seus limites.

体は剣で出来ている。

血潮は鉄で心は硝子。

幾たびの戦場を越えて不敗。

ただの一度の敗走もなく、

ただの一度の勝利もなし。

担い手はここに孤り、剣の丘で鉄を鍛つ。

ならば、わが生涯に意味は不要ず。

この体は、無限の剣で出来ていた。

Fate/Stay Night, Kinoko Nasu

MARTINS, Gustavo Micael Gomes. Aqui se está fazendo muy grande nação: gerúndio *versus* infinitivo preposicionado no português brasileiro oitocentista. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo; 2019.

RESUMO

Introdução: Essa pesquisa averiguou a utilização de estruturas *estar + a + infinitivo* e *estar + gerúndio* no português brasileiro de cinco regiões e suas relações com a construção da identidade linguística brasileira. As estruturas acima descritas são tomadas como variantes altamente representativas de seus respectivos dialetos do português, o europeu e o brasileiro.

Objetivo: Analisar o uso das referidas formas no português brasileiro oitocentista em relação ao discurso linguístico então vigente. **Método:** Se iniciou o trabalho pela observação do discurso linguístico do período através da leitura de textos sobre história e língua portuguesa publicados no período imperial. Tendo sido essa parte feita, dados foram extraídos de romances de quinze autores nascidos no período imperial, três de cada uma das cinco regiões de interesse (Bahia, Minas Gerais, Pernambuco, Rio de Janeiro e São Paulo). Uma vez extraídos, os dados foram classificados de acordo com critérios linguísticos e sociais e analisados estatisticamente.

Resultados: O uso das variantes se mostrou altamente diverso, com taxas de uso significativamente diferentes sendo apresentadas por autores da mesma região ou afiliação política. Aspectos geográficos, por outro lado, apresentaram uma correlação significativa com o uso da variável, ocorrendo um maior emprego de *estar + a + infinitivo* no sudeste, e particularmente no Rio de Janeiro, do que no nordeste. Fatores linguísticos como a presença de interpolação e a presença de certos tipos de palavras nos arredores do termo também apresentaram uma correlação significativa com o uso da variável. **Conclusões:** Em termos gerais, o discurso e a prática linguística parecem diferir, sendo o uso da variável mais fortemente ligado a aspectos regionais e linguísticos.

PALAVRAS-CHAVE: Gerúndio. Infinitivo preposicionado. Português brasileiro. Brasil Império. Sociolinguística. Mudança linguística.

MARTINS, Gustavo Micael Gomes. Here a mighty nation is in the making: gerund *versus* prepositioned infinite in XIXth century brazilian portuguese. São Paulo: Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo; 2019.

ABSTRACT

Introduction: This research dove into the variation between *estar* + *gerund* and *estar* + *a* + *infinitive* in the XIXth century Brazilian Portuguese of five major Imperial Provinces (Bahia, Minas Gerais, Pernambuco, Rio de Janeiro and São Paulo) and its relations with the Imperial nation building project. The two aforementioned constructions are understood by native speakers to be highly pronounced markers of the Brazilian and European varieties of Portuguese, respectively, a situation which was at least partially mirrored during the Brazilian Empire era. **Objectives:** To investigate whether the above mentioned constructions where as strongly associated with a “Brazilian identity” during Imperial Times as they are now. **Methodology:** A conceptualization of the XIXth century mindset on nationhood and language was built through the Reading of contemporary discourses on history, culture and language published by key figures of Brazilian intellectuality. From this, fifteen of authors split among the five regions where elected as “model speakers” and some of their work taken as material basis for the corpus. The data acquired from this was then categorized according to social and linguistic criteria and the results were then analyzed statistically analyzed. **Results:** The actual usage of the two variants has shown itself to be highly diversified, its usage rates varyingly significantly even among author of the same region or political affiliation. Regional variation, however, when taken by itself was found out to be significant both when taken in the larger macro regional level (southeast *versus* northeast) but also among provinces, with a higher rate of usage of *estar* + *a* + *infinitive* in the southeast and especially in the province of Rio de Janeiro. Linguistic factors like the presence of interpolation or certain types of surrounding words have also been be significantly correlated to the variable albeit the reason for that remains inconclusive. **Conclusions:** In a general sense, linguistic discourse and linguistic praxis strongly differ from each other and the variation between the two constructions appears to be more strongly tied to regional and linguistic factors than social factors.

KEYWORDS: Gerund. Prepositioned Infinitive. Brazilian Portuguese. Imperial Brazil. Sociolinguistics. Linguistic Change.

Tabela 1: Distribuição de dados por autor	93
Tabela 2: Regressão logística por autor	95
Tabela 3: Modelo de regressão logística por autor (Dom Casmurro).....	97
Tabela 4: Distribuição de dados por região.....	98
Tabela 5: Realização de infinitivo gerundivo ao longo do tempo,	99
Tabela 6: Modelo de regressão logística por região	100
Tabela 7: Distribuição de dados por província.....	100
Tabela 8: Modelo de regressão logística segundo província.....	101
Tabela 9: Modelo de regressão logística de Minas Gerais e São Paulo contra Rio de Janeiro	102
Tabela 10: Modelo de regressão logística de Rio de Janeiro vs outras províncias	102
Tabela 11: Modelo de regressão logística segundo nascimento.....	104
Tabela 12: Modelo de regressão logística por ano de publicação	106
Tabela 13: Modelo de regressão logística por ano de publicação e nascimento	106
Tabela 14: Distribuição de dados segundo unidade precedente	108
Tabela 15: Modelo de regressão logística por unidade precedente	109
Tabela 16: Distribuição de dados por unidade seguinte	110
Tabela 17: Modelo de regressão logística por unidade seguinte	111
Tabela 18: Modelo de regressão logística por unidade precedente e seguinte.....	112
Tabela 19: Distribuição de dados por colocação pronominal	114
Tabela 20: Modelo de regressão logística por colocação pronominal.....	114
Tabela 21: Distribuição de dados por tipo de unidade sintática.....	115
Tabela 22: Modelo de regressão logística por tipo sintático.....	117
Tabela 23: Modelo de regressão logística por função sintática com dados limpos.....	118
Tabela 24: Distribuição de dados por número do verbo	119
Tabela 25: Modelo de regressão logística por número do verbo.....	120
Tabela 26: Distribuição de dados por tempo do verbo	121
Tabela 27: Modelo de regressão logística por tempo do verbo.....	121
Tabela 28: Distribuição de dados por ocorrência de interpolação.....	122
Tabela 29: Modelo de regressão logística por interPOlação.....	123
Tabela 30: Distribuição de dados por tipo semântico do verbo	125
Tabela 31: Modelo de regressão logística por tipo semântico do verbo.....	127

Tabela 32: Modelo de regressão logística com grupos de fatores linguísticos	129
Tabela 33: Modelo de regressão logística com região, nascimento e publicação	133
Tabela 34: Modelo de regressão logística por ano de nascimento e publicação (nordeste) ..	134
Tabela 35: Modelo de regressão logística por ano de nascimento e publicação (sudeste)	134
Tabela 36: Modelo de regressão logística com autor e grupos de fatores linguísticos.....	135
Tabela 37: Modelo de regressão logística com província e grupos de fatores linguísticos ...	135
Tabela 38: Modelo de regressão logística com região e grupos de fatores linguísticos	136

Gráfico 1: Distribuição geral de dados por autor	94
Gráfico 2: Distribuição de dados por região	99
Gráfico 3: Distribuição geral de dados por província.....	101
Gráfico 4: Taxa de emprego de eIP por ano de nascimento.....	104
Gráfico 5: Taxa de emprego de eIP por ano de publicação	105
Gráfico 6: Distribuição de dados por unidade precedente	108
Gráfico 7: Distribuição total dos dados por unidade precedente.....	109
Gráfico 8: Distribuição de dados por unidade seguinte	110
Gráfico 9: Ocorrências de dados por unidade seguinte	111
Gráfico 10: Distribuição de dados por tipo de sentença	116
Gráfico 11: Distribuição total de tipos de sentenças.....	116
Gráfico 12: Distribuição de dados por número do verbo.....	120
Gráfico 13: Distribuição de dados por tempo do verbo	121
Gráfico 14: Distribuição de dados por ocorrência de interpolação	122
Gráfico 15: Distribuição de dados por tipo semântico do verbo	126
Gráfico 16: Efeitos de ano de nascimento do autor e publicação da obra de acordo com região	133

SUMÁRIO

1: INTRODUÇÃO	14
2: PRESSUPOSTOS TEÓRICOS.....	17
2.1: Linguística histórica, sociolinguística, linguística sociohistórica.....	17
2.2: A variável em questão	28
3: EXPOSIÇÃO HISTÓRICA E HISTORIOGRÁFICA.....	36
3.1: A história e a historiografia do Brasil	36
3.1.1: Rocha Pita e a história no setecentos	39
3.1.2: O século XIX e o nascimento do Brasil	44
3.2: Identidade e discurso linguístico no Brasil.....	50
4: PERFIS DOS AUTORES	69
4.1: Perfis	70
4.2.1: Amadeu Amaral	70
4.2.2: Bernardo Guimarães.....	72
4.2.3: Carneiro Vilela.....	72
4.2.4: Franklin Távora.....	73
4.2.5: Inês Sabino	74
4.2.6: Joaquim Felício dos Santos	75
4.2.7: Júlia Lopes de Almeida	75
4.2.8: Júlio Ribeiro.....	76
4.2.9: Lindolfo Rocha	77
4.2.10: Luísa de Camargo Pacheco.....	78
4.2.11: Machados de Assis	78
4.2.12: Manuel Antônio de Almeida	79
4.2.13: Manuel Benício	80
4.2.14: Vicente Félix de Castro	80
4.2.15: Xavier Marques.....	81
5: APRESENTAÇÃO DOS DADOS	82
5.1: Preditores extralinguísticos	82
5.2: Preditores linguísticos	84
5.3: Distribuição de dados por autor	93
5.4: Distribuição de dados por região e estado.....	98
5.5: Distribuição de dados por ano de nascimento do autor e ano de publicação	103

5.6: Distribuição de dados por unidade precedente e seguinte.....	107
5.7: Distribuição de dados por tipo de sentença	115
5.8: Distribuição de dados por número e tempo do verbo e ocorrência de interpolação	119
5.11: Distribuição dos dados por tipo semântico do complexo verbal	125
5.12: Análises multivariadas	127
6: CONCLUSÕES	138
REFERÊNCIAS	140

1: INTRODUÇÃO

Em 1822, quase duzentos anos atrás, ocorreu o evento que se convencionaria chamar de “Independência do Brasil” – a separação política entre a colônia americana e sua metrópole europeia. O fato definitivo foi a culminação de uma série de eventos que ocorriam desde o estabelecimento da colônia e que se tinham acelerado conforme o agravamento da crise econômica portuguesa levava a coroa lusitana a aumentar a taxaço sobre a população brasileira. Do século XVIII para frente, revoltas começavam a ocorrer no Brasil, algumas delas de tendências liberais e republicanas, porém foi o modelo monárquico o grande vencedor da contenda, no século XIX, estabelecendo uma dinastia imperial que não duraria até o fim do século.

A criação de um estado de proporções geográficas mais ou menos definidas (ajustes seriam feitos por um tempo mais, nem sempre segundo a vontade do governo central, mas em grande parte o território do Brasil pouco mudaria desde a independência) e um sistema político não eram as únicas necessidades para se construir um projeto nacional. Se o estado em suas dimensões física e política pode ser entendido como uma máquina, faltava um fantasma que lhe animasse e unisse. Havia o país, mas não a nação.

Muitos foram os aspectos empregados na consolidação de um espírito nacional e não há dúvidas de que de cada deles seria passível de grandes discussões – desde os aspectos raciais até a busca por tradições folclóricas - mas não se pretende aqui explicar toda a máquina do mundo, mas apenas uma pequena engrenagem de um de seus subsistemas, aquele que por muitos foi chamado de “a língua nacional”.

Nascido colônia e querendo ser nação aos moldes europeus, o Brasil Império elegeu o português como sua língua nacional. Havia outras línguas no país, muitas das quais por décadas foram as principais formas de comunicação entre povos, mas essas acabariam suplantadas pelo idioma lusitano (salvo em comunidades isoladas), representando pouca ameaça real ao domínio da língua oficial. Se havia uma ameaça à identidade da língua do Brasil essa não eram os falares dos povos nativos ou sequestrados, mas o da Lusitânia. “Como poderia o Brasil expressar as próprias ideias se usava a língua de seus dominadores?”, diziam alguns, inspirados pela ciência de então. Era preciso que se criasse uma língua para a nação, uma que trouxesse à tona o espírito e o gosto de seus falantes. Para que tal língua existisse era necessário encontrar as diferenças

entre ela e sua base, isso é, sua identidade natural, fosse essa orientada pelas formas de dizer ou pelas palavras únicas a cada lado do atlântico. Todos esses elementos foram empregados de modos diversos para que se tentasse criar, em discurso, um verdadeiro Português do Brasil, e é a intenção aqui se averiguar um desses modos de dizer.

Assim, nas páginas porvir, se fará um exame de como um certo fato da língua portuguesa foi empregado num discurso e numa prática de nacionalidade. Esse fato, a mera realização de um certo tipo de perífrase, é apenas um detalhe mínimo do repertório linguístico da “última flor do lácio”, porém é um fato que foi invocado por alguns autores para marcar uma identidade linguística. Para explorar como esse fenômeno se deu, este trabalho será dividido em seis capítulos, o primeiro dos quais é o presente, um resumo de suas ideias e de sua estrutura geral. Um guia, por assim dizer, do que espera a seguir.

No segundo capítulo, os pressupostos teóricos centrais do trabalho serão apresentados, assim como o fato linguístico de interesse – a variação entre duas formas de realização de um aspecto progressivo. É nesse momento que se apresentarão as ideias fundamentais e as justificativas centrais do método usado.

Seguindo-se a esse, o terceiro capítulo explorará o universo conceitual da historiografia brasileira oitocentista e os discursos linguísticos correntes na intelectualidade de então. Será por meio desses que se tentará compreender quais ideias e tradições governavam as relações entre a intelectualidade do período, o que pensavam sobre sua língua e seu país e quais projetos carregavam em seus textos. Esse tipo de imersão é fundamental para uma análise desse tipo, pois é apenas através dela que se poderia compreender o discurso oficial. Nesse mesmo sentido, o quarto capítulo é uma coletânea de breves resumos das vidas dos autores cujas obras serviram de fonte a esse projeto.

Após isso, o quinto capítulo tratará da apresentação e análise dos dados obtidos, de acordo com os métodos detalhados no segundo capítulo. É nesse momento que se apresentarão os critérios utilizados para classificar o *corpus*, se fará as primeiras análises exploratórias e os testes estatísticos sobre essas. Por fim, o capítulo seis traz as conclusões sobre as análises realizadas, resumindo os problemas e fatos encontrados ao longo da pesquisa e oferecendo alguns pontos que se poderão erguer para experiências futuras.

Ao seguir esse processo e essa ordem, se espera tentar entender melhor, mesmo que só um pouco, como surgiu essa coisa tão peculiar que se costuma chamar de “cultura brasileira”,

lançando-se mão de todas as ferramentas que se tem disponíveis, desde a leitura interpretativa da história até testes estatísticos. Trata-se, é claro, de um projeto que de partida se pretende complexo, ainda que o material que utilize não necessariamente o seja – é apenas um detalhe linguístico, afinal. O sucesso da empreitada, como se verá no fim, foi parcial e ainda que não se tenha encontrado respostas definitivas, alguns dados de interesse foram revelados e esses podem indicar caminhos futuros a se seguir.

2: PRESSUPOSTOS TEÓRICOS

Onde se apresentam os pilares teóricos que sustentam essa pesquisa, os aspectos gerais da teoria sociolinguística e da linguística histórica e o material linguístico sobre qual se trabalhou.

2.1: LINGUÍSTICA HISTÓRICA, SOCIOLINGUÍSTICA, LINGUÍSTICA SOCIOHISTÓRICA

Ao longo de toda sua história, incontáveis questões assombraram e encantaram a humanidade, forjando na imaginação coletiva uma complexa cosmologia de mistérios e narrativas, tudo em busca de dar alguma forma de explicação às coisas do mundo. É desse modo que se formam nas mitologias as histórias sobre a origem do céu e da terra, sobre o nascer dos primeiros seres humanos, a aurora de suas cidades e de seus reis e, como não poderia deixar de ser, de suas línguas.

As discussões sobre a língua ocupam o pensamento humano desde a antiguidade clássica, onde já se via a natureza da relação entre os nomes e as coisas ser discutida por Platão (*circa* 420 a.C. – *circa* 340 a.C.) em seu *Crátilo*. Na Índia antiga, Panini (*circa* IV a.C.), o gramático, dedicou parte significativa de sua vida a explorar a natureza da linguagem, desenvolvendo um projeto complexo de descrição das regras do sânscrito (HARRIS & TAYLOR, 1997). Que mesmo na antiguidade se discutisse a natureza fundamental da língua não é surpreendente, embora não deixe de ser interessante, e essas primeiras discussões servirão como um prelúdio a todo o futuro de um campo. Nesse momento, talvez um dos pontos mais interessantes a se manter em mente seja a discussão presente em Varrão (116 a.C. – 27 a.C.) sobre a questão do uso coletivo como origem da regularidade e da irregularidade a despeito da liberdade dos indivíduos em poder empregar termos livremente (HARRIS & TAYLOR, 1997, p. 46).

O período medieval veria as discussões sobre a língua tomarem outros rumos e certas ideias acabariam por surgir nesse momento, como a possibilidade de que línguas fossem aparentadas e de que todas viessem de uma única língua mãe comum, a qual muitos, por razões mais religiosas do que de qualquer outra ordem, acreditavam ser o hebraico (ECO, 2008). Já no

século XVIII, começam a surgir trabalhos que se embasavam na comparação efetiva de estruturas e palavras, buscando, com isso, argumentar o parentesco entre línguas bastante distintas. É nesse período que surge o trabalho de William Jones (1746 – 1794), emissário britânico na Índia que observou similaridades entre sânscrito, grego e latim (HARRIS & TAYLOR, 1997, p. xviii). Apesar dos diversos erros presentes nos métodos e do pouco interesse linguístico, estritamente falando, William Jones ainda é reconhecido como um marco no desenvolvimento da linguística como uma ciência. O projeto de William Jones e seu método comparado acabou servindo de prenúncio ao que se tornaria a ciência linguística propriamente dita, uma vez que essa, no século XIX, se pautaria pela comparação do mesmo modo que outras ciências o fariam. Houve no 1800 um movimento geral em direção ao comparativismo, sendo o darwinismo talvez o maior expoente disso, mas outros campos, como a antropologia e a linguística, também usariam os mesmos métodos. Não por menos, August Schleicher (1821-1868), um dos nomes mais notáveis do período, tem um livro intitulado *Die Darwinsche theorie und die sprachwissenschaft*¹. O próprio uso da *stammbaum* ou *árvore genealógica de línguas* por Schleicher em seu trabalho, assim como a noção geral de que línguas podem ser comparadas para que se reconstrua seus antecessores tem paralelos com os princípios gerais da biologia oitocentista. É nesse século também que conceitos como o “índo-europeu” (ou “índo-germânico”) ganham força dentro do pensamento linguístico, juntamente a uma série de outras ideias que começam a pautar a relação entre línguas, povos e sociedades de maneira mais determinística. William Dwight Whitney (1827-1894), linguista americano consideravelmente influente na tradição brasileira, é talvez um dos expoentes mais claros desse tipo de linha de pensamento, advogando que a qualidade de uma língua seria um fator determinante no sucesso de uma nação.

Estabelecido o tempo como uma força fundamental da natureza e a mudança como um resultado inevitável de sua ação, o pensamento sobre a língua passou a buscar a relação entre os dois ao mesmo passo em que tentava recuperar o estado primeiro das línguas, a pedra angular da Torre de Babel, ambos objetivos aos quais a comparação servirá muito bem. Assim, como se faria para as famílias de animais, as línguas serão traçadas em árvores genealógicas e suas estruturas, seus “órgãos”, como alguns chamarão, serão comparados para definir os graus de parentesco.

¹ “A teoria darwinista e a ciência da linguagem”, em tradução liberal.

Não cabe aqui uma exposição prolongada das minúcias do pensamento de Schleicher ou dos linguistas do período, mas basta dizer que junto à sua noção de mudança viriam também uma série de outras ideias cujo impacto se sentirá na ciência da linguagem por mais de um século. Dentre essas, se destaca a ideia de que as línguas teriam um percurso fundamental que as levaria de um a outro estado dentro do sistema de organização morfológica proposto por Humboldt, indo do mais “básico” e “truncado” isolacionismo até a “perfeição” da morfologia declinativa e daí de volta ao começo num processo decadente. (SCHLEICHER, 1863 [1983]). A ideia dessa progressão acabará por assombrar linguistas em todo o mundo mesmo uma vez estabelecidas novas formas de pensar a linguagem, mas não será a única herança que o pensamento linguístico do oitocentos trará a essa.

Ao definir seu projeto de ciência da linguagem, Schleicher o faz com vistas de formar uma ciência aos moldes da modernidade, baseada na observação e comparação necessários a uma ciência natural cujo objeto é um fenômeno “sujeito a toda a série de fenômenos que abarcamos sob o nome de ‘vida’” e cujos organismos “não foram jamais dirigidos pela vontade dos homens; [mas] surgiram e se desenvolveram por si mesmos de acordo com leis precisas, elas envelheceram e morreram” (SCHLEICHER, 1863 [1983], p. 21, tradução própria). Tal ideal de uma ciência linguística que se pudesse realizar segundo os preceitos e suposições de uma ciência natural, completamente isolada do elemento humano e governada tão e tão somente por forças fundamentais do mundo, será caro à ciência oitocentista, fato que não deixará de trazer impactos na lógica apresentada pelos autores de então. É por acreditar que a língua é um fenômeno de tal ordem que Schleicher a tomará como o mais adequado critério para a classificação dos povos – isso e, ironicamente, sua qualidade como um fenômeno fundamentalmente humano (SCHLEICHER, 1863 [1983]).

Na mesma linha de Schleicher estavam outros cientistas do período, como Whitney, Müller e Diez. Diez, inclusive, será o responsável pela publicação da mais importante gramática histórica das línguas românicas produzida dentro dos princípios da linguística comparada, a “*Grammatik der romanischen Sprachen*”. Algo mais tarde, na América, projetos pautados pela linguística comparada do oitocentos buscarão a classificação das línguas ameríndias, com maior ou menor afincamento, de acordo com aspectos regulares de cada uma delas. Em capítulos futuros será dada maior atenção à repercussão do projeto comparatista no pensamento americano. Por hora basta dizer que se por um lado esse pouco se realizou como uma prática científica, as teorias foram amplamente utilizadas para justificar uma série de políticas questionáveis.

O princípio da mudança como uma força inabalável, proposto já em Schleicher, acabaria por se consolidar ainda mais quando inserido dentro do projeto dos neogramáticos, ainda no XIX, segundo os quais os processos de mudança linguística operariam como forças absolutas dentro de sua esfera de influência – é o princípio da regularidade da mudança. Tal como os linguistas comparados que os procederam, os neogramáticos tinham como seu objetivo maior explicar a mudança das línguas e para tanto viam no indo-europeu sua principal fonte de dados observáveis. Ainda nesse momento, a gramática comparada será o principal produto intelectual dos linguistas, ainda que certas questões de outro cunho comecem a ganhar forma.

Nesse período, no pensamento de Hermann Paul (1846-1921), vêm à tona considerações que acabarão apagadas sob a sombra da tradição que tomará os holofotes da ciência linguística a partir de 1916, mas que previam questionamentos visíveis nas obras de autores futuros. Em seu *“Prinzipien der Sprachgeschichte”*, Paul delineia as nuances de suas considerações linguísticas, tecendo críticas aos comparatistas, autores de obras muito próximas das gramáticas descritivas, defendia ele, e propondo um novo modo de pensar a língua. De maior interesse para a narrativa aqui proposta é o que Paul dirá sobre a variação linguística, ou ao menos sobre algo que se pareceria muito com ela. Sua noção da gradação da mudança toma um passo a mais do que a apresentada por Schleicher, considerando que cada ato particular de fala, isso é, cada iteração do processo de dizer, é em si um passo em direção à mudança sob a força dos processos gerais. Nesse jogo, o indivíduo falante existe com um poder restrito de influência sobre sua linguagem, suas possibilidades de intervenção limitadas pela natureza maior da mudança. Ainda que o indivíduo exista, como o próprio Hermann Paul proclamará em sua rejeição a uma “psicologia dos povos”, sua força empalidece ante as pressões da mudança como um todo (PAUL, 1888).

Ainda que reconheça, como o fizeram Schleicher e os neogramáticos, a língua como um objeto natural, Paul não nega a existência do falante e seu papel nos processos comunicativos - sua própria definição do objeto do linguista levaria a tal conclusão, uma vez que esse seria não a língua como uma gramática, mas como o conjunto total dos sons produzidos por todos os indivíduos falantes (PAUL, 1888, p. 2). As considerações de Paul e seu desejo de acessar o maior número possível de eventos de fala no maior número possível de condições, espelha aquele que outros trarão no futuro, porém seu projeto acabaria por não ir a cabo, sufocado pela contundente força da linguística Saussuriana.

Há qualquer coisa de irônico que tenha sido Saussure, ou ao menos certa leitura que se fez de sua obra e certa linguística que dessa se originou, o responsável por mover as atenções do pensamento linguístico dos processos de mudança e evolução onde até então se concentravam para o momento presente. Irônico pois que Saussure era ele mesmo um linguista dedicado à história cujo trabalho teórico, mais tarde confirmado com a descoberta do hitita, permitiu exorcizar fantasmas que por décadas haviam assombrado o estudo das línguas indo-europeias. A proposição, no *Cours de linguistique générale*, da dicotomia diacronia e sincronia e sua defesa do estudo da língua como um sistema de regras estruturadas acabaria por dar origem ao estruturalismo (SAUSSURE, 1916 [1995]), vertente dominante da ciência da linguagem até a metade do século XX.

Com a ascensão do estruturalismo ao patamar de “verdadeira ciência da linguagem”, os estudos históricos acabaram relegados ao segundo plano. É na guisa do projeto estruturalista que se verão publicados, nos Estados Unidos, os primeiros grandes estudos sobre as línguas nativas da América do Norte. Foi nesse momento que surgiram importantes nomes da linguística descritiva, como Franz Boas (1858-1942), Edward Sapir (1884-1939) e Leonard Bloomfield (1887-1949). O período estruturalista foi uma era fértil para a linguística, vendo se expandir o leque de famílias descritas e línguas estudadas, fato que não em pouco se deveria a um *zeitgeist* científico. A antropologia crescia nesse mesmo período, impulsionada pela onda globalizadora que se seguiu à Primeira e Segunda guerras mundiais. Naturalmente, o estudo de culturas não poderia ignorar o estudo das línguas por elas faladas e a lógica estruturalista, seus métodos de descrição e análise, se prestavam bem aos interesses desse grande projeto cultural (ROBINS, 1967).

Após anos de domínio estruturalista, um novo paradigma linguístico surgiu na década de 1950, quando um certo Avram Noam Chomsky (1928-presente), um jovem cientista formado pela Universidade da Pensilvânia, publicou *Syntactic Structures*. Essa seria uma obra de grande importância para a linguística, o primeiro pilar do que se tornaria a linguística gerativa. Da obra gerativa, interessa aqui apenas sua suposição sobre a mudança, a qual se poderia resumir como a conceitualização dessa como uma falha de transmissão similar a uma mutação genética ocorrida entre duas gerações, isso é, a mudança ocorre uma vez que um indivíduo falha em reproduzir perfeitamente o sistema da geração que lhe gerou (CHOMSKY, 2009). Com o rápido ganho de força do gerativismo a partir da década de 1960, muito da pesquisa em linguagem se voltou para seu arcabouço teórico, e os estudos da mudança não

foram diferentes nesse aspecto. Importantes pesquisadores, como Lightfoot e Kiparsky, publicaram estudos sobre mudança a partir de uma visão gerativista.

Todavia, algo ocorria em paralelo ao crescimento do paradigma Chomskyano. Uma inquietação profunda persistia na mente de certos pesquisadores cujos interesses seguiam em uma direção menos cartesiana. Dentre esses estava William Labov (1927-presente). Em 1963, Labov publicou um artigo intitulado “*The Social Motivation of Sound Change*” no qual analisava a variação da pronúncia de certas vogais na ilha de Martha’s Vineyard. O fenômeno, Labov defendia, é um caso de mudança em progresso, onde a população mais jovem da ilha apresenta uma maior taxa de centralização na pronúncia de ditongos. Porém a grande chave da questão estaria não simplesmente no tempo, mas na sociedade e a pronúncia de uma forma mais ou menos centralizada seria indicativo de uma postura do falante quanto à ilha – aqueles que mais se sentiam ligados às tradições locais usavam a pronúncia tradicional em contraposição àqueles cujo apreço à ilha era menor (LABOV, 1963). Assim começa a sociolinguística moderna.

Labov faria ainda outros trabalhos na mesma direção, como o famoso estudo das lojas de departamento de Nova Iorque (LABOV, 1986), logo estabelecendo as bases da sociolinguística. No nível teórico, a questão central do campo estava no desejo de incluir nos estudos da mudança os aspectos sociais, isso em vista de uma suposição básica – sendo a língua utilizada por seres humanos, é natural que eles tenham opiniões sobre ela e uma vez que essas opiniões possam ser positivas ou negativas sobre certos jeitos de falar, surge um sistema de estímulo-resposta que leva à mudança. Não se trata de uma formulação completamente inédita já que, em certo sentido, ela retoma proposições feitas por Saussure, Varrão e Paul sobre como as pressões da comunidade impelem o indivíduo a usar sua língua de tal e tal maneira, mas a ideia de que essa pressão possa ser recusada por um grupo de indivíduos é uma novidade pertinente. Mas ainda mais do que isso, a grande inovação de Labov foi metodológica. O estabelecimento do conceito de “variável linguística”, isso é, um elemento presente na língua que pode ser realizada de duas ou mais maneiras sistemicamente equivalentes, mas socialmente distintas fornece um meio quantificável de medir relações sociais por meio da língua.

Uma vez que se estabelece a existência da variação, seus efeitos na mudança são óbvios. Não por menos, Labov publicou três volumes sobre mudança linguística, aptamente intitulados “*Principles of Linguistic Change*” (LABOV, 1994, 2001, 2010). A base de tal consideração é uma progressão lógica da premissa de sobrevivência pelo uso. Assim, enquanto

o gerativismo compreende a mudança como resultante, primordialmente, de um erro de cópia, a teoria da variação entende o problema em termos de pressões sociais e linguísticas que afetam o menor ou maior emprego de formas em competição. Aquelas formas mais utilizadas são passadas aos grupos que se inserem no universo da língua, assim se tornando, para eles, a forma natural de dizer.

Desde os anos 1960, as práticas da sociolinguística tiveram desenvolvimentos consideráveis. A entrevista sociolinguística se tornou uma das principais formas de aquisição de dados, modelos de análise estatística passaram a ser empregados para garantir a significância de diferenças percebidas em conjuntos de dados e as variáveis sociais, originalmente limitadas a categorias macrossociais pré-definidas, como classe social e gênero, foram ampliadas para permitir uma análise sobre categorias e agrupamentos locais. Sobre isso, Eckert (2012, 2016), traça um panorama das várias abordagens da sociolinguística, as quais divide em três “ondas”, ou padrões gerais no comportamento dos estudos sociolinguísticos. A primeira dessas ondas compreenderia os estudos que se dedicaram a averiguar as relações entre usos linguísticos e fatores macrossociais tais como raça, gênero e classe. A segunda, por sua vez:

[...] delved under the large social aggregates of the First Wave, uncovering local categories and configurations under which change takes place”²

e a Terceira:

“departs from the first two by viewing variation as a system of signs, whose meanings emerge in their role in styles that enact social personae or types”³
(ECKERT, 2016, p. 69).

Em termos gerais, se poderia compreender a diferença entre a primeira e segunda ondas como uma diferença de escopo. Se à primeira onda interessavam fatores gerais, universais e aplicáveis a qualquer comunidade, na segunda onda tais categorias, ainda que continuem tendo certa relevância, passam ao pano de fundo enquanto questões relativas à comunidade específica ganham destaque. Disso se origina o estudo realizado por Eckert numa

² “[...] mergulhou sob os grandes agrupamentos sociais da Primeira Onda, desvelando categorias e configurações locais sob as quais a mudança ocorre”

³ “se afasta das primeiras duas ao ver a variação como um sistema de signos cuja significação emerge de seus papéis em estilos que encenam personae ou tipos”

comunidade escolar de Detroit, o qual se poderia separar em duas etapas diferentes: observação etnográfica e análise linguística. Na primeira dessas etapas, Eckert se dedicou a observar a comunidade de interesse, verificando suas relações internas e as interações entre seus membros, formulando a partir das informações recolhidas categorias relevantes para tal comunidade. Uma vez que realizado tal processo de observação e estabelecidas categorias de interesse, se passa a observar padrões de uso linguístico empregados pelos falantes englobados nas categorias (ECKERT, 1989).

Se a segunda onda se distinguiria da primeira devido ao grau de aproximação da comunidade alvo e aos métodos de recolha e análise que necessariamente resultariam dessa maior aproximação, a terceira onda se afasta das outras ao assumir um papel ativo do indivíduo falante no emprego de estruturas linguísticas para a construção de determinadas *personae* sociais, isso é, a terceira onda assume que, em algum nível, um falante é um agente de uso linguístico, capaz de mobilizar estratégias comunicativas para atingir certos efeitos em seus interlocutores (ECKERT, 2016).

Num exemplo do que seria um estudo de terceira onda, Podesva (2007) analisa o emprego de manipulações de frequência fundamental na construção de *personae* sociais por um indivíduo em particular. Nesse estudo, o indivíduo em questão, Heath, foi observado em três situações diferentes, um churrasco com amigos, o atendimento de um paciente em seu consultório e uma conversa por telefone com seu pai. Dentre essas três situações, Heath apresentaria um maior uso de *falsetto* durante o churrasco (9,07%) em contraste à conversa com um paciente e com seu pai (3,72% e 3,85%, respectivamente). Tal fato, Podesva argumenta, seria resultante da tentativa por parte de Heath de criar distintas *personae* nessa situação, empregando o *falsetto* com maior liberdade durante o churrasco como um dos elementos que constituem a *persona* de uma “diva” (*ibid.*).

De modo algum se deve pensar, porém, que as três “ondas” substituem umas às outras ou mesmo que um estudo não possa considerar ao mesmo tempo categorias gerais e locais. Tais possibilidades existem e são realizadas desde o estudo de Labov sobre Martha’s Vineyard quando ele liga o uso da variável não a classe social, mas à opinião de um falante sobre o lugar onde vive e assume mesmo a posição de que a variável possa ser instrumentalizada.

Outra direção na qual se desenvolveram os estudos sociolinguística é não aquela relativa ao modo de aproximação da comunidade de interesse, mas à localização de tal comunidade no espaço-tempo. Uma vez que a tecnologia humana ainda não permite que se

estude comunidades vindouras, a possibilidade restante é o estudo de comunidades passadas, isso é, a aplicação do método sociolinguístico para a análise da mudança linguística em sentido *diacrônico*. A ideia do tempo e da mudança já existia na sociolinguística desde Labov na forma do “tempo aparente”. Fundamentalmente, essa ideia dita que a análise de indivíduos de idades diferentes permitiria ver tendências do fenômeno ao longo do tempo, uma vez que os momentos de aprendizagem da língua divergiriam em indivíduos de idades diferentes. Ainda assim, mesmo num cenário como esse, a aquisição de dados se faz no tempo presente e embora um elemento temporal tenha sido inserido, a análise é puramente sincrônica. Porém, desenvolvimentos verdadeiramente diacrônicos aconteceram na linguística desde a década de 1980, como os estudos em linha com aqueles realizados por Romaine (1982).

Naturalmente, tal divisão é, antes de mais nada, didática. Não há razão lógica para se separar os fenômenos de mudança dos fenômenos de variação, ambos são parte de um mesmo processo maior, do mesmo modo que mutações genéticas, seleção natural e evolução biológica são um contínuo no progresso da vida terrestre. Se é possível que se separe variação e mudança, isso é tão e tão somente devido a uma estratégia heurística.

Deixando isso de lado, é possível aceitar a separação do processo de evolução linguística em variação e mudança e nesses termos o interesse dos trabalhos de Romaine se revela. Na sua origem no oitocentos, a linguística histórica tomou uma postura epistemológica aonde não caberia a consideração do social. Como realizada por Whitney e Schleicher, a ciência da mudança linguística presumia uma língua destituída de falantes cuja evolução se daria por regras e processos naturais. E embora certos autores como Hermann Paul tenham se interessado pelo indivíduo, um sistema real para o estudo do efeito desse na mudança não tomou forma naquele momento. Em contrapartida a essa tradição, a linguística sociohistórica iniciada por Romaine busca no elemento social parte das razões das transformações pela qual passam as línguas do mundo. Em seu seminal “*Socio-historical Linguistics: its status and methodology*”, Romaine (1982) emprega métodos tipicamente relacionados à sociolinguística, como o emprego de análises estatísticas de *corpora* e consideração de diferentes estilos, para investigar o uso e desenvolvimento de sentenças relativas do inglês médio escocês. Após a proposta original de Romaine, a linguística sociohistórica floresceu em outros contextos e com outros métodos (*vide* CONBOY, 2010; VAN DER WAL & RUTTEN, 2013; VAN DER WAL & NEVALAINEN, 2014). Mais recentemente, Conde-Silvestre (2016) lançou mão de técnicas da chamada “terceira onda” para analisar cartas pessoais de uma família inglesa do século XV, os Stonor. Nessa análise, Conde-Silvestre interpreta os Stonor como uma *comunidade de práticas*,

um agrupamento de indivíduos que compartilham certas ideias e interesses, e a partir disso realiza uma análise de seus hábitos linguísticos como manifestos nas cartas. Embora existam várias lacunas e pontos questionáveis nessa abordagem, seus méritos são notáveis, pois ela insere considerações importantes sobre a transmissão de modas linguísticas dentro de um grupo no estudo de seu desenvolvimento linguístico.

A interface entre sociolinguística e linguística histórica é frutífera, tanto já se sabe, mas não é de toda uma estrada sem espinhos. Muito pelo contrário, as dificuldades são muitas e altamente complexas. A suposição central da sociolinguística de que se pode correlacionar fatos da língua com fatos da sociedade exige que se compreenda profundamente o grupo alvo, os modos como as diversas partes dele interagem, o que reconhecem como “si mesmo” e o que tomam como “outro”. E no caso de certas sociedades, essa tarefa traz alguns complicadores adicionais. Tal é o caso do Brasil.

Invadido por portugueses em 1500, o território que se tornaria o Brasil se insere no contexto das colônias americanas formadas durante a Era das Navegações e, como essas, enfrentou uma longa história de exploração, abuso e dominação por parte da metrópole. Sendo a principal colônia portuguesa numa América do Sul cercada por territórios submetidos à Espanha, o Brasil se coloca numa posição sociocultural bastante particular na história do mundo mesmo antes do processo que o levaria a se tornar a única nação do chamado Novo Mundo a se declarar um Império. Não obstante, a intrincada rede de relações entre os povos que formam o país, suas dimensões colossais, a sua história de povoação e desenvolvimento econômico e cultural e as especificidades de seu processo de independência e *nationbuilding* resultam em uma situação social complexa. Nesse âmbito, como ocorreu em outras nações forçadas sob o jugo da dominação estrangeira, a discussão sobre a língua falada no Brasil ganhará no oitocentos densos contornos de nacionalismo. Em vista disso, o presente trabalho visa verificar como certos grupos de falantes empregaram certas formas linguísticas em seu discurso acerca da língua brasileira, assim como tais formas eram de fato realizadas por falantes de português nascidos no Brasil.

O cruzamento de uma proposta de análise sociolinguística cujas bases se centram em um conceito como “identidade cultural” com uma outra sociolinguística dedicada à dimensão histórica traz consigo um número de complicadores. Como se viu em Eckert (1989), uma parte fundamental de estudos de segunda e terceira onda é a familiarização com as categorias e relações específicas à comunidade alvo, seja essa um grupo de amigos, os alunos de uma escola

em Detroit ou imigrantes nordestinos que vivem em São Paulo⁴. Tal identificação exigiria reconhecer os valores relevantes para o grupo em questão, seus relacionamentos com outros grupos, sua visão quanto ao mundo e todo o cenário conceitual compartilhado por tais indivíduos. Em outras palavras:

“[...] o estudo da língua é um campo multidisciplinar, um campo no qual a Linguística normal é indispensável, mas no qual outras disciplinas, como a sociologia, a antropologia social, a educação, a folclorística e a poética são também indispensáveis” (HYMES, 2001, pg. Vii, tradução própria).

Se tal familiarização com o ambiente social dentro do qual uma língua é usada exige de um sociolinguista dedicado a questões sincrônicas um esforço considerável, uma vez que se decida operar sobre um grupo temporalmente distante, uma nova ordem de problemas surge. A falta de acesso à comunidade falante implica a necessidade de um trabalho arqueológico de reconstrução dos sistemas e relações sociais então vigentes, o que resulta, por sua vez, numa série de perguntas difíceis de se responder. Quais grupos eram tomados como modelo por esses falantes? Quais valores sociais eram desejáveis? Quais eram rejeitados? Quão diversas eram as situações sociais então comuns? Que regras regiam o comportamento desses grupos nessas situações? Quem era e quem desejava ser aquele que vivia nesses tempos idos? Todas essas são perguntas cabíveis a um estudo da realidade social de um grupo qualquer, mas difíceis de responder uma vez que se pense sobre um grupo extemporâneo. Se adicione a isso uma dimensão individual e o problema ganha contornos de pesadelo.

Naturalmente, não seria difícil se realizar um estudo sociohistórico de bases puramente variacionistas. Uma vez que se tenha dados o bastante e se trabalhe apenas categorias externas ao falante e relativamente universais, tais como sexo, gênero e local de origem, bastaria apenas a classificação e análise estatística para que se pudesse gerar um panorama da variação dentro

⁴ Ressalte-se aqui que o próprio conceito de uma “comunidade” não é algo fácil de se tratar, uma vez que uma mesma comunidade não se limita a indivíduos que dividam um mesmo espaço geográfico, especialmente quando se considera o mundo moderno, onde a presença de meios de comunicação instantâneos permitem que grupos em locais muito distantes formem redes de contato bastante próximas. Não apenas isso, mas um mesmo indivíduo normalmente faz parte de mais de um grupo em qualquer dado momento, daí a necessidade de se considerar diferentes *personae* – isso é, pessoas sociais empregadas por indivíduos em momentos diferentes (*vide* PODESVA, 2007, 2011).

da língua. O que se deve questionar, contudo, é a realidade de tais dados como uma representação de uma situação social. Ainda que se possa constatar com certa facilidade que homens e mulheres de um determinado período passado realizassem seus atos linguísticos de forma diferente, estabelecer as razões por trás de tal fato e os mecanismos sociais envolvidos não é tarefa trivial. Se, hipoteticamente, um certo homem desse período apresentasse um padrão de uso mais próximo daquele de uma mulher, como tal fato deveria ser interpretado pelo linguista? Seria isso um signo de feminilidade ou um instrumento empregado mais comumente por mulheres, mas cujo valor não seria imediatamente classificável como “feminino”? Tomando de base a proposta do estudo presente, se for descoberto um determinado padrão de uso “brasileiro” e outro “português” e se for encontrado um falante brasileiro que lança mão de um padrão português seria adequado assumir, de imediato, que tal falante se identificaria mais com Portugal do que com o Brasil?

Responder a tais perguntas, ou sequer erguer hipóteses sobre elas, exige um considerável esforço de familiarização com o período alvo. Dada a natureza da mudança linguística, da variação e dos dados utilizados em pesquisas históricas inevitavelmente se deverá manter extrema cautela quanto às decisões tomadas. Por outro lado, o emprego de métodos de análise quantitativa tem se provado extremamente poderoso na elucidação de questões sociais sincrônicas ao mesmo tempo em que a linguística de *corpus* e a análise quantitativa de *corpora* tem sido aplicadas para lidar com processos diacrônicos (*vide* NEVALAINEN, 2006, NOBBELS & RUTTEN, 2014).

Em resumo, o que se ergue aqui é primordialmente uma proposta de exercício de método, uma tentativa de se cruzar os avanços em dois campos da sociolinguística em busca de uma melhor compreensão das relações entre fatores sociais complexos e a evolução das línguas humanas. Em certos termos, se busca aqui criar uma ponte entre fenômenos de ordem sincrônica e diacrônica, se empregando métodos já testados a uma situação limite.

2.2: A VARIÁVEL EM QUESTÃO

Cunha (1986), discutindo sobre a natureza inovadora ou conservadora do português brasileiro em relação à sua matriz europeia, ergue uma questão quanto à dificuldade em se tratar

tais conceitos no contexto de uma língua de formação tão heterogênea. O português brasileiro, afinal, é uma língua falada por mais de duzentos milhões de pessoas num território verdadeiramente continental e de formação díspar. Mas ainda que se assumisse um certo português brasileiro padrão, seja esse uma média dos incontáveis falares brasileiros ou uma sublimação conceitual de suas características mais típicas, haveria espaço para questionar se aplicar uma dessas categorias – inovador ou conservador – seria cabível à língua como um todo. Se por um lado diversas das estruturas sintáticas do português brasileiro seriam muito próximas da norma do português médio, como defendem Galves *et al.* (2006), de sua fonologia e de muito de sua morfossintaxe não se poderia dizer o mesmo. Em oposição a essas mudanças, a realização do aspecto imperfectivo via gerúndio seria um elemento de conservação do português brasileiro, remetendo à norma mais antiga do português médio, em contrapartida à forma comum no português europeu, a combinação de uma preposição *a* e um infinitivo, fenômeno que alguns chamam de “infinitivo gerundivo” (MOTHÉ, 2006, 2007) ou de “infinitivo preposicionado” (BARBOSA & COCHOFEL, 2005). Sem seguir o alinhamento teórico das autoras, manteremos o nome “infinitivo preposicionado” (daqui para frente, IP).

O favorecimento do IP em português europeu em contextos onde o português brasileiro apresentaria um gerúndio pode ser constatada com grande facilidade. Basta pegar qualquer texto contemporâneo escrito em Portugal para que se verifique a alta taxa de ocorrência de tal forma. Vide, por exemplo, o trecho a seguir, retirado de um site de notícias português:

NieR: Automata **está a celebrar** um ano de existência, sendo esse um dos motivos pelos quais **está a ser** tão falado, mas também chamou a atenção pela conquista de um prémio D.I.C.E., onde nenhum responsável da Square Enix esteve presente (ênfases adicionadas)

Pouco surpreendente seria que, fosse essa mesma notícia veiculada num site brasileiro, se visse nos dois pontos destacados um gerúndio. Tal diferença de estilo foi constatada já no século XIX por Paranhos da Silva, segundo quem o uso de um IP seria o suficiente para que um leitor brasileiro identificasse a aderência de um texto à norma portuguesa e o rejeitasse (PARANHOS DA SILVA, 1881, p. 279). A rejeição suposta por Paranhos da Silva é um fato de interesse particular, uma parte de um discurso mais amplo que corria entre a intelectualidade brasileira no Segundo Reinado, mas isso será tratado no momento certo. Por ora, o fato

importante é que essa é uma distinção já antiga. Quão antiga, porém, é uma questão válida para o momento.

Mothé (2007), em sua dissertação de mestrado, analisou a evolução do uso do IP ao longo dos séculos XIX e XX no Brasil e em Portugal, identificando um crescimento significativo no emprego dessas formas em Portugal na década de 1920. Considerando, porém, a citação anterior de Paranhos da Silva, tal declaração soaria estranha. Tamanho reconhecimento de uma forma como específica de uma determinada variedade da língua dificilmente se originaria de uma taxa pequena de uso em Portugal. A afirmação de Paranhos da Silva, colocada em tons de senso comum, parece indicar uma familiaridade considerável do autor com as duas variantes, ou ao menos uma consciência de um uso maior de uma forma ou outra de um ou outro lado do Atlântico. Ademais, as constatações levantadas por autores modernos acerca do uso dessa variável no XIX se baseiam, invariavelmente, em textos escritos, cujo estilo tende ao conservadorismo estrutural. Não seria de se surpreender que houvesse nesse material uma certa resistência a utilizar uma forma um pouco mais inovadora. De fato, quando analisando dados de língua falada, a mesma Mothé (2007) encontra uma maior taxa de ocorrência de IP do que em dados retirados de textos escritos do mesmo período.

Os fatos até aqui listados permitem se assumir, com certo grau de segurança, que o IP tenha se manifestado inicialmente na fala, como de fato argumenta Mothé (2007), e que apenas ao longo do tempo tenha se transferido para a escrita. De modo algum isso seria um fato novo já que, como dita a máxima latina, as palavras voam e a escrita permanece. Assim, ainda que uma datação exata das origens de tal variante na fala de indivíduos, sejam esses brasileiros ou europeus, seja impossível de definir, permanece o fato de que essa provavelmente seria anterior às suas primeiras ocorrências textuais. Barbosa (2008) aponta ocorrências do fenômeno em cartas de comércio do século XVIII e pistas do que seriam primeiras ocorrências disso já no século XIII, em “A Demanda do Santo Graal”. Ainda que o último caso não se confirme, dados presentes no *corpus Tycho Brahe* mostram a existência dessa variação desde pelo menos o século XVII:

Nisto se detinha, e em outros empregos de grande serviço de Deus, e delRei aquele espírito de fogo, e de luz; mas dali continuava a mandar por aqueles rios, e matos aqueles pescadores, e caçadores de almas incansáveis, e valorosos (BARROS, 1675)

Com esses fatos postos em jogo, cabe um certo recuo para que se defina com mais clareza o que é a variável de que tratamos aqui. Uma definição ampla de IP colocaria qualquer ocorrência de uma preposição *a* seguida de um infinitivo como uma ocorrência válida, porém essa definição não poderia ser utilizada para os interesses da pesquisa atual, *vide* os casos a seguir:

- 1) Eu vi João **a rir**.
- 2) É certo que ele correu **a ver** de onde vinha o barulho.
- 3) Não sei o que isso tem **a ver** com o assunto.

Dos três casos listados acima, apenas 1 poderia ser considerado uma ocorrência válida para a pesquisa, pois seria a única que se substituída por um gerúndio manteria o mesmo sentido geral. Veja:

- 1b) Eu vi João **rindo**.
- 2b) É certo que ele correu **vendo** de onde vinha o barulho.
- 3b) Não sei o que isso tem **vendo** com o assunto.

Na sentença 1b, a interpretação primária de que o sujeito (“eu”) viu um outro indivíduo realizando uma determinada ação (“rindo”) se mantém a despeito da troca do IP por um gerúndio. Em 2b, porém, o sentido original do complexo *a* + infinitivo não é o mesmo. A substituição da preposição presente nessa sentença por uma outra, como *para*, clarifica a questão e deixa clara a impossibilidade:

- 2c) É certo que ele correu **para ver** de onde vinha o barulho.

Nesse caso, a preposição é uma marca de finalidade, indicando o intuito da ação anterior, isso é, a razão pela qual o sujeito correu. Do mesmo modo, a substituição do complexo a + infinitivo de 3 por um gerúndio faz com que o sentido original da sentença seja perdido. Ao mesmo tempo, não serão todos os casos de gerúndio que se poderão tomar como equivalentes a um IP. Observe-se:

4) Estou **morrendo/a morrer**

5) Quem **rindo/a rir** bate, **rindo/a rir** morre

6) Ele **vindo/a vir**, eu saio

Se em 4 a troca do gerúndio por um IP pouco muda na interpretação da frase, o mesmo não valeria para 6 onde há uma restrição na leitura. A interpretação inicial de 6 como uma construção condicional equivalente a “se ele vier” não licencia o uso de um IP. A única leitura possível para essa estrutura como um IP seria relativamente estranha com uma interpretação como “Ele estando vindo”.

Travaglia (2008) aponta que “o aspecto expresso pelo gerúndio parece depender do tipo de oração desenvolvida, a que a reduzida do gerúndio corresponde e do tempo flexional em que está o verbo principal”. Haveria diversos casos onde, dependendo da combinação de fatores, o gerúndio pode ou não apresentar uma leitura progressiva, sendo esses os casos onde poderia ser considerado sistematicamente equivalente aos usos de IP. Complicando ainda mais a situação, há leituras não óbvias do gerúndio, casos em que uma interpretação progressiva poderia ser feita junto a uma outra leitura terminada. Veja:

7) **Chegando** em casa, tirei os sapatos.

Em 7, embora uma leitura progressiva seja possível, essa concorre com uma segunda leitura onde o gerúndio marcaria o aspecto acabado e seria, portanto, equivalente a uma perífrase verbal gerúndio + particípio (“**tendo chegado** em casa”). Em resumo, o espectro

significativo do gerúndio é consideravelmente amplo, apenas alguns trechos dele sendo equivalentes ao IP.

Todavia, todas essas interpretações são feitas com base no português contemporâneo, ao qual a pesquisa atual não se refere. Tratar de casos como esses para o português oitocentista traz à frente da questão as possibilidades de interpretação específicas do período. Não é segredo ou novidade que um dos grandes problemas no tratamento de questões morfossintáticas em visão diacrônica é a disparidade de intuições entre o analista e os falantes do período. Para minimizar esse problema, certas precauções devem ser tomadas. Mothé (2006) optou por elencar uma classe de ocorrências que chamou de ambíguas, isso é, casos de IP ou gerúndio que poderiam ou não ser equivalentes à variante oposta (e, portanto, relevantes para sua pesquisa). Essa compreende aproximadamente metade dos seus dados totais. Tal medida não será seguida aqui, pois adiciona um patamar de desconfiança dificilmente desejável para a análise, ambiguidade essa que poderia danificar o grau de confiabilidade de todos os resultados obtidos. Ainda assim, uma medida deve ser tomada e essa deve atender aos seguintes critérios:

- I) **Critério de qualidade** - reduzir as ocorrências de ambiguidade
- II) **Critério de representatividade** - reter o espectro aspectual da variável
- III) **Critério da significação** - reter a relevância sociolinguística da variável
- IV) **Critério da quantidade** - Permitir a obtenção de um número significativo de dados

Em detalhe, o critério I já foi explicado. A presença de um número excessivo de casos ambíguos reduz a segurança de qualquer modelo estatísticos que se possa construir, assim como de qualquer análise que se faça em cima desse. Embora casos ambíguos possam ser de interesse como pontos de entrada para variação, o risco trazido por confiar profundamente na intuição linguística do pesquisador adiciona um contrapeso inaceitável a esse ganho potencial. O critério II se presta a interesses linguísticos. O espectro aspectual do gerúndio é amplo e embora discutir esse fato não seja de muito interesse para a presente pesquisa (de fato, esse não será um dos preditores nos modelos aqui usados), é desejável que esse seja representado tanto quanto possível para que não se perca nuances possíveis do fenômeno. O critério III é talvez o mais simples, representando tão e tão somente uma restrição de segurança para evitar sobrelimitação.

Se as medidas tomadas para redução das ambiguidades diminuïrem a variável a casos excessivamente limitados, se perde a segurança de que essa seja relevante, uma vez que se reduz a variável a uma mera fração da amplitude total. Variação linguística é um fenômeno complexo e multidimensional e toda restrição gera perda de significância – reduzir a variável em excesso simplesmente faz com que se perca parte de seu significado. O critério IV, por fim, é autoexplicativo – uma pesquisa será apenas significativa se obtiver um número mínimo de ocorrências para que se possa realizar análises inferenciais. Poucos dados resultam em modelos pouco confiáveis, embora o que constitua “um número significativo de dados” possa ser posto em questão. Ainda assim, como regra geral, mais dados são mais desejáveis do que menos dados, uma vez respeitados todos os outros critérios.

Em vista disso, a medida mais adequada para garantia dos quatro critérios acima descritos foi a delimitação da variável não a todas as ocorrências de IP e gerúndio, mas tão e tão somente às ocorrências desses com o auxiliar *está*, isso é, a perífrases *estar + IP* e *estar + gerúndio* (daqui em diante, eIP e eGER). Mothé (2007) demonstra uma taxa maior de uso IP em construções perifrásticas com auxiliares verbais, sendo os principais desses *estar* e *continuar*. O auxiliar *continuar* oferece um problema interessante, uma vez que se poderia argumentar que ele por si só carregaria, em algum nível, tons de progressividade. A ver:

8) Eu continuo **tendo/a ter** que correr com minha dissertação

9) João continua **morrendo/a morrer** de dor.

10) Você continua **dando/a dar** exemplos estranhos

Ainda que se aceite o fato de que o aspecto em português é fundamentalmente composicional, aceitar os casos com o auxiliar *continuar* pode mascarar casos que de outro modo não seriam considerados funcionalmente equivalentes. Ademais, as perífrases de *estar* contam com uma maior transferência de material dentro do espectro aspectual. Travaglia (1981) aponta que as perífrases de *estar* “marcam os aspectos **imperfectivo, cursivo, não acabado e durativo** [...] exceto nos pretéritos perfeito e mais-que-perfeito do indicativo que expressam os aspectos **perfectivo e durativo**” (TRAVAGLIA, 1981, p. 192), sendo possível realizar uma substituição do gerúndio por um IP nos quatro primeiros casos, como visível nos exemplos a seguir:

- 11a) Ken **estava lutando** naquele dia (imperfectivo)
- 11b) Ken **estava a lutar** naquele dia
- 12a) **Estou jogando** há horas (cursivo)
- 12b) **Estou a jogar** há horas
- 13a) **Estou saindo** com Monika (não acabado)
- 13b) **Estou a sair** com Monika
- 14a) Mário **está tentando** entrar no cano há muito tempo (durativo)
- 14b) Mário **está a tentar** entrar no cano há muito tempo

Tais possibilidades atestam o alto grau de equivalência funcional e o estatuto de ambas como variantes de uma mesma variável, uma vez que virtualmente toda a abrangência semântica de eGER poderia ser reproduzida por vias eIP. Esse grau de paridade entre as duas formas permite, portanto, que se realizasse de modo mais adequado e seguro uma comparação entre elas, uma vez que, não havendo, *a priori*, uma divergência de ordem linguística, a opção por uma ou outra seria convencional e estilística. É esse fato que nos permite, portanto, selecionar essa variável para os fins de uma análise de cunho mais social.

Em resumo, a variação entre eGER e eIP é um fato bastante antigo na língua portuguesa e as duas formas possuem uma abrangência conceitual consideravelmente alta que lhes permite serem usadas uma pela outra em diversos contextos. A existência dessa variação, porém, não indica como essa foi utilizada (ou se sequer o foi) pelos falantes de então. É disso que trata o capítulo seguinte, assim como de outros aspectos particulares da questão social dessa pesquisa.

3: EXPOSIÇÃO HISTÓRICA E HISTORIOGRÁFICA

Onde é feita a imersão teórica no universo conceitual do século XIX a partir de uma abordagem de três vias: uma histórica, uma historiográfica e uma metalinguística.

3.1: A HISTÓRIA E A HISTORIOGRAFIA DO BRASIL

Como definido na sessão anterior, o presente trabalho tem por base os pressupostos teóricos da sociolinguística variacionista em sua vertente histórica, como estabelecido por Suzane Romaine, sendo seus interesses voltados para a averiguação de aspectos do uso da língua num determinado contexto histórico e como esses se ligariam a processos sociais maiores. Para esta pesquisa em particular, o interesse central é averiguar as relações possíveis entre identidade linguística e uso linguístico no tocante ao período referente ao Império do Brasil.

Sendo a presente pesquisa dedicada a um período histórico não concorrente com sua execução, certas dificuldades surgem, das quais uma se destaca: como analisar uma população a cujas opiniões e sistemas de valor não se tem acesso? Do ponto de vista de uma pesquisa sociolinguística que se delimitasse aos interesses da chamada “primeira onda” (ECKERT, 2016), tal pergunta não se colocaria. Todavia, uma vez que o questionamento se dê em torno de algo que se poderia chamar de “identidade”, a compreensão do ambiente sociocultural onde os falantes existiam se torna parte necessária da pesquisa, pois é o que permite penetrar no sistema de significação social desses e compreender, portanto, a língua por eles falada. Em outras palavras, sendo a identidade um resultado de interações sociais, é necessário compreender as relações internas existentes no grupo de interesse, a avaliação de suas opiniões quanto a si mesmo e os outros que o cercam e no modo como essas se constituem em contextos variados ou ao menos nos contextos relevantes para a investigação que aqui se pretende fazer. Se tal exercício etnográfico exige do pesquisador de grupos contemporâneos um esforço notável e um grande rigor metodológico, àquele que decide se curvar sobre o passado recaem dificuldades ainda maiores: tudo que lhe é possível acessar, quando muito, são registros limitados e invariavelmente turvos de uma comunidade já extinta ou profundamente modificada pelas forças do tempo. A incapacidade de se imergir no sistema significativo de uma época se não através de resquícios deixados por determinados indivíduos limita muito quantidade de

dados acessíveis. É então necessário ao pesquisador lançar mão de todos os recursos que lhe estiverem disponíveis para tentar reconstruir o cenário social, cultural e linguístico do período que pretende estudar, refazendo o passado a partir das pistas deixadas por aqueles que viveram outrora. Todo trabalho histórico é, portanto, um trabalho de reconstrução dedutiva. Por vezes, acessar esse passado remoto é algo virtualmente impossível, como quando lidando se lidando com culturas ágrafas ou extintas há muito tempo. Felizmente, esse não é o caso do Brasil. Formado a partir de um modelo europeu de civilização, o Brasil se vê no século XIX em meio a um processo “civilizatório”, buscando a construção de um universo cultural próprio e de uma história nacional. É nesse contexto que se verão publicadas diversas histórias do Brasil, assim como um largo número de opúsculos, memórias, livros e artigos acerca do povo e civilização tropical e a partir desses se poderia reconstruir o ambiente cultural e político que constituía a nação.

Naturalmente, não se pretende aqui que o material supracitado seja a representação fac-símile de todo o universo cultural brasileiro. Esse jamais poderia ser o caso, seja para o Brasil ou outro país, no século XIX, XX ou no presente. Todavia, no contexto de uma nação recém-independente e ansiosa por reconhecimento de sua validade como uma entidade política, a produção cultural ganha realces de força e ainda que não seja imagem dos anseios e desejos da população comum, carrega em si os princípios segundo os quais a intelectualidade do período imaginava a nação. Será no Império, afinal, que se fundarão as bases do sistema educacional brasileiro, além de diversas instituições dedicadas a construir a história e traçar os rumos do país.

Para tentar fundamentar a pesquisa aqui pretendida, se tomará, em primeiro lugar, a historiografia como discutida por autores do período, assim como aspectos históricos que levariam a essas historiografias. Se propõe, portanto, uma investigação de mão dupla, se analisando o que era dito e tomado como história no XVIII e no XIX e o que a história sociopolítica como compreendida hoje pode revelar sobre a origem das visões passadas. A história é tomada como ponto de partida em razão de sua importância como instrumento político, especialmente no século XIX, quando o interesse sobre essa passa a vir em razão de um determinado projeto de governo. Fontana (1998) aponta que o interesse histórico do 1800 virá a serviço da contrarrevolução, confiando ao conceito de “nação” o papel antes ocupado pelo direito natural e pelo poder divino atribuído ao Rei e, por extensão, às instituições de governo, o papel de manter a coesão do estado e suas relações de poder. Nesse contexto, os novos fundamentos de governo se acharão “no fortalecimento da ideia de nação, entendida como a

comunidade dos homens que compartilham a mesma história e cultura [...] e encarnada na pessoa do monarca constitucional” (FONTANA, 1998, p. 117). Tal busca por uma nação e seus símbolos levará, no XIX, ao estabelecimento dos movimentos românticos europeus, os quais buscarão na recuperação do passado medieval e no culto aos marcos geográficos do território símbolos centrais da nação (*ibid.*). No Brasil, contudo, tais elementos se provarão de mais difícil utilização. Tendo sido colonizado apenas no XVI, portanto já depois do medievo mítico, o Brasil não contará com um conjunto de mitos fundadores de suas fronteiras – não haverá no Brasil um Rei Artur ou um Carlos Magno, figuras que sintetizam em si determinados ideais de nação. Serão, assim, a narrativa do descobrimento, como no “Caramuru”, de Santa Rita Durão, e no “Vila Rica”, de Cláudio Manoel da Costa, e a figura do índio, como no “O Guarani”, de Alencar, ou no “Y-Juca Pirama”, de Gonçalves Dias, a servir ao propósito de fundamentar o passado histórico da nação. A questão indígena, e em menor escala questão a negra, serão elementos centrais da formação da identidade brasileira, sendo arrogada pela intelectualidade do período e tendo um espaço considerável no pensamento da época. As teses de miscigenação, como propunha von Martius (1845), formarão um dos pilares da imagem do Brasil, mas não sem problemas particulares. No âmbito intelectual do século XIX, onde a pureza das línguas e dos povos era tomada como elementos organizadores das nações, a nascença híbrida de uma nação é um desafio aos “laços naturais” (ROSSOLILLO, 1998) que dariam forma e razão às nações. Não menos dificultoso é o tamanho colossal do Brasil em contraparte aos estados nacionais europeus. A história, nesse âmbito, ocupará um papel central no estabelecimento de uma identidade única que consiga perpassar não apenas todos os povos que compunham o Brasil como também suas diversas regiões. É uma criação “de cima para baixo” (TAVARES, 2003, p.25).

Tal história geral do Brasil (será esse o título que Varnhagen dará à sua obra) se via obrigada a enfrentar as agitações políticas do XVIII e do XIX pré-1822. Se no século XVIII se tomará certos eventos como oriundos de espíritos avessos à ordem, no XIX, Tiradentes e Tomás Antônio Gonzaga serão vistos como patronos da independência brasileira, a despeito das opiniões e ideias que eles mesmos tivessem. E se por um lado se deverá fazer tais considerações quando se lendo essas histórias a partir de seu momento de produção, quando esse é desconsiderado e os fatos narrados são observados a partir do ponto de vista contemporâneo, outras questões inevitavelmente surgirão. Pois se no XIX se tomará certos eventos como manifestação de uma nacionalidade nascente, como um conflito entre o interesse colonial e algum outro, não se pode tomar por garantido que esse outro fosse de qualquer modo

“nacionalista”, “nativista” ou “brasileiro”, sem que se faça sobre isso algumas ressalvas. Cabe ainda um questionamento similar, porém não de todo igual: ainda que se assumisse que tais eventos fossem de fato nativistas, eles poderiam ser tratados como representações de um único movimento, isso é, de um nativismo *brasileiro*? Seria esse, de fato, o estatuto reivindicado? Silva (1997) coloca tal questão em termos bastante claros:

“[...] se tomássemos o período colonial, afirma-se que [Manuel] Beckman, no momento de sua execução, teria proclamado ‘Morro pelo Maranhão’ – representaria isso ‘brasilidade’ ou ‘maranhanidade’? O que dizer então das inconfidências – seriam afirmações ‘nativistas’ ou ‘regionalistas’? ou talvez ‘separatistas’? [...] os Pernambucanos de 1817, 1824 seriam nativistas ou separatistas? ‘Pernambucanidade’ ou ‘brasilidade’?” (SILVA, 1997, p. 87)

Em vista disso, se veria necessária uma reanálise dos conceitos identitários então vigentes, tanto quanto são esses de máximo interesse à pesquisa aqui conceituada. Aquilo que é compreendido como nação ou pátria, termos nem sempre intercambiáveis, constaria como tal para aqueles por trás de movimentos revolucionários setecentistas? oitocentistas? É natural que a resposta a essa questão fosse relevante aos interesses presentes, porém sua complexidade em vista da diversidade de tais movimentos impede que se possa levá-la a cabo aqui. Se manterá a pergunta no horizonte de pensamento da corrente pesquisa, porém, para efeitos práticos, se considerará os discursos do período como fundamentalmente válidos, a despeito de dissonâncias que esses possam ter com os eventos aos quais se referem. A análise aqui proposta será, portanto, fundamentalmente *historiográfica* ao invés de *histórica*.

3.1.1: Rocha Pita e a história no setecentos

Michel de Certeau define a história como produzida a partir do século XVI como um instrumento político empregado por aqueles que detém o poder para legitimar suas ações e estabelecer uma *práxis* governamental (CERTEAU, 1982). A história seria, portanto, primordialmente um instrumento discursivo de ordem política utilizado para a justificativa teórica das ações do governo. Dada essa definição de história, se considera essencial a

compreensão do discurso vigente no século XVIII na nação brasileira uma vez que será a essa que a história oitocentista deverá responder.

De princípio, reside nessa proposta uma dificuldade quanto às fontes disponíveis, uma vez que a inexistência de imprensas no Brasil até 1808 e a necessidade de autorizações de órgãos oficiais ligados ao governo português para que se pudesse publicar obras limitavam em muito a exposição de ideias que contrariassem o modelo de história hegemônica. Não somente isso, mas o interesse na história do Brasil, então ainda apenas uma colônia portuguesa, era consideravelmente reduzido, sendo o primeiro texto do tipo a alcançar sucesso considerável a “História da América Portuguesa”, de Sebastião da Rocha Pita, publicado em 1730 sob os auspícios da Academia da História Portuguesa e da Censura Inquisitorial. Em 1849, a Revista do IHGB publicou um artigo memorial acerca da “História da América Portuguesa”, onde destaca a importância histórica da obra de Rocha Pita, em especial por se tratar de uma história do Brasil escrita por um brasileiro: que “O Brazil no entanto carecia de uma historia [...] e maior gloria lhe caberia se fosse essa história escripta por um de seu filho [*sic*]” (Revista do IHGB, 1849).

O ineditismo da obra, em termos da origem de seu produtor, contudo, esbarra nas visões apresentadas pelo texto, em especial àquelas referentes aos movimentos revolucionários que haviam tomado lugar no Brasil, mas também em outros pontos. Uma vez feitos os adequados encômios ao então Rei de Portugal, Dom João V, advertências e prólogos, sua narrativa se inicia descrevendo o estado da coroa Portuguesa no período anterior ao “descobrimento”⁵ do Brasil antes de prosseguir, nos parágrafos seguintes, a uma descrição geográfica do “opulento Império do Brasil” (ROCHA PITA, 1730, p. 5). Ao estabelecer de início as razões da exploração, Rocha Pita enquadra a história do Brasil como uma parte da história maior de Portugal – daí seu título não ser *História do Brasil*, como se verá nas histórias do oitocentos, mas *História da América Portuguesa*. O termo “América Portuguesa”, assim como outro sinônimo usado por Rocha Pita, “nossa América” demonstram a concepção de mundo e identidade do autor e sua compreensão do lugar histórico e social do Brasil; uma colônia, primordialmente, e ele um português, ainda que nascido no Brasil. Não parece existir, em qualquer momento da história como contada por Rocha Pita, uma concepção de uma “nação

⁵ Tal termo é aqui utilizado apenas em viés histórico, não havendo espaço neste trabalho para uma discussão mais longa acerca da validade do uso de tal palavra no contexto da chegada dos exploradores portugueses ao Brasil no XVI.

brasileira”, fato visível no tratamento dos movimentos nativistas e revolucionários ocorridos no território brasílico aos quais ele dá atenção. Esses Rocha Pita tratará sempre em termos negativos, mais como perturbações do que movimentos com qualquer nível de organização. Dirá ele sobre a revolta paulista de 1641:

No caminho encontrou Antonio de Albuquerque aquella **insolente turba**; e querendo persuadir aos mais poderosos della desistissem do impulso [revoltoso], em que **commetiaõ taõ grande offensa contra Deos, e tanto delicto contra El Rey**, lhe deraõ taõ pouca attenção, e mostrara tal porfia, que quando o Governador intentava reprimirlhes com palavras o furor, se vio muy arriscado a experimentado por obras, porque determinavaõ prendello (ROCHA PITA, 1730, p. 380, ênfases adicionadas)

Sucede logo adiante o texto a tratar de outro momento de confronto entre os brasileiros nativos e os colonos portugueses, durante a revolução pernambucana de 1710, à qual dará os seguintes termos como sua “origem e causa”:

São os Pernambucanos, **naturalmente altivos**; não permittiaõ, que no Senado da Camera da Cidade de Olinda entrassem pessoas de outra esfêra, que a da Nobreza daquella Provincia. Achavaõ-se no Recife (porto, e feira de todas as suas Povoações) muitos homens ricos, aos quaes o trato mercantil fizera poderosos, e não podiaõ alcançar os cargos da Governança da Republica, ainda que alguns os chegaraõ depois a conseguir, más com traça tal, e tanto trabalho, que esta dificuldade os obrigou a pertenderem fazer villa aquelle lugar, para lograreм os seus moradores as mesmas dignidades. Representavaselhes fácil a empreza pela opulência do Recife, que em templos, e casas igualava à Cidade de Olinda, e em números de morádores a excedia, porque os estragos, que padecera na guerra dos Holandezes, haviaõ diminuído, e arruinado a sua grandeza (ROCHA PITA, 1730, p. 383, ênfase adicionada).

Sobre a revolta baiana em resposta à instituição do imposto alfandegário de 10% sobre a entrada de fazendas nas províncias brasileiras, dirá:

Tratava Pedro de Vasconcellos de estabelecer este direito na Alfandega da Bahia, quando alterada a mayor parte dos homens de negocio, tendo prevenido ao Juiz do Povo, seus Misteres, e quantidade de plebe apparecerão juntos na Praça de Palacio, na manhã do dia dezanove de Outubro 5 e mandando o Juiz do Povo tocar incessantemente o sino da Cidade, foy concorrendo de varias partes **tanta gente vil**, que em breves instantes se virão cheas a Praça, e as ruas visinhas a ella (ROCHA PITA, 1730, p. 402, ênfase adicionada).

Por fim, acerca da Revolta de Vila Rica, em 1722, diz a história de Rocha Pita que:

Aquelles ânímos orgulhosos, feros, e inimigos do socego, se démorarão alli [na Vila do Carmo] dezaseis dias, com o pretexto de novas duvidas, que se lhes offereciaõ, sendo o fim rebellar aos moradores da Villa do Carmo com muitos projectos, que lhes faziaõ ; e **naõ o podendo conseguir, obrarão taes desordens, que se vio em termos aquelle Paiz de huma grande ruina**, havendo-se o Conde com prudência superior aos seus poucos annos, e com dissimulaçaõ taõ util ao serviço Réal, como conveniente à resoluçaõ, que intentava tomar contra os culpados (ROCHA PITA, 1730, p. 427-428, ênfases adicionadas).

Entre classificações como “gente vil” ou “inimigos do sossego”, Rocha Pita valora todo movimento que visa desafiar a ordem de mundo estabelecida pelo sistema colonial, ainda que esses não tenham necessariamente interesses separatistas, como uma afronta a tudo que há de bom. Os laços que ligariam Brasil e Portugal seriam naturais e qualquer ameaça a eles uma “afronta a Deus e delito contra El Rei”. Sua visão das populações nativas do Brasil também difere muito daquela que se verá esposada pela intelectualidade oitocentista, a qual lançará mão desse elemento como um pilar da cultura brasileira propriamente dita⁶. Tal visão, porém, é profundamente diferente daquela apresentada por Rocha Pita em sua história oitocentista, segundo a qual:

⁶ Cabe lembrar, outra vez, o texto de Martius (1845), assim como os movimentos de cunho indianista que se repetirão no Brasil tanto em seu momento romântico, no XIX, quanto no modernista, já no XX.

Todo este vastíssimo corpo, que temos mostrado estava possuído, e habitado de inculta gentildade, dividida em innumeraveis Nações, **algumas menos feras, mas todas barbaras: não tinhaõ culto de Religião, idolatravaõ à gula, e servião ao appetite, sem regimen de ley, ou de razaõ** ; tinhaõ principaes, a quem davaõ moderada obediencia, que mais era respeito, que sogeiçaõ, repugnantes à doutrina Evangélica, que lhes prérgou o glorioso Apostolo S. Thomé (ROCHA PITA, 1730, p. 33, ênfase adicionada)

Se a história é tomada como narrativa de legitimação de um determinado poder, a visão endossada por Rocha Pita em seu texto demonstra claro favorecimento ao governo monárquico lusitano que então dominava o país. Sua história se inicia, não sem razão, com o estado da coroa portuguesa no período anterior ao descobrimento – constrói assim a história, que não é do Brasil, mas da “nossa portuguesa américa”, como uma história de continuidade em relação a Portugal onde qualquer evento que se volte contra a monarquia é tratado em vias de rebeldia e vileza.

Se os fatos históricos mostram que por um lado começavam a se constituir movimentos nativistas⁷ no Brasil já no XVIII, por outro a história oficial, mesmo aquela escrita por nativos do território brasileiro, ainda é uma onde prepondera uma visão colonial na qual o Brasil é, antes de mais nada, um território vassalo e extensão de Portugal, sua história nada mais do que parte do escopo maior da história portuguesa. Conceitos que viriam a ser, mais tarde, identificados como parte fundamental de uma identidade cultural brasileira, como a máxima do Brasil fruto da união de povos não são aparentes nesse período. A figura do indígena, quando surge, é reduzida a inimigo ou instrumento, enquanto o negro é tomado como se muito ferramenta ou mero fato a-histórico. Certas formas prévias do que viria a ser a interpretação nacionalista da figura indígena existem, de fato, no XVIII, porém limitadas ao âmbito da literatura, como no Caramuru, de Santa Rita Durão, publicado já no final do século⁸.

⁷A questão quanto a quão cabível é o termo “nativista” a alguns dos movimentos revoltosos do XVIII, porém, não é fechada.

⁸ E ainda nesse espaço, seria possível questionar quanto disso representaria uma concepção prévia da ideia de um “Brasil-Nação” e quanto não seria apenas resultado de determinadas normas estilísticas exigidas pelo gênero épico.

3.1.2: O século XIX e o nascimento do Brasil

No século XIX, o processo de independência do Brasil começa a se delinear mais claramente. Em 1808 ocorreu a fuga da família real Portuguesa ao Brasil, resultado da perseguição dela pelas forças Napoleônicas então em expansão pela Europa. Em 1815, o Império Português se tornou o Reino Unido de Portugal, Brasil e Algarves, alteração que concede ao território brasileiro o *status* de Reino dentre as posses ultramarinas e capital do Império Português. Em 1816, se iniciou a Missão Artística ordenada por Dom João VI por meio da qual artistas franceses viriam ao Brasil estabelecer um modelo cultural para a nação. A família real se manteve no Brasil até 1820, quando retornou para Portugal em razão das revoltas que ali ocorriam. Dois anos mais tarde, em 1822, teve lugar, oficialmente, a declaração de independência do Brasil.

O afastamento político definitivo entre colônia e metrópole não se deu como um evento abrupto, mas sim como a conclusão de uma série de eventos anteriores que vinham a aumentar a insatisfação geral da população brasileira quanto à administração portuguesa⁹. Ao longo do século XVIII, após se ver destituído de suas posses asiáticas, Portugal passou para o Brasil a tarefa de sustentar sua economia, aumentando taxações e mantendo maior controle da produção e da administração americana (PRADO JUNIOR, 1970). As tensões resultantes disso levaram a diversos movimentos, especialmente no final do século, contra a regência colonial. Já nos entornos da independência, disso resultou um afastamento profundo entre as duas nações. Sodré (1978) apresenta o seguinte dado acerca das relações econômicas entre Brasil e Portugal:

Em 1805, haviam entrado no porto do Rio de Janeiro, sob regime de monopólio, 810 navios portugueses; o número caíra para 642, no ano seguinte; voltara a ascender, em 1808, chegando a 765. Pois em 1810 entrariam 1214, sem discriminação de bandeira. Essa discriminação mostra, por outro lado, o declínio do comércio português metropolitano: em 1820, chegavam ao Rio de Janeiro apenas 57 barcos lusos (SODRÉ, 1978, p. 139).

⁹ Termo utilizado aqui apenas na mais abrangentes de suas acepções.

Com a quebra dos laços econômicos e, logo em seguida, dos laços políticos, o Brasil se viu efetivamente independente de Portugal em termos governamentais. Assim se iniciou o processo de construção da matriz cultural brasileira por vias das artes e da história. Será nesse período que se verá fundado, sob a égide do Conselho Administrativo da Sociedade Auxiliadora da Indústria Nacional, o Instituto Histórico e Geográfico Brasileiro (doravante, IHGB), cuja primeira reunião se dará no Museu Nacional, em outubro desse mesmo ano (CALLARI, 2001). Seus membros eram “elementos oriundos da burocracia estatal, logo comprometidos com a ordem que representavam” (*ibid.*). Será por meio do IHGB que se verão publicadas no período medial do XIX diversas histórias e corografias, tanto do Brasil quanto de suas províncias.

A fundação do IHGB foi um marco no desenvolvimento da cultura letrada brasileira, sendo o Instituto um dos principais responsáveis pela formação da história oficial do Império. Já nas primeiras edições da Revista do IHGB (doravante, RIHGB) se verão publicadas memórias de brasileiros ilustres, textos sobre a escravatura e, mais notavelmente, o programa histórico do instituto, na qual ele é declarado “o representante das ideias de ilustração que em diferentes épocas se manifestaram em nosso continente”, título dado pelo visconde de São Leopoldo ao artigo que abre o segundo número da RIHGB. Nesse se lê uma admoestação do governo colonial, tomado por opressor de ideias e temeroso da perda do controle da nação, algo que se vê repetido em diversos discursos pelo longo do período Imperial, ao mesmo tempo em que crescem os elogios históricos a figuras anteriormente rejeitadas pela história oficial. Nessas linhas, se encaixa “Os Mártires Pernambucanos”, do Padre Dias Martins (identificado por si mesmo como um “luso-pernambucano”), uma coletânea de biografias acerca dos indivíduos envolvidos nas revoluções pernambucasas de 1710 e 1817. O texto, em sua essência, se pretende a revisitar a história das revoltas acima citadas, reinterpretando sua natureza e qualidade, assim como a de seus participantes, como nobres e patrióticas, visão profundamente dissonante daquela exibida por Rocha Pita e a história colonial.

Percebe-se ainda, nesse período, a ampla discussão acerca do que comporia uma história e literatura legitimamente brasileiras, ideias que serão caras ao projeto do IHGB, e que são claramente colocadas por Abreu e Lima no prefácio de seu “Compêndio da História do Brasil”, de 1843:

Não ha muito eu pela imprensa se negou a naturalidade de um dos nossos melhores poetas, só porque em suas poesias fallava do Tejo e do Douro, de

pastores e cajados, da faia e da oliveira, em vez do Parahyba e do Rio Doce, do laço e das bolas, da jaboticada e da mangabeira. Isto [...] **é a meu ver forte argumento para provar, que aquella parte da literatura portuguesa, obra de escriptores Brasileiros, pertence-lhe exclusivamente, porque tem todo o cunho de sua nacionalidade.** [...] Sem os estorvos que o zelo indiscreto dos Portuguezes nos punha sempre por diante, para impedir o rápido vôo de nossa inteligência, **devemos cuidar de reparar o tempo perdido, dando principio á obra da nossa regeneração intelectual, e preparando os elementos de uma literatura propriamente brasileira** (ABREU E LIMA, 1843, ênfases adicionadas).

É seguindo essa mesma razão histórica e buscando essa mesma construção de uma identidade “propriamente brasileira” que será publicada pela RIHGB, em 1845, a tese de Martius acerca da escritura da história do Brasil. Tomado aqui como marco teleológico de um processo iniciado muito antes, o texto de von Martius se insere, portanto, num ambiente social, político e histórico onde a discussão e a reivindicação da história ocupam lugar central num projeto iniciado décadas antes. Sua proposta do Brasil como uma nação formada pelo encontro de três povos vem a cabo da criação de marcos identitários que se possam facilmente utilizar para distinguir a história brasileira da portuguesa. O povo brasileiro é, portanto, não mais português, mas fruto da miscigenação, e por isso alheio às reivindicações e poderes lusitanos. Porém, sendo Portugal a nação a trazer ao país a lei e o rei e sendo dotado do poder civilizatório e moral, caberia a seus descendentes o direito de conduzir o destino da nação. Essa é a tese central apresentada por Martius em seu ensaio. Não obstante, ao escrever os meios de sua história, Martius não pretende esconder sua aliança ao projeto Imperial, dedicando parte de seu texto a tratar, diretamente, da questão dos movimentos liberais, então plenamente em voga nos círculos de discussão política:

O Brasil está afeto em muitos membros de sua população de **idéias políticas imaturas**. Ali vemos republicanos de todas as cores, ideólogos de todas as qualidades. É justamente entre estes que se acharão muitas pessoas que estudarão com interesse uma história de seu país natal; **para eles, pois, deverá ser calculado o livro, para convencê-los por uma maneira desta da inexequibilidade de seus projetos utópicos, da inconveniência de discussões licenciosas dos negócios públicos, por uma imprensa desenfreada, da necessidade de uma Monarquia em um país onde há um tão grande**

número de escravos¹⁰. Só agora principia o Brasil a sentir-se como um todo unido. Ainda reinam muitos preconceitos entre as diversas províncias; estes devem ser aniquilados por meio de uma instrução judiciosa; cada uma das partes do Império deve tornar-se cara às outras, deve procurar-se provar que o Brasil, país tão vasto e rico em fontes variadíssimas de ventura e prosperidade civil, alcançará o seu mais favorável desenvolvimento, se chegar, firmes os seus habitantes na sustentação da Monarquia, a estabelecer, por uma sábia organização entre todas as províncias, relações recíprocas (MARTIUS, 1845, p. 54, ênfases adicionadas).

A seleção desse texto como vencedor de um concurso para uma teoria da historiografia brasileira revela, por si só, o posicionamento político do IHGB e a presença do próprio Imperador em reuniões do órgão reforça os laços entre o Instituto e o governo vigente. De fato, o IHGB constituía um papel de continuidade com as ideias fundamentalmente eurocêntricas adotadas pelo governo lusitano. Ademais, suas propostas teóricas se colocam contra os movimentos republicanos do período que veem a regência como um governo inválido para a recém-nascida nação brasileira (SODRÉ, 1978). A criação do IHGB vem como parte de um projeto amplo do governo imperial de legitimação de seu controle sobre o território brasileiro, assim como da criação de uma identidade nacional mais facilmente utilizável para a promoção das políticas pretendidas pelo Império (SILVA, 2015), tendo a função de “construir uma história da nação, recriar um passado, solidificar mitos de fundação, ordenar fatos buscando homogeneidade em personagens e eventos até então dispersos” (SCHWARCZ, 1993, p. 101). Consta no IHGB uma prevalência da discussão histórica, geográfica e biográfica, conquanto se pautará suas considerações antropológicas num profundo darwinismo social onde o negro ocupava o papel mais baixo, o indígena uma posição ambígua, mas redimível, e o branco o lugar de civilizador (SCHWARCZ, 1993). Não por menos, Martius dá a cada um desses grupos o espaço que dá em seu ensaio: cabem aos portugueses brancos nove páginas, aos índios cinco e aos negros duas.

Ideias do tipo, onde a natureza racial de um povo e as qualidades do clima de um país ocupam um papel determinístico na formação de uma nação são preponderantes nos textos

¹⁰ Tal menção ao número de escravos presente no Brasil parece remeter ao medo constante nos países escravistas de que se vissem nessas situações similares às ocorridas no Haiti durante sua própria revolução. Para uma discussão acerca desse tema, ver Nascimento (2008).

publicados pela revista e constituirão uma parte significativa do campo imagético do movimento romântico que lhe é contemporâneo. Tal natureza é melhor sintetizada na *Canção do Exílio* de Gonçalves Dias, cuja exaltação da natureza brasílica não se distingue daquela feita pelos textos de viajantes europeus que conheceram o Brasil no XIX, dentre os quais constam Ferdinand Denis, Maria Graham, James Henderson e o próprio Martius. Terão os estrangeiros, aliás, papel imenso na construção do romantismo brasileiro, tendo sido convidados por Dom João VI para estabelecer uma escola neoclássica que se deveria se opor à estética barroca do XVIII – é a chamada Missão Artística Francesa¹¹.

Nesse sentido, temos a história como escrita por Varnhagen, no XIX, que se prestará a um claro projeto nacional alinhado às intenções imperiais. Sua “História Geral do Brasil”, publicada originalmente em dois volumes em 1854 e 1857, ainda que narre as condições que levaram Portugal a se projetar aos mares para a conquista ultramarina que levaria à descoberta do Brasil, não toma esses fatos como ponto de partida de sua versão da história. Seu primeiro capítulo se dedica a uma descrição geográfica do Brasil: descreve sua extensão, sua topografia geral e as árvores, plantas e animais encontrados no território. Os capítulos seguintes se dedicarão à apresentação dos povos indígenas do país antes de prosseguir, enfim, na sessão V, à história que levaria à chegada dos Portugueses ao Brasil. Por longo, sua narrativa se dedicará às questões indígenas, seja na descrição dos povos nativos do território brasileiro ou dos conflitos entre esses e os colonizadores. E embora esses sejam em muito representativos do que se supõe aqui como uma historiografia romântica, a parte de maior interesse para a discussão aqui presente seja a visão representada por Varnhagen quanto às revoltas e tentativas de revolução ocorridas durante o XVIII às quais dedicará grande atenção.

Dentre essas, a que recebe maior atenção é a ocorrida em Minas Gerais, em 1789, e usualmente chamada “Conjuração Mineira”, a qual Varnhagen tratará em diversos momentos como uma prévia da Independência brasileira. Sua narrativa, de fato, chega ao ponto de naturalizar os eventos de 1822 uma vez que, discutindo a culpabilidade de Manuel Antônio Gonzaga e Luís da Silva, considera que “toda a culpa, se havia, se reduziria a serem ambos muito ilustrados, verem claro o que se passava no mundo, e preverem os sucessos que, segundo a ordem natural, tinham de acontecer um dia” (VARNHAGEN, 1956, p. 319). A tentativa de

¹¹ O movimento modernista e até a fundação da Universidade de São Paulo terão também grande influência da intelectualidade francesa.

revolta, em si, é tratada em termos de um fracasso inevitável, ainda que bem-intencionado. Concluirá ele nos seguintes termos a sua narrativa da tentativa de revolução:

“Curvemos a cabeça ao decreto da Providência, que, à custa do próprio sangue **dos mártires do patriotismo**, veio a conduzir-nos à única situação, em que podemos, sem novos ensaios, procurar ser felizes, **e fazer-nos respeitar como nação**” (VARNHAGEN, 1956, p. 323, ênfases adicionadas).

Já a descrição de Varnhagen da Guerra dos Mascates, conflito ocorrido em Pernambuco entre os comerciantes portugueses de Recife e a velha aristocracia de Olinda, toma claro partido de Olinda, cujas causas considera “verdadeiramente patrióticas” (VARNHAGEN, 1956, p. 320). E se ainda fez ressalvas quanto aos eventos da Inconfidência Mineira, de Mascates Varnhagen não falará que não em termos elogiosos, apontando continuamente os erros nas ações dos mercadores portugueses durante os eventos. Num dos momentos talvez mais reveladores de uma consciência nacional presente, Varnhagen apontará que os moradores de Olinda se viam ofendidos pelo fato de que vinham “aqueles de outras terras mandarem nas suas” (*ibid.*, p. 323). Ainda que tal concepção seja apresentada mais de cem anos após os fatos ocorridos e não possa, portanto, ser tomada como um fato relevante quanto à percepção dos envolvidos na Guerra dos Mascates, ela demonstra, por outro lado, a noção oitocentista de uma distinção Brasil-Portugal forte o suficiente para legitimar a tomada de certos eventos passados como conflito nativistas, ainda que tal conceito não estivesse, necessariamente, em uso naquele contexto.

Parece se configurar como padrão no XIX, portanto, a rejeição do modelo de história vigente no XVIII como parte do processo de dissidência sociocultural entre Brasil e Portugal. Contudo, tal rejeição é feita apenas em termos e a partir de uma determinada série de teorias e modelos estéticos primariamente importados. A historiografia oficial se prestará, durante o Império, à constituição da unidade nacional visando criar elos naturais entre as diversas regiões ao mesmo tempo em que proclama uma “alma brasílica” imanente, nascida dos fatos naturais do Brasil, quaisquer que sejam esses. Revisita-se conflitos passados e se passa a enxergar neles uma prévia da independência e disso se verá o crescente interesse nas histórias regionais. Se constitui, por fim, o conflito entre republicanos e monarquistas, reflexo esse de conflitos anteriores entre nativistas e mantenedores da servidão brasileira a Portugal, fonte esse de uma

parte significativa dos embates que se verão ocorrer no Brasil nesse século e cuja forte repressão foi prioridade para a corte imperial. Se, portanto, no XVIII a história do Brasil era escrita em subserviência à história de Portugal, tal relação já não se pode mais verificar no XIX. Conquanto que seja ainda uma dinastia portuguesa e uma razão eurocêntrica a comandar a historiografia nacional, suas questões mais fortes parecem se voltar agora ao papel do Brasil como um império, que mesmo herdeiro da Lusitânia se diz e se percebe como algo diferente.

3.2: IDENTIDADE E DISCURSO LINGUÍSTICO NO BRASIL

Tendo a história e a historiografia sido tratados na seção anterior, resta agora o terceiro pilar de sustentação: o discurso linguístico. Toma-se como base para a hipótese defendida nesta pesquisa o princípio de que a língua ocupa um papel (particularmente complexo) na constituição daquilo que se poderia compreender como “identidade nacional”, especialmente no contexto da transição operada na passagem do século XVIII para o século XIX entre o modelo de governança monárquico e o estabelecimento das nações como concebidas modernamente. Ainda que não seja de qualquer modo original, a ideia de tal relação não é também em nada óbvia ou direta – há línguas divididas em várias nações tanto quanto há nações divididas em várias línguas – todavia pareceria natural a um indivíduo moderno, membro de uma nação formada pelo modelo ocidental assumir que a língua faz parte da identidade cultural de seu país. Que essa conexão entre língua e nacionalidade, ou de outra forma a “identidade de grupo”, seja importante se poderia ver através dos discursos de movimentos separatistas ao redor do mundo, como no caso da Irlanda em fins do XIX e ao longo do XX, onde uma busca pela renascença gaélica veio junto a desejos de independência política (HUTCHINSON, 1987), assim como a situação Catalã, onde a língua ocupou desde a *renaixença* um papel central na discussão de sua relação com a Espanha (CONVERSI, 2010). Mesmo relatos retirados da narrativa bíblica já demonstravam compreensão metalinguística e a capacidade de estabelecer quem é o outro a partir de sua fala: soldados gileaditas, conta a história, exigiam daqueles que passassem por seus territórios que falassem uma determinada palavra, *shibboleth* (שבולת), e pelo modo de pronunciar julgavam se esses eram aliados ou futuros cadáveres.

Como um elemento de identificação cultural, portanto, a língua detém para si um papel considerável na construção de grupos, sejam esses uma tribo urbana, uma cidade, uma região ou uma nação. Nessa última categoria, a construção de estados-soberanos, o século XIX

apresentará talvez os casos mais explícitos. Durante a formação dos estados modernos europeus, foram dois os modelos principais de transição entre o mundo monárquico e a modernidade republicana: de um lado o francês, cuja abrangência se dará sobre todos aqueles nascidos na República e que a ela desejam pertencer e que, portanto, não se basearia em uma procedência linguística, étnica ou cultural, e do outro o modelo alemão, baseado em pertencimento natural, ao qual o falar de uma língua de base germânica é tão central quanto a herança racial (HOBBSBAWN, 1989 [2013]). E ainda que o modelo francês não tenha se baseado primordialmente na língua como seu critério-fundador “não há dúvidas de que para a maior parte dos Jacobinos um francês que não falasse francês era algo suspeito e de que, na prática, o critério etno-linguístico era geralmente aceito” (*ibid.*, p. 21). Mesmo fora do modelo ocidental, questões linguísticas ocuparão lugar central nas discussões de nacionalidade. Em Hong Kong, território com uma complexa história de domínio estrangeiro e um lugar peculiar no mundo globalizado, a mudança na tradução do nome de uma personagem de uma famosa franquia de jogos, Pokémon, de sua antiga base em cantonês para um de base mandarim levantou protestos por parte de fãs. No campo científico, a nacionalidade ou origem de um falante é, desde Labov, um critério de notável importância e o estudo realizado pelo mesmo em Martha's Vineyard demonstrou a capacidade individual de se mobilizar características linguísticas para se demonstrar alinhamento a uma determinada valorização da ilha e de seu modo de vida (LABOV, 1963).

Antes que se possa avançar mais na discussão, porém, cabe discutir o que se compreende aqui como “língua nacional”, e por extensão como “nação”. O *Dicionário de política*, em sua definição de nação como uma “pessoa coletiva” aponta traços comuns dos indivíduos que a comporiam, dentre os quais a língua (ROSSOLILLO, 1998). Mais interessante é sua definição da nação como “ideologia de estado” onde se encontrará o seguinte exemplo:

O caráter ideológico da Nação explica também as mudanças de enfoques — nas diferentes situações histórico-políticas — com que são abordados seus diversos e contraditórios conteúdos representativos. Por ser a ideologia de um Estado, precisará adaptar-se no seu conteúdo às diferentes exigências da razão de Estado. Por isto, quando a Alsácia era objeto de disputa entre a França e a Alemanha, a Nação era, para os franceses, o grupo dos que “querem viver juntos”, enquanto era definida, pelos alemães, com base na comunhão de língua e de costumes; assim, antes da Primeira Guerra Mundial, Trento e Trieste eram italianas porque seus habitantes eram de língua italiana, enquanto a partir do

término da Primeira Guerra Mundial, o Tirol do Sul é italiano porque se acha situado dentro dos "limites naturais" da Itália. (*ibid.*, p. 797).

Do exemplo citado por Rossolillo, retira-se o aspecto primordialmente funcional da língua na formação dos estados modernos – sua utilização como instrumento de identificação e mobilização política e como justificativa de autoridade. Não por coincidência, o método comparativo, modelo de análise científica cuja metodologia e teoria se sustentam na comprovação de laços ancestrais entre idiomas, será o modelo de análise linguística escolhido pela intelectualidade germânica por boa parte do XIX.

Para além disso, uma “língua nacional” pode apenas ser compreendida à luz da existência de uma “nação” que a fale – assim, o português apenas pode ser considerado a língua nacional de Portugal uma vez que exista um estado Português. A relação entre “língua” e “nação” é, portanto, ambígua uma vez que em certos contextos uma língua será usada como argumento para a criação de uma nação, conquanto em outros o oposto se dará. E se complicadores diversos são visíveis na Europa, o caso de nações surgidas a partir de antigas colônias apresenta problemas adicionais, ainda mais se considerando certas ideias tidas quanto à língua dentro da ciência do período. Em seu cerne, tais complicadores nascem de duas questões que se poderiam formular do seguinte modo: “pode um povo se distinguir de outro, em seu pensamento e sua cultura, usando a mesma língua que esse?” e “pode uma língua mudar sem decair?”.

A primeira dessas perguntas passa a ser elaborada no âmbito da ciência linguística oitocentista, quando surgem as proposições quanto à conexão da linguagem e as ideias de quem a fala. Isso é, que a relação entre algo que se conceitua como “língua” e outra coisa que se chamará de “espírito”. Nesse ambiente, se viram ser publicadas declarações como a seguinte:

Há, portanto, um elemento de delimitação no aprendizado de língua, mas é um elemento do qual o aprendiz é completamente inconsciente. Qualquer que seja a língua aprendida primeiramente por ele, essa é para ele o modo natural e necessário de pensar e falar; ele não concebe nenhum outro como sequer possível [...]. Nós, é claro, olhando de fora, podemos às vezes achar razões para nos entristecer, dizendo: “Eis aqui um homem de capacidades muito além da média da comunidade degradada à qual ele pertence, tivesse apenas ele, em justiça a essas capacidades, nascido onde uma língua mais elevada o tivesse desenvolvido no que ele poderia ser... (WHITNEY, 1887, p. 22)

Whitney, autor desse trecho, será um dos nomes de maior influência na produção linguística do oitocentos tardio no Brasil. Dentro de tal ambiente intelectual, portanto, falar a língua de seu colonizador é assumir o seu modo de ver e pensar, internalizar suas relações e, no mais extremo, o ser. Daí um certo crítico português considerar Gonçalves Dias um poeta português e uma enciclopédia alemã tratar Mark Twain como um humorista inglês (HABERLY, 1973) e João Ribeiro dizer que galicismos são “resultado da educação do povo pelo espírito francês” (RIBEIRO, 1889, p. 301).

Guardado no trecho de Whitney acima citado se encontra também a chave para que se compreenda a segunda pergunta proposta, isso é, se é possível haver mudança sem degeneração. Pois se uma língua guarda em si maior ou menor possibilidade de desenvolver as ideias, ou o espírito, daqueles que a falam, o desejável seria invariavelmente que se falasse uma língua elevada. Todavia, reside o demônio metafórico entre uma necessidade e um medo: a necessidade de se afirmar a identidade própria nação por meio de sua língua e o medo de que a afirmação dessa identidade por vias de sua diferença implicasse em atrofiamento de suas qualidades. Naturalmente, essa não é uma questão que se põe para um estado independente e antigo, mas sim para um que visa se desvencilhar de antigas relações de poder. Tal é o caso das colônias independentes.

Se nas revoluções nacionais realizadas na Europa a língua operou ora como elemento unificador de determinadas populações, ora como excludente de certos grupos, sua função no âmbito das revoluções nacionais americanas não poderia de modo algum ser igual a essa. Se caberia na Europa o apelo à ancestralidade da língua e sua ligação fundamental com os povos que as falam e os territórios onde são faladas, no ambiente americano tal apelo esbarraria em problemas diversos. Como seria possível que um estado independente, ao cabo de sua liberação política, se mantivesse empregando a língua dos seus antigos dominadores? Esse problema foi discutido de formas mais ou menos diversas por todo o continente americano, porém ao fim do processo de independência das colônias não se rejeitará na América Inglesa o uso do inglês, ou na América Espanhola o uso do castelhano nem tampouco na América Portuguesa o do português. Pois que sejam as línguas coloniais um laço contínuo entre as nações libertas e seus antigos captores, essas são ainda línguas europeias, faladas havia séculos nesses países, símbolos da cultura ocidental e desenvolvida à qual nenhuma alternativa realista existe, ao menos de acordo com um certo ideal de mundo. Rejeitar as línguas coloniais a favor de línguas

nativas, por exemplo, seria rejeitar a civilização e abraçar a “barbárie”. Assim, o que se tem nas nações do Novo Mundo é uma questão complexa, onde se vê em diversos campos do pensamento oitocentista a ânsia de criação de uma nação comparável às europeias, dotada de meios e instituições próprias, com uma história e uma literatura únicas e distinguíveis daquelas dos antigos colonizadores, porém onde a língua oficial não se altera de princípio.

Uma vez que a presente pesquisa se baseia na existência de uma distinção percebida pelos falantes oitocentistas entre algo que se conceitualizava como “português brasileiro”, “português do Brasil”, “Língua Nacional” ou um de muitos outros nomes dado a esse objeto, e outra coisa compreensível como “português europeu” ou “português de Portugal”, é necessário que se defina o que tais termos significariam naquele momento histórico. Se faz necessário, portanto, uma investigação das considerações feitas no XIX por aqueles intelectuais dedicados à questão da língua, que se não exaustiva, como jamais poderia ser dada a amplitude do período e do problema, deve ser ao menos exemplar.

Limita-se os interesses aqui apenas ao oitocentos tardio dado alguns fatores históricos (e historiográficos) particulares, o primeiro dos quais é a inexistência de uma imprensa formal no Brasil antes de 1808 – uma proibição real impediria a formação de imprensa na Colônia até então, ano em que a família real Portuguesa chega ao Brasil fugindo das forças napoleônicas. Ademais, as discussões linguísticas por boa parte do oitocentos se darão ainda dentro do modelo da Gramática Filosófica, cujos interesses particulares não se prestariam a uma discussão nacionalista, sendo apenas a partir do último quartel do século que se introduzem no Brasil estudos inspirados pela linguística comparada, mais afeita às discussões nacionalistas (FÁVERO & MOLINA, 2006).

Mais do que isso, a discussão acerca do português brasileiro no oitocentos tardio gira em muito ao redor de uma idealização da língua em função de um projeto educacional que se iniciava. Esse aspecto é amplamente explorado por Mattos e Silva (2004), segundo quem o ideal central da língua brasileira defendido pelo governo tem historicamente sido um de homogeneidade materializada não apenas na supremacia do português como língua de cultura como também em uma idealização do próprio português brasileiro como algo único e invariável. Em menor ou maior grau, esse mesmo discurso pode ser visto na fala de vários autores que estabelecem uma distinção entre um português verdadeiro e puro e um outro português “miscigenado”. Como se verá adiante, essas ideias se correlacionam fortemente com um ideário

geral do oitocentos que se prolongou até o século XX e cujos conceitos guiaram as ideias higienistas do projeto civilizatório brasileiro.

Tendo tudo isso dito, cabe olhar para alguns eventos importantes.

Data oficialmente de 1822 a independência brasileira. Em 1824 será outorgada a primeira constituição do Império. Em 1827 a Faculdade de Direito de São Paulo é fundada, em 1837 o colégio Pedro II e em 1838 o Instituto Histórico Geográfico Brasileiro. Começa assim o processo de construção da intelectualidade brasileira propriamente dita. Assim, a formação de intelectuais e autores dentro do paradigma social brasileiro e não mais luso-brasileira se dará apenas a partir da década de 1840, sendo os autores anteriores formados ainda sob a égide colonial. Não por menos, até 1865 serão publicadas obras gramaticais em português modeladas a partir da Gramática Filosófica e só em 1881, com a *Grammatica Portuguesa* de Júlio Ribeiro se fundará no Brasil a tradição comparada.

Nesse período inicial, anterior à *Grammatica Portuguesa*, as questões específicas acerca da variante brasileira do português serão pouco discutidas. Sotero dos Reis será aquele que mais espaço dedicará a apontar especificidades do português brasileiro e pontos onde essa variante divergiria da europeia. Seus apontamentos, porém, serão ainda poucos e restritos e a própria essência da tradição filosófica, centrada na noção do “gênio imutável”, pouco se prestaria a essas questões. A mudança, de fato, pouco existe no imaginário dos autores dessa escola, tanto sendo que se poderá ver dito, ainda em 1837, que o português não seria uma língua herdeira do latim devido à diferença de sua colocação sintática (SARAIWA, 1837).

Mesmo na *Grammatica Portuguesa* de Júlio Ribeiro, porém, o espaço para uma discussão de cunho centralmente nacionalista será bastante reduzido. Menções ao português do Brasil na obra, seja como dialeto, língua ou qualquer outro valor similar, se limitam a menções anedóticas da língua dos “caipiras de São Paulo” e seus erros e conservadorismos. Nessa mesma gramática, um dos nomes homenageados será o de Camilo Castelo Branco, autor de outro modo aviltado por Baptista Caetano em suas defesas do português brasileiro.

Na *Revista Brasileira*, importante veículo da intelectualidade do país durante o Império, as discussões linguísticas encontrarão seu auge entre 1879 e 1881, durante a fase em que foi editada por Nicolau Midosi, quando foram publicados em suas páginas artigos de João Ribeiro, Paranhos da Silva, Júlio Ribeiro, Baptista Caetano e outros nomes importantes da linguística brasileira. Nesses textos, o pensamento acerca da língua do Brasil será muito mais

fecundo e argumentos de diversas sortes serão feitos sobre o caráter da “língua nacional”, seu estatuto e suas peculiaridades.

Como um todo, será apenas no XIX tardio que o nacionalismo linguístico se manifestará claramente na produção intelectual brasileira. Nesse movimento, o interesse pelas culturas indígenas terá grande força. Dicionários e gramáticas dedicarão espaço a listar o léxico de origem americana inserido no português brasileiro. Na literatura “como atestam, entre outras, as obras de José de Alencar (1829-1877) [...], o léxico foi uma ferramenta essencial para a impressão de brasilidade aos textos artísticos: garantia-se a cor local, sobretudo com o uso de palavras provenientes do tupinambá” (COELHO, 2008, p. 150). Na linguística, diversos autores envolvidos nas polêmicas acerca do português brasileiro demonstrarão interesse pela questão indígena, como João Ribeiro e Macedo Soares, de quem se verão publicados opúsculos sobre filologia e glotologia indígena. Tal interesse tomará formas mais diversas, desde a valorização do papel dessas línguas na criação de um léxico próprio brasileiro ao desejo proclamado de que seus falantes formem, junto aos europeus, uma raça brasileira propriamente dita, como propagam von Martius (1847) e Paranhos da Silva (1881). O pensamento romântico brasileiro desse período advogava mesmo a mescla como fundação de sua individualidade nacional, mitificada por autores do período, dentro os quais se destaca José de Alencar, que trará como conclusão de seu *O Guarani* um nascimento simbólico para o Brasil: após o dilúvio que varre a Terra, restam Peri e Cecília, Adão e Eva do Novo Mundo.

Processo análogo de valorização do indígena como símbolo se dará em outros países da América do Sul, onde se verá publicado um livro inspirado pela gramática comparada onde se pretenderá que o quíchua é uma língua “ariana” e, portanto, que os indígenas peruanos seriam, em algum nível, indo-europeus (LOPEZ, 1871).

O elemento indígena como fator identitário e formador da raça, porém, esbarra no problema da miscigenação, elemento aviltado pelas teorias raciais oitocentistas, para as quais os indivíduos mestiços eram também um problema e “personificavam a ‘degeneração’” (SCHWARCZ, 1957 [1993], p. 44), seja essa a degeneração genética, social ou linguística. A mistura, dirão alguns autores do período, seria desfavorável à índole do português e arriscaria fazer com que a “última flor do lácio”, forçada a se encaixar nas “manias” das línguas indígenas (e africanas) se deteriorasse. Assim, ao mesmo tempo em que serve para formar o caráter próprio do Brasil e de sua língua, a miscigenação era também um risco para o bom funcionamento de ambos.

Em termos da língua, a solução para tal problema foi encontrada na localização da mistura numa parte e da pureza em outra. Assim, seria possível que as importações léxicas, seja de línguas indígenas, seja de línguas africanas, fossem vistas positivamente pelos autores do período posto que seriam índice da nacionalidade brasileira, necessárias para a representação da realidade do novo país (COELHO, 2003) sem que fosse um contrassenso propor que fossem as duas línguas, o português brasileiro e o europeu, ainda uma só. Daí ser possível que se fale, como o fez Silvio Romero, que o português do Brasil não era, em 1880, um dialeto do português europeu, mas a mesma língua que esse à qual apenas se tinham adicionado algumas palavras e expressões (ROMERO, 1880). Se poderia diferenciar, assim, a mistura boa e desejável, aquela que importa termos e imagens da língua dos nativos americanos e permite que o brasileiro ganhe autonomia, daquela baixa e ruim quando à guisa dos bárbaros rebaixa-se a sintaxe da língua pátria. Isso demonstra bem Baptista Caetano Almeida Nogueira (1881), ao se posicionar sobre a língua que defende:

“[...] e eu tenho de defender a linguagem brasileira, acoimada de incorrecta e logo de principio vou topar com a dificuldade penosa de distinguir a linguagem brasileira, que defendo, desse outro falar hybrido, que não é nem protuguez [*sic*], nem brasileiro, desse falar mascavado de francez, de inglez, de africano e de não sei que mais, que predomina na corte e nas cidades chamadas cultas” (ALMEIDA NOGUEIRA, 1881, p. 8, grifo adicionado)

É perceptível nesse trecho uma consciência metalinguística onde se coloca como elementos não apenas distintos como contrários o “verdadeiro português” por ele defendido e “outro português” cuja característica central seria seu hibridismo. Outros autores colocarão a questão em termos diferentes, mas sem escapar a essa mesma razão central quanto à natureza do brasileiro como um verdadeiro português.

Dentre os artigos publicados na *Revista Brasileira* consta uma polêmica entre Arthur Barreiros, Paranhos da Silva e Pinheiro Júnior acerca da colocação pronominal. Nesta, Paranhos da Silva responderá ao artigo de Arthur Barreiros onde esse, mencionando uma anedota sobre um político brasileiro, delineará uma série de regras quanto à colocação do pronome, as mais das quais incidem na recomendação de posposição. Contrário a essa opinião, Paranhos da Silva

opõe o português brasileiro ao europeu, renegando o modelo lusitano e incitando o uso de formas nativas ao Brasil.

Nessa mesma revista Paranhos da Silva publicará, em 1881, um outro artigo em que arguirá sobre a dialeção do português do Brasil, a qual, ele propõe, não existiria de fato uma vez que se compreenda dialeção como afastamento de um vernáculo. Citando Max Muller, Paranhos da Silva defende a tese de que se houve dialeção, essa se deu primordialmente no “hodierno português”, enquanto o brasileiro seria “a língua portuguesa antiga, a língua conhecida como *lengua de las flores*” (PARANHOS DA SILVA, 1881a, p. 282). Prosseguirá ele listando uma série de processos fonéticos do português brasileiro aos quais justifica como remontando ao português trazido à América durante a colonização. Nesse sentido, o português do Brasil seria mais português que o próprio lusitano. O mesmo Paranhos da Silva responderá, em outro artigo, às teses segundo as quais o clima brasileiro seria sua ruína, algo rejeitado completamente por ele tanto quanto a proposta de importação de populações europeias, ao qual Paranhos da Silva oferece a antítese de Martius de que o Brasil “devia ser colonizado pelos milhões de índios existentes no interior” (PARANHOS DA SILVA, 1881b, p. 413).

A publicação de tais declarações numa revista que se intitula *Revista Brasileira* e em cujas páginas se viram lançadas obras de Machado de Assis, José de Alencar, Sylvio Romero e outros patronos das letras brasileiras, indicia o interesse da intelectualidade do período na busca por uma identidade nacional própria que não seja mais pautada pelas leis portuguesas, sejam essas culturais, literárias, sociais ou mesmo linguística. E nesse espaço de discussão, a língua terá grande destaque. Em todas as edições da *Revista* no período Midosi se publicará ao menos um artigo cujo tema é a língua brasileira, muitas vezes mais – serão quatro apenas no quinto tomo do segundo ano. Comum à maior parte desses textos é a rejeição das regras gramaticais portuguesas e a afirmação de uma qualidade própria ao falar brasileiro, que se não superior ao português europeu, como disse Paranhos da Silva (1880), é ao menos digno de sustentar uma literatura própria.

As discussões gramaticais presentes na *Revista Brasileira* não se limitarão a ela, porém, e ganharão as páginas de livros. Dentre esses, consta o “*Rascunhos sobre a grammatica da língua portuguesa*”, de Baptista Caetano. Essa obra se apresenta como uma resposta do autor a críticas feitas por certos intelectuais quanto à língua utilizada por escritores brasileiros, dando destaque especial à polêmica da colocação pronominal, sobre a qual se deterá por mais de uma centena de páginas.

A obra de Baptista Caetano é uma das mais claras demonstrações do nacionalismo linguístico brasileiro do período, uma defesa inflexível da “língua nacional” contra os “donos da língua”, isso é, os críticos portugueses. Logo em suas primeiras páginas se lê que “tanto se elogia e gaba a pureza dos escriptores portugueses, quanto se martela a incorrecção dos brasileiros” (ALMEIDA NOGUEIRA, 1881, p. 3). A oposição entre incorrecção e pureza retoma, novamente, a noção oitocentista da mudança degenerativa onde a divergência geraria necessariamente melhora ou piora. Que assim o seja fica claro logo no parágrafo seguinte, quando mesmo Baptista Caetano define a língua que visa defender como não sendo um outro “falar hybrido, que não é nem portuguez, nem brasileiro”, um “falar mascavado de francez, de inglez, de africano e de não sei que mais, que predomina na corte e nas cidades chamadas cultas” (*ibid.*, p. 3), mas como uma outra coisa. Esse “falar hybrido”, defende ele mais tarde, é aquele dos “habituês”, que “vem-nos de lá dos nossos tios [os portugueses], no ror (é de C. Branco) de romances franceses traduzidos para vernáculo, que entupem as estantes do Gabinete Portuguez e as da Bibliotheca Fluminense” (*ibid.*, p. 164).

O problema das importações francesas, os chamados galicismos, é recorrente na crítica ao português brasileiro, mas segundo a defesa de Baptista Caetano, se se poderia encontrar galicismos em Alencar ou algum outro autor brasileiro como diriam os censores portugueses, apenas em Pinheiro Chagas, um dos aviltados rigoristas, se acharia “novecentos e noventa e nove mil galicismos esparramados” (*ibid.*, p. 179). Para demonstrar isso, Baptista Caetano traduz para francês trechos de autores portugueses com o único intuito de apontar “estapafúrdio estylo que parece tradução litteral de livro francez” (*ibid.*, p. 4). Argumento similar usou Urbano Duarte em crônica publicada na *Gazeta Literária*, em 1902 onde diz que “a degenerescência fônica e mórfica introduzida pelo galicianismo, não só na parte lexicológica, como na própria economia orgânica da língua portuguesa, é muito mais corruptora e perversiva do que as alterações do chamado brasileirismo” (DUARTE, 1902).

O fundamento da crítica de Baptista Caetano aos galicismos se dá na noção de naturalidade, resumida por ele numa simples máxima: “para mim não é bonito o que não é natural” (ALMEIDA NOGUEIRA, 1881, p. 4). A oposição entre naturalidade e imitação percorre sua fala e adianta argumentos similares utilizados por João Ribeiro anos depois. Subjaz no discurso de ambos que a expressão linguística é manifestação do pensamento e, portanto, do espírito de quem a fala. Nesse sentido, seriam os portugueses mais corrompidos que os brasileiros, pois mais de sua língua se teria deixado influenciar pelo francês, se afastando por isso da índole (ou do gênio, talvez dissessem anos antes) do verdadeiro português, crime maior

que se poderia cometer numa língua. Frente à acusação de ser a próclise pronominal fruto de galicismo, Baptista Caetano responde se tratar esse padrão de colocação na realidade de “neo-latinismo e muito conforme com a regra de Paulino de Sousa, 73, que, notamos já, é de imensa generalidade” (*ibid.*, p. 141).

Desta feita, uma vez que se excluem os vícios, as mudanças sofridas pelo português, defende Baptista Caetano, seriam resultado natural do transplante de uma língua para um novo mundo, posto que quando inserida num novo ambiente, uma língua se transformará, não só em seus sons, mas em suas formas. Das mudanças sonoras causadas, diz ele citando Teixeira de Mello, pelo “influxo do sol intertropical” e a “indolência sul-americana”, resultariam necessariamente mudanças sintáticas pois “quem innova algumas [formas] pode innovar muitas, pode innovar todas” (*ibid.*, p. 211). A diferenciação, portanto, é “real, é fatal, não há meio algum de evita-la, tem de crescer pelo simples decurso do tempo” (*ibid.*, p. 6). E o português europeu também não estaria imune ao jugo do tempo e “si ha incorrecção do modo de falar dos brasileiros, tambem o ha no modo de fallar do portuguez moderno” (*ibid.*, p. 4).

Resulta de tudo isso que Baptista Caetano veja os censores do português brasileiro como “tyrannos” e autointitulados “donos da língua”, cujas opiniões não são nada mais que “estapafúrdias arrogantes”. Para demonstrar que esse é o caso, Baptista Caetano responde a regras de colocação pronominal propostas por um número de autores, achando nelas contínuas falhas as quais lista sem cessar. De Teixeira de Mello retira trecho em que este diz poder identificar a educação de um autor, brasileira ou portuguesa, “só pela collocação dos pronomes caprichosa no brasileiro e methodica no portuguez” (*ibid.*, p. 8), frase que repetirá diversas vezes ao longo do texto com crescente ironia. E em precedência à sua análise de casos específicos de críticas feitas pelos “rigoristas” e “donos da língua”, põe ele a questão nos seguintes termos:

A prova de que não podemos saber à quantas andamos, é a incoherencia das regras, que dão. Vêm um bastante lido e versado nos clássicos, estabelece certas regras; eis quando, chega outro, que tambem manuseou os clássicos e estatue preceitos sinão inteiramente diferentes dos do outro, pelos menos bastante discordantes, e às vezes discordantes entre si os do mesmo mestre (*ibid.*, p. 11).

Culminará sua resposta num recuo à *Grammatica Philosophica* de Joaquim Soares Barbosa, “o mestre soberano que dicta a lei (desgraçadamente) no domínio da língua portuguesa” (*ibid.*, p. 15). Entre críticas a autores e gramáticos portugueses, Baptista Caetano forja seu discurso em uma incessante invectiva ao modo europeu, contra os “puristas” e a favor dos “brasileirismos”, contra o antigo modelo intelectual de estudo da linguagem e tudo que ele representa. Classifica a si mesmo como “insubordinado” e a seus opositores como “tyrannos”, acusa-os de arbitrariedade nos julgamentos e de hipocrisia, diz que “em tudo e por tudo os homens chalaceam da fala brasileira [...] dizem chufas e nada mais, pois elles não sabem o que censuram” (*ibid.*, p. 135).

O clamor de Baptista Caetano é, em essência, uma rejeição aos moldes da estilística lusitana, assim como de sua língua e suas regras. Seu discurso invoca ativamente a necessidade de um divórcio linguístico como o foi o divórcio político, relação explicitada no momento em que o autor invoca a imagem de José Bonifácio de Andrada e da constituição. Seja na fala, no pensamento ou na literatura, o Brasil deverá ser diferente de Portugal e não mais pautar sua evolução em “importações” de Portugal. Paranhos da Silva, a quem Baptista Caetano se refere diversas vezes ao longo de seu texto, publica num artigo na *Revista Brasileira* um pensamento similar ao dizer que:

Finalmente, me parece que o pronome se pode colocar sempre antes do verbo; mas só às vezes antes dele [*sic*]. Por isso também me atrevo a lhes aconselhar que, não contrariando a gramática geral, não usem na escrita linguagem diferente da que todos falam no Brasil; mas se conformem com o gênio do nosso idioma, isto é, segundo Whitney, com o resultado das preferências da sociedade brasileira (PARANHOS DA SILVA, 1880)

A ideia recorrente é a da diferença linguística como signo da diferença “espiritual”, advinda essa das divergências sociais e mesológicas existentes entre Brasil e Portugal que com o tempo se impuseram sobre a índole da língua. A mesologia, ou clima, será um aspecto reiterado na fala do período, mas que alguns tomarão como recebendo demasiado papel nas teorias de mudança. Assim pensa João Ribeiro, que dirá que a influência do clima é tida em exagero por muitos e que o único aspecto da língua em que ela seria inegável seria a prosódia, ou sotaque (RIBEIRO, 1897). Seja como for, o clima brasileiro será tema não apenas nos

estudos da língua, mas também na antropologia e na literatura, onde comporá um dos pilares do ideário romântico brasileiro.

Sobre outro pilar da fundação nacional do Brasil, a questão indígena, pouco se lerá nos Rascunhos de Baptista Caetano, todavia parte significativa de seu trabalho intelectual foi dedicado ao estudo de línguas nativas americanas, sobre as quais publicou opúsculos e artigos na *Revista Brasileira*, donde chama atenção para a necessidade do estudo dessas para a compreensão da história do Brasil (ALMEIDA NOGUEIRA, 1880). Ainda que não manifestado em sua resposta aos “rigoristas”, o elemento indígena reside no pensamento de Baptista Caetano e dentre as misturas que formam a “língua mascavada” por ele criticada, não constarão essas. Suas considerações sobre esse assunto podem ser mais bem avaliadas através de seu opúsculo nos *Rascunhos de Sciencia*, onde apresentará suas teorias de afiliação das línguas indígenas do Brasil, de modo similar ao que o fez Lopez no *Races aryennes du Pérou*, diferindo desse em sua reduzida formalização de dados e em não tentar defender uma origem indo-europeia para alguma língua brasileira (ALMEIDA NOGUEIRA, 1876). O opúsculo de Baptista Caetano será, curiosamente, um dos poucos trabalhos brasileiros a operar o método comparativo para uma tentativa de classificação linguística.

Ao cabo dos pontos até aqui levantados, é possível ver manifesto um certo nacionalismo linguístico na voz de Baptista Caetano, nacionalismo esse que se pauta em termos do “espírito” da nação, da “naturalidade” da fala, da “inevitabilidade da mudança”, de sua necessidade de emancipação das regras dos “gramaticistas” e “rigoristas” e contra o discurso gramatical vindo da antiga metrópole. Seu ambiente conceitual é, portanto, o de um clamor de independência de ideias e formas, ainda que não necessariamente de língua. Em nenhum momento Baptista Caetano tratará o português brasileiro como um dialeto ou uma língua própria, a menos que se considere o português europeu também como um dialeto.

Outro autor de considerável relevo para a época e cuja obra tematizou longamente a questão do português brasileiro foi João Ribeiro. Intelectual, membro da Academia Brasileira de Letras e professor do colégio Dom Pedro II, foi responsável pela publicação de um número de obras acerca de questões do país, dentre as quais serão aqui consideradas sua *Grammatica Portugueza*, de 1889, e *A Língua Nacional*, de 1933. São ambas obras tardias, a segunda das quais publicada mais de quatro décadas após o fim do Império do Brasil, todavia permanecem relevantes dada sua representatividade de questões fundamentais da discussão nacional-linguística que perpassou o crepúsculo imperial e prosseguiu para a realidade republicana. Não

obstante, a produção em livro de João Ribeiro se inicia em 1884, sendo sua formação, portanto, feita no âmbito do mundo Imperial. Por fim, considera-se que sua relevância para a discussão do nacionalismo linguístico brasileiro é por demais importante para ser ignorada.

Ribeiro define “gramática” como “a coordenação de leis ou regras, segundo as quaes uma língua é falada ou escripta” (RIBEIRO, 1889, p. 5). Dedicará vocábulos para definir também a “grammatica histórica” e a “grammatica comparativa”, nisto revelando sua base teórica, não surpreendentemente para o período, consolidada sobre os termos da tradição histórico-linguística oitocentista. Dedicará maior parte dessa obra a discussões gerais de questões puramente gramaticais. De cunho educacional mais do que histórico ou comparativo, tratará o livro dos “brasileirismos” somente em notas de rodapé, à exceção de sua quadragésima quarta lição, donde abordará tal ponto juntamente a anomalias gramaticais, idiotismos e dialetos. Dialetos que, definirá Ribeiro, seriam “os modos de falar ou escrever a língua portuguesa nas diversas regiões onde foi ella implantada” e que do português seriam o galego, o indo-português e o “áfrico” (RIBEIRO, 1889, p. 306). Em contrapartida a esses que seriam dialetos verdadeiros, o brasileiro é apenas “impropriamente chamado dialeto” (*ibid.*) ainda que esse difira do lusitano por “diferenças notáveis de prosodia e de syntaxe, por um vocabulário novo de termos tupis-guaranis e africanos”. Curiosamente, em um ponto à frente na gramática se lerá que “a reacção litteraria de dous seculos nunca pôde obstar nem diminuir a dialectação do portuguez no Brazil”. Esse argumento em muito remete àquele oferecido por Silvio Romero (1880) para quem a dialeção do português brasileiro ainda não tinha acontecido, mas poderia eventualmente se dar. A discussão quanto à classificação do brasileiro como dialeto ou não levou à pena as mãos de outros autores brasileiro e grandes oscilações se dão quanto a seu uso ou o de alguma outra denominação do brasileiro, seja como linguagem, língua ou norma, mas como um todo poucos serão os que defenderão que o português do Brasil fosse uma língua diferente do europeu (COELHO, 2008).

Mais à frente, ao listar os “brasileirismos sintáticos”, João Ribeiro adjunta ao texto uma longa nota de rodapé onde discute as causas da dialeção sofrida pelo brasileiro, as principais das quais seriam “as novas necessidades da vida colonial, as condições climatéricas e topográficas, as relações constantes com os povos originários índios e com os africanos” e secundariamente “o povoamento das fronteiras por povos castelhanos, a disseminação dos ciganos expulsos de Portugal” (*ibid.*, p. 310). Apontará, por fim, que os elementos “tupi e o africano” tenderão ao desaparecimento decido à “extinção da imigração negra e pelo caldeamento das raças”. Em ordem inversa a esses processos, dirá ele que no sul a imigração

de outros povos europeus resultará, dentro de um século, que “o sul do Brasil destruirá a unidade ethnica da pátria brasileira, se outras circunstancias não se oppozerem à evolução que já se vai notando desde agora” (*ibid.*, p. 310). Em outra obra, o *Diccionario Grammatical*, João Ribeiro classificará como “desfavorável ao português” de certas regiões fossem habitadas por números muito grandes de africanos e indígenas, pois disso resultaria o falar “viciadíssimo” dos habitantes de tais regiões (RIBEIRO, 1897, p. 354).

De certo que não são esses elementos de darwinismo-social únicos ao pensamento de João Ribeiro, mas topoi do pensamento oitocentista, amplamente manifestados nas práticas científicas do período e, ainda mais que isso, constituintes da razão nacionalista brasileira. Todavia, reside na manifestação brasileira dessas ideias um aspecto ambíguo. Por um lado, seria o contato com as línguas indígenas e africanas que daria ao português brasileiro sua expansão de ideias e a garantia de suas qualidades individuais. Por outro, porém, dessa mesma influência resultam “vícios” e “decadência”. Mesmo a imigração de populações de origem europeia, parte do projeto racial da nação, surge com ressalvas, uma vez que o contato com outras nações europeias surge como ameaça à “unidade ethnica da pátria brasileira” (RIBEIRO, 1889, p. 35).

Não em muito se difere a *Grammatica* de João Ribeiro daquelas outras produzidas no período sob a orientação do método histórico-comparado. Sua defesa do português brasileiro, contudo, já nessa época se demonstra com notável força. Não apenas dedicará ele mais espaço que Júlio Ribeiro para tratar da questão como também erguerá considerações político-sociais que no outro autor não se viam manifestadas e que virão à primeira página em sua obra quatro décadas futura, a não sem razão chamada *A língua nacional*.

Publicado no século XX, já não no contexto da transição entre Império e República, mas naquele de uma República inserida no entreguerras, a obra é parte do que se poderia compreender como nacionalismo da alvorada do novo século, marcado pela modernização de estruturas e discursos e uma ânsia de inserção do Brasil no círculo das nações industriais. Será um mero ano depois que, no centenário da Independência, se fará a semana de arte moderna e só nove anos mais tarde se iniciará a Era Vargas, período marcado pela invasão de interesses nacionalistas em todos os âmbitos da vida pública brasileira. Como apontado por Schwarcz, nos anos 1920 a questão da raça ressurge no discurso da nação, descambando em propostas eugênicas e um certo desgosto para com a miscigenação, herança da razão darwinista que imperava em certa linha do pensamento oitocentista (SCHWARCZ, 1957 [1993]).

Em tal ambiente, essa segunda obra de João Ribeiro se introduz com maior ímpeto e força em seu discurso acerca da língua do Brasil. Se na *Grammatica* seu discurso se respaldava no poder inerente do método científico e sua linguagem era como um todo consideravelmente neutra, em *A Língua Nacional*, João Ribeiro deixará de lado qualquer possível neutralidade ou pretensão do texto em si é uma compilação de análises filológicas de expressões brasileiras das mais diversas, mas sua introdução é uma longa defesa do português brasileiro, da validade de suas especificidades e ainda mais da necessidade do fim da “submissão das palavras” brasileiras pelas portuguesas. Diz Ribeiro que “parece incrível que a nossa Independência ainda conserve essa algema nos pulsos, e que a personalidade de americanos pague tributo à submissão das palavras” (RIBEIRO, 1933, p. 8). Assim, cria o texto de partida um laço necessário entre a Independência política, à época já solidificada, e a Independência linguística. Nisso Ribeiro retorna a ideias já expostas pelo pensamento do oitocentos tardio segundo os quais seria a língua a manifestação maior da mente, sua qualidade a qualidade dos povos que a falam. Daí ser possível dizer que “corrigindo-nos [os brasileiros], estamos de facto a mutilar ideias e sentimentos que não são pessoas. Já não é a língua que apuramos, é o nosso espírito que sujeitamos a servilismo inexplicável” (RIBEIRO, 1933, p. 8). Soam ainda aqui ecos de Whitney, distando a visão de Ribeiro apenas que nem o português lusitano nem o brasileiro seriam línguas bárbaras – nem o poderiam ser, sendo línguas europeias e herdeiras de Roma – mas sendo cada qual a língua de uma nação diferente, com realidades sociais, políticas, geográficas e humanas distintas, as “corrigendas e erratas vernáculos” resultariam no entrave de “todo o progresso nacional” (RIBEIRO, 1933, p. 14). Essa equivalência língua/espírito utilizada para confrontar o português do Brasil com aquele de Portugal existia já na *Grammatica*, visível quando Ribeiro diz que os são o “resultado da educação do povo pelo espírito francês” (RIBEIRO, 1889, p. 301). Assim, o fim da sujeição das palavras seria necessário para o fim da sujeição do espírito, isso é, do pensamento brasileiro.

Petri, Schneiders e Scherer (2015) lerão *A Língua Nacional* na chave de um “discurso descolonizador” que estaria em sintonia com outros discursos similares propostos em outras nações americanas. De fato, uma seção significativa do capítulo homônimo *d'A Língua Nacional* será dedicada a uma citação de Rupert Hughes, cujas palavras Ribeiro tomará, *mutatis mutandis*, como descritivas das necessidades e ânsias brasileiras. Em ambas as falas figura o desejo de uma independência intelectual que permita às antigas colônias afirmar seu pertencimento no rol das grandes nações do mundo moderno. Há, contudo, uma tensão aqui presente, um paradoxo fundamental: pois se importar e estruturas e palavras é, de fato, se

“educar pelo espírito” de outro povo, ter como língua nacional aquela do dominador seria manter-se sob o julgo de sua cultura e poder intelectual. Consciente, talvez, desse problema, Ribeiro faz o seguinte apontamento:

“A Lingua nacional é essencialmente a língua portugueza, mas enriquecida na America, emancipada, e livre nos seus proprios movimentos. Com esse intuito, e n’esse fundamento, foram escriptas as paginas do nosso livro que não inculcam lingua nova, mas revelam os matizes, as variações, e a originalidade do pensamento americano” (RIBEIRO, 1933, ênfases adicionadas).

Lido em termos de uma teoria cultural, esse trecho se mostra em consonância com as teses históricas e raciais pregadas desde a origem do Império. Pois que seja sim o Brasil uma nação independente, nascida da miscigenação de três raças, mas sua fundação social é portuguesa. Assim como é a língua nacional “essencialmente a língua portuguesa, mas enriquecida na America” é a nação brasileira fundamentalmente uma nação europeia ainda que a seu branco originário se tenha somado o cobre americano e qualquer coisa do negro africano.

Ribeiro prossegue a defesa rejeitando que se tome como erros certas particularidades do português do Brasil, visto que “todos os phenomenos são especialmente legítimos” e que “todos os factos de linguagem, cá e lá, têm equal excellencia, como phenomenos” (*ibid.*, p. 10). Usando-se do cientificismo, esse argumento visa legitimar os usos brasileiros, dentre os quais será mencionado a amplamente discutida colocação pronominal, assim como as expressões populares às quais a obra se dedicará. Todos esses são vistos por Ribeiro como indícios da evolução própria da língua nacional, necessários para a “independência do nosso pensamento e de sua imediata expressão” (*ibid.*, p. 16), mas não como o suficiente para criar um dialeto nem como um sinal de degeneração. São o mero resultado de uma nova mente, de um novo espírito e de um novo povo em um novo mundo. Porém, opostos a esses usos válidos, fruto da criatividade americana, estariam os “erros inexcusáveis” que permeiam certos falares brasileiros, os quais Ribeiro não definirá nesse ponto, mas que se poderia inferir, sem risco, como o falar “viciadíssimo” dos caipiras (RIBEIRO, 1897). Se some a isso sua definição de gramática descritiva como “a arte que ensina a falar e escrever **correctamente** uma língua” (RIBEIRO, 1889, p. 2, ênfase adicionada) e disto se percebe que há no pensamento de João

Ribeiro um limite implícito nas liberdades que uma língua pode tomar antes que incorra em vício e erro. Enquanto a importação de elementos lexicais e a alteração de certos elementos sintáticos é admissível, mudanças mais bruscas na sintaxe são ameaças à estabilidade da língua.

Ameaças, sim, pois a língua é “um superorganismo em progresso ou em decadência, e sempre em movimento” (RIBEIRO, 1889, p. 301) e “as palavras corrompem-se mais profundamente pela terminação”, razão pela qual “se perderam os casos em língua tão corrompida quanto devia ser o latim falado pelos estrangeiros bárbaros” (*ibid.*, p. 114, ênfase adicionada). Se reitera aqui o papel daquele que fala na evolução da língua, mais especificamente em sua derrocada, em sua corrupção, em sua devastação. A adjetivação negativa é reveladora do medo carregado por João Ribeiro, simbólico do medo maior de seu período de que a mistura eventualmente levasse à degeneração.

Em vista das teorias darwinistas imperantes no XIX, se torna possível conciliar a necessidade de diferenciação do português brasileiro e que essa seja feita em elementos mais externos (o léxico, a pronúncia) ou em termos internos que possam ser, sem muita dificuldade, argumentados como tendências ancestrais. Será assim que se justificará o padrão brasileiro de colocação pronominal, “uma corrente arcaica e muito profunda que havia extirpado da língua todos os esdrúxulos, quase todas as palavras dactylicas do italiano e do latim, renovadas depois, um pouco pedantesamente, pelos letrados do renascimento” (RIBEIRO, 1933, p. 13). Tudo aquilo que não possa ser justificado, contudo, deve ser tomado como erro e expurgado. Assim, a evolução própria da língua, signo da independência, é válida apenas até certo ponto além do qual o afastamento da matriz portuguesa traz o risco de “degeneração”. Para evitar esse mal, para garantir o bem-estar da língua e do povo, é que servirão a gramática e o ensino.

Em resumo, a mecânica da definição do português como uma língua nacional e uma língua de cultura no século XIX poderia ser entendida como uma que ao mesmo tempo nega e afirma aspectos únicos da língua, designando para partes mais externas o papel de criar a “personalidade” ou “espírito próprio” da língua brasileira em contraste contra o seu núcleo duro, o qual seria extremamente português e tradicional. Essa é uma definição que não escapa da conceitualização de nação proposta por Martius, segundo a qual a raça branca comporia o eixo cultural e moral da nação ao qual as culturas negra e indígena seriam adicionadas como pouco mais que um termo acessório, um “tempero tropical” que se posto em quantidade excessiva poderia estragar a “sopa brasileira”. Nesse mesmo sentido, a unidade nacional é outro mito central para a “ideia do Brasil” e ainda que regionalismos acabem por ocupar certo papel na

construção da nação no século XX, elementos culturais foram empregados para tentar se criar a ilusão de um estado homogêneo. Assim, a língua do Brasil, como proposta e pretendida pela intelectualidade oitocentista não é diferente do próprio Brasil, um número de partes de constituição altamente diversa mantida sob a ilusão de igualdade.

4: PERFIS DOS AUTORES

Como será visto, diversos dos autores utilizados são muito pouco famosos e a informação disponível sobre eles é mínima. De outros, em contraste, se conta com uma vasta fortuna crítica. Essa discrepância na quantidade e qualidade das informações disponíveis, contudo, não é prejudicial uma vez que este trabalho assuma uma postura que prioriza a variável e não os autores – isso é, a suposição inicial é a de uma correlação real entre a variável (eIP *versus* eGER) e o conflito Portugal *versus* Brasil. A partir da utilização relativa de um ou outra forma, o autor revelaria uma tendência mais ou menos voltada para certa via. Obviamente, é pouco provável que uma relação tão clara se mostre, porém, esse é um fato desejável dentro dos objetivos propostos de testar até onde o método é capaz de ir. Não se pode negar que uma pesquisa mais dedicada a interesses de terceira onda encontraria maior sucesso se focasse em análises mais detalhadas de um número menor de indivíduos, porém, tal abordagem viria ao custo de qualquer possibilidade de generalização. De um modo ou de outro, associar certos usos linguísticos a determinadas práticas requer estudos consideravelmente mais amplos mesmo para variáveis simples de grupos modernos – em se tratando de uma variável como a aqui utilizada e um grupo consideravelmente complexo (se é que se poderia sequer chamar o apanhado de autores aqui utilizados de um “grupo”) tentar supor ligações diretas entre variável e opiniões seria temerário.

A despeito disso, algum nível de conhecimento acerca dos autores utilizados é desejável, ainda que somente para que se abra possibilidades de questionamento. Nenhum desses perfis será, é claro, exaustivo ou mesmo muito longos. São nada mais que breves familiarizações com os pontos principais indivíduos de interesse.

4.1: CRITÉRIOS PARA SELEÇÃO DE AUTORES E OBRAS

Esse trabalho se baseia em quinze obras, cada uma das quais produzida por um autor e, distribuídos entre cinco províncias, sendo três autores para cada província. Para que um autor fosse escolhido como um informante válido para esse trabalho, ele deveria ser um autor que tenha nascido ou vivido uma grande parte de sua infância num das cinco províncias de interesse no período que compreende o regime imperial, isso é, entre 1822 e 1888. Sendo esse um critério

altamente abrangente, o número resultante de autores válidos é consideravelmente alto, muito mais do que se poderia analisar adequadamente. Para filtrar esses números, se estabeleceu alguns critérios adicionais.

Dentro do possível se tentou selecionar autores culturalmente relevantes e reconhecidos como exemplares dentro da literatura de suas respectivas regiões. Esse critério foi posto em frente dado à natureza desse estudo, que é fundamentalmente voltado à cultura intelectual e letrada do Brasil Império. Baseado nisso, selecionar autores tomados como exemplares permite ter um meio de entrada seguro dentro da população alvo. Ademais, se estabeleceu por razões práticas que obras com um número de ocorrências menor que 9 não poderiam ser utilizadas. Por fim, dentro do possível, se tentou selecionar ao menos uma autora do sexo feminino de cada província, escolha feita para que o estudo possa ter uma maior representatividade da realidade linguística, mas também para tentar resgatar alguma memória de autoras de outro modo esquecidas.

Quanto às obras, foram escolhidas as obras mais relevantes de cada autor, desde que essas atendessem ao critério de ocorrências mínimas e estivessem disponíveis. No caso de mais de uma obra ser reconhecida, se selecionou aquela com o maior número de ocorrências totais. As edições selecionadas foram, quando possível, as versões mais antigas disponíveis em formato digital. O formato digital foi preferido pela possibilidade de extração automática de dados, mais precisa e mais rápida que a extração manual. Dentro desses critérios, se selecionou os autores listados e descritos a seguir.

4.2: PERFIS

4.2.1: Amadeu Amaral

Capivari (atual Monte Mor), 1875 – São Paulo, 1929

Autor nascido no oeste Paulista, se mudou para São Paulo em 1892, nos primeiros anos da primeira República. Lá, trabalhou no *Correio Paulistano*, publicação oficial do Partido Republicano Paulista (PRP), até 1905, quando se desligou do órgão. Em 1910, passou a escrever para *O Estado de São Paulo*, publicação opositora ao PRP, Partido que então dominava a cena

política paulista. Suas relações com *O Estado de São Paulo* se mantiveram ao longo do restante de sua vida e lhe garantiram espaço onde publicar sua futura obra literária e ensaística.

Em 1917, se filiou à Liga Nacionalista de São Paulo, partido orientado pelo projeto da Liga de Defesa Nacional, fundada no ano anterior por Olavo Bilac, pela qual concorreu a deputado pelo quarto Distrito Eleitoral em 1920, defendendo uma plataforma política baseada no voto secreto e universal. Perdeu a eleição para Júlio Prestes, o candidato do PRP.

Em 1919, foi eleito para a ABL, ocupando a cadeira que fora de Gonçalves Dias e Olavo Bilac – fato que foi justificado pelo seu interesse pelos assuntos nacionais. Em 1922, após publicar uma série de denúncias contra as práticas políticas do PRP, Amadeu Amaral se mudou para o Rio de Janeiro, onde permaneceu até 1926, quando retornou em São Paulo. Lá, se manteve pelo resto da vida.

Autor polivalente, foi responsável por textos jornalísticos, ensaios políticos, poesia, prosa, estudos folclóricos e linguísticos. Nessa última categoria se encaixa *O Dialeto Caipira*, de 1920. Orientada pela então já formada linguística (o “*Cours de linguistique générale*” é de 1916), a obra faz uma descrição do dialeto caipira de São Paulo em suas dimensões fonética, morfológica, sintática e lexicológica, podendo ser assim compreendida como uma manifestação da tradição dialetológica que crescia nos estudos linguísticos do Brasil no início do século XX.

Em 1921, publicou na *Revista do Brasil* o artigo “A língua nacional”. Tal texto, fundamentalmente um ensaio, apresenta a visão de Amadeu Amaral sobre a dimensão social e política da língua. A língua, ele diz, é “a manifestação mais extensa e mais profunda da alma multiforme de uma nacionalidade, porque é obra anônima e inconsciente de inúmeras gerações” (AMARAL, 1921, p. 26). Tal aspecto político não surpreende – toda a carreira de Amadeu Amaral será pautada na discussão dos caminhos da República – e se revela de pronto na frase que conclui o texto: “a língua é a própria nacionalidade” (*ibid*, p. 31).

A obra aqui utilizada foi *Memorial de um passageiro de bonde*, série de crônicas publicada postumamente em 1938. A versão utilizada aqui é a original, publicada pelas Edições Cultura Brasileira e digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina.

4.2.2: Bernardo Guimarães

Ouro Preto, 1825 – Ouro Preto, 1884

Autor de grande renome por suas novelas, muitas das quais permanecem populares mesmo quase dois séculos depois de seu nascimento, Bernardo Guimarães nasce numa família de intelectuais, seu pai e irmãos também homens de letras. Aos 22 anos, ingressou na Faculdade de Direito de São Paulo, se formando em 1852.

Em 1859, se mudou para o Rio de Janeiro, onde escreveu para o jornal *A Actualidade*. Retornando às Minas Gerais, passou a lecionar no Liceu a partir de 1866, mas pouco depois voltou para Ouro Preto, onde permaneceu até sua morte.

Sua produção literária é contínua e considerável em tamanho, tendo publicado onze romances apenas entre 1868 e 1883, além de ampla produção poética e crítica em uma fase anterior (SOUZA, 2012). Considerado um representante do romantismo mineiro, sua obra atingiu grande popularidade em seu tempo e além e nela se vê manifestado amplo sentimento abolicionista, tema abordado em algumas de suas mais notáveis obras, como aquela que o aqui representa, *A escrava Isaura*, de 1875. O livro, como a maioria das outras obras de suas outras obras, foi publicado pela carioca B. L. Garnier e foi adaptada para a televisão em duas ocasiões. A versão utilizada é aquela digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina a partir da edição original de 1875.

4.2.3: Carneiro Vilela

Recife, 1846 – Recife, 1913

Autor cuja amplitude e profundidade literária não pode ser exagerada, Carneiro Vilela atuou na produção de poesia, romances, contos, ensaios e textos jornalísticos. Nascido já no segundo reinado, Lucena Filho (2016) o associa à geração de 1870, fase da vida intelectual brasileira marcada pela profunda desilusão com o fracasso do projeto imperial em suprir as reformas necessárias para a modernização do estado nacional brasileiro.

Herdeiro das famílias Carneiro e Vilela, ambas envolvidas na vida política de Pernambuco (as duas tiveram membros envolvidos na revolta de 1817 e, mais tarde, em cargos públicos), Carneiro Vilela foi aluno da Faculdade de Direito de Recife entre os anos de 1862 e 1866. Foi nesse período que teve início sua carreira dramática e jornalística, tendo participado da fundação de um periódico estudantil, o *Ilustração Acadêmica*.

Em 1876, em razão da grande popularidade de sua peça *Os maçons*, se mudou para Belém, no Pará, onde irá permanecer até 1879. A referida peça, vale notar, foi de enorme repercussão, positiva por partes, mas extremamente negativa por aqueles ligados às instituições católicas por ela atacadas. Já em 1880, se mudou para o Rio de Janeiro, onde publicou textos na *Revista Brasileira*, exerceu o cargo de juiz por um mês e acabou envolvido em casos de diversas ordens. É nesse período que adquiriu o jornal *A Gazetinha*, no qual serviu como editor. Retornou para o Recife em 1885, onde permaneceu exercendo o jornalismo e a pintura.

No seu retorno para o Recife, publicou aquela que seria sua obra mais reconhecida, *As tragédias do Recife*, também chamada de *A Emparedada da Rua Nova*. Nessa mesma fase, sua obra jornalística incluía diversas charges cáusticas contra o governo imperial e instituições religiosas, ambos vistos pelo autor como causadores da falta de progresso no Brasil. Após o fim do regime Imperial, continuou publicando e defendendo reformas institucionais na economia, educação e administração da república. Permaneceu nessas mesmas vias pelo resto da vida e morreu em 1913 (LUCENA FILHO, 2016).

De sua vasta obra, se selecionou o romance *Noêmia*, publicado em 1894 pela imprensa do jornal pernambucano *A Província*, em edição original presente na coleção Mindlin.

4.2.4: Franklin Távora

Baturité, 1842 – Rio de Janeiro, 1888

Nascido no Ceará, aos cinco anos se mudou junto aos pais para Pernambuco, onde recebe educação formal e, mais tarde, cursou direito na Faculdade de Direito de Pernambuco. Durante sua formação, participou de diversos jornais, se envolveu em polêmicas com a Igreja e participou da maçonaria. Em seu período como jornalista, se associou fortemente aos movimentos liberais que se espalhavam pelo país.

Em 1874, se mudou para o Rio de Janeiro, onde assumiu um cargo público. Foi também no Rio de Janeiro que entrou em contato com outros autores junto aos quais fundou a Academia Brasileira de Letras. Sua produção literária continuou nesse período e em 1876 foi publicado *O Cabeleira*, romance que aqui foi selecionado para representar a obra de Távora.

Ainda no Rio de Janeiro, Franklin Távora se envolveu na direção da *Revista Brasileira*, mantendo o controle sobre a revista entre 1879 e 1881. Sua produção para a revista se voltou para a representação de uma cultura nordestina – ou “do norte”, como ele chamava. Em consonância com sua política liberal, sua produção literária absorveu ideias do racionalismo naturalista e se opunha ao romantismo que lhe antecedeu. Após deixar a direção da *Revista Brasileira*, ingressou no Instituto Histórico e Geográfico Brasileiro, permanecendo como um membro ativo do órgão e dentro dele avançando seu projeto folclorístico. Morreu em 1888, no Rio de Janeiro (BERTIOLLI, 2008).

A edição de *O Cabeleira* aqui utilizada é a de 1902, publicada por H. Garnier e presente na coleção Mindlin da Universidade de São Paulo.

4.2.5: Inês Sabino

Salvador, 1853 – Rio de Janeiro, 1911

Oriunda da Bahia, nasceu numa família abastada, o que lhe permitiu estudar na Europa, na Inglaterra e na França por boa parte da vida. Retornou ao Brasil após a morte do pai, quando passou a morar em Recife. Mais tarde, se mudou para São Paulo e então para o Rio de Janeiro, onde permaneceu até sua morte (QUINLAN, 1998). Desde o seu retorno para o Brasil, se manteve ativa nas letras, publicando artigos e textos literários em jornais.

Do momento em que retornou para o Brasil, esteve envolvida com a causa abolicionista, tomando parte na sociedade Aves Libertas, além de ter uma dedicação notável para o estudo da figura feminina. Sobre esse segundo tema, publicou *Mulheres Ilustres do Brasil*, uma série de biografias sobre figuras femininas e sua importância na história brasileira. Morreu em 1911, no Rio de Janeiro.

A obra de Sabino utilizada nesse trabalho é a coletânea *Contos e Lapidações*, publicada em 1891, já no período carioca de sua obra. A edição utilizada é a original, publicada pela

Laemmert & Co. e disponibilizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina.

4.2.6: Joaquim Felício dos Santos

Serro, 1828 – Diamantina, 1895

Nasceu no interior de Minas Gerais em 1828, filho de um funcionário público da antiga hierarquia colonial. Teve os estudos básicos em Diamantina, mas cursou o ensino superior na Faculdade de Direito de São Paulo. Formado, retornou a Diamantina em 1850, quando passou a lecionar no Ateneu São Vicente de Paula e no Seminário Episcopal.

A partir da década de 1860, iniciou carreira política, sendo eleito deputado em 1864, mas falhando em conseguir a cadeira de senador por Minas Gerais em 1883. Nesse período, já tinha iniciado a publicação de seus textos – literários ou não – escrevendo para jornais de Minas Gerais. Em 1860, fundou *O Jequitinhonha*, jornal que receberia parte significativa de sua obra. Tal jornal, vale destacar, se orienta sob a ótica liberal-republicana, tendo sua primeira edição sido lançada às vésperas das eleições imperiais daquele ano.

Educado como liberal, sua posição política vai se radicalizando ao longo dos anos, ao ponto de rejeitar a possibilidade de ser eleito senador em 1889 em vista da continuidade do sistema monárquico. Morreu poucos anos depois, em 1895 (NOVAES, 2014).

De sua obra, usa-se aqui o romance *Acaiaca*, de 1866, em edição publicada em 1894 pela Tipografia do jornal Estado de Minas, versão digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina.

4.2.7: Júlia Lopes de Almeida

Rio de Janeiro, 1862 – Rio de Janeiro, 1934

Filha de portugueses, mas nascida no Rio de Janeiro, Júlia Lopes de Almeida foi uma das mais bem reconhecidas autoras do século XIX. Devido à orientação política de seus pais, políticos de linha liberal, estudou por alguns anos na Europa. Foi uma escritora ativa, publicando textos no *Jornal do Comércio* e a *Gazeta de Campinas*, além de outros periódicos. Sua ligação com a Europa, dada pelos seus pais e pelos seus estudos no velho continente, a permitiram trasladar entre as duas terras, publicando obras no Brasil e em Portugal (FANINI, 2009).

Educada como liberal, sua obra é uma das mais notáveis manifestações do feminismo oitocentista no Brasil, tendo mesmo participado do Primeiro Congresso Internacional Feminista, em 1922. Escreveu romances, contos, poesias e textos práticos. Morreu em 1934, no Rio de Janeiro, onde recebeu homenagem da Academia Brasileira de Letras. Sua produção aqui é representada por *A Intrusa*, romance de 1905, publicado em Portugal, em edição de 1908 da companhia editorial Francisco Alves, do Rio de Janeiro, disponível na coleção Mindlin.

4.2.8: Júlio Ribeiro

Sabará, 1845 – Santos, 1890

Anteriormente, se tratou da obra de Júlio Ribeiro em sua dimensão linguística – foi ele o responsável por publicar a primeira gramática brasileira orientada pela linguística histórica. Que se fale agora de sua pessoa.

Filho do norte-americano George Washington Vaughan e da professora Maria Francisca Ribeiro Vaughan, realizou estudos primários em Minas Gerais, antes de se mudar para o Rio de Janeiro, onde por três anos cursou a Escola Militar, a qual largou para exercer estudos liberais. Passou daí a se dedicar ao jornalismo, ao magistério e à literatura.

Sua formação liberal passou por muito daquilo que seriam as marcas tradicionais dos intelectuais de mesma orientação àquele tempo: a prática maçônica e a defesa da república como modelo de organização social. Sua literatura também seguiu nessa mesma direção, se alinhando à tradição Naturalista, favorita dos autores que mais apreciavam o cientificismo Darwiniano de então. A obra de Ribeiro que aqui se usou, “A Carne”, de 1888, se encaixa desde o título nesse molde estético.

Como jornalista, dirigiu um considerável número de periódicos, os mais dos quais foram fundados em São Paulo, estado que havia se tornado sua residência desde muito anos. Teve em Sorocaba como sua cidade de residência, mas organizou periódicos nas outras duas principais cidades do Estado, Campinas e São Paulo. Como jornalista, representou a defesa de seus interesses políticos e religiosos, jamais se esquivando dos confrontos com os monarquistas, grupo dominante da política Sorocabana em seu tempo. Morreu no ano de 1890, em Santos (SILVA, 2016).

A edição de *A Carne* aqui utilizada é a segunda, de 1896, publicada pela Livraria Teixeira e disponível na coleção Mindlin.

4.2.9: Lindolfo Rocha

Grão Mogol, 1862 – Jequié, 1911

Meira & Rebello e Almeida (2016), em artigo-resumo sobre a vida e obra de Lindolfo Rocha, o definem como um dos nomes esquecidos da literatura mineira, um autor que apesar de seu valor artístico acabou à margem da tradição literária brasileira.

A vida de Lindolfo Rocha é como sua obra, pouco se sabe dela. As autoras relatam sua origem pobre, um filho mulato, órfão de pai e de cuja formação básica nada se conhece. Em 1880, se mudou para a Bahia para estudar e poucos meses depois ingressou na Faculdade de Direito de Recife, se formando bacharel em 1882. Diplomado, permaneceu na Bahia por um ano como juiz-preparador, antes de se mudar para Jequié, cidade onde fez vida. Lá, exerceu magistério por cinco anos, quando então se tornou advogado, profissão que exerceu até sua morte, em 1911.

Como tantos outros, exerceu o jornalismo em vida e fez dos periódicos a primeira casa de muito de sua obra. Sua literatura é definida como regionalista, o sertão nordestino, nos tempos de outrora assolado pela fome e a miséria, é o palco principal de sua obra. Seu *Magnum opus* e obra que aqui o representa, *Maria Dusá*, se encaixa nessa ordem de fatos, se dispondo a representar as agruras da zona de garimpo do norte de Minas Gerais.

A edição de *Maria Dusá* aqui utilizada é a versão digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina

a partir da edição original publicada em 1910 pela Livraria Chardron de Lello e Irmão, em Portugal.

4.2.10: Luísa de Camargo Pacheco

Campinas, ? – Campinas, ?

Muito pouco se sabe da pessoa de Luísa de Camargo Pacheco afora aquilo que é dito na introdução do romance que serve de fonte para esse trabalho. Como muitas autoras oitocentistas, sua pessoa e obra foram em grande parte perdidas para o tempo. É sabido que estudou no colégio de Ytu, como diz Basílio Magalhães no prefácio à obra aqui considerada e que provavelmente nasceu em Campinas. De sua obra, se pode inferir que recebeu educação formal e que sabia inglês e francês – línguas que utiliza ao longo da narrativa – e que era católica. Além disso, nada mais se descobriu.

A obra da enigmática autora usada para fins desse trabalho é o romance *Alice*, publicado pela tipografia Livro Azul de Castro Mendes & irmão, da cidade de Campinas, em 1903.

4.2.11: Machados de Assis

Rio de Janeiro, 1839 – Rio de Janeiro, 1908

Dentre os autores aqui estudados, nenhum dispensaria apresentações tanto quanto Joaquim Maria Machado de Assis, homem de letras tomado por muitos, corretamente ou não, como o maior expoente da literatura brasileira.

Oriundo da então capital do Império brasileiro, seus pais eram figuras de pouco renome, ele um brasileiro negro, ela uma portuguesa, possivelmente branca, nascida nos Açores. Sendo fruto de uma família pobre, pouco teve em termos de educação formal e diferente de muitos dos autores aqui estudados, não teve formação superior. Ainda assim, teria tido professores informais, sob os quais estudou os clássicos e aprendeu francês (CALDWELL, 1970). Sua carreira como escritor se iniciou ainda na adolescência com a publicação de poesias na revista

Marmota Fluminense. Pelos anos que se seguiram, se envolveu com o jornalismo e a produção teatral, sendo admitido em 1862 para a o Conservatório Dramático Brasileiro, órgão imperial de promoção do teatro (*ibid.*). A relação entre Machado de Assis e o estado não se encerrou aí – ao longo de sua vida, ele exerceu diversos cargos públicos e se aproximou até mesmo do Imperador.

Com o avanço de sua carreira como escritor, entrou em contato com outras figuras importantes da literatura brasileira, consolidando amizades e desafetos. Muitos de seus textos foram publicados na *Revista Brasileira*, espaço onde, como se viu anteriormente, a discussão da língua, literatura e nacionalidade brasileiras ocupou espaço central. Além da *Revista Brasileira*, Machado de Assis foi também um dos membros fundadores da Academia Brasileira de Letras, órgão que reuniu em sua formação original figuras como Olavo Bilac, José Veríssimo e João Ribeiro. Morreu em 1908, aos 69 anos.

Dentre as diversas obras escritas por Machado de Assis, se optou por usar *Memórias Póstumas de Brás Cubas*, romance originalmente publicado na *Revista Brasileira* ao longo de 1880. A edição utilizada é aquela digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina a partir da versão publicada pela Editora Nova Aguilar em 1994.

4.2.12: Manuel Antônio de Almeida

Rio de Janeiro, 1830 – Macaé, 1861

Médico carioca e filho de um tenente morto quando o futuro escritor contava 11 anos. Coursou a Faculdade de Medicina do Rio de Janeiro a partir de 1849 e em 1852 começou a publicar em partes seu único romance, *Memórias de um sargento de milícias*. Operou cargos públicos na Tipografia Nacional e na Fazenda Real do Rio de Janeiro. Sua vida se encerrou num naufrágio em 1861, quando pretendia iniciar carreira política.

Usou-se aqui sua única obra completa, *Memórias de um sargento de milícias*, em edição digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina a partir da edição original da Tipografia Brasiliense, publicada em 1854 (SERPA, 1945).

4.2.13: Manuel Benício

Pernambuco, 1861 - ?, ?

Filho de militar, estudou um ano na Faculdade de Direito de Recife, largando o curso incompleto e passando a cursar a Escola Militar do Rio de Janeiro. Largou a academia militar, assumindo então cargo público no Rio de Janeiro. Como jornalista, serviu como correspondente de guerra para os jornais *O Tempo* e *Jornal do Commercio*. Foi em seu período nesse segundo que realizou trabalho jornalístico sobre a Guerra de Canudos, obra que inspiraria o livro *O Rei dos jagunços*, o qual aqui se utiliza (BLAKE, 1902). A edição utilizada é a fac-similar publicada pelo Senado Federal em 1997 e digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina.

4.2.14: Vicente Félix de Castro

Areias, 1822 – Silveiras, 1878

Sacramento Blake, no *Diccionario Bibliographico*, coloca Areias como local de nascença do “pai do romance paulista”. Filho de comerciantes letrados do Vale Paraíba, sua família se envolveu com o partido liberal e a revolução de 1842 e se viu punida ao fracasso dessa (CASTRO, 1878). O evento seria marcante na vida de Vicente Félix e sua ligação com o liberalismo permaneceria por toda a vida.

Sua carreira como escritor se iniciou em 1859, com publicações de folhetins. Em 1861, publicou “Os mistérios da Roça”, editado em Guaratinguetá. A década de 1860 será uma de muitos frutos para Vicente Félix de Castro, sendo nela publicados ainda dois outros romances, o mais tardio dos quais, “História de um voluntário paulista”, de 1869, foi aqui tomado como representante de sua obra. Sua carreira literária continuou por toda sua vida, sendo o seu último romance, “Os homens de sangue ou os sofrimentos da escravidão”, publicado meros cinco anos antes de sua morte. Como o título deixa claro, a escravidão é o seu tema central, assunto sem dúvidas muito caro à intelectualidade do Império Tardio.

A edição de sua obra aqui utilizada é a de 1871, publicada pela Tipografia Bananalense, disponível na coleção Mindlin.

4.2.15: Xavier Marques

Itaparica, 1861 – Salvador, 1942

Segundo ocupante da cadeira 28 da Academia Brasileira de Letras, político em certos momentos, jornalista em outros, mas eternamente um escritor, Xavier Marques foi um típico homem de letras de seu tempo, a despeito da humilde origem de filho de barqueiro.

Foi na Bahia que cursou o ensino básico e lá onde começou sua carreira como jornalista, publicando textos no *Jornal de Notícias* e no *Diário da Bahia*. Sua produção literária se inicia na década de 1880 e prossegue continuamente até sua morte, compreendendo prosa, poesia, ensaios e crítica, além de um texto tardio sobre a língua nacional. Esse, “A cultura da língua nacional”, é texto de cunho mais político do que científico, sendo alimentado pelo sentimento nacionalista em voga nos anos 1930.

Além de jornalista e escritor, Xavier Marques foi também político, tendo sido eleito como deputado estadual pela Bahia, em 1915, e como deputado federal em 1921. Deixa a vida política para trás após esse segundo mandato, retomando a carreira jornalística no *Ilustração Brasileira*, publicando nesse comentários políticos que desvelam seu liberalismo (RAMOS JR, 2006).

De sua obra, se utilizou aqui a coletânea *Simples Histórias*, em edição original, publicada em 1886 pela Tipografia da Bahia e digitalizada pelo Núcleo de Pesquisas em Informática, Literatura e Linguística da Universidade Federal de Santa Catarina.

5: APRESENTAÇÃO DOS DADOS

Nesse capítulo são apresentados os dados utilizados, assim como os critérios e métodos aplicados para sua obtenção e classificação e as razões práticas e teóricas que justificam tais decisões.

5.1: PREDITORES EXTRALINGUÍSTICOS

Dada a imensa extensão territorial do Brasil e a complexa realidade histórica da formação de suas diversas regiões, se viu necessária a delimitação de zonas de interesse para essa pesquisa. São dois os intuitos centrais de tal divisão: (i) contemplar a realidade específica de cada uma das regiões de interesse e como essa se relacionaria com os usos da variável linguística; (ii) reduzir o escopo da pesquisa aqui proposta a uma escala factível tanto em termos de disponibilidade de material base quanto para os limites temporais disponíveis para a realização da presente pesquisa. Com base nesses critérios, se optou pelo estabelecimento de uma divisão geopolítica bidimensional, uma de natureza macrorregional e outra de natureza provincial. Em termos da divisão macrogeográfica, se decidiu pelo estabelecimento de dois eixos, nordeste e sudeste, sendo esses compostos, no nível provincial, por Bahia e Pernambuco (nordeste) e São Paulo, Minas Gerais e Rio de Janeiro (sudeste).

A divisão aqui proposta se dá por razões primariamente históricas e políticas. Desde o século XVIII, o discurso vigente sobre o Brasil colocava as então capitanias aproximadamente equivalentes a essas regiões como centros da vida pública e econômica da nação. Cláudio Manuel da Costa, em seu poema épico “Vila Rica”, lista quatro como as capitanias de maior importância no Brasil: Bahia, Pernambuco, Rio de Janeiro e São Vicente, sendo essa última então equivalente ao território de Minas Gerais e São Paulo (COSTA, 1773). Em termos populacionais, Botelho (2001) aponta como, desde 1808, essas cinco províncias eram as mais populosas do Brasil, concentrando mais da metade da população brasileira. Já no censo de 1823, esse número cresce para mais de 60%, o que se mantém no censo geral do Império, de 1872 (BOTELHO, 2001). Segundo o IBGE, ainda hoje essa realidade segue inalterada, como visível nos dados do censo de 2010, onde o sudeste aparece concentrando 42,1% da população total, enquanto o nordeste abriga 27,8%. O sul, em terceiro lugar, tem pouco mais da metade da

representatividade do nordeste, com 14,4%. Se observados os estados em separado, se verá que a maior parte da população brasileira, em 1872, se concentrava, respectivamente, em Minas Gerais (20,5%), Bahia (13,9%), Rio de Janeiro (10,7%), Pernambuco (8,5%) e São Paulo (8,4%), situação que se mantém até 1900, quando o Rio Grande do Sul ultrapassa Pernambuco. Ainda assim, não será o suficiente para reverter a hegemonia das cinco províncias tradicionais, que, em 2010, concentram 52,2% da população brasileira total (IBGE, 2012).

Em termos históricos, a divisão macrorregional representa dois momentos da estrutura de poder político e econômico na sociedade colonial brasileira, um que se prolonga do início da colonização até meados do século XVIII e outro que vai desse ponto até o presente. No período inicial de colonização, durante aquilo que se convencionalmente compreende como o Ciclo do Açúcar, o nordeste concentrava o capital econômico, sendo Pernambuco e Bahia os principais polos de produção açucareira (PRADO, 1970 [2004]). Junto ao poder econômico, vinha o poder político, e nesses séculos iniciais esse se concentrava pelo nordeste. A situação começa a se alterar no período das Invasões Holandesas no século XVII, conflito que forçaria a coroa Portuguesa a mudar seus planos para a grande colônia americana. A partir do século XVIII, com a descoberta do ouro no sudeste, o centro das atenções da colônia migra para o sudeste, resultando no estabelecimento de Minas Gerais, em 1720, e na mudança da capital para o Rio de Janeiro, em 1763 (*ibid.*). É o começo do Ciclo do Ouro. A mudança na estrutura geopolítica colonial vem acompanhada de uma mudança no fluxo migratório, levando ao despovoamento de regiões do nordeste e à maior entrada de estrangeiros no sudeste (*ibid.*). Junto aos migrantes, vai sua língua, e é a partir desse momento que o português começa a ser mais ouvido em Minas Gerais (RODRIGUES, 1996).

Assim, o estabelecimento do eixo macrorregional sudeste-nordeste visa contemplar as diferenças resultantes dos diferentes momentos e modos de colonização das duas regiões. Por sua vez, o eixo microrregional tenta captar as especificidades de cada uma das áreas que compõem o eixo macrorregional, considerando as nuances de suas respectivas formações e funções dentro do universo brasileiro. Ainda que Minas Gerais, São Paulo e Rio de Janeiro componham a parte do sudeste brasileiro aqui contemplada, cabe às duas primeiras áreas um papel de produtor de exportações, seja o ouro no XVIII ou o café no XIX, e uma variante linguística própria, o caipira. Ao Rio de Janeiro, em contrapartida coube a partir do XVIII o papel de centro político e cultural da nação, sediando o governo colonial, e a família real durante sua fuga, e as principais instituições culturais do Império, como o Colégio Pedro II, o Arquivo Imperial, a Biblioteca Fluminense e o IHGB. Não obstante, seria óbvio para qualquer falante

nativo de português que as línguas faladas nesses três estados teriam peculiaridades próprias, tanto em termos de sua pronúncia quanto de seu léxico e mesmo de sua sintaxe.

A categoria autor visa meramente contemplar diferenças estilísticas que estariam presentes nos textos dos diferentes indivíduos cujo material foi empregado na obtenção dos dados dessa pesquisa. Junto a essa, foram incluídas as categorias ano de nascimento e ano de publicação. A primeira se baseia em hipótese similar à apresentada por Galves *et al.* (2006) em artigo analisando construções sintáticas do português seiscentista no qual as autoras argumentam que a utilização do ano de nascimento do autor serviria para uma importante chave para a avaliação de fenômenos históricos, uma vez que estaria mais fortemente associado à norma adquirida. Kato (2000) apresenta um argumento similar. O ano de publicação foi considerado por razões filológicas e históricas. O século XIX foi o início da era editorial brasileira e a possibilidade de que certas normas e tradições de editoração acabem por influenciar as práticas linguísticas dos autores não pode ser completamente descartada. Um passo além seria inserir como fator de influência o local de publicação ou a companhia editorial envolvida, porém tais fatos exigiriam um *corpus* maior e que possivelmente não se conseguiria obter.

5.2: PREDITORES LINGUÍSTICOS

Em termos de linguísticos, os seguintes preditores foram selecionados para a pesquisa: (i) unidade precedente, (ii) unidade seguinte, (iii) tipo de sentença, (iv) número, (v) tempo, (vi) ocorrência de interpolação e (vii) categoria semântica do verbo principal.

Como visível nas categorizações listadas, os dados não foram classificados quanto a aspecto, decisão tomada por razões meramente metodológicas. O modelo atual de classificação se divide entre seis preditores linguísticos e cinco preditores sociais para um total de onze preditores. Embora modelos diferentes possam permitir a combinação de séries de fatores, a adição de mais categorias exigiria um número maior de dados para garantir a confiabilidade do sistema geral. Assim, se escolheu priorizar as categorias listadas acima em detrimento de outras. Apesar disso, não há dúvidas de que uma versão mais completa desse estudo poderia (ou mesmo deveria) incluir tais fatores. Essa tarefa, porém, caberá a trabalhos futuros uma vez que, embora

importantes, os fatores linguísticos são secundários aos interesses atuais. Embora os fatos de ordem linguística sejam claramente importantes, os fatores sociais são a prioridade desse estudo.

As categorias (i) e (ii) referem-se às unidades linguísticas (palavras) imediatamente anteriores e seguintes à uma ocorrência da variável em foco. Considerou-se como “palavra” qualquer unidade morfosintática livre. Esse critério foi aplicado especificamente para que pronomes oblíquos pudessem ser tratados como palavras, visto sua importância na discussão sobre o português no mil e oitocentos e a alta liberdade de colocação dessas unidades.

Essas duas classes foram utilizadas pois abordagens sequenciais de análise linguística têm sido empregadas com considerável sucesso no campo do processamento de língua natural para identificação de padrões formulaicos de uso (*vide* QUINNOU, CELLIER, CHARNOIS & LEGALLOIS, 2012, CONDE-SILVESTRE, 2016). Uma lógica similar é utilizada na análise de variáveis linguísticas, onde o contexto precedente e seguinte influencia a realização (ou não) de determinada variante (*vide* OUSHIRO, 2016). A extensão dessa linha de pensamento para o plano sintático pode, em princípio, parecer questionável e um contra-argumento básico poderia partir da diferença dos modos de execução, isso é, dizer que a influência do contexto na realização de um ou outro som se dá por razões articulatórias que facilitam ou dificultam a execução de certos movimentos enquanto um processo análogo não existiria na sintaxe. Todavia, análises probabilísticas oferecem um campo frutífero para a discussão linguística, fato sobre o qual Manning (2003) disserta bem ao apontar a considerável flexibilidade das estruturas linguísticas e como análises probabilísticas têm sido amplamente empregados no campo de processamento de linguagem natural.

Ambos os fatores, *unidade precedente* e *unidade seguinte*, seguem a mesma classificação tipológica categorial, sendo essa: artigo, adjetivo, advérbio, preposição, pronome, pronome oblíquo, substantivo e verbo. Originalmente, houve uma separação tríplice dos dados do tipo pronome entre pronomes, pronomes possessivos e pronomes oblíquos. As categorias pronome e pronome possessivo, porém, foram fundidas numa única categoria de pronome. O objetivo da junção foi fortalecer o contraste com a categoria pronome oblíquo, a qual é de interesse particular para a discussão apresentada.

Tal interesse particular provém do estatuto particular que o pronome oblíquo têm na consciência linguística dos falantes do XIX. A chamada “polêmica da colocação pronominal” foi tema recorrente no discurso de diversos autores do período ao ponto de que alguns declarassem que bastava a colocação do pronome para que se identificasse onde nasceu ou tinha

estudado certo autor. Mais detalhes dessa discussão podem ser encontrados no capítulo 3.3. De um ponto de vista teórico, a posição do pronome oblíquo em relação ao verbo, nos casos onde esse ocorra, poderia acabar resultando num efeito de ativação que resultaria numa maior chance de seleção de eGER ou eIP. Efeitos similares têm sido observados em estudos de percepção sociolinguística, onde elementos linguísticos ou não linguísticos exercem efeito no modo como certas variantes e indivíduos são avaliados e realizados (*vide* HAY & DRAGER, 2010). Do outro lado, estudos no campo da cognição têm observado que padrões de seleção de palavras são afetados por termos anteriores (*vide* JURAFKSY *et al.*, 2001). Partindo dessas duas visões, a inclusão dos fatores (i) e (ii) se fez necessária, fato que acabou e provando significativamente importante na fase de análises inferenciais.

Tipo de sentença, a categoria de análise (iii), refere-se ao tipo sintático da oração em que se encontra cada ocorrência. Barbosa (1999) e Mothé (2007) utilizam ambos um modelo de análise sintática tradicional em suas análises, tendo encontrado um maior favorecimento do uso de estruturas com infinitivo em sentenças dos tipos **temporal**, **condicional** e **modal**. Essa categoria se divide em:

(i) Absoluta

“Sinhá **está se fiando** muito naquela sonsa” (Bernardo Guimarães)

(ii) Adjetiva

“Mas o pior é que o intestino anda funcionando meio a matroca de tudo, uma sensação de cansaço pelo corpo todo, essa sensaçõzinha amolante e gostosa de um corpo que **está pedindo** cama ou rede, que é melhor” (Amadeu Amaral)

(iii) Aditiva

“Papai o jantar está na mesa e mai Anna **está perguntando** si o doente quer um caldo” (Carneiro Vilela)

(iv) Adversativa

“Tenho vontade de tentar a sorte na chapada nova mais a mulher **está repunando**” (JLR)

(v) Apositiva

“Não e difícil adivinhar o teu pensamento neste instante: com toda a certeza **estás pensando** no que acabaste de ler” (Luísa Carneiro Pacheco)

(vi) Assindética

“Ela saiu, ele limpou os olhos, **esteve algum tempo a pensar** em coisas distantes, depois com um suspiro voltou ao seu catalogo” (Júlia Lopes de Almeida)

(vii) Comparativa

“Felix Modesto observava tudo com uma atenção profunda refletida como si **estivesse estudando** o terreno e reunindo os dados precisos para resolver algum problema rural” (Carneiro Vilela)

(viii) Completiva

“És capaz de adivinhar, querida Alice, em que e que **estou pensando** agora?” (Lúcia Carneiro Pacheco)

(ix) Condicional

“Se eu **estivesse a escrever** um romance em vez de um unico capitulo da vida real contaria minudentemente ao leitor as luctas em que a pobre Zina debateu se e triumphou” (Xavier Marques)

(x) Conformativa

“A nota porém característica que os distingue é ter o queixada o queixo branco, como **está vendo**” (Júlio Ribeiro)

(xi) Explicativa

“A questão me interessava porque **estou tratando** de redigir uma reclamação para a imprensa contra certas irregularidades do serviço”
(Amadeu Amaral)

(xii) Final

“E olhe o que lhe digo se não **está pegando** fogo, está se ajuntando lenha para isso” (Manuel Antônio de Almeida)

(xiii) Objetiva

“Adeus, senhor Leôncio, eu me retiro porque seria altamente inconveniente indigno e ridículo da minha parte **estar a disputar** com o senhor por amor de uma escrava” (Bernardo Guimarães)

(xiv) Principal

“**Estou só sentindo** que sou o alvo da atenção e da magnanimidade de alguém” (Júlia Lopes de Almeida)

(xv) Predicativa

“dize-me cá em que quarto **está dormindo**” (Júlia Lopes de Almeida)

(xvi) Temporal

“Com efeito, um dia de manhã, **estando a passear** na chácara, pendurou se me uma ideia no trapézio que eu tinha no cérebro”
(Machado de Assis)

Das categorias sintáticas apresentadas, diversas apresentaram uma quantidade extremamente baixa de ocorrências sendo descartadas das análises estatísticas, assim como os dados referentes a essas. As categorias mantidas ao final foram: Absoluta, Adjetiva, Aditiva, Assindética, Completiva, Explicativa, Objetiva, Principal, Predicativa e Temporal.

Número (iv) e tempo (v) são categorias referentes tão e tão somente ao número do verbo (se singular ou plural) e ao tempo em que esse se dá (presente, pretérito e futuro). Dentro

do *corpus* houve poucas ocorrências de dados no futuro de modo que essas foram eliminadas do conjunto para evitar que sujassem as análises.

A categoria (vi) interpolação se refere à ocorrência ou não de material linguístico entre o auxiliar e o verbo principal. Tal categoria foi incluída em resposta à ocorrência de dados como os seguintes:

- a) Nhê-nhá *está **lhe** chamando* cá pr'a dentro
- b) E ela *está **ali** dançando*
- c) Que diabo *está **o Sr.** a ler?*

Nos três exemplos acima, se percebe a interpolação de elementos de diferentes classes à estrutura perifrástica, respectivamente um pronome oblíquo, um advérbio e um sujeito. Como visível nos casos (b) e (c), a ocorrência de interpolação pode incluir também inversão de ordem sintática, movendo elementos em direção à margem esquerda (b) ou direita (c).

Casos como (a) e (b), contudo, devem receber uma certa ressalva, uma vez que a estruturas eGER presentes nessas poderiam não configurar um caso de perífrase verdadeira. Veja-se que na sentença (b) o segundo verbo (dançando) poderia ocupar apenas um papel de complemento do verbo principal, isso é, uma leitura como a seguinte seria possível:

- (a) Ela *está ali* [dançando]

Se tal leitura for considerada a leitura primária, não se poderia tomar tal sentença como um caso interpolado. De fato, não se poderia sequer considerar ela um caso de perífrase verdadeira, mas apenas uma sentença superficialmente similar a essa. Todavia, mesmo que o segundo verbo seja tomado como um complemento e não como o verbo principal da sentença, ele poderia ainda assim ser substituído por um IP, resultando em algo da ordem de (a):

- a) Ela *está ali* [a dançar]

Não apenas isso, mas a mesma sentença poderia também ser traduzida para uma forma não-perifrástica como a seguinte:

- a) Ela [dança] ali.

O último caso listado, embora em primeira vista estranho, se revela próximo aos anteriores em casos hipotéticos como os seguintes:

- a) (Alice, chegando a um lugar e vendo alguém que não esperava ver) O que você faz aqui?
- b) (Rin, ouvindo alguém dizer algo que ela estranha) O que é isso que você me diz?

Tanto em (a) em (b), seria possível substituir os verbos por perífrases de Gerúndio ou IP sem que se perca o sentido central, assim se obtendo sentenças como as seguintes:

- a) (Alice, chegando a um lugar e vendo alguém que não esperava ver) O que você **está fazendo/a fazer** aqui?
- b) (Rin, ouvindo alguém dizer algo que ela estranha) O que é isso que você **está me dizendo/está a me dizer**?

Assim, estruturas com interpolação são, em resumo, estruturas que podem ser lidas não como casos de perífrase *stricto sensu*, mas que se comportam como estruturas perifrásticas com inversão.

Por fim, o grupo de fatores (vii), *tipo semântico do verbo*, opera com as categorizações apresentadas por Halliday (2004). Se considerou inicialmente utilizar as categorias empregadas por Vendler (1967), porém, devido à natureza perifrástica da variável utilizada, o modelo de

Vendler, mais ligado ao *aspecto verbal*, e à utilização anterior do modelo de Halliday por Mothé (2006, 2007), esse modelo foi preterido.

O modelo proposto por Halliday conceitualiza a língua em termos de **processos**, os quais representam os diferentes tipos de eventos e experiências sensíveis pelos quais uma entidade pode passar (HALLIDAY, 2014). Essas diferentes estruturas se desenvolvem a partir da percepção do indivíduo quanto a fenômenos ocorridos no mundo e sobre si mesmo e se dividem em seis categorias: materiais, comportamentais, mentais, verbais, relacionais e existenciais.

Processos do tipo material se referem a eventos ocorridos a ou ações tomadas sobre entidades do mundo, ainda que essas não sejam entidades necessariamente materiais. Trata-se de processos que ocorrem a partir de um ponto externo. Em uma sentença como “A Britânia foi conquistada pelos francos”, a construção “foi conquistada” é material pois representa uma ação tomada por um agente (os francos) sobre uma entidade (a Britânia). Processos materiais não necessariamente pressupõem um agente que opera sobre um paciente de modo que em uma sentença como “aquela pessoa caiu”, o verbo “caiu” também seria classificado como um processo material uma vez que é um evento externo ocorrido sobre um paciente. Ex:

Sinhazinha, ponderou a mucama, o que lhe fez mal foi o cheiro forte do vidro que vassunce estava pondo na roupa; a mim também me tonteou (B. Guimarães)

Processos mentais seguem uma direção oposta e representam acontecimentos “de dentro para fora”, isso é, processos internos do paciente, como o verbo “querer” em uma sentença como “Eu não quero dinheiro”. Processos desse tipo representam o “mundo interno” e a experiência subjetiva e emotiva de uma entidade.

O senhor está estranhando alguma coisa na minha pessoa? (A. Amaral)

Processos relacionais representam estruturas que demonstram relações entre fatos e entidades, servindo para ligar unidades de sentido. O verbo “ser” numa sentença como “todos

os homens são mortais” indicam uma relação ao aplicar o predicado “mortais” ao sujeito “todos os homens”.

Ao domingo ou se estiver chovendo ou na roça em qualquer ocasião de lazer pode tentar ver se o soneto sai (M. de Assis)

Processos comportamentais, por sua vez, descrevem manifestações (em geral físicas) de processos internos, sendo, portanto, algo que existe entre a categoria dos processos mentais e materiais. Algo como “rir” em “a mulher riu do idiota” é comportamental ao ser a manifestação material (um ato físico) de um processo mental (um sentimento ou emoção).

Bem estou vendo (F. Távora)

Processos verbais são indicativos de ações de fala e expressam atos comunicativos de entidades. Um exemplo de um processo desse tipo seria o verbo “falar” em uma sentença como “me fale sobre o seu dia”.

Rita, que estivera a conversar com o positivo na porta, voltou-se a uma interjeição da senhora (L. J. Rocha)

Por fim, processos existenciais recobrem sentenças e verbos usados para indicar a existência de uma entidade ou fenômeno, como, por exemplo, “tem” em uma sentença como “tem muitas pessoas que gostam de filmes de ação”.

A classificação segundo os tipos de processo de Halliday, porém, não pode ser considerada inerente a um tipo de verbo, mas sim a uma construção ou sentido, uma vez que

um mesmo verbo pode, em condições diversas, ser representativa de processos diferentes. Assim, um verbo como *ter* pode surgir em uma construção do tipo mental como “*Eu tenho medo de fantasmas*” ou em uma sentença material como “*Eu tenho dinheiro*”.

Para a classificação do tipo semântico não se considerou apenas o verbo principal, mas o valor do complexo perifrástico como um todo, isso é, de toda a unidade estar + gerúndio ou estar + IP.

5.3: DISTRIBUIÇÃO DE DADOS POR AUTOR

Observemos em primeiro lugar a tabela acima. Nela, se vê a distribuição dos dados de acordo com autor. As colunas da esquerda mostram os números totais de ocorrências, enquanto as colunas da direita apresentam o percentual dessas.

TABELA 1: DISTRIBUIÇÃO DE DADOS POR AUTOR

Autor	ESTADO	GER	GER (%)	IP	IP (%)	TOTAL
Inês Sabino	BA	10	90.9	1	9.1	11
Joaquim Felício dos Santos	BA	9	100	0	0	9
Xavier Marques	BA	6	46.2	7	53.8	13
Bernardo Guimarães	MG	56	77.8	16	22.2	72
Júlio Ribeiro	MG	36	90	4	10	40
João Lindolfo Rocha	MG	87	92.6	7	7.4	94
Carneiro Vilela	PE	22	75.9	7	24.1	29
Franklin Távora	PE	59	95.2	3	4.8	62
Manuel Benício	PE	23	88.5	3	11.5	26
Júlia Lopes de Almeida	RJ	33	82.5	7	17.5	40
Machado de Assis	RJ	13	46.4	15	53.6	28
Manual A. de Almeida	RJ	24	85.7	4	14.3	28
Amadeu Amaral	SP	9	64.3	5	35.7	14
Luísa de Camargo Pacheco	SP	17	100	0	0	17
Vicente Félix de Castro	SP	10	90.9	1	9.1	11
TOTAL		414	83.8	80	16.2	494

Um primeiro dado que se deve notar é a considerável disparidade nos totais de dados – um terço dos autores (João Lindolfo Rocha, Bernardo Guimarães, Franklin Távora, Júlia Lopes de Almeida e Júlio Ribeiro) concentram 62% das ocorrências. Tal fato não se dá simplesmente devido a uma maior extensão das obras analisadas uma vez que a obra de Carneiro Vilela é a mais longa dentre todas, porém, ainda assim apresenta menos de um terço

das ocorrências identificadas na obra de João Lindolfo Rocha. Há uma clara diferença estilística entre os autores, alguns dos quais tendem a usar essa variável muito mais do que outros.

Como é visível nos dados acima, as ocorrências de eGER contemplam a vasta maioria dos dados, um total de 414 (83,8%) de 494 (de modo que eIP representa 16,2% do total). Tal padrão é consideravelmente diferente daquele que Mothé (2007) encontra no século XIX e XX no Brasil, cerca de 10% dos dados, assim como é maior também que a taxa encontrada por ela nos dados escritos de Portugal (também cerca de 10%). Contudo, como tende a ser costumeiro quando se tratando de qualquer avaliação numérica, médias ofuscam uma complexidade subjacente significativamente maior. Embora o padrão geral apresente uma diferença extremo de uso do eIP, as divergências entre os usos de autores específicos é notável.

GRÁFICO 1: DISTRIBUIÇÃO GERAL DE DADOS POR AUTOR

O gráfico 1 mostra com mais clareza o grau de divergência entre autores específicos. Nele, as partes escuras das barras representam ocorrências de eGER, enquanto as partes claras representam ocorrências de eIP. A largura das barras é ajustada de acordo com o número de ocorrências total dentro do *corpus* de modo tal que as barras mais largas representam um número maior de ocorrências do que as mais estreitas.

Chama a atenção, em primeiro lugar, as barras referentes às obras de Joaquim Felício de Souza e de Lúisa de Camargo Pacheco, ambas apresentando totalidade de ocorrências de

eGER, um caso extremo que não tem reflexo na direção oposta – não há, afinal, qualquer autor que apresente um uso categórico de eIP. Todavia, ambos os conjuntos são relativamente pequenos, sendo que João Felício dos Santos representa o menor número de ocorrências dentro do *corpus* com apenas nove casos totais. Luísa de Camargo Pacheco apresenta um número maior, mas ainda assim abaixo da média geral de ocorrências, com um total de dezessete casos.

Em contrapartida a esses dois, há os casos de Xavier Marques e Machado de Assis, ambos apresentando emprego de eIP a uma taxa de mais de 50% (53,8% e 54,6%, respectivamente). Comparados ao padrão geral, esses são dois casos extremos que seguem numa direção completamente oposta àquela de todos os outros autores. Mais desconcertante é o fato de que Xavier Marques é um autor oriundo do nordeste, região que, como se verá adiante, emprega significativamente menos eIP do que o sudeste. Não apenas isso, mas a remoção dos dados referentes a esses dois autores reduz a taxa global de emprego de eIP de 16,2% para 12,8%. Em nota similar, a remoção dos conjuntos com uso categórico de eGER move o padrão geral para cima, para 17,1%. Se as ocorrências referentes aos quatro autores são removidas, esse número cai para 13,6%.

Uma análise estatística mais apurada, porém, revela uma situação interessante. A análise abaixo, assim como todas as outras apresentadas nesse trabalho, foram realizadas utilizando o pacote R (R Development Core Team, 2008) usando o RStudio (RStudio Team, 2016) e o pacotes lmerTest (KUZNETSOVA, BROCKHOFF & CHRISTENSEN, 2017), RMS (HARREL JR, 2018), dplyr (WICKHAM *et alli*, 2018) e effects (FOX, 2003), enquanto os gráficos foram desenhados, parcialmente, utilizando o pacote sjPlot (LUDECKE, 2018).

TABELA 2:REGRESSÃO LOGÍSTICA POR AUTOR

<i>lrm(formula = Tipo de ocorrência ~ Autor)</i>				
	Coef	S.E.	Wald Z	Pr(> Z)
Intercept	-0.5878	0.5578	-1.05	0.292
B. Guimarães	-0.665	0.6257	-1.06	0.2879
C. Vilela	-0.5573	0.7067	-0.79	0.4303
F. Távora	-2.3911	0.8133	-2.94	0.0033**
I. Sabino	-1.7148	1.1879	-1.44	0.1489
J. F. do Santos	-9.3397	47.715	-0.2	0.8448
J. L. Almeida	-0.9628	0.6959	-1.38	0.1665
J. Ribeiro	-1.6094	0.7674	-2.1	0.036*
L. C. Pacheco	-9.3397	34.7199	-0.27	0.7879
J. L. Rocha	-1.9322	0.6822	-2.83	0.0046**

M. de Assis	0.7309	0.6743	1.08	0.2784
M. A. de Almeida	-1.204	0.7764	-1.55	0.121
M. Benífico	-1.4491	0.8294	-1.75	0.0806
V. F. de Castro	-1.7148	1.1879	-1.44	0.1489
X. Marques	0.7419	0.7878	0.94	0.3463

Dos quinze autores que compõem o *corpus* de trabalho atual, apenas três foram selecionados como causando uma diferença estatisticamente relevante do valor médio do grupo num alfa de 0.05 – Franklin Távora, Júlio Ribeiro e Lindolfo Rocha. Nos três casos, o coeficiente apresentado é negativo, indicando que os dois autores apresentam uma tendência relativamente menor de emprego de eIP. Os três são também aqueles com os maiores coeficientes negativos, salvo por Lúcia Carneiro Pacheco e Joaquim Felício dos Santos. Os dois últimos apresentam uma distribuição igual de dados com um uso categórico de eGER, o que, somado ao número relativamente baixo de dados apresentados pelos dois, acaba por resultar em um erro padrão consideravelmente alto. Mais interessante que isso é que os dois autores com padrão mais distoante, Machado de Assis e Xavier Marques, não tenham sido considerados estatisticamente relevantes, apesar de seu coeficiente levemente positivo, único entre os autores. Em todo caso, é preciso ressaltar que os autores selecionados como estatisticamente relevantes estão entre aqueles que apresentaram o maior número de ocorrências. Embora os padrões apresentados aqui sejam estatisticamente relevantes dentro de certa escala, não se pode deixar de questionar se haveria mudanças caso um número maior de dados pudesse ter sido obtido de cada autor. Ainda que isso escape ao escopo da atual pesquisa, uma expansão interessante seria tentar reproduzir as análises aqui realizadas, mas com uma distribuição igualitária dos dados, uma vez que isso eliminaria penalizações aplicadas por diferenças no total de produções por autor. Se é suposto que as taxas gerais de cada autor se manteriam iguais, é altamente provável que os valores de significância mudassem nesse cenário.

Levar tais fatores em consideração levanta uma consideração metodológica geral, mas com repercussões específicas para estudos do tipo presente. Não se questiona aqui que o uso de métodos estatísticos para a validação de análises é um fator central à toda pesquisa de *corpus*, porém, mesmo uma leitura cursória das informações antes apresentadas indica que uma confiança absoluta nesses, ou, mais gravemente, apenas nos valores de significância, induziria a erros de interpretação. O que a análise apresentada anteriormente indica é tão e tão somente que dentro do conjunto de dados apresentados, tais e tais distribuições são significativamente diferentes de outras. Contudo, isso não significa, pura e simplesmente, uma verdade sobre os

autores, mas apenas sobre o que aqui se usou deles. Como dito antes, uma mudança no *corpus* poderia, sim, trazer algumas mudanças. Para exemplificar isso, se decidiu por extrair dados de *Dom Casmurro*, em edição digitalizada disponibilizada pelo grupo de estudos de literatura e linguística da Universidade Federal de Santa Catarina. Nessa, foram encontradas quarenta e seis (46) ocorrências de eGER contra vinte e duas (22) ocorrências de eIP, uma diferença que se mostra significativa ($p < 0.01$) em relação aos dados presentes em Memórias Póstumas, embora não altere os resultados de significância de modo algum, como visível no teste de regressão logística abaixo:

TABELA 3: MODELO DE REGRESSÃO LOGÍSTICA POR AUTOR (DOM CASMURRO)

<i>lrm(formula = Tipo de ocorrência ~ Autor)</i>				
Intercept	Coef	S.E.	Wald Z	Pr(> Z)
B. Guimarães	-0.5878	0.5578	-1.05	0.2920
C. Vilela	-0.6650	0.6257	-1.06	0.2879
F. Távora	-0.5573	0.7067	-0.79	0.4303
I. Sabino	-2.3911	0.8133	-2.94	0.0033**
J. F. do Santos	-1.7148	1.1879	-1.44	0.1489
J. L. Almeida	-9.3258	47.3849	-0.20	0.8440
J. Ribeiro	-0.9628	0.6959	-1.38	0.1665
L. C. Pacheco	-1.6.094	0.7674	-2.10	0.0360*
J. L. Rocha	-9.3258	34.4797	-0.27	0.7868
M. de Assis	-1.9322	0.6822	-2.83	0.0046**
M. A. de Almeida	0.7309	0.6743	1.08	0.2784
M. Benício	-1.2040	0.7764	-1.55	0.1210
V. F. de Castro	-1.7148	1.1879	-1.44	0.1489
X. Marques	0.7419	0.7878	0.94	0.3463

Como os dados mostram, embora a diferença dentro de um determinado autor exista, essa não altera o padrão geral. Ainda assim, fica claro que a norma anteriormente vista em Machado de Assis com um uso majoritário de eIP não é absoluta dentro da obra do autor e a inclusão de mais obras suas ao *corpus* poderia acabar revelando um padrão muito mais próximo do que se observa no todo. Isso mostra que os critérios de seleção de fontes também são um elemento central para uma pesquisa. Sem dúvidas, há problemas resultantes de como os dados aqui apresentados estão distribuídos, problemas esses derivados da organização do *corpus* e de sua montagem. O estabelecimento da regra “uma obra por autor” resulta, necessariamente, em uma disparidade de ocorrências, mas também numa ilusão de regularidade que, como visto, não

necessariamente se sustenta. Pareceria óbvio, portanto, que seria mais desejável ter obtido um número se não equivalente, parecido de ocorrências, ainda que retiradas de obras diversas. Dito isso, também é óbvio que essa abordagem alternativa traria consigo certas dificuldades técnicas, uma vez que a quebra da simetria entre obra e autor obrigaria a incluir novos fatores nas análises, alguns dos quais não seriam facilmente contempláveis por vias estatísticas. Tais são, por exemplo, fatores estilísticos que poderiam causar divergência entre uma e outra obra e onde e quando essas foram editadas. Ainda mais que isso, em se tratando de um estudo de documentos antigos, dificuldades adicionais acabam surgindo uma vez que o número de obras acessíveis é finito e existe a despeito da vontade do pesquisador. Enquanto estudos realizados a partir da produção de falantes, como é o caso em muitos estudos de sociolinguística sincrônica, podem condicionar suas perguntas e táticas de entrevista para a produção de certas unidades, algo especialmente fácil para variáveis fonéticas, as mesmas possibilidades não existem quando todo o material já foi produzido. Para pesquisas de maior profundidade histórica, esse risco é ainda maior. Todas as considerações levantadas até aqui devem ser mantidas em mente também para os resultados subsequentes.

5.4: DISTRIBUIÇÃO DE DADOS POR REGIÃO E ESTADO

Como dito antes, os dados foram divididos espacialmente em dois níveis, um nível regional Nordeste X Sudeste e um nível provincial (Bahia x Minas Gerais x Pernambuco x Rio de Janeiro x São Paulo). Esses dois fatores são centrais para a pergunta de pesquisa aqui proposta. Verifiquemos, primeiro, como os dados se revelam de acordo com a distribuição regional.

TABELA 4: DISTRIBUIÇÃO DE DADOS POR REGIÃO

	eGER	eGER %	eIP	eIP	TOTAL
Nordeste	207	88.1	28	11.9	235
Sudeste	207	79.9	52	20.1	259
TOTAL	414	83.8	80	16.2	494

GRÁFICO 2: DISTRIBUIÇÃO DE DADOS POR REGIÃO

Como visível na tabela 3 a região sudeste apresenta uma taxa de emprego de eIP aproximadamente duas vezes maior do que a região nordeste (20,1% contra 11,9%). Se comparados com os valores encontrados por Mothé (2007) em dados escritos de português brasileiro ao longo dos séculos XIX e XX, esses valores parecem extremamente altos. Veja:

TABELA 5: REALIZAÇÃO DE INFINITIVO GERUNDIVO AO LONGO DO TEMPO,

Tempo	(Cambiáveis + Ambíguos) percentual
Fase 1 (1808-1840)	7%
Fase 2 (1841-1870)	6%
Fase 3 (1871-1900)	3%
Fase 4 (1901 – 1924)	5%
Fase 5 (1925 – 1949)	7%
Fase 6 (1950 – 1969)	3%
Fase 7 (1970 – 1985)	1%
Fase 8 (1986 – 2000)	12%
Total	5%

Adaptado de Mothé (2007, p. 151).

Se ressalve o fato de que os dados acima se referem a todo tipo de construção com *a* + infinitivo, enquanto a taxa de uso de combinação de *estar* + infinitivo gerundivo (o que aqui se chama de eIP) é de 5% (Mothé, 2007, p. 148). Infelizmente, o trabalho não apresenta a evolução temporal da distribuição de dados segundo o tipo de auxiliar, mas mesmo sem esses dados, a diferença é notável, especialmente quando os valores das três primeiras fases, grosso modo equivalentes ao recorte temporal que aqui foi utilizado, são comparados com os valores encontrados para o sudeste. Ainda que se considere a diferença dos materiais usados na

composição dos *corpora* (cartas e anúncios de jornais em Mothé, romances aqui), tamanha disparidade de dados levanta questões, ainda mais quando se mantém em mente os percentuais de emprego de eIP vistos em cada autor particular, os quais no mais das vezes se mantinham acima dos 10%. A hipótese de que tal divergência se daria pelos gêneros textuais utilizados é a primeira que se deveria erguer, mas não se poderia ir muito além quanto a isso, não sem que se recolhesse mais material e se fizessem outras análises, algo que não caberia às questões, métodos e espaço deste trabalho. Novamente, resta apenas deixar a questão para qualquer mente inquisitiva que no futuro queira encontrar sua resposta.

Deixando de lado esse problema e voltando o foco à distribuição encontrada aqui, caberia a pergunta quanto à relevância estatística da diferença entre sudeste e nordeste. Essa, revelada de forma clara na tabela e gráfico anteriores, se confirma numa análise estatística, como se vê a seguir:

TABELA 6: MODELO DE REGRESSÃO LOGÍSTICA POR REGIÃO

	Coef.	S.E.	Wald Z	Pr(> Z)
Intercept	-2.0005	0.2014	-9.94	<0.0001***
Sudeste	0.6190	0.2542	2.44	0.0149**

O que o modelo acima nos diz é que em relação ao nordeste, o sudeste apresenta uma tendência positiva de emprego de eIP com as chances dessa distribuição ser mero acaso muito abaixo da margem de significância de 0.05. Em outras palavras, não se trata de mero acaso, mas de uma diferença real. Ou ao menos o é dentro do escopo do *corpus* utilizado. Em vista disso, algumas hipóteses já poderiam ser levantadas, mas antes de ir a elas, há de se tomar o cuidado de olhar para as distribuições por província.

TABELA 7: DISTRIBUIÇÃO DE DADOS POR PROVÍNCIA

	eGER	eGER%	eIP	eIP%	TOTAL
Bahia	103	87.3	15	12.7	118
M. Gerais	101	83.5	20	16.5	121
Pernambuco	104	88.9	13	11.1	117
R. de Janeiro	70	72.9	26	27.1	96
S. Paulo	36	85.7	6	14.3	42
Total	414	83.8	80	16.2	494

GRÁFICO 3: DISTRIBUIÇÃO GERAL DE DADOS POR PROVÍNCIA

Tal como é o caso das distribuições por região, a por província mostra uma variação significativa entre cada uma das cinco áreas consideradas, sendo Pernambuco a província com o menor índice de emprego (11,1%) em oposição ao Rio de Janeiro, que tem o maior emprego (27,1%). Um teste de regressão logística feito sobre esses dados apresenta o seguinte resultado:

TABELA 8: MODELO DE REGRESSÃO LOGÍSTICA SEGUNDO PROVÍNCIA

	Coef.	S.E.	Wald Z	Pr(> Z)
Intercept	-1.9267	0.2764	-6.97	<0.0001***
M. Gerais	0.3073	0.3692	0.83	0.4052
Pernambuco	-0.1528	0.4036	-0.38	0.7051
R. de Janeiro	0.9363	0.3593	2.61	0.0092**
S. Paulo	0.1349	0.5204	0.26	0.7954

Traduzindo a tabela acima, o *intercept* e o Rio de Janeiro apresentam uma diferença estatisticamente significativa dentro de um alfa de 0.05, tendo o intercept (composto por dados de referentes à Bahia e com ocorrências de eGER) uma tendência negativa de emprego de eIP e o Rio de Janeiro uma tendência positiva. O tamanho severamente reduzido dos dados referentes a São Paulo é indesejável e levemente preocupante, mas terá que ser deixada de lado no momento, apenas se devendo reiterar tudo aquilo que foi dito em referência à escassez de dados de certos autores.

Seja como for, a síntese dos dados para região e dos dados para província nos diz que embora exista uma diferença significativa entre sudeste e nordeste, uma análise que considera as cinco unidades federativas em separado não mostra grandes distinções entre todas elas, mas apenas entre duas. A diferença entre as duas regiões não é surpreendente e, de fato, se encaixa com aquilo que se conhece da história brasileira e sua povoação, com um nordeste colonizado mais cedo que o sudeste. De fato, Rodrigues (1986) aponta que até o século XVIII, as línguas gerais ocupavam a maior parte da vida linguística na região que eventualmente se tornaria São Paulo e Minas Gerais, sendo apenas a partir do projeto das Minas de Ouro que o português falado pelos colonos lusitanos começou a suplantar as línguas americanas. O mesmo, contudo, não se pode dizer do Rio de Janeiro, que já desde antes gozava de importância política dentro do Brasil. Não apenas isso, mas a própria coroação do Rio de Janeiro como capital da colônia e, mais tarde, sua recepção da corte portuguesa no famoso 1808 dão a essa província feições peculiares dentre aquelas do sudeste. Esses fatos somados permitem supor que a norma vigente no Rio de Janeiro seria, possivelmente, diversa daquela que se encontraria no resto do sudeste, algo que os dados comprovam:

TABELA 9: MODELO DE REGRESSÃO LOGÍSTICA DE MINAS GERAIS E SÃO PAULO CONTRA RIO DE JANEIRO

	Coef.	S.E.	Wald Z	Pr(> Z)
Intercept	-1.6619	0.2139	-7.77	<0.0001***
R. de Janeiro	0.6715	0.3139	2.14	0.0324*

O modelo acima nos diz que o padrão de uso de eIP empregado no Rio de Janeiro difere significativamente ($p < 0.05$) daquele empregado em Minas Gerais e São Paulo, sendo a norma carioca mais propensa à utilização de eIP do que as outras duas, tomadas como um único elemento. Essa diferença se acentua quando se contrasta o Rio de Janeiro às outras quatro províncias:

TABELA 10: MODELO DE REGRESSÃO LOGÍSTICA DE RIO DE JANEIRO VS OUTRAS PROVÍNCIAS

	Coef.	S.E.	Wald Z	Pr(> Z)
Intercept	-1.8517	0.1464	-12.65	<0.0001
R. de Janeiro	0.8613	0.2723	3.16	0.0016**

Nesse cenário, a diferença é ainda mais estatisticamente significativa ($p < 0.01$) e o coeficiente de emprego relativamente maior ($\sim 0,67$ no modelo contrastado com Minas Gerais e São Paulo e $\sim 0,86$ no modelo contrastado com todas as outras quatro províncias). Todos esses fatos apontam numa mesma direção: o caso do Rio de Janeiro é peculiar dentro do contínuo brasileiro.

Lido em conjunto com os fatos extralinguísticos conhecidos, a história contada por esses dados é a de uma região cujo estilodifere severamente do resto do Brasil, o que não surpreende, ainda que leve a um paradoxo. No século XIX, o Rio de Janeiro era o centro do poder brasileiro, lar do trono e da alta casta da intelectualidade, totem do qual emanava a norma e as suposições do que seria a nacionalidade brasileira. É no Rio de Janeiro, afinal, que nascem a Academia Brasileira de Letras e o Instituto Histórico e Geográfico Brasileiro, dois órgãos cuja *raison d'être* era tomar os frutos da terra e do homem e com eles forjar o Brasil. Ainda assim, é lá que se encontra a taxa de uso que mais se aproxima daquela vista na antiga metrópole.

É natural que isso traga certas questões e não há aqui qualquer pretensão de que tal leitura seja uma verdade absoluta. Os problemas delineados antes quanto à escolha do *corpus* continuam válidos e nesse momento a sombra do individualismo volta pôr em xeque conclusões que de outro poderiam ser tiradas. Ainda que seja um fato que os dados referentes ao Rio de Janeiro diferem em muito dos apenas um conjunto limitado de dados foi utilizado e apenas um único fato linguístico foi analisado. Assim, o que há aqui é uma primeira pista de que pode existir uma diferença considerável nos padrões gerais do Rio de Janeiro, diferença que não encontra reflexo nas outras regiões do Brasil.

5.5: DISTRIBUIÇÃO DE DADOS POR ANO DE NASCIMENTO DO AUTOR E ANO DE PUBLICAÇÃO

Como explicado anteriormente, os fatores ano de nascimento do autor e ano de publicação da obra foram incluídos nessa pesquisa para que se pudesse averiguar o desenvolvimento das tendências de uso ao longo do período analisado, a hipótese geral sendo que haveria um crescimento do emprego de eIP no começo do século e uma queda desse no fim do mesmo. Uma vez mais, e mais do que em qualquer outro caso, deve se ressaltar que o número de autores utilizados aqui é consideravelmente baixo para esse fim, sendo os dados

apresentados aqui, portanto, meramente ilustrativos. Os dados obtidos segundo o ano de nascimento podem ser visualizados no gráfico a seguir:

GRÁFICO 4: TAXA DE EMPREGO DE EIP POR ANO DE NASCIMENTO

Os valores no eixo vertical indicam a taxa de emprego de eIP numa escala de 0 a 1, onde 0 indica emprego nulo dessa variável e 1 emprego total. Como pode ser visto, há uma variação considerável ao longo do tempo, ocorrendo dois picos consideráveis em 1839 e 1864, datas dos nascimentos de Machado de Assis e Xavier Marques. Ambos os picos se seguem imediatamente por uma queda leve no representante seguinte (Franklin Távora em 1842 e Amadeu Amaral em 1875) e então por uma queda mais drástica (Júlio Ribeiro em 1845 e Lúcia Carneiro Pacheco em 1884). A aplicação de uma regressão logística sobre essa distribuição retorna o seguinte resultado:

TABELA 11: MODELO DE REGRESSÃO LOGÍSTICA SEGUNDO NASCIMENTO

	Coef.	S.E.	Wald Z	Pr(> Z)
Intercept	15.794149	14.019310	1.127	0.260
Nascimento	-0.009440	0.007592	-1.243	0.214

Por extenso, o que isso diz é que embora aparente haver uma redução no emprego de eIP ao longo do tempo (*estimate* ~ -0.009), esse não é um resultado estatisticamente significativo ($p > 0.05$). Outro problema que suja os dados acima é a assimetria de distribuição, com três autores tendo nascido no ano de 1862, fato que não se repete para nenhum outro ano. Certas décadas são também significativamente mais bem representadas do que outras, com a década de 1860 contando quatro autores enquanto às décadas de 1850, 1870 e 1880 são representadas por apenas um. O desequilíbrio nas distribuições é um dos fatores que inevitavelmente acabou danificando a qualidade dos dados e impedindo que resultados significativos fossem obtidos. Por ora, esses dados devem ser deixados de lado para que se possa partir ao fator seguinte, o ano de publicação, cuja distribuição pode ser vista a seguir:

GRÁFICO 5: TAXA DE EMPREGO DE EIP POR ANO DE PUBLICAÇÃO

A curva presente no gráfico acima apresenta um crescimento gradual entre as década de 1860 e 1870, culminando em 1876. Após esse pico, há um declínio geral que se prolonga da década de 1880 até a década de 1920, quando há, em 1922, um novo salto. A esse salto, segue-se uma nova queda no último marco temporal utilizado aqui, 1938.

Algumas diferenças devem ser notadas entre os dados referentes ao nascimento e publicação das obras, a primeira, e talvez mais importante, sendo o espaço de tempo coberto por cada uma dessas variáveis. Enquanto os anos de nascimento variam dentro de uma faixa de pouco menos de 60 anos (*grosso modo* a duração do período Imperial), o período de publicação

das obras se estende por pouco mais de setenta anos, alcançando o período entreguerras e a Era Vargas. Essa maior dispersão nos dados resulta em uma cobertura geral menor de cada período e, portanto, uma figura menos detalhada dos desenvolvimentos específicos, problemas agravados pelo pouco número de obras avaliadas. Novamente, a aplicação de uma regressão logística sobre esses dados traz um resultado pouco animador:

TABELA 12: MODELO DE REGRESSÃO LOGÍSTICA POR ANO DE PUBLICAÇÃO

	Estimate	Std. Error	z value	Pr(> z)
Intercept	-10.096899	12.231526	-0.825	0.409
Ano de publicação	0.004469	0.006465	0.691	0.489

Dessa vez, a tendência revelada é de uma influência levemente positiva no emprego de eIP para obras mais tardias, porém com uma chance de aleatoriedade próxima de 50%. Em outras palavras, o *corpus* deste trabalho não permite que se faça uma análise em separado da influência que o ano de nascimento do autor e publicação da obra teriam no emprego de eIP. A solução óbvia para tal problema seria a obtenção de um *corpus* maior e mais equilibrado, com representação balanceada de todas as décadas. Outra possibilidade que pode revelar informações novas sem a necessidade de ampliação do conjunto de dados é o cruzamento desses dois fatores, o qual traz a seguinte resposta:

TABELA 13: MODELO DE REGRESSÃO LOGÍSTICA POR ANO DE PUBLICAÇÃO E NASCIMENTO

	Estimate	Std. Error	z value	Pr(> z)
Intercept	17.90203	14.03700	1.275	0.202187
Ano de publicação	0.05729	0.01589	3.606	0.000311
Ano de nascimento	-0.06924	0.01857	-3.729	0.000192

Quando o ano de nascimento do autor e o ano de publicação da obra são analisados em conjunto através de um modelo de regressão logística generalizada, ambos os fatores se revelam altamente significativos ($p < 0.01$), mantendo a direção já descoberta anteriormente: nascimento tem uma relação inversamente proporcional ao emprego de eIP, enquanto publicação tem uma relação diretamente proporcional ao emprego de eIP. Esse resultado advém da não correlação entre ano de nascimento e ano de publicação, isso é, o fato de que as curvas desses dois fatores não são equivalentes, certos autores podem ter publicado mais ou menos

cedo em suas vidas. O que um resultado como esse aparece indicar é que não apenas a norma adotada pelo falante, mas também o processo editorial pelo qual a obra passou acabam influenciando na seleção de variante, fato esperado e conforme com as expectativas iniciais. Surpreende, contudo, que o efeito do ano de publicação seja positivo, ainda que pouco, uma vez que a hipótese inicial seria a de redução global no uso ao longo dos dois eixos temporais conforme uma norma favorável a eGER se consolida no Brasil. Todavia, é fato que formas eIP ainda são utilizadas em PB, assim como outras estruturas do tipo IP, logo mesmo passados séculos da separação, a norma brasileira não é categórica quanto a esse fenômeno. Acesso a normas editoriais e informações extralinguísticas análogas poderiam ajudar a esclarecer esse resultado, mas é pouco provável que manuais de escrita e edição do período tenham sobrevivido até os dias presentes. Resta apenas confiar que o cruzamento dessas informações com outros fatores possa trazer mais revelações.

5.6: DISTRIBUIÇÃO DE DADOS POR UNIDADE PRECEDENTE E SEGUINTE

Até aqui, as análises tinham se focado nos fatores extralinguísticos – autor, região, província, ano de nascimento e ano de publicação. Como um apanhado geral, ficou claro que o fenômeno é complexo e todos os fatores acabaram se mostrando estatisticamente relevantes, seja quando analisados sozinhos ou em combinação com outro fator da mesma ordem. Que tal seja o caso não chega a ser surpreendente – é natural que um fenômeno complexo e socialmente carregado sofra influência de várias direções. Agora, porém, é hora de deixar um pouco de lado os fatores sociais e voltar os holofotes aos fatos linguísticos.

A análise seguirá a ordem listada na apresentação dos preditores linguísticos, começando, portanto, por **unidade precedente** e **unidade seguinte**. Como já foi explicado anteriormente, essas são categorias que se referem, respectivamente, à palavras diretamente anterior e posterior à perífrase em foco. Por ordem lógica, cabe olhar primeiro aos dados de unidade precedente, visíveis na tabela e no gráfico que seguem:

TABELA 14: DISTRIBUIÇÃO DE DADOS SEGUNDO UNIDADE PRECEDENTE

	eGER	eGER %	eIP	eIP %	Total
#	56	90.3	6	9.7	62
Adjetivo	5	71.4	2	28.6	7
Advérbio	48	81.4	11	18.6	59
Artigo	6	85.7	1	14.3	7
Conjunção	62	88.6	8	11.4	70
Pr. Oblíquo	31	83.8	6	16.2	37
Pr. Relativo	21	67.7	10	32.3	31
Pronome	51	87.9	7	12.1	58
Preposição	22	84.6	4	15.4	26
Substantivo	86	78.2	24	21.8	110
Verbo	26	96.3	1	3.7	27

GRÁFICO 6: DISTRIBUIÇÃO DE DADOS POR UNIDADE PRECEDENTE

O primeiro fato que cabe notar é a disparidade nos totais de unidades precedentes, com quatro classes de palavra (Substantivo, conjunção, início absoluto de palavra e advérbio) totalizando 241 (cerca de 49%) ocorrências, quase metade das quais (110) são somente de substantivos. O gráfico abaixo apresenta essa distribuição:

GRÁFICO 7: DISTRIBUIÇÃO TOTAL DOS DADOS POR UNIDADE PRECEDENTE

Aparenta haver, portanto, uma preferência por uma combinação estrutural de substantivo + eGER/eIP, sem que esse fator influa na distribuição geral, uma vez que a proporção entre as variáveis precedidas por substantivo é próxima da média geral. A aplicação de uma regressão logística sobre esse fator revela o seguinte cenário:

TABELA 15: MODELO DE REGRESSÃO LOGÍSTICA POR UNIDADE PRECEDENTE

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.11	0.05 – 0.25	<0.001
Adjetivo	3.73	0.59 – 23.59	0.161
Advérbio	2.18	0.75 – 6.35	0.151
Artigo	1.56	0.16 – 15.18	0.704
Conjunção	1.20	0.39 – 3.69	0.745
Pronome Oblíquo	1.81	0.54 – 6.08	0.340
Pronome	1.28	0.40 – 4.06	0.674
Pronome Relativo	4.44	1.44 – 13.75	0.010
Preposição	1.70	0.44 – 6.60	0.445
Substantiv	2.60	1.00 – 6.77	0.050
Verbo	0.36	0.04 – 3.13	0.354
Observations	493		
Cox & Snell's R ² / Nagelkerke's R ²	0.033 / 0.057		

Apenas dois tipos de itens precedentes se revelam importantes para a seleção de variante numa leitura univariada, sendo esses pronome relativo ($p < 0.01$) e substantivo ($p < 0.05$), tendo pronome relativo e substantivos efeitos positivos no emprego de eIP em oposição a início absoluto de palavra, com efeito negativo. Antes de avançar para interpretações desses

números, vale observar a outra categoria análoga, *unidade seguinte*, cuja distribuição se dá como mostrado adiante:

TABELA 16: DISTRIBUIÇÃO DE DADOS POR UNIDADE SEGUINTE

	GER	GER %	eIP	eIP %	Total
#	72	93.5	5	6.5	77
Adjetivo	5	83.3	1	16.7	6
Advérbio	45	93.8	3	6.3	48
Artigo	65	78.3	18	21.7	83
Conjunção	48	85.7	8	14.3	56
P. Oblíquo	4	23.5	13	76.5	17
P. Relativo	2	100	0	0	2
Pronome	20	80	5	20	25
Preposição	85	81.7	19	18.3	104
Substantivo	39	84.8	7	15.2	46
Verbo	29	96.7	1	3.3	30

GRÁFICO 8: DISTRIBUIÇÃO DE DADOS POR UNIDADE SEGUINTE

De modo similar a unidade precedente, a categoria unidade seguinte também apresenta uma concentração significativa de ocorrências ao redor de um número pequeno de fatores, nesse caso preposição, artigo, final absoluto de palavra e conjunção, que somados correspondem a 320 das 494 ocorrências (aproximadamente 65% do total). Observe-se a distribuição no gráfico abaixo:

GRÁFICO 9: OCORRÊNCIAS DE DADOS POR UNIDADE SEGUINTE

Novamente, fica claro que certas combinações são preferidas para a estrutura perifrástica geral, seja ela eIP ou eGER. O passo seguinte, portanto, é tentar identificar se existiria uma preferência para cada variante específica, o que pode ser verificado por meio de um teste de regressão logística. Esses são os resultados encontrados:

TABELA 17: MODELO DE REGRESSÃO LOGÍSTICA POR UNIDADE SEGUINTE

<i>Predictors</i>	<i>Odds Ratios</i>	TP	<i>p</i>
		<i>CI</i>	
(Intercept)	0.07	0.03 – 0.17	<0.001
Adjetivo	2.88	0.28 – 29.62	0.374
Advérbio	0.96	0.22 – 4.21	0.957
Artigo	3.99	1.40 – 11.35	0.010
Conjunção	2.40	0.74 – 7.78	0.144
Pronome Oblíquo	46.80	11.07 – 197.80	<0.001
Pronome	3.60	0.95 – 13.68	0.060
Preposição	3.22	1.14 – 9.05	0.027
Substantivo	2.58	0.77 – 8.68	0.125
Verbo	0.50	0.06 – 4.43	0.531
Observations	492		
Cox & Snell's R ² / Nagelkerke's R ²	0.094 / 0.160		

Nesse teste, a variável pronome relativo foi eliminada dada a sua baixa taxa de ocorrência (apenas dois casos dentro de um universo de 494) e a “sujeira” resultante desse fato. Com essa ressalva, observe-se o modelo acima. Como ocorreu para unidade precedente, vários fatores foram selecionados como correlatos com a variável, sendo eles artigo ($p < 0.01$),

pronome oblíquo ($p < 0.001$) e preposição ($p < 0.05$), todos apresentando uma correlação positiva, a maior das quais é o de pronome oblíquo.

Por fim, antes de prosseguir às análises, um último teste precisa ser feito quanto ao efeito combinado das duas categorias, unidade precedente e unidade seguinte. Em um caso como esse, onde dois fatores distintos podem colimitar a seleção um do outro, seria necessário testar a correlação de ambas as variáveis. Todavia, a parca quantidade de dados e o grande número de combinações possíveis impede que essa possibilidade seja testada. De fato, seria necessário um *corpus* ordens de magnitude maior do que o que aqui se obteve para que se pudesse avaliar tal interação. Portanto, tudo que se pode fazer com o material disponível é um teste de regressão logística multivariado dos dois grupos de fatores. A realização desse retorna o seguinte resultado:

TABELA 18: MODELO DE REGRESSÃO LOGÍSTICA POR UNIDADE PRECEDENTE E SEGUINTE

<i>Predictors</i>	<i>Odds Ratios</i>	TP	<i>p</i>
		<i>CI</i>	
(Intercept)	0.03	0.01 – 0.12	<0.001
Adjetivo (prec)	2.95	0.34 – 25.57	0.327
Aadvérbio (prec)	2.82	0.89 – 8.89	0.077
Artigo (prec)	0.93	0.06 – 15.60	0.959
Conjunção (prec)	1.40	0.42 – 4.69	0.587
Pronome Oblíquo (prec)	1.91	0.50 – 7.27	0.343
Pronome (prec)	1.08	0.31 – 3.81	0.903
Pronome Relativo (prec)	6.90	2.04 – 23.39	0.002
Preposição (prec)	1.72	0.40 – 7.50	0.468
Substantivo (prec)	3.17	1.13 – 8.94	0.029
Verbo (prec)	0.60	0.07 – 5.54	0.656
Adjetivo (seg)	2.75	0.25 – 29.74	0.406
Advérbio (seg)	0.87	0.19 – 3.91	0.851
Artigo (seg)	3.83	1.31 – 11.16	0.014
Conjunção (seg)	2.60	0.78 – 8.69	0.122
Pronome Oblíquo (seg)	60.09	13.44 – 268.62	<0.001
Pronome Relativo (seg)	0.00	0.00 – Inf	0.983
Pronome (seg)	4.96	1.24 – 19.83	0.024
Preposição (seg)	3.13	1.09 – 9.04	0.035
Substantivo (seg)	2.47	0.71 – 8.55	0.153
Verbo (seg)	0.54	0.06 – 4.93	0.584
Observations	493		
Cox & Snell's R^2 / Nagelkerke's R^2	0.131 / 0.223		

Uma vez mais se excluiu o fator pronome relativo da categoria unidade seguinte. Dessa vez, os fatores marcados como estatisticamente relevante para unidade precedente foram pronome relativo ($p < 0.01$) e substantivo ($p < 0.05$), enquanto para unidade seguinte foram marcados artigo ($p < 0.05$), pronome oblíquo ($p < 0.001$), pronome ($p < 0.05$) e preposição ($p < 0.05$), tendo todos eles efeitos *positivos* na seleção de eIP. Se repete o fato de que pronome oblíquo seguinte é o fator com o maior peso na seleção de variável, tendo, nesse caso, um efeito ainda maior do que quando a os grupos de fatores foram analisados em separado.

No capítulo 3.2 se descreveu o discurso sociolinguístico vigente no Império Brasileiro e o papel que tais discussões tiveram entre a intelectualidade. Um elemento recorrente nas discussões sobre a identidade do dialeto brasileiro era a colocação do pronome oblíquo, a qual, segundo Batista Caetano, Camilo Castelo Branco dizia ser suficiente para identificar a origem ou local de educação de um certo autor (ALMEIDA NOGUEIRA, 1876), preferindo os autores portugueses o pronome posposto e os brasileiros o pronome preposto. De fato, provavelmente não há tópico sobre dialetologia que tenha rendido mais discussões e gerado mais polêmicas do que a questão da colocação pronominal. O caso sendo esse, não haveria muita dificuldade em afirmar que tal variável ocuparia um lugar de destaque no repertório linguístico de brasileiros e portugueses. Tomando esse fato como verdade, seria possível afirmar que a presença de uma variável fortemente associada ao PE como é a posposição acabaria por influir na organização textual dos autores e os levando a selecionar¹² também na variável *realização de gerúndio* uma forma mais associada ao PE.

Para averiguar esses dados mais a fundo, os casos onde ocorrem os pronomes oblíquos foram separados do montante total e classificados de acordo com a posição do clítico, se proclítica ou enclítica, obtendo-se a seguinte tabela de distribuições:

¹² É preciso, porém, ressaltar aqui que a palavra *seleção* não é utilizado com um significado de uma escolha consciente (embora não seja impossível que tal seja o caso), mas meramente para descrever o resultado da ação. Não caberia aqui discutir quanto do processo de seleção linguística é consciente ou não, ou se consciência sequer se faria sentido falar em consciência como um elemento nesse processo.

TABELA 19: DISTRIBUIÇÃO DE DADOS POR COLOCAÇÃO PRONOMINAL

	eGER	eIP	Total
Ênclise	3	12	15
Próclise	31	7	38
Total	34	19	53

De partida, fica visível uma maior tendência à ênclise junto a eIP e uma maior tendência à próclise em eGER, fato que se comprova estatisticamente:

TABELA 20: MODELO DE REGRESSÃO LOGÍSTICA POR COLOCAÇÃO PRONOMINAL

Coefficients	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	13863	0.6455	2148	0.031742*
Próclise	-28744	0.7693	-3736	0.000187***

Como fica claro, a relação entre colocação pronominal e a variável dependente é altamente significativa para próclise ($p < 0.001$) e ênclise ($p < 0.05$), sendo a próclise mais favorável a eGER e a ênclise a eIP. Os dois fenômenos se mostram, portanto, em consonância quanto à representação de uma norma brasileira (que usaria de próclise e eGER) em oposição a uma norma não-brasileira (que usaria ênclise e eIP). É preciso destacar, contudo, que esses fenômenos podem ser ambos resultados de uma mudança mais geral na norma, iniciada, possivelmente, por uma série de alterações maiores na língua. Quaisquer que sejam as razões para essa situação, a correlação entre os dois fenômenos se mantém e deve ser mantida em mente.

As outras correlações encontradas são mais difíceis de explicar, sendo mais apropriado deixar suposições acerca delas para etapas seguintes da análise. Não se pode descartar a possibilidade de que ela sejam resultantes não exatamente de si mesmas, mas de sua interação com outros elementos linguísticos, algo que será testado no momento adequado. Por ora, adiante.

5.7: DISTRIBUIÇÃO DE DADOS POR TIPO DE SENTENÇA

Como explicado antes, essa categoria se refere à posição específica do complexo perifrástico, considerando a função desse dentro da sintaxe geral da ocorrência. Dentro dessa definição, os dados se distribuem como visível a seguir:

TABELA 21: DISTRIBUIÇÃO DE DADOS POR TIPO DE UNIDADE SINTÁTICA

	eGER	eGER %	eIP	eIP%	Total
Absoluta	72	88.9	9	11.1	81
Adjetiva	62	87.3	9	12.7	71
Aditiva	11	61.1	7	38.9	18
Adversativa	7	87.5	1	12.5	8
Apositiva	2	100	0	0	2
Assindética	30	83.3	6	16.7	36
Causativa	1	100	0	0	1
Completiva	13	86.7	2	13.3	15
Comparativa	9	100	0	0	9
Condicional	0	0	1	100	1
Conformativa	1	100	0	0	1
Explicativa	19	76	6	24	25
Final		75	1	25	4
Objetiva	61	88.4	8	11.6	69
Principal	99	83.9	19	16.1	118
Predicativa	15	83.3	3	16.7	18
Subjetiva	1	100	0	0	1
Temporal	8	50	8	50	16

GRÁFICO 10: DISTRIBUIÇÃO DE DADOS POR TIPO DE SENTENÇA

Uma vez mais, é possível observar que as ocorrências gerais se concentram ao redor de algumas poucas funções sintáticas, dessa vez oração principal, oração absoluta, oração subordinada adjetiva e oração subordinada objetiva. Juntas, essas quatro categorias correspondem a aproximadamente 70% do total de dados obtidos. Observe o gráfico a seguir:

GRÁFICO 11: DISTRIBUIÇÃO TOTAL DE TIPOS DE SENTENÇAS

A disposição dos dados ao longo de um número consideravelmente grande de fatores dentro da categoria *função sintática* e a alta concentração desses dados em um subconjunto pequeno desses fatores acaba por resultar em um número de pontos sendo representados por um número mínimo de dados, apenas um ou dois em certos casos. Como anteriormente explicado, esse tipo de dado é problemático para uma análise estatística, sua inclusão significando uma redução na confiabilidade dos modelos utilizados. Ainda assim, por critério de completude, a análise desse preditor será feita inicialmente com todos os dados. Um modelo univariado de seleção da variável por tipo de sentença traz os seguintes resultados:

TABELA 22: MODELO DE REGRESSÃO LOGÍSTICA POR TIPO SINTÁTICO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.14	0.07 – 0.29	<0.001
ADJT	1.06	0.38 – 2.93	0.916
ADTV	4.90	1.49 – 16.14	0.009
ASND	1.30	0.40 – 4.23	0.667
CMPL	1.17	0.22 – 6.09	0.855
EXPL	2.47	0.77 – 7.91	0.128
OBJT	0.89	0.31 – 2.55	0.830
PRCP	1.45	0.61 – 3.43	0.400
PRED	1.62	0.38 – 6.79	0.513
TEMP	8.00	2.33 – 27.41	0.001
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.048 / 0.079		

O primeiro fato que deve chamar a atenção nos dados acima é a presença de diversos casos de erro padrão extremamente desviantes da norma. Esses casos, como esperado, representam as categorias com uma combinação de poucas ocorrências e uso inequívoco de uma única variante. Apesar disso, orações absolutas, adjetivas e temporais foram selecionadas como estatisticamente relevantes para a seleção de variável, tendo as temporais e as adjetivas um efeito positivo na seleção de eIP, em contraste a absolutas. Quando a análise é repetida após a eliminação dos dados problemáticos, os resultados não se alteram, fato resultante do pouco número de ocorrências removidas.

Em todo caso, a relevância da categoria *oração subordinada temporal* e seu papel positivo na seleção de eIP condiz com os resultados encontrados por Mothé (2016), a qual ainda acabou por encontrar efeito significativo em orações subordinadas condicionais, fato que aqui

não se reproduz, talvez em razão dos poucos dados do tipo. Embora esse seja um achado interessante, é preciso questionar sua significância e, mais do que isso, sua realidade. Embora essa seção trate especificamente da categoria *função sintática* é preciso abrir uma exceção e tratar dessa ao lado das categorias anteriores, *unidade precedente* e *unidade seguinte*. Isso porque esses três grupos de fatores interagem em um nível profundo, estabelecendo limites uns para os outros – isso é, o tipo de oração ao mesmo tempo restringe e é restringido pelo tipo de palavra que o circunda. Infelizmente, testar interações entre esses grupos não é possível com os dados disponíveis, mas a aplicação de um modelo de regressão logística multivariado é cabível. Quando se faz isso, os resultados a seguir se revelam:

TABELA 23: MODELO DE REGRESSÃO LOGÍSTICA POR FUNÇÃO SINTÁTICA COM DADOS LIMPOS

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.03	0.01 – 0.14	<0.001
ADV	3.40	1.02 – 11.40	0.047
PRL	72.85	8.98 – 590.71	<0.001
SUB	3.66	1.21 – 11.07	0.021
POB	33.95	6.66 – 172.93	<0.001
PRO	8.40	1.79 – 39.44	0.007
PRP	3.49	1.06 – 11.45	0.040
ADJT	0.09	0.01 – 0.65	0.017
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.186 / 0.308		

Algumas informações importantes surgem aqui. A primeira delas é a redução do efeito da categoria oração temporal e sua perda de significância ($p > 0.05$). De fato, dentre todos os tipos de oração, apenas orações adjetivas mantém sua significância estatística dentro desse modelo, porém com uma direção de efeito inversa ao que se tinha encontrado anteriormente. Enquanto o modelo univariado apresentava orações adjetivas com um leve efeito positivo na realização de eIP, a análise multivariada desse fator junto a *unidade precedente* e *unidade seguinte* encontra um efeito negativo maior do que o efeito positivo anterior.

A informação mais importante que esse modelo nos traz parece ser a pouca influência que o tipo de oração parece ter para a seleção dessa variável, especialmente em contraste à relativa importância de *unidade precedente* e *unidade seguinte*. Não apenas são fatores dessas categorias que detém os maiores efeitos na seleção, estão nelas também o maior número de

fatores relevantes. Se esses dados forem levados à sua conclusão lógica, seria necessário assumir que a influência da sintaxe na seleção de variante, ao menos no tocante à essa variável, é na realidade muito pequena, praticamente inexistente. De modo algum se pretende afirmar que essa é uma verdade universal da língua e ainda restam análises importantes para se fazer nesse próprio trabalho, mas nada disso faz desse um resultado menos intrigante.

5.8: DISTRIBUIÇÃO DE DADOS POR NÚMERO E TEMPO DO VERBO E OCORRÊNCIA DE INTERPOLAÇÃO

Essa seção irá apresentar as distribuições dos dados para as três categorias linguísticas binárias de *número*, *tempo* e *interpolação*. A classe *tempo*, como apresentado anteriormente, classifica dados apenas a partir de sua localização no eixo passado, presente e futuro, ignorando diferenças de aspecto. Apesar da intenção inicial de separar essa categoria em três, apenas um dado no futuro surgiu dentro do *corpus* inviabilizando essa organização. Esse dado único, portanto, foi colocado junto ao conjunto de dados do presente e a disposição foi reclassificada para passado e não-passado.

Uma vez que essas são todas categorias com apenas dois fatores e essa seção visa apenas apresentar as distribuições, sem maiores incursões explicativas, todas elas foram unidas numa única parte para melhor organização do texto geral. Nesses termos, observe-se o primeiro agrupamento de dados – número do verbo:

TABELA 24: DISTRIBUIÇÃO DE DADOS POR NÚMERO DO VERBO

	eGER	eGER %	eIP	eIP%	Total
Plural	46	80.7	11	19.3	57
Singular	368	84.2	69	15.8	437

GRÁFICO 12: DISTRIBUIÇÃO DE DADOS POR NÚMERO DO VERBO

Como é possível ver, a vasta maioria das ocorrências encontradas foi com verbos no singular. Como o gráfico 12 mostra, a diferença entre as proporções de eIP e eGER não diferem muito entre um caso e outro e uma análise univariada confirma que qualquer disparidade observada é muito provavelmente resultado do acaso.

TABELA 25: MODELO DE REGRESSÃO LOGÍSTICA POR NÚMERO DO VERBO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.24	0.12 – 0.46	<0.001
Singular	0.78	0.39 – 1.59	0.500
Observations	494		
Cox & Snell's R ² / Nagelkerke's R ²	0.001 / 0.002		

Um valor de *p* de 0.5 indica uma chance de 50% de que os resultados sejam aleatórios, um número muito acima de qualquer margem de aceitável e que elimina qualquer possibilidade de que o número da perífrase, por si só, esteja afetando a seleção de variável.

O tempo do verbo, por sua vez, segue uma distribuição mais balanceada, com 226 ocorrências de presente contra 118 ocorrências de não presente. Essa é uma distribuição esperada uma vez que a tradição do romance ocidental elencou o pretérito como seu tempo narrativo padrão. A diferença entre as proporções, por sua vez, é um pouco maior do que para

a classe número, sendo que eIP representa 14,1% das ocorrências de verbos no presente e 23,4% das ocorrências no não-presente.

TABELA 26: DISTRIBUIÇÃO DE DADOS POR TEMPO DO VERBO

	eGER	eGER %	eIP	eIP%	Total
Presente	226	85.9	37	14.1	263
Pretérito	118	76.6	36	23.4	154

GRÁFICO 13: DISTRIBUIÇÃO DE DADOS POR TEMPO DO VERBO

Uma vez mais, uma regressão logística pode ser utilizada para medir o efeito real.

TABELA 27: MODELO DE REGRESSÃO LOGÍSTICA POR TEMPO DO VERBO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	<i>p</i>
		<i>CI</i>	
(Intercept)	0.15	0.11 – 0.21	<0.001
Pretérito	1.71	1.06 – 2.78	0.029
Observations	494		
Cox & Snell's R ² / Nagelkerke's R ²	0.010 / 0.016		

Diferentemente de *número do verbo*, tempo parece se ter uma correlação significativa com a seleção de variável ($p < 0.05$) com o pretérito apresentando um coeficiente positivo para a realização de eIP, embora com um efeito reduzido. Algumas considerações precisam ser feitas

sobre esse ponto, porém, novamente acerca de restrições trazidas pelo material utilizado. Uma vez que o tempo de narração padrão do romance é o passado, os dados de não-passado usados aqui são, em sua maioria, retirados de diálogos de personagens e embora o autor do texto seja o mesmo tanto para os trechos de narração quanto para os trechos de diálogo, não se pode ignorar que falas tendem a apresentar um tom e uma direção diferentes. Existe, portanto, a possibilidade de que a diferença verificada entre passado e não-passado seja na realidade uma diferença entre trechos de narração e trechos de diálogo. Infelizmente, esse não foi um fato controlado pela classificação utilizada aqui, então não se pode fazer nada além de erguer conjecturas sobre efeitos possíveis e deixar tal consideração para trabalhos futuros que também se utilizem de romances em seus *corpus*.

O último preditor a ser apresentado nessa seção é *interpolação*, segundo o qual os dados se distribuem como pode ser visto a seguir:

TABELA 28: DISTRIBUIÇÃO DE DADOS POR OCORRÊNCIA DE INTERPOLAÇÃO

	eGER	eGER %	eIP	eIP%	Total
Não	349	86%	58	14%	57
Sim	65	75%	22	25%	437

GRÁFICO 14: DISTRIBUIÇÃO DE DADOS POR OCORRÊNCIA DE INTERPOLAÇÃO

Como foi com *tempo* e *número*, *interpolação* também apresenta uma diferença significativa no total de dados entre as duas possibilidades, com a concentração dessa vez

estando em não ocorrência de interpolação, isso é, em interpolação negativa. Ao mesmo tempo, 25% dos dados com interpolação positiva são de eIP contra 14% dos dados com interpolação negativa. É uma diferença considerável e que se confirma diante da prova estatística:

TABELA 29: MODELO DE REGRESSÃO LOGÍSTICA POR INTERPOLAÇÃO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	<i>P</i>
		<i>CI</i>	
(Intercept)	0.17	0.13 – 0.22	<0.001
S	2.04	1.17 – 3.56	0.012
Observations	494		
Cox & Snell's R ² / Nagelkerke's R ²	0.012 / 0.020		

A divergência entre interpolação positiva e negativa se mostra estatisticamente significativa ($p < 0.05$) e a interpolação positiva tem uma correlação positiva com a ocorrência de eIP. É preciso, contudo, notar que o modelo utilizado para classificar material interveniente não considera a quantidade ou o tipo desse, algo que poderia apresentar diferenças na interpretação final do fato. Infelizmente, uma vez que a quantidade de ocorrências de interpolação é relativamente baixa, uma classificação mais detalhada correria o risco de diluir a informação, acabando por causar uma perda geral da qualidade das análises. Apesar disso, há certos fatos notáveis que se precisa destacar sobre esse preditor.

Um primeiro ponto notável é a não ocorrência de pronomes oblíquos em posição de interpolação em sentenças com eIP. Em sua maioria, os casos de interpolação que se detectou ocorreram com advérbios ou pronomes, como nos exemplos a seguir:

- a) Há de **estar aí a chorar** metida em algum canto tornou o compadre (M. Antônio de Almeida)
- b) Era um contra senso **estar sempre a recair**, a ocupar-se de uma criatura vulgar, comuníssima, que lhe não merecia ódio, com a qual não valia a pena perder um pensamento (J. Ribeiro)
- c) **Aí estou eu a fundir** todos os contrastes como se esses nomes e pessoas não fossem mais do que modos de ser da minha afeição anterior (Machado de Assis)

Em contraste, com eGER ocorrem diversos casos de interpolação com pronomes oblíquos, como a seguir:

- a) Pois vale a pena **estar se matando** por isso (L. Jacinto Rocha)
- b) Eu **estou me regalando** (J. Lopes de Almeida)

Casos do tipo presente nas sentenças *a* e *b* acima representam 50% dos casos de interpolação presentes nos dados de eGER em contraste com 0% dos casos de interpolação com eIP, uma diferença altamente significativa ($p < 0.001$) e que parece indicar uma relação entre a colocação do clítico e a seleção da variável. A natureza exata dessa relação, contudo, não é algo que seja possível confirmar nesse momento, mas sua existência é clara o suficiente e o fato de se tratar de um fenômeno linguístico, a próclise, também altamente associado ao português brasileiro pode dar pistas do que está acontecendo. O equivalente eIP das duas sentenças acima seria, numa primeira expectativa, respectivamente:

- a) Pois vale a pena **estar a se matar** por isso
- b) Eu **estou a me regalar**

Todavia, nos casos de interpolação com eIP encontrados no *corpus* jamais ocorre a presença de material interveniente entre a preposição e o verbo no infinitivo de modo que, observado esse padrão, as formas dessas sentenças acabariam por ser:

- a) Pois vale a pena **estar-se a matar** por isso
- b) Eu **estou-me a regalar**

A não ocorrência de tais possibilidades e a posição inequívoca do material interpolado em eIP indicam uma ligação mais forte entre a preposição e o verbo infinitivo. No subcapítulo sobre unidades precedentes e seguintes se estabeleceu que eIP se correlaciona com a ênclise enquanto eGER se correlaciona com a próclise, indicando que ambos os fenômenos podem ser

parte de uma mesma norma ou resultarem de um mesmo conjunto de mudanças mais gerais. Considerando esses fatos e sabendo que o pronome oblíquo interpolado é, na realidade, um pronome proclítico do verbo principal, a realização do clítico interpolado na única posição disponível em eIP (entre o verbo estar e a preposição) seria um pronome enclítico – algo desfavorecido por muito do discurso linguístico brasileiro. A realização de uma estrutura como em *a* e *b* acabaria sendo, portanto, duplamente estranha à norma brasileira, o que ajudaria a explicar sua não ocorrência dentro do *corpus*.

5.11: DISTRIBUIÇÃO DOS DADOS POR TIPO SEMÂNTICO DO COMPLEXO VERBAL

Em primeiro lugar, observe-se a distribuição dos dados a seguir a seguir:

TABELA 30: DISTRIBUIÇÃO DE DADOS POR TIPO SEMÂNTICO DO VERBO

	EGER	EGER %	EIP	EIP%	TOTAL
Comportamental	182	83.1	37	16.9	219
Existencial	2	100	0	0	2
Mental	35	81.4	8	18.6	43
Material	161	87	24	13	185
Relacional	4	100	0	0	4
Verbal	30	73.2	11	26.8	41

GRÁFICO 15: DISTRIBUIÇÃO DE DADOS POR TIPO SEMÂNTICO DO VERBO

Como ocorreu com todos os outros preditores, quando os dados são organizados de acordo com o tipo semântico, uma diferença gritante surge entre alguns tipos, altamente representados, em contraste a outros. Essa sendo uma categoria dividida em um número menor de fatores (apenas seis), essa concentração pode ser dividida em três partes, a primeira compreendendo verbos materiais e comportamentais (os quais abrigam aproximadamente 82% do total de dados), a segunda os mentais e verbais (somados, cerca de 17%) e, por fim, a terceira contendo os verbos existenciais e relacionais (o 1% restante).

Em termos das taxas relativa de eIP, alguns fatos precisam ser notados. O primeiro deles é a necessidade de se ignorar os valores de 0% visíveis nos fatores da categoria 3. Muito possivelmente, essa é uma coincidência estatística resultante da pouca ocorrência de verbos desse tipo. Para as categorias restantes, há uma variação pequena nas médias de emprego, salvo pelos verbos da categoria verbal, 26,8% dos quais são casos de eIP, em comparação a 18,6% da segunda categoria com mais ocorrências, os verbos do tipo *mental*. Uma vez mais, cabe aplicar um teste de regressão logística para averiguar a relevância estatística dessa diferença. Seguem os resultados:

TABELA 31: MODELO DE REGRESSÃO LOGÍSTICA POR TIPO SEMÂNTICO DO VERBO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.20	0.14 – 0.29	< 0.001
Mental	1.12	0.48 – 2.62	0.786
Material	0.73	0.42 – 1.28	0.274
Verbal	1.80	0.83 – 3.92	0.136
Observations	488		
Cox & Snell's R ² / Nagelkerke's R ²	0.010 / 0.016		

Os dados das classes *relacional* e *existencial* foram removidos em razão de sua escassez, porém, um teste realizado com a inclusão desses não apresenta qualquer diferença. Como pode ser visto, nenhum dos tipos de verbos se mostra significativo dentro de um alfa de 0.05, indicando uma não correlação entre o tipo semântico do verbo e a seleção da variante, ao menos em termos de uma análise univariada.

5.12: ANÁLISES MULTIVARIADAS

Nas seções anteriores, as diversas classes de fatores foram analisadas isoladamente, salvo pelas classes *unidade precedente*, *unidade seguinte* e *tipo de sentença*, para as quais uma análise multivariada era do mais imediato interesse. A partir dos resultados encontrados nesses testes anteriores, é possível tirar algumas conclusões iniciais.

- Apenas três itens da categoria *autor* (Franklin Távora, Júlio Ribeiro e Joaquim Lindolfo Rocha) apresentaram uma correlação com a seleção de variável, todos os quais apresentaram preferência por eGER;
- Em termos de região, o sudeste apresentou uma tendência significativamente maior de emprego de eIP em relação ao nordeste;
- Já para as províncias, Rio de Janeiro e Bahia foram os únicos que mostraram uma diferença relevante, com o Rio de Janeiro apresentando uma tendência positiva de emprego de eIP e a Bahia apresentando uma tendência negativa;
- Quando vistos em separado, nem o ano de nascimento do autor nem o ano de publicação da obra se mostraram relevantes, mas ao ser avaliados em paralelo, se descobriu que o ano de nascimento apresenta um efeito inversamente relativo na

realização de eIP, enquanto o ano de publicação apresenta um efeito diretamente relativo;

- e) Tanto a unidade precedente quanto a unidade seguinte apresentaram uma relação estatisticamente significativa com a seleção de variável, embora apenas em alguns fatores (pronomes relativos e substantivos, para unidade precedente e pronomes oblíquos, artigos, preposições) tenham se mostrado assim;
- f) Se vista por si mesma, a categoria *tipo de sentença* é relevante, ainda que apenas dois de seus fatores (subordinada aditiva e subordinada temporal) tenham um efeito. Todavia, quando olhada lado a lado com *unidade precedente* e *unidade seguinte*, essa situação se altera e apenas um tipo de sentença se mantém relevante;
- g) Não se verificou diferenças significativas entre dados no singular e dados no plural;
- h) O tempo pretérito parece favorecer a ocorrência de eIP, porém, esses dados são ligeiramente problemáticos;
- i) A ocorrência de interpolação tem uma correlação positiva com a ocorrência de eIP;
- j) A interpolação de pronomes oblíquos ocorre apenas com eGER;
- k) Nenhum tipo semântico apresentou uma correlação com a variável.

Assim, enquanto todas as categorias extralinguísticas analisadas se mostraram relevantes numa primeira fase de análise, o mesmo não se pode dizer das categorias intralinguísticas, das quais apenas as categorias *unidade precedente*, *unidade seguinte* e *interpolação* apresentam significância estatísticas sem ressalvas. Esse segundo fato parece indicar que a ordenação dos itens, isso é, a *sintaxe da frase*, em sentido mais amplo, é o fator linguístico que mais influencia a seleção de uma ou outra variante. Não menos interessante é o fato de que o maior efeito encontrado entre todas as análises univariadas foi na presença de um pronome oblíquo como unidade seguinte, um fato brevemente explorado anteriormente.

Com essas informações em mente, se pode enfim partir para as análises multivariadas. A partir desse ponto, serão cruzadas as diversas classes de fatores na tentativa de se obter o melhor modelo explanatório possível, além de testar interações entre certos fatos. Uma prévia disso pôde ser vista nas análises de *tipo de sentença*. A cada modelo apresentado, se tentará levantar alguns pontos e hipóteses explicativas para esses padrões. Todavia, é preciso ressaltar que a combinação específica de variável e método aqui utilizados não possuem pares na produção científica, limitando o escopo do que se pode atingir em termos de uma explicação.

Num cenário ideal, se poderia cruzar esses dados com outros, de tipo similar, resultantes de pesquisas anteriores. De um modo ou de outro, não se espera aqui explicar todos os fatos que serão encontrados, mas se tentará ao menos pôr em destaque questões por eles suscitadas.

Uma vez que em certo nível tal exercício foi feito, parece razoável começar as análises multivariadas pelos grupos de fatores linguísticos. Esses, como se viu anteriormente, apresentaram um grau de correlação geralmente menor do que os grupos de fatores sociais. Começemos por esses:

TABELA 32: MODELO DE REGRESSÃO LOGÍSTICA COM GRUPOS DE FATORES LINGUÍSTICOS

<i>Predictors</i>	<i>Odds Ratios</i>	TP	<i>P</i>
		<i>CI</i>	
(Intercept)	0.02	0.01 – 0.11	<0.001
ART	3.57	1.05 – 12.12	0.041
POB	29.77	5.81 – 152.63	<0.001
PRO	8.71	1.77 – 42.95	0.008
PRP	3.72	1.10 – 12.63	0.035
ADV	3.80	1.12 – 12.86	0.032
PRL	76.53	9.08 – 645.20	<0.001
SUB	3.90	1.27 – 11.96	0.017
ADJT	0.07	0.01 – 0.56	0.012
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.207 / 0.343		

O modelo acima avalia as classes de fatores linguísticas em paralelo, eliminando todos aqueles dados problemáticos para permitir uma melhor visualização dos efeitos individuais de cada fator no resultado final. Com esse modelo em mãos, um fato salta às vistas: dentre todos os fatores estatisticamente relevantes, apenas um deles não pertence às categorias *unidade precedente* ou *unidade seguinte* e esse é *oração subordinada adjetiva*. Dentre os fatores relevantes, aqueles com o maior efeito são, em ordem, *pronome relativo precedente*, *pronome oblíquo seguinte* e *oração subordinada adjetiva*. Nenhum desses é um fato novo – no geral, esses foram os mesmos resultados encontrados quando das análises das categorias às quais pertenciam esses fatores. Fato novo é a irrelevância sistemática de todas as outras classes de fatores, mesmo aquelas que, em análise univariada, tinham se mostrado estatisticamente significantes. É em contraste a esse fato novo que a manutenção da importância das categorias anteriores deve ser lida. Note-se ainda que alguns outros fatores ganham importância nessa análise; nominalmente *substantivo precedente* ($p < 0.05$), *advérbio precedente* ($p < 0.5$),

pronome seguinte ($p < 0.01$) e *preposição seguinte* ($p < 0.05$). Todos, portanto, fatores ligados às palavras circundantes do verbo.

A proposição inicial das classes *unidade precedente* e *seguinte* se baseava em dois fatos de ordem distintas, o primeiro sendo a utilização de classes análogas em análises de variáveis fonológicas e o segundo a ampla utilização de métodos baseados em sequenciamento de palavras no campo da linguística computacional e em abordagens probabilísticas. Havia, contudo, uma terceira suposição por trás dessa seleção, sendo essa a que justificou a separação de pronomes e pronomes oblíquos. O resultado esperado era que a posição do pronome oblíquo pudesse exercer efeito na seleção da variável, uma vez que esse fato ocupava lugar de destaque na identidade linguística brasileira. Como os resultados mostram, a suposição se confirmou, ainda que parcialmente: a ocorrência de um pronome oblíquo posposto traz, sim, um efeito positivo na realização de eIP, a variável que aqui se definiu como “mais portuguesa”. Porém, ao mesmo tempo, a preposição do pronome oblíquo não tem efeitos significativos. Trata-se, portanto, de uma associação unilateral entre dois itens identificados com o PE.

Em vista disso, haveria um impulso natural em se tentar associar os dois elementos como uma confirmação da tese proposta, se argumentando, talvez, em termos de uma teoria da ativação, mas certos fatos constroem tal primeira euforia. Antes de mais nada, que embora essa não seja categoria com o maior efeito na construção, *pronome relativo precedente* ocupa essa posição e não se poderia estabelecer um laço entre pronomes relativos e qualquer tipo de “portuguesidade” ou “não-brasilidade” que fosse. Outros elementos também envolvidos na sintaxe e que poderiam oferecer pistas sobre a estrutura não foram inseridos no modelo ou na classificação dos dados, elementos tais como ordem de constituintes, presença de inversão ou pistas lexicais. É preciso lembrar ainda do corolário básico da análise estatística: “correlação não indica causalidade”. No caso em mãos, uma possibilidade para a situação dos pronomes oblíquos seria uma simples co-ocorrência de dois elementos que marcariam uma mesma variante linguística. Se essa for a hipótese assumida, não se poderia falar em influência de um elemento sobre o outro.

Há uma questão teórica profunda por trás dessa situação, uma questão que se volta sobre a natureza específica da correlação encontrada. De um lado, não seria impossível postular,

como descrito antes, que a presença de um dos elementos¹³ levaria à ativação de outro. Sanchez *et al.* (2015) apontam para um fenômeno similar no inglês neozelandês, onde a pronúncia de vogais varia entre um modelo neozelandês e um modelo australiano. No caso apresentado pelas autoras, a presença de elementos lexicais associados com a Austrália tem um efeito positivo na realização das variantes australianas das vogais, indicando que esses causariam uma resposta cognitiva, favorecendo a realização das palavras associadas. Muito interessante é que, como no caso da relação entre pronome oblíquo e eIP, a presença de itens neutros ou Neozelandeses não apresenta um efeito inverso, isso é, não tem um efeito negativo na realização das variantes australianas nem um efeito positivo na realização de variantes neozelandesas. Com os dois casos indo na mesma direção, se poderia construir um argumento sobre a direção da influência de elementos contextuais, conceitualizando uma distinção entre elementos mais ou menos salientes dentro da qual as partes da cultura do próprio falante seriam tomadas como um padrão e, portanto, apenas elementos reconhecidos como “externos” ou “diferentes” resultando em uma resposta clara. É uma possibilidade intrigante e digna de exploração, porém é também uma possibilidade que exige testes rigorosos e uma metodologia específica para que seja confirmada ou refutada.

Quanto ao problema dos outros fatores que demonstram um efeito significativo mesmo sem uma associação cultural forte, algumas coisas se pode dizer. Em primeiro lugar, há o ponto levantado anteriormente: não se pode garantir que não haja algo mais ocorrendo disfarçado por trás desses dados. Ainda que esse não fosse o caso e a influência dos outros fatores precedentes e seguintes seja causada tão e tão somente por eles mesmos, não se poderia assumir que ela fosse advinda das mesmas razões para todos os tipos de unidades, especialmente quando alguns desses tipos parecem se correlacionar com outros elementos também considerados estatisticamente significativos.

Em termos desse último ponto, vale destacar a situação de *pronome relativo precedente* e *oração subordinada adjetiva*, onde os dois elementos aparecem muitas vezes juntos. Como os dados mostram, enquanto *pronome relativo precedente* apresenta não apenas um efeito positivo, mas o maior efeito positivo de correlação com eIP, a relação com *subordinadas adjetivas* é inversa – seu efeito de correlação com eIP é o mais negativo entre

¹³ É necessário fazer essa suposição já que nada garante que a seleção de posposição não se dá em razão da presença de eIP e não o oposto. Devido às suas propriedades matemáticas, testes de correlação desse tipo são incapazes de identificar a direção de influência (ou se sequer *há* uma influência e não uma mera correlação).

todos. Dificilmente esse é um resultado aleatório, mas um resultado da interação entre esses dois elementos, algo que só poderia ser testado em um *corpus* exponencialmente maior do que se dispõe aqui.

Concluindo os comentários do modelo acima, embora haja certas pistas de uma relação mais profunda entre certos elementos, qualquer afirmação mais forte sobre esses seria pretensiosa e descabida. Fatos interessantes surgiram, a maior correlação da variável com termos circundantes do que com a estrutura sintática sendo o mais notável, e esses indicam a necessidade de novas investigações e considerações sobre a escolha de fatores condicionadores.

Tudo isso dito, o próximo passo é avaliar os fatores sociais. Esses são em menor número do que os linguísticos e devido a sua natureza exigem um menor número de testes multivariados. Isso se dá devido à distribuição das categorias uma vez que muitas delas são colineares, como é o caso de *região* e *província*. Uma vez que a segunda é nada mais do que uma divisão mais exata daquela, realizar testes com as duas juntas não traria qualquer benefício. Do mesmo modo, testar qualquer uma dessas duas classes junto a *autor* seria inútil já que *autor* é, novamente, apenas mais uma subdivisão de *região*. *Autor* e *nascimento*, assim como *autor* e *publicação* são também combinações sem validade científica já que cada autor equivale a apenas um *nascimento* e é representado por apenas uma obra, isso é, por apenas um valor de *publicação*. Naturalmente, esse não seria o caso se o presente trabalho trabalhasse com mais obras de cada autor.

Com os dados disponíveis, as únicas combinações de fatores sociais possíveis são entre *região*, *província*, *ano de nascimento* e *ano de publicação*. Como se viu antes, todas essas quatro classes de fatores se mostraram relevantes, ainda que *ano de nascimento* e *ano de publicação* tenham o sido apenas quando lidas em conjunto. Uma vez que *região* e *província* não podem ser analisadas conjuntamente, restam apenas as combinações entre essas duas, separadamente, com *nascimento* e *publicação*. Entretanto, devido ao pouco número de pontos em cada uma das cinco províncias, uma análise feita com a combinação dessas com os dois conjuntos de fatores temporais se mostrou infrutífera. Em outras palavras, a única combinação que revela qualquer nível de significância estatística é entre *região*, *nascimento* e *publicação*. Quando esses três grupos são alinhados e testados a partir de uma interação entre *região* e uma combinação dos outros dois, se tem os seguintes resultados:

TABELA 33: MODELO DE REGRESSÃO LOGÍSTICA COM REGIÃO, NASCIMENTO E PUBLICAÇÃO

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	43912957	0.00 – 2358	0.184
Sudeste	0.00	0.00 – 6762...	0.229
Nascimento	0.84	0.74 – 0.95	0.004
Publicação	1.14	1.06 – 1.24	0.001
Sudeste:Nascimento	1.15	1.01 – 1.30	0.033
Sudeste:Publicação	0.91	0.83 – 0.99	0.021
Observations	494		
Cox & Snell's R ² / Nagelkerke's R ²	0.		

O modelo acima indica que não há uma diferença significativa entre as regiões, mas que nascimento ($p < 0.01$) e publicação ($p < 0.001$), assim como suas interações com *região* são relevantes ($p < 0.05$ para ambos). Ao observar o gráfico de efeitos dessas interações, se pode notar uma diferença considerável entre os efeitos observados em cada uma das regiões:

Gráfico 16: Efeitos de ano de nascimento do autor e publicação da obra de acordo com região

Como se pode ver no gráfico 16, tanto a região sudeste quanto a região nordeste apresentam efeitos na mesma direção, com o nascimento tendo uma correlação inversa à realização de eIP e publicação tendo uma correlação direta, porém em ambos os casos é notável a maior sensibilidade da região nordeste. Se os dados são separados de acordo a região e o modelo é aplicado em separado, essas informações se confirmam:

TABELA 34: MODELO DE REGRESSÃO LOGÍSTICA POR ANO DE NASCIMENTO E PUBLICAÇÃO (NORDESTE)

<i>Predictors</i>	TP		
	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	4391...	0.00 – 2358...	0.184
Nascimento	0.84	0.74 – 0.95	0.004
Publicação	1.14	1.06 – 1.24	0.001
Observations	235		
Cox & Snell's R ² / Nagelkerke's R ²	0.05		

TABELA 35: MODELO DE REGRESSÃO LOGÍSTICA POR ANO DE NASCIMENTO E PUBLICAÇÃO (SUDESTE)

<i>Predictors</i>	TP		
	<i>Odds Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)	62.73	0.00 – 26310...	0.796
Nascimento	0.96	0.93 – 1.00	0.027
Publicação	1.04	1.00 – 1.07	0.027
Observations	259		
Cox & Snell's R ² / Nagelkerke's R ²	0.023 / 0.036		

Como pode ser visto, *nascimento* e *publicação* são estatisticamente relevantes para as duas regiões, porém, os efeitos para o nordeste são aproximadamente quatro vezes mais fortes, um padrão que parece indicar que a região nordeste apresenta uma resistência maior que o sudeste à forma inovadora eIP. Esse era o comportamento esperado por várias razões, dentre as quais o histórico da região. Como foi em capítulos anteriores, a região nordeste teve um histórico considerável de conflitos com o domínio português e embora a Bahia tenha sido o ponto inicial de entrada da coroa no Brasil no 1808, os focos de imigração portuguesa no Brasil eram todos no sudeste, primeiro para Minas Gerais e São Paulo, no tempo das minas de ouro e mais tarde para o Rio de Janeiro (PRADO JUNIOR, 1989). Não seria surpreendente que esses fatores somados resultassem numa resistência ao uso de uma forma que já então era marcadamente portuguesa. Uma afirmação tão grande não poderia ser feita tomando por base apenas uma variável e nesse momento os limites do método se revelam com mais força. Parece claro, portanto, que uma abordagem que lançasse mãos de um conjunto de variáveis com valores similares permitiria se verificar mais adequadamente a questão em jogo.

A combinação de fatores sociais e linguísticos parece suscitar ainda mais essa direção de inquérito. Observe-se os três modelos a seguir, os quais cruzam informações linguísticas com *autor*, *província* e *região*, respectivamente:

TABELA 36: MODELO DE REGRESSÃO LOGÍSTICA COM AUTOR E GRUPOS DE FATORES LINGÜÍSTICOS

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.03	0.00 – 0.54	0.017
Franklin Távora	0.02	0.00 – 0.28	0.003
Lindolfo Rocha	0.06	0.01 – 0.47	0.008
Xavier Marques	25.63	1.38 – 477.32	0.030
Pronome relativo (prec)	1316.20	65.82 – 26319.93	<0.001
Substativo (prec)	5.46	1.49 – 19.97	0.010
Pronome Oblíquo (seg)	54.52	6.80 – 437.43	<0.001
Pronome (seg)	16.72	2.60 – 107.60	0.003
Preposição (seg)	6.53	1.50 – 28.33	0.012
Oração adjetiva	0.01	0.00 – 0.18	0.001
Classe Verbal	3.71	1.02 – 13.44	0.046
Tempo Pretérito	3.82	1.63 – 8.98	0.002
Interpolação positiva	3.84	1.55 – 9.51	0.004
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.341 / 0.564		

TABELA 37: MODELO DE REGRESSÃO LOGÍSTICA COM PROVÍNCIA E GRUPOS DE FATORES LINGÜÍSTICOS

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.00	0.00 – 0.04	<0.001
Minages Gerais	6.34	2.10 – 19.07	0.001
R. de Janeiro	4.92	1.72 – 14.05	0.003
Advérbio (prec)	4.04	1.09 – 14.95	0.037
Pronome relativo (prec)	393.05	29.43 – 5249.43	<0.001
Substativo (prec)	5.02	1.53 – 16.49	0.008
Artigo (seg)	4.83	1.28 – 18.24	0.020
Conjunção	4.70	1.06 – 20.90	0.042
Pronome Oblíquo	44.26	7.46 – 262.62	<0.001
Pronome	16.38	3.02 – 88.72	0.001
Preposição	5.31	1.40 – 20.08	0.014
Adjetivo	0.06	0.01 – 0.62	0.018
Tempo Pretérito	2.27	1.14 – 4.52	0.020
Interpolação positiva	2.14	1.00 – 4.59	0.050
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.259 / 0.429		

TABELA 38: MODELO DE REGRESSÃO LOGÍSTICA COM REGIÃO E GRUPOS DE FATORES LINGÜÍSTICOS

<i>Predictors</i>	<i>Odds Ratios</i>	TP	
		<i>CI</i>	<i>p</i>
(Intercept)	0.00	0.00 – 0.03	<0.001
Sudeste	6.44	2.80 – 14.80	<0.001
Advérbio (prec)	4.04	1.09 – 14.95	0.037
Pronome Oblíquo (prec)	5.95	1.03 – 34.35	0.046
Pronome Relativo (prec)	308.78	24.95 – 3821.01	<0.001
Substantivo (prec)	5.66	1.72 – 18.66	0.004
Artigo (seg)	3.79	1.05 – 13.75	0.042
Pronome Oblíquo (seg)	36.92	6.68 – 204.04	<0.001
Pronome (seg)	13.86	2.61 – 73.71	0.002
Preposição (seg)	4.54	1.24 – 16.63	0.023
Oração Adjetiva	0.06	0.01 – 0.62	0.018
Tempo Pretérito	2.15	1.08 – 4.27	0.028
Interpolação (sim)	2.32	1.10 – 4.92	0.028
Observations	417		
Cox & Snell's R ² / Nagelkerke's R ²	0.252 / 0.417		

Como se pode ver, há pouca mudança entre um modelo e outro. Em todos os casos, os padrões percebidos em análises anteriores se mantêm: em termos de categorias linguísticas, unidades circundantes são as categorias com maior correlação e o maior número de fatores estatisticamente relevantes. Em vários casos, os efeitos percebidos anteriormente se mostram intensificados, como ocorre com *pronome relativo precedente* e *pronome oblíquo seguinte*. Em todos os casos, essas, assim como *oração subordinada adjetiva* ainda apresentam os maiores efeitos.

As únicas alterações notáveis que ocorrem são a relevância de *interpolação* que surge nos três modelos e relevância da categoria semântica *verbal* na tabela 33. A interpolação e a possível razão de sua correlação foi comentada anteriormente, porém o tipo semântico *verbal* é fato novo ainda que seu efeito seja o menor dentro todos os fatores considerados significativos e que a relevância seja limitada ao modelo baseado em autores específicos. Um caso parecido ocorre com *pronome oblíquo precedente* e *artigo seguinte* no modelo baseado em região, todos com um valor de *p* acima de 0.04.

Como um todo, a pouca variação entre os modelos confirma as suspeitas anteriores. Há uma diferença significativa nos três níveis de organização, o regional, o provincial e o individual, ainda que não em todas as províncias e autores. No nível individual, não há uma correlação forte entre a posição política do autor e o seu uso da variável. Mesmo autores de

postura altamente liberal e nacionalista mostram uso da forma que seria menos brasileira, ao ponto de uma das maiores taxas de emprego de eIP ser encontrada na obra de Xavier Marques. E mesmo dentro da obra de um mesmo autor é possível encontrar taxas significativamente díspares de uso da variável, ora pendendo para um favorecimento de eIP, ora para um favorecimento de eGER. Fatores linguísticos, por sua vez, apresentam correlações mais fortes, embora não nas categorias esperadas e com uma relação material ainda pouco clara.

6: CONCLUSÕES

No capítulo inicial, o objetivo desse trabalho foi definido como tentar encontrar uma correlação entre usos linguísticos o emprego de eIP/eGER e a relação de regiões e autores com um conceito que se poderia compreender como “brasilidade” durante o período imperial. Através da leitura de textos de história e linguística, se definiu que essa é uma questão iniciada desde o nascimento do Império, mas que ganhou força no Segundo Reinado. Nessa fase, a discussão sobre a língua, pautada nos termos da linguística comparada, então considerada a forma mais científica de estudo da língua, tomou um lugar notável nos periódicos e nas discussões sobre a cultura brasileira, especialmente entre literatos, sendo feita especialmente em termos do que comporia a língua nacional brasileira. A resposta encontrada pela maioria dos pensadores do período foram os pontos de divergência, fossem esses lexicais ou gramaticais, entre a forma lusa e a brasílica da língua, confirmando a hipótese inicial.

Em termos da variável que se pretendeu estudar, os resultados encontrados foram parcialmente os esperados. Há um uso geral maior de eGER em relação a eIP, porém esse uso apresenta variação em diversos níveis e mesmo em obras diferentes de um mesmo autor, fator que contradiz o discurso de uma língua monolítica que imperava entre a intelectualidade da período. Trata-se de um fenômeno pautado por um alto grau de individualização e granularidade, ainda que também afetado por certos critérios gerais. A correlação encontrada com a colocação pronominal aponta ainda para uma cooperação desses dois fenômenos num processo de desenvolvimento do que se poderia entender como a norma culta do português brasileiro. Esse fato exige novas pesquisas para que possa ser adequadamente compreendido e descrito ainda mais quando se considera outras mudanças que pudessem estar ocorrendo concomitantemente. Esse é um período de instabilidade na sociedade, suas instituições e sua língua.

Os fatores linguísticos utilizados, como um todo, se mostraram pouco reveladores, indicando uma falha metodológica na seleção de critérios e modelos de análise. Embora os agrupamentos utilizados sigam majoritariamente resultados de pesquisas anteriores, o resultado final é pouco satisfatório. Se tomados em interpretação mais literal, os resultados das análises apontam para uma correlação muito mais forte com itens imediatamente circundantes do que com a estrutura sintática, embora as duas categorias se correlacionem. Testes mais sofisticados seriam necessários para confirmar com precisão o tipo de interação que ocorre, o que, por sua vez, exigiria um *corpus* não somente mais extenso como também mais variado em termos de

estruturas avaliadas. Ainda assim, é possível que se considere os resultados encontrados como indicativos de uma realidade subjacente e pesquisas futuras poderiam encontrar sucesso ao considerarem a organização imediata da sentença, aqui representada apenas por palavra precedente e seguinte, em consideração. Outros aspectos linguísticos podem ser defendidos com mais segurança, como a interpolação, cuja correlação com a variável e com a colocação pronominal fortalece a ideia de um conjunto de variáveis que estivessem formando o português brasileiro. Como um todo, os fatores linguísticos tiveram correlações fortes com a variável, mas ainda é necessário que se investigue mais a fundo a natureza de tais correlações e suas verdadeiras implicações para a língua. Um estudo vertical, dedicado a traçar a evolução da forma, seria altamente interessante nesse caso pois permitiria que se acompanhasse o desenvolvimento dessas estruturas ao longo do tempo.

Se entendida como um experimento, um teste de método para resolução de um determinado problema, essa pesquisa mostrou um sucesso moderado. Sua intenção original de tentar averiguar a correlação entre opiniões políticas e uso de uma variável não se confirmou, porém esse fato poderia ser relacionado muito mais com a variável selecionada do que com a teoria em si. Ainda que um certo discurso linguístico existisse, esse não encontra reflexo nos diversos autores, ao menos não do modo como foi delimitada aqui e não em isolamento. É possível que se analisada em conjuntura com outras variáveis também entendidas pelos autores da época como muito indicativas de uma regionalidade, a realização de progressivo se mostrasse mais significativa. Outras técnicas de avaliação também poderiam ser empregadas para se avaliar os padrões encontrados, tais como a resenha das obras utilizadas ou uma análise que compreendesse mais de um trabalho do mesmo autor. Esses caminhos alternativos permitiram que se medisse não apenas como um mesmo material linguístico pode ser empregado em situações diferentes, mas também que se julgasse a obra inserida em seu contexto a partir da recepção que ela encontrou em seu momento de publicação. A entrada cultural, que não se poderia dizer verdadeiramente antropológica, não está errada, mas precisa ainda ser refinada.

Tal trabalho, contudo, pertence ao futuro, e esse texto se encerra aqui, respondendo a menos perguntas do que inicialmente desejava, ainda que mais do que se esperava. No fim, o Império do Brasil continua tão ou mais misterioso do que era quando todo esse projeto se iniciou, para o bem e para o mal, mas se de toda essa lama se puder tirar uma grama de ouro, essa bastará como prova da possibilidade e esse projeto terá tido qualquer ínfimo sucesso.

REFERÊNCIAS

ABREU E LIMA, J. I. de. **Compendio da História do Brasil**. Rio de Janeiro: Laemmert, 1843.

ALMEIDA NOGUEIRA, Baptista Caetano. Apontamentos sobre o Abaíenga também chamado Guarani ou Tupi ou Lingua Geral dos Brasis, in: **Ensaio de Sciencia**. Rio de Janeiro: Brown & Evaristo, 1876, pp. p. 1-78.

_____. **Rascunhos sobre a grammatica da língua portugueza**. Rio de Janeiro: Typographia de A. dos Santos, 1881.

BARBOSA, Afrânio Gonçalves. Fontes escritas e história da língua portuguesa no Brasil: as cartas de comércio no século XVIII, in: LIMA, Ivana Stolze; CARMO, Laura do. **História social da língua nacional**. Rio de Janeiro: Edições Casa de Rui Barbosa, p. 181-211, 2008.

BARBOSA, Pilar; COCHOFEL, Fátima. O infinitivo preposicionado em PE, in: **Actas do XX Encontro Nacional da Associação Portuguesa de Linguística**, ed. I. Lisboa: Associação Portuguesa de Linguística, 2005, pp. 387–400.

BLAKE, A. C. A. S. Dicionario Bibliographico brasileiro, v.6, p. 375-376. Rio de Janeiro: Typographia Nacional, 1883-1902. Disponível em: <http://www2.senado.leg.br/bdsf/item/id/221681>

BOTELHO, Tarcísio Rodrigues. História da população brasileira: balanços e perspectivas. **Cadernos de história**, v. 6, n. 7, p. 30-48, 2001.

CALLARI, Cláudia Regina. Os Institutos Históricos: do Patronato de D. Pedro II à construção do Tiradentes, in: **Revista brasileira de História**, São Paulo, v. 21, n. 40, p. 39-82, 2001.

CASTRO, Vicente Félix. Uma página da revolução de 1842 na província de S. Paulo, in: LISBOA, José Maria. **Almanaque literário de São Paulo para o ano de 1978**. São Paulo: Tipografia da Província de São Paulo, 1978.

CAMPBELL, Lyle & POSER, William J. **Language Classification: History and Method**. Cambridge: Cambridge University Press, 2008.

CERTEAU, Michel de. **A escrita da História**. Rio de Janeiro: Editora Forense Universitária, 2ª edição, 1982.

CHOMSKY, A. Noam. **Cartesian Linguistics: a Chapter in the History of Rationalist Thought**. Cambridge: Cambridge University Press, 3ª edição, 2009.

_____. *Syntactic Structures*. Berlin-Nova Iorque: Mouton de Gruyter: 2002, 12ª ed.

COELHO, Olga Ferreira. Léxico, Ideologia e a historiografia linguística no século das identidades, in: **Revista Letras**, Curitiba, n. 61, especial, p. 153-166, 2003.

_____. Os nomes da língua: configuração e desdobramentos do debate sobre a língua brasileira no século XIX. **Revista do Instituto Estudos Brasileiros**, nº 47, p. 139-160, 2008.

CONDE-SILVESTRE, Camilo. *A 'third-wave' historical sociolinguistic approach to late Middle English correspondence: Evidence from the Stonor Letters* in: RUSSIA, Cinzia. **Current Trends in historical sociolinguistics**. Varsóvia/Berlin: De Gruyter Open, 2016, p. 46-64.

CONVERSI, Danielie. *Language or race?: The choice of core values in the development of Catalan and Basque nationalisms*, in: **Ethnic and Racial Studies**, v. 13, n. 1, p. 50-70, 1990.

CUNHA, Celso. Conservação e inovação no português do Brasil. **O Eixo e a Roda: Revista de Literatura Brasileira**, n. 5, p. 199-230, 1986.

DARWIN, Charles. *On the Origin of Species by means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*. Londres: John Murray, 1859.

DIEZ, Friedrich C. *Der grammatik der romanischen sprachen*. Bonn: E. Weber, 1858.

DUARTE, Urbano. **As duas línguas**. 1902. Disponível em: <http://academia.org.br/abl/media/memoria6.pdf>. Acesso em 26/12/2017.

ECKERT, Penelope. *Jocks and burnouts: Social categories and identity in the high school*. Nova Iorque: Teachers College Press, 1989.

_____. *Three Waves of variation study: the emergence of meaning in the study of sociolinguistic variation*, in: **Annual Review of Anthropology**, 2012, p. 85-100.

_____. *Variation, meaning and social change* in: COUPLAND, Nikolas. **Sociolinguistics: Theoretical Debate**. Cambridge, Nova Iorque: Cambridge University Press, 2016, p. 68-85.

FANINI, Michele Asmar. Júlia Lopes de Almeida: entre o salão literário e a antessala da Academia Brasileira de Letras, in: **Estudos de Sociologia**, v. 23, n. 45, 2009, pp. 317-338.

FAVERO, Leonor Lopes & MOLINA, Márcia A. G. **As concepções linguísticas no Século XIX: A gramática no Brasil**. Rio de Janeiro: Editora Lucerna, 2006.

FONTANA, Josep. *História: análise do passado e Projeto Social*. Bauru: Edusc, 1998.

FOX, John. *Effect Displays in R for Generalised Linear Models*. **Journal of Statistical Software**, n. 8, v. 15, 2003, pp. 1-27. URL <http://www.jstatsoft.org/v08/i15/>.

GALVÃO, Bruno. <http://www.eurogamer.pt/articles/2018-03-14-nier-automata-recebe-premio-d-i-c-e-e-yoko-taro-agradece-ao-seu-estilo>

GALVES, Charlotte; NAMIUTI, Cristiane; PAIXÃO DE SOUSA, Maria Clara. Novas perspectivas para antigas questões: revisitando a periodização da língua portuguesa, in: *Grammatische Strukturen des Europäischen Portugiesisch*, p. 45-75, 2006.

GUIMARÃES, Manoel Luis Salgado. Nação e civilização nos trópicos: o Instituto Histórico e Geográfico Brasileiro e o projeto de uma História Nacional, in: **Caminhos da Historiografia**, v1, n1, 1988: p. 5-27.

HABERLY, David T. O Nacionalismo Linguístico no Brasil Pós-Colonial: um Estudo Comparativo. **Revista do Instituto de Estudos Brasileiros**, São Paulo, n. 14, p. 23-34, 1973. Disponível em: <<https://www.revistas.usp.br/rieb/article/view/69794>>. Acesso em: 03/01/2018.

HARRELL JR, Frank E. rms: Regression Modeling Strategies. R package version 5.1-2, 2018. <https://CRAN.R-project.org/package=rms>

HARRIS, Roy & TAYLOR, Talbot J. *Landmarks in Linguistic Thought I: The Western Tradition from Socrates to Saussure*. Londres: Routledge, 2ª ed, 1997

HAY, Jennifer & DRAGER, Katie. *Stuffed toys and speech perception*, in: *Linguistics* v. 48, i. 4, 2010, pp. 865-892.

HOBBSBAWN, ERIC. *Nations and Nationalism since 1780: programme, myth, reality*. Nova Iorque: Cambridge University Press, 2ª ed., 2013.

https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/default_atlas.shtm

HUTCHINSON, John. *The Dynamics of Cultural Nationalism: The Gaelic Revival and the Creation of the Irish Nation State*. Taylor & Francis Routledge, 1987.

- HYMES, Dell. *Foundations in Sociolinguistics*. Londres: Travistock Publications Ltd., 2001.
- JURAFSKY, Daniel; BELL, Alan; GREGORY, Michelle & RAYMOND, William D. *Probabilistic Relations between Words*, in: BYBEE, Joan & HOPPER, Paul (orgs.). *Frequency and the Emergence of Linguistic Structure*. Amsterdã/Filadélfia: John Benjamins Publishing Co., 2001, p. 229-254.
- KATO, Mary Aizawa. A restrição de mono-argumentalidade da ordem VS no português do Brasil in: **Fórum Linguístico** v. 2, n. 1, p. 97-127, 2000.
- KUZNETSOVA, A; BROCKHOFF, PB; CHRISTENSEN, RHB. *lmerTest Package: Tests in Linear Mixed Effects Models*. **Journal of Statistical Software**, n. 82, v.13, 2017, pp. 1-26.
- LABOV, William. *The social motivation of a sound change*. **Word**, v. 19, n. 3, p. 273-309, 1963.
- _____. *The social stratification of (r) in New York City department stores*. In: *Dialect and Language Variation*. 1986. p. 304-329.
- _____. **Principles of Linguistic Change, volume 1: Internal Factors**. Malden/Oxford: Blackwell, 1994.
- _____. **Principles of Linguistic Change, volume 2: Social Factors**. . Malden/Oxford: Blackwell, 2001.
- _____. **Principles of Linguistic Change, volume 3: Cognitive and Cultural Factors**. Wiley-Blackwell, 2010.
- LOPEZ, Vicente Fidel. *Les races aryennes du Pérou*. Paris: Librairie A. Franck, 1871.
- LUCENA FILHO, Marcio. **Carneiro Vilela: Língua de ‘navalha’ e pena de ‘ponta de faca’**. Tese de doutorado em História. Universidade Federal de Pernambuco, 2016.
- LÜDECKE, D. *sjPlot: Data Visualization for Statistics in Social Science*, 2018. doi: 10.5281/zenodo.1308157 (URL: <http://doi.org/10.5281/zenodo.1308157>), R package version 2.6.2, <URL: <https://CRAN.R-project.org/package=sjPlot>>.
- MACEDO, Joaquim Manoel de. **Lições de História do Brasil para uzo dos alumnos do Imperial Collegio de Pedro Segundo**. Rio de Janeiro: Tipografia Imparcial, 1861.

MARTINS, Joaquim Dias. **Os mártires pernambucanos vítimas da liberdade nas duas revoluções ensaiadas em 1710 e 1817**. Recife: Tipografia de F. G. de Lemos e Silva, 1853.

MARTIUS, Karl Friedrich Philipp von. Como se deve escrever a história do Brasil. in: **RIHGB**, t. 6, Rio de Janeiro: IHGB, p. 381-403, 1844.

MOTHÉ, Núbia Graciella Mendes. **Gerúndio versus Infinitivo Gerundivo**: Brasil e Portugal em contraste nos séculos XIX e XX, in: **Estudos Lingüísticos**, v. 35, p. 1554-1563, 2006.

_____. **Variação e Mudança Além e quem Mar: gerúndio versus infinitivo gerundivo no português dos séculos XIX e XX**. Dissertação de Mestrado em Língua Portuguesa. UFRJ, Rio de Janeiro, 2007.

MATTOS E SILVA, Rosa Virgínia. **“O português são dois...”**: Novas fronteiras, velhos problemas. São Paulo: Parábola, 2003.

NASCIMENTO, Washington Santos. “São Domingos, o grande São Domingos”: repercussões e representações da Revolução Haitiana no Brasil escravista (1791-1840) in: **Dimensões**, v. 21, 2008. Disponível em: <http://www.publicacoes.ufes.br/dimensoes/article/viewFile/2486/1982>. Acesso em 25/06/2017.

NEVALAINEN, Terttu, & RAUMOLIN-BRUNBERG, Helena. **Historical sociolinguistics: language change in Tudor and Stuart England**. Taylor & Francis, 2016.

NEVALAINEN, Terttu. *Vernacular universals?*. in: NEVALAINEN, Terttu; KLEMOLA, Juhani & LAITINEN, Mikko. **Types of Variation: Diachronic, Dialectical and Typological Interfaces**, v.76, 2006, p. 351-369.

NOBELS, Judith, & GIJSBERT Rutten. *Language norms and language use in seventeenth-century Dutch: negation and the genitive*, in: RUTTEN, Gijsbert; VOSTERS, Rik & VANDENBUSSCHE, Wim. **Norms and usage in language history, 1600-1900**. John Benjamins Publishing co., 2014.

NOVAES, Eder Liz. **Joaquim Felício dos Santos**: republicanismo e cultura historiográfica (1869-1871). 2014. 196 f. Dissertação (Mestrado em História) - Instituto de Ciências Humanas e Sociais, Universidade Federal de Ouro Preto, Mariana, 2014.

PARANHOS DA SILVA, José Jorge. A collocação dos pronomes in: **Revista Brasileira**, a. 2, t. 5, 1880, p. 496-499.

_____. A imigração perante a fisiologia. **Revista Brasileira**, a. 3, t. 9, p. 502-510, 1881.

_____. Uma questão de linguística, in: **Revista Brasileira**, a. 3, t. 9, p. 133-137, 1881.

PAUL, Hermann. *Principles of the History of Languages*. Londres: Swan Sonneck, Lowrey & Co. 1888.

PESSOA DE MELLO, Urbano Sabino. **Apreciação da revolta praieira em Pernambuco**. Rio de Janeiro: Tipografia do Correio Mercantil de Rodrigues e cia., 1849.

PETRI, Verli & SCHNEIDERS Caroline M. & SCHERER, Amanda. “A língua nacional” de João Ribeiro: alguns apontamentos in: **Polifonia**, Cuiabá, MT, v. 22, n. 31, p. 53-72, 2015.

PODESVA, Robert J. *Phonation type as a stylistic variable: The use of falsetto in constructing a persona* in: **Journal of sociolinguistics**, v. 11, n. 4, p. 478-504, 2007;

_____. *The California vowel shift and gay identity*, in: **American speech** v. 86, n. 1, p. 32-51, 2011.

PRADO JUNIOR, Caio. **História econômica do Brasil**. São Paulo: Brasiliense, 1970, 2ª ed.

QUINLAN, Susan Canty. Inês Sabino e as personagens femininas na *Belle époque* in: **Letras de Hoje**, Porto Alegre, v. 33, n. 3, p. 17-25, 1998.

RAMOS JR, José de Paula. Xavier Marques: beletismo e narrativa crítica in: **Revista USP**, n. 110, pp. 115-125, 2006.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

RIBEIRO, Cristina Betioli. **Um norte para o romance brasileiro: Franklin Távora entre os primeiros folcloristas**. 2008. [200] p. Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. Disponível em: <<http://www.repositorio.unicamp.br/handle/REPOSIP/270280>>. Acesso em: 10 ago. 2018.

RIBEIRO, Júlio. **Grammatica Portugueza (3º anno)**. 3ª edição, Rio de Janeiro: Livraria Clássica de Alves & C, 1889. Consultada na *CLB - Biblioteca Virtual das Ciências da Linguagem no Brasil*.

RIBEIRO, João. **Diccionario Grammatical**. 2ª edição, São Paulo: Livraria Clássica de Alves & Comp., 1897.

_____. **A Língua Nacional: Notas Aproveitáveis**. 2ª edição, São Paulo: Companhia Editora Nacional, 1933. Consultada na *CLB - Biblioteca Virtual das Ciências da Linguagem no Brasil* <http://www.labeurb.unicamp.br/bvclb/obr018> [Fonte: Biblioteca do Instituto de Estudos da Linguagem – IEL, UNICAMP]

ROBINS, R. H. **Short History of Linguistics**. Londres: Longman, 1967.

ROCHA PITA, Sebastião. **Historia da America Portugueza**. Bahia: Imprensa Economica, 1878.

RODRIGUES, Aryon Dall'Igna. As línguas gerais sul-americanas. **PAPIA: Revista Brasileira de Estudos Crioulos e Similares**, São Paulo, v. 4, n. 2, p. 6-18, 1996. Disponível em: <<http://abecs.net/ojs/index.php/papia/article/view/143/210>>. Acesso 15 abr. 2011.

ROMAINE, Suzanne. **Socio-historical linguistics. Its status and methodology**, London: Cambridge University Press, 1982.

ROMERO, Silvio V. S. R. A poesia popular, capítulo 8. **Revista Brasileira**, a. 2, t. 6, p. 206-235, 1880.

ROSSOLILO, Francesco. Nação. in: BOBBIO, Norberto & MATTEUCCI, Nicola & PASQUINO, Gianfranco. **Dicionário de Política**. Brasília: Editora Universidade de Brasília, p. 795-799, 1998.

SANCHEZ, K., HAY, J., & NILSON, E. (2015). Contextual activation of Australia can affect New Zealanders' vowel productions in: **Journal of Phonetics**, 48, 76–95.

SARAIVA, Francisco de S. Luiz. Memória em que se pretende mostrar, que a língua portugueza não he filha da latina, nem esta foi em tempo algum a lingua vulgar dos lusitanos. **Memorias da Academia**, Classe de Sciencias Moraes, e Bellas Letras, t. XII, p. I. Lisboa: Typografia da Academia Real das Sciencias de Lisboa, 1837.

SAUSSURE, Ferdinand de. **Mémoire sur le système primitif des voyelles dans les langues indo-européennes**. Leipzig: B. G. Teubner, 1879. Disponível em: <https://gallica.bnf.fr/ark:/12148/bpt6k729200/>. Acesso em: 14/01/2019.

_____. *Cours de linguistique generale*. Paris: Éditions Payot & Rivages, 1995.

SCHLEICHER, A. *Die ersten Spaltungen des indogermanischen Urvolkes* in: *Allgemeine Monatsschrift für Wissenschaft und Literatur*, 3, 1853, p. 786-787

_____. t. *Darwinism tested by the Science of language* in: KOERNER, Konrad. *Linguistics and evolutionary theory: three essays by August Scheleicher, Ernst Haeckel and Wilhelm Bleek*. Amsterdã/Filadélfia: John Benjamins Publishing co. 1983, p. 12-66

SCHWARCZ, Lilia Moritz. **O espetáculo das raças: Cientistas, Instituições e Questão Racial no Brasil, 1870-1930**. São Paulo: Companhia das Letras, 1993.

SERPA, Phocion. Manuel Antonio de Almeida. **Revista Iberoamericana**, v. 9, n. 18, p. 325-356, 1945.

SILVA, Ivanilson Bezerra. Júlio Ribeiro: leitura sobre a trajetória de um intelectual maçom e protestante na cidade de Sorocaba na segunda metade do século XIX. *Revista Confluências Culturais*, Joinville, v. 1, n. 1, p. 29-45, sep. 2012. ISSN 2316-395X. Disponível em: <<http://periodicos.univille.br/index.php/RCCult/article/view/65/63>>. Acesso em: 16 jan. 2019.

SILVA, Leandro Maciel. A escrita da história no Brasil oitocentista: o lugar da província do Ceará. **Saeculum – Revista de História**, n. 33, João Pessoa, p. 339-356, 2015.

SILVA, Rogério F. **Colônia e Nativismo**. São Paulo: Hucitec, 1997.

SODRÉ, Nelson Werneck. **As razões da independência**. São Paulo: Civilização Brasileira, 1978.

SOUZA, Luana Batista de. **Grande é o poder do tempo: colação entre testemunhos de O Seminarista de Bernardo Guimarães**. 2012. Dissertação (Mestrado em Filologia e Língua Portuguesa) - Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2012. doi:10.11606/D.8.2012.tde-20082012-115854. Acesso em: 2019-04-11.

TAVARES, Marcelo dos Reis. Oliveira Lima e a fundação da nacionalidade brasileira por Dom João VI. Franca: UNESP-FHDSS, 2003.

TRAVAGLIA, Luiz Carlos. **O aspecto verbal no português**. 5ª ed. Uberlândia: Universidade Federal de Uberlândia, 2014.

VARNHAGEN, Francisco Adolfo de. **História geral do Brasil**. Revisão e notas de Rodolfo Garcia. 5ª ed. São Paulo: Edições Melhoramentos, 1956, tomo IV.

VENDLER, Zeno. *Linguistics in philosophy*. 1969.

WHITNEY, William Dwight. *The life and growth of languages: an outline of linguistic science*. Nova Iorque: D. Appleton and Company, 1887.

WICKHAM, Hadley; FRANÇOIS, Roman; HENRY, Lionel & MÜLLER, Kirill. dplyr: A Grammar of Data Manipulation. R package version 0.7.8, 2018. <https://CRAN.R-project.org/package=dplyr>