UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE LETRAS E LINGUÍSTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGUÍSTICOS

NEUBIANA SILVA VELOSO BEILKE

POMMERSCHE KORPORA: UMA PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE *CORPORA* DIALETAIS

UBERLÂNDIA
2016

NEUBIANA SILVA VELOSO BEILKE

POMMERSCHE KORPORA: UMA PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE *CORPORA* DIALETAIS

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Letras e Linguística da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de mestre em Estudos Linguísticos.

Área de Concentração: Estudos em Linguística e Linguística Aplicada.

Linha de Pesquisa: Teoria, descrição e análise linguística.

Orientador: Prof. Dr. Guilherme Fromm
Coorientadora: Profa. Dra. Adriana Cristina Cristianini

UBERLÂNDIA
2016

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

B422p 2016

Beilke, Neubiana Silva Veloso, 1984-
Pommersche Korpora : uma proposta metodológica para compilação de corpora dialetais / Neubiana Silva Veloso Beilke. - 2016.
285 f. : il.

Orientador: Guilherme Fromm.
Coorientadora: Adriana Cristina Cristianini.
Dissertação (mestrado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Estudos Linguísticos.
Inclui bibliografia.

1. Linguística - Teses. 2. Linguística de corpus - Teses. 3. Pomeranios - Brasil - Teses. I. Fromm, Guilherme. II. Cristianini, Adriana Cristina. III. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Estudos Linguísticos. IV. Título.

CDU: 801

NEUBIANA SILVA VELOSO BEILKE

POMMERSCHE KORPORA: UMA PROPOSTA METODOLÓGICA PARA COMPILAÇÃO DE *CORPORA* DIALETAIS

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos do Instituto de Letras e Linguística da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de mestre em Estudos Linguísticos.

Uberlândia, 27 de Julho de 2016.

Prof. Dr. Guilherme Fromm – UFU – Orientador

Profa. Dra. Adriana Cristina Cristianini – UFU – Co-orientador(a)

Prof. Dr. Ariel Novodvorski – UFU – Examinador

Profa. Dra. Stella Esther Ortweiler Tagnin – USP – Examinador(a)

In memoriam,
a Wilhelm Friedrich August Ferdinand Beilke,
ancião que iniciou o clã dos Beilke no Brasil,
por sua coragem em atravessar, em 1857, até o outro lado do oceano e
enfrentar um Brasil desconhecido, como tantos imigrantes pomeranos o fizeram.

AGRADECIMENTOS

Ao concluir mais uma etapa da vida acadêmica que escolhi trilhar, sinto-me realmente grata por tudo e por todos, por ter conseguido vencer os percalços que a trajetória impõe a todos nós, que somos sujeitos às determinações e ao tempo cronológico. Sinto-me grata também pela oportunidade de aprender sempre e poder demonstrar que, diante dos desafios, o melhor que pude fazer foi enfrentá-los. Então, de fato, chegou a hora de agradecer...

Agradeço, primeiramente, a Deus, por ter me permitido ser! E ter criado a mim de modo que desde criança me interessasse por sua "Babel de línguas".

Agradeço à minha mãezinha, por me amar incondicionalmente, por me dar força e por sempre se preocupar com minha saúde ao passar tantas horas estudando.

Agradeço ao meu querido orientador, Guilherme Fromm, pela paciência, pelas orientações e conselhos, por ser sempre tão disciplinado, comprometido e presente nos momentos de cumprir com os compromissos do mestrado.

Agradeço à minha co-orientadora, Adriana Cristina Cristianini, pela atenção, pelo companheirismo e por ser também uma amiga.

Agradeço, em especial, aos componentes da banca examinadora, Stella Esther Ortweiler Tagnin e Ariel Novodvorski, por terem aceitado o compromisso de ler meu trabalho, avaliá-lo e contribuir para meu crescimento acadêmico.

Agradeço a todos aqueles que foram meus professores durante o curso das disciplinas da pós-graduação, pois aprendi muito com eles, com suas experiências e com as aulas incríveis que já deixaram saudades.

Agradeço ao Instituto de Letras e Linguística e ao Programa de Pós Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia bem como à sua coordenação.

Agradeço à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela bolsa de estudos concedida.

Agradeço às secretárias Virgínia D. Ávila e Luana A. Silva e à equipe de servidores que me atendeu durante os dois anos do curso de mestrado.

Agradeço a todos os meus colegas de mestrado e de estudo, como Márcio Issamu Yamamoto, Lucas Peixoto, Danila Carvalho, Solange Cardoso e Célia Davi de Assunção, especialmente, aos meus queridos amigos e companheiros Daniela Faria Grama e Raphael Marco Oliveira Carneiro, por terem feito do curso de mestrado um tempo muito agradável.

Agradeço também pela oportunidade de participar das proveitosas reuniões dos grupos de pesquisa PLEX, GPS e GPELC.

Agradeço a todos aqueles que, direta ou indiretamente, contribuíram para a realização deste trabalho, particularmente a Stéfano Paschoal, que ofereceu importantes contribuições na qualificação deste trabalho, e também a Hans-Joachim Osterheld, Dieter Böhnke, Luís Eduardo Carmo, Rúdio Pieper, Nelson Pieper, Hermes Pieper e à toda a família Pieper. À família Gaede pela contribuição e pela doação de um livro. Ao amigo e pesquisador Danilo Kuhn Silva e à minha querida amiga e colaboradora Marta Hinz Treptow.

Agradeço, com muito carinho e admiração, aos sujeitos-entrevistados e participantes desta pesquisa, que possuem um conhecimento mais rico do que imaginam e que sempre estarão em meu coração. Além deles, agradeço a todos os pomeranos, a quem também dedicamos a nossa pesquisa, pois são a principal inspiração do nosso trabalho. Pomeranos estes que têm lutado pelo resgate e pela valorização de sua cultura, para manter viva a fala pomerana no contexto brasileiro.

Gostaria de poder não me esquecer de ninguém, caso o tenha feito, sintam-se todos reconhecidos com minha sincera gratidão.

"(...) Vasto é o mundo
Bastantes são os sonhos
O que queremos
Nós, pomeranos,
é simplesmente sermos também!
Sermos por muito tempo!"

(KALK, C. Blåg Sei. Mar Azul. Poesias de um Pomerano - livro póstumo).

RESUMO

O presente estudo teve como objetivo principal realizar a compilação de *corpora* do pomerano no Brasil, com a adoção de um amplo espectro de fontes, para a composição de um conjunto formado por diversos *corpora* escritos e um *corpus* oral, a fim de contribuir para a documentação do pomerano. Nossos dois objetivos específicos foram: tentar identificar a sobrevivência ou o desaparecimento do pomerano nas regiões alvo de nossa pesquisa – os Vales do Rio Doce/MG e do Rio Pardo/RS; e comparar brevemente o pomerano coletado nessas duas regiões. Para alcançar os referidos objetivos, o estudo teve como fundamentação teórica a Lexicologia, a Linguística de *Corpus* (LC), precisamente, e as teorias que fundamentam a abordagem-metodologia da LC. E, dada a natureza interdisciplinar desta pesquisa, contemplamos também as noções da Sociogeolinguística (SGL), campo de estudo que subsidiou nossa coleta de dados orais. Os princípios do empirismo e da perspectiva descritiva também inspiraram nosso trabalho. Entre nossos referenciais, encontram-se Biderman (1978, 1992, 2001), Sinclair (1991, 2004), Fillmore (1992), Barbosa (1991), Crystal (1991), Sánchez (1995), McEnery e Wilson (1996), Berber Sardinha (2000, 2004, 2009), Tognini-Bonelli (2001), Tagnin (2001, 2004, 2005, 2010), Stubbs (2001), Brum-de-Paula e Espinar (2002), Fromm (2003, 2010), Feitosa (2005), Lemnitzer e Zinsmeister (2006), Perini (2006, 2008), Gonzalez (2007), Cristianini (2007, 2012), Teixeira (2008), Zavaglia e Welker (2013), Marcuschi (2008), Raso e Mello (2009), Parodi (2010), Cardoso (2010), Viana (2010), Alves (2011), Mello (2012), Araújo (2012), McEnery (2013), Takano (2013), Almeida (2014), Von Borstel (2014), Novodvorski e Finatto (2014) e Novodvorski (2015). Nossa metodologia envolveu a coleta, compilação, transcrição e tratamento de dados escritos e orais, bem como a conversão da escrita para um padrão único. Adaptamos os questionários para a coleta de dados orais, o Questionário Sociolinguístico (QS) e o Questionário Semântico Lexical (QSL), este último para a versão em alemão *Lexikalisch-Semantischer Fragebogen* (LSF). Também organizamos todo o material compilado como uma unidade, realizamos a codificação, nomenclatura, agrupamento, separação e reagrupamento dos dados. Formulamos quatro hipóteses para esta pesquisa: a cogitação de que o pomerano não é ágrafo, pois a coleta e compilação de *corpora* demonstra a existência de formas escritas; a ideia de que os falantes são encontrados, em grande maioria, no meio rural, podendo ser inexistente, em algumas localidades, a presença de falantes de pomerano na zona urbana; a interferência do contato de pomeranos com outras etnias; e, a consideração de que o contato dos descendentes de pomeranos com o português tenha permitido o surgimento de uma variedade brasileira do pomerano. Os resultados permitiram a confirmação de três das hipóteses. Ao final, chegamos a um conjunto de *corpora,* chamado *Pommersche Korpora* (PK), que contém quatorze *corpora* escritos e um *corpus* oral. Esboçamos análises prévias sobre fatos linguísticos direcionados pelos *corpora.* Verificamos, com base nas evidências encontradas no PK, a existência de um pomerano que apresenta influências portuguesas, alemãs e dialetais (de outras variedades germânicas), por isso o chamamos de *Brasilianisch-Pommersch*, ou seja, uma variedade brasileira do pomerano. Encerramos nosso trabalho lançando perspectivas futuras para o estudo do pomerano.

Palavras-chave: Pomerano. *Corpora*. Linguística de *Corpus*. Sociogeolinguística. *Brasilianisch-Pommersch.*

ABSTRACT

The aim of the present study was mainly to compile *corpora* of Pomeranian in Brazil, with the adoption of a wide range of sources for the composition of a set of various textual *corpora* and spoken *corpus* in order to contribute to the Pomeranian documentation. Our two specific objectives were to try to identify the survival or disappearance of the Pomeranian in the target regions of our research – Vale do Rio Doce/MG and Vale do Rio Pardo/RS – and briefly compare the Pomeranian collected in these two regions. To achieve the aforementioned objectives, the study is theoretically based on Lexicology and *Corpus Linguistics* (CL), especially the theories behind the approach-methodology of CL. In addition, owing to the interdisciplinary nature of this research, we have contemplated the notions of Sociogeolinguistics (SGL), field of study that supported our collection of oral data. The principles of empiricism and descriptivism have also inspired our work. Among our references are the studies of Biderman (1978, 1992, 2001), Sinclair (1991, 2004), Fillmore (1992), Barbosa (1991), Sánchez (1995), McEnery e Wilson (1996), Berber Sardinha (2000, 2004, 2009), Tognini-Bonelli (2001), Tagnin (2001, 2004, 2005, 2010), Stubbs (2001), Brum-de-Paula e Espinar (2002), Fromm (2003, 2010), Feitosa (2005), Lemnitzer e Zinsmeister (2006), Perini (2006, 2008), Gonzalez (2007), Cristianini (2007, 2012), Teixeira (2008), Zavaglia e Welker (2013), Marcuschi (2008), Raso e Mello (2009), Parodi (2010), Cardoso (2010), Viana (2010), Alves (2011), Mello (2012), Araújo (2012), McEnery (2013), Takano (2013), Almeida (2014), Von Borstel (2014), Novodvorski e Finatto (2014) e Novodvorski (2015). Our methodology involved the collection, compilation, transcription and processing of textual and spoken data as well as the writing conversion to a single standard. It was also necessary to adapt the questionnaires to collect the spoken data, which are the Sociolinguistic Questionnaire (SQ) and the Lexical-Semantic Questionnaire (LSQ), the latter in the German version is called *Lexikalisch-Semantischer Fragebogen* (LSF). Additionally we organized all the compiled data as a unit, carried out the coding, classification, grouping, separation and regrouping of the data. The following four research hypotheses were formulated: the Pomeranian is not agraphic; the speakers are found in rural areas, and the presence of Pomeranian speakers in urban areas may be non-existent in some locations; the Pomeranians intermingling with other ethnic groups may allow linguistic interference; the contact of Pomeranian's descendants with the Portuguese language has allowed interference to the point of there being a Brazilian variety of Pomeranian speech. The results point to the confirmation of three hypotheses. At last, we have made a group of *corpora* called *Pommersche Korpora* (PK) that contains fourteen textual *corpora* and one spoken *corpus*. We also made some previous analysis about language facts driven by the *corpora*. It is apparent that our work enabled the detection of the existence of a Pomeranian variety in Brazil with Portuguese, German and dialectal (other German varieties) influences, in other words, the existence of a *Brasilianisch-Pommersch*, based on evidences found in the PK. We finish our work giving some future directions for the study of Pomeranian.

Keywords: Pomeranian. *Corpora*. *Corpus* Linguistics. Sociogeolinguistics. *Brasilianisch-Pommersch.*

LISTA DE FIGURAS

Figura 1: Mapa da Pomerânia. 32
Figura 2: Livro de Registro da "Agência Official de Colonisação". 34
Figura 3: Registro Eclesiástico da IECLB-SCS/RS. 35
Figura 4: Mapa de localidades onde há presença de pomeranos no Brasil 56
Figura 5: Arquivo etiquetado contendo palavras soltas em pomerano. 105
Figura 6: Documento de conversão e convenção da escrita - Versão 1. 110
Figura 7: Documento de conversão e convenção da escrita - Versão 2. 111
Figura 8: Resposta do IBGE sobre Vila Neitzel/MG. 120
Figura 9: Mapeamento do pomerano no Brasil. O Vale do Rio Doce/MG. 121
Figura 10: Escola no interior de Vila Neitzel e placas de identificação do povoado rural. 122
Figura 11: Mapeamento do pomerano no Brasil. O Vale do Rio Pardo/RS 126
Figura 12: Mapeamento do pomerano no Brasil. A Serra dos Tapes e a Lagoa dos Patos/RS. 129
Figura 13: Mapeamento do pomerano no Brasil. Espírito Santo/ES 130
Figura 14: Questionário Semântico-Lexical em alemão, excerto 139
Figura 15: Exemplo de etiquetagem, antes da retirada das perguntas do *corpus* oral. 147
Figura 16: Exemplo de etiquetagem retirado dos *corpora* escritos. 148
Figura 17: Arquivos dos *corpora* escritos com a codificação dos nomes. 152
Figura 18: Arquivos do *corpus* oral com a codificação dos nomes. 153
Figura 19: Logotipo do *Pommersche Korpora.* 161
Figura 20: Organograma com o detalhamento da estrutura do *Pommersche Korpora.* 162
Figura 21: Estatísticas do *Corpus* de Inscrições dos Túmulos Pomeranos. 165
Figura 22: Exemplos de inscrições em lápides pomeranas em Itueta/MG e Vera Cruz/RS. 165
Figura 23: Itens *Hapax Legomena* do *Corpus* de Inscrições dos Túmulos Pomeranos. 166
Figura 24: Estatísticas do *Corpus* de Livros de Registros Eclesiásticos. 168
Figura 25: Lista de Palavras do *Corpus* de Livros de Registros Eclesiásticos. 169
Figura 26: Exemplo de documento do *Corpus* de Registros Eclasiásticos. 170
Figura 27: Variações nos itens para designar casamento no *Corpus* de Livros de Registros Eclesiásticos. 171
Figura 28: Estatísticas do *Corpus* de Cartas Pessoais. 172
Figura 29: Exemplo de documento do *Corpus* de Cartas Pessoais. 172
Figura 30: Lista de palavras do *Corpus* de Cartas Pessoais. 173

Figura 31: Exemplos de documento do *Corpus* de Receitas Culinárias. 174
Figura 32: Estatísticas do *Corpus* de Receitas Culinárias 174
Figura 33: Lista de palavras do *Corpus* de Receitas Culinárias. 175
Figura 34: Estatísticas do *Corpus* do Jornal Folha Pomerana. 176
Figura 35: Exemplo de documento do *Corpus* do Jornal Folha Pomerana. 176
Figura 36: Lista de palavras do *Corpus* do J. Folha Pomerana. 177
Figura 37: Estatísticas do *Corpus* de Textos Diversos da Internet. 178
Figura 38: Exemplo do *Corpus* de Textos Diversos da Internet. 178
Figura 39: Lista de palavras do *Corpus* de Textos Diversos da Internet. 179
Figura 40: Imagem do Documentário *Dai Pomerer: Dai gang fon ainem folk.* 180
Figura 41: Estatísticas do *Corpus* de Legendas de Documentários 181
Figura 42: Lista de palavras do *Corpus* de Legendas de Documentários. 182
Figura 43: Estatísticas do *Corpus* de Trabalhos Acadêmicos. 183
Figura 44: Exemplo do *Corpus* de Trabalho Acadêmicos. 183
Figura 45: Lista de palavras do *Corpus* de Trabalhos Acadêmicos. 184
Figura 46: Estatísticas do *Corpus* de Textos Religiosos. 185
Figura 47: Exemplos do *Corpus* de Textos Religiosos. 185
Figura 48: Lista de palavras do *Corpus* de Textos Religiosos. 186
Figura 49: Composição das telas de frequência e lematização dos itens *God*, *Got*, *Jeesus* e *Jesus*. 187
Figura 50: Estatísticas do *Corpus* de Músicas Pomeranas. 188
Figura 51: Exemplo do *Corpus* de Músicas Pomeranas. 189
Figura 52: Palavras relevantes no *Corpus* de Músicas Pomeranas. 189
Figura 53: Estatísticas do *Corpus* Literário Pomerano. 190
Figura 54: Exemplo do *Corpus* Literário Pomerano. 191
Figura 55: Lista de palavras do *Corpus* Literário Pomerano. 191
Figura 56: Estatísticas do *Corpus* de Livros. 192
Figura 57: Exemplo do *Corpus* de Livros. 193
Figura 58: Lista de Palavras do *Corpus* de Livros. 193
Figura 59: Estatísticas do *Corpus* de *Sprüche* Diversos. 195
Figura 60: Exemplo do *Corpus* de *Sprüche* Diversos. 195
Figura 61: Lista de Palavras do *Corpus* de *Sprüche* Diversos. 196
Figura 62: Estatísticas do *Corpus* de Palavras Soltas. 197
Figura 63: Exemplo do *Corpus* de Palavras Soltas. 198

Figura 64: Lista de palavras do *Corpus* de Palavras Soltas. 199
Figura 65: Diário doado por família pomerana da Zona Rural Norte/MG. 200
Figura 66: Estatísticas do PKE – Total dos *corpora* escritos carregados no WST. 202
Figura 67: Estatísticas do *Pommersch Korpus* Oral. 203
Figura 68: Lista de palavras do PKO. 204
Figura 69: Estatísticas do PK total. 205
Figura 70: A partícula "*deet"* localizada na *Wordlist* e lematizada no *Pommersche Korpora*. 223
Figura 71: Linhas de concordâncias de *deet* lematizado. 223

LISTA DE QUADROS E TABELAS

Quadro 1: Quadro comparativo de formas de escritas pomeranas. 112
Tabela 1: Resposta da consulta ao IBGE sobre dados oficiais de Vila Neitzel/MG 119
Quadro 2: Delimitação dos pontos de pesquisa. 131
Tabela 2: Cálculo do número de entrevistas e participantes. 134
Quadro 3: Legendas dos códigos de identificação dos conjuntos dos *corpora* escritos e oral 150
Quadro 4: Ordem dos códigos de identificação do conjunto dos *corpora* com exemplos. 152
Quadro 5: Agrupamento do conjunto dos *corpora* do pomerano 156
Quadro 6: Levantamento das rádios com programação em pomerano. 201
Tabela 3: Dados quantitativos parciais e totais do *Pommersche Korpora.* 207
Quadro 7: Respostas ao LSF/QSL que apresentaram interferência de outras variedades germânicas. 217
Quadro 8: Indícios do *Brasilianisch-Pommersch*, a variedade brasileira do pomerano no PK. 218
Quadro 9: Exemplos de frases do *Brasilianisch-Pommersch* extraído do PKO. 219

LISTA DE GRÁFICOS

Gráfico 1: Famílias falantes e não-falantes de pomerano presentes no leste de Minas.....117
Gráfico 2: Famílias falantes e não-falantes de pomerano que emigraram do leste de Minas 118
Gráfico 3: Proporção de imigrantes pomeranos em VC/RS conforme registros eclesiásticos. 167
Gráfico 4: Proporção dos *corpora* do PK, segundo o número de *Tokens*. 206

LISTA DE ABREVIATURAS E SIGLAS

ALIB – Atlas Linguístico do Brasil

ALMA – Atlas Linguístico-Contatual das Minorias Alemãs na Bacia do Prata

BP – Brasilianisch-Pommersch ou a variedade brasileira do pomerano

CEP – Comitê de Ética em Pesquisa

DiWA – Digitaler Wenker-Atlas

DOC – Documento de texto com formatação

GPDG – Grupo de Pesquisa em Dialetologia e Geolinguística

HD – Hochdeutsch

HR – Hunsrückisch ou Hunsriqueano Riograndense

IBGE – Instituto Brasileiro de Geografia e Estatística

IDS – *Institut für Deutsche Sprache* – Instituto para a Língua Alemã

LC – Linguística de *Corpus*

LCA – Linguística de *Corpus* Aplicada

LSF – Lexikalisch-Semantischer Fragebogen

ONG – Organização Não-Governamental

PD – Plattdeutsch ou Plattdüütsch

PK – Pommersche Korpora

PKO – Pommersch Korpus Oral

PKE – Pommersche Korpora Escritos

PLN – Processamento de Linguagem Natural

PTB – Português Brasileiro

QS – Questionário Sociolinguístico

QSL – Questionário Semântico Lexical

SGBD – Sistema de Gerenciamento de Banco de Dados

SGL – Sociogeolinguística

TXT – Plain Text – texto plano

VOT – *Voice Onset Time* – tempo de ataque de vozeamento

WST – *WordSmith Tools*®

SUMÁRIO

APRESENTAÇÃO **20**

1. INTRODUÇÃO **23**

1.1 Nosso tema e objeto de estudo 23

1.2 Justificativa 23

1.3 Hipóteses 27

1.4 Objetivos Geral e Específicos 29
1.4.1 Objetivo Geral 29
1.4.2 Objetivos Específicos 30

1.5 Organização da dissertação 30

2. O POMERANO NA EUROPA E NO BRASIL **32**

2.1 O Pomerano na Europa 32

2.2 História da Imigração: como os pomeranos vieram parar aqui? 33

2.3 O pomerano no Brasil e sua definição: língua ou dialeto? 39
2.3.1 Nosso posicionamento 45
2.3.2 "Alemão", termo genérico para representar línguas germânicas do grupo alemão 49

2.4 Mapeamento e levantamento de pesquisas 55

3. FUNDAMENTAÇÃO TEÓRICA **61**

3.1. O Léxico e a Lexicologia 61

3.2 A relação com as evidências, a descrição e o empirismo 64

3.3 A Linguística de *Corpus* - LC 67
3.3.1 Breve Histórico da LC **67**
3.3.2 As definições de LC e a nossa concepção de LC **70**
3.3.3 A Concepção de Língua da LC **76**
3.3.4 Tipos de abordagens em LC: *Corpus-based* e *Corpus-driven* **77**
3.3.5 Prerrogativas do uso de *corpora* nas pesquisas **79**
3.3.6 Princípios fundamentais e conceitos importantes **80**
3.3.6.1 Planejamento e construção de *corpus* 81
3.3.6.2 Representatividade e extensão 81
3.3.6.3 Balanceamento 82
3.3.6.4 Frequência e lista de frequência 83
3.3.6.5 *Type* e *Token* 83
3.3.6.6 Lematização 84
3.3.6.7 Concordanciador 84
3.3.6.8 *Corpus* escrito e oral 84

3.3.6.9 *Corpora*.....91
3.3.6.10 Tipos de *Corpus*.....92

3.4 A complexidade das múltiplas fontes e as esferas de realização linguística presentes nos *corpora*: diamesia, gêneros, tipos, domínios discursivos e suportes.....93

3.5 A Sociogeolinguística como fundamentação para uma abordagem-metodologia de compilação de *corpus* oral dialetal.....96

4. DESENVOLVIMENTO: PROCEDIMENTOS METODOLÓGICOS PARA CONSTITUIÇÃO DOS *CORPORA* DO POMERANO.....103

4.1 Primeira fase: constituição dos *corpora* escritos.....103
4.1.1 Coleta dos dados escritos - fontes.....105
4.1.2 Processos de compilação: transcrição, digitalização, digitação.....106
4.1.3 Tratamento dos dados escritos.....107
4.1.3.1 Conversão da escrita.....108
4.1.3.2 Convenção da escrita.....109

4.2 Segunda fase: constituição do *corpus* oral.....112
4.2.1 Metodologia para coleta dos dados orais.....113
4.2.1.1 Coleta dos dados orais - entrevistas..... 113

4.2.2 Da seleção das localidades e suas caracterizações segundo a SGL 114
4.2.2.1 Minas Gerais.....115
4.2.2.1.1 Vila Neitzel e Itueta.....115
4.2.2.2 Rio Grande do Sul.....123
4.2.2.2.1 Vera Cruz.....123
4.2.2.2.2 Arroio do Tigre.....125
4.2.2.2.3 São Lourenço do Sul.....127
4.2.2.2.4 Canguçu.....127
4.2.2.3 Outras localidades: Santa Maria de Jetibá/ES.....129
4.2.2.4 Esclarecimento sobre a caracterização das localidades.....131
4.2.2.5 Contatos-chave.....132

4.2.3 Plano de recrutamento: dos critérios de inclusão e exclusão, segundo as variáveis da SGL e as exigências do CEP.....132

4.2.3.1 Da justificativa do número de sujeitos participantes da pesquisa e da quantidade de entrevistados.....133
4.2.3.2 Dos gêneros dos sujeitos participantes das entrevistas.....135
4.2.3.3 Faixas etárias contempladas.....135
4.2.3.4 Escolaridade.....137

4.2.4 Dos procedimentos para a realização das entrevistas e da descrição dos riscos ...137

4.2.4.1 Medidas para contenção dos riscos.....137
4.2.4.2 Preparação dos instrumentos de coleta.....137
4.2.4.3 Testagem do método: realização de entrevistas piloto.....140
4.2.4.4 O método de Wenker.....140
4.2.4.5 Do equipamento para gravação das entrevistas.....141

4.2.5 Da aplicação dos questionários – entrevistas para coleta dos dados orais 142
4.2.5.1 Da duração das entrevistas 143

4.2.6 Tratamento dos dados orais 144
4.2.6.1 Método de transcrição 144
4.2.6.2 Etiquetagem parcial 146
4.2.6.3 Exclusão de dados antroponímicos 149

4.3 Organização dos *corpora* escritos e do *corpus* oral como uma unidade 149
4.3.1 Codificação e nomenclatura dos arquivos 149
4.3.2 Agrupamento, separação e reagrupamento segundo a classificação em diamesias ou gêneros ou domínio discursivo ou suportes 154

4.4 Recursos 157

4.5 Listagem 158

4.6 Arquivamento de segurança do conjunto dos *corpora* 159

5. RESULTADOS 160

5.1 Resultado Geral: A constituição de um conjunto de dados a partir dos *corpora* do pomerano 160

5.2 Resultado da coleta de diversas fontes de dados pomeranos 162

5.3 Os *Pommersche Korpora Escritos* como resultados da primeira fase de compilação 163
5.3.1 *Corpus* de Inscrições dos Túmulos Pomeranos – CITP 163
5.3.2 *Corpus* dos Livros de Registros Eclesiásticos – CLRE 166
5.3.3 *Corpus* de Cartas Pessoais – CCP 171
5.3.4 *Corpus* de Receitas Culinárias – CRC 173
5.3.5 *Corpus* do Jornal Folha Pomerana – CJFP 175
5.3.6 *Corpus* de Textos Diversos da Internet – CTDI 177
5.3.7 *Corpus* de Legendas de Documentários – CLD 179
5.3.8 *Corpus* de Trabalhos Acadêmicos – CTA 182
5.3.9 *Corpus* de Textos Religiosos – CTR 184
5.3.10 *Corpus* de Músicas Pomeranas – CMP 187
5.3.11 *Corpus* Literário Pomerano – CLP 190
5.3.12 *Corpus* de Livros – CL 192
5.3.13 *Corpus* de *Sprüche* Diversos – CSD 194
5.3.14 *Corpus* de Palavras Soltas – CPS 196
5.3.15 *Corpus* de diários – Resultado não alcançado 199
5.3.16 *Corpus* de material de Rádios – *Scripts* – Resultado não alcançado 200

5.4 Resultados quantitativos do PKE em conjunto 202

5.5 O *Pommersch Korpus Oral – PKO* como resultado da segunda fase de compilação. 203

5.6 Resultados quantitativos do PK (PKE e PKO) 204

5.7 Resultados referentes aos objetivos específicos 207

6. ANÁLISE DOS RESULTADOS 211

6.1 Classificação e tipologia geral dos *corpora* 211
6.1.1 Classificação geral e tipologia dos *corpora* escritos, o PKE 211
6.1.2 Classificação e tipologia do *corpus* oral, o PKO 213

6.2 Testagem e avaliação das hipóteses de pesquisa 214

6.3. Análise do resultado das entrevistas interativas para coleta de dados orais 220

6.4 Outros resultados e hipóteses direcionados pelos *corpora*: esboços 221

6.5 Análise geral dos resultados 226

7. À GUISA DE CONSIDERAÇÕES FINAIS: DESDOBRAMENTOS FUTUROS E POSSIBILIDADES PARA O *POMMERSCHE KORPORA* 228

7.1 Possibilidades futuras para os *Pommersche Korpora* 229

REFERÊNCIAS BIBLIOGRÁFICAS 233

APÊNDICES 245

APÊNDICE A - Levantamento de algumas famílias pomeranas residentes e falantes de pomerano no leste de Minas Gerais (censo informal) 245

APÊNDICE B - Levantamento de algumas famílias pomeranas que emigraram do leste de Minas Gerais (censo informal) 246

APÊNDICE C – Procedimento metodológico para transcrições e compilações: Documento de conversão e convenção da escrita dos *corpora* do pomerano 247

APÊNDICE D - Lista de textos coletados e que compõem o PKE - *Pommersche Korpora Escritos* 249

APÊNDICE E – Questionário Semântico-Lexical (QSL) / Lexikalisch-Semantischer Fragebogen (LSF) 254

APÊNDICE F – Questionário Sociolinguístico – QS 268

ANEXOS 279
ANEXO A - Parecer consubstanciado do CEP - nº.1145346 279
ANEXO B - Die 40 (endgültigen) Sätze Georg Wenkers (1880) 283
ANEXO C - Mapas da Pomerânia 285

APRESENTAÇÃO

Nesta seção, apresento as motivações, de caráter pessoal, que me levaram a escolher o pomerano como tema desta pesquisa. Em virtude disso, opto por utilizar a primeira pessoa do singular para descrevê-las.

O contato, desde os 12 anos de idade, com um canal de TV alemã, o *Deutsche Welle* (DW), as observações dos sons e da escrita que ali se revelavam bem como a interação, desde essa época, com vizinhos e amigos, descendentes alemães provenientes da região sul do Brasil, fizeram com que eu me interessasse pela língua alemã. Além disso, com 14 anos, conheci alemães nativos, por meio de visitas que fizeram à comunidade evangélica da qual eu fazia parte, e, assim, estreitei os contatos orais com a língua, de modo a praticá-la face a face. A partir disso, pude testar o processo de aprendizagem do alemão que já tinha sido iniciado por mim, como autodidata, ao observar os padrões assistidos no DW e transcrevê-los e ao buscar material didático em alemão, como livros e apostilas, na biblioteca municipal da cidade de Uberlândia, visto que na época eu não tinha acesso à internet.

Aos quinze anos, comecei a estudar alemão na Central de Línguas da UFU, em um curso intensivo que tinha o objetivo de apresentar o conteúdo de um semestre em um mês. Dessa forma, além de me apaixonar definitivamente pelo idioma, percebi que o "alemão" dos descendentes do sul que eu conhecia era diferente do alemão ensinado no curso.

Logo, conheci um pomerano, com quem me casei aos 20 anos de idade, e, a partir disso, estreitei meu relacionamento com sua avó e seu pai, falantes de pomerano como língua materna. Assim, entrei em contato mais efetivo com o falar pomerano e fui percebendo algumas variações em relação ao alemão-padrão.

Em 2012, ao fazer estudos genealógicos e visitar cemitérios de imigrantes no sul do Brasil para reconstruir a árvore genealógica da família Beilke, proveniente da Pomerânia, me deparei com muitas inscrições nos túmulos, contendo escritas alemãs e suas variações e, então, pensei: por que não partir da morte para a vida e tentar reconstruir a histórias dessas pessoas ao pesquisar a língua que falavam? Certamente, estar diante de fontes tão ricas e quase inexploradas deu-me a inspiração necessária e forte motivação para começar a pesquisar o pomerano.

Percorri muitas cidades no Brasil em busca do pomerano e descobri uma região no leste de Minas Gerais com a presença de pomeranos. Sim, pomeranos em Minas Gerais! E, peregrinei por várias localidades no país, como na região do Vale do Rio Pardo/RS, onde também encontrei pomeranos.

Com as experiências que obtive ao conhecer os pomeranos no Brasil, identifiquei que eles possuem apreço pela natureza, são caprichosos nos cuidados com o lar, são pessoas simples, prezam pelo bem-estar em detrimento de consumismos, possuem muita sabedoria de vida e são bastante acolhedores. Aliás, demonstraram-se contentes por ser objeto de minha pesquisa. Acrescento que, nos lugares em que estive, o tempo parecia passar de forma diferente, sem a escravidão do relógio. Se, num primeiro momento, os pomeranos parecem desconfiados pelo estranhamento, no momento seguinte já estão prontos a fazer amizade e a ajudar ao próximo com solicitude e fraternidade. São falantes de uma variedade de língua tão rica, que talvez nem imaginem o quanto possuem um conhecimento precioso. Esses são os pomeranos e essas são as minhas primeiras impressões, extraídas de meus diários de campo.

Por tudo isso, decidi voltar à universidade, após sete anos sem estudar, para prestar a seleção do curso de mestrado e desenvolver esta pesquisa sobre o pomerano. Aqui estão as razões para uma patrocinense pesquisar o pomerano! *Up gooh*![1]

[1] Expressão equivalente às expressões *auf geht's* e *let's go*, significa "vamos lá!". Tradução nossa. Doravante todas as traduções serão nossas, exceto quando indicada uma autoria diferente, conforme as normas.

Velas brancas se embalam sobre o mar azul,
Gaivotas brancas voam na altura azul;
florestas azuis coroam dunas de areias brancas
Pomerânia, meu desejo está voltado a ti!

Agora estou em peregrinação, logo estou aqui, logo ali;
mas dentre todas as outras sou sempre levado a seguir adiante.
Até que em ti eu encontre novamente a minha paz,
eu envio a ti minhas canções, oh terra natal!

Weiße Segel wiegen sich auf blauer See,
weiße Möwen fliegen in der blauen Höh,
blaue Wälder krönen weißer Dünen Sand:
Pommerland, mein Sehnen ist dir zugewandt.

Jetzt bin ich im Wandern, bin bald hier, bald dort,
doch aus allen andern treibt's mich immer fort:
Bis in dir ich wieder finde meine Ruh',
send' ich meine Lieder dir, o Heimat, zu[2].

[2] *Das Pommernlied* de Gustav Adolf Pompe (1851) – Segunda e última estrofes retiradas do poema que contém cinco estrofes no total. O poema acrescido da melodia de Karl Gross (1818) ficou conhecido como o hino da Pomerânia. A menção às cores azul e branca são referências às cores da bandeira da Pomerânia.

1. INTRODUÇÃO

Nesta seção inaugural, apresentamos nosso tema, a fim de identificar nosso objeto de estudo: o pomerano. Em seguida, versamos sobre a justificativa da nossa pesquisa, informamos quais são as hipóteses que levantamos e expomos quais são os objetivos, geral e específicos, do nosso trabalho. Por último, demonstramos a forma como organizamos toda a nossa dissertação, sua estrutura e suas sequências.

1.1 Nosso tema e objeto de estudo

A pesquisa que nos propomos a desenvolver tem como tema o pomerano, que é uma variedade linguística germânica falada no Brasil e praticamente extinta nas regiões em que tivera origem – o território onde hoje estão o nordeste da Alemanha e o norte da Polônia.

Nosso objetivo geral é compilar *corpora*[3] – materiais autênticos – da variedade brasileira do pomerano para constituição dos *corpora* do pomerano. Portanto, nosso tema é o pomerano no Brasil e nossa proposta é coletar dados, organizá-los e permitir a descrição do nosso objeto de estudo por meio da Linguística de *Corpus*.

A referência à variedade pomerana é encontrada de diversas formas: *Pommersch,* forma alemã, termo coloquial para se referir a todos os dialetos pomeranos; *Pomeranian*, variedade inglesa nos EUA; *Plattdüütsch,* para os pomeranos-brasileiros; *Pomerisch*, como aparece no dicionário pomerano-português. E ainda é possível encontrar na internet as denominações *Pommerschplatt*, *Pommeranisch* e "alemão-pomerano".

1.2 Justificativa

Esta proposta de trabalho parte da consciência de que precisamos agir frente à possibilidade de desaparecimento da variedade pomerana, devido às constatações através do nosso contato, pelas visitas de campo e pelos estudos de autores como Pessoa (1995), Vogt (2001), Seyferth (2003), Dück (2005; 2008), Maltzahn (2011), Heinemann (2013), entre outros, que indicam que o pomerano corre risco de desaparecimento, visto que as novas gerações não estão fazendo uso do pomerano e que, ao morrerem os falantes mais idosos, o pomerano entra cada vez mais em declínio e pode cair em desuso, pela diminuição gradual do número de

[3] *Corpora*, termo originário do latim, é o plural de *corpus*, que em sentido restrito, significa um conjunto de textos. Definimos o que é *corpus* e o que são *corpora*, no capítulo 3, na seção 3.3.6, dos "princípios fundamentais e conceitos importantes", diretamente nos subitens 3.3.6.8 e 3.3.6.9.

falantes. Não estamos dando um veredicto ou lançando previsões, apenas salientando a necessidade de medidas de revitalização, ensino-aprendizagem e fomento do pomerano.

Ressalvamos também que, em localidades fora do eixo principal de pesquisas sobre o pomerano no Brasil, a saber Espírito Santo e Santa Catarina[4], outras regiões, como as que escolhemos estudar – o leste de Minas e o interior do Rio Grande do Sul –, são pouco abordadas por pesquisas em estudos linguísticos. Acreditamos que nessas localidades, as pesquisas sobre o pomerano ainda não foram satisfatoriamente realizadas, por exemplo, carecem de um censo linguístico para levantamento do número de falantes e de estudos que identifiquem se os falantes são ativos ou passivos, qual o vocabulário ativo, se as novas gerações ainda estão aprendendo o falar, em qual estágio está o processo de desaparecimento ou manutenção da fala pomerana, quais os sentidos, usos, contatos, etc. Por isso, acreditamos ser importante voltar o nosso olhar para localidades menos privilegiadas e abordadas por pesquisas. Nossa pesquisa justifica-se pela necessidade de fazer um registro do pomerano e colaborar para sua documentação.

Outra razão em contribuir para a documentação do pomerano está no fato de que, na região da antiga Pomerânia oriental, onde hoje é a Polônia, o pomerano oriental não é mais falado (GRANZOW, 2009), salvo alguns descendentes dos pomeranos orientais que ainda reúnem-se em encontros periódicos para falar em pomerano (WANGERINER TREFFEN, 2015).

As problemáticas de pesquisa impõem-se a partir das lacunas deixadas por outros trabalhos e por questões que não tenham sido satisfatoriamente estudadas em determinadas localidades, sob o nosso ponto de vista. É isso que fizemos ao escolher estudar o pomerano abordando também as localidades menos exploradas ou referenciadas como lugares de presença pomerana, nas quais a população é remanescente da imigração pomerana, mesmo que não tenha ainda recebido grandes menções e reconhecimentos pela mídia e pelas pesquisas, como é o caso de Itueta/MG.

Acreditamos que um conjunto de *corpora* do pomerano, pensado como acervo, pode permitir a replicação de pesquisas sobre essa variedade de língua, tanto em estudos baseados na LC quanto em estudos com outras abordagens e metodologias. Consideramos que é possível

[4]As cidades mais conhecidas por conter falantes de pomerano são Pomerode/SC e Santa Maria de Jetibá/ES, mas há várias outras localidades no Brasil nas quais o pomerano também é falado, tanto dentro desses estados quanto em outros estados do Brasil, conforme levantamento que segue no capítulo 2.

o desenvolvimento de futuras descrições e análises do pomerano com base no levantamento de dados empíricos reunidos em *corpora*.

Portanto, este trabalho justifica-se pela necessidade de criar condições e opções para o estudo do pomerano e pela necessidade de reunião e de compilação de registros, a fim de fornecer dados para a produção de materiais didáticos e/ou descritivos, por exemplo, que possibilitem, inclusive, outros estudos a respeito dessa variedade e o ensino do pomerano.

Consideramos também que há uma necessidade de possibilitar meios de acesso a textos originalmente escritos em pomerano, que mostram o uso real da língua e, assim, oferecer métodos para que se proponha futuramente o ensino do pomerano com base na escrita germânica, como forma de resgate da cultura escrita que foi perdida – da qual acreditamos que as gerações anteriores tinham conhecimento – e como uma medida de fomento ao pomerano, combatendo assim o processo de desaparecimento devido à falta de aprendizagem e de uso do pomerano pelas novas gerações.

Para discutir, brevemente, a efetividade dessa possibilidade de um ensino que resgate o pomerano, vamos citar o exemplo do galês (*Cymraeg* ou *Welsh*). Em *North Wales*, na região norte do país de Gales, o galês[5] entrou em declínio e precisou ser fortemente reavivado. Segundo Morris (2013), o galês sofreu um processo de anglicização, intensificado por volta de 1536, quando algumas terras galesas foram anexadas à coroa inglesa. O inglês tornou-se a língua dos negócios, e com essa necessidade, os galeses precisaram se tornar bilíngues. Assim, o galês foi perdendo o prestígio linguístico, a subjugação do País de Gales pela Inglaterra levou a um declínio do galês. A Revolução Industrial também acelerou o declínio da língua e foi preciso que o governo estabelecesse leis que protegessem a língua e elaborasse um planejamento para sua revitalização. No final do século XIX, passou a haver a escolaridade obrigatória em galês.

Segundo o censo realizado em 1891 pelo governo britânico, o galês era falado por 51,2% dos habitantes, que eram monolíngues em galês. Em 2011, apenas 19% da população total do país de Gales falava galês e já era bilíngue. Houve um declínio acentuado em um século. Então, as medidas de revitalização da língua, principalmente na área da educação, resultaram no desenvolvimento de um sistema de ensino em galês, que é consideravelmente abrangente, pois

[5] Língua celta do grupo das línguas celtas do sul, do grupo *P-Celtic*, pertencente à família das línguas Indo-Europeias. Era falada antes da ocupação romana da ilha e sobreviveu após as invasões dos anglo-saxões. A língua *welsh* é falada por cerca de 500.000 mil pessoas, na maior parte da Grã-Bretanha, principalmente no país de Gales, onde é uma língua minoritária. Mas também em outras partes do mundo para onde foram imigrantes, como na região da Patagônia, na Argentina.

Morris (2013) relata que hoje é possível que uma criança estude em escolas galesas durante toda a sua vida escolar.

Há redes de escolas de galês em todo o país e é possível para uma criança ir do maternal até a pós-graduação estudando na língua galesa. O galês no ensino superior foi recentemente reforçado com a criação do *Coleg Cymraeg Cenedlaethol* (Colégio Nacional Galês), que financia e promove cursos de ensino superior galeses e facilita a formação de docentes em galês. No País de Gales, existem também programas de rádio e televisão em galês.

Um sistema de ensino do pomerano no Brasil, semelhante ao que foi implantado no País de Gales, seria uma grande conquista. Mas, infelizmente, os pomeranos ainda estão lutando por condições mínimas para o ensino do pomerano, como subsídios financeiros para treinamento e contratação de professores, desenvolvimento de materiais didáticos e espaço nas grades curriculares. Podemos exemplificar isso por meio de iniciativas de alguns professores para ensinar o pomerano nas escolas, como fez Kuhn Silva (2012).

A compilação dos *corpora* do pomerano, a partir dos textos autênticos encontrados e do vocabulário extraído de registros falados e transcritos, permitirá a futura criação de um banco de dados linguísticos do pomerano, reunindo informações e fontes primárias sobre o tema.

O pomerano ainda não foi satisfatoriamente examinado, especialmente com foco em algumas regiões do Brasil. No caso de Itueta/MG, é recente a descoberta desses descendentes e há poucos trabalhos sobre o assunto. De forma geral, os pomeranos ficaram isolados em seu grupo por muito tempo até serem descobertos pela mídia (SEIBEL, 2010). Granzow (2009, p.161-162) também relatou ter encontrado comunidades de pomeranos bem fechadas e isoladas.

No caso específico de Minas Gerais, houve um isolamento geográfico, devido à necessidade de se atravessar o Rio Doce de balsa para se chegar até eles, em períodos em que as estradas da zona rural norte de Itueta não estavam transitáveis. O pomerano falado por eles pode apresentar peculiaridades – como o contato com o regionalismo mineiro do português – que mereçam um estudo comparativo com aquele falado na região do Vale do Rio Pardo/RS - espaço povoado por muitos imigrantes pomeranos, que ainda guarda marcas da cultura de seus colonizadores. Seus descendentes ainda vivem nessas localidades, incorporados à cultura brasileira sem abandonar suas manifestações linguísticas, que precisam ser verificadas e descritas.

Não foram localizados meios de promoção e estudo do pomerano em algumas regiões do Brasil, como no entorno do Vale do Rio Doce no leste de Minas Gerais e no Vale do Rio Pardo no Rio Grande do Sul, por isso um estudo comparativo entre o pomerano falado nessas regiões pode ser realizado, com base na variação do uso de itens lexicais que podemos encontrar

por meio da compilação de *corpora*. Essas localidades situam-se fora do eixo principal de pesquisas sobre o pomerano no Brasil (a saber, Espírito Santo e algumas localidades sulistas como Pomerode e Blumenau). É necessária e desejável uma pesquisa que identifique melhor a situação do pomerano nessas regiões e também iniciativas que promovam o resgate e o fomento do pomerano nas regiões menos pesquisadas.

A partir de nossas reflexões, baseadas nas ideias de Barbosa (1991), Biderman (2001) e Beilke (2013), inferimos que estudar o pomerano é importante para a comunidade e para os pesquisadores, pois o pomerano, como variedade de língua, comporta toda uma forma de conceber o mundo e expressa uma forma de pensar, sentir, agir, organizar, nomear, significar, interagir e experenciar. O pomerano carrega consigo toda uma riqueza cultural e um conhecimento histórico-linguístico, que justifica a razão de seu registro, preservação e fomento. Lembramos também a importância da língua para a constituição identitária de um povo, um grupo social e/ou etnia, bem como para a transmissão de modos de viver e se comunicar às novas gerações. Estudar o pomerano também é útil para reafirmar a importância do uso linguístico e da pluralidade de falas que mantêm o léxico ativo e enriquece as relações humanas; como forma de resistência às padronizações prescritivas e imposições monolinguísticas. Resgatar e registrar o pomerano também é uma forma de oportunizar o conhecimento do vocabulário pomerano por outras culturas, enquanto herança cultural e tesouro linguístico dos pomeranos e como patrimônio da sociedade. É importante para conhecermos a forma que o léxico assume na cultura pomerana. Enfim, salvaguardar o léxico pomerano, ao incentivar e fomentar o seu uso e registro, é uma forma de ampliar o conhecimento das relações entre língua, sociedade e cultura.

Em virtude de tudo que foi mencionado, justificamos a relevância da nossa pesquisa e propomos um estudo do pomerano por meio da LC, que nos fornece métodos e ferramentas úteis para trabalhar com um conjunto de dados linguísticos escritos e orais do pomerano.

1.3 Hipóteses

Levantamos quatro hipóteses que estão discriminadas abaixo:

Primeira hipótese

O pomerano não é ágrafo no Brasil, pois a coleta e compilação de *corpora* demonstra que existe uma forma de escrita com traços dialetais pomeranos, presentes em cartas e textos advindos de descendentes de pomeranos.

O conhecimento de uma grafia germânica foi se perdendo ao longo das gerações que se sucederam pela não continuidade do estudo em língua alemã e também devido ao processo de escolarização brasileiro e à aprendizagem do português. *Corpora* escritos do pomerano demonstram a presença de alguma escrita pomerana, portanto, não houve caso de agrafia, mas a perda de cultura escrita.

Segunda hipótese

No início da imigração, as comunidades pomeranas estabeleceram-se no campo, em áreas rurais para os trabalhos na área da agricultura. Como os pomeranos ainda são considerados um grupo predominantemente camponês, acreditamos que os falantes encontram-se apenas no meio rural, sendo inexistente a presença deles na zona urbana de algumas localidades, como Itueta/MG. Essa hipótese aplica-se aos dois principais pontos de coleta: Vale do Rio Pardo/RS e Vale do Rio Doce/MG.

Terceira hipótese

Nos pontos de pesquisa selecionados como principais – Vale do Rio Pardo/RS e Vale do Rio Doce/MG – houve contato de pomeranos com outras etnias, portanto, os dados de *corpora* apresentam evidências desse contato interdialetal e também evidências de variação lexical na comparação entre essas regiões. Essa hipótese deve-se ao fato de que houve casamentos interétnicos entre alemães de regiões diferentes, também contato entre eles nas colônias e nas localidades em que residiam os imigrantes e, posteriormente, seus descendentes pomeranos.

Quarta hipótese

O contato dos descendentes de pomeranos no Brasil com o português e o prestígio do português como língua oficial do país permitiram uma interferência português-pomerano a ponto de o pomerano falado atualmente no Brasil já representar uma variedade brasileira do pomerano, na qual o contato de línguas[6] tenha propiciado usos com empréstimos lexicais do

---

[6] O contato de línguas é "um fenômeno com que muitas etnias convivem no dia-a-dia de sua história linguística, pois as faces das línguas/variedades são reveladas nesse encontro e põem à mostra a complexidade e dinamicidade de uma língua natural em funcionamento. Várias situações complexas surgem no processo da interação discursiva e se tornam desafios para os estudos linguísticos. Embora o fenômeno seja de natureza linguística, existem fatores sociais, psicológicos, culturais, históricos, políticos e geográficos que influenciam a configuração linguística de um grupo social. [...] Na medida em que se intensifica o uso da língua majoritária, há uma tendência a integrar os elementos da outra língua na fala dos bilíngues. Isso é possível, uma vez que as línguas em contato estão transitando no repertório linguístico dos bilíngues e, muitas vezes, na interação discursiva. Os elementos

português no sistema pomerano. A hipótese, nesse caso, é a de que os *corpora* do pomerano apresentam indícios na formação de frases e na inovação de itens lexicais que demonstrem essa relação de "mistura" pomerano-português.

1.4 Objetivos Geral e Específicos

Detalhamos, a seguir, qual é o objetivo geral e quais são os objetivos específicos do nosso trabalho.

1.4.1 Objetivo Geral

O nosso objetivo geral e principal é a compilação de *corpora* do pomerano para a composição de um conjunto formado por *corpora* escritos e um *corpus* oral, ou seja, coletar dados que contenham léxico do pomerano e constituir um acervo de *corpora* que possibilite a futura criação de um banco de dados do pomerano, a fim de contribuir para o registro, para a documentação e para a descrição do léxico do pomerano no Brasil.

Integra o nosso objetivo principal a própria coleta, compilação[7], organização, classificação e a transcrição de material escrito e oral do pomerano, visto que se trata de proposta ambiciosa, tendo em vista a dificuldade de transcrição de dados orais dialetais. E, também, a necessidade de desenvolver uma metodologia apropriada para conseguir constituir esses *corpora*.

Ressalvamos que um levantamento do léxico pomerano por meio da compilação de *corpora* e que não se restrinja à coleta no estado do Espírito Santo é algo que ainda não foi feito (conforme justificativa apresentada no tópico 1.2 e discussão apresentada no Capítulo 2, tópico 2.4), sendo necessário desenvolver etapas metodológicas, um método de transcrição, de normatização, além dos instrumentos para a coleta de dados orais, como a elaboração/adaptação dos questionários, por exemplo. Por tudo isso, o objetivo central do nosso trabalho limita-se à própria criação do conjunto de *corpora* do pomerano.

---

pertencentes à língua nativa e à outra língua se unem e se compõem, atualizando a fala de acordo com o contexto e a necessidade do falante" (TAKANO, 2013, p. 68 e 75).

[7] Consideramos que coleta é colher material e compilação é transferir os materiais compilados para o computador por meio da digitação, de forma que são dois procedimentos diferentes. Em nossa pesquisa foi necessário realizar trabalho de campo para coletar material pessoalmente, para depois transpor os materiais coletados para o computador. Nessa perspectiva, a coleta e a compilação só coincidem quando é realizada a compilação direta (termo que definimos no tópico 4.1.2) como comumente ocorre na Linguística de *Corpus*.

1.4.2 Objetivos Específicos

Os objetivos específicos de nossa pesquisa são dois. O primeiro é verificar material compilado em pomerano, a fim de tentar identificar por meio de indícios escritos e falados a sobrevivência ou o desaparecimento do pomerano nas localidades de coleta de dados dentro das regiões abrangidas pela pesquisa, no caso, dos Vales do Rio Doce/MG e do Rio Pardo/RS.

No caso das regiões nas quais focamos, uma das necessidades mais urgentes é averiguar a possibilidade de ainda encontrar falantes ativos de pomerano, em lugares em que grande parte da população é descendente da etnia pomerana, mas não tem sido objeto de estudos. Para alcançar esse objetivo, nos propomos a realizar entrevistas e a verificar a existência de algum material, antigo ou recente, escrito em pomerano.

O nosso segundo objetivo específico é comparar, brevemente, o pomerano coletado nas duas regiões principais que nossa pesquisa enfoca (Minas Gerais e Rio Grande do Sul), durante o processo de compilação dos *corpora* e de geração e comparação das listas de palavras dos *corpora*, para verificar proximidades e distanciamentos entre o léxico pomerano utilizado nas referidas regiões, além de permitir testar minimamente as potencialidades futuras do nosso conjunto de dados.

Neste período, realizamos a introdução do nosso traballho, justificamos a importância da nossa pesquisa, informamos quais são as hipóteses que levantamos e os nossos objetivos, geral e específicos.

A seguir, descrevemos a forma como esta dissertação está organizada.

1.5 Organização da dissertação

Neste momento, realizamos a descrição de como nossa dissertação está organizada. Esta dissertação está organizada em sete capítulos. O capítulo um é este e refere-se à apresentação do tema da nossa pesquisa.

No capítulo dois, tratamos do pomerano na Europa, sua origem e fazemos uma breve contextualização do nosso objeto de estudo. Em seguida, resumimos algumas informações sobre a imigração germânica que trouxe essa variedade linguística para o Brasil. Então, falamos do pomerano no Brasil, momento no qual discutimos a definição do pomerano e nos posicionamos a respeito. Nessa seção, enfrentamos a polêmica de o pomerano ser língua ou dialeto. Posteriormente, partimos para a apresentação de algumas pesquisas realizadas sobre o pomerano no Brasil, citando brevemente alguns autores que são referência em relação ao tema.

No capítulo três, construímos nossa fundamentação teórica, tratando pontualmente de alguns conceitos básicos como Lexicologia e léxico, Linguística de *Corpus* e Sociogeolinguística. E encerramos o capítulo, indicando como as teorias, que subsidiam as abordagens-metodologias que escolhemos, baseiam o nosso trabalho de constituição dos *corpora* do pomerano no Brasil.

No capítulo quatro, iniciamos o desenvolvimento efetivo do trabalho, descrevendo detalhadamente nossa metodologia. Como foram os processos de compilação e construção dos *corpora*, o passo a passo, as dificuldades, os problemas encontrados e as soluções adotadas, as fases de coleta (que foram duas), os instrumentos de coleta do *corpus* oral (que foram os questionários QS – Questionário Sociolinguístico e o LSF – *Lexikalisch-Semantischer Fragebogen*, versão alemã do QSL – Questionário Semântico Lexical), além da ferramenta utilizada, o *WordSmith Tools*® (WST). Falamos dos procedimentos de conversão e convenção da escrita, devido à diversidade de formas escritas do pomerano e da organização dos *corpora* do pomerano.

No capítulo cinco, expomos os resultados do nosso trabalho, mostrando o conjunto dos *corpora* constituídos, sua nomeação, logotipo, arquitetura, demonstramos com algumas imagens os *corpora* constituídos, descrevemos seus dados estatísticos, bem como descrevemos detalhadamente cada um dos *corpora* compilados e fazemos análises superficiais de cada um.

No capítulo seis, fazemos a avaliação dos resultados. Primeiramente, classificamos os *corpora* compilados, depois, esboçamos algumas análises preliminares a partir do *Pommersche Korpora* (PK) e comprovamos como os dados observados e extraídos do PK permitem falar em variedade brasileira do pomerano - VBP, ou, *Brasilianisch-Pommersch* (BP). Tratamos também, neste momento, de esclarecer as hipóteses que foram confirmadas e/ou negadas pelos resultados da pesquisa. Além disso, fazemos análises breves de fatos para os quais o PK nos direcionou.

No capítulo sete, a última seção deste trabalho, à guisa de considerações finais, fazemos uma prospecção dos passos futuros para dar continuidade a este trabalho, fazemos um balanço do trabalho que realizamos e encerramos nossa dissertação indicando algumas possibilidades de desdobramentos a partir desta pesquisa.

Ao fim do presente volume, encontram-se referências, apêndices e anexos contendo dados relevantes e complementares sobre nossa pesquisa.

A serguir, vamos tratar do pomerano e sua contextualização, na europa, no Brasil, sua definição, mapeamento e pesquisas sobre o tema.

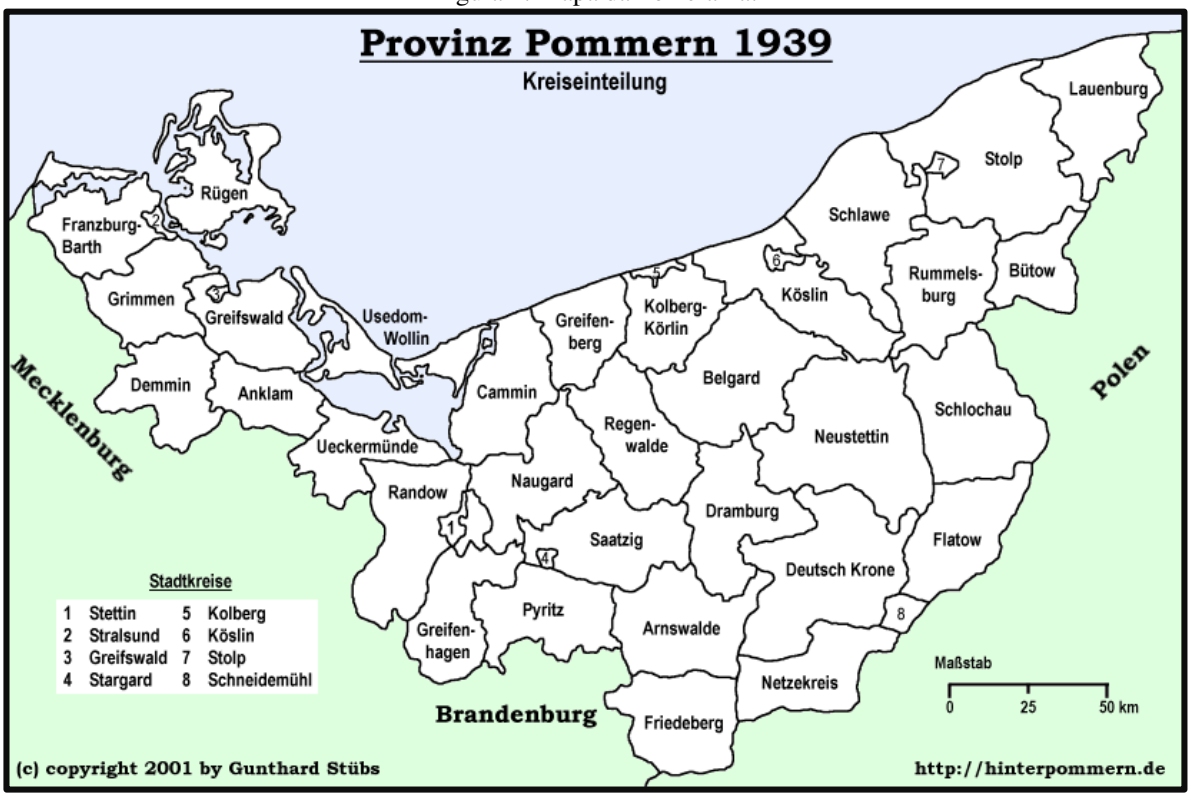
2. O POMERANO NA EUROPA E NO BRASIL

Neste capítulo vamos falar brevemente do pomerano na Europa, seu lugar de origem. Em seguida, trataremos do processo de imigração, para explicar como os pomeranos estabeleceram-se no Brasil. Na sequência, definiremos o pomerano e trataremos do nosso posicionamento a respeito dele. Por fim, apresentaremos um levantamento do pomerano no Brasil e faremos um breve balanço das pesquisas realizadas no país, sobre o pomerano.

2.1 O Pomerano na Europa

O pomerano é oriundo da Pomerânia ou *Pommerland*, um território da região nordeste do Reino da Prússia (atual Alemanha); porém a maior parte dessas terras foi perdida pelos autóctones pomeranos alemães ao fim da Segunda Guerra Mundial e hoje pertence à Polônia. A Pomerânia era composta pelas *Vorpommern* e *Hinterpommern*, a Pomerânia anterior (ocidental) e a Pomerânia posterior (oriental), conforme é possível visualizar na Figura 1, a seguir:

Figura 1: Mapa da Pomerânia.



Fonte: http://hinterpommern.de.

No mapa apresentado na Figura 1, a Pomerânia anterior está limitada a oeste por *Mecklenburg* e a Pomerânia posterior, limitada a leste por *Polen*, a Polônia. E a sudoeste limitada pelo estado alemão de *Brandenburg*. A divisão entre as Pomerânias situava-se

passando pelos territórios limítrofes de Usedom-Wollin, Ueckermünde, Randow e terminava dividindo o território entre Greifenhagen e Pyritz. A linha divisória pode ser visualizada por meio dos dois mapas constantes no ANEXO C, expostos ao final deste volume.

A *Hinterpommern* (Pomerânia posterior) já existia como província desde 1653 e permaneceu até 1815 quando, unida à região anterior, tornou-se a Província Pomerana da Prússia (*Preußische Provinz Pommern*, 1815–1945).

A região oriental da Pomerânia foi alvo de muitas disputas ao longo da história. Segundo Böttcher (2005), a Pomerânia Oriental deixou definitivamente de fazer parte do território alemão em 1945, com o estabelecimento da linha *Oder-Neisse*[8]. Ainda segundo Böttcher, na configuração política atual da Alemanha, restou apenas a Pomerânia anterior (*Vorpommern*). Atualmente, essa região é chamada *Mecklenburg-Vorpommern*, pois formou um estado regional unindo-se administrativamente a Mecklemburgo. Conforme explica Böttcher (2005), com o fim da segunda guerra mundial, a população da Pomerânia oriental foi aniquilada ou expulsa para as regiões central e oeste da Alemanha. Atualmente, predomina na região da antiga Pomerânia oriental população de etnia e língua polonesas. Essa região teve suas fronteiras confirmadas recentemente, no ano de 1990, por meio de um tratado germano-polonês.

Vieram para o Brasil imigrantes pomeranos das duas regiões da Pomerânia, mas a maioria dos imigrantes era procedente da região posterior da Pomerânia, ou seja, a *Hinterpommern* (Granzow, 2009), que possuía uma extensão territorial e densidade populacional maior que a Pomerânia ocidental. A Pomerânia oriental era uma região agrária e tinha na pesca, nos frutos do mar e na agricultura, seus principais gêneros econômicos.

2.2 História da Imigração: como os pomeranos vieram parar aqui?

A imigração pomerana para o Brasil data de mais de 150 anos. Em 2009, foram realizadas as comemorações dos 150 anos de Imigração Pomerana no Brasil. Conforme Granzow (2009, p. 11), o marco oficial da imigração pomerana é 1859: "no longínquo dia 28 de junho de 1859, aportaram em Vitória 117 imigrantes saídos do porto de Hamburgo em 27 de abril daquele mesmo ano". Porém, se consultarmos alguns documentos, como as listas de passageiros dos navios e as confrontarmos com a origem dos imigrantes, contidas nos registros

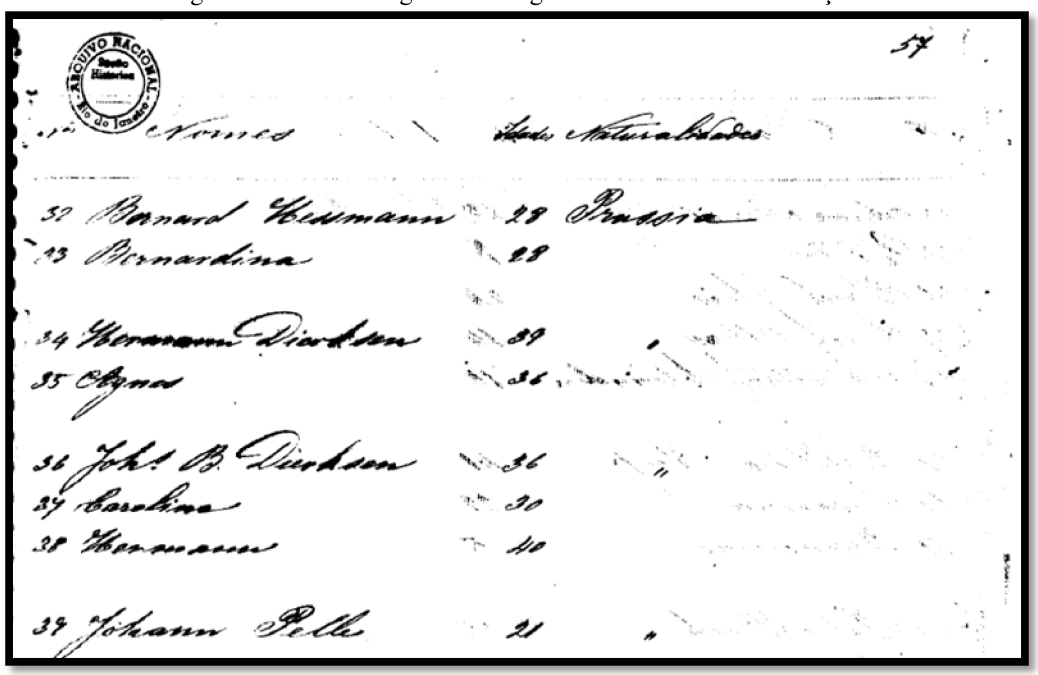
[8] Retornando a Böttcher (2005), podemos esclarecer que a Linha *Oder-Neisse*, foi uma forma de demarcar as fronteiras políticas alemãs, após 1945, com o fim da Segunda Guerra Mundial. O estabelecimento da Linha se deu por meio do marco geográfico dos rios Oder e Neisse, situados na divisa entre o nordeste da Alemanha e o noroeste da Polônia. As fronteiras sofreram uma alteração posterior que implicou na perda das cidades de Stettin e Swinemünde pela Alemanha, que passaram à jurisdição da Polônia.

eclesiásticos, por exemplo, encontraremos famílias pomeranas que imigraram antes de 1859 (como em 1858 para São Lourenço do Sul/RS, por exemplo). Ainda segundo o autor, "na década de 1870, novos imigrantes pomeranos chegaram ao Estado [do Espírito Santo]" (GRANZOW, 2009, p. 11).

Observamos que não existem dados precisos sobre a data em que os primeiros pomeranos vieram para o Brasil, porque em muitos casos eram registrados apenas como prussianos nos documentos de entrada no país, o que dificulta também os cálculos sobre a quantidade de pomeranos que emigraram para o Brasil. Segundo Granzow, "encontramos muita literatura sobre a imigração alemã no Brasil, mas quase nada sobre a imigração pomerana" (GRANZOW, 2009, p. 93-94). O autor também confirma nossa afirmação de que os pomeranos eram identificados como prussianos: "de acordo com um relatório de 1860, moravam entre os alemães 174 prussianos, como se referiam aos pomeranos na época" (GRANZOW, 2009, p. 177).

Para exemplificar o que afirmamos, exibimos, a seguir, duas fontes primárias; na primeira (Figura 2) é informada a entrada dos imigrantes no Brasil, listando a naturalidade de todos eles simplesmente como "Prússia", conforme é visto na imagem a seguir.

Figura 2: Livro de Registro da "Agência Official de Colonisação".



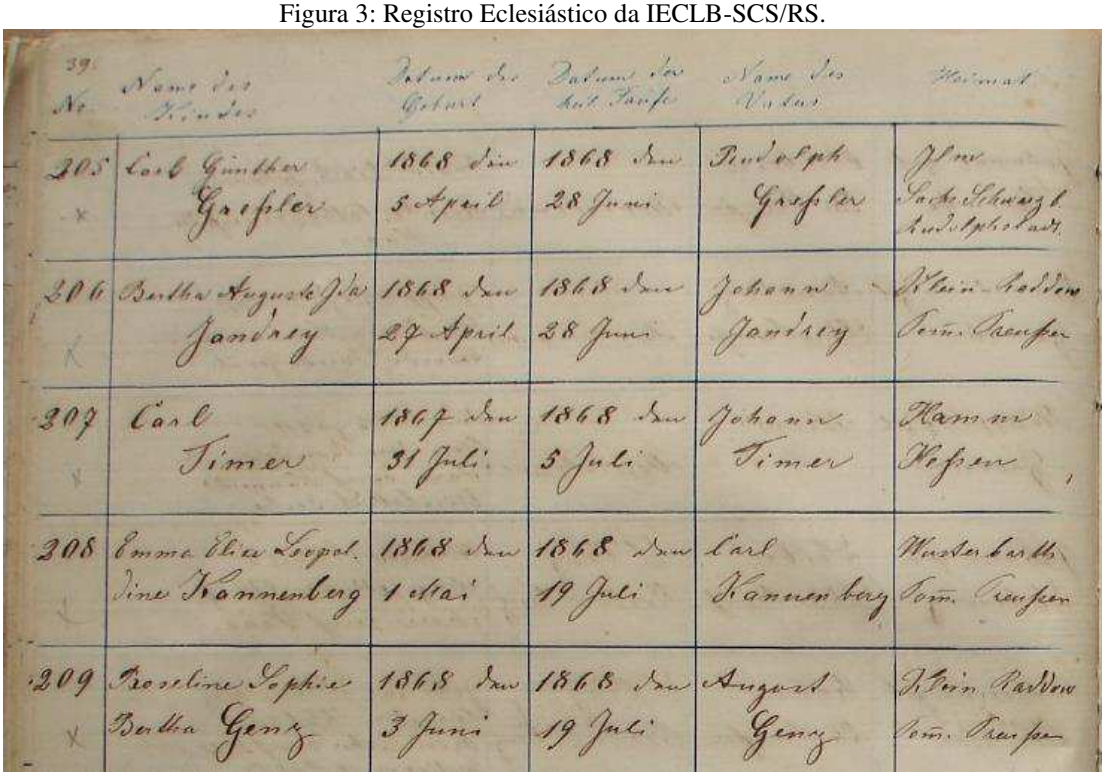
Fonte: Arquivo Nacional (1859). Obtido por meio da Lei de Acesso à Informação.

No documento exibido na Figura 2, a primeira informação à esquerda é o número do registro na lista de imigrantes que entraram no país, seguido do nome do imigrante;

transcrevemos a primeira linha, como exemplo: *nº 32 Bernard Hessmann, Idade 28, Naturalidades: Prussia* (seguidos de dois traços símbolos de idem, abaixo de Prússia na coluna "Naturalidades", para indicar a repetição da origem dos outros imigrantes listados).

Na segunda fonte primária (exibida na Figura 3), na qual constam os registros de batismos dos filhos dos imigrantes, já nas colônias brasileiras, é citada a origem dos imigrantes, conforme segue:

Figura 3: Registro Eclesiástico da IECLB-SCS/RS.



39 No.	Name des Kindes	Datum der Geburt	Datum der Taufe	Name des Vaters	Heimat
205 ×	Carl Günther Graefler	1868 den 5 April	1868 den 28 Juni	Rudolph Graefler	Pfarr Sachs Schwarzb. Rudolphstadt
206 ×	Bertha Auguste Joh. Jandrey	1868 den 29 April	1868 den 28 Juni	Johann Jandrey	Klein Radow Pom. Preußen
207 ×	Carl Timer	1867 den 31 Juli	1868 den 5 Juli	Johann Timer	Hamm Posen
208 ×	Emma Elise Leopol. Line Kannenberg	1868 den 1 Mai	1868 den 19 Juli	Carl Kannenberg	Neuenbarth Pom. Preußen
209 ×	Caroline Sophie Bertha Genz	1868 den 3 Juni	1868 den 19 Juli	August Genz	Klein Radow Pom. Preußen

Fonte: Acervo de Registros Eclesiásticos do Vale do Rio Pardo/RS (1878).

Na figura acima, podemos observar que, na sexta coluna, contada da esquerda para a direita, intitulada *Heimat* (terra natal/pátria), constam nos números de registro 206, 208 e 209, apenas *Pom. Preußen* (Pomerânia, Prússia), o que indica a região pomerana da Prússia, mas, ainda assim, não identifica a localidade de origem com precisão.

O Brasil começava a se abrir para receber imigrantes desde o início do século; o governo português, em 1808, já tomava medidas para incentivar a imigração como "o decreto de 1808 que estabelecia o direito de propriedade de terra a estrangeiros" (DIETRICH, 2001, p. 5).

De acordo com Dietrich (2001), entre 1815 e 1816, houve o fechamento comercial da Europa oriental e a elevação das tarifas aduaneiras, o desemprego na Prússia aumentou, o sistema agrário estava em crise e o custo de vida estava alto. Em condições precárias de fome e falta de recursos, os camponeses começaram a ver a emigração como solução, pois não tinham

expectativa de trabalho. As famílias que trabalhavam no ramo das atividades comerciais também foram prejudicadas e, sem condições para exercer seu comércio, também decidiram emigrar.

Segundo Rodrigues (2008), a primeira imigração de germânicos para o Brasil, de uma maneira organizada[9], foi a de imigrantes suíços e se deu a partir da criação da colônia de Nova Friburgo, em 1819, na região da então capital Rio de Janeiro. Geralmente, a imigração alemã para o Brasil é dividida em três fases. Os pomeranos são identificados como pertencentes à segunda fase de imigração, salvo exceções e os problemas de identificação da origem (aqueles que vieram antes, identificados como prussianos), conforme já foi explicado neste trabaho.

Dentre as motivações para o Brasil atrair e receber imigrantes, podemos listar, retornando aos trabalhos de Dietrich (2001) e Rodrigues (2008), a necessidade de serviços e profissões que aqui ainda não existiam, de mão de obra qualificada, de produção de ferramentas e de técnicas de trabalho, inclusive técnicas de construção civil e, principalmente, de plantio. A imigração também era estimulada no período colonial para ocupar, colonizar e proteger vastos territórios do país considerados ainda inexplorados.

Resumindo as ideias de Rodrigues (2008, p. 37), além das mudanças no regime do trabalho, com o fim do regime escravista no Brasil, o país tinha inúmeras demandas internas, não só a necessidade de ocupação do território, bem como a passagem da economia do açúcar, que era o primeiro artigo na balança comercial, para a economia do café. O café era uma economia crescente e um produto de exportação. Esse conjunto de fatores favoreceram ações efetivas no recrutamento da força de trabalho. O governo brasileiro e também os fazendeiros já percebiam o problema da escassez de mão de obra.

Segundo Witter (1986), a respeito da segunda fase da imigração germânica para o Brasil:

> A imigração alemã foi retomada após 1845, tendo em vista os interesses da política de colonização do Império, que podem ser constatados em vários relatórios e obras publicadas por diplomatas, conselheiros e técnicos do Ministério da Agricultura, Comércio e Obras Públicas: os alemães eram considerados bons agricultores, imigrantes ideais para povoar vazios demográficos no regime da pequena propriedade; por outro lado, os estados alemães, principalmente a Prússia, eram vistos como países de emigração, e

[9] Ressalvamos que consta na historiografia a chegada de alemães dentre a frota de Cabral, em 1500, segundo confirma Tresspach (2013). Também vieram alemães em 1530, com Martim Afonso de Sousa (que almejava assentar germânicos no Brasil para trabalhar nos engenhos de açúcar) para se estabelecerem na Vila São Vicente (hoje região de SP). Hans Staden que veio em 1549 era alemão e lembramos também que Maurício de Nassau-Siegen (que chegou ao Brasil em 1637) era alemão. A data oficial da imigração alemã (diversas etnias alemãs) para o Brasil é 25 de julho de 1824.

> acreditava-se que parte do fluxo dirigido para os Estados Unidos da América poderia ser desviado para o Brasil, caso fossem tomadas medidas adequadas (WITTER, 1986, p. 1-2).

Deste modo, localizamos a chegada de pomeranos, entre 1856 e 1858, aproximadamente, posicionada na segunda fase de incentivo à imigração, voltada ao trabalho nas lavouras de cana de açúcar e principalmente café.

Em meados do século XIX, no contexto do Império Brasileiro, as lavouras estavam em plena expansão, tínhamos um aumento dos cafeicultores ao passo que estava escassa a mão de obra escrava, em parte, devido às leis como, por exemplo, o Bill Aberdeen[10] decretado pela Inglaterra, e a Lei Eusébio de Queiroz, em 1850, que determinava a extinção do tráfico de escravos. Essa lei não acabou com a escravidão[11], mas dificultou a aquisição de escravos, enfraquecendo o sistema da época, contexto que gerou a busca por alternativas para obtenção de mão de obra para as lavouras brasileiras. A respeito disso, Witter (1986) explica:

> A crise da mão-de-obra no Brasil exigia soluções urgentes, a fim de que a 'grande lavoura' não sofresse demasiadamente a carência do elemento humano necessário ao seu funcionamento. As soluções para o problema foram sendo tentadas. A primeira alternativa foi a de estabelecer-se uma corrente contínua de imigração europeia, destinada a suprir o déficit de trabalhadores, principalmente nas áreas de expansão dos cafezais (WITTER, 1986, p. 1-2).

Com base na citação acima, percebemos que nessa época a crise de mão de obra era um problema fundamental para o desenvolvimento da economia brasileira. Expostas as razões que permitiram que imigrantes, dentre eles pomeranos, fossem desejáveis no Brasil, vejamos agora o contexto que propiciou o interesse de esses imigrantes viajarem para o Brasil.

Na Europa, o que teria influenciado a emigração para o Brasil e quais as condições que instigaram os pomeranos a emigrar? Em geral, as obras que tratam da emigração e da imigração descrevem a necessidade de subsistência como principal fator de impulsionamento para emigrar e imigrar, bem como a pressão das condições sociais, econômicas e políticas de seus locais de origem.

Consideramos útil listar alguns fatores que contribuíram para a emigração de prussianos para o Brasil, com base em Rodrigues (2008): crises econômicas, população numerosa e pouca terra para plantio, altos impostos e fome impeliam os camponeses a se aventurar nas Américas.

---

[10] Promulgado pela Grã-Bretanha, em 1845, proibia o comércio de escravos entre África e a América.
[11] A abolição da escravidão só ocorreu efetivamente a partir de 1888, por isso lembramos que o fim do tráfico de escravos não foi o fim da escravidão.

O Brasil foi divulgado como uma “terra de oportunidades” e de esperanças de uma vida melhor. Por isso, muitos camponeses vieram na esperança de obterem um pedaço de terra para praticar a agricultura familiar.

A crise na Europa e as motivações dos europeus (dentre eles teutos pomeranos) para deixarem sua terra natal são questões levantadas por Martins e Cohen (2000). Segundo as autoras, houve uma “febre de emigrar”, pois os camponeses buscavam trabalho nas cidades, mas havia muitos trabalhadores desempregados, um verdadeiro excesso de mão de obra. Os pequenos agricultores estavam endividados, houve secas e muitas plantações foram perdidas, ocorria um empobrecimento crescente da população. Nesse contexto, começaram a ser distribuídos folhetos de propaganda em regiões pobres, anunciando a abertura do Brasil para a imigração. As autoras relatam também que os governos locais davam suporte para o recrutamento desses camponeses para servirem de mão de obra em outros países.

Conforme o relato de Alves (2003), a situação socioeconômica dos camponeses, nos estados alemães, era muito precária. Segundo a referida autora, a sociedade de meados do século XIX ainda era muito agrária e o setor industrial crescia, mas muitos camponeses não estavam preparados para tais mudanças; era um tempo de crise. Conforme Alves (2003), os camponeses perderam suas terras em muitas reformas agrárias. Além disso, muitos ofícios desapareceram, devido à industrialização. Tais condições desfavoráveis os impeliam a emigrar, pois acreditavam que “no Brasil garantiriam pelo menos a alimentação diária” (ALVES, 2003, p. 170).

A mudança de regime de trabalho no Brasil foi um processo durante o qual coexistiram, por um tempo, um resquício de mão de obra escrava e a implantação de mão de obra dos imigrantes europeus. Era um momento de profundas modificações no Brasil e também no modo de vida desses imigrantes, que estabeleceram aqui suas raízes e contribuíram para a construção da economia e, também, da cultura brasileira. No que se refere à língua e à cultura, houve a presença de variantes em cada região do nosso país onde os pomeranos e outros imigrantes habitaram.

A respeito dos motivos que levaram à imigração dos prussianos (dentre eles os pomeranos) bem como a respeito das características do processo histórico da imigração, Sturz (1862; 1868) confirma que a crise agrária no Reino da Prússia, a realização das propagandas atrativas sobre o Brasil e a numerosa população na Europa constituíram fatores que propiciaram a vinda dos imigrantes germânicos para o Brasil.

Granzow (2009) esclarece o contexto da época e explica em linhas gerais o processo de imigração:

> Propagandistas do Brasil chegaram à Alemanha e fizeram acordos com as grandes empresas marítimas para o transporte de imigrantes e organizaram associações de imigração e colonização. O imperador do Brasil, da casa dos Habsburgos, deu preferência aos alemães, por sempre terem se mostrado bons trabalhadores e colonizadores. Quando D. Pedro I se casou com a princesa Leopoldina, foi influenciado por ela a procurar colonizar as terras brasileiras com imigrantes da Alemanha e da Áustria. Depois das guerras napoleônicas, a Alemanha havia se tornado um país pobre e por esta razão as propagandas sobre o Brasil tiveram grande repercussão. A primeira colonização alemã ocorreu na Bahia em 1818 e, em homenagem à princesa, recebeu o nome de Leopoldina. Porém, ganhou força apenas em 1824. O navio holandês "*Argo*" trouxe agricultores e soldados que desembarcaram no Rio de Janeiro no dia 13 de janeiro de 1824. [...] para o Rio Grande do Sul [...] chegou lá no dia 25 de julho de 1824. [...] Os primeiros imigrantes alemães vieram da região sul da Alemanha: Baden, Württemberg e Baviera, inclusive do Böhmerwald. Mais tarde vieram imigrantes das regiões: Hesse, Rio Reno e Hunsrück. Na Prússia, de acordo com as leis vigentes na época, todos os cidadãos precisavam de uma autorização oficial para emigrar para outro país. Os agentes responsáveis pela autorização de emigração na Prússia até ficaram, temporariamente, proibidos de conceder este documento para pessoas que pretendiam emigrar para o Brasil. [...] Por esta razão, a imigração de famílias pomeranas só se intensificou a partir de 1859, com maior ênfase nas décadas de 1860 e 1870. [...] O que mais desejavam era poder possuir um pedaço de terra própria. [...] Porém, a maioria mesmo, era de famílias que trabalhavam como diaristas nas grandes fazendas da Pomerânia Oriental. Um grande número de imigrantes pomeranos saiu das seguintes regiões: Köslin, Kolberg, Belgard, Regenwalde, Greifenberg, Schievelbein e Neustettin. Em número menor, das regiões de Stolp, Schlawe, Naugard e Bublitz. [...] No meio dos grupos de imigrantes que vieram da Alemanha a partir de 1859, os pomeranos representavam a maioria. [...] prussianos do leste e oeste, pomeranos, bem como imigrantes de Mecklemburgo, Schleswig-Holstein, Oldenburgo, Holanda, Renânia, Baden, Baviera, Boêmia e Suíça. Não mencionei o total das províncias prussianas, mas todas estão representadas aqui. No entanto os pomeranos são a maioria (GRANZOW, 2009, p.166, 167 e 171).

Assim, entre os germânicos, um grande de número de pomeranos veio para o Brasil. Aqui eles estabeleceram raízes inicialmente no Espírito Santo, Santa Catarina e Rio Grande do Sul, mas hoje existem descendentes de pomeranos espalhados por todo o território brasileiro, como no Paraná, em Rondônia e em Minas Gerais.

2.3 O pomerano no Brasil e sua definição: língua ou dialeto?

Acreditamos haver uma lacuna no que se refere a uma melhor identificação do pomerano. No Brasil, ainda não foram desenvolvidos estudos suficientes que comprovem a afirmação de Tressmann de que o pomerano é uma língua (2008, p. 10, 12), separada e diferente do grupo das línguas alemãs. Não localizamos análises do pomerano nos níveis sintático, semântico, lexicogramatical, etc. que apresentem estudos científicos e comprovações empíricas

e documentais que validem tal posicionamento. Existe uma intrigante unanimidade no meio das produções acadêmicas e pesquisas independentes no Brasil quando se refere a retomar o histórico linguístico do pomerano e sua definição. Ocorre que, são sempre retomadas as afirmações de Tressmann, sem qualquer senso crítico ou revisão bibliográfica, muito menos é feita uma checagem de outras produções teóricas dentro e fora do Brasil e nem mesmo se busca reconstruir os fatos com base em fontes primárias.

A existência de mais estudos colaboraria para elucidar essa problemática. Alguns elementos, como o nível de inteligibilidade, semelhanças nos sistemas lexicais e fonológicos, análise sintática, etc., poderiam contribuir para o esclarecimento dessa questão.

Exemplo dessa problemática é a afirmação de que o pomerano era ágrafo até a publicação do dicionário de pomerano-português no Brasil. Para Tressmann (2010), a "língua pomerana", no Brasil, "só passou a ter forma escrita no ano 2000, quando a pesquisa ficou mais forte no estado capixaba e houve um reconhecimento oficial, que permitiu mais incentivos nos âmbitos municipal, estadual e federal". Tressmann produziu um dicionário pomerano-português, publicado em 2006.

A contribuição de Tressmann para o pomerano é inestimável, pois sua proposta de escrita dicionarizada registra o léxico pomerano por meio da documentação de cerca de 16 mil verbetes, além de reunir informações enciclopédicas. O autor também possui produções na área da cultura pomerana (2005), aborda a questão do bilinguismo pomerano-português (1998) e também discute políticas linguísticas (2009).

Porém, a afirmação de que o pomerano era uma língua ágrafa não se sustenta, devido à existência de cartas antigas em pomerano pertencentes aos descendentes (fonte primária). Ademais, consultamos a Biblioteca *online* de Greifswald[12] e encontramos literatura pomerana em escrita pomerana, por exemplo Dräger (1878), Bernhard (1878), Raeck (1969), etc. Também não podemos desconsiderar que desde 1588 já existia uma bíblia (*Barther Bibel*) em escrita pomerana.

O posicionamento de Tressmann quanto à distância entre pomerano e alemão fica evidente no excerto a seguir:

---

[12] Greifswald é uma cidade pomerana, localizada na costa do Mar Báltico, entre as ilhas de Rügen e Usedom, no estado de Mecklemburgo-Pomerânia Ocidental. Em Greifswald está uma das mais antigas Universidades alemãs, a Ernst-Moritz-Arndt-Universität, fundada em 1456. É possível consultar o acervo da biblioteca da Universidade de Greifswald. Disponível em: <http://www.uni-greifswald.de/bibliothek.html>. Acesso em: 10/05/2016.

> O termo "Baixo-Alemão" (em Alemão *Niederdeutsch*, *Plattdeutsch*), embora muito popular na Alemanha, inclusive entre os círculos de estudos, sugere que as línguas nativas das planícies da Europa Central proviriam do Alemão, ou seriam suas variedades dialetais. Neste sentido, postulamos que o termo Baixo-Saxão (Inglês: *Low Saxon*) é o mais acertado para identificar a subfamília linguística à qual pertence o Pomerano e as demais línguas das terras baixas da Europa Central. [...] O quadro aponta, que o Pomerano, o Neerlandês, o Escocês e o Inglês, por exemplo, descendem de uma língua em comum: o Saxão antigo, que deu origem a várias subfamílias linguísticas. O Pomerano, especificamente, é uma língua baixo-saxônica, isto é, uma língua saxônica das terras baixas da região do Mar Báltico, Europa. Já o Inglês é uma língua anglo-saxônica, derivada do Saxão antigo e do Anglo, por isso aparentado também com o Pomerano. O Alemão, no entanto, pertence a um outro grupo de línguas; descende do Alto-Alemão antigo (das regiões altas, montanhosas da Alemanha e da Suíça), que se originou do Gótico (TRESSMANN, 2008, p. 10-14).

Para que se afirme que o pomerano é uma língua (TRESSMANN, 2008 p.10, 12) e não uma variação dialetal dentro do conjunto das línguas alemãs, é necessário, primeiramente, que se desenvolvam estudos do léxico, da sintaxe, da fonética, da morfologia, da estrutura da fala pomerana e que também se desenvolva um detalhado estudo comparativo com as mesmas esferas de outras línguas, não só do alemão-padrão. Enfim, o pomerano precisa ser estudado não só em nível lexical, mas em todas as esferas que o compõem, todas igualmente importantes.

É necessário também que haja uma explicitação da metodologia utilizada e que seja baseada em critérios científicos, em dados autênticos e evidências empíricas. Também é importante que não sejam desconsiderados séculos de produção linguística e historiográfica no lugar onde o pomerano surgiu, na Europa, onde existe uma longa tradição em estudos de Filologia e Dialetologia.

No Brasil, há muitos estudos em andamento, mas ainda não avançamos a ponto de confirmar a afirmação de que o pomerano é uma língua autônoma, o que seria, no máximo, uma hipótese. Na Europa, as fontes alemãs indicam que o pomerano é um *Mundart* ou *Dialekt* (HERRMANN-WINTER, 1998; VOLLMER, 2008).

Com base nas palavras de Tressmann (2008; 2010), podemos avaliar um breve histórico da origem da língua pomerana que o autor faz. Segundo ele, o pomerano é uma língua que pertence ao tronco indo-europeu e da família das línguas germânicas, formado pelo baixo-saxônico e pelo westfaliano. Dentre algumas línguas que para Tressmann influenciaram o pomerano, poderíamos citar: o *Plattdeutsch* (segundo o autor formado pelo saxão antigo e baixo-alemão médio), *Ostniederdeutsch* (baixo-alemão oriental, com influências frísias), *Baltendeutsch* (alemão báltico), *Niederpreußisch-Hochpreußisch* (baixo e alto prussiano), *Schlesisch* (silésio) e *Westpreußisch* (prussiano ocidental).

A nosso ver, seria preciso conhecer bem todas essas variedades linguísticas acima para comprovar que elas formaram o pomerano, podendo assim comparar diversos aspectos entre elas e o pomerano falado hoje no Brasil. Além disso, precisaríamos recorrer a estudos históricos e linguísticos que demonstrassem a trajetória acima sugerida para afirmar que essa confluência de variedades formou o pomerano. Contudo, é interessante notar que todas essas variedades linguísticas e formas dialetais acima mencionadas por Tressmann são línguas germânicas e alemãs, o que só confirma a aproximação do pomerano em sua origem com os dialetos[13] provenientes do grande grupo do baixo-alemão.

Para Tressmann (2010), o pomerano desenvolveu-se antes da constituição do Reino da Prússia (1701-1918); o autor também diz que o alemão entrou na Pomerânia bem mais tarde em relação à constituição do pomerano:

> A partir do ano 1400, o *Pomerisch* ou Pomerano, formado a partir de línguas pertencentes à subfamília Baixo-Saxão, se solidifica e passa a ser língua corrente na Pomerânia. Esta é a língua que mais tarde foi levada para o Brasil, especialmente o Pomerano oriental. A entrada do Alemão na Pomerânia se dá mais tarde, a partir de 1530, com a Reforma luterana (TRESSMANN, 2008, P.12).

A afirmação acima sugere então que na Pomerânia, antes de 1530, o alemão (o alto-alemão) era desconhecido. Para Kappler (1999, p. 90) já se utilizava o termo *deutsch* (alemão), de modo genérico, desde o século VIII d.C.

Tressmann (2010) afirma que o pomerano obteve status de língua autônoma desde o século XIII e foi uma "língua franca" até o século XVI, devido ao seu uso pelos comerciantes que circulavam nas regiões da Liga Hanseática. Entretanto, com a crise e queda do comércio na região do Mar Báltico, deixou de ser tão falado, visto que o território passou a ser do domínio

[13] Do ponto de vista do funcionamento linguístico não há algo que deprecie o dialeto ou o faça inferior com um sema negativo, o que diferencia uma língua de um dialeto, em poucas palavras é o status histórico. Para afirmar isso, nos fundamentamos em Coseriu (1982), ele explica que um modo comum e tradicional de falar é um sistema de isoglosas realizável no próprio falar e um sistema de isoglosas completo, ou seja, realizável como atividade linguística, é uma língua. E o conceito geral de língua, por sua vez é o de sistema de isoglosas comprovadas em uma atividade linguística completa; o falar e o entender de vários indivíduos de acordo com uma tradição historicamente comum. O autor explica ainda que todo sistema que pode funcionar no falar e funciona como atividade linguística é uma língua. Ele afirma que entre dialeto e língua não há diferença de substância ou natureza, ou seja, um dialeto é uma língua, pois possui um sistema fônico, gramatical e léxico. Deste modo, consideramos o pomerano um dialeto, por ser tributário de uma língua histórica, o "alemão", mais especificamente, o baixo-alemão. Conforme Coseriu: "existe entre língua e dialeto, diferença de estatus histórico (real ou atribuído) [...] o termo dialeto [...] uma língua incluída dentro de um uma língua maior que é justamente, uma língua histórica (um idioma)". (COSERIU, 1982, p. 11-12).

do Reino da Prússia. Em 2008, o autor já afirmava sobre o pomerano ter sido língua autônoma e língua franca:

> [...] O quadro comparativo sugere, entre outras coisas, que o Pomerano é uma língua autônoma. [...] Até o início do século 16, o Baixo-Saxão era língua franca em toda a costa do mar Báltico (Liga Hanseática), tanto na forma oral como escrita. Na Europa medieval, o Baixo-Saxão era língua franca em toda a costa do mar Báltico e do mar do Norte, entre os anos 1200 e 1600, e também possuía escrita e todos os direitos de uma língua verdadeira. Na época, o Pomerano era também conhecido por Baixo-Saxão (TRESSMANN, 2008, p.14,16).

Walter (1997) refere-se a uma forma de baixo-alemão que foi língua franca por volta do século XIII:

> No norte, uma língua supra-regional mista havia se formado desde o século XII ou XIII. Uma espécie de baixo-alemão unificado nasceu então das necessidades da Liga Hanseática, [...] Essa língua perdeu seu caráter supra-regional na Alemanha no fim do século XV, no momento em que se iniciou o declínio econômico e político da Hansa (WALTER, 1997, p. 274).

No excerto mencionado acima, Walter (1997) não afirma diretamente que tenha se tratado da variedade oriental do pomerano nem que fosse língua autônoma, porém apresenta o surgimento dessa forma linguística como uma iniciativa de unificação linguística dentro do grupo do baixo-alemão. Reconhecemos, contudo, a possibilidade do pomerano, ou um dos dialetos pomeranos, ter sido língua franca, devido ao intenso comércio no mar Báltico e pela comprovação da existência da Liga Hanseática.

Tressmann (2008; 2010) justifica que, para ele, a "língua" se formara em regiões diferentes da Alemanha e também de forma distinta em relação ao alemão. O autor afirma que o prussiano e o alemão foram impostos aos pomeranos por meio das escolas e das igrejas.

Concordamos que, geograficamente, o território da época em que os imigrantes vieram (considerando o marco de 1859) e o território atual da Alemanha são bem diferentes. Porém, quando retomamos a história da localização geográfica do pomerano, percebemos que havia diversos dialetos e línguas em contato na Europa; não havia, por exemplo, a unificação sob a forma de Estado Nacional como conhecemos hoje. Aquela que chamamos hoje de Alemanha, compreendia, na época, um território maior e era chamado de Reino da Prússia.

Na concepção de Tressmann (2008), o alemão e o pomerano não são inteligíveis entre si. Nós acreditamos que, se compararmos um pomerano do interior que nunca teve contato com o alemão-padrão da Alemanha atual, de fato, pode haver um estranhamento inicial, mas ainda

assim há um conjunto de itens lexicais compartilhados, conforme comprova o dicionário do próprio autor. Observamos que, mesmo com uma grafia que se distancia da grafia do alemão, aproxima-se do dinamarquês com caracteres escandinavos e possui terminações suecas, existem itens no referido dicionário pomerano (verbos, substantivos, pronomes, etc.) próximos ao alemão-padrão. Citamos como exemplo os seguintes casos que se apresentam respectivamente em alemão-padrão e em pomerano: *alles* (BEAU, 1983, p. 669) – *ales* (TRESSMANN, 2006, p.13); *Blitz* (BEAU, 1983, p. 740) – *blits* (TRESSMANN, 2006, p.57); *bringen* (BEAU, 1983, p. 746) – *bringa* (TRESSMANN, 2006, p.66); *lachen* (BEAU, 1983, p. 924) – *lacha* (TRESSMANN, 2006, p. 284); *Lampe* (BEAU, 1983, p. 925) – *lamp* (TRESSMANN, 2006, p. 285); *schwarz* (BEAU, 1983, p. 1045) – *sward* (TRESSMANN, 2006, p. 471), etc.

Sobre a inteligibilidade e não-inteligibilidade entre pomerano do Brasil e alemão-padrão, Granzow cita casos de compreensão entre essas variedades, como em "todos dali, adultos e crianças, entenderam as minhas histórias em alemão e pude fazer a leitura e contá-las sem problemas" (2009, p. 65) ; "durante a aula fiz uma palestra e li algumas histórias do meu livro, em alemão, que foram entendidas sem problemas" (GRANZOW, 2009, p. 112-113) e "inclusive ali em Rio Pequeno podia-se falar tanto em alemão quanto em pomerano, pois todos compreendiam os dois idiomas" (GRANZOW, 2009, p. 131).

Mas Granzow também cita casos de não-compreensão: "com eles faço novamente o teste e leio uma história bem divertida em alemão. Da maioria não surge qualquer reação e apenas alguns dirigentes da igreja sorriem timidamente. Isto comprova mais uma vez que as pessoas dali não entendiam mais o alemão e sim apenas o pomerano" (GRANZOW, 2009, p. 62).

Por isso advertimos que, em um evento comunicativo, existem diversas variáveis que influenciam o processo compreensivo, por exemplo, capacidade individual do ouvinte, dicção do falante, audição do ouvinte, acústica local, influências de outros contatos linguísticos, cansaço, nível de atenção, etc. A inteligibilidade entre pomerano e alemão-padrão é um tema controverso e necessita de mais estudos para melhor esclarecimento.

Tressmann (2008) localiza o pomerano como possuindo uma origem diferente da comumente atribuída ao baixo-alemão; para ele trata-se de uma língua pertencente ao grupo das línguas baixo-saxônicas, denominadas, segundo ele, imprecisamente, de baixo-alemãs. O autor prossegue afirmando que:

> O Pomerano é uma língua baixo-saxônica, isto é, uma língua saxônica das terras baixas da região do Mar Báltico. Também integram o grupo das línguas

> baixo-saxônicas o Vestfaliano, o Platt Menonita, o Saxônio, o Neerlandês, entre outras. O Inglês e o Escocês são, por sua vez, línguas anglo-saxônicas, também aparentadas com o Pomerano (TRESSMANN, 2008, p.10).

Os estudos de Walter (1997) demonstram que o inglês, embora também pertença ao grupo do germânico, tomou um caminho diferente, em seu percurso histórico, geográfico e linguístico. Embora reconheçamos que o pomerano possui algumas pronúncias próximas ao inglês, para nós, isso não o invalida como forma proveniente do baixo-alemão. A seguir, nos posicionaremos sobre a classificação do pomerano.

2.3.1 Nosso posicionamento

Concordamos que o pomerano pertence ao tronco indo-europeu e à família das línguas germânicas, porém, como variedade dialetal, sem relação de preconceito linguístico, ele faz parte do grupo do baixo-alemão, mesmo que tenha sido influenciado pelo baixo-saxão, uma vez que este também é uma variedade baixo-alemã. Lembramos, inclusive, que *Sachsen*, a Saxônia, é um território que ainda hoje faz parte do território alemão.

O pomerano é uma variedade linguística germânica, descendente do Germânico do oeste. A variedade mais próxima do pomerano que ainda é falada na Alemanha é o pomerano anterior mecklemburguense (*Mecklenburgisch-Vorpommersch*), embora não seja exatamente a mesma variedade dialetal, possivelmente é a forma viva mais próxima hoje do pomerano ocidental na Alemanha, catalogada pelo dicionário do *Mecklenburgisch-vorpommerschen* (HERRMANN-WINTER, 2003).

Existe também um dicionário pomerano mais amplo publicado na Alemanha, chamado de *Das Pommersche Wörterbuch*, o dicionário pomerano (HERRMANN-WINTER; VOLLMER, 1999).

Em seus estudos lexicográficos e geolinguísticos, ao produzir os dicionários acima referidos e fazer um atlas do pomerano na Europa bem como um levantamento lexicográfico do pomerano na Europa, Vollmer (1999, 2008) e Herrmann-Winter (1998, 1999, 2013) referem-se ao pomerano como *Pommersches Plattdeutsch* (baixo-alemão pomerano), portanto, dentro do grupo do alemão.

Esclarecemos, com base em Hermann-Winter (1998), que havia dialetos pomeranos, no plural, sendo que as denominações *ostpommersche niederdeutsche Mundarten* (dialetos baixo-alemães pomeranos orientais) e *niederdeutschen vorpommersche Mundarten* (dialetos baixo-alemães pomeranos anteriores), também colocam as formas do pomerano enquanto variedades dialetais.

Para nós, o pomerano está dentro do grupo do baixo-alemão, uma das variedades das terras planas do norte da Europa. Inclusive, concordamos que podemos nos referir ao pomerano como *Pommersches Plattdeutsch*, ou seja, o baixo-alemão pomerano, dentro do grupo maior do baixo-alemão do leste e do subgrupo linguístico do Germânico do Oeste:

> Os povos do mar do Norte: frisões, anglos e uma parte dos saxões são representados hoje pelas populações que falam o frisão e o inglês, e que devem ser distintas das que a partir do Norte da Alemanha, do Elba e do Reno, espalharam-se, para o Oeste. [...] Só muito mais tarde é que se diferenciaram o baixo-alemão, fonte do neerlandês e do *Plattdeutsch*, e o alto-alemão, de onde deriva o alemão de hoje. Os linguistas designam todas essas línguas com o nome de germânico do Oeste [...]. No primeiro milênio de nossa era ocorreu a diferenciação das línguas germânicas em diversas variedades, que se dividiram em três grandes grupos, usando como critério as direções aproximadas da saída de suas migrações: germânico do leste, cuja descendência não sobreviveu; germânico do norte, que reúne as línguas escandinavas; germânico do oeste, o mais importante, subdividido em dois subgrupos: alemão, nas duas formas do baixo-alemão e do alto-alemão (a Teutônia); e o anglo-frisão, que deu origem ao inglês (nas ilhas Britânicas) e ao frisão (hoje essencialmente nos Países Baixos) (WALTER, 1997, p.246, 254).

Acreditamos que as transformações linguísticas que os pomeranos e seus descendentes viveram ao longo do tempo, em solo brasileiro, permitiram um ambiente favorável ao surgimento de uma variedade com características peculiares de pronúncias, usos e sentidos, pois são características inerentes às línguas a variação e a mudança (LABOV, 1972).

Para nós, o pomerano pode ser referido como língua em oposição ao português, quando é citado no discurso em constraste com a língua portuguesa, pois, nesse caso, são sistemas linguísticos e origens históricas diferentes. Também é aceitável ser referido como língua, por uma questão identitária dos pomeranos no Brasil, visto que, do ponto de vista do funcionamento linguístico, não há diferenças que releguem o pomerano a um *status* inferior. Portanto, não vemos problema em referí-lo como língua, entretanto, discordamos de que seja língua autônoma e separada do grupo alemão. Em relação ao alemão, o pomerano se configura como um *Dialekt* (dialeto) ou *Mundart* (tipo de fala), pois não podemos desconsiderar todo o percurso que o relaciona à língua alemã, como língua histórica, nem mesmo desvinculá-lo por completo da tradição histórica-linguística-cultural do qual ele é tributário, com base em toda a discussão acima apresentada.

Para Tressmann, ao contrário, o pomerano não pode ser denominado genericamente de língua alemã ou dialeto alemão:

> O vocábulo a que nos referimos como Baixo-Saxão ou Saxão das Terras Baixas é largamente conhecido como "Baixo-Alemão" (*Niederdeutsch*, *Plattdeutsch*; *Plattdüütsch*), especialmente com referência às suas variedades faladas na Alemanha. No entanto, como atesta Hahn (2002), apesar do fato de o termo gozar de popularidade na Alemanha, inclusive assim é muitas vezes referido na literatura linguística, a expressão "Baixo-Alemão" não pode ser considerada a escolha mais acertada para nomear esta subfamília. Em primeiro lugar, Baixo-Saxão e Alemão descendem de línguas diferentes, respectivamente, Saxão antigo e Alto-Alemão antigo. Além disso, antes de o Baixo-Saxão começar a perder prestígio em relação ao Alemão, muitos de seus falantes ainda se referiam à língua como "Saxão". [...] O estudo histórico-comparativo entre línguas germânicas nos possibilitou identificar o Pomerano como uma língua pertencente à família linguística Germânica (ocidental) e à subfamília Baixo-Saxão (oriental), que por sua vez originou-se do Saxão antigo. Preferimos o termo Baixo-Saxão a "Baixo- Alemão", especialmente porque o Pomerano e o Alemão descendem de línguas diferentes, respectivamente, Saxão antigo e Alto-Alemão antigo. [...] O Pomerano, o Vestfaliano e o Platt Menonita, por exemplo, são incompreensíveis entre si, não podendo ser considerados, portanto, uma língua única, ou seja, denominados genericamente de "Baixo- Saxão", nem tampouco de "Alemão", ou de "dialetos alemães" ou "línguas alemãs" (TRESSMANN, 2008, p. 16, 20).

Porém, outras referências sempre localizam o pomerano como um dialeto proveniente do grupo do baixo-alemão, referindo-se ao pomerano como dialeto alemão. Citamos algumas dessas referências a seguir.

Bost (2010), ao escrever uma matéria para um jornal alemão, refere-se ao pomerano como um dialeto baixo-alemão, que no norte da Alemanha está quase morto, e na Província Pomerânia, hoje Polônia, já está extinto. Segundo o autor, esse dialeto está sendo atualmente reavivado fortemente no Espírito Santo[14].

Höhmann e Savedra (2011), ao escreverem um artigo a respeito dos pomeranos no Espírito Santo, também se referem ao pomerano como uma variedade linguística baixo-alemã-oriental, um dialeto alóctone desenvolvido qual um dialeto alemão que prevaleceu na respectiva comunidade linguística[15]. E outra menção é a de Glüsing (2013), ao escrever para a revista alemã *Der Spiegel*, sobre os pomeranos no Brasil, citando em especial Santa Maria de Jetibá/ES, refere-se ao pomerano como baixo-alemão pomerano e afirma que ele é um dialeto baixo-alemão do leste[16].

---

[14] No original: *Der pommersche niederdeutsche Dialekt, der in Norddeutschland fast ausgestorben ist, und in der Provinz Pommern im heutigen Polen ganz erloschen ist, wird in Espírito Santo (Brasilien) zurzeit mit großem Engagement wiederbelebt* (BOST, 2010).

[15] No original: *Das Pommerische, eine ostniederdeutsche Sprachvarietät, [...] eines allochthonen Dialekts entwickelt [...] welcher deutsche Dialekt in der jeweiligen Sprachgemeinschaft vorherrschte*. (HÖHMANN; SAVEDRA, 2011, p.284-285).

[16] No original: *Pommersch Platt ist ein ostniederdeutscher Dialekt* (GLÜSING, 2013).

Herrmann-Winter (1998), ao citar os dialetos pomeranos do oeste e do leste refere-se a eles como dialetos ou *Mundarten* do grupo do *Niederdeutsch*. Também nos referimos ao pomerano comumente como uma variedade linguística para não deixar impressão de sentido pejorativo. A expressão "dialeto" pode ou não ser utilizada com relação de preconceito, mas isso depende do contexto de uso. Em nosso caso, não há qualquer visão de inferioridade. Conforme, Walter o baixo-alemão tem tanto valor quanto o alto-alemão:

> Não se deve pensar que essas denominações tradicionais correspondam a julgamentos de valor. Designam apenas as variedades de germânico que se desenvolveram nas planícies do norte da Europa (baixo-alemão) e nas regiões montanhosas da Europa central (alto-alemão): a língua do "país plano" de um lado e a das montanhas de outro (WALTER, 1997, p. 275).

A respeito da forma de se referir ao pomerano, no contexto brasileiro, baseados nas transformações linguísticas e históricas ao longo do tempo, no distanciamento do lugar de origem, no contato com outras variedades germânicas e com o português no Brasil, e, transformado por esses contatos linguísticos e culturais, optamos por nomeá-lo como *Brasilianisch-Pommersch* (BEILKE, 2014), ou seja, um pomerano-brasileiro, uma variedade brasileira do pomerano, que "mesclou" as variedades oriental e ocidental do pomerano, pois, em solo brasileiro, os pomeranos de ambas as regiões da Pomerânia ficaram juntos e não separados em áreas ocidental e oriental como acontecia nos territórios de suas respectivas origens, antes de imigrarem para o Brasil.

Embora a maioria dos pomeranos fossem provenientes da Pomerânia oriental, a nomenclatura que propomos visa considerar os contatos de línguas, com variedades alemãs e com o português. Exemplo do contato interétnico e da influência dos contatos de línguas no pomerano, encontramos em Granzow (2009):

> O pomerano dele era um pouco diferente do meu. Ele contou que já viajou muito entre as colônias pomeranas e constatou que muitos pronunciavam o pomerano de uma forma diferente, principalmente na entonação do "ü". Por exemplo, a mesma palavra poderá soar diferente "*Hühner*" ou "*Heuhner*", ambos significam *galinha.* Ou "*Höhner*", "*Heiner*" ou "*Hauhner*" que significa *galo*. Para ponte alguns dizem: "*Brigg*, *Briech*, *Brög* ou *Brüch*". [...] Estas diferenças de algumas palavras se dão por influência dos "*hunsrücker*". Além disto, também moravam algumas famílias holandesas na região, o que misturava o pomerano com o holandês [...] O pomerano daqui já não é tão original e é misturado com outros dialetos, o que é natural, pois nas redondezas moravam descendentes de várias etnias da Alemanha (GRANZOW, 2009, p. 67-68, 81, 129).

Esse é um indício de uma de nossas hipóteses, pois o relato do autor revela que o contato entre descendentes de germânicos de regiões diferentes acontecia a ponto de influenciar nos seus falares. Porém, falta a comprovação empírica por meio de dados extraídos de *corpora*, conforme propomos nesta pesquisa.

Na próxima seção, discutimos o termo "alemão", como forma de nos referirmos às línguas germânicas.

2.3.2 "Alemão", termo genérico para representar línguas germânicas do grupo alemão

O termo alemão, muitas vezes, é usado de uma forma genérica para se referir a toda uma variedade de línguas que descendem do grupo das línguas germânicas, sendo como se o sufixo *deutsch* as agregassem sob um domínio comum. Exemplo disso é o termo *Brasildeutsch* utilizado por Von Borstel (2011).

Bossmann (1953, p. 96) há muito tempo já fazia referência à diversidade de variedades teutônicas no Brasil como "*deutsch-brasilianische Mischsprache*" ou mescla linguística teuto-brasileira. Essa nomenclatura congrega variedades germânicas faladas pelos descendentes de diversas regiões da Prússia, bem como suas influências pelo contato com outras línguas no Brasil. Esse autor mencionava o *deutsch* ou alemão de um modo mais genérico, referindo-se aos elementos comuns que agrupavam as variedades, sem, contudo, desconsiderar as características próprias e sem estabelecer qualquer relação de preconceito linguístico em relação a alguma variedade específica. É como se houvesse uma referência, senso comum, a um "alemão genérico".

Esclarecemos que concordamos com Tressmann no que tange a não afirmar diretamente que o pomerano é o alemão (*Hochdeutsch*, doravante HD); de fato, reconhecemos que não são iguais e comparar o pomerano com o alemão-padrão de hoje, ensinado normativamente nas escolas, é descontextualizar ambas as variedades. No entanto, consideramos que o pomerano é tributário histórica e linguisticamente ao grupo das línguas alemães, como *Plattdüütsch* (baixo-alemão) pomerano.

Quando o alemão é oposto radicalmente ao pomerano, percebemos que há uma confusão do que se entende por "alemão", pois não é preciso ver o "alemão genérico" (que pode estabelecer diferentes relações de sentidos) como algo igual ao alemão-padrão. No Brasil, o termo "alemão" foi muitas vezes utilizado (por Bossmann, pelo senso comum, pelos próprios falantes, etc) como forma de se reportar às diferentes variedades linguísticas de minorias étnicas alemãs e não necessariamente como referência ao alemão-padrão. O termo "alemão", de forma

genérica, foi muito utilizado na historiografia e também na oralidade no Brasil. Ao vivenciar o contato com falantes de pomerano presenciamos esse fato em inúmeras ocorrências.

Entre o alto-alemão e o alemão-padrão existe, segundo Walter (1997), uma pequena diferença; foi a partir do primeiro que houve um processo que culminou na convenção do segundo, que hoje chamamos de alemão-padrão.

O alemão-padrão surgiu do processo de prestígio no uso do alto-alemão e de uma evolução do saxão, muito utilizado por figuras importantes, como os reis da Saxônia, também adotado por Lutero; era uma forma muito utilizada e foi sendo convencionado devido a necessidade de padronização, por haver diversos falares alemães. O alto-alemão ganhou prestígio, mas continuaram existindo os dialetos. Com base nos estudos de Walter (1997), podemos retomar um breve histórico, para entender como do alto-alemão originou-se o alemão-padrão:

> [...] uma língua comum acabaria finalmente catalisando a diversidade dialetal da Idade Média na Alemanha. [...] Devido à situação política da Alemanha, [...] o aparecimento de uma língua comum naquele país foi muito mais tardio, já que cada região cultivava com estima e devoção exclusivas sua própria variedade dialetal até o século XVI, [...], houve duas tentativas de uniformização, e a primeira tentativa ocorreu no norte [...] No sul, foi apenas no final do século XIV que começou a se formar uma língua comum na região central e oriental, zona de cruzamento de povos que falavam diversos dialetos alto-alemães. Surgiu da necessidade de se dispor de uma língua que facilitasse a comunicação entre as várias chancelarias imperiais, e parece que a de Carlos IV (1347-1378), cuja sede ficava em Praga, teve um papel muito importante em sua elaboração devido à sua posição central [...] Essa língua de chancelaria, no início essencialmente escrita, revelava ao mesmo tempo uma tendência à uniformização das formas e uma procura de harmonização na grafia. Foi essa língua administrativa da Alemanha Central que foi o ponto de partida do alemão comum. Mas antes, foi preciso renunciar ao latim. No fim do século XIV e paralelamente ao latim, que ainda era favorito, foi sendo elaborada pouco a pouco uma língua escrita que era uma combinação de vários usos da Alemanha Central em torno de Leipzig, Erfurt e Dresden, cidades onde se misturavam povos de origem variada. Em meados do século seguinte, foi a língua dessa mesma região que Lutero, [...], escolheu naturalmente para sua tradução da Bíblia (1521-1522). Mas essa escolha teve outra justificativa: essa língua mista, que já era a da chancelaria saxã, era também a que, segundo ele, podia ser entendida tanto na Alta Alemanha quanto na Baixa Alemanha. Era efetivamente uma língua "acima dos dialetos", encontrada também em diversas outras chancelarias principescas e municipais no início do século XVI. Logo, ela já estava um pouco uniformizada quando foi difundida nos países germânicos [...] ainda que a língua escrita estivesse unificada, cada região continuava a falar um dialeto diferente. [...] mas essa língua só começou a ser ensinada nas universidades no século XIX (WALTER, 1997, p. 274-278).

Conforme explicamos acima, precisamos compreender que há múltiplas formas de alemão, as "línguas alemãs" que são faladas no Brasil, são heranças dos imigrantes de diversas regiões diferentes do que chamamos hoje de Alemanha, mas na época era um reino formado por diversos ducados, por nações dentro de um território que hoje está espacialmente reduzido e é uma região unificada, sob a forma de um estado nacional, de uma divisão política com fronteiras, hino, bandeira e governo que une uma multiplicidade de culturas germânicas ali congregadas. Era justamente a origem teuta e as falas alemãs (em suas diversas formas dialetais) que agrupavam as pessoas da Prússia, o que permitiu sua denominação genérica enquanto alemães. Conforme Müller (1981) informa:

> A Alemanha não existia como unidade nacional. Havia reinados, principados, ducados, independentes entre si. A língua identificava e unificava todos e aí pode-se falar em Alemanha. Lutero ao traduzir a bíblia criou uma língua oficial que propiciasse a leitura da bíblia a todos, até então havia diversos dialetos. Ao uniformizar a língua, havia um elo comum entre todos os departamentos políticos vindos da Idade Média. Neste sentido, ao falarmos em imigrantes da Alemanha, antes de 1871, ano da unificação formalizada por Bismark, referimo-nos às pessoas de fala alemã. Os passaportes da época registram a origem das pessoas como sendo da Prússia, de Schleswig-Holstein, Renânia, Hesse ou Pomerânia. Como todos falavam a mesma língua, a História apenas registra "alemães" (MÜLLER, 1981, p. 45).

A Alemanha foi tardiamente unificada, em 1871, por Otto von Bismarck. E até mesmo o nome Alemanha não é uma tradução literal de *Deutschland*, que vem de *Thorland* (Wiesner, 1989, p.326), terra do *Thor*. Alemanha vem de *Alamania*, que era uma região específica dentro da Germânia. E Germânia é como a Alemanha é referida até hoje em inglês; *Germany*. Uma tradução próxima ao sentido literal de *Deutschland* é "Terra dos teutos[17]". De modo que não dá para desconsiderar que os pomeranos são teutos, descendentes dos povos germânicos teutônicos e, portanto, alemães.

Houve um processo de transformações históricas e políticas, isso alterou as fronteiras da região, de modo que, a Alemanha atual não é tão ampla como era na época da Prússia. A Pomerânia oriental era uma região que existia dentro do contexto territorial e político da época do reino prussiano. Por isso acreditamos que pensar o pomerano dentro do grupo do alemão, em um sentido mais amplo, não é uma ofensa à identidade pomerana, faz parte das sucessões

[17] Ou terra dos povos. Teutos vem do latim *teutones*, do germânico *theudo*, povo e do indo-europeu *Teuta*, povo. Segundo Cunha (2007p, 768), "Teutos, plural de teuto: do latim *teutõnes*. Teutão adj. sm. diz-se de, ou indivíduo dos teutões". Segundo Bueno (1974, p. 3966), teuto é "o primeiro elemento de compostos com a ideia de teutões, germânicos, alemães. Teutões – s. m. pl. Povo antigo da Germânia".

históricas. Segundo Walter, “os limites linguísticos raramente coincidiam com as fronteiras dos Estados. As situações atuais resultam da história dos povos que se sucederam nesses territórios desde o primeiro milênio a.C”. (WALTER, 1997, p. 243). Não há nesse sentido uma relação de preconceito, trata-se apenas do mesmo grupo de descendentes de povos germânicos.

Vale ressaltar, com base em Zilly (2013) que havia um sentimento de "germanidade", de "ser alemão”. Antes do século XIX, já estava presente a ideia de nação alemã. Zilly (2013) explica que “a Alemanha não existia, como estado, mas existia sim, como ideia de uma nação, não no sentido moderno de nação, mas havia um sentimento de que os alemães formavam uma coletividade principalmente cultural, não política, não jurídica, [...]”. Com base nisso, entendemos que havia uma coletividade de sujeitos de língua alemã que se sentia unida, não no sentido político ou de Estado Nacional, como vivemos no período contemporâneo, mas havia, de modo geral, um sentimento de unidade compartihado socialmente, entre os “alemães”.

Não encontramos nenhuma comprovação de algum fato histórico que indique uma inimizade tão grande entre alemães e pomeranos. Exceto pelas torturas que sofreram no Brasil, por serem confundidos com alemães nazistas, conforme exposto no documentário Bate-Paus (2005), mas essas injustiças foram praticadas por autoridades brasileiras e não por alemães da Alemanha contemporânea. Esclarecido isso, acreditamos que não há justificativas para que a etnia pomerana seja vista como separada dos povos alemães. Granzow fala inclusive na relação entre grupos diferentes de alemães: “Depois chegamos numa região onde os ‘*hunsrücker*’ e os pomeranos convivem pacificamente” (GRANZOW, 2009, p. 80).

Acreditamos que nem todas as minorias étnicas alemãs (germânicas) presentes no Brasil reivindiquem uma separação da origem histórica, étnica, cultural e linguística alemã, o que não fere a identidade desses sujeitos[18], embora reconheçam que sua fala é diferente do alemão moderno, padronizado e ensinado nas escolas. Os suábios, por exemplo, reconhecem que sua identidade cultural está interligada com a história da Alemanha, embora o grupo mais amplo dos suábios do Donau, esteja também relacionado com o que hoje são os países da Áustria, Hungria, Romênia e ex-Iugoslávia. Reconhecem que são, “de maneira geral, povos de etnia e cultura germânicas que têm sua origem principal no sudoeste e no oeste do território que hoje corresponde à Alemanha”, segundo a Fundação Cultural Suábio-Brasileira (2014).

[18]Adotamos o termo sujeito para nos referir aos pomeranos, com base em Santos (1998, p. 183 *apud* SOARES, 2012, p.96) que desde o fim da década de 1990, já utilizava esse termo no contexto da Geolinguística, pois pois acredita que a linguagem é orientada pela visão de mundo dos sujeitos (SANTOS, 2011 *apud* SOARES, 2012, p. 55). Alguns trabalhos no campo da Sociogeolinguística também adotam esse termo, como CRISTIANINI (2012).

Quanto à referência aos pomeranos como povos eslavos em uma origem remota (WILLE, 2011, p. 16), por serem descendentes dos *Wendes*, esse fato remonta ao século VI d.C. e Tressmann em nota reconhece que não foram encontrados vestígios da língua *Wende* no pomerano falado atualmente no Brasil, quando ressalva que:

> Estudos da influência do Eslavo sobre o Pomerano e outras variedades do Baixo-Saxão encontramos em Teuchert (1958). Segundo a pesquisa em andamento, constatamos que a influência do Eslavo sobre o léxico do Pomerano falado no Espírito Santo não mais se manteve, a não ser em nomes de famílias (TRESSMANN, 2008, p. 20, n. 2).

Sobre os pomeranos serem considerados descendentes de povos eslavos, esclarecemos que um estudo consistente baseado em fontes primárias e fundamentações históricas foi produzido por Neményi (2015) no qual o autor reorganiza a questão da origem dos supostos eslavos e acredita que alguns equívocos na historiografia fizeram com que surgisse o que ele chama de "*Der Slawen-Mythos*", o mito eslavo. O autor esclarece que se trata de mera denominação linguística e afirma com muita segurança que os *Wendes* eslavos eram descendentes dos vândalos germânicos, winedos e godos e de outras tribos germânicas do leste e explica que os eslavos são, de fato, os germânicos do leste[19].

Afastar o pomerano do alemão por motivações políticas atuais ou por posicionamentos ideológicos pessoais não muda o fato de que o pomerano ou *Pommersches Plattdeutsch*, como o próprio nome indica, está dentro do grupo do baixo-alemão.

Tressmann afirma que o pomerano é mais próximo do Neerlandês (chamado comumente de Holandês) do que do alemão, quando diz que "podemos constatar certo grau de uniformidade entre os léxicos, por exemplo, do Pomerano e o Neerlandês" (2008, p.15; 2010).

Porém, lembramos que o alemão designa um grupo maior, a ponto de que o *dutch* (modo como se chama o neerlandês em inglês) está dentro do grupo do *düütsch* (modo como os pomeranos se referem ao alemão).

O *Niederländisch* (Neerlandês) está dentro do grupo do *Niederdeutsch,* sinônimo de *Plattdeutsch* na Europa, ou seja, baixo-alemão, conforme comprova Walter:

> A etapa seguinte conduzirá à descrição das línguas resultantes do baixo-alemão: o neerlandês nos Países Baixos, ao qual estão ligados o Plattdeutsch na Alemanha e as variedades de neerlandês faladas na Bélgica e na França

[19] No original: *Die slawischen Wenden entpuppen sich als Nachkommen der germanischen Wandalen, Wineder, Goten und anderer ostgermanischer Stämme [...] erklärt die Slawen als das, was sie wirklich sind: Ostgermanen* (NEMÉNYI, 2015).

> (chamadas flamengo nesses dois países). [...] O Plattdeutsch no Norte da Alemanha e o Neerlandês inclusive o flamengo estão dentro do grupo do baixo-alemão, que faz parte do Teutônico, que descende do Germânico do Oeste dentro das línguas Germânicas, do indo-Europeu [...] E o cúmulo da confusão é quando ele se conscientiza de que Dutch, que em inglês designa a língua dos Países Baixos, é originalmente a mesma palavra que Deutsch - , que, em alemão, designa a língua alemã [...] A palavra inglesa dutch significava também alemão até o fim do século XVI, e ainda pode ter esse valor na América. Foi só quando os Países Baixos tornaram-se um país independente que a palavra dutch, por volta do século XVII e apenas em inglês, teve seu sentido restringido (WALTER, 1997, p.256-256; 294).

Lembramos que o alemão-padrão que os pomeranos seriam supostamente "obrigados" pelo governo prussiano a falar, continuou a ser utilizado nas pregações e eventos eclesiásticos, mesmo depois que eles já estavam vivendo em território brasileiro: "Portanto, para os pomeranos em uma igreja evangélica, em primeiro lugar, o culto deve ser ministrado em alemão" (GRANZOW, 2009, p.125). Embora saibamos que no Brasil não estariam mais sob a tutela oficial de um governo prussiano que os obrigasse a falar somente na forma padrão, havia laços identitários que os unia enquanto pessoas de uma origem semelhante e de uma cultura comum, conforme o excerto de Granzow acima nos informou, a opção pelo culto em alemão e não em português ou outra língua, indica minimamente que havia alguma relação de elo ou de reconhecimento pelo papel do alemão entre os pomeranos.

Também ressalvamos que o alemão-padrão da época era influenciado por traços dialetais, fato comprovado por análise de cartas antigas que contém traços dialetais e não eram escritos totalmente em alemão-padrão. Primeiramente, porque geralmente um gênero tão pessoal como uma carta é escrito na nossa língua materna e na qual mais nos identificamos, principalmente quando se dirigem a um ente familiar. Segundo, devido ao fato de que as regras ortográficas do alemão-padrão só foram publicadas em 1901, depois que muitos pomeranos já estavam vivendo no Brasil. Encontramos em Walter (1997) essa referência:

> A partir de 1830, com a escola obrigatória e o desenvolvimento da imprensa, a língua alemã generalizou-se e, com a fundação do Império Alemão em 1871, sua uniformização foi verdadeiramente regulamentada. A ortografia estava a partir de então fixada de acordo com uma norma tornada pública em 1901 com um dicionário oficial para a ortografia alemã (WALTER, 1997, p. 278).

Acreditamos que havia um léxico mínimo comum compartilhado entre *Hoch* e *Platt* (alemão-padrão e baixo-alemão pomerano). Para defender essa posição, admitimos, com Walter (1997), que no processo de padronização, alguns traços característicos do baixo-alemão foram incorporados ao alemão-padrão:

> A norma do alemão baseia-se em princípio nas formas do alto-alemão, mas faz algumas concessões ao baixo-alemão das regiões do Norte. Os manuais de pronúncia ensinam, por exemplo, que as consoantes /p/, /t/ e /k/ no início da palavra são aspiradas, ou seja, seguidas de um leve [h]. Ora, esse é um traço geral dos falares baixo-alemães do norte, mas muito raro nos falares alto-alemães do sul (WALTER, 1997, p. 279).

Assim, acreditamos que os pomeranos transitavam entre os dois códigos ou variantes em situação de estabilidade, de forma que isso não lhes diminuía a sua identidade pomerana. Comprovamos esse fato, porque muitos descendentes de pomeranos são filhos de casamentos mistos, ou seja, entre alemães de outras regiões e pomeranos: "[...] já encontramos muitos casais que ele era "*hunsrücker*" e ela pomerana. Se não fosse pelos nomes: Seibel, Knippel, etc., muitas vezes ficaria difícil descobrir a procedência" (GRANZOW, 2009, p. 80); e, também, porque quando em contato com a comunidade pomerana identificamos muitos bilíngues em pomerano e alto-alemão.

2.4 Mapeamento e levantamento de pesquisas

Alguns fatos históricos contribuíram para um declínio da quantidade de falantes de pomerano no Brasil. O governo brasileiro, durante o Estado Novo (1937-1945), proibiu a fala de outras línguas no país através de medidas como o Decreto-Lei nº. 383, de 18 de abril de 1938. A política da época priorizava a nacionalização da educação e padronização da língua oficial. Esse foi um momento histórico crucial para as línguas provenientes de processos de imigração, pois as ações da Era Vargas foram determinantes no declínio das línguas alóctones.

No Brasil ainda existem muitas comunidades significativamente numéricas (TRESSMANN, 2008; GRANZOW, 2009; BEILKE, 2013) de falantes de pomerano, em várias regiões. Então, com um enfoque específico no baixo-alemão pomerano, realizamos um levantamento prévio, como tentativa de localizar essas comunidades.

Como esboço de mapeamento do pomerano, apresentamos a seguir um mapa (Figura 4) que elaboramos por meio do programa de geoprocessamento *ArcGis*. No mapa, identificamos as principais localidades do Brasil nas quais existe a presença de descendentes de pomeranos:

Figura 4: Mapa de localidades onde há presença de pomeranos no Brasil.



Fonte: Beilke (2013). ArcGis.

Por meio do mapa (Figura 4) é possível perceber que não são apenas nas regiões do Espírito Santo e Santa Catarina em que há pomeranos presentes no Brasil. Visitamos quase

todas essas localidades pessoalmente e também cruzamos informações de prefeituras, sites, pesquisas, redes sociais, jornais e etc. para esboçar o mapeamento proposto.

As regiões do Vale do Rio Pardo/RS, incluídos os municípios de Santa Cruz, Vera Cruz, Candelária, Sinimbu, Rio Pardinho e no Vale do Rio Doce/MG, incluídos os municípios de Itueta, Resplendor, Mutum e o povoado rural Vila Neitzel/MG são lugares com presença de descendentes da etnia pomerana. Os registros eclesiásticos e os registros de imigração são fontes da formação étnica dessas regiões.

Os relatos de Granzow também confirmam a presença de pomeranos na região do Vale do Rio Pardo/RS:

> Também eu já tinha uma programação para outro local de nome Sinimbu, bem como ainda precisava voltar para fazer a palestra em Rio Pequeno, na casa dos Wartchow. [...] A palestra em Sinimbu Alto transcorreu de forma semelhante à de Rio Pequeno, outra colonização autêntica de pomeranos. [...] Ali também falei alternadamente o alemão e o pomerano e ambos eram compreendidos. [...] Uma segunda colonização de colonos pomeranos se deu no Rio Grande do Sul, na região perto de Santa Cruz (GRANZOW, 2009, p. 131; 175).

Porém, trata-se de regiões que carecem de estudos linguísticos sobre a manutenção da fala pomerana, não há cooficialização ou outras iniciativas que fomentem o pomerano. Por isso, nessas regiões, ainda há muito a fazer para catalogar, registrar e estudar o pomerano.

Segundo estudos já publicados de Beilke (2013), a partir do levantamento realizado em visitas de campo, estudo de fontes primárias, pesquisas bibliográficas etc., existem descendentes de pomeranos nas localidades inseridas na região do Vale do Rio Pardo/RS e na região do Vale do Rio Doce/MG; e isso nos deu respaldo para efetuarmos o trabalho exposto nesta dissertação referente ao mapeamento das regiões com presença de pomeranos e à proposta de verificar a situação da variedade pomerana em Minas Gerais e Rio Grande do Sul.

Quanto ao levantamento específico dos descendentes de pomeranos em Minas Gerais, podemos afirmar que eles estão presentes no leste do estado, no povoado chamado Vila Neitzel e na zona rural norte de Itueta. Conforme Pessoa (1995), alguns pomeranos se deslocaram para Minas Gerais, "especialmente para Santo Antonio, Itueta, Aimorés e Resplendor, região próxima a Baixo Gandu" (PESSOA, 1995, p. 79). Segundo Antunes (2011) estima-se que havia aproximadamente 2.000 pomeranos na região (de Itueta e Vila Neitzel), mas, devido à evasão populacional, no fim da década de 90, apenas 50% (aproximadamente) permanecem nas duas localidades. Dessa forma, um levantamento linguístico específico precisa ser feito nessa região. São famílias que foram primeiro para o Espírito Santo e atravessaram a fronteira mineira, em

busca de terras, além de familiares das primeiras famílias pomeranas que vieram para Minas Gerais fugidos da Segunda Guerra Mundial.

Em Rondônia, no município de Espigão do Oeste[20], a cooficialização do pomerano estava em fase de aprovação até 1995, depois disso não conseguimos obter confirmação da existência da lei aprovada. Fato é que, nessa região do estado de Rondônia, vivem algumas famílias pomeranas que migraram do leste de Minas Gerais e da fronteira com o Espírito Santo em busca de terras. Houve também emigração de pomeranos de Minas Gerais para os Estados Unidos em busca de melhores condições de trabalho. Essa evasão populacional se deu no início da década de 70, como aponta Pessoa (1995, p. 84).

Segundo nota de Granzow (2009, p. 188, n. 2), os pomeranos no Espírito Santo têm uma presença significativa em: Santa Maria de Jetibá, Santa Leopoldina, Domingos Martins, Afonso Cláudio, Laranja da Terra, Baixo Guandu, Itarana, Itaguaçu, São Gabriel da Palha, Vila Pavão, Pancas, Grande Vitória, entre outros. Alguns municípios do estado capixaba obtiveram a cooficialização do pomerano. A Emenda Constitucional nº 11/2009, de 25/04/2011, instituiu o pomerano e o alemão como línguas secundárias no estado do Espírito Santo.

Em Pomerode/SC, houve a cooficialização, mas se referindo ao alemão de forma genérica, sendo que o pomerano também é falado lá. No município de Canguçu/RS, o pomerano já foi oficializado e em Arroio do Padre, no interior do Rio Grande do Sul, até 2013 a cooficialização estava em andamento, depois não conseguimos mais obter fontes sobre o assunto que comprovem a existência da lei aprovada. Contudo, notamos, ao buscar dados, que são poucas as iniciativas de fomento fora do eixo Espírito Santo – Santa Catarina. É inegável a presença de descendentes da etnia pomerana em outras regiões do nosso país.

Há localidades, como São Lourenço do Sul/RS e Canguçu/RS, com relevante presença de pomeranos, em que as pessoas falam o pomerano em casa, entre amigos, nos cultos religiosos, nas escolas e em outras esferas da vida pública e privada. Nas referidas cidades, os pastores Münchow, Wendler, Vorpagel e Blank, por exemplo, já traduziram para o pomerano algumas histórias bíblicas para crianças; o *Bíbliha Aventures* (SBB, 2012), e também traduziram o evangelho de Lucas, pois possuem um projeto de tradução de textos bíblicos do novo testamento para o pomerano[21].

---

[20] No município de Espigão do Oeste/RO, o número de falantes de pomerano está diminuindo, conforme aponta Pessoa (1995), que desenvolveu um estudo sobre os pomeranos de Rondônia.

[21] O grupo de pastores Arnildo Münchow, Hilbert Wendler, Milton Vorpagel e Renato Blank trabalham no projeto de tradução de textos bíblicos para o pomerano desde 2012 e possuem o apoio da SBB e da IELB. Disponível em: <http://www.jornaltradicao.com.br/site/content/variedades/index.php?noticia=11283>. Acesso em 02 de junho 2016.

Quanto ao que se refere aos diversos trabalhos desenvolvidos no Brasil e referências que abordam o pomerano, fizemos um levantamento dos principais trabalhos realizados.

Um dos trabalhos mais mencionados é o de Bahia (2000), que aborda a identidade, a magia e a religião entre camponeses pomeranos do Espírito Santo. Em sua tese de doutorado, a antropóloga trata da religiosidade e das práticas mágicas entre pomeranos, fala da relação com a terra, com a língua e descreve alguns ritos de passagem pomeranos, como o nascimento, o casamento e a morte, descrevendo o imaginário pomerano sob o seu ponto de vista. Assim, a autora versa a respeito da relação entre identidade social e étnica e suas formulações culturais no contexto pomerano.

Benincá (2008) estuda as dificuldades de crianças pomeranas na aquisição de alguns fonemas do Português. Ela estuda crianças bilíngues em pomerano-português e detecta casos de inversões, substituições, problemas de conjugações, trocas de fonemas e também casos de hipossegmentação.

Schaeffer (2012) elaborou a descrição fonética e fonológica do pomerano falado no Espírito Santo. A autora também se refere ao pomerano como língua. Schaeffer descreveu a estrutura silábica do pomerano. Ela concluiu que o pomerano possui 18 fonemas consonantais e 13 fonemas vocálicos, bem como as oclusivas surdas possuem aspiração e as sonoras possuem um índice médio de VOT (*Voice Onset Time* – tempo de ataque de vozeamento) próximo às oclusivas surdas do PTB.

A primeira publicação contendo um esboço de conjugação de verbos do pomerano foi realizada em 2012, por Kuhn Silva, em uma escrita transliterada e como material didático para ensino do pomerano para crianças da Escola Municipal de Ensino Fundamental Germano Hübner, situada no terceiro distrito de Santa Tereza, uma localidade rural de São Lourenço do Sul/RS. Na referida escola é desenvolvido o Projeto Pomerando, que além de ensinar o pomerano e incentivar o uso da variedade, também produz materiais didáticos coletivos, desenvolvidos entre as crianças e seus professores, por exemplo, cartazes e o livro do projeto publicado em 2012.

Mujica (2013) realizou um estudo sobre a identidade linguística dos pomeranos da comunidade de Santa Augusta (São Lourenço do Sul/RS) em uma perspectiva da Sociolinguística laboviana e analisou casos de bilinguismo pomerano-português em diferentes faixas-etárias. No distrito de Santa Augusta também é desenvolvido o Projeto Pomervida, organizado por professoras da Escola Municipal de Ensino Fundamental Martinho Lutero.

Em uma perspectiva antropológica, Krone (2014), da Universidade Federal de Pelotas, realizou um estudo sobre a região de São Lourenço do Sul/RS e abordou a identidade pomerana,

o patrimônio cultural, a relação dos pomeranos com a culinária, as festividades e outras manifestações culturais pomeranas. O autor desenvolveu, ao longo de sua dissertação, a noção de "Pomeraneidade", como um modo de viver e de ser pomerano e, também, de reconstruir a identidade e a cultura pomeranas.

Küster (2015) abordou, em sua dissertação, o Programa de Educação Escolar Pomerana (PROEPO), fez um estudo de caso de uma escola de ensino fundamental em Santa Maria de Jetibá/ES e realizou um estudo sobre a cultura e a "língua" pomerana na perspectiva da educação. A autora também está envolvida no Inventário da Língua Pomerana (ILP) recentemente aprovado (em 2016) pelo Instituto do Patrimônio Histórico e Artístico Nacional – IPHAN. Na primeira etapa, o levantamento linguístico vai abordar 13 Municípios do Espírito Santo, mas poderá ser expandido para outras regiões do Brasil.

Sobre a atual estimativa de falantes de pomerano no Brasil, ainda não existem dados precisos, mas Granzow, quando publicou a primeira edição de seu livro em 1975, afirmou que:

> [...] no interior do Rio Grande do Sul existiam mais de 100.000 descendentes de pomeranos, em Santa Catarina aproximadamente 40.000 e no Espírito Santo aproximadamente 100.000. Destes, 250.000 habitantes, quase a metade ainda falava o pomerano no seu dia-a-dia, em casa. (GRANZOW, 2009, p. 186-187).

Os dados acima precisam ser atualizados, mas já nos dão uma noção a respeito do número de descendentes de pomeranos e de falantes de pomerano no Brasil.

Neste capítulo, abordamos a origem europeia do pomerano, a imigração pomerana, debatemos a definição do pomerano, discutimos aspectos históricos, linguísticos, e, fizemos um mapeamento do pomerano no Brasil, além de listarmos algumas obras produzidas a respeito do tema.

No próximo capítulo, vamos nos deter nas fundamentações teóricas da nossa pesquisa, enfocando a LC e outras abordagens que a complementam.

3. FUNDAMENTAÇÃO TEÓRICA

Neste capítulo expomos as noções que fundamentam nosso trabalho e norteiam nossa reflexão. Trataremos primeiramente de noções básicas de léxico e lexicologia. Também refletimos sobre algumas definições de LC, alguns de seus princípios e conceitos, bem como nos posicionaremos a respeito. Além disso, lançaremos mão de outras abordagens, referenciais teóricos e conceitos que embasam nosso trabalho, como a perspectiva descritiva na Linguística e a relação com as evidências, o empirismo, a representatividade, a Sociogeolinguística como subsídio para compilação de *corpus* oral, dentre outros conceitos.

3.1. O Léxico e a Lexicologia

Nosso objeto de estudo, o léxico pomerano, situa-se na Linguística, dentro da área do conhecimento da Lexicologia.

Nós, seres humanos, sujeitos, etnias, comunidades linguísticas, nomeamos e conceituamos nossas experiências por meio da nossa "cognição da realidade" e *Weltanschauung* (concepção de mundo), expressas por meio do léxico. De acordo com Biderman (2001, p. 13-14), "o léxico de uma língua natural constitui uma forma de registrar o conhecimento do universo" e "[...] pode ser identificado com o patrimônio vocabular de uma dada comunidade linguística ao longo de sua história". Mas, segundo a autora, "os modelos formais dos signos linguísticos preexistem, portanto, ao indivíduo" (BIDERMAN, 2001, p. 14) e o indivíduo-falante processa cognitivamente "o vocabulário nomeador das realidades cognoscentes juntamente com os modelos formais que configuram o sistema lexical" (BIDERMAN, 2001, p.14).

Assim, o léxico assume formas distintas nas várias línguas e culturas, com dimensões significativas próprias e organizações em sistemas semântico-léxico-gramaticais diferentes:

> Assim, podemos afirmar que o homem desenvolveu uma estratégia engenhosa ao associar palavras a conceitos, que simbolizam os referentes. Portanto, os símbolos, ou signos linguísticos, se reportam ao universo referencial. [...] O processo de cognição e de apropriação do conhecimento assumiu formas distintas, ou seja, os sistemas lexicais das numerosíssimas línguas naturais (vivas ou mortas), processo de conceptualização universal, as línguas constituem sistemas muito distintos e variados. [...] Em suma, o universo conceptual de uma língua natural pode ser descrito como um sistema ordenado e estruturado de categorias léxico-gramaticais. As palavras geradas por tal sistema nada mais são do que rótulos, através dos quais o homem interage cognitivamente com o seu meio. [...] Embora se atribua à Semântica o estudo das significações linguísticas, a Lexicologia faz fronteira com a Semântica, já

> que, por ocupar-se do léxico e da palavra, tem que considerar sua dimensão significativa (BIDERMAN, 2001, p.13-14;16).

O léxico é o conjunto de palavras pertencentes a uma língua, e, quando essas palavras são materializadas em um texto, oral ou escrito, são chamadas de vocabulário. De acordo com Biderman (1992), o léxico é:

> [...] o tesouro vocabular de uma língua, incluindo a nomenclatura de todos os conceitos linguísticos e não linguísticos e de todos os referentes do mundo físico e do universo cultural do presente e do passado da sociedade. Esse tesouro constitui um patrimônio da sociedade, juntamente com outros símbolos da herança cultural (BIDERMAN, 1992, p. 399).

O léxico é também detalhado, conforme o conceito formulado por Zavaglia e Welker (2013), como o conjunto rico e dinâmico de todas as palavras de uma língua que possuem organização enunciativa interdependente. Podemos acrescentar, conforme exposto pelos autores, que o léxico é entendido como:

> [...] o conjunto de todas as palavras de uma língua, também chamadas de lexias. As lexias são unidades de características complexas cuja organização enunciativa é interdependente, ou seja, a sua textualização no tempo e no espaço obedece a certas combinações. [...] dele faz parte a totalidade das palavras, desde as preposições, conjunções ou interjeições, até os neologismos, regionalismos ou terminologias, passando pelas gírias, expressões idiomáticas, provérbios ou palavrões (ZAVAGLIA; WELKER, 2013).

A Lexicologia, por sua vez, é o ramo da Linguística que se dedica ao estudo científico do léxico, uma ciência que estuda tanto as unidades lexicais como a organização do léxico. Ainda sobre a abrangência da Lexicologia, é possível esclarecer que:

> [...] Cada palavra remete a particularidades relacionadas ao período histórico em que ocorre, à região geográfica a que pertence, à sua realização fonética, aos morfemas que a compõem, à sua distribuição sintagmática, ao seu uso social e cultural, político e institucional. Desse modo, cabe à Lexicologia dizer cientificamente em seus variados níveis o que diz o léxico, ou seja, a sua significação. Ao lexicólogo, especialista da área, incumbe levar a termo essa tarefa tão complexa sobre uma ou mais línguas (ZAVAGLIA; WELKER, 2013).

Biderman (2001) define a Lexicologia como a ciência que "tem como objetos básicos de estudo e análise a palavra, a categorização lexical e a estruturação do léxico" (BIDERMAN, 2001, p.16); é a aplicação científica ao estudo do léxico, que inclusive, "faz fronteira com

ciências tais como a Dialetologia e a Etnolinguística; nessas áreas interdisciplinares fizeram-se estudos sobre Palavras e Coisas, isto é, sobre as relações entre a língua e a cultura" (BIDERMAN, 2001, p.16).

Segundo Biderman (2001, p.15), "tanto a lexicologia quanto a lexicografia são disciplinas que enfocam o seu objeto de estudo, o léxico, de modos distintos, porém, ambas têm como principal finalidade a descrição desse mesmo léxico". O nosso estudo situar-se-á dentro da primeira, a lexicologia, como ciência da descrição lexical; porém, reconhecemos que a lexicografia, enquanto processo de descrição, documentação e definição do léxico, pode ser futuramente meio de aplicação de um conjunto de *corpora* do pomerano, que poderá culminar na produção de uma obra lexicográfica baseada em *corpora*.

Ainda no sentido de definir a Lexicologia, e, para ampliar sua conceituação, corroboramos Biderman (2001) com o pensamento de Barbosa (1991):

> No campo das definições, pode-se dizer que a lexicologia é o estudo científico do léxico, isto é, propõe-se a estudar o universo de todas as palavras de uma língua, vistas em sua estruturação, funcionamento e mudança, cabendo-lhe, entre outras tarefas: definir conjuntos e subconjuntos lexicais; examinar as relações do léxico de uma língua com o universo natural, social e cultural; conceituar e delimitar a unidade lexical de base – a lexia – bem como elaborar os modelos teóricos subjacentes às suas diferentes denominações; abordar a palavra como um instrumento de construção e detecção de uma "visão de mundo", de uma ideologia, de um sistema de valores, como geradora e reflexo de sistemas culturais; analisar e descrever as relações entre a expressão e o conteúdo das palavras e os fenômenos daí decorrentes (BARBOSA, 1991 *apud* ANDRADE, 2001, p.191).

No excerto acima Barbosa (1991) delineia as definições de léxico e Lexicologia, descreve, em sua conceituação, o processo pelo qual uma língua permite dar nomes aos seres e objetos, classificar, agrupar, identificar semelhanças e discriminar traços distintivos, para estruturar (cognitiva-cultural e linguisticamente) o mundo que nos cerca.

Nosso trabalho sobre o pomerano está situado entre os estudos do léxico baseados em *corpora*, bem como nos tratamentos eletrônicos do léxico, visto que propomos justamente a criação dos *corpora* do pomerano brasileiro para permitir uma descrição lexical dessa variedade.

Alves (2011) aborda como a Linguística de *Corpus* (doravante LC) pode estar a serviço do estudo do léxico, acredita que é possível realizar descrições e estudos lexicais por meio da LC. A autora menciona o exemplo da elaboração de obras lexicográficas baseadas em *corpora*

e reconhece o ganho que o aparato da LC traz para quem deseja estudar o léxico, tornando o processo "menos braçal e mais ágil" (ALVES, 2011, p. 90).

Se a compilação de *corpora* permite o estudo do léxico por meio da descrição baseada nas evidências reunidas, então a LC é uma forma de abordagem do léxico e sua metodologia também permite a organização e o estudo do léxico. O tema da relação com as evidências, a descrição e o empirismo será abordado na seção que se segue.

3.2 A relação com as evidências, a descrição e o empirismo

Sinclair (1991) destaca o surgimento de novas evidências sobre como a linguagem é usada e afirma também que surgem novos paradigmas descritivos influentes, utilizando os elementos fornecidos pela extensa análise de *corpus*. O autor também expõe, em sua referida obra, uma série de ganhos advindos do uso de *corpora* nos estudos sobre a linguagem. Compilamos algumas de suas ideias, conforme listadas a seguir:

> A análise de textos autênticos, falados e escritos, e, em particular, o processamento computacional de textos revelaram padrões linguísticos não esperados […] A grande diferença tem sido a disponibilidade de dados […] a grande novidade foi o registro de evidências completamente novas sobre como a língua é usada […] e as evidências objetivamente compiladas de textos são enormes e sistemáticas […] A língua mostra-se diferente quando examinada extensivamente (SINCLAIR, 1991, p. XVII; I; 2; 4; 100; 489)[22].

Em 2004, Sinclair continua descrevendo os desdobramentos da relação entre os estudos e as evidências contidas nos *corpora*. O autor reitera as mudanças no modo de lidar com as evidências. Para ele, temos que "cultivar uma nova relação entre as ideias que temos e as evidências que estão diante de nós" (SINCLAIR, 2004, p. 17), pois "com grandes *corpora* e potentes computadores estamos nas fronteiras de uma nova visão da língua, onde podemos apreciar toda a sua complexidade sem sermos impedidos pelos detalhes" (SINCLAIR, 2004, p. 166)[23]. Com base nisso, acreditamos que a LC configura uma forma diferenciada de visualizar e explorar a linguagem.

---

[22] No original: *Analysis of extended naturally occurring texts, spoken and-written, and, in particular, computer processing of texts have revealed quite unsuspected patterns of language […] The big difference has been the availability of data […] [The] major novelty was the recording of completely new evidence about how language is used […] and the evidence compiled objectively from texts is huge and systematic […] The language looks different when you look at a lot of it at once* (SINCLAIR, 1991, p. XVII; I; 2; 4; 100; 489).

[23] No original: *With large corpora and powerful computers we are at the frontiers of a new view of language, where we can appreciate its full complexity without getting hampered by the detail* (SINCLAIR., 2004, p. 166).

Carter (2003), ao introduzir a obra de Sinclair (2004), corrobora as ideias do autor, impelindo-nos a confiar no texto, confiar no potencial que existe nos dados:

> As paisagens dos estudos linguísticos estão mudando diante de nossos olhos como resultado das possibilidades propiciadas por *corpora* e pela linguística computacional. E com cada novo avanço ficará cada vez mais claro que o texto poderá e será confiável (CARTER, 2003, p. 6)[24].

Deste modo, Carter (2003) nos mostra como as possibilidades de estudar as evidências linguísticas foram ampliadas pela contribuição dos computadores e esse fato, segundo Carter (2003) e Sinclair (2004), configura novas situações, permite novos estudos com um novo olhar para os dados. Com base nos referidos autores, acreditamos estar diante de uma forma inovadora de abordagem da língua, proporcionada pela LC, a qual leva-nos a descobrir novas hipóteses e nos permite propor novas reflexões sobre a língua em uso.

A necessidade de uma Linguística descritiva é apontada de forma veemente por Perini (2006; 2008) quando ele menciona que falta a ampla descrição de dados linguísticos que sejam observados, organizados e sistematizados para estudo. Perini (2006; 2008) destaca que há carência de dados coletados para se desenvolver uma descrição antes de se propor teorizações, ou seja, o autor chama a atenção para a questão de que teorias são elaboradas sem estarem ligadas, suficientemente, a fatos linguísticos. É importante esclarecer que Perini (2006; 2008) não despreza a importância da teorização, mas dá ênfase ao desenvolvimento de uma linguística descritiva, de maneira a valorizar o uso concreto da língua.

Ao expor os princípios da linguística descritiva (2006) e seus estudos de gramática descritiva (2008), Perini lista algumas características que o linguista deve possuir, a saber: se interessar pela língua como ela é, e não como ela deveria ser; partir sempre de fatos linguísticos para sua análise e não confundir suas hipóteses com os fatos; registrar e descrever como é que as pessoas realmente falam e/ou escrevem; e, sistematizar os fatos da língua encontrados. Observamos que esses princípios também estão presentes na LC.

Embora Perini (2006; 2008) não esteja se referindo diretamente à LC e nem a *corpora* eletrônicos, ele reconhece que o uso de *corpus*, de maneira geral, tem a vantagem de neutralizar os desejos do pesquisador. Pois, se o linguista se valer apenas da introspecção, isso pode influenciar subjetivamente os resultados de sua análise.

---

[24] No original: *The landscapes of language study are changing before our eyes as a result of the radically extended possibilities afforded by corpus and computational linguistics. And with every new advance it will be ever more likely that the text can and will be trusted* (CARTER, 2003, p. 6).

A LC configura para nós uma forma de abordar a língua que permite a descrição, que possui uma metodologia descritiva, porque dá primazia aos dados e às evidências empíricas em detrimento da introspecção do pesquisador.

Lembramos que a LC comporta princípios para obtenção de dados, tanto orais como escritos, com base em linguagem natural, autêntica e na coleta de fontes primárias. Permite, assim, uma boa sistematização dos dados linguísticos e uma descrição da língua objeto do *corpus* de estudo a fim de checar o uso na prática, fornecendo-nos uma base confiável de dados "para construir e testar eventuais teorias" (PERINI, 2006. p. 11).

A LC responde, ao menos parcialmente, às queixas de Perini, quando ele se refere à falta de estudos descritivos, pois para ele é preciso dar "ênfase no levantamento de dados como a maneira de fundamentar as teorias" (PERINI, 2008, p.14). Assim, a LC é uma forma de resposta às queixas de que os estudos linguísticos carecem de mais evidências, de que é preciso elaborar "bases de dados cuidadosamente colhidos, sistematizados e descritos, cobrindo áreas significativamente grandes de línguas naturais particulares", pois a investigação linguística depende de dados "suficientemente amplos e adequados" (PERINI, 2008, p.14).

Pelo seu empirismo e pela sua capacidade de mobilizar uma grande quantidade de dados linguísticos autênticos, a LC constitui uma metodologia-abordagem muito potente. Parodi afirma que a LC apresenta: "oportunidades revolucionárias para a descrição, análise e ensino de discursos de todo tipo […] caracteriza-se por dar sustento à investigação da língua em uso […]" (PARODI, 2010, p.14)[25].

O empirismo é uma das características mais relevantes para a LC, vejamos o que significa a abordagem empírica da LC:

> Com a Linguística de *Corpus*, dados empíricos são fornecidos, tornando a pesquisa mais fundamentada em fatos "concretos" do que especulação. Muitos detalhes que frequentemente "escapam" aos olhos do leitor podem ser percebidos com a ajuda dos dados quantitativos, ajudando a ressaltar certas características dos textos (FEITOSA, 2005, p. 35).

Deste modo, percebemos que a abordagem empírica é um meio de fundamentar a pesquisa em fatos concretos. De acordo com Berber Sardinha (2004, p. 30), "na linguística, empírico significa dar primazia aos dados provenientes da observação da linguagem". Na LC, a abordagem empírica é um pressuposto, pois coletar dados autênticos e trabalhar com a

---

[25] No original: *"oportunidades revolucionarias para la descripción, análisis, y enseñanza de discursos de todo tipo. [...] se caracteriza por brindar sustento a la investigación de la lengua en uso [...]"* (PARODI, 2010, p.14).

linguagem em uso permite-nos uma observação que "revela uma quantidade surpreendente de evidências linguísticas" (BERBER SARDINHA, 2004, p. 384).

Isso está diretamente ligado à noção de linguagem adotada pela LC, pois os dados permitem verificar traços que se repetem, padrões de comportamento linguístico, e assim, atestar se existem regularidades sistemáticas e quantificá-las, confirmando inclusive a hipótese de que não são aleatórias, ou, ainda, refutar hipóteses sobre o funcionamento da língua.

Essas hipóteses podem ter sido previamente estabelecidas, em um estudo *corpus-based* (baseado em *corpus*) ou podem ser levantadas hipóteses à posteriori, a partir da observação dos dados coletados, visto que os resultados de um *corpus* compilado podem direcionar o estudo, nesse caso, um estudo *corpus-driven* (direcionado pelo *corpus*), trataremos desses dois tipos de abordagem em LC no tópico 3.3.4.

Seguindo essa perspectiva, podemos também dizer que a LC permite trabalhos muito variados, com abordagens distintas, e pode ser desenvolvida de modo a instrumentalizar diferentes teorias.

A seguir, retomamos o histórico da LC, a definimos e abordamos alguns conceitos.

3.3 A Linguística de *Corpus* - LC

Nesta seção, discorremos brevemente sobre o histórico da LC, depois, apresentamos um levantamento a respeito das definições que explicam o que é a LC e nos posicionamos em relação aos conceitos expostos. Em seguida, exporemos qual é a concepção de linguagem da LC. Depois disso, faremos uma distinção entre as abordagens baseadas em *corpus* e direcionadas pelo *corpus*. Por fim, trataremos em sequência: dos princípios fundamentais da LC, das prerrogativas de uso dos *corpora* na visão da LC e definiremos alguns conceitos básicos.

3.3.1 Breve Histórico da LC

Com base em Berber Sardinha (2000), delineamos um breve histórico da LC. Em linhas gerais, a descrição da linguagem por meio de *corpora* se desenvolveu no século XX. Entre os linguistas que já se dedicavam à descrição da linguagem por meio de *corpora*, podemos citar, por exemplo, Franz Boas (1911), embora nesse caso ainda não se tratava de *corpus* eletrônico.

Porém, uma das dificuldades era justamente o estudo manual dos dados dos *corpora*, exemplo disso encontramos em Käding (1897). Sua pesquisa sobre a ortografia do alemão precisou de cinco mil analistas para cobrir a quantidade de dados (11 milhões de palavras) a serem verificados (BERBER SARDINHA, 2000).

Segundo Berber Sardinha (2000), no fim da década de 50, houve uma mudança de paradigma na Linguística. Teorias racionalistas da linguagem, como as ideias de Chomsky, publicadas em 1957, relegaram pesquisas empíricas a segundo plano, pois nessa perspectiva, não era preciso coletar dados, já que "os dados necessários estavam na mente e eram acessíveis por meio da introspecção" (BERBER SARDINHA, 2000, p. 324). Na perspectiva chomskyana, o foco é o estudo da competência e, na LC, ao contrário, o foco é investigar o desempenho.

A principal crítica contra os trabalhos que utilizavam *corpora* remetia à não-confiabilidade de uma grande quantidade de dados serem analisadas por meios manuais, mas alguns estudos baseados em *corpora* persistiram e novos *corpora* foram sendo compilados.

Em 1959, o *Corpus* SEU (*Survey of English Usage*), com 1 milhão de palavras, foi compilado e etiquetado manualmente por Quirk e sua equipe. Em 1989, o mesmo foi transformado em *corpus* eletrônico por Svartvik e sua equipe da *Lund University* (Suécia). A parte falada do SEU converteu-se no *Survey of Spoken English* (SSE) e esse *corpus* ficou conhecido como *London-Lund Corpus,* o maior *corpus* falado (500 mil palavras) em língua inglesa nos anos 90, embora o primeiro *corpus* eletrônico de língua falada (220 mil palavras) seja atribuído a John McHardy Sinclair. O SEU inspirou a criação de outros *corpora* eletrônicos.

A revolução, nesse âmbito, deu-se justamente em passar da coleta, compilação e análise manual para a informatizada. Um marco importante para a LC, foi a criação do *Corpus Brown* (*University Standard Corpus of Present-Day American English*). O *Corpus* Brown foi lançado em 1964 e continha 1 milhão de palavras, foi o primeiro *corpus* linguístico compilado eletronicamente.

O advento do computador nos anos de 1960 trouxe uma contribuição fundamental e as pesquisas puramente racionalistas perderam um pouco de prestígio. Assim, o contexto tornou-se favorável ao desenvolvimento da LC.

Havia a necessidade de analisar uma grande quantidade de dados de modo confiável, e foi a partir dessa lacuna que a LC foi se desenvolvendo. O avanço das tecnologias da informação e do computador contribuiu para o desenvolvimento e a multiplicação dos estudos com *corpora*, pois passaram a ser compilados, manipulados e analisados eletronicamente. Segundo Berber Sardinha (2000), a entrada em cena dos microcomputadores pessoais, nos anos 80, também impulsionou a pesquisa linguística baseada em *corpus*.

O uso do computador nas análises linguísticas permitiu com que os linguistas desenvolvessem métodos de análise e observação com o uso da tecnologia e permitiu que

passassem a verificar uma grande quantidade de fenômenos como um cientista observa materiais pela lente de um microscópio.

Essa analogia, que faz uma comparação na qual o computador está para a LC como o microscópio e o telescópio estão para as ciências naturais, foi formulada orginalmente por Stubbs (2001):

> Uma vez que os padrões são encontrados apenas em grandes *corpora*, eles são observáveis apenas com a ajuda da tecnologia computacional. Isso não é um problema em si mesmo: muitas descobertas nas ciências naturais dependem de observações que podem ser feitas apenas com a ajuda de instrumentos como microscópios e telescópios (STUBBS, 2001, p. 168)[26].

Assim, o uso do computador permitiu ampliarmos nosso olhar, verificar fatos em uma amplitude muito maior do que antes, revelando fenômenos novos, ainda inexplorados na linguagem, pois observar tantos fenômenos "a olho nu" seria impraticável sem a lente microscópica da metodologia da LC. Além disso, a LC nos permite observar padrões sistemáticos de uso da língua na sociedade, o que para Stubbs (2001) era impossível explicar por meio de um comportamento individual inconsciente.

A primeira grande obra que adotou "Linguística de *Corpus*" no título foi *English Corpus Linguistics – Studies in honour of Jan Svartvik (*AIJMER; ALTENBERG, 1991).

Conforme conta Berber Sardinha (2000), em 1998 é publicado um manual de LC (BIBER et al., 1998) com uma perspectiva americana da área, pois até então a maioria dos trabalhos eram produzidos da Europa.

Segundo Berber Sardinha (2004) a LC é uma das áreas de pesquisa de linguagem mais ativa nos últimos anos. No Brasil, conforme o autor, um dos primeiros *corpora* do português foi o *corpus* compilado por Biderman (1969)[27].

A LC produz trabalhos diversificados e abre possibilidade para o estudo da linguagem em diversas áreas como estudos lexicais, sintáticos, literários, etc.

---

[26] No original: *Finally, since the patterns are to be found only in large corpora, they are observable only with the help of computer technology. This is no problem in itself: many findings in the natural sciences depend on observations which can be made only with the help of instruments such as microscopes and telescopes* (STUBBS, 2001, p. 168).

[27] Berber Sardinha cita o *corpus* do *Frequency dictionary of Portuguese words* como um dos primeiros *corpora* eletrônicos de português, contendo 500 mil palavras de português europeu. Porém, o autor não deixa claro se trata-se do mesmo *corpus* compilado por Biderman e/ou qual o nome do *corpus* compilado por ela no Brasil.

Devido à multiplicação de estudos com *corpora* e do avanço da LC, podemos afirmar que "passamos da idealização para a sistematização da observação das evidências" (NOVODVORSKI, 2015) do nosso objeto de estudo, no caso, a língua, seja ela qual for.

No próximo tópico, levantamos algumas definições de LC e formulamos nossa concepção sobre a área.

3.3.2 As definições de LC e a nossa concepção de LC

Realizamos um levantamento de diversas definições formuladas a respeito da LC. Vamos, neste momento, nos deter em expô-las para posteriormente nos posicionar em relação a como concebemos a LC.

Nessas definições a LC aparece como:

a) *Campo de criação e análise*: "A LC é um campo que se dedica à criação e análise de *corpora*" (BERBER SARDINHA, 2009, p.7);
b) *Campo disciplinar*: "[...] da Linguística de *Corpus*, campo disciplinar que agrega inovações paradigmáticas calcadas na interlocução dos estudos linguísticos com princípios da Estatística e da Ciência da Computação" (MELLO, 2012, p. 53);
c) *Área*: "área que dedica-se à exploração da linguagem por meio de evidências empíricas" (BERBER SARDINHA, 2004, p. 3);
d) *Área de pesquisa*: "a Linguística de *Corpus* vem despontando como uma importante área de pesquisa" (FEITOSA, 2005, p. 34);
e) *Área do conhecimento*: "A Linguística de *Corpus* é uma área do conhecimento que estuda a linguagem por meio da utilização de grandes quantidades de dados empíricos relativos ao efetivo uso da linguagem, com o auxílio do computador" (GONZALEZ, 2007, p. 8);
f) *Ramo específico do saber* ou *disciplina*: "A disciplina que possibilita essa investigação denomina-se Linguística de *Corpus*. A Linguística de *Corpus* (LC) oferece uma metodologia que veio facilitar muito a identificação das unidades convencionais da língua" (TAGNIN, 2005, p. 21).
g) *Metodologia para investigações empíricas das línguas(gens)*: "uma metodologia para a investigação das línguas e da linguagem, a qual permite levar a cabo investigações empíricas em contextos autênticos e que se constitui em torno de certos princípios reguladores poderosos […]" (PARODI, 2010, p.15)[28].

[28] No original: *stricto sensu, como una metodología para la investigación de las lenguas y del lenguaje, la cual*

h) *Descrição de enunciados de línguas naturais* e *elaboração teórica com base na análise de textos autênticos*: "Como Linguística de *Corpus* chamou-se à descrição de enunciados de línguas naturais, seus elementos e estruturas, e a partir disso, a elaboração de construções teóricas com base em análises de textos autênticos, que estão reunidos em *corpora*" (LEMNITZER; ZINSMEISTER, 2006, p. 10)[29].

Em todas essas definições acima reunidas, a LC é vista de uma maneira mais ampla, como fomentadora de pesquisas e produtora de conhecimento, não se reduzindo a uma metodologia. No tópico "h" acima apresentado, Lemnitzer e Zinsmeister (2006), que fizeram um estudo abrangente sobre trabalhos desenvolvidos na LC do alemão, afirmam que a descrição que a LC permite é voltada para elaborações teóricas a partir de descrições baseadas em textos autênticos; isso nos remete aos anseios de Perini (2008) ao falar da necessidade das descrições. Acreditamos que esses anseios são atendidos pela LC (conforme apresentamos no item 3.2).

Parodi (2010), conforme o tópico "g", vê a LC como uma metodologia investigativa com princípios reguladores poderosos. No entanto, para outros autores, a LC pode ser definida apenas como metodologia. Para McEnery e Wilson (1996) a LC é o "estudo da língua baseado em exemplos de usos linguísticos na 'vida real' […] Linguística de *Corpus* é uma metodologia que pode ser usada em quase qualquer área da Linguística, mas ela não constitui uma área da Linguística em si mesma" (MCENERY; WILSON, 1996, p. 1-2)[30]. Embora a LC não tenha alcançado o patamar de ser uma subárea dentro da árvore do conhecimento da Linguística, a visão de McEnery e Wilson parece limitar a LC, por reduzi-la a uma metodologia, embora reconheçamos que de fato a LC possa ser usada apenas como uma metodologia, para diferentes bases teóricas.

Entretanto, não podemos desconsiderar que são produzidos conhecimentos como resultados da utilização dessa "metodologia"; embora todas as metodologias possam produzir algum conhecimento como resultado de sua aplicação, enfatizamos o olhar peculiar da LC. Para nós, a LC se configura como um modo diferenciado de ver a língua.

---

permite llevar a cabo investigaciones empíricas en contextos auténticos y que se constituye en torno a ciertos principios reguladores poderosos (PARODI, 2010, p.15).

[29]No original: *Als Korpuslinguistik bezeichnet man die Beschreibung von Äußerungen natürlicher Sprachen, ihrer Elemente und Strukturen, und die darauf aufbauende Theoriebildung auf der Grundlage von Analysen authentischer Texte, die in Korpora zusammengefasst sind* (LEMNITZER; ZINSMEISTER, 2006, p. 10).

[30]No original: *the study of language based on examples of 'real life' language use […] Corpus linguistics is a methodology that may be used in almost any area of linguistics, but it does not truly delimit an area of linguistics itself* (MCENERY; WILSON, 1996, p. 1-2).

Assim, visualizar e operar os dados linguísticos sob a lente da LC leva a resultados diferentes, se comparados à análise dos mesmos dados sob outra metodologia-abordagem. Desse modo, acreditamos que a LC não se restringe a um método, mas é também uma abordagem da língua. É como se o caminho percorrido alterasse o ponto de chegada: não que altere os dados em si, mas a forma de percebê-los e analisá-los.

Para corroborar nossa visão, lembramos a afirmação de Berber Sardinha de que a LC é "uma perspectiva, isto é, uma maneira de se chegar à linguagem" (2004, p.37).

Novodvorski e Finatto (2014) também legitimam nossa concepção de LC, de que a mesma é uma forma diferenciada de ver e abordar a língua, pois afirmam que a LC também é um modo de compreender a língua "[...] que permite apreciar um objeto de estudo sob um ângulo diferenciado, que se constitui uma nova área de pesquisa, com abordagens e métodos próprios. [...] espécie de 'caminho de ida e volta', amarrado à importância da observação empírica dos fatos linguísticos" (NOVODVORSKI; FINATTO, 2014, p. 9-10). Analogia na qual os *corpora* são vistos como "um território vasto o suficiente e propício para a descoberta de evidências" (NOVODVORSKI; FINATTO, 2014, p. 10). Perspectiva na qual a língua é concebida como "um sistema probabilístico de combinatórias, no qual uma unidade se define pelas associações que mantém com outras unidades" (NOVODVORSKI; FINATTO, 2014, p. 10).

Ou seja, há uma forma de conceber a LC de maneira mais ampla, reconhecendo seus atributos e contribuições para o conhecimento linguístico. A partir de nossa reflexão sobre todas essas definições de LC aqui mobilizadas, nossa posição é clara: para nós a LC é uma abordagem-metodologia de princípios descritivos, que se fundamenta em dados autênticos e se relaciona com as evidências de maneira ampla, utiliza metodologia de pesquisa empírica, e permite a produção de conhecimentos variados fundamentados na realidade linguística, além de nos guiar a investigar hipóteses não pré-meditadas e permitir a descoberta e a comprovação de fatos linguísticos.

Consideramos a LC como um campo interdisciplinar que, ao ser ao mesmo tempo abordagem e metodologia, produz inúmeros conhecimentos inovadores sobre as línguas. Essa abordagem-metodologia permite produzir os mais diversificados estudos e olhares sobre a linguagem em geral.

Fromm (2010, p.1-2) lista algumas dentre as muitas aplicações da LC: frequência das palavras mais comuns da língua; frequência das classes gramaticais; variações morfossintáticas; comparação de colocações na língua; reconhecimento de lexias compostas e complexas, n-gramas; fraseologismos; regência verbal e nominal; seleção de nomenclatura e dados para uma

obra terminológica; criação de dicionários gerais mono ou multilíngues; verificação de modalidades de tradução em *corpus* mono ou bilíngue; dialetologia; base de dados para tradutores; ensino de língua estrangeira; elaboração de material didático; avaliação de traduções literárias, técnicas e jornalísticas (*corpora* paralelos); subsídios para Análise do Discurso; análises estilísticas; estudos pragmáticos; estudos semânticos etc.

Nesse sentido, podemos, pois, listar alguns estudos com aplicações da LC. Exemplos de estudos em linguística forense são encontrados em Coulthard (1964, apud BERBER SARDINHA, 2009), de avaliações literárias aliadas à LC são encontrados em Gonçalves (2008), de estudos de metáforas por meio da LC em Berber Sardinha (2009), de fraseologia e LC em Tagnin (2011), terminologia e LC em Bevilacqua (2013), estudos contrastivos em tradução com o método da LC, em Ferreguetti, Pagano e Figueredo (2012), e diálogo com a linguística computacional e com o Processamento de Linguagem Natural (PLN) em Pinheiro (2007). Além desses exemplos, também podemos destacar pesquisas diferenciadas baseadas em LC: Fernández Pérez (2011), que compilou e estudou um *corpus* de fala infantil e Nascimento e Tagnin (2014), que realizaram uma análise da tradução dos efeitos sonoros de filmes franceses nas legendas para surdos e ensurdecidos utilizando a metodologia da LC e o subsídio do WST. Assim, listamos apenas alguns tipos de estudo já desenvolvidos no âmbito da LC e sua interdisciplinaridade.

É fato que têm sido produzidos muitos estudos descritivos com base em *corpora* e que resultam em produtos de aplicação, como ferramentas computacionais, cursos, etc. Esse fato já permite que se fale em Linguística de *Corpus* Aplicada (doravante LCA), conforme Mukherjee e Rohrbach (2006, apud ALMEIDA, 2014).

Almeida explica melhor a relação entre estudos que desenvolvem uma descrição da linguagem e a LCA:

> Mukherjee e Rohrbach (2006) consideram que a Linguística de *Corpus* Aplicada se encontra na encruzilhada da linguística descritiva com base no *corpus*, com a pesquisa em aquisição de segunda língua ou língua estrangeira e a pedagogia da língua. A encruzilhada de Mukherjee e Rohrbach (2006) se traduz, de certa forma, nas palavras de Sinclair (2004, p.271): novas evidências, provenientes da linguística descritiva com base no *corpus*, novas prioridades, reveladas pela nova visão da língua e, novas atitudes, que seria a aplicação pedagógica do *corpus*. As pesquisas com *corpora* têm influenciado enormemente a compilação de uma nova geração de dicionários, gramáticas e livros didáticos (ALMEIDA, 2014, p.39).

Podemos citar como exemplo de trabalho que envolve descrição e LCA a própria Almeida (2014), que gerou listas de inadequações relativas ao uso de verbos frequentes no

inglês e, a partir dessas listas, extraídas do seu c*orpus* de estudo, realizou uma aplicação pedagógica, com base em sua análise de resultados, para seus alunos aprendizes de inglês. Outro exemplo refere-se aos autores Murakami e Tagnin (2014), que aplicaram a LC ao ensino, ao compilar um *corpus* de especialidade e preparar o conteúdo programático e o material didático para um curso rápido que tinha o objetivo de preparar alunos para um exame de proficiência em língua inglesa. Conforme Murakami e Tagnin (2014) explicam, eles aplicaram a pesquisa em LC até para definir o que ensinar e como ensinar, conseguindo realizar, de forma eficaz, uma rápida revisão gramatical da língua inglesa, por meio da observação dos fenômenos linguísticos em circunstâncias factuais de uso, demonstrando a abordagem descritiva e pedagógica que a LC possui. Por fim, mencionamos Santos e Brito (2014), que discorrem sobre a Mesa Cartesiana – um produto da LCA desenvolvido para uso no processo de transposição de documentos impressos (no caso, manuscritos históricos do tipo jurídico) para o formato digital, com o intuito de formar *corpora* linguísticos eletrônicos. Os pesquisadores realizaram essa atividade com fidedignidade e praticaram a fotografia de modo cientificamente controlado, a fim de viabilizar e contribuir para o desenvolvimento da LC.

Além dessas aplicações, podemos imaginar uma infinidade de possibilidades para a LC. Por exemplo, a compilação de *corpora* dialetais, de variedades pouco abordadas por estudos em LC, como uma comparação entre um *corpus* do Galês[31] (de *North Wales*) e do Gaélico[32] (de *South Wales*), que permitiria a elaboração de hipóteses sobre a sintaxe dessas variedades, estudos contrastivos entre as duas variedades, comparação de similaridades e traços distintivos, percepções de padrões recorrentes de uso, etc. Ou seja, estamos afirmando que a LC permite produzir conhecimentos inovadores sobre uma grande quantidade de dados e fatos linguísticos.

A respeito da eficácia da LC, apoiamo-nos em Fillmore (1992, p.35), pois ele afirma que "cada *corpus* que tive a chance de examinar, mesmo pequeno, ensinou-me fatos que não poderia imaginar encontrar de nenhum outro modo"[33]. Ocorre que usos até então despercebidos podem sobressaltar ao linguista de *corpus* que, diante de uma tela de concordâncias, por exemplo, percebe variações de uso que são evidenciadas pelas recorrências das próprias variações, que seriam sistemáticas e padronizadas. Assim, poderíamos intuir sobre a padronização da linguagem e indicar que a repetição pode representar um padrão.

---

[31]Localizamos dois *corpora* do Galês. Disponíveis em: <http://corpws.cymru/?lang=en> e <http://bangortalk.org.uk/>. Acesso em: 23/05/2016.
[32]Localizamos um *corpus* do Gaélico. Disponível em: <http://www.dasg.ac.uk/en>. Acesso em: 23/05/2016.
[33]No original*: [...] every corpus that I've had a chance to examine, however small, has taught me facts that I couldn't imagine finding out about in any other way* (FILLMORE, 1992, p. 35).

O privilégio da pesquisa em LC está na sua capacidade de nos permitir enxergar e examinar informações e fatos sobre as línguas, aos quais talvez não teríamos acesso de outro modo, devido à quantidade que agrega e à eficiência dos programas e ferramentas que operacionalizam o trabalho com *corpora,* que embora jamais dispense o papel de um analista humano, realiza contagens, organizações e trabalhos que não seriam factíveis a um ser humano, por meio de trabalho manual.

Disso depreendemos o grande potencial descritivo da LC, e sua forma específica de perceber e percorrer os fatos linguísticos, configurando uma forma diferenciada de abordagem e de visão da língua. Nessa perspectiva, a LC nos permite realizar uma validação empírica dos dados e sua legitimação científica, bem como o fato de operar justamente com uma visão descritiva e não prescritiva da língua(gem), além de estimar a probabilidade de ocorrências, baseada na frequência dos usos reais. Desta maneira, percebemos que a LC indica evidências linguísticas surpreendentes e é, de fato, uma ampliação da nossa capacidade de abordar a língua.

Diante de toda essa discussão, concluímos que a LC é uma opção lógica para o estudo do léxico, pois coletamos textos, que permitem o estudo do léxico em seu contexto de uso, por exemplo, por meio das listas de concordâncias onde visualizamos um item lexical no contexto e no cotexto.

Por isso, podemos estudar o léxico em vários âmbitos e em maior profundidade, abrindo inúmeras possibilidades de explorar os conteúdos e analisá-los, tanto em nossa língua materna quanto na investigação de outras línguas.

Em uma geração de fluidez da informação, de comunicação de massa, em que tudo acontece de maneira muito rápida, a LC vem atender às necessidades de se conseguir trabalhar com uma grande quantidade de documentos em menor tempo e, assim, podemos lançar um olhar para esses dados de uma forma mais ampla e eficiente.

Encontramos em Fromm (2003) respaldo para afirmar que é possível produzir conhecimento sobre o funcionamento das línguas, com base nos fatos evidenciados pela análise de *corpora*, sob os princípios da LC:

> [...] acreditamos que há a real possibilidade de construção de teorias linguísticas baseadas em fatos, passíveis de serem re-analisados. Essas teorias não se baseariam em soluções (ou mais especificamente, exemplificações) criadas por autores, mas sim em colocações autênticas da língua em estudo, conferindo-lhes um caráter científico (FROMM, 2003, p. 76).

Com base nessa afirmação, podemos salientar que não só é possível a construção de teorias baseadas nos fatos linguísticos observados em *corpora*, como também que a LC já

possui uma concepção de linguagem firmada na recorrência dos padrões percebidos nos *corpora*. É sobre essa concepção que vamos tratar no próximo tópico.

3.3.3 A Concepção de Língua da LC

Segundo Gonzalez (2007), a determinação de padrões de língua pode contribuir de forma relevante para a investigação da padronização do léxico.

Padrões são todas as palavras e estruturas regularmente associados. "Um padrão pode ser identificado se uma combinação de palavras ocorre com relativa frequência, se é dependente de uma palavra específica, e se há um significado claro associado" (BERBER SARDINHA, 2004, p. 39).

A padronização da língua pode ser percebida pela recorrência de combinações de uso, além de permitir que observemos que mesmo as variações são sistemáticas. Ao analisar os padrões, podemos notar que o funcionamento da estrutura aponta para uma dependência entre léxico e gramática, conforme Berber Sardinha (2004) explica:

> [...] estruturas que se repetem significativamente mostram sinais de ser um padrão lexical ou léxico-gramatical. A linguagem forma padrões que apresentam regularidade (estáveis em momentos distintos, isto é, têm frequência comparável em *corpora* distintos) e variação sistemática (correlacionam-se com variedades textuais, genéricas, dialetais etc.) [...] a não-trivialidade da investigação da frequência de ocorrência de traços linguísticos (lexicais, sintáticos, semânticos, discursivos), pois é pelo conhecimento da frequência atestada que se pode estimar a probabilidade teórica (BERBER SARDINHA, 2004, p. 31-32).

A regularidade dos padrões e a variação sistemática, que permitem inferir sobre a probabilidade de ocorrências na língua, podem ser examinadas por meio de estudos. A visão da linguagem como probabilidade tem em Halliday seu maior expoente. Segundo Gonzalez (2007, p. 18-20), "um dos maiores expoentes da visão probabilística é Halliday (1991, 1992, 1993)", que desenvolveu estudos sobre probabilidades em sistemas linguísticos (em 1950, ao preparar uma gramática da língua chinesa) e realizou pesquisas que atestam o potencial do uso das probabilidades para as descrições linguísticas.

A visão da língua como sistema probabilístico adotada pela LC implica, segundo Gonzalez (2007, p. 20), "em repensarmos o conceito de 'escolha' na língua em uso", e considerar que as probabilidades fazem parte dos sistemas linguísticos. Essa visão amplia nosso entendimento e enriquece nosso conhecimento sobre o funcionamento das línguas. Além disso, fomenta a elaboração de hipóteses e permite as descrições linguísticas.

3.3.4 Tipos de abordagens em LC: *Corpus-based* e *Corpus-driven*

A distinção entre as abordagens *Corpus-based* (baseada em *corpus*) e *Corpus-driven* (direcionada pelo *corpus*) foi feita pela primeira vez por Tognini-Bonelli (2001).

A respeito da abordagem *Corpus-based*, segundo as reflexões de Tognini-Bonelli (2001), podemos observar que:

> [...] o termo *Corpus-based* é utilizado para se referir a uma metodologia que se utiliza principalmente de *corpus* para expor, testar ou exemplificar teorias e descrições que foram formuladas antes de grandes *corpora* terem sido disponibilizados para informar estudos linguísticos. [...] se um *corpus* era para ser usado para avaliar teorias, a teoria teria de ser posta de uma forma explícita para que aqueles aspectos do *corpus* que ela cobria pudessem ser distinguidos daqueles em que a teoria não cobria, ou discordava das evidências. Tal relação entre teoria e dados é a relação clássica em Linguística (TOGNINI-BONELLI, 2001, p. 65)[34].

Para Tognini-Bonelli (2001) é necessário considerarmos que as evidências podem ou não confirmar as categorias previamente estabelecidas em um estudo *Corpus-based*, que é uma visão mais restrita, por tratar a evidência do *corpus* como meio para extração de exemplos e checagem de hipóteses, mas a evidência trazida à luz pode fazer com que as formulações prévias sejam rejeitadas ou confirmadas, porém a evidência não pode ser ignorada. A abordagem baseada em *corpus*, geralmente, se limita a quadros teóricos prévios e corre o risco de tentar "encaixar a evidência" nesse quadro.

Ainda assim, existem muito trabalhos que reconhecem a importância de um estudo baseado em *corpus* e para muitos estudos a abordagem *Corpus-based* parece ser a melhor opção. Taylor e Francis ([20--]), por exemplo, falam em estilística de *corpus* (*Corpus stylistics*) e afirmam que, para a necessidade de pesquisa que tinham, a *Corpus-based approach* foi a melhor ferramenta que puderam encontrar para atender aos seus objetivos:

> Nossa decisão em usar uma abordagem baseada em *corpus* deveu-se ao fato de ser a melhor ferramenta que encontramos para realizar o tipo específico de investigação que tínhamos em mente. [...] Isso levou à ideia de desenvolver

---

[34] No original: *[...] the term corpus-based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study. [...] If a corpus was to be used to evaluate one of this class of theories, the theory would have to be put into an explicit form so that those aspects of corpus patterning that it covered could be distinguished from those where the theory did not cover, or was at variance with, the evidence. Such a relationship between theory and data is the classical one in linguistics* (TOGNINI-BONELLI, 2001, p.65).

> um método de anotação textual sistemática e replicável que seria usado amplamente (TAYLOR; FRANCIS, [20--], p. 6).[35]

Mas, se direcionado pelo *corpus*, o estudioso pode postular novas hipóteses. Tognini-Bonelli (2001) reconhece que a observação leva à hipótese que leva à generalização e assim podemos fazer afirmações teóricas a respeito do funcionamento linguístico.

A autora ressalva ainda, que na LC, a metodologia de verificação dos dados, não é mecânica, mas mediada constantemente pelo linguista, que continua comportando-se como um linguista e aplica o seu conhecimento, experiência e inteligência em todas as fases durante o processo de análise. A abordagem *Corpus-driven*, direcionada pelo *corpus*, pode extrair categorias linguísticas sistematicamente, a partir dos padrões recorrentes e distribuições regulares, observados a partir da linguagem no contexto.

A respeito da abordagem *Corpus-driven* para a LC, Tognini-Bonelli (2001) explica que:

> Na abordagem direcionada por *corpus*, o linguista usa um *corpus* para além da seleção de exemplos para confirmar um argumento linguístico ou validar uma teoria. Em uma abordagem direcionada por *corpus*, o compromisso do linguista é com a integridade dos dados como um todo, e as descrições objetivam ser amplas em relação às evidências do *corpus*. O *corpus*, portanto, é mais do que um repositório de exemplos usados para apoiar teorias pré-existentes ou determinar probabilisticamente um sistema já bem definido. As teorizações são totalmente consistentes com e refletem diretamente as evidências fornecidas pelo *corpus* (TOGNINI-BONELLI, 2001, p.84)[36].

Tognini-Bonelli vê a abordagem *Corpus-driven* de modo mais otimista, pois para ela, nesse âmbito o *corpus* não se limita a ser um repositório de exemplos, para apoiar teorias previamente formuladas, mas, ao contrário; as demonstrações teóricas vão ser totalmente coerentes com as evidências fornecidas pelo *corpus*.

Assim, a autora afirma que a LC é uma nova abordagem, e essa é uma das razões pelas quais a confiabilidade do *corpus* suscita preocupação. Mas, ela mesmo, nos lembra que isso pode ser remediado por meio do monitoramento constante do *corpus*, como sustenta Biber

---

[35] No original: […] *our decision to use a corpus-based approach was because it was the best tool we could find to carry out the particular kind of investigation we had in mind […] This led to the idea of developing a method of systematic and replicable textual annotation which would be used comprehensively* (TAYLOR; FRANCIS, [20--], p. 6).

[36] No original: […] *the corpus-driven approach to corpus linguistics, where the linguist uses a corpus beyond the selection of examples to support linguistic argument or to validate a theoretical statement. In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus* (TOGNINI-BONELLI, 2001, p.84).

(1994 *apud* TOGNINI-BONELLI, 2001, p. 88). A evidência do *corpus* se relaciona com a descrição e com a teoria exatamente da mesma forma que qualquer outra evidência científica faz, passa por procedimentos de amostragem, testagem e etc.; o que só aumenta a autoridade e a confiabilidade das evidências encontradas no campo. Com base nessas reflexões, Tognini-Bonelli (2001) afirma que a teoria não existe independente da evidência. E a LC, principalmente a abordagem direcionada pelo *corpus*, demonstra essa relação de "permitir a formulação de constructos teóricos com base nas evidências linguísticas" (TOGNINI-BONELLI, 2001, p.68-88).

No entanto, acreditamos também que existem estudos *Corpus-based* e *Corpus-driven* ao mesmo tempo, pois os resultados do *corpus* podem ser surpreendentes a ponto de redirecionar as hipóteses, fazendo com que sejam reformuladas ou ainda que preservadas as primeiras hipóteses, podem permitir o surgimento de novas hipóteses diante da riqueza dos dados fornecidos pelos *corpora*.

Posto isto, fica clara a importância desse arsenal para estudos linguísticos. Quando utilizamos um *corpus* ou diversos *corpora* para desenvolver análises linguísticas, estamos fazendo a Linguística de *Corpus*, mas, para isso, alguns princípios precisam ser observados. É sobre isso que tratamos no próximo tópico.

3.3.5 Prerrogativas do uso de *corpora* nas pesquisas

Ao falar das prerrogativas do uso de *corpora* eletrônicos, Tagnin (2001, p.1-2) lista algumas das principais vantagens proporcionadas por sua utilização:

1. É de fácil acesso, seja por meio de CD-ROM ou *online*;
2. Possibilita manipulação de grande quantidade de dados;
3. Dá acesso a dados observáveis e verificáveis;
4. Permite repetir uma experiência, confirmando intuições e/ou conclusões;
5. Permite estudos de frequência que podem oferecer informações importantes tanto com referência aos dados encontrados [como as palavras mais comuns na língua alvo] quanto aos não encontrados;
6. Facilita a obtenção de dados quantitativos/estatísticos;
7. Permite generalizações, que podem, por exemplo, servir de base para regras gramaticais;
8. Permite a confecção de material de ensino, seja no âmbito de aprendizado de língua seja no da tradução;
9. Permite ao não nativo acesso a dados autênticos (TAGNIN, 2001, p.1-2).

Gostaríamos de complementar a lista acima, acrescentando mais algumas vantagens que percebemos no uso de *corpora*:

1. Permite identificar padrões de uso (lexicais, colocacionais, etc.);
2. Permite um olhar diferenciado sobre as línguas (leitura vertical, palavras em contextos, checagem da lista de concordâncias que são dispostas de forma compreensível e sistemática, comparação simultânea entre as linhas de diversos textos, etc.);
3. Permite desenvolver estudos qualitativos;
4. Permite obter alto grau de objetividade e confiabilidade, pois contém um método de análise isento de opinião e de julgamento prévio;

Portanto, concluímos que o estudo de um *corpus* ou de um conjunto de *corpora*, escrito e/ou oral, com o auxílio de um computador, ajuda-nos a superar algumas limitações humanas pela amplitude, bem como obter maior representatividade. Se considerarmos a vastidão do material mobilizado, nos permite consultar fontes primárias em uma quantidade abundante. Permite-nos também observar combinações que são de fato utilizadas em contextos reais de uso.

Importa-nos o uso que os falantes fazem da língua no sistema comunicativo, o modo como utilizam-na no cotidiano. O uso de *corpora* nos estudos linguísticos, sob os princípios da LC, permite examinar um vasto conteúdo reunido a partir do comportamento linguístico real e proveniente da fala espontânea e da escrita autêntica de pessoas reais.

Na próxima seção, definimos alguns princípios e conceitos sob a perspectiva da LC.

3.3.6 Princípios fundamentais e conceitos importantes

No âmbito da LC, para que um conjunto de dados linguísticos seja um *corpus*, existem alguns princípios que devem ser considerados. Segundo Berber Sardinha (2004, p.18-19), esses princípios ou critérios são: a origem, pois os dados devem ser autênticos e escritos [ou falados] por falantes nativos [da língua alvo]; o propósito: os dados devem ser objeto de estudo linguístico; a composição: os dados devem ser escolhidos e colhidos com critério; a formatação: os dados devem ser legíveis por computadores; a representatividade, em que os dados devem ser representativos de uma língua ou de uma variedade linguística, o que na prática significa dizer que o *corpus* deve ser o maior possível e a extensão: o material deve ser vasto para ser representativo.

Além dos princípios acima definidos com base em Berber Sardinha (2004), vamos, na subseção a seguir, apresentar alguns princípios e/ou conceitos como planejamento e construção de *corpus*, representatividade, extensão, balanceamento, lematização, concordanciador, lista de frequência, e detalhar alguns conceitos como *corpus* (oral e escrito), *corpora* e tipos de *corpus*.

3.3.6.1 Planejamento e construção de *corpus*

Uma questão importante que deve ser observada para a compilação de um *corpus* é o planejamento. As etapas de planejamento, execução e tratamento dos dados devem ser previamente formuladas e organizadas a fim de que a compilação do *corpus* (escrito e/ou oral) possa atender aos seus propósitos. É o que explica Fromm:

> Antes de esquematizarmos um projeto para um *corpus*, devemos ter em mente quais aplicações práticas queremos retirar desse *corpus*. Existe uma variada gama de análises linguísticas que podemos fazer a partir dele. [...] A construção de um *corpus* ou *corpora*, gerais ou específicos, requer um grande planejamento prévio por parte do pesquisador. A falta desse poderá invalidar os dados obtidos na futura pesquisa (FROMM, 2003, p. 70; 76).

Tendo essa recomendação em mente, além dos critérios que devem ser observados para compilação de *corpora* escritos e orais, também podemos considerar princípios interdisciplinares e métodos auxiliares, de outros ramos da linguística, para planejarmos e esquematizarmos a compilação de *corpora*.

A construção de *corpus*, segundo McEnery (2013, p. 5) é "o processo de planejamento de um *corpus*, coleta de textos, codificação do *corpus*, organização e armazenamento de metadados, anotação de textos quando necessário e possivelmente adição de etiquetagem linguística[37]. Concordamos com a definição do autor, exceto pela adição de anotação, que a nosso ver, não faz parte da construção de um *corpus* e sim do tratamento dos dados do *corpus*, sendo um filtro adicional que altera o modo original como os dados estavam dispostos quando encontrados e coletados. A anotação faz parte do que chamamos de tratamento dos dados, conceito que definiremos no capítulo 4, no qual descrevemos nossa metodologia.

3.3.6.2 Representatividade e extensão

Quanto à representatividade, consideramos que deve ser avaliada do ponto de vista dos objetivos aos quais os *corpora* se referem. Tagnin (2010) esclarece que não há um consenso sobre esse conceito, mas que o *corpus* deve ser representativo daquilo que pretende estudar. Assim, entendemos que o *corpus* deve atender minimamente à necessidade para o qual foi criado e que, nas palavras da referida autora, "cabe ao criador do *corpus* estabelecer os critérios que garantam essa representatividade" (TAGNIN, 2010, p. 360). A nosso ver, um dos principais

---

[37] No original: *The process of designing a corpus, collecting texts, encoding the corpus, assembling and storing the metadata, marking up the texts where necessary and possibly adding linguistic annotation* (MCENERY, 2013, p. 5).

critérios que garante a representatividade é a autenticidade dos dados coletados e não necessariamente a extensão do *corpus*.

Deste modo, acreditamos que nem sempre a representatividade estará diretamente ligada à extensão do *corpus*, pois um *corpus* de especialidade, de uma área nova, como a Linguística Forense[38], por exemplo (segundo Berber Sardinha, 2000; 2004), não poderá ser muito extenso, até que novos materiais sejam produzidos de forma autêntica sobre o assunto. No caso de *corpora* dialetais, quando se tratar de variedades em processo de desaparecimento, dificilmente se obterá um mega-*corpora* em extensão, porém ainda assim se fará importante reunir material sobre o dialeto em um *corpus*, visto que se trata de uma forma de acervo, com material legítimo que poderá documentar o mesmo.

Ainda sobre a questão da representatividade dos *corpora*, Tognini-Bonelli (2001) explica que, geralmente, ela é avaliada comparando o leque de variabilidade de uma amostra com as características de uma população alvo. Embora a autora afirme que a representatividade possa ser medida apenas em termos relativos (estatísticos), questionamos se a representatividade não deveria também ser avaliada em termos qualitativos. Para Tognini-Bonelli, um *corpus* que é tomado para ser representativo é projetado para ser usado como base para generalizações sobre o sistema linguístico.

Quanto ao critério do *corpus* ser o mais vasto e extenso possível, parecendo estar o tamanho relacionado indispensavelmente à representatividade, lembramos que, atualmente, são considerados os objetivos de cada pesquisa antes de determinar o tamanho do *corpus*. Outrossim, se um *corpus* pretende salvaguardar, descrever e estudar um dialeto em processo de desaparecimento, seu tamanho provavelmente não será extenso, mais ainda assim será significativo e, desde que atenda aos objetivos propostos pelo pesquisador, poderá ser considerado, sim, um *corpus* representativo.

3.3.6.3 Balanceamento

Quanto ao princípio de um *corpus* precisar ser balanceado, contendo cada um de seus textos um tamanho igual ou aproximado, observamos que, em casos específicos, não é possível atingir uma mesma quantidade de textos de cada tipo e gênero para a coleta e nem que esses textos sejam do mesmo tamanho.

---

[38] Em 2014, durante a disciplina de "Lexicologia, Lexicografia e Terminologia", tentamos compilar um *corpus* de Linguística Forense em inglês e em português, porém não encontramos uma quantidade relevante de artigos sobre o assunto, principalmente em língua portuguesa.

Tendo em vista que, no caso da compilação de *corpus* ou *corpora*, de uma variedade em risco, qualquer texto que se obtenha é importante para permitir o conhecimento do léxico em questão. Observamos que, no nosso caso, o "desequilíbrio" (não-balanceamento) do conjunto de *corpora* do pomerano, que nos propomos a compilar, é diretamente proporcional à sua representatividade, ou seja, quanto mais diversificados são os conteúdos e tipologias das amostras, mais representatividade conseguimos obter acerca do pomerano, tendo em vista o objetivo de documentar o léxico e de resgatar a variedade linguística.

3.3.6.4 Frequência e lista de frequência

Segundo Gonzalez (2007), a frequência é importante para a investigação da linguagem do ponto de vista probabilístico, e "é preciso ter em mãos um instrumento que nos permita aferir essa probabilidade. Esse instrumento básico é a lista de frequência" (GONZALEZ, 2007, p.15).

A lista de frequência é valiosa para o estudo de um *corpus* no âmbito da LC, pois nela há "uma lista de todos os itens de um *corpus* juntamente com a frequência de ocorrência de cada um" (MCENERY, 2013, p.6)[39] dentro do *corpus* de estudo, permitindo-nos verificar quais são os itens mais e menos frequentes.

Para observar a lista de frequências é preciso gerar primeiro a lista de palavras, que oferece a relação das palavras em ordem alfabética ou em ordem de frequência. Além de obter essa função de listar as frequências no *corpus*, a lista de palavras também permite verificar a quantidade total de palavras nos textos (*tokens*/itens) e a quantidade de palavras diferentes nos textos (*types*/formas) a partir da função estatística. A função estatística, acessada a partir da lista de palavras também fornece a relação *type*/*token*, opção presente na maioria dos programas que operam sob os princípios da LC. As estatísticas também oferecem o número de sentenças e parágrafos existentes no *corpus*.

3.3.6.5 *Type* e *Token*

Nesta seção, vamos definir o que é um *type* e o que é um *token*.

Segundo McEnery (2013, p. 9), um *type* é a "forma individual de uma palavra. Qualquer diferença de forma cria uma palavra de um tipo diferente. Todos os itens constituídos dos mesmos caracteres são considerados exemplos de uma mesma forma"[40].

---

[39] No original: *Frequency list - A list of all the items of a given type in a corpus together with a count of how often each occurs* (MCENERY, 2013, p. 6).

[40] No original: *Type – A single particular wordform. Any difference of form makes a word a different type. All tokens comprising the same characters are considered to be examples of the same type* (Idem, p. 9).

Ainda segundo o autor, o *token*, ou item, é "cada instância particular e individual de uma palavra em um texto ou *corpus*" (MCENERY, 2013, p. 9)[41].

3.3.6.6 Lematização

A lematização é o ato de reunir sob um *lemma* suas formas flexivas e derivadas dentre os itens de um *corpus*, em outras palavras, consiste em **"**agrupar palavras relacionadas com a mesma base de palavra ou radical, diferindo apenas pela inflexão"[42] (MCENERY, 2013, p. 6). A lematização também pode ser utilizada para agrupar um mesmo item que tenha sido escrito de maneiras diferentes, como por exemplo, sua forma padrão e suas variações.

3.3.6.7 Concordanciador

O Concordanciador é um instrumento que serve para mostrar as linhas de concordâncias a partir de um nódulo, que é uma palavra escolhida, que fica na posição central da linha de concordância. As concordâncias são os contextos imediatos que cercam a palavra desejada. A linha de concordância pode ser expandida indo até ao texto no qual a palavra escolhida está originalmente inserida. Por meio da palavra em contexto, vários tipos de análise podem ser feitas.

Segundo Gonzalez (2007, p. 16), "uma prática frequente dos Linguistas de *Corpus* é a observação das palavras em seu contexto original. Esse procedimento possibilita evidenciar claramente o significado e/ou o uso contextualizado das palavras".

Há muitos outros conceitos importantes na LC que se referem a parâmetros, princípios e funções de análise, porém, optamos aqui por definir apenas aqueles que são fundamentais e/ou que serão mobilizados dentro do nosso trabalho. Por isso, consideramos desnecessário nos deter em numerosas definições que não serão retomadas no presente trabalho.

3.3.6.8 *Corpus* escrito e oral

Segundo Sinclair, *corpus* é uma "coleção de textos" (1991, p. 171), e nesse sentido, já existia antes da Linguística de *Corpus*, mas no caso desta, é uma coleção de textos "de

---

[41] No original: *Token - Any single, particular instance of an individual word in a text or corpus* (MCENERY, 2013, p. 9).
[42] O lemma é definido originalmente por McEnery como: *A group of words related to the same base word differing only by inflection* (MCENERY, 2013, p. 6).

ocorrências de linguagem natural" (SINCLAIR, 1991, p. 171), ou seja, autêntica, escolhidos para caracterizar um estado ou variedade de linguagem.

Berber Sardinha (2004, p. 16-17) define *corpus* como "uma coletânea de textos reunida com um propósito definido de ser usado como base para a pesquisa linguística". O autor estabelece também que *corpus* é um artefato produzido para a pesquisa, com textos autênticos, com "porções de linguagem que são planejadas, selecionadas, organizadas, de acordo com critérios linguísticos explícitos, a fim de serem usadas como uma amostra da linguagem e armazenadas em formato legível por computador" (BERBER SARDINHA, 2004, p. 16-17).

Embora essa seja uma das explicações mais completas do que é um *corpus*, não revela de que tipos podem ser os textos, e sua definição aplica-se tanto as características de um *corpus* escrito quanto oral. Por isso consideramos que essas características que definem *corpus* podem ser complementadas pela definição de Biderman (2001, p. 79): "*Corpus* constitui um conjunto homogêneo de amostras da língua de qualquer tipo (orais, escritos, literários, coloquiais, etc.). [...] A análise dos dados linguísticos de um *corpus* deve permitir ampliar o conhecimento das estruturas linguísticas da língua que eles representam".

Um *corpus,* escrito ou oral*,* constitui-se quando selecionamos um corpo de documentos, de textos, de transcrições, fontes primárias, etc., da realidade, ou seja, autênticos e produzidos por humanos. A partir desse material composto como um conjunto, debruçamos-nos sobre o mesmo a fim de realizar um estudo linguístico, tendo em vista que a seleção e organização das amostras de linguagem já foram feitas com os critérios específicos para um estudo linguístico, sob o método e abordagem da LC.

Há diversos autores que definem *corpus*. De um modo geral, não observamos grandes variações entre as definições, apenas constatamos que os conceitos complementam-se. Em síntese, *corpus* "é uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina" (TAGNIN, 2005, p. 21). Os textos podem ser aqui definidos como "amostra de linguagem falada ou escrita delimitada segundo critérios dos compiladores do *corpus*" (BERBER SARDINHA, 2004, p. 21, n. 3), de modo que a decisão do compilador, segundo seus objetivos e critérios, irá interferir diretamente nos conteúdos que farão parte do seu *corpus*, que podem ser provenientes de dados da fala ou coletados de dados na forma escrita.

A referida coletânea de textos é definida também como "um conjunto de unidades textuais, como uma coleção finita de um universo infinito" (PARODI, 2010, p.24), mas não é uma única instância comunicativa, pode representar várias instâncias, e, deve

indispensavelmente reunir os dados que representa em formato eletrônico (no âmbito da LC), para possibilitar que os dados sejam investigados com o uso de ferramentas computacionais.

Embora não haja contradição entre essas definições, que se completam para o entendimento do que é um *corpus*, acreditamos que a definição de Parodi (2010) é esclarecedora: "[...] podemos falar em um *corpus*, no contexto da LC, no mínimo como um conjunto de textos que é formado pelo menos por dois ou mais textos, [...] deve conter [...] certos traços definitórios, limitado somente por características inerentes à natureza dos mesmos" (PARODI, 2010, p. 25)[43].

A partir dos conceitos apresentados, concluímos que *corpus* é uma coleção de textos escritos ou falados, um conjunto de dados de partes de uma língua ou dialeto, coletados criteriosamente a partir de fontes autênticas, provenientes de exemplos reais de uso linguístico natural, os quais são organizados para a finalidade de um estudo linguístico e que a partir de dois textos (ou arquivos de amostras da linguagem) já se pode falar em um *corpus*.

Até aqui nos detemos em definir e avaliar as características de *corpus*, sem, contudo, distinguir características específicas para *corpus* oral. Passaremos agora a tratar especificamente deste tipo de *corpus*, devido à importância que esse conceito tem dentro no nosso trabalho e da nossa proposta metodológica de compilação de *corpora* dialetais, não só de textos escritos, como também de um *corpus* oral.

Um *corpus* também pode ser formado por conjunto de dados linguísticos provenientes do uso oral. Desde que os dados que o compõe sejam coletados a partir de material autêntico na língua alvo.

Um *corpus* oral pode ser alinhado, no qual é possível escutar a fala simultânea à sua escrita, para ser estudado preservando as características próprias da fala, que sem a sonoridade ficam limitadas à escrita. Porém, na impossibilidade de realizar um alinhamento, um *corpus* oral também pode ser composto por transcrições de fala, em que se estuda o *corpus* em seu ambiente escrito, porém com base no material coletado e transcrito originalmente da fala.

Para a compilação de um *corpus* oral é preciso ter em conta alguns critérios, além de observar os mesmos princípios que a LC estabelece para a compilação de dados escritos, como, por exemplo, a autenticidade e a representatividade. Também é preciso observar que sejam dados espontâneos e que as transcrições sejam padronizadas para todos os dados orais coletados.

---

[43] No original: *conjunto de textos que está formado por al menos dos o más textos [...] debe contener [...] ciertos rasgos definitorios, limitado solo por características inherentes a la naturaleza de los mismos* (PARODI, 2010, p. 25).

Corpora orais igualmente devem ser sistematizados segundo os critérios da LC: originalidade, propósito, formatação etc. Porém, quanto à extensão, no caso de um *corpus* oral torna-se mais difícil devido a mão de obra necessária para as transcrições, que são, geralmente, difíceis e demoradas, mas ainda assim, um *corpus* oral é representativo de algum âmbito do uso linguístico, ainda mais espontâneo por ser proveniente da fala, em geral, menos monitorada do que a escrita.

Do nosso ponto de vista, um *corpus* oral tem sua autenticidade mais facilmente atestada, pois as transcrições revelam a forma do uso espontâneo da língua, com maior liberdade sintática nas expressões, e porque a origem do conteúdo advém da coleta que se faz na pesquisa de campo, com falantes da língua-alvo do *corpus*. De modo que, conforme afirma Sánchez (1995), um *corpus* oral é igualmente capaz de propiciar resultados vários e úteis para a descrição e análise linguística.

> Um *corpus* linguístico é um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade de maneira que sejam representativos do total do uso linguístico ou de algum de seus âmbitos e dispostos de tal modo que possam ser processados por computador, com a finalidade de obter resultados vários, úteis para a descrição e a análise (SÁNCHEZ, 1995, p. 5)[44].

Neste trecho, podemos observar que tanto um *corpus* escrito quanto um oral precisam estar em um formato legível por computador, o que torna as transcrições indispensáveis para que os dados constituam um *corpus* e também que sejam organizados e sistematizados.

Ao citar Crystal (1991), Parodi (2010), esclarece ainda melhor a questão do *corpus* oral, pois informa que um *corpus,* enquanto uma coleção de dados linguísticos, também pode ser composto de transcrições de fala gravada:

> Uma coleção de dados linguísticos, seja de textos escritos seja de transcrições de fala gravada, que pode ser utilizada como ponto de partida para descrições linguísticas ou como meio de verificação de hipóteses acerca de uma língua (CRYSTAL, 1991, p.32 apud PARODI, 2010, p. 21)[45].

---

[44] No original: *Un corpus lingüístico es un conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con el fin de obtener resultados varios útiles para la descripción y el análisis* (SÁNCHEZ, 1995, p. 5).

[45] No original: *Una colección de datos lingüísticos, ya sea de textos escritos o de transcripciones de habla grabada, los que pueden ser utilizados como punto de partida para descripciones lingüísticas o como un medio de verificación de hipótesis acerca de una lengua* (CRYSTAL, 1991, p.32 apud PARODI, 2010, p. 21). Esclarecemos

Aqui fica claro que um *corpus* oral também possui o potencial de permitir descrições linguísticas e verificação de hipóteses sobre uma língua, não sendo esta uma característica exclusiva de *corpora* escritos.

Para compilação de um *corpus* oral é preciso enfrentar algumas dificuldades adicionais: além de adotar uma metodologia para a coleta dos dados, também é necessário optar por um método de transcrição para alcançar o rigor metodológico e a representatividade dos dados da fala. Brum-de-Paula e Espinar (2002) tratam dessas dificuldades e da importância do estudo das produções orais:

> [...] realização linguística submetida aos imponderáveis de uma tarefa que se desenvolve em tempo real. Em pesquisas cujo objeto é a oralidade, o pesquisador necessita prever problemas metodológicos e teóricos adicionais. De fato, a fim de empreendê-las precisamos notadamente efetuar uma coleta de dados e uma transcrição das gravações efetuadas. [...] o oral conquistou um espaço importante e credível dentro dos estudos sobre a linguagem (BRUM-DE-PAULA; ESPINAR. 2002, p. 1).

Transcrever um *corpus* oral para realizar um estudo a partir da oralidade permite, segundo as autoras, "o acesso a sistemas linguísticos imersos no ambiente em que eles se originam, se transformam ou desaparecem" (BRUM-DE-PAULA; ESPINAR, 2002, p.2). Porém, segundo as autoras, devido ao processo de coleta e transcrição, que torna mais demorada a composição de um *corpus* oral, eles não são tão numerosos quanto os *corpora* compilados com base em textos escritos.

A transcrição deve ser desenvolvida com base nos objetivos da pesquisa e, também, considerando a possibilidade de que outros trabalhos possam ser desenvolvidos posteriormente, com base nas transcrições. No caso de um *corpus* para ser estudado sob os princípios da LC, é indispensável que as transcrições sejam legíveis por computador e pelo *software* adotado, pois as transcrições devem permitir a geração de listas de palavras, o acesso às linhas de concordâncias e o levantamento estatístico dos dados do *corpus*.

Sobre a complexidade da transcrição e suas características, nos apoiamos em nossos referenciais, os quais consideram que a transcrição não é uma operação mecânica, mas "uma verdadeira reconstituição perceptiva das condições de produção, pois não é empreendida

---

que encontramos o volume original da *Cambrige Enciclopedia of Language*, de Crystal, porém a edição de 1997 não contém o mesmo conteúdo.

durante a situação comunicativa e a regulação intersubjetiva de seus participantes" (BRUM-DE-PAULA; ESPINAR, 2002, p. 11).

Decorre dessa reconstituição perceptiva das falas a importância metodológica do trabalho a ser empreendido quando se decide compilar um *corpus* oral. Reconhecemos, porém, que a língua escrita não dá conta da complexidade da oralidade, que é rica em pausas, interrupções, "idas e vindas", que possui uma prosódia, uma sonoridade específica de cada falante, etc. Nesse sentido, Brum-de-Paula e Espinar afirmam que, "em relação a outros tipos de produção, o texto oral é abundante, variável e, consequentemente, mais difícil de ser conservado, representado e manipulado" (BRUM-DE-PAULA; ESPINAR, p.10).

Um notável exemplo de *corpus* oral e de rigor metodológico é o *Corpus* C-Oral-Brasil[46], que é o *corpus* da fala espontânea do PTB, representativo da diatopia do estado de Minas Gerais e, em particular, da área urbana de Belo Horizonte. O *C-Oral Brasil* é um *corpus* de referência do PTB falado informal. Esse *corpus* considera a variação diafásica e diastrática, e as interações entre falantes de faixas socioculturais diferentes.

Segundo Raso e Mello (2009), a opção foi pela transcrição de base ortográfica, mas várias adaptações foram introduzidas para capturar alguns fenômenos da fala do PTB, pois os autores buscam investigar o que é próprio da fala e específico de uma dada língua/cultura. Sendo assim, trata-se de um *corpus* com alinhamento fala-escrita (som e transcrição). A equipe de transcritores passou por um treinamento e o projeto possui um sistema de validação[47] das transcrições.

Mello (2012) também reconhece que o tempo de produção de um *corpus* oral é necessariamente longo e que depende de uma equipe bem articulada e treinada. Ela lista algumas características próprias da fala que dificultam as transcrições e detalha as questões metodológicas que envolvem seu trabalho:

> A própria natureza dos dados linguísticos orais dificulta a sua compilação em *corpora*. [...] os dados orais carecem de tratamento muito minucioso e específico na sua captura, após um complexo período de arquitetura e definição do *corpus* que virão a compor. [...] a fala espontânea raramente se dá em forma estritamente monológica. No universo da existência humana, a fala se presta muito mais à interação dialógica e multilógica, o que acarreta a sobreposição de turnos, interferências, falsos começos etc. e, consequentemente, sua maior dificuldade de representação em forma

[46] Para melhores informações consultar: <http://www.c-oral-brasil.org/>. Acesso em: 14/04/2016.
[47] A validação é *stricto sensu* a ação de tornar algo válido. Entendemos por validação o processo de conferência e confirmação dos dados por uma autoridade no assunto. É uma forma de revisar, atestar e legitimar o trabalho realizado. Doravante todas as vezes em que citarmos "Validação" será a partir desta definição que formulamos.

> ortográfica. [...] As decisões relacionadas ao código de transcrição da fala não são triviais e exigem dos pesquisadores um grande esforço metodológico para que se atinjam soluções satisfatórias, que cumpram o papel de explicitar fenômenos característicos da fala, sem contudo dificultar ao extremo a leitura dos dados pelo usuário do *corpus* (MELLO, 2012, p. 38-39).

Deste modo, fica claro o quanto as opções metodológicas e o rigor são importantes para a compilação de um *corpus* oral, que deve sempre manter o foco em seus objetivos, visto que, fazendo das palavras de Mello (2012) as nossas, o rigor metodológico é o maior trunfo da LC, e sem ele não podemos compilar *corpora*. A autora, nos brinda com uma máxima, "um *corpus* vale de acordo com o seu objetivo" (MELLO, 2012, p. 32) e com base em Lüdeling e Kytö, nos lembra que "o valor de um *corpus* só pode ser medido em função de seu sucesso em atender os propósitos para os quais foi criado" (LÜDELING; KYTÖ, 2008, 2009 apud MELLO, 2012, p. 32).

Existem também os *corpora* orais dialetais, que são constituídos de transcrições de fala de variedades não-padrão e/ou de variedades linguísticas que não possuem o *status* histórico de língua autônoma. Um exemplo de *corpus* oral dialetal é o *Corpus* Dialetal da Língua Espanhola, da variedade falada na Argentina. Araújo (2012), além de estabelecer uma relação entre a compilação de um *corpus* oral e a Dialetologia, viabilizou material da língua espanhola de modo que representasse as sete regiões dialetais da Argentina, utilizando, inclusive, material proveniente de entrevistas radiofônicas. O autor considerou as variáveis escolarização, idade, procedência geográfica e gênero dos participantes. Esse *corpus* diatópico da Argentina foi constituído por enunciados da modalidade oral da língua e considerou as variedades hispano-americanas.

Outro exemplo de *subcorpus* oral é o CRPC – *Corpus* de Referência do Português Contemporâneo[48] (2010) é um *corpus* eletrônico do português europeu, do brasileiro e de países do continente africano de língua portuguesa. O *corpus* foi inicialmente pensado para ser um *corpus* equilibrado, mas acabou por se tornar um *corpus* monitor. Ele contém textos da segunda metade do século XIX até 2006, embora a maioria dos textos seja posterior ao ano de 1970. É composto por 309,8 milhões de palavras provenientes de textos escritos e 1,6 milhões de palavras provenientes de transcrições de gravações de registros orais. Acreditamos que trata-se de um conjunto de *corpora*, pois possui *corpora* escritos que abrangem diversos gêneros de

[48] Para mais informações consultar:
<www.clul.ul.pt/sectores/linguistica_de_*corpus*/*corpus*_oral_pf_publicado.zip>. Acesso em: 14/04/2016.

textos escritos: literário, jornalístico, técnico, científico, didático, folhetos, decisões judiciais, sessões parlamentares, etc. E também é constituído por um *corpus* oral que inclui o discurso formal e o informal e diferentes tipos de interação: monólogos, diálogos, conversas, telefonemas, leituras e etc.

3.3.6.9 *Corpora*

Corpora é o plural de *corpus*. Quando temos mais de um conjunto de textos, escritos ou falados, chamamos de *corpora*, que pode ser estudado em sua totalidade ou de maneira mais restrita, com a escolha de um ou outro *corpus*.

No nosso caso, propomos compilar *corpora*, porque são vários conjuntos de textos de áreas e gêneros diferentes, em pomerano, e um conjunto de textos transcritos de entrevistas com fala pomerana, ou seja, um *corpus* oral.

Especificamente no nosso caso, a pretensão é formar uma unidade com todos esses conjuntos que seria diversos *corpora* e um *corpus* oral formando um conjunto de conjuntos, ou ainda, uma coletânea de *corpora*, a fim de aumentar a representatividade de um acervo dialetal, como o que propomos, e, ainda, ser diversificado para contribuir melhor para o resgate do pomerano.

Dentro desta vasta gama de textos que compõem cada *corpus*, pode haver *corpus* que contenha fragmentos, trechos de textos, frases soltas e até palavras soltas. Berber Sardinha (2000) afirma que não é qualquer conteúdo que pode ser considerado um *corpus*, "tais como sentenças soltas" (BERBER SARDINHA, 2000, p. 17-18), pois "a linguagem natural autêntica não é formada de fragmentos desconexos e, portanto, sentenças soltas não seriam representantes da linguagem" (BERBER SARDINHA, 2000, p. 18). Porém, o mesmo autor considera que haveria uma exceção, caso o *corpus* fosse criado exatamente para esse propósito, de ser uma coletânea de frases soltas e extendemos essa exceção ao caso de um conjunto de palavras soltas, criado com esse propósito.

No nosso caso, optamos por considerar frases soltas e até palavras soltas como material aceitável para compilar *corpus*, pois a intenção é coletar o maior número de itens lexicais possíveis para resgatar o léxico pomerano. Também complementamos com a explicação do próprio Berber Sardinha (2000), de que se quisermos constituir um *corpus* geral de uma língua, deve-se fazer uma escolha a mais variada possível e assim, optar por incluir o maior número possível de registros encontrados na língua-alvo. Esta foi a opção que fizemos e que, portanto,

adequa-se aos princípios da LC, do ponto de vista da representatividade e da autenticidade, considerando os objetivos e o contexto específico da "língua-alvo" desta pesquisa.

3.3.6.10 Tipos de *Corpus*

Existem variados tipos de *corpora*, já que a compilação de um *corpus* depende dos objetivos de sua criação. Com base em nossa pesquisa bibliográfica, listamos abaixo alguns tipos mais comuns de *corpora* e seus respectivos conceitos. Dentro dos parâmetros utilizados pela LC, podemos destacar:

a) ***Corpus* geral**: "também conhecido como *corpus* de língua geral, compõe-se de uma coletânea de textos com a finalidade de permitir a pesquisa de uma determinada língua, sem levar em conta distinções genéricas, varietais, dialetais, lexicais, etc" (GONZALEZ, 2007, p. 10)**.**
b) ***Corpus* especializado/*Corpus* de especialidade**: é um *corpus* "desenvolvido para atender às necessidades específicas de um trabalho de pesquisa em particular, de acordo com os objetivos propostos" (GONZALEZ, 2007, p. 10). Geralmente, compreende uma variedade dentro da língua. Segundo Gonzalez (2007), pode ser composto pela linguagem de uma determinada área de interesse como o internetês, ou ainda, a linguagem de um domínio específico, como a linguagem jurídica.
c) ***Corpus* de estudo**: é o *corpus* que compilamos para atender aos objetivos da nossa pesquisa e tendo como língua-alvo aquela que pretendemos estudar.
d) ***Corpus* de referência**: é um *corpus* no mínimo 5 vezes (BERBER SARDINHA, 2004, p.100-102) maior que o *corpus* de estudo, funciona como um parâmetro para o *corpus* de estudo e é carregado para gerar as palavras-chaves em contraste com a lista de palavras do *corpus* de estudo.
e) ***Corpus* de aprendizes**: a partir de nossa leitura de Almeida (2014), inferimos que é um *corpus* elaborado a partir de material produzido por aprendizes de uma língua e permite detectar dificuldades de aprendizagem. Também permite verificar a influência da língua materna no processo de aprendizagem dos aprendizes, percebida por meio do *corpus* de suas produções. O *COrELE – Corpus* Oral de Espanhol Língua Estrangeira (NOVODVORSKI et all, 2014) é um exemplo de *corpus* de aprendizes.
f) ***Subcorpus*:** uma parte de um *corpus*, que "pode ser fixa ou mutável (dinâmica, isto é, flexível durante a análise)" (BERBER SARDINHA, 2004, p. 16).

g) ***Corpora* comparáveis**: são *corpora* compilados a fim de ser comparados entre si, coletados usando o mesmo método de amostragem. "São compostos de dois ou mais *subcorpora*, cada um contendo textos originalmente escritos em uma dada língua ou variante linguística, mas com características semelhantes, tais como tipologia textual, gênero, tamanho, data de publicação, área e subárea temáticas" (TEIXEIRA, 2008, p. 164).

h) ***Corpora* paralelos**: são aqueles compostos por textos originalmente escritos numa língua, acompanhados de suas respectivas traduções para uma ou mais línguas. Os *corpora* paralelos podem ser unidirecionais ou bidirecionais. Para um *corpus* paralelo ser útil é fundamental alinhar os textos originais e suas traduções, "quando as sentenças do original e da(s) traduç(ão/ões) podem ser vistas lado a lado ou com uma linha abaixo da outra" (TEIXEIRA, 2008, p. 164).

i) ***Corpora* bilíngues**: é um conjunto contendo textos em duas línguas diferentes.

j) ***Corpora* multilíngues**: é um conjunto contendo mais de dois *corpora*, no qual cada *subcorpus* é composto em uma língua diferente. O *Comet* (2001) e o *Reuters corpora* (2000) são exemplos de *corpora* multilíngues.

k) ***Corpus* dialetal**: é um *corpus* composto de material coletado a partir de uma variação diatópica ou de uma variedade linguística não-padrão, por exemplo, de um dialeto italiano, alemão, etc.

l) ***Corpus* oral**: é o *corpus* composto de dados provenientes da fala, que pode ser da fala alinhada à escrita, ou somente da fala transcrita. Este tipo de *corpus* está melhor detalhado no tópico 3.3.6.8.

3.4 A complexidade das múltiplas fontes e as esferas de realização linguística presentes nos *corpora*: diamesia, gêneros, tipos, domínios discursivos e suportes

Quando se pretende compilar *corpora* dialetais de uma variedade em risco, é fundamental ampliar o leque de opções para que nenhum dado linguístico seja desprezado e com ele possibilidades de resgate e documentação do léxico. Por isso, consideramos todos os gêneros, tipos, suportes e diamesias como fontes para compilação de *corpora* do pomerano.

Para lidar com toda essa diversidade de formas, origens e conteúdos de dados é preciso buscar fundamentos e critérios que subsidiem a organização dos dados e seu agrupamento de forma que constituam *corpora* aptos a serem estudados.

A primeira separação possível, no caso de um conjunto de *corpora* proveniente de meios diferentes, é a separação por diamesia. A diamesia é "a esfera da variação linguística no que se refere ao meio físico-ambiental e canal através do qual a língua é usada" (BERRUTO, 1993, p.8).

Geralmente, os *corpora* ou textos de um *corpus* são organizados por ordem cronológica ou por grupos de gêneros principais. Uma alternativa seria separar e organizar os dados coletados por diamesia. Ou seja, pela diamesia em que foram coletados originalmente, tendo em vista que, no caso do *corpus* oral, por exemplo, a diamesia original é falada, mas os estudos podem ser realizados a partir das transcrições, no caso de não-alinhamento som-transcrição.

A diamesia pode ser dividida em diamesia escrita e diamesia falada. Porém, nem sempre dividir o conjunto de *corpora* em diamesias escritas e faladas é suficiente, devido a variedade dos materiais coletados. Portanto, acreditamos que diamesia não se confunde com o conceito de suporte, que são variados e que fixam e veiculam a informação linguística.

Utilizamos as contribuições teóricas de Marcuschi (2008) como aporte teórico para trabalhar com a variedade de instâncias comunicativas presentes nos *corpora*, pois, segundo ele, "toda ação linguística real recorre a algum gênero textual e são formas de realizar linguisticamente objetivos específicos em situações sociais particulares" (MARCUSCHI, 2008, p. 154).

A seguir, vamos trazer as definições a respeito das principais formas de realização linguística, abordadas por Marcuschi (2008), que contribuíram para a elaboração dos critérios utilizados na organização e separação dos dados a fim de constituir os *corpora* desta pesquisa.

O gênero textual, por exemplo, designa os textos materializados em situações comunicativas recorrentes. "Os gêneros textuais são os textos que encontramos em nossa vida diária e que apresentam padrões sociocomunicativos característicos, constituindo listagens abertas" (MARCUSCHI, 2008, p.155).

Como exemplos de gêneros textuais, podemos citar sermão, carta, bilhete, reportagem, notícia jornalística, receita culinária, etc. Segundo Marcuschi (2008, p.155), os gêneros são "formas textuais escritas ou orais bastante estáveis, histórica e socialmente situadas". Os gêneros textuais possuem também uma grande heterogeneidade tipológica, mas, consoante Marcuschi (2008), não devemos concebê-los de maneira estanque, "não são estruturas rígidas, pois são corporificados na linguagem como entidades dinâmicas" (MARCUSCHI, 2008, p.156), pois gênero e tipo são formas constitutivas do texto em funcionamento.

O autor apresenta ainda três tipos de gêneros: gêneros autorais, que são os textos que mantêm um caráter de autoria pelos traços de estilo, caráter pessoal e se situam em especial na

literatura, política, etc.; gêneros rotineiros, que são os comuns de nosso dia-a-dia, por exemplo, entrevistas radiofônicas, jornalísticas, etc; e gêneros conversacionais, que são "são os gêneros de menor estabilidade e sem uma organização temática previsível como as conversações" (MARCUSCHI, 2008, p. 160).

O tipo textual designa uma "espécie de construção teórica definida pela natureza linguística de sua composição (aspectos lexicais, sintáticos, tempos verbais, relações lógicas, estilo)" (MARCUSCHI, 2008, p. 154). Segundo o autor, os tipos textuais abrangem categorias como narração, argumentação, exposição, descrição e injunção e "o conjunto de categorias para designar tipos textuais é limitado" (MARCUSCHI, 2008, p.154).

Além das diamesias, gêneros e tipos textuais, também há o domínio discursivo, definido por Marcuschi (2008, p. 155) como "uma esfera da atividade humana no sentido *bakhtiniano*", que indica as instâncias discursivas. Segundo o autor, "não abrange um gênero em particular, mas dá origem a vários deles" (MARCUSCHI, 2008, p. 155). Como exemplo de domínio discursivo temos o discurso jurídico, o discurso religioso, etc. O autor lembra que os gêneros são entidades dinâmicas, históricas, sociais, situadas, comunicativas e ligados a domínios discursivos.

Quanto à questão do suporte de gêneros textuais, Marcuschi (2008) define como:

> [...] um lócus físico ou virtual com formato específico que serve de base ou ambiente de fixação do gênero materializado como texto. Pode-se dizer que suporte de um gênero é uma superfície física em formato específico que suporta, fixa e mostra um texto. Essa ideia comporta três aspectos: o suporte é um lugar (físico ou virtual); b) suporte tem formato específico e c) suporte serve para fixar e mostrar o texto (MARCUSCHI, 2008, p. 174-175).

Deste modo, entendemos que um jornal, um livro ou uma revista são suportes, pois são portadores de gêneros, de modos de manifestação e de difusão (enunciados orais, papel, etc.) dos gêneros. Quando o suporte se refere a um meio para transporte da mensagem é possível confundir com o conceito de diamesia mencionado anteriormente, porém não são exatamente o mesmo conceito, pois Marcuschi (2008) explica que além de ser o meio, o suporte é também o modo de circulação, fixação e de consumo dos gêneros. O suporte também contribui para identificar o gênero, pois uma informação transmitida por meios de fixação e transmissão diferentes altera o gênero de um bilhete para um telegrama, um email, etc.

Portanto, concluímos com a síntese de Marcuschi (2008) de que "todos os textos realizam algum gênero e que todos os gêneros comportam uma ou mais sequências tipológicas e são produzidos em algum domínio discursivo que, por sua vez, se acha dentro de uma

formação discursiva" (MARCUSCHI, 2008, p.177), sendo que os textos sempre se fixam em algum suporte pelo qual chegam à sociedade.

Para identificar os gêneros presentes nos diversos *corpora* e também os tipos textuais predominantes, seria necessária uma análise atenta dos dados compilados, realizada a partir de leituras completas de cada texto. Portanto, se optássemos em organizar dados compilados a partir de variadas fontes e origens, contendo diversas esferas de realização linguística sob a perspectiva de Marcuschi (2008), essa decisão implicaria em identificar os gêneros e tipos textuais dos conteúdos dos *corpora*, porém, neste trabalho, não nos propomos a realizar essa tarefa, que dependeria de um longo estudo dos tipos textuais e um conhecimento do conteúdo de cada texto dos *corpora*.

Nesse sentido, nossa opção teórico-metodológica é pela organização dos dados dos *corpora* por um misto de categorização dessas realizações linguísticas. Por exemplo, no caso do *corpus* oral transcrito, consideramos necessário separá-lo fisicamente das outras esferas por ter sido coletado originalmente em diamesia falada. No caso dos *corpora* escritos, identificamos, de maneira superficial, os principais gêneros para fazer a separação de cada *corpus*. Assim, temos *corpus* de cartas, de receitas, de músicas etc. No caso do *corpus* das inscrições coletadas nas lápides dos cemitérios, as lápides são o suporte, portanto o suporte é o critério de separação. Já o *corpus* de Internet é um suporte, veículo e serviço ao mesmo tempo.

A organização dos dados compilados será detalhada no capítulo referente à metodologia (Capítulo 4), e a separação concretizada será exposta no capítulo referente aos resultados deste trabalho (Capítulo 5).

Na próxima seção, vamos expor alguns princípios da Sociogeolinguística, um campo interdisciplinar que adotamos como forma de desenvolver o planejamento necessário para a coleta e compilação do nosso *corpus* oral, além de nos subsidiar por seus princípios, enquanto fundamentação de nossa proposta.

3.5 A Sociogeolinguística como fundamentação para uma abordagem-metodologia de compilação de *corpus* oral dialetal

Neste trabalho, adotamos o aporte teórico-metodológico da abordagem interdisciplinar da Sociogeolinguística (doravante SGL). Neste momento, apresentamos o levantamento conciso de conceitos básicos da Geolinguística e da Sociolinguística. Na sequência, definimos a abordagem teórico-metodológica que escolhemos para subsidiar a coleta de dados orais que constituem o nosso *corpus* oral.

A SGL, por sua nomenclatura, já sugere abranger os princípios e variáveis da Sociolinguística e da Geolinguística. Para esclarecer porque escolhemos nos fundamentar na SGL e qual o diferencial que ela possui, vamos primeiramente definir as duas grandes áreas que a compõem.

Adotando a definição objetiva de Cardoso, podemos dizer que a Geolinguística ou Geografia Linguística é um método da Dialetologia que se incumbe de "recolher de forma sistemática o testemunho das diferentes realidades dialetais refletidas nos espaços considerados" (CARDOSO, 2010, p. 46).

Conforme Takano (2013, p.94), a Geolinguística é "uma das áreas que se propõe a investigar o universo linguístico, que retrata a situação das línguas/variedades das regiões, ao estudar a correlação entre os fenômenos linguísticos e o espaço geográfico". Ou seja, é o estudo empírico das variedades linguísticas na relação com as diatopias.

Segundo Cristianini (2012, p.23), a Geolinguística consiste num "método em que há a aplicação de um questionário a um conjunto de sujeitos, numa rede de pontos", em um espaço geográfico definido e permite "visualizar as relações entre o ambiente geográfico e a difusão e distribuição espacial dos fenômenos linguísticos" (CRISTIANINI, 2012, p. 23). Tradicionalmente, realiza estudos cartográficos dos dialetos para "determinar a repetição topográfica dos fenômenos" (CRISTIANINI, 2012, p.23) linguísticos e se há uma norma diatópica em uma dada localidade.

Retornando ao trabalho de Cristianini (2012) lembramos que os estudos geolinguísticos "têm se preocupado com a análise das variações diatópicas levando em consideração determinadas variáveis sociais como idade, gênero e escolaridade" (CRISTIANINI, 2012, p. 21). Porém, ela reconhece que "muitas outras são as variáveis que atuam nos usos, costumes e fenômenos linguísticos" (CRISTIANINI, 2012, p. 21). Deste modo, a Sociogeolinguística permite a abrangência de uma gama mais ampla de variáveis.

Pela relação que a Geolinguística estabelece com os espaços físicos, ela nos faz repensar as definições de rural e urbano. É fundamental definirmos como entendemos e adotamos as concepções de zona urbana e zona rural neste trabalho, visto que nos propomos a coletar dados nesses dois meios físicos de vida e habitação.

Quando nos referirmos a pontos de pesquisa localizados em zona urbana, faremos menção a municípios dentro do distrito urbano, nos quais as pessoas moram em residências uma ao lado da outra, em que há presença de comércios, prefeitura e outros serviços básicos característicos de uma cidade, mesmo que pequena e mesmo que sua economia esteja baseada nos produtos agrícolas produzidos em seu entorno.

Por outro lado, quando fizermos menção a pontos de pesquisa localizados em zona rural, faremos referência a propriedades rurais, sítios, chácaras, fazendas e até mesmo pequenas vilas no interior, entre as propriedades, onde há plantações, campos, criação de animais, rios, represas e vales que cortam as localidades rurais, situadas fora do distrito urbano. Neste estudo, o acesso à tecnologia e aos meios de transporte não são diferenciais do meio urbano e do meio rural, pois, atualmente, mesmo em propriedades agrárias do interior, é possível encontrarmos acesso à internet, estradas e meios de transporte.

Com o objetivo de atender às peculiaridades desta pesquisa, além de adotarmos o método da Geolinguística, também recorremos a elementos da Sociolinguística e reconhecemos, por isso, nossa perspectiva interdisciplinar, pois é o modo pelo qual encontramos subsídios para estudar o objeto em questão, ou melhor, para compilar *corpus*, qual seja, um *corpus* oral dialetal do pomerano no Brasil.

A Sociolinguística é aqui adotada como uma abordagem que considera as relações entre língua e sociedade, não se restringindo às variáveis linguísticas, mas também considerando as variáveis sociais e relacionando ambas. Segundo Von Borstel (2014), é essencial observar o contexto social, cultural, étnico, religioso, político e econômico em uma sociedade ou comunidade de fala. Sobre o surgimento dos estudos sociolinguísticos, a autora informa que:

> Quando se começou a observar os aspectos sociais da língua percebeu-se que a função comunicativa, de interação dos sujeitos, não ocorria de forma homogênea, mas pelo contrário, de maneira heterogênea, tão diferente de pessoa para pessoa, de região, de idade, de escolaridade entre outros fatores sócio-psicológicos e culturais, adotando o modelo de análise da Sociolinguística e da Teoria da Variação Linguística (VON BORSTEL 2014, p. 507-508).

Ambas, a Sociolinguística e a Geolinguística, possuem métodos já consolidados e replicados de coleta de dados orais por meio dos instrumentos que são os questionários. A Sociolinguística utiliza comumente o método da conversação livre/espontânea.

Quando dispomo-nos a realizar uma coleta e compilação de dados dialetais orais, alguns princípios devem ser observados, como, por exemplo, a importância do contexto histórico e geográfico em que vivem os sujeitos falantes alvos do levantamento. A adaptação que fizemos dos questionários atentam para essas variáveis.

Consideramos os fatores históricos, demográficos, culturais, econômicos e sociais para definição dos pontos de coleta e elaboração dos questionários adaptados a partir dos modelos do Atlas Linguístico do Brasil (ALIB, 2014) e de Cristianini (2007).

O Questionário Sociolinguístico (doravante QS) consiste em coletar dados pessoais sobre o sujeito-entrevistado como idade, local de nascimento, e, também informações sobre o modo de vida dos falantes, a sua relação com os meios de comunicação, aspectos culturais e etc. Portanto, o QS deve ser adaptado ao contexto da variante ou dialeto alvo. O Questionário Semântico Lexical (doravante QSL) consiste num método para entrevista em que as questões são onomasiológicas e objetivam coletar dados do léxico utilizado cotidianamente pelos falantes.

Para compor nossas amostras e constituir nosso *corpus* oral, seguimos os parâmetros da Geolinguística e da Sociolinguística, todavia, dada a peculiaridade desta pesquisa, fizemos algumas adaptações/alterações nos critérios de seleção dos sujeitos da pesquisa (estes critérios estarão detalhados no capítulo referente à metodologia deste trabalho).

Os estudos e debates desenvolvidos pelo grupo de pesquisa GPDG – Grupo de Pesquisa em Dialetologia e Geolinguística –, segundo afirmam Cristianini e Encarnação (2009, p. 91 apud CRISTIANINI, 2012, p.26), permitiram o surgimento do termo SGL, que surgiu em 2004, para designar os estudos geolinguísticos que consideram fatores tanto geográficos quanto sociais para coleta, registro e análise de dados linguísticos. As pesquisadoras e os integrantes do grupo de pesquisa chegaram à conclusão de que muitos trabalhos já vêm utilizando as duas abordagens juntas, ou seja, na prática muitos atlas linguísticos foram produzidos considerando também as variáveis sociais e unindo o método da Sociolinguística ao método da Geolinguística.

O diferencial dos trabalhos que assumem a SGL é que visam retratar não só a variação dos fenômenos nas diferentes regiões topográficas, mas também "retratar a cultura, as crenças, as memórias e o modo de ver o mundo dos integrantes de uma comunidade linguística" (CRISTIANINI, 2012, p. 30). Segundo Cristianini (2012), essas pesquisas "revelam aspectos que se reportam à atualização do léxico num processo de mudança linguística e à compreensão das subjacências presentes a cada designação" (CRISTIANINI, 2012, p. 30).

Sobre as características da SGL, Cristianini (2012, p. 26) esclarece que o "interesse da SGL não se restringe à elaboração de atlas linguísticos [...]", mas "continua-se tendo como ponto de partida o levantamento dos indicadores sociais e o mapeamento histórico-geográfico da área a ser pesquisada" (CRISTIANINI, 2012, p. 26). Ademais, a SGL fornece elementos (caracterização das localidades, critérios de seleção, etc) para a constituição da rede de pontos e da relação dos sujeitos candidatos à entrevista. Nessa perspectiva, a coleta de dados é efetuada pessoalmente com a aplicação do QS e do QSL. Ressalvamos que um ponto fundamental na

concepção da SGL é a consideração de que os resultados obtidos por esse processo de coleta "são registros da interação que ocorre no momento da pesquisa" (CRISTIANINI, 2012, p. 26).

Deste modo, a SGL considera que existem variáveis associadas ao uso da língua que, por sua vez, adaptam-se conforme o contexto social e são carregadas de significado, pois existe uma relação entre as variedades de fala e as identidades culturais e sociais. Com base em sua experiência de campo e no trabalho com a SGL, Cristianini (2012) observa que:

> [...] A língua está relacionada à questão cultural do seu usuário e da comunidade da qual o sujeito faz parte, visto que a língua é o principal elemento de interação social. Os sentidos se dão nessa interação entre os sujeitos, que estão situados num determinado tempo, num dado espaço, e pertencem a um grupo (CRISTIANINI, 2012, p. 21).

Na perspectiva da autora, é possível que as crenças, os costumes e as ideologias interfiram nas escolhas linguísticas dos sujeitos, bem como a autoimagem e a concepção de mundo dos mesmos. Segundo Cristianini (2012, p. 28), "a maneira com que o sujeito se vê na sociedade e a maneira como ele vê o pesquisador também influenciam na sua escolha lexical".

Um ponto fundamental na concepção da SGL é a interação entre sujeito-entrevistado e sujeito-entrevistador[49] (SOARES, 2009; 2012). A SGL preconiza o momento da entrevista como um momento de interação face a face, uma relação estabelecida no momento do diálogo e na interlocução com o sujeito-entrevistado.

Assim, a entrevista interativa pelo método da SGL, mesmo com todas as variáveis que agem sobre o fenômeno da fala, permite um contato pessoal, próximo e, por isso mesmo, natural e espontâneo, com os falantes alvo da pesquisa.

Dessa forma, os questionários são aplicados como um diálogo dirigido, são utilizados como roteiro, pois temos consciência de que não é possível atingir uma objetividade ideal nem excluir totalmente a subjetividade.

Acreditamos que a consideração das variáveis socioculturais propicia um ambiente de maior naturalidade no momento da entrevista interativa, em que a capacidade de coleta de conteúdos lexicais da língua alvo é otimizada. Deste modo, a metodologia da SGL resulta em

---

[49] Escolhemos utilizar os termos sujeito-entrevistado e sujeito-entrevistador, fundamentados em Soares, conforme ela: "cada um desses sujeitos – o entrevistado e o entrevistador – exerce uma função social específica, portanto com papéis discursivos determinados. Logo, o gênero entrevista tal como outros gêneros, abarca características linguísticas e sociais próprias, determinando uma função aos sujeitos que dele participam e interagem [...] escolhemos identificar esses sujeitos como sujeito-entrevistado e sujeito-entrevistador" (SOARES, [2009], 2012, p.56).

uma maneira eficiente de coleta de dados da fala espontânea e, consequentemente, do léxico alvo do empreendimento.

A SGL também faz algumas concessões no que se refere a se reorganizar no momento da entrevista, não sendo obrigatório seguir os roteiros e questionários *stricto sensu*, pois o mais importante é permitir a naturalidade, sem que a entrevista se torne um ritual mecânico, podendo causar um retraimento do sujeito-entrevistado. Por exemplo, quando o sujeito não entende a pergunta formulada, podemos voltar à ela de outro modo, ou dialogar com o sujeito-entrevistado, sem, contudo, induzi-lo às respostas.

Consideramos que a SGL pode trazer contribuições importantes para a compreensão dos fenômenos linguísticos, visto que compreende não só as variáveis diatópica (diferença de área topográfica, geográfica), diageracional (diferença de idade), diastrática (diferença de estrato social) e diasexual (diferença de gênero), tradicionalmente consideradas, como também casta, país de origem, religião, escolaridade, pressuposições cognitivo-culturais, bilinguismo, contato de línguas, variáveis psicológicas, etc. Logo, as novas variáveis que podem surgir devem ser consideradas e avaliadas.

Quando é feita a transcrição e inicia-se o tratamento e a análise dos dados, a abordagem da SGL não se restringe a aspectos quantitativos, mas também abrange os aspectos qualitativos e pode ou não resultar na produção de atlas linguísticos. Os dados também podem ser aproveitados para outras análises, pois foram coletados sob uma base metodológica criteriosa e constituem dados legítimos. Eles constituem dados atuais e são evidências empíricas para o desenvolvimento de estudos descritivos. Aqui se unem as abordagens-metodologias da SGL e da LC, no que se refere a coletar dados orais por meio da primeira e analisar esses dados por meio da segunda.

Neste trabalho, como um todo, fica estabelecida uma relação entre a LC e a SGL, pois ambas são abordagens empíricas e descritivas. Consideramos que ambas oferecem meios eficientes para coleta e compilação de conteúdo lexical, permitem a percepção das variações linguísticas, e permitem resultados produtivos e analisáveis com as ferramentas da LC que opera com grandes quantidades de dados.

Concluímos que a SGL, aqui definida como campo interdisciplinar e como abordagem-metodologia, é fundamentação eficiente para atender aos objetivos propostos no presente trabalho, especificamente no que se refere à coleta de dados orais para a constituição de um *corpus* oral dialetal.

Neste capítulo, abordamos noções teóricas e temas fundamentais em nosso trabalho, tendo como ponto central aqueles que giram em torno da compilação de *corpora*. Iniciamos por

definir o léxico e a ciência que o estuda, bem como abordamos os conceitos de LC, nossas concepções a respeito de sua abordagem e da importância de realizar pesquisas com base empírica. Também apontamos a contribuição da SGL para este trabalho.

Por tudo isso, torna-se oportuno e eficaz compilar dados do pomerano, pois seria inconsistente tentar formular uma prescrição para o pomerano com base apenas nas intuições dos falantes e desconsiderar o uso linguístico real (autêntico) no contexto brasileiro.

Consideramos que nosso estudo em LC, tendo como objeto a variedade brasileira do pomerano, poderá contribuir para a ampliação da sua descrição lexical, tendo em vista que "o léxico de qualquer língua constitui um vasto universo de limites imprecisos e indefinidos" (BIDERMAN, 1978, p.139) e também que "o léxico de cada uma das línguas é tão rico e dinâmico que mesmo o melhor dos linguistas não seria capaz de enumerá-lo" (ZAVAGLIA; WELKER, 2013).

No próximo capítulo, vamos detalhar os procedimentos metodológicos que adotamos e desenvolvemos para alcançar o objetivo de compilação de *corpora* escritos e *corpus* oral do pomerano.

4. DESENVOLVIMENTO: PROCEDIMENTOS METODOLÓGICOS PARA CONSTITUIÇÃO DOS *CORPORA* DO POMERANO

Neste capítulo, tratamos sobre o desenvolvimento efetivo do nosso trabalho como meio de alcance dos nossos objetivos. Por isso, vamos detalhar os procedimentos metodológicos adotados, desenvolvidos e aplicados. Descrevemos o passo-a-passo de todas as etapas metodológicas que embasaram a coleta de dados escritos e orais do pomerano para a compilação e constituição dos nossos *corpora*.

Esclarecemos que encontramos alguns percalços no caminho, que fazem parte do processo de pesquisa, por isso não os omitimos. Esses contratempos foram produtivos, na medida em que permitiram que reagíssemos a eles, produzindo procedimentos que atendessem às nossas demandas, a fim de darmos conta da complexidade linguística e sociocultural da constituição de *corpora* escritos diversos e de um *corpus* oral dialetal.

Nossa etapa metodológica de pesquisa dividiu-se em duas fases principais: a primeira fase foi a coleta de dados em diamesia escrita, o que foi feito pessoalmente, em visitas de campo, por meio da internet e por meio de sessões de fotografias em arquivos e cemitérios. A segunda fase foi a coleta de dados orais por meio da realização de 21 entrevistas em seis municípios de três estados do Brasil.

Para cada uma das duas fases, realizamos um planejamento: o roteiro das atividades, a pesquisa prévia sobre as localidades, a preparação de material, o contato prévio com pessoas das comunidades e a elaboração de um diário de campo, a fim de que ajudasse posteriormente na caracterização das localidades e nesta descrição metodológica.

Nos tópicos seguintes deste capítulo, detalharemos cada uma das etapas referentes às duas fases desta pesquisa.

4.1 Primeira fase: constituição dos *corpora* escritos

Ao iniciar a busca por textos pomeranos, enfrentamos uma dificuldade inicial, pois percebemos que seria difícil encontrá-los e que se o critério de coleta dos textos fosse limitado a apenas uma categoria, nos restringiríamos e encontraríamos pouco material. Por isso, decidimos não nos limitar a gêneros, meios e suportes convencionais, pois inicialmente tínhamos em mente livros antigos e cartas pessoais de pomeranos.

Então, percebemos que deveríamos ampliar o leque de opções e não ser tão restritivos em relação aos meios e formas de escrita. Com isso, passamos a considerar qualquer forma de escrita que tivesse conteúdo lexical pomerano e textos que fossem utilizados pelos pomeranos,

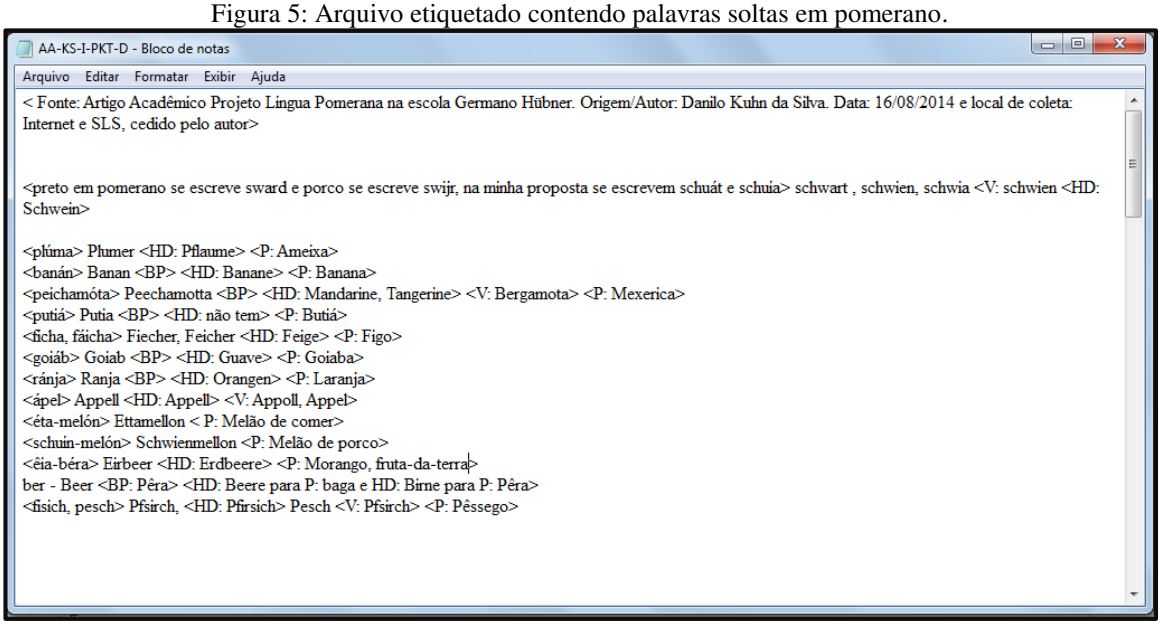
que fossem identificados por eles como textos pomeranos, e/ou, que fossem identificados por nós como textos do baixo-alemão pomerano. A partir disso, passamos a considerar também: formas de escrita não dicionarizadas; escritas transliteradas[50]; material encontrado na internet; bíblias antigas em baixo-alemão, etc.

Também ressalvamos que embora a compilação de *corpora* consista em coletar textos e não palavras soltas, não desprezamos materiais nos quais encontramos apenas algumas palavras soltas em pomerano. Justificamos nossa escolha pela dificuldade de encontrar textos originalmente escritos em pomerano, pelo contexto do pomerano ser o de um dialeto em processo de desaparecimento e pela utilidade destas palavras para compor a *Wordlist*. Reconhecemos, porém, que as palavras soltas não podem gerar linhas de concordâncias e nem ser trabalhadas em contexto, mas contribuem para o resgate e a documentação do léxico pomerano. E lembramos também que são alguns arquivos dos *corpora* que contêm palavras soltas e não todos, nem a maioria.

A seguir, apresentamos um exemplo, na Figura 5, de um arquivo de texto (etiquetado) do *corpus* de artigos acadêmicos, no qual contém palavras soltas (pois consideramos somente as palavras pomeranas presentes no artigo):

[50] Definimos aqui que transcrição refere-se a escrever uma língua seguindo o vínculo existente entre os sons e as formas gráficas já convencionadas para a língua padrão de origem, a língua histórica da qual variou, ou, a língua viva mais próxima da variedade que se quer registrar. Já a transliteração seria o "convencionamento" de formas gráficas que partem da língua do registrante, do proponente da forma escrita, ou da língua majoritária em contato com a variedade minoritária a ser registrada. Na transliteração entendemos que os sons vão ser grafados conforme interpretados pelo ouvinte, com base no código linguístico que ele conhece, com base numa convenção prévia que ele já tem internalizada. Assim, aquele que translitera vincula os sons a determinadas formas gráficas pelas quais fora alfabetizado. Se optarmos pela transcrição, ela deve ser detalhadamente explicada, com fundamentação das escolhas realizadas, e com base na convenção da língua padrão mais próxima da variedade a ser grafada, a qual essa escrita se vincula.

Figura 5: Arquivo etiquetado contendo palavras soltas em pomerano.



Fonte: Elaboração própria.

Contudo, a coleta continuou sendo criteriosa e foi sempre orientada com os princípios da LC em mente, por exemplo, a autenticidade, a originalidade, a produção natural, dentre outros. Nesta fase, foi necessário realizar trabalho de campo para buscar materiais escritos em pomerano com pessoas das comunidades, em arquivos de igrejas, em museus, com outros pesquisadores, fazer cópia de partes de livros em bibliotecas, etc.

De modo geral, o risco do desaparecimento do pomerano, principalmente pelo pouco ou nenhum uso entre as novas gerações (conforme já exposto no capítulo 2), justifica a nossa opção pela coleta de materiais diversificados.

4.1.1 Coleta dos dados escritos - fontes

Coletamos textos escritos em pomerano fotografando descrições de um museu em Santa Maria de Jetibá/ES, fotografando alguns *Wandschoener* e *Huusspruch*[51] encontrados em residências, fotografando lápides de cemitérios de imigrantes, aquelas que se referiam a túmulos de pomeranos e que continham inscrições com indícios de traços dialetais.

Também coletamos cartas, letras de músicas, lembranças, anotações e receitas culinárias doadas por pomeranos em cidades como Canguçu/RS, São Lourenço do Sul/RS e na zona rural

[51] Na cultura pomerana *Wandschoener* são enfeites colocados na parede com inscrições. São feitos, geralmente, de tecido e bordados. Quando uma frase é escrita diretamente na parede chama-se *Wandhuus*, *Wandspruch* ou *Wandmoola*. Os quadros com *Huusspruch* são inscrições com provérbios pintados, geralmente, acima da porta principal das casas, com frases para abençoá-las. Existe também o *Bauernmoolerei* quando são frases pintadas ou entalhadas em tábuas de madeira.

de Itueta/MG. Também fotografamos algumas páginas[52] de bíblias antigas em baixo-alemão pomerano, coletamos matérias jornalísticas e calendários pomeranos na Internet, assinamos o jornal *Folha Pomerana*, e buscamos os mais diversificados meios de obter textos, partes de textos e/ou algum material escrito em pomerano.

Durante uma visita, fomos honrados com a doação de um diário de um sujeito já falecido; sua filha relatou que ele desejava que aquele diário fosse útil para algo. Não conseguimos identificar, com certeza, se ele era descendente de alemães de outra região, ou se da Pomerânia, mesmo porque a família em questão era originária de casamentos mistos (conforme já explicamos no capítulo 2, tópico 2.3.2). Contudo, esse diário é riquíssimo em itens lexicais e em potencial para estudos linguísticos, pois é bilíngue "alemão"-português e mostra o processo de aprendizagem do português. O diário contém indícios de traços dialetais e um alemão não padronizado, além de ter sido escrito utilizando algumas letras de gótico cursivo. Ele data de 1916, ou seja, do início do século XX.

Enfim, com todos os exemplos acima citados e, principalmente, com o exemplo do diário, fica clara a riqueza e a diversidade do material coletado em diamesia escrita.

4.1.2 Processos de compilação: transcrição, digitalização, digitação

A compilação desses materiais foi feita de diversas formas: digitando manualmente enquanto observava os textos de cartas, receitas e congêneres. Depois, a digitação foi conferida atentamente. Esse procedimento foi adotado quando se tratava de material com conteúdo menor.

No caso de livros inteiros em pomerano, os textos foram escaneados em formato PDF e já em via digital foram tratados por um programa leitor de caracteres ópticos (OCR - *Optical Character Recognition*), que transforma a imagem em texto. Em seguida, os textos foram salvos em extensão doc (documento de texto com formatação) e foram revisados, mas esse processo implicou em redigitar algumas partes que sofreram distorção de caracteres, de modo que, nem sempre utilizamos esse método. Algumas vezes consideramos a digitação manual dos livros mais produtiva. Ainda assim, o método de aplicação de OCR contribuiu no processo de compilação.

[52] Fotografamos apenas algumas páginas das bíblias em baixo-alemão porque levaríamos muito tempo para fazer o procedimento com cada página, dependíamos do empréstimo por parte dos proprietários e também porque não teríamos tempo hábil para compilar as bílbias inteiras durante esta pesquisa.

Em outros casos, o conteúdo textual (escrito) foi transcrito a partir de material pictórico como fotografias com conteúdo de textos em pomerano. Nesses casos, as fotografias foram coladas em um arquivo doc para observarmos a imagem enquanto digitamos os textos que continham nela.

Outros textos da internet foram compilados e salvos, primeiramente em extensão doc, depois foram identificados por meio de cabeçalhos e, em seguida, foram salvos em formato txt (*plain text* – texto plano). A esse procedimento chamamos de compilação direta. Para a constituição dos nossos *corpora* escritos, foi utilizada a compilação direta somente nos casos dos materiais encontrados na internet e também dos materiais obtidos em formato PDF copiável, ou seja, a partir dos quais era possível selecionar e copiar seu conteúdo escrito.

Também compilamos textos que estavam no formato digital, mas foi preciso reescrê-los e digitá-los novamente, pois as formas de escrita não eram legíveis pelo programa de LC que adotamos, o WST. Essa foi uma grande dificuldade enfrentada durante nosso trabalho, pois além do fato de não serem legíveis pelo WST, também não estavam padronizados a ponto de permitir as futuras comparações. Por isso, decidimos tratar os dados desenvolvendo um método de tratamento de dados e de normatização que descrevemos no próximo tópico.

4.1.3 Tratamento dos dados escritos

Vamos inicialmente definir o que chamamos de tratamento dos dados. Para nós, tratamento de dados é todo e qualquer trabalho de manuseio e/ou ação com a fonte linguística após sua coleta. O tratamento de dados é uma série de ações necessárias para constituição de *corpora* para estudos em LC. Dentre essas ações, listamos a própria compilação, além da transcrição, da conversão e convenção da escrita (caso necessário, devido às variações) e da análise de dados.

Também consideramos como tratamento dos dados a inserção de cabeçalhos, qualquer modificação no arquivo (sem alterar o conteúdo lexical), organização, salvamento em extensão txt, limpeza, aplicação de *stoplist*[53], etiquetagem e quaisquer anotações de análise e comentários do compilador e/ou analista, dentre outras alterações, a partir da forma exata na qual um texto foi coletado originalmente.

Deste modo, concluímos que o tratamento dos dados é aquilo que faz com que um *corpus* deixe de ser um "*corpus* cru", que são, segundo Novodvorski (2013, p.66), "os textos

[53] Lista para ignonar automaticamente a ocorrência de itens no *corpus*, caso seja útil aos objetivos da pesquisa.

em sua versão original", passando a ser, para nós, um *corpus* tratado, porque os textos dele já não estão mais em sua versão primária, e consideramos que já foram tratados, pois, nesse momento, o *corpus* já está organizado em outra etapa do processo de composição para estudo, sob os preceitos da LC.

Em suma, realizar o tratamento dos dados é deixá-los organizados e em condições de serem legíveis por programas de computador e operáveis pelo linguista de *corpus*. Poderá haver outros tratamentos dos dados dos *corpora* que precisarão ser realizados, dependendo do conteúdo que os componha e também dependendo dos objetivos da pesquisa. No nosso caso, cujo objetivo principal é compor os *corpora* do pomerano para sua descrição lexical, o tratamento da escrita pomerana, denominado por nós de conversão e convenção, foi indispensável, a fim de tornar possível o estudo do pomerano por meio da LC.

Quanto ao tratamento das diversas escritas pomeranas que encontramos nos textos, explicamos detalhadamente, nas subseções (4.1.3.1 e 4.1.3.2) a seguir, em que consistem esses procedimentos metodológicos de tratamento de dados linguísticos, chamados de conversão e convenção, que foi preciso criar e desenvolver neste trabalho, como solução para conseguir obter os *corpora* do pomerano.

4.1.3.1 Conversão da escrita

Chamamos de conversão da escrita a ação de reescrever os textos pomeranos conforme o padrão germânico, sem, contudo, utilizar caracteres escandinavos, principalmente a letra /å/ (lê-se a-*ablaut*), chamada popularmente de "a coroado" em português. Essa letra está presente na língua dinamarquesa, por exemplo. Os textos com a escrita pomerana dicionarizada no Brasil por Tressmann (2006), que utiliza o /å/, não puderam ser legíveis pelo programa WST, conforme os testes que realizamos durante a compilação.

No nosso caso, o programa precisava ser configurado em língua alemã. Mas, como a letra /å/ pertence ao dinamarquês e o WST não configura duas ou mais línguas ao mesmo tempo, notamos que o programa apresentou caracteres distorcidos e que alguns itens lexicais foram perdidos, impedindo a legibilidade dos textos. Por isso, fizemos a conversão de todos os casos de /å/ por "oo", pois o som em pomerano é de um /o/ longo e alto. Além disso, encontramos na literatura pomerana casos de duplo /o/ nas posições em que Tressmann (2006), decidiu grafar como /å/. Além disso, em dinamarquês, conforme Walter (1997, p. 268) essa letra representa originalmente uma substituição de "aa" por /å/, justamente para conferir uma identidade à língua dinamarquesa, visto que a grafia com duplo /a/ era empréstimo do alemão. E, por isso,

concluímos que no pomerano, o som de /å/ não é o mesmo que em dinamarquês. Citamos, por exemplo, o caso de *Hågel* (TRESSMANN, 2006, p.182), ou seja, *Hoogel* (granizo/chuva de pedra), em pomerano, e, *Hagel* em alemão.

Também fizemos a conversão da escrita dicionarizada nos casos de itens que encontramos escritos de outras formas na literatura pomerana, anteriores ao dicionário, visto que a estética gráfica da escrita pomerana dicionarizada no Brasil aparenta uma distância do pomerano em relação ao alemão, maior do que ela realmente é. Esses casos estão convencionados e justificados em um documento auxiliar que criamos, como um passo do procedimento metodológico de convenção (ver Figura 07).

Também fizemos a conversão da escrita no caso de textos pomeranos encontrados em escrita transliterada ou escrita conforme o método de Wiesemann (2008), ou seja, escritas conforme a interpretação gráfica de falantes do pomerano alfabetizados somente em língua portuguesa. Nesse caso, os testes demonstraram que essas escritas eram legíveis pelo WST, porém, decidimos converter a escrita para padronizar somente uma forma de escrita nos *corpora*, pois do contrário não conseguiríamos detectar as repetições nas listas de frequência da *Wordlist*.

Inserimos abaixo um pequeno excerto do nosso trabalho de conversão da escrita, à guisa de exemplificação:

> **a) Fonte original:**
> Deych yexiht fó Noé is pasit zeya fél yóó […]
> Noé vee a gaua meyx, vat léva dee up ayna veylt
> vat ful xléht lüya vee (SOCIEDADE BÍBLICA DO BRASIL - SBB, 2012).
>
> **b) Tratamento da fonte:**
> Dees geschichte vo <von> Noeh is passiert seehr vel joohr […] Noeh wär a <ein> gauer mensch, wat leewa de up einer werlt wat vull schlecht lüür wär (BEILKE, 2014).

O procedimento de conversão reescreve todas as lexias de conteúdo semântico igual, porém com variações na forma de grafia, para uma mesma forma de escrita, padronizada no processo paralelo à conversão, que é a convenção, ou seja, normatização das decisões de escrita. Assunto sobre o qual explicamos na próxima subseção.

4.1.3.2 Convenção da escrita

A convenção é a padronização, dentro do nosso trabalho, das decisões que tomamos em relação à conversão da escrita. Esse procedimento mantém em um mesmo padrão a conversão que fizemos. Na convenção é que justificamos (motivados, inclusive, pela recorrência das

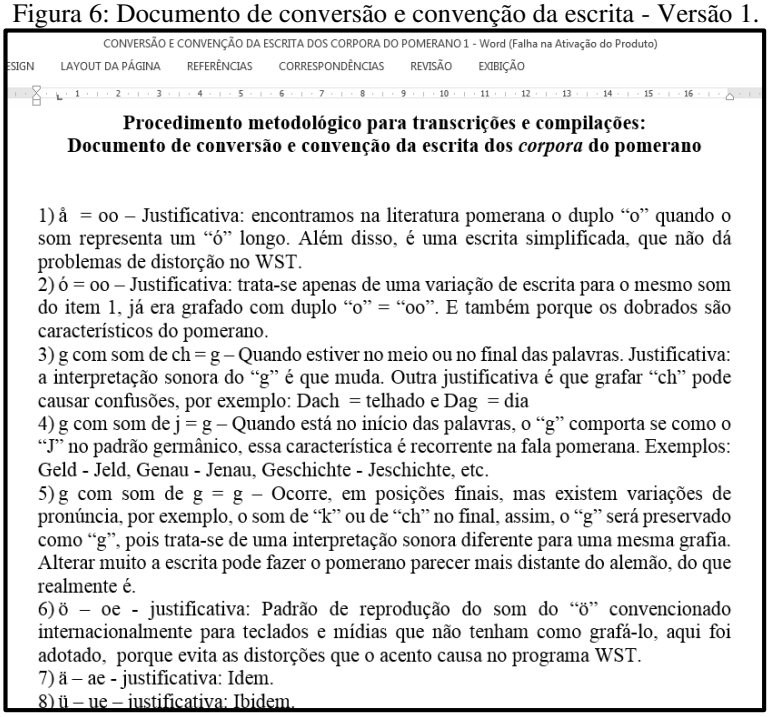
formas de escrita dos *corpora* "crus") a escrita que decidimos adotar. A convenção da escrita é um procedimento metodológico que, além de explicar as escolhas que fizemos e os critérios para fazê-las, dá coerência e uniformidade ao trabalho de compilação dos nossos *corpora*.

Esclarecemos que não criamos uma forma de escrita para o pomerano e nem propomos uma nova forma de escrever o pomerano. Nossa reescrita é apenas um recurso necessário, dentro do contexto da nossa pesquisa, de acordo com os objetivos que definimos nesta dissertação e para conseguir alcançá-los.

Ponderamos também que os signos são arbitrários no que tange à capacidade de representar a fala, que é tão rica e complexa, e que o vínculo "fala-escrita" pode ser problemático, dependendo do ponto de vista. Qualquer escrita sempre registra a fala com algum nível de perda do conteúdo sonoro, toda escrita é uma convenção. Segundo Beilke (2016), "o vínculo que relaciona som/escrita também parece não ter motivação intrínseca e sim convenção dos suportes mecânicos que damos para a nossa memória ao registrar e documentar na escrita tais formas sonoras". A autora lembra também que "o próprio Saussure fala em princípio de hábito coletivo e/ou convenção" (BEILKE, 2016).

A seguir, podemos visualizar um trecho do documento, ilustrado pela Figura 6, no qual constam as decisões que tomamos na execução dos procedimentos de conversão e convenção da escrita pomerana para a compilação dos nossos *corpora*.

Figura 6: Documento de conversão e convenção da escrita - Versão 1.


CONVERSÃO E CONVENÇÃO DA ESCRITA DOS CORPORA DO POMERANO 1 - Word (Falha na Ativação do Produto)

ESIGN LAYOUT DA PÁGINA REFERÊNCIAS CORRESPONDÊNCIAS REVISÃO EXIBIÇÃO

Procedimento metodológico para transcrições e compilações:
Documento de conversão e convenção da escrita dos *corpora* do pomerano

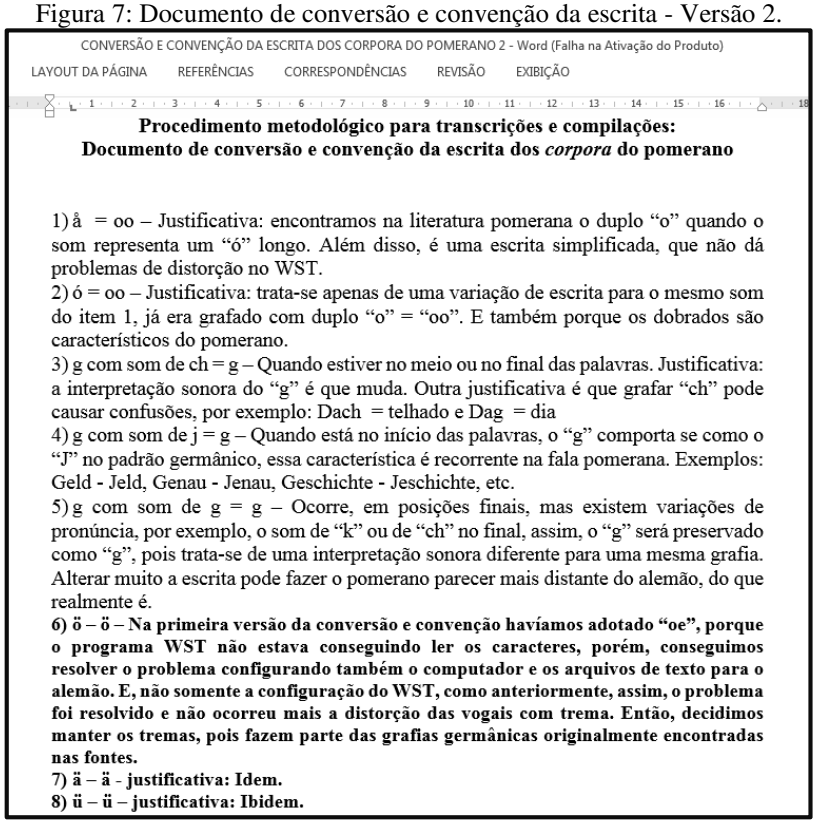
1) å = oo – Justificativa: encontramos na literatura pomerana o duplo "o" quando o som representa um "ó" longo. Além disso, é uma escrita simplificada, que não dá problemas de distorção no WST.
2) ó = oo – Justificativa: trata-se apenas de uma variação de escrita para o mesmo som do item 1, já era grafado com duplo "o" = "oo". E também porque os dobrados são característicos do pomerano.
3) g com som de ch = g – Quando estiver no meio ou no final das palavras. Justificativa: a interpretação sonora do "g" é que muda. Outra justificativa é que grafar "ch" pode causar confusões, por exemplo: Dach = telhado e Dag = dia
4) g com som de j = g – Quando está no início das palavras, o "g" comporta se como o "J" no padrão germânico, essa característica é recorrente na fala pomerana. Exemplos: Geld - Jeld, Genau - Jenau, Geschichte - Jeschichte, etc.
5) g com som de g = g – Ocorre, em posições finais, mas existem variações de pronúncia, por exemplo, o som de "k" ou de "ch" no final, assim, o "g" será preservado como "g", pois trata-se de uma interpretação sonora diferente para uma mesma grafia. Alterar muito a escrita pode fazer o pomerano parecer mais distante do alemão, do que realmente é.
6) ö – oe - justificativa: Padrão de reprodução do som do "ö" convencionado internacionalmente para teclados e mídias que não tenham como grafá-lo, aqui foi adotado, porque evita as distorções que o acento causa no programa WST.
7) ä – ae - justificativa: Idem.
8) ü – ue – justificativa: Ibidem.

Fonte: Elaboração própria.

Na primeira versão da conversão e convenção havíamos adotado escritas que não mantinham os tremas (por exemplo, “oe” para “ö”) porque o programa WST não estava conseguindo ler os caracteres. Mas foi necessário elaborar uma segunda versão da conversão e convenção, porque conseguimos resolver o problema, ao configurar o computador e os arquivos de texto para o alemão, e não somente a configuração do WST, como anteriormente. Assim, não ocorreu mais a distorção das vogais com trema. Então, decidimos mantê-los, pois fazem parte das grafias germânicas originalmente encontradas nas fontes.

A seguir, podemos visualizar um trecho da segunda versão do documento de conversão e convenção da escrita, ilustrado pela Figura 7. O documento na íntegra pode ser visualizado por meio do APÊNDICE C.

Figura 7: Documento de conversão e convenção da escrita - Versão 2.


CONVERSÃO E CONVENÇÃO DA ESCRITA DOS CORPORA DO POMERANO 2 - Word (Falha na Ativação do Produto)

LAYOUT DA PÁGINA REFERÊNCIAS CORRESPONDÊNCIAS REVISÃO EXIBIÇÃO

Procedimento metodológico para transcrições e compilações:
Documento de conversão e convenção da escrita dos *corpora* do pomerano

1) å = oo – Justificativa: encontramos na literatura pomerana o duplo “o” quando o som representa um “ó” longo. Além disso, é uma escrita simplificada, que não dá problemas de distorção no WST.
2) ó = oo – Justificativa: trata-se apenas de uma variação de escrita para o mesmo som do item 1, já era grafado com duplo “o” = “oo”. E também porque os dobrados são característicos do pomerano.
3) g com som de ch = g – Quando estiver no meio ou no final das palavras. Justificativa: a interpretação sonora do “g” é que muda. Outra justificativa é que grafar “ch” pode causar confusões, por exemplo: Dach = telhado e Dag = dia
4) g com som de j = g – Quando está no início das palavras, o “g” comporta se como o “J” no padrão germânico, essa característica é recorrente na fala pomerana. Exemplos: Geld - Jeld, Genau - Jenau, Geschichte - Jeschichte, etc.
5) g com som de g = g – Ocorre, em posições finais, mas existem variações de pronúncia, por exemplo, o som de “k” ou de “ch” no final, assim, o “g” será preservado como “g”, pois trata-se de uma interpretação sonora diferente para uma mesma grafia. Alterar muito a escrita pode fazer o pomerano parecer mais distante do alemão, do que realmente é.
6) ö – ö – Na primeira versão da conversão e convenção havíamos adotado “oe”, porque o programa WST não estava conseguindo ler os caracteres, porém, conseguimos resolver o problema configurando também o computador e os arquivos de texto para o alemão. E, não somente a configuração do WST, como anteriormente, assim, o problema foi resolvido e não ocorreu mais a distorção das vogais com trema. Então, decidimos manter os tremas, pois fazem parte das grafias germânicas originalmente encontradas nas fontes.
7) ä – ä - justificativa: Idem.
8) ü – ü – justificativa: Ibidem.

Fonte: Elaboração própria.

Acima é possível visualizar que os itens de seis a oito foram alterados. Com isso, os arquivos dos *corpora* que já haviam sido convertidos conforme a primeira versão da convenção foram alterados novamente. O procedimento foi passar os textos do txt para doc, utilizar os comandos localizar e substituir e depois salvar novamente em txt.

Os procedimentos acima descritos permitiram a coleta e compilação de dados para a constituição dos nossos *corpora* escritos. Após todas essas etapas de coleta, compilação e tratamento de dados, os *corpora* foram organizados. Tratamos sobre a organização dos *corpora*

no tópico 4.3, visto que essa organização foi realizada juntamente com a organização do *corpus* oral e descrevemos esse passo de uma só vez.

Esclarecemos que salvamos uma versão em doc antes da conversão e convenção das formas de escrita encontradas, a fim de permitir um estudo comparativo das formas de escritas pomeranas originalmente encontradas, e para preservar os dados dialetais originais. Também salvamos uma versão em doc após os procedimentos supracitados, e uma versão totalmente tratada, em txt, além de salvarmos a *Wordlist* dos *corpora* escritos e alguns exemplos de linhas de concordâncias dos nossos testes.

Segue, no Quadro 1, um comparativo de algumas formas de escrita que encontramos nos *corpora* e em obras lexicográficas:

Quadro 1: Quadro comparativo de formas de escritas pomeranas.

COMPARAÇÕES ENTRE FORMAS DE ESCRITA DO POMERANO			
***Corpora* escritos**	**Das Pommersche Wörterbuch, Hermann-Winter (1999)**	**Pomerisch-Portugijsisch Wöirbauk, Tressmann (2006)**	**Projeto Pomerando, Kuhn Silva (2012)**
Haus, Hause, Hus, Huus	Huus	Huus	Hus
Hagel, Hoogel, Hågel, Hoochel	Hågel	Hågel	Hóhal
Wiet, Weiss, Witt	Wiet	Wit	vit
Futéla, votela, fortela, fotela	vertellen	fortela	futéla
Frau, Fruh, Fruu, Fruuh, Fruch, Fruug	Fru	Fruug	Fruch
Kerl, Keirl, Keyrl, Kejrl	Kierl	Keirl	Kêil
Schaffe, Schoppe, Schopa, Schoppa	Schoppe	Schåp	Schop

Fonte: Elaboração própria.

No quadro acima é possível visualizar, à esquerda, as diversas formas de escrita dos mesmos itens lexicais, porém com as variações encontradas nos textos pomeranos e, à direita, também é possível visualizar as diferentes propostas de escrita já publicadas.

A respeito dos procedimentos realizados na segunda fase da pesquisa, da constituição do *corpus* oral dialetal do pomerano, trataremos a seguir.

4.2 Segunda fase: constituição do *corpus* oral

A segunda fase da pesquisa foi aquela em que fomos a campo, em busca de obter dados da fala pomerana, para que pudessem ser transcritos e assim compilarmos um *corpus* com conteúdo coletado originalmente em diamesia falada. Antes de iniciarmos essa fase, o primeiro

passo foi a submissão do projeto ao comitê de ética, o que exige a preparação de um roteiro, delineando a pesquisa para atender às especificações exigidas pela plataforma, na qual se submete os pedidos de autorização. Os parâmetros do comitê nos ajudaram a desenvolver uma metodologia mais detalhada.

O projeto foi submetido ao Comitê de Ética em Pesquisas com seres humanos – CEP da Universidade Federal de Uberlândia/UFU e obteve sucesso na aprovação com resposta favorável, conforme parecer consubstanciado CEP nº.1145346 (ANEXO A).

Além desse procedimento prévio, para a constituição do nosso *corpus* oral e para conseguirmos êxito nas entrevistas, muitos outros procedimentos foram necessários durante a segunda fase da nossa pesquisa, os quais passamos agora a descrever.

4.2.1 Metodologia para coleta dos dados orais

Ao compilarmos um *corpus* oral, devemos definir parâmetros de coleta, compilação e transcrição, de modo que os resultados possam ser aplicados e que os estudos possam ser replicados. O desenho (ou a arquitetura) do *corpus* oral deve estabelecer o perfil dos entrevistados, a quantidade de sujeitos que participarão das entrevistas, os gêneros, faixas etárias, as localidades de coleta, as técnicas para obtenção dos dados e o equipamento a ser utilizado.

Neste momento, ocuparemo-nos do detalhamento dos procedimentos adotados para a coleta e compilação do *corpus* oral. Descreveremos, nos subitens a seguir, o método, os critérios da pesquisa e os passos, trataremos das localidades selecionadas, da caracterização dos pontos de pesquisa e do plano de recrutamento, segundo as variáveis da SGL, dos procedimentos para a realização das entrevistas e dos riscos, do planejamento e preparação dos instrumentos de coleta, da aplicação dos questionários e do método de transcrição dos dados orais. Enfim, seguem-se todos os procedimentos referentes à constituição do *corpus* oral dialetal do pomerano.

4.2.1.1 Coleta dos dados orais - entrevistas

Optamos por coletar dados orais do pomerano por meio da realização de entrevistas diretas e interativas com falantes de pomerano. Os falantes de pomerano são os descendentes dos imigrantes que vieram para o Brasil, há mais de um século, e que vivem nas localidades alvo desta pesquisa. O método consistiu em realizar entrevistas face-a-face com desenvolvimento de diálogos entre sujeito-entrevistado e sujeito-entrevistador, a partir de

questões previamente formuladas. Durante as entrevistas, gravamos os depoimentos orais, para que, posteriormente, pudéssemos escutá-los e transcrevê-los.

4.2.2 Da seleção das localidades e suas caracterizações segundo a SGL

Conforme já discutido no Capítulo 3, a SGL considera os aspectos geográficos, culturais e históricos como relevantes nos fenômenos linguísticos e considera um amplo espectro de variáveis, não somente as variáveis diatópica, diasexual e diageracional, mas também as variáveis situacionais, psicológicas e psicossociais, como a constituição identitária, a autodeterminação, etc. Neste tópico, descrevemos as variáveis básicas; as características geográficas, econômicas e históricas das localidades.

Foram inicialmente selecionadas quatro localidades para constituírem os pontos de pesquisa, sendo duas no estado de Minas Gerais e duas no estado do Rio Grande do Sul, a saber, Vila Neitzel/MG, Itueta/MG, Vera Cruz/RS e Arroio do Tigre/RS. Porém, posteriormente foi necessária uma ampliação dos pontos de pesquisa, para viabilizar a coleta, e para contornar algumas contingências com as quais nos deparamos durante o processo, como o fato de não termos encontrado falantes disponíveis para entrevistas em algumas localidades (Vera Cruz/RS, zona rural e urbana). Contudo, mantivemos o foco nessas localidades, pois lembramos que é um dos nossos dois objetivos específicos, o segundo deles, conforme o item 1.4.2: realizar uma breve comparação entre Minas e Sul, por meio dos dados coletados nessas regiões.

As localidades incluídas na ampliação dos pontos de pesquisa foram São Lourenço do Sul e Canguçu, ambas no Rio Grande do Sul, pois, ao coletar dados escritos nessas localidades, tomamos conhecimento de que havia um considerável número de falantes de pomerano e precisávamos substituir o município de Vera Cruz (pontos rural e urbano), da região do Vale do Rio Pardo, onde não encontramos falantes.

Também escolhemos Santa Maria de Jetibá/ES como um ponto de pesquisa extra, pois é a localidade popularmente referida como a capital pomerana no Brasil e tem relevância no cenário nacional de pesquisas sobre o pomerano que foram realizadas nas áreas da Educação, da Linguística e da Antropologia, conforme citamos no capítulo 2. Lembramos que foi em Santa Maria de Jetibá onde surgiu o dicionário pomerano-português. Então, decidimos fazer uma entrevista na localidade, a fim de obtermos um parâmetro de comparação com as regiões que são nosso escopo de pesquisa, Minas e Sul, e também para verificarmos se há interferência do alemão-padrão no pomerano falado no município. Nessa perspectiva, embora nosso foco seja as localidades de Minas e Sul, temos o intuito de verificar a presença do pomerano em Santa

Maria de Jetibá, já que é referência quando o assunto é o pomerano. Vale destacar que o acréscimo de dados provenientes do referido município na constituição dos nossos *corpora* é enriquecedor, na medida em que possibilita o desenvolvimento futuro de estudos comparativos, entre o pomerano falado na localidade e em outras regiões do Brasil.

A seguir, faremos a descrição das características principais dos pontos de pesquisa que abrangemos. Esse procedimento faz parte do método da SGL, como já explicamos no início deste tópico. Esclarecemos que as informações sobre os municípios foram extraídas dos nossos diários de campo, produzidos nas visitas exploratórias, com base nos dados fornecidos pelas prefeituras municipais e também com base nos dados fornecidos a nós, em consulta ao IBGE (Instituto Brasileiro de Geografia e Estatística, doravante IBGE), pela Lei de acesso à informação.

4.2.2.1 Minas Gerais

No estado de Minas Gerais, escolhemos duas localidades para a coleta de dados, por isso caracterizamos os dois pontos de pesquisa do estado em uma única seção, tendo em vista que Vila Neitzel é circunscrita a Itueta, ou seja, faz parte da jurisdição e do território (na zona rural norte) do referido município, localizado no leste do estado. Há outros municípios mineiros com presença de pomeranos, conforme apontamos no capítulo 2, porém não foi possível abranger um número maior de pontos de coleta no contexto desta pesquisa. A seguir, caracterizamos Vila Neitzel e Itueta, onde realizamos coleta de dados orais e escritos em pomerano.

4.2.2.1.1 Vila Neitzel e Itueta

Em Minas Gerais, selecionamos Vila Neitzel, localizada na zona rural norte de Itueta, situada ao leste do estado. Embora nomeada como Vila, possui o status oficial de povoado e recebeu este nome por causa da família Neitzel, que foi uma das primeiras famílias a chegar na região, cedendo, posteriormente, parte de suas terras para o povoado. Vila Neitzel é separada de Itueta pelas águas do Rio Doce. São aproximadamente 20 quilômetros de estrada de terra, mas a travessia também é realizada por balsa[54].

[54] Segundo informantes locais, a travessia pelo Rio Doce para a zona rural norte de Itueta/MG é feita por balsa, pois é o meio de transporte mais fácil, mesmo após o rompimento da barragem de Mariana/MG que afetou o Rio Doce e todo o leste de Minas. Devido à distância e à precariedade das estradas, os moradores utilizam a balsa como meio de transporte.

A localidade de Vila Neitzel e toda a região situada ao longo do Vale do Rio Doce está próxima à fronteira com Baixo Guandu/ES. Segundo Pessoa (1995, p. 79), "alguns pomeranos se deslocaram para Minas Gerais, especialmente para Santo Antonio, Itueta, Aimorés e Resplendor, região próxima a Baixo Gandu", em busca de terras mais acessíveis para o trabalho na agricultura.

A topografia é própria de um vale, a região é rodeada por montanhas rochosas. Em Vila Neitzel, predomina o modo de subsistência agrícola e a vida colonial. A localidade possui características rurais. É uma comunidade formada por pequenos produtores, as lavouras de café constituem a principal economia de subsistência, mas também plantam inhame, batata e milho. Lembramos que já apresentamos nossa definição de rural e urbano no capítulo 3.

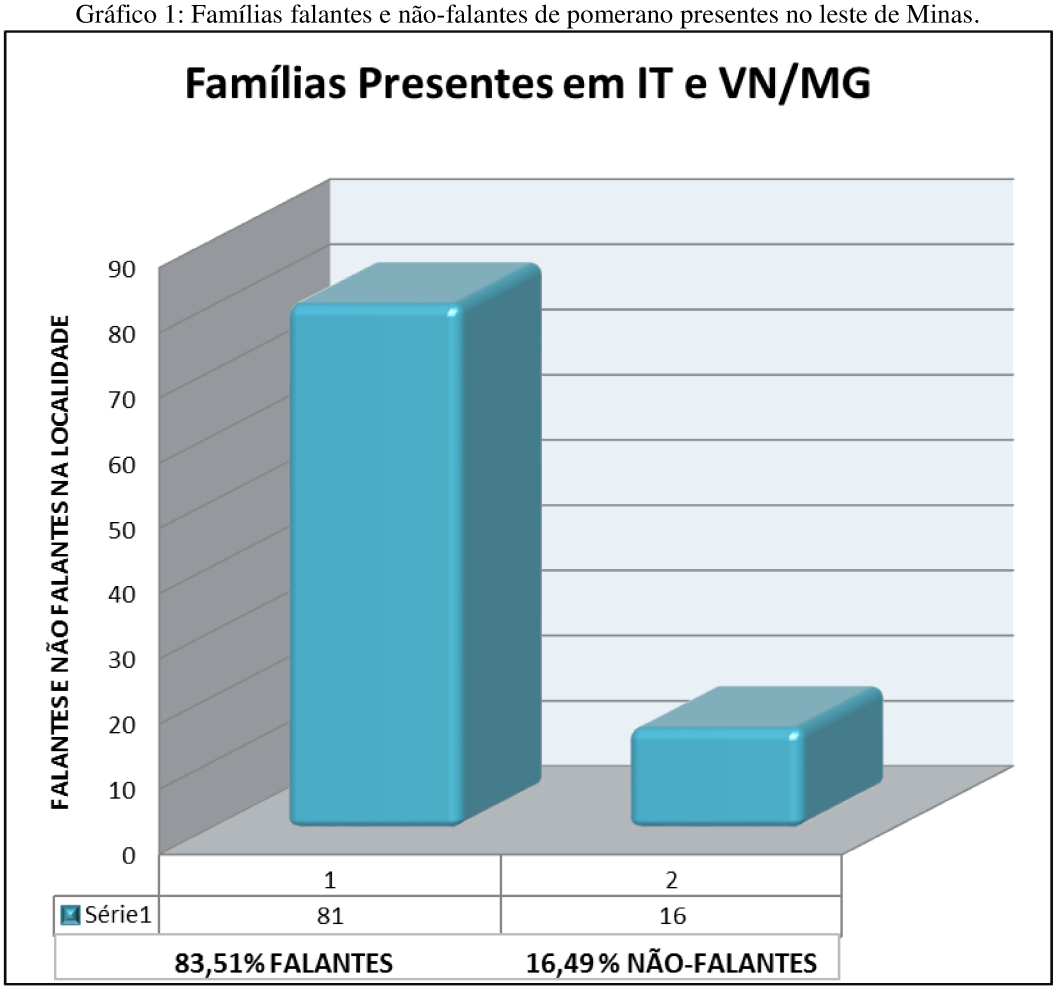
Antunes (2011) informa que na região havia cerca de 2 mil descendentes de pomeranos, que iniciaram a ocupação da região entre 1915 e 1920. Segundo o autor, na década de 1950 houve uma nova onda de migração, elevando o número de pomeranos para quase 4 mil, considerando toda a zona rural norte de Itueta.

Realizamos um levantamento e cruzamos algumas informações sobre os pomeranos dessa região a fim de verificar os dados acima mencionados. Após algumas pesquisas e viagens ao campo, verificamos que existem descendentes de pomeranos presentes na região leste de Minas Gerais, tanto em Vila Neitzel quanto no restante da zona rural norte de Itueta e também na zona urbana de Itueta. Estimava-se em torno de 2.000 pomeranos presentes nessa região, mas, devido à evasão populacional nas décadas de 70 e de 90 para Rondônia, em busca de terras para cultivo e também à evasão para os Estados Unidos, em busca de melhores condições de trabalho (PESSOA, 1995), estimamos que atualmente apenas 50% de pomeranos (aproximadamente) ainda estão nas localidades mineiras, ou seja, cerca de 1.000 pomeranos. Chegamos a essa estimativa com base na análise dos dados de Pessoa (1995), Antunes (2011), Beilke (2013) e nas informações obtidas em nossos contatos com o ex-prefeito de Itueta, além do nosso censo prévio que verificou uma amostragem de quase 10% da quantidade de pomeranos referidos no cálculo sobre os 1.000 descendentes de pomeranos (mencionados por PESSOA, 1995 e ANTUNES 2011) que permaneceram nas localidades do leste mineiro, visto que verificamos algumas dessas famílias pessoalmente em nossas visitas de campo.

No intuito de esclarecer essa questão, organizamos e checamos algumas listas de famílias da localidade. Realizamos um censo informal datado de 2014, sobre 140 famílias (baseado em uma listagem de 180 famílias), a partir do qual elaboramos dois gráficos que apontam a porcentagem de falantes de pomerano na região (Gráfico 1) e fazem uma estimativa sobre os pomeranos que emigraram da região (Gráfico 2).

A relação numérica, em porcentagem, de famílias falantes de pomerano e não-falantes de pomerano, na região leste de Minas Gerais, especificamente no entorno do Vale do Rio Doce e de Itueta, incluindo o município e toda a zona rural norte, pode ser visualizada por meio do Gráfico 1, que elaboramos e inserimos a seguir:

Gráfico 1: Famílias falantes e não-falantes de pomerano presentes no leste de Minas.



Fonte: Elaboração própria.

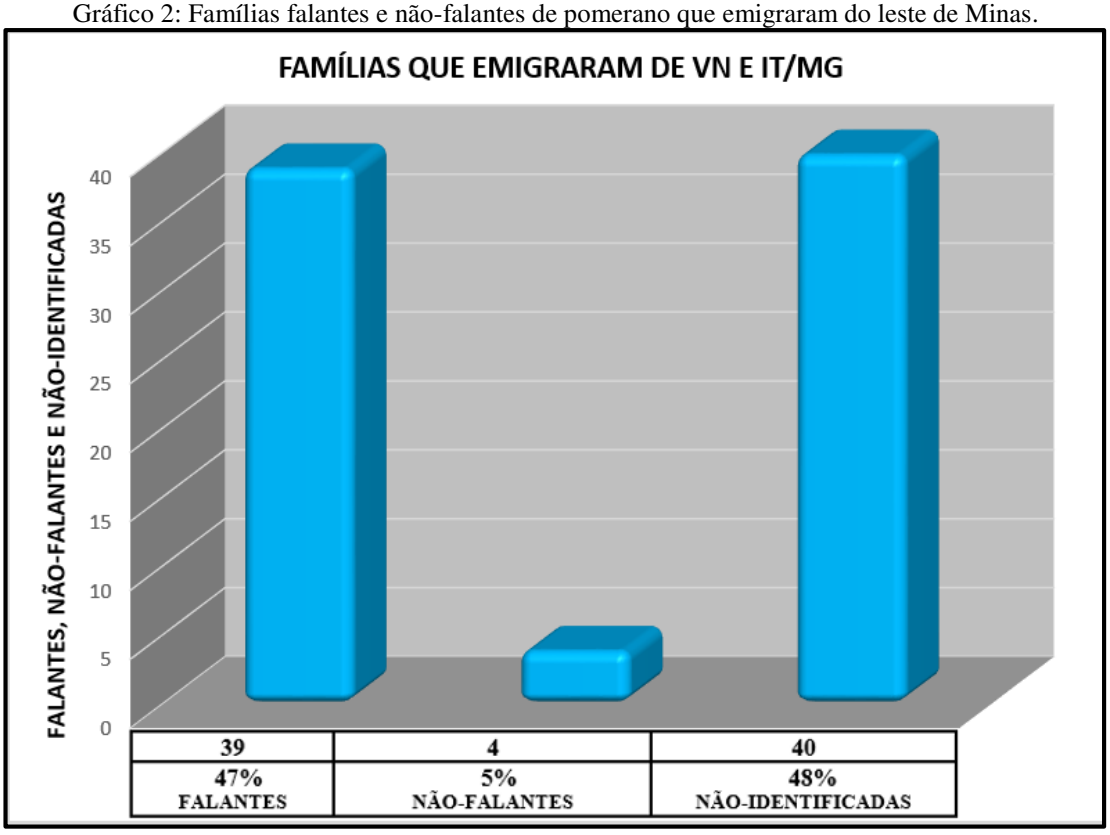
O gráfico acima indica que de um total de 97 famílias residentes nas localidades mineiras pesquisadas (Itueta e Vila Neitzel), 16 famílias não possuíam mais falantes de pomerano, porém 83,51% das famílias que conseguimos abranger, ainda falavam o pomerano até a data do levantamento, em 2014. Esse cálculo foi baseado em um levantamento anterior realizado por um morador da região, pesquisador autônomo[55], que começou a fazer o levantamento das famílias pomeranas – com as quais ele tinha contato direto e sobre as quais ele tinha conhecimento – em listagens organizadas por escrito. Quando estivemos em contato

[55]Não informamos os nomes do ex-prefeito e do pesquisador autônomo neste trabalho, pois eles foram circunstantes em algumas entrevistas realizadas para coleta de dados orais. Conforme as determinações do CEP, não podem ser informados dados que possam trazer risco de identificação dos sujeitos-participantes da pesquisa.

com ele, as listagens nos foram cedidas para continuarmos o trabalho, e com a ajuda do referido pesquisador e também do ex-prefeito de Itueta, terminamos de organizar as listagens; visitamos o máximo de famílias que conseguimos, conferimos a situação das mesmas e fizemos os cálculos em relação às famílias abordadas por nosso censo prévio informal.

Ao realizar nosso pequeno censo informal, pretendíamos saber quais famílias pomeranas ainda residiam em Itueta, seja na zona urbana ou rural, que ainda falassem o pomerano e as famílias que não o falassem mais. Mas o objetivo desse levantamento também foi, além de confirmar os dados de Pessoa (1995) e Antunes (2011), saber quais famílias "pomeranas-mineiras" haviam migrado para Rondônia e se na época da migração, falavam ou não o pomerano. Por isso, também elaboramos um gráfico sobre as famílias pesquisadas que emigraram da região, como podemos verificar no Gráfico 2, a seguir:

Gráfico 2: Famílias falantes e não-falantes de pomerano que emigraram do leste de Minas.



Fonte: Elaboração própria.

O censo informal mencionado aponta que o total de famílias que emigraram do leste de Minas Gerais foi 83. Elas foram identificadas em nossos dados por seus sobrenomes pomeranos. Conseguimos constatar apenas que 5% das famílias que deixaram a região já não preservavam mais o uso do pomerano quando saíram de lá, pois os falantes da região e nossos informantes, que estavam em contato com as famílias em questão, relataram que os

componentes delas só falavam português, e não mais o pomerano. Não conseguimos verificar a situação (falante ou não-falante de pomerano) de um grande número de famílias (40 famílias do total de 83), as quais constam no Gráfico 2 como "não-identificadas. Em relação a isso, a única afirmação que podemos fazer com segurança é a de que todas são famílias pomeranas e ex-residentes da região.

Uma de nossas fontes para a coleta desses dados foi o ex-prefeito de Itueta, informante principal na região, além de ser nosso contato-chave no leste de Minas Gerais. Além do levantamento que ele e o pesquisador autônomo da região começaram, podemos ressalvar que o informante em questão é profundo conhecedor da região, pois sempre viveu em Itueta. Ele foi um colaborador para a obtenção de informações, conhecia e tinha contato com as famílias que permaneceram e que emigraram, coletou dados sobre os pomeranos, tentou organizar projetos de resgate da cultura pomerana e possui uma notável memória. Esse informante é pomerano, falante de pomerano e alemão e nos acompanhou em visitas a muitas famílias pomeranas para a coleta e checagem dos dados.

Os dados do nosso censo informal estão mais bem expostos em tabelas que identificam as famílias por seus sobrenomes pomeranos: são as listagens das famílias pomeranas presentes na região, bem como das famílias que emigraram da região. As duas listagens estão disponíveis no final deste trabalho, como APÊNDICES A e B.

Realizamos uma consulta para obter um levantamento oficial da densidade demográfica em Vila Neitzel, por meio da lei de acesso à informação, solicitando os dados ao IBGE, porém, conforme mostra a Tabela 1, a seguir, só conseguimos obter dados do município de Itueta e de seu distrito, chamado Quatituba.

Tabela 1: Resposta da consulta ao IBGE sobre dados oficiais de Vila Neitzel/MG

DADOS ORGANIZADOS DA RESPOSTA DO IBGE				
Situação do domicílio (população) residente, conforme o censo demográfico de 2010				
Código IBGE	**Município e Distrito**	**Total**	**Urbana**	**Rural**
3134103	Itueta - MG	5.830	3.299	2.531
313410305	Itueta - Itueta - MG	3.226	1.324	1.902
313410310	Quatituba - Itueta – MG	2.604	1.975	629
Observação: Dados obtidos através da Lei de acesso à informação. Consulta realizada em: 14/06/2014.				

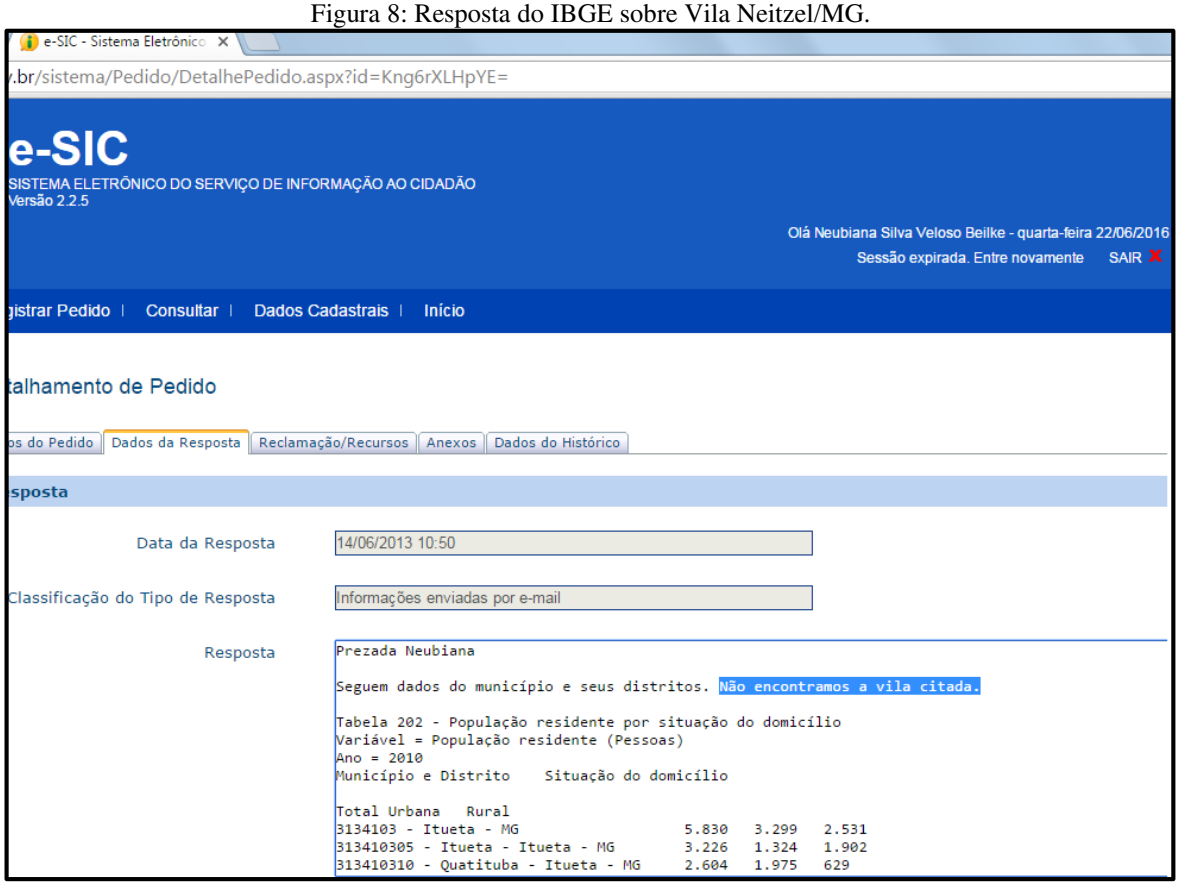
Fonte: Elaboração própria.

Com base na Tabela 1, percebemos que o município de Itueta possui baixo índice populacional. Na zona rural de Itueta, observamos que há uma expressiva quantidade de habitantes. Os dados informados pelo IBGE e organizados no Tabela 1, acima, considera a

população do município de Itueta em sua totalidade, depois somente o distrito sede, e por último o distrito Quatituba.

Os dados populacionais formais do povoado de Vila Neitzel não foram informados pelo IBGE, que afirmou, em resposta à nossa solicitação, que não localizou a vila em seus registros, conforme comprova a Figura 8, a seguir:

Figura 8: Resposta do IBGE sobre Vila Neitzel/MG.



Fonte: Elaboração própria.

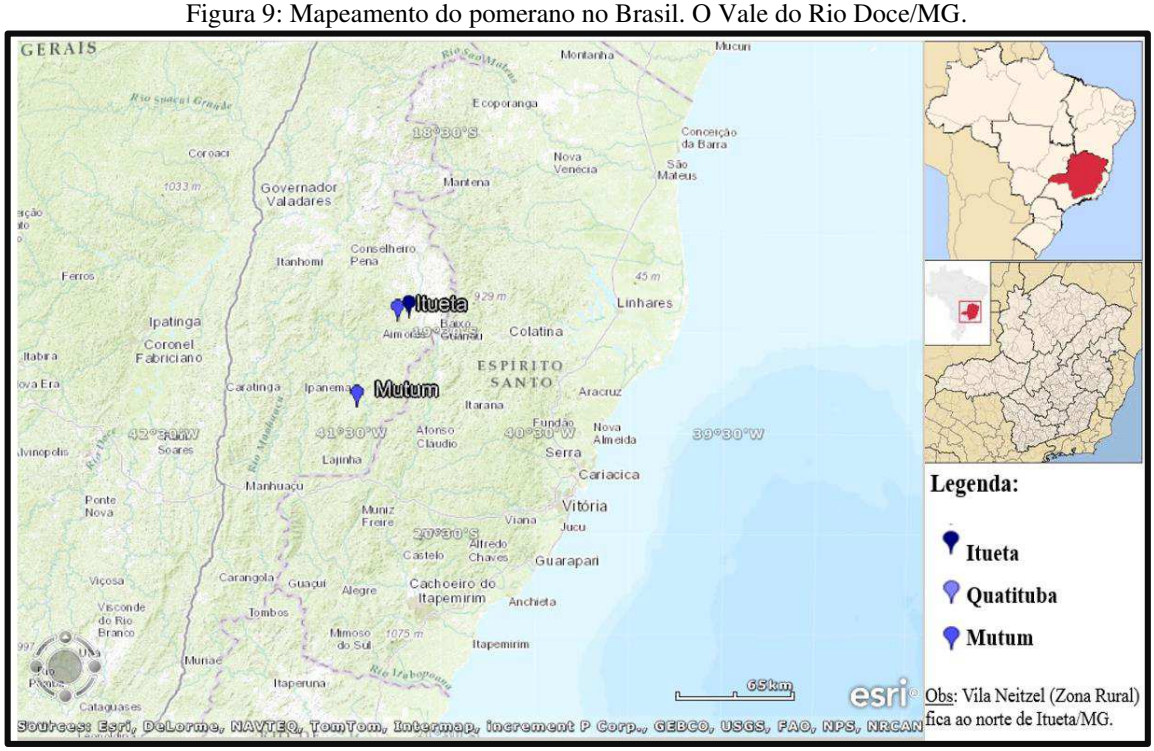
Não fomos satisfatoriamente informados pelo IBGE, se a população de Vila Neitzel está incluída na contagem da população rural de Itueta, mas acreditamos ser possível que tenham sido contabilizados quando da visita dos recenseadores à zona rural do município, pois Vila Neitzel está geograficamente situada dentro da zona rural da localidade e seria difícil não percebê-la.

Quanto ao distrito de Quatituba, mencionado na tabela acima, recebemos informações no campo, de que havia falantes de pomerano na localidade, na qual realizaríamos uma entrevista, mas foi preciso cancelar por falta de tempo para abranger mais uma localidade.

A seguir, expomos um mapeamento (Figura 9) da região leste de Minas Gerais, mais especificamente da região do Vale do Rio Doce, que elaboramos por meio do *ArcGis*, a fim de

evidenciar alguns municípios com presença de população pomerana. Dentre as localidades, damos destaque a Itueta e Quatituba, conforme a legenda. Vila Neitzel não pôde ser localizada pelo programa de geoprocessamento utilizado, mas o povoado está situado ao norte do município de Itueta.

Figura 9: Mapeamento do pomerano no Brasil. O Vale do Rio Doce/MG.



Fonte: Beilke (2013). ArcGis.

No mapa acima é possível ver a representação de Mutum, localidade onde algumas famílias pomeranas estabeleceram-se. Encontramos reportagens *online* sobre a presença de famílias pomeranas em Mutum, migrantes de Itueta e região. Segundo Antunes (2011), em Minas Gerais, além de Itueta e Vila Neitzel, também há falantes de pomerano em Resplendor e Aimorés, no entanto, não foi possível checarmos pessoalmente a informação do autor, tampouco a da internet sobre Mutum.

Como Vila Neitzel não pôde ser visualizada no mapa, só pudemos identificá-la através de fotos que marcam a direção da localidade rural. Visitamos uma escola no interior de Vila Neitzel e constatamos que algumas crianças ainda falam algumas poucas palavras em pomerano e relatam que os seus pais falam o pomerano em casa. Porém, não entrevistamos crianças em nossa pesquisa, pois não solicitamos essa permissão ao CEP. Na escola visitada, Escola Municipal Barra do Joazeiro, não há nenhuma forma de ensino do pomerano, conforme relatos dos professores, por falta de incentivos, pois não possuem material didático e nem professor.

A seguir, podem ser visualizadas, por meio da Figura 10, fotografias da referida escola e das placas de identificação de Vila Neitzel:

Figura 10: Escola no interior de Vila Neitzel e placas de identificação do povoado rural.



Fonte: Arquivo do autor. Beilke (2014).

Acontece anualmente em Vila Neitzel, desde 2006, um encontro de tocadores de concertina (Encontro da Cultura Pomerana e Festival de Concertina de Itueta-MG, que até o ano de 2013 – 8ª edição – era realizado na fazenda Pitelkow) que é divulgado como forma de resgate da cultura pomerana no local. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla VN-MG.

Sobre Itueta, o segundo ponto de pesquisa selecionado em Minas Gerais, vamos tecer mais algumas observações a seguir.

Itueta, que já caracterizamos em parte, junto à Vila Neitzel, também está localizada no leste de Minas Gerais, no Vale do Rio Doce. O município fez parte de Resplendor e, em 1948, foi emancipado. Na verdade, Itueta, nosso ponto de pesquisa, é chamada de Nova Itueta, pois foi realocada, na década de 1990, devido à inundação que sofreu para a construção da represa do Rio Doce. Optamos por chamar a cidade apenas de Itueta. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla IT-MG.

Acreditamos que muito material documental dos pomeranos deve ter sido deixado para trás e/ou perdido quando da transposição da cidade, pois, durante as entrevistas, os pomeranos de Itueta relataram que havia grande quantidade de pessoas "antigas" (como eles disseram) descendentes de pomeranos na Velha Itueta, e manifestaram um grande saudosismo em relação à antiga cidade.

Em Itueta foi elaborado um projeto desenvolvido pela ONG Rede Vidas, com o objetivo de incentivar o pomerano entre os mais jovens. O projeto foi chamado de "Língua Mutter: O Resgate da Cultura Pomerana". Porém, quando visitamos a localidade, percebemos que tal proposta ainda não tinha sido colocada em prática[56].

Acreditamos que testes mais específicos precisam ser feitos com as famílias antes de determinar o nível de pomerano que preservam ou não, se são falantes ativos, passivos, etc. Temos conhecimento de que existem muitos descendentes de pomeranos que compreendem em pomerano, porém costumam responder e falar em português. Reconhecemos, porém, que esses dados carecem ainda de validação científica. Mas fizemos o levantamento com o objetivo de começar a identificar a situação linguística na localidade, caracterizar as diatopias e possibilitar futuras averiguações a respeito.

4.2.2.2 Rio Grande do Sul

Passamos agora a caracterizar os pontos de pesquisa selecionados no Vale do Rio Pardo e fora dele, na Serra dos Tapes, todos no Sul do Brasil, no estado do Rio Grande do Sul. São eles: Vera Cruz, Arroio do Tigre, São Lourenço do Sul e Canguçu, escolhidos como pontos para coleta de dados orais por meio das entrevistas.

4.2.2.2.1 Vera Cruz

O município de Vera Cruz/RS está localizado na região do Vale do Rio Pardo. O distrito era subordinado ao município de Santa Cruz do Sul e tornou-se independente em 1959. Foi escolhido, porque embora seja constituído por descendentes de origem étnica predominantemente pomerana, não há muitas pesquisas de levantamento linguístico sobre o pomerano na localidade. Segundo o censo demográfico de 2010 sua população é de 23.983 habitantes, distribuídos entre 10.663 habitantes residentes na zona rural e 13.320 habitantes

[56]Contactamos a ONG Rede Vidas para nos informar a respeito do ano de proposição do projeto Língua Mutter e para obter informações a respeito de sua implementação, porém, nossos emails nunca foram respondidos.

residentes na zona urbana. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla VC-RS.

Os registros eclesiásticos, históricos e também os dados do IBGE que levantamos indicam que Vera Cruz começou a ser colonizada em 1850, com a chegada dos primeiros imigrantes alemães. Essa localidade passou por muitas alterações toponímicas. Inicialmente, foi chamada de Faxinal de Dona Josefa (1858), depois Linha Dona Josefa e Linha Vila Teresa, que era um núcleo urbano o qual, unido à Linha Dona Josefa, passou a constituir o distrito denominado de Vila Tereza, em 1889. Em 1938, passou a chamar-se Tereza e, por fim, Vera Cruz, que é o nome atual.

A economia de base de Vera Cruz é a agricultura, o seu entorno possui características rurais, o seu principal gênero agrícola é o fumo, pois há muitas plantações de tabaco na região, mas também milho, mandioca, arroz irrigado, batata inglesa, soja, cana-de-açúcar, dentre outros. A localidade também possui indústrias e um pequeno comércio, além de estabelecimentos com culinária alemã e feiras de artesanatos. A população costuma participar da *Oktoberfest*, que acontece em Santa Cruz do Sul, cidade famosa por realizar a segunda maior festa deste gênero no Brasil. Esses dados têm origem no nosso diário de campo, proveniente das visitas exploratórias que realizamos.

Nosso objetivo era coletar dados na zona rural e na zona urbana de Vera Cruz, porém nos deparamos com um percalço. Em duas visitas à localidade, não conseguimos encontrar falantes de pomerano na zona urbana e não conseguimos estabelecer um contato-chave para adentrar nas comunidades da zona rural.

Não encontramos falantes, embora os dados históricos (DUMMER, 2009; GRANZOW, 2009) apontem a presença de colonização pomerana naquela região. Inclusive, os dados dos registros eclesiásticos, que constituem um de nossos *corpora* escritos, indicam a maioria de descendentes de pomeranos nessa localidade (ver Gráfico 3, tópico 5.3.2). Não conseguimos encontrar nenhum falante de pomerano para entrevistar, porém, pretendemos, futuramente, se tivermos oportunidade de continuar nossa pesquisa, retornar e realizar novas buscas, explorar mais a região urbana e, principalmente, a zona rural.

Decidimos realizar a caracterização do município, justamente por ele ser pouco pesquisado pelo tipo de abordagem que propomos e porque também coletamos materiais escritos nessa localidade.

Diante do contexto encontrado em Vera Cruz/RS, tivemos que substituir esse ponto de pesquisa por outro, com a presença de falantes de pomerano disponíveis para entrevista. Então,

realizamos entrevistas em São Lourenço do Sul e em Canguçu, que estão localizados fora da região do Vale do Rio Pardo, na região da serra dos Tapes.

4.2.2.2.2 Arroio do Tigre

Arroio do Tigre/RS é um município situado na microregião centro da serra rio-grandense, ao norte do Vale do Rio Pardo. Arroio do Tigre foi oficialmente fundado em 1963.

Escolhemos essa cidade, pois nos informamos sobre a presença de descendentes de pomeranos na localidade por meio de registros eclesiásticos e visitas à região, anteriores ao início da pesquisa. Ao visitarmos a cidade, seus cemitérios e checar os registros, constatamos que algumas famílias ainda residentes na região são descendentes de imigrantes originários da região do *Hunsrik* e da Pomerânia. A partir disso, inferimos que na localidade falantes de diferentes variedades alemãs conviveram juntos em Arroio do Tigre, como de *Hunsrückscher* (HR) e pomerano, por exemplo. Arroio do Tigre possui expressiva presença de descendentes de hunsriqueanos. E existem famílias com ascendência mista, filhos de pais pomeranos e mães hunsriqueanas ou pais hunsriqueanos e mães pomeranas. A cidade é abrangida pelo projeto ALMA[57].

Segundo o censo demográfico de 2010 (IBGE), a população total é de 12.648 habitantes, dividindo se em 6.686 habitantes na zona rural e 5.962 habitantes na zona urbana. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla AT-RS.

O município começou a ser colonizado por volta de 1900, quando chegaram as primeiras famílias de origem germânica, boa parte migrantes de Santa Cruz do Sul e de outras regiões próximas. Vila Tigre, como era denominada, começou a se formar por volta de 1920. Até 1929, pertencia ao município de Soledade, quando passou a ser distrito de Sobradinho. Sua área é constituída de territórios antes correspondentes a Sobradinho, Soledade e Espumoso.

O principal produto agrícola consiste nas plantações de fumo, mas também há plantação de milho e soja dentre os principais gêneros. Uma das principais indústrias é a de malharia, mas o município também possui indústrias de móveis, calçados, metais e cerâmica.

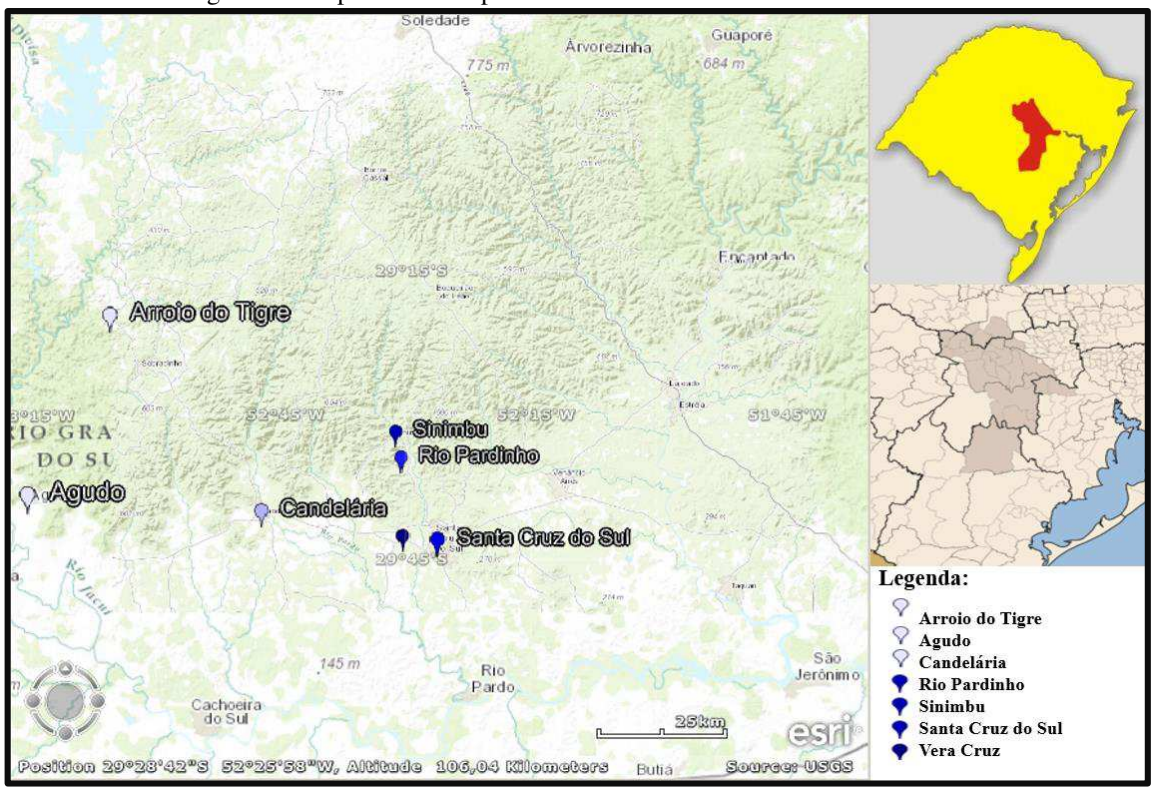
---

[57]Atlas Linguístico-Contatual das Minorias Alemãs na Bacia do Prata. Mais informações disponíveis em: <http://www.ufrgs.br/projalma/oqueeh/apresentacao.html>. Acesso em 02 de junho 2016.

O cemitério evangélico dos imigrantes possui muitas lápides com inscrições nos túmulos, embora as mais antigas já estejam destruídas. Em visita exploratória constatamos os restos de túmulos com inscrições e fotografamos todos os que ainda estavam parcialmente ou totalmente legíveis.

A seguir, apresentamos um mapa da região do Vale do Rio Pardo, que elaboramos demarcando inclusive algumas das localidades mencionadas em nossa pesquisa.

Figura 11: Mapeamento do pomerano no Brasil. O Vale do Rio Pardo/RS



Fonte: Beilke (2013). ArcGis.

No mapa da Figura 11, acima, é possível verificar que, na região do Vale do Rio Pardo, o município de Vera Cruz/RS não é identificado pelo geoprocessamento do ArcGis por seu nome. O ArcGis omite o nome de municípios de porte muito pequeno, indicando apenas aqueles de maior extensão e/ou população, mas é possível encontrar as pequenas localidades por meio das coordenadas. Para isso, tivemos que inserir as coordenadas geográficas de satélite e posicionar o marcador, além de gerar a legendagem com o nome do município, que fica a, aproximadamente, 8,2 Km de Santa Cruz do Sul. No mapa da Figura 11 é possível localizar o ponto de pesquisa Arroio do Tigre/RS ao norte do Vale do Rio Pardo.

4.2.2.2.3 São Lourenço do Sul

São Lourenço do Sul/RS é um município situado à beira da Lagoa dos Patos, "ao pé" da Serra dos Tapes, fundado em 1884. Foi escolhido pela formação étnica pomerana ser relevante na localidade e em substituição ao ponto Vera Cruz/RS (zona urbana). Segundo os dados do IBGE, do censo demográfico de 2010, a população total é de 43.111 habitantes, dividindo-se em 18.874 habitantes residentes na zona rural e 24.237 habitantes residentes na zona urbana. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla SLS-RS.

Os primeiros imigrantes pomeranos chegaram à cidade em 18 de Janeiro de 1858 (quando seu território ainda pertencia ao município de Pelotas/RS). A economia de São Lourenço do Sul é fomentada pela agricultura, pela pesca e também pelo turismo. Na localidade existe o Caminho Pomerano, uma rota para se conhecer mais a história e a cultura pomerana. O pomerano é reconhecido nessa localidade como "legado cultural resultante do processo da imigração e colonização alemã e como língua materna dos descendentes da etnia pomerana" (Prefeitura Municipal de São Lourenço do Sul). Porém, não encontramos uma lei, decreto do município ou projeto de cooficialização do pomerano. Em 2015, aconteceu uma audiência pública no município para debater os direitos dos pomeranos, enquanto povo tradicional.

Visitamos escolas de distritos rurais da localidade em que são desenvolvidos projetos para o ensino do pomerano. As Secretarias Municipais de Cultura e Educação costumam apoiar iniciativas culturais e educativas para a promoção do pomerano no município. Há muitos falantes de pomerano na localidade. Um dos pontos que chamou nossa atenção e que registramos no nosso diário de campo foi a vida cultural ativa, inclusive com bandas que cantam em pomerano e produzem músicas em pomerano.

4.2.2.2.4 Canguçu

Canguçu/RS é um município também situado na região da Serra dos Tapes. Entre seus limites territoriais estão outros dois municípios com presença numerosa de descendentes de pomeranos, São Lourenço do Sul e Pelotas. O município foi fundado em 1857. Foi escolhido pela formação étnica pomerana ser relevante na localidade e em substituição ao ponto Vera Cruz/RS (zona rural). Segundo os dados do IBGE, do censo demográfico de 2010, a população total é de 53.259 habitantes, dividindo-se em 33.565 habitantes residentes na zona rural e 19.694 habitantes residentes na zona urbana. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla C-RS.

Segundo os registros do município, a chegada dos primeiros imigrantes pomeranos data de 1858, quando Rheingantz[58] comprou terras (em 1856) e fomentou a vinda de imigrantes para a região, pois fizera um acordo com o Império brasileiro para a colonização do território. Os imigrantes se instalaram numa área que abrangia a região onde hoje estão Canguçu, São Lourenço do Sul e Pelotas (a Colônia São Lourenço).

A economia de subsistência é a agricultura familiar, pois há muitas pequenas propriedades rurais. Um dos principais gêneros de plantio é o fumo, mas também há plantio de feijão, batata, sorgo, arroz e cebola.

Uma das informações mais relevantes sobre Canguçu, que extraímos do nosso diário de campo, é que foi nessa localidade que pudemos ouvir pessoas falando em pomerano naturalmente nas ruas, sem perceberem que estavam sendo observadas. Também identificamos alguns estabelecimentos comerciais com nomes em pomerano.

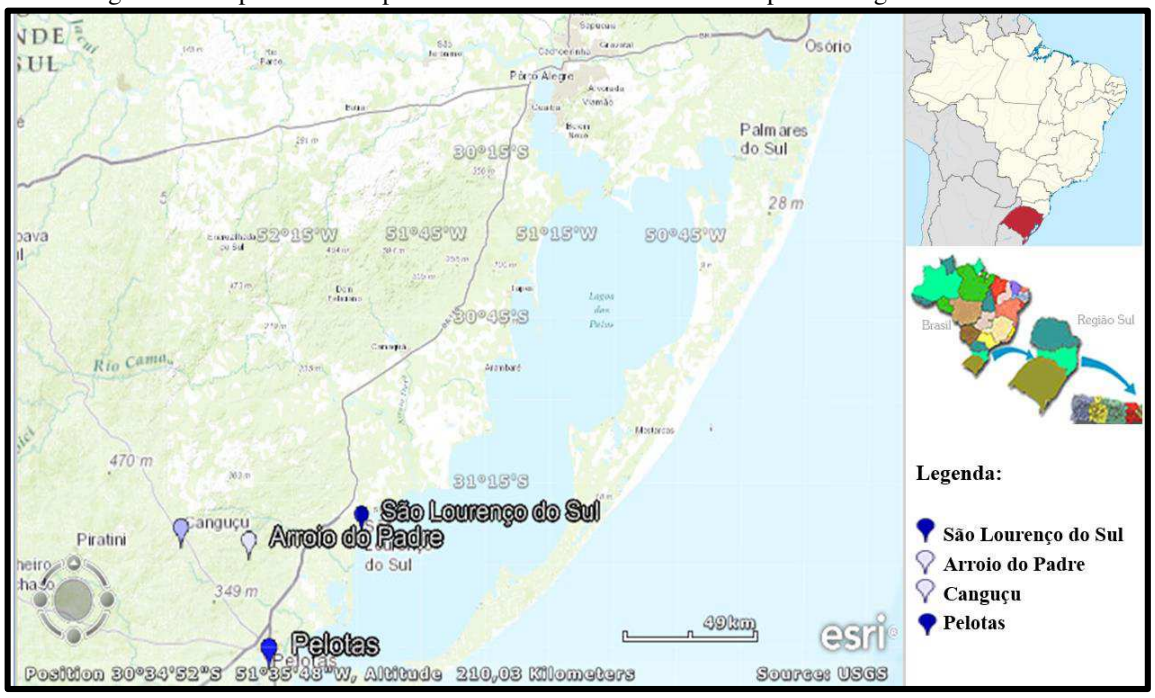
Esta localidade é reconhecida pela presença da fala pomerana, onde existe uma lei de cooficialização, nº 3.473/2010, promulgada pela câmara municipal de vereadores.

O trabalho de tradução dos evangelhos para o pomerano está sendo desenvolvido em Canguçu, por pastores luteranos. Na localidade há transmissão de programas de rádio com conteúdo falado em pomerano.

A seguir, por meio da Figura 12, apresentamos um mapa da região da serra dos Tapes e da Lagoa dos Patos, no Rio Grande do Sul:

[58] Segundo Granzow (2009, p. 137), Rheingantz, era um organizador da colonização no sul do Brasil. Ele era responsável pela distribuição de terras na região da Serra dos Tapes no Rio Grande do Sul (Colônia São Lourenço, atual região de Canguçu, São Lourenço do Sul e Pelotas); ele distribuía as colônias de terra por números. Sobre ele, Granzow informa que: "Em 1856 Jakob Rheingantz adquiriu oito milhas quadradas de terras em Pelotas, para lá criar uma colonização. Ele se responsabilizou de, em cinco anos, fazer as medições das terras e fixar lá pelo menos 1.440 colonos. O governo pagava por cada imigrante, com idade entre 10 e 45 anos, um valor de 30 mil réis por ano. O dinheiro da viagem era adiantado aos colonos pelo Sr. Rheingantz" (GRANZOW, 2009, p. 174).

Figura 12: Mapeamento do pomerano no Brasil. A Serra dos Tapes e a Lagoa dos Patos/RS.



Fonte: Beilke (2013). ArcGis.

Por intermédio do mapa acima é possível localizar os municípios de São Lourenço do Sul e Canguçu, nos quais realizamos coletas de dados orais como alternativa às localidades em que não encontramos falantes (Vera Cruz, zona rural e urbana), por isso substituímos por falantes dessa região.

4.2.2.3 Outras localidades: Santa Maria de Jetibá/ES

Santa Maria de Jetibá é um município da região serrana do estado do Espírito Santo e limita-se com outras cidades capixabas onde também há falantes de pomerano (como Santa Teresa e Santa Leopoldina, por exemplo). O município, fundado oficialmente em maio de 1988, era conhecido até 1943 como Jequitibá.

Escolhemos Santa Maria de Jetibá como ponto de pesquisa extra, devido às entrevistas que precisaram ser descartadas (duas entrevistas realizadas na zona rural de Itueta), devido à relevância do município, à existência de população predominantemente pomerana, à possibilidade de comparação com os municípios do nosso escopo e às justificativas anteriormente mencionadas no tópico 4.2.2.

Segundo os dados do IBGE, do censo demográfico de 2010, a população total é de 34.176 habitantes, dividindo-se em 22.379 habitantes na zona rural e 11.797 habitantes na zona urbana. Enquanto ponto de coleta desta pesquisa, a localidade é identificada nos dados (transcrições e textos doados) pela sigla SMJ-ES.

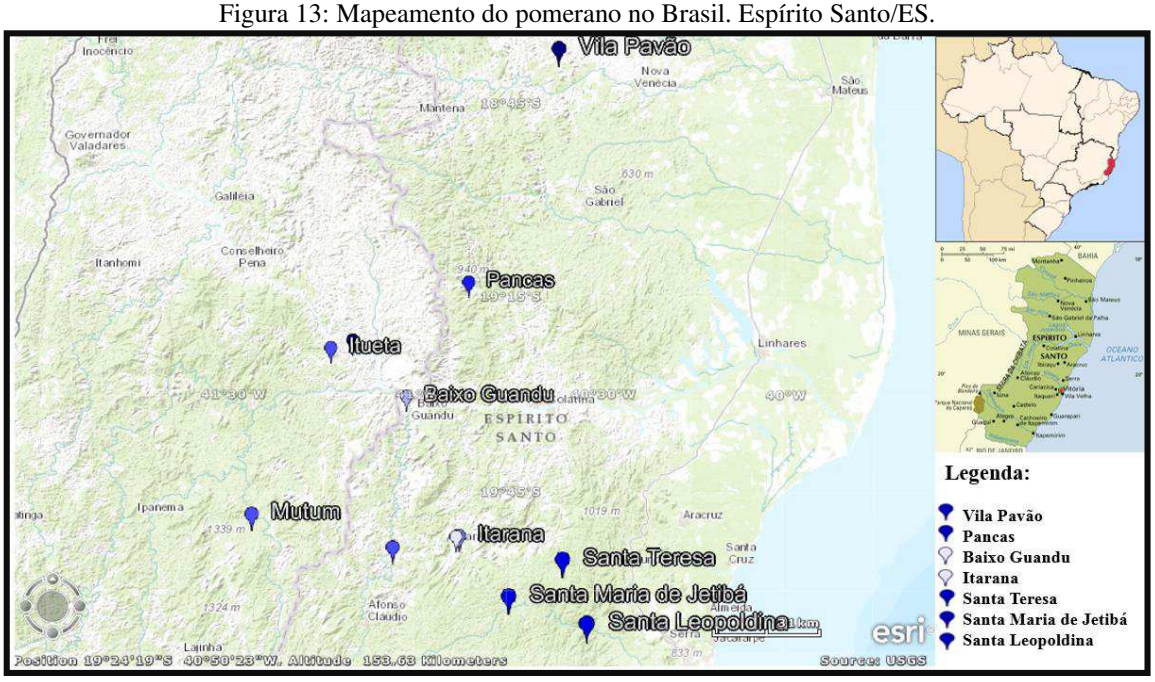
O município formou-se a partir da chegada dos primeiros imigrantes pomeranos na região, por volta de 1859, os quais chegaram para a Colônia de Santa Leopoldina, a região que hoje equivale, em sua maior parte, ao município de Santa Maria de Jetibá, segundo Granzow (2009, p.11). Na década de 1870, novos imigrantes pomeranos chegaram à região.

Sua economia é baseada na agricultura familiar e é um dos municípios que mais produzem legumes e verduras no estado do Espírito Santo. Além da produção de hortifrutigranjeiros, a economia local é movimentada pelas plantações de café, pela avicultura e pelo turismo.

O pomerano é cooficial no município (Lei 1136/2009), no qual há festas tradicionais pomeranas, programa de ensino do pomerano nas escolas, rádio com programação em pomerano e um museu pomerano. Também é possível notar fachadas e placas de identificação de lugares públicos escritos em pomerano, como, por exemplo, a prefeitura e a rodoviária. A 27ª Festa Pomerana na cidade foi realizada entre os dias 28 de Abril a 1º de Maio de 2016. Também foi em Santa Maria de Jetibá que foi publicado o dicionário pomerano-português em 2006.

Produzimos um mapa (Figura 13) para localizar a cidade de Santa Maria de Jetibá no Espírito Santo, conforme a figura seguinte:

Figura 13: Mapeamento do pomerano no Brasil. Espírito Santo/ES.



Fonte: Beilke (2013). ArcGis.

O ponto de coleta de dados orais Santa Maria de Jetibá está representado no mapa acima, situado entre Santa Teresa e Santa Leopoldina.

4.2.2.4 Esclarecimento sobre a caracterização das localidades

Esclarecemos que concedemos maior enfoque na caracterização da região de Minas Gerais por reconhecer essa lacuna e necessidade, pois como já esclarecemos no capítulo 2, tratam-se de localidades pouco abordadas por pesquisas sobre o pomerano, principalmente no campo da Linguística.

Também informamos que "ponto de pesquisa" e "localidades" são sinônimos dentro do contexto deste trabalho. Porém, nem sempre que citamos uma localidade, como cidade, significa que ela foi abordada como um ponto de pesquisa por nós, apenas aquelas que definimos como local de coleta de dados, tanto orais quanto escritos. Cada ponto de pesquisa ou localidade, tem para nós, um subponto ou sublocalidade, que é a zona rural em torno do município escolhido, incluindo os distritos em meio à zona rural.

Doze foram as localidades estabelecidas previamente como pontos de pesquisa, incluindo seus subpontos (zonas rurais). Apresentamos, no Quadro 2 a seguir, todas as localidades e as siglas por nós propostas para indicar cada ponto e subponto de coleta de dados orais:

Quadro 2: Delimitação dos pontos de pesquisa.

LOCALIDADES DEFINIDAS COMO PONTOS DE COLETA DE DADOS	
PONTOS e SUB-PONTOS	SIGLAS
Itueta/MG	IT-MG
Zona Rural Norte de Itueta/MG	ZRN-IT-MG
Vila Neitzel/MG	VN-MG
Vera Cruz/RS*	VC-RS
Zona Rural de Vera Cruz/RS**	ZR-VC-RS
Arroio do Tigre/RS	AT-RS
Zona Rural de Arroio do Tigre/RS***	ZR-AT-RS
Canguçu/RS	C-RS
Zona Rural de Canguçu/RS	ZR-C-RS
São Lourenço do Sul/RS	SLS-RS
Zona Rural de São Lourenço do Sul/RS****	ZR-SLS-RS
Santa Maria de Jetibá/ES*****	SMJ-ES
Observações: *Não foram encontrados falantes. **Idem. ***Não entrevistamos sujeitos na ZR, só na urbana. ****Em SLS as coletas foram realizadas nas zonas distritais, em meio à ZR. *****Nos restringimos à área urbana, pois tratou-se de entrevista extra para comparativo.	

Fonte: Elaboração própria.

Conforme observamos acima, dos 6 municípios somando 12 localidades (zona urbana e zona rural) previamente estabelecidos, foram coletados dados orais efetivamente em 8 localidades incluindo seus distritos e zonas rurais.

4.2.2.5 Contatos-chave

Para contatar os pomeranos e para adentrar nas comunidades, foi necessário primeiro contatar um pomerano que informasse previamente algumas pessoas sobre a nossa pesquisa, para checar a disponibilidade e aceitação de participar das entrevistas. O contato-chave de cada localidade foi uma pessoa que nos acompanhou na maioria das residências e das propriedades rurais onde moravam os pomeranos entrevistados e/ou nos apresentou aos candidatos a sujeitos-participantes das entrevistas interativas.

4.2.3 Plano de recrutamento: dos critérios de inclusão e exclusão, segundo as variáveis da SGL e as exigências do CEP

A abordagem da SGL considera uma vasta amplitude de variáveis e possui um caráter interdisciplinar. Suas variáveis permitem identificar os fenômenos linguísticos nos espaços geográficos, bem como relacionar os elementos socioculturais a esses fenômenos. A SGL considera os processos de interação presentes no momento da coleta de dados e também orienta a escolha dos participantes das entrevistas e os critérios de inclusão e exclusão dos sujeitos.

Com base nisso, foram consideradas: a variável diatópica, porque escolhemos coletar dados de diferentes regiões; a variável diasexual, porque decidimos entrevistar pomeranos de ambos os gêneros; a variável diafásica, porque consideramos mais de uma faixa etária; a variável psicológica, considerada para o planejamento e aplicação das entrevistas, para que obtivéssemos naturalidade, interação e evitar o constrangimento e o retraimento dos participantes. As variáveis culturais também foram consideradas, a fim de evitar questões muito alheias ao contexto pomerano. Não consideramos a variável diastrática, pois não estabelecemos critérios de inclusão ou exclusão para algum estrato social, e nem mesmo a escolaridade foi um fator determinante para inclusão ou exclusão de participantes, pois não queríamos limitar nosso campo de abrangência e nem dificultar a localização de falantes.

As variáveis aqui consideradas foram tomadas apenas para a seleção dos participantes. E elas poderão ser consideradas posteriormente quando da análise dos dados do *corpus* oral. Porém, como este trabalho limita-se a compilação de *corpora*, a consideração das variáveis figura apenas como critérios e métodos para coleta de dados orais e não como objeto de análise.

4.2.3.1 Da justificativa do número de sujeitos participantes da pesquisa e da quantidade de entrevistados

Escolhemos 20 sujeitos para participar das entrevistas, descendentes de pomeranos e falantes de pomerano, nascidos nas localidades selecionadas ou que, ao menos, tenham vivido durante a metade de suas vidas nessas localidades. Estabelecemos a porcentagem da quantidade de sujeitos participantes das entrevistas da seguinte forma: 60% de mulheres em cada localidade, correspondente ao número de três sujeitos participantes do gênero feminino e 40% de homens em cada localidade, correspondente ao número de dois sujeitos participantes do gênero masculino.

Atualmente, estima-se que a população pomerana no Espírito Santo gire em torno de 120 mil descendentes e, em termos de Brasil, talvez, ultrapasse 300 mil pomeranos (TRESSMANN, 1998). Assim, consideramos 20 participantes um número mínimo necessário, pois ainda assim, é apenas 0,66%, ou seja, menos de 1% do total de possíveis falantes de pomerano no Brasil. Então, decidimos que dez indivíduos por região, sendo dez em Minas Gerais e dez no Rio Grande do Sul, seria a quantidade mínima necessária, além de termos decidido realizar uma entrevista extra em Santa Maria de Jetibá/ES, decisão tomada durante o percurso das entrevistas. Também levamos em consideração os custos das viagens, pois o custo da pesquisa seria inviável para trabalhar com menos participantes, já que o custo dos deslocamentos, equipamentos, hospedagens e etc, seriam praticamente os mesmos para se trabalhar com mais ou menos sujeitos nas regiões.

Para justificar a quantidade de sujeitos participantes da entrevista, recorremos à pesquisa de Takano (2013). A pesquisadora realizou um estudo semelhante àquele que aqui propomos e catalogou a variedade nipo-brasileira no entorno de Brasília/DF. No percurso de sua pesquisa, Takano (2013) identificou que a quantidade mínima ideal era de cinco participantes por localidade, concordamos com a pesquisadora, tendo em vista a necessidade de considerarmos o contexto linguístico de desaparecimento da variedade em questão e a dificuldade de encontrar sujeitos com o perfil necessário para participar das entrevistas.

As entrevistas foram realizadas no interior rio-grandense e também no leste mineiro, sendo que a maioria dos falantes foi encontrada na zona rural, mas também os encontramos na zona urbana e nos distritos. Portanto, o parâmetro inicial foi de cinco participantes para cada uma das localidades, considerando para a contagem, a zona rural inclusa, como uma sublocalidade.

Realizamos incialmente as 20 entrevistas pré-definidas com sujeitos que se identificavam como pomeranos, porém, durante as entrevistas percebemos que um sujeito era falante de outra variedade alemã e não o pomerano, no caso era suábio (s*chwäbisch*). Confirmamos isso ao perguntar sobre o local de origem de seus pais. Aconteceu também que outro sujeito que se autodenominou pomerano, falava hunsriqueano. Portanto, das vinte entrevistas realizadas descartamos duas, no sentido de que não as transcrevemos e elas não compõem o nosso *corpus* oral.

A seguir, apresentamos a Tabela 2, com o número de entrevistas realizadas:

Tabela 2: Cálculo do número de entrevistas e participantes.

CÁLCULO DA QUANTIDADE DE ENTREVISTAS REALIZADAS		**PARTICIPANTES***
Cálculo	Nº de entrevistas efetivadas e localidade ZR – Zona Rural + ZU – Zona Urbana	Gênero e Nº de entrevistados
SUBTOTAL 1	10 ENTREVISTAS – RS (8 ZR e 2 ZU)	6-F e 4 M
SUBTOTAL 2	10 ENTREVISTAS – MG (6 ZR e 4 ZU)	6 F e 4 M
SUBTOTAL 3	20 ENTREVISTAS (RS e MG) (14 ZR e 6 ZU)	20 PARTICIPANTES (12 F e 8 M)
EXTRA	1 ENTREVISTA (ES) - ZU	2 PARTICIPANTES F (entrevista dupla)
SUBTOTAL 4	21 ENTREVISTAS (RS, MG e ES) (14 ZR e 7 ZU)	22 PARTICIPANTES (14 F e 8 M)
EXCLUSÃO	2 ENTREVISTAS (MG) - ZR	1 falante de hunsriqueano (F) e 1 falante de suábio (M)
TOTAL:	19 ENTREVISTAS TRANSCRITAS E UTILIZADAS (12 ZR e 7 ZU)	20 PARTICIPANTES falantes de pomerano (13 F e 7 M)

Legenda: Lê-se "F" para gênero feminino e "M" para gênero masculino.

Observações: * Sujeitos-participantes da entrevista como entrevistados, excetuando-se circunstantes e sujeito-entrevistador.

Fonte: Elaboração própria.

Na tabela acima, já estão contabilizadas as localidades (dentro das respectivas regiões) em que não encontramos falantes disponíveis para entrevista (Vera Cruz, por exemplo) e as que possuem somente falantes na zona rural. Além disso, apresentamos o número total de entrevistas realizadas e o número de entrevistas que, de fato, foram utilizadas para esta pesquisa.

Segundo a SGL, a quantidade de participantes é determinada a fim de permitir a verificação da frequência e da distribuição regular. A abordagem-metodologia da SGL permite,

por exemplo, verificar se haveria frequência mais elevada num determinado ponto ou em outro. Ela também permite verificar a porcentagem de falantes, as abstenções, se existe uma norma pomerana, etc. Nossa proposta para este trabalho limitou-se ao levantamento para a compilação de um *corpus* oral, porém esses dados poderão ser futuramente analisados para verificação da distribuição e se existe uma norma pomerana regional e suprarregional.

4.2.3.2 Dos gêneros dos sujeitos participantes das entrevistas

Foram entrevistados três sujeitos do gênero feminino e dois sujeitos do gênero masculino, em cada localidade. Constatamos, por meio de uma sondagem prévia realizada nas localidades, que a maioria dos sujeitos disponíveis para participar das entrevistas era do gênero feminino, porque havia mais mulheres que falavam o pomerano e que poderiam conversar conosco nas regiões alvo da pesquisa. Esse fato foi confirmado quando realizamos efetivamente as entrevistas.

Consideramos que a presença no *corpus* oral de maior número de sujeitos participantes do gênero feminino pode permitir uma maior quantidade de conteúdo lexical proveniente do vocabulário feminino, influenciando o léxico presente no *corpus* oral, de modo que, seja mais próximo a esse universo, visto que não conseguimos encontrar a mesma quantidade de falantes de pomerano do gênero masculino disponíveis para as entrevistas. Em todas as localidades visitadas para coleta de material oral, encontramos mais mulheres falantes do que homens falantes de pomerano, disponíveis para a entrevista. Porém, ressalvamos que a presença de população feminina mais numerosa em algumas localidades (segundo Censo do IBGE, 2010), pode ter interferido no fato termos encontrado mais falantes mulheres. Logo, não é possível atribuir fatores linguísticos para explicar essa questão, sem um estudo científico prévio que demonstre que, de fato, as mulheres preservam mais o pomerano do que os homens[59].

4.2.3.3 Faixas etárias contempladas

[59]Para obter informação oficial a respeito do assunto, consultamos o Censo de 2010 (IBGE), por meio da Lei de Acesso à Informação, para nos informar sobre a população dos municípios pesquisados sob o critério gênero. Em Itueta/MG, Vera Cruz/RS, a população do gênero femino é maior em relação ao gênero masculino. Em Arroio do Tigre/RS, Canguçu/RS e São Lourenço do Sul/RS a população do gênero masculino é maior, porém com pequena diferença. Já em Santa Maria de Jetibá/ES, a população do gênero masculino é maior, em quantidade expressiva. Porém, mesmo em Santa Maria de Jetibá, encontramos com maior facililidade sujeitos do gênero feminino para entrevistar.

Outro critério de inclusão é a faixa etária, de acordo com as classificações próprias da metodologia da Geolinguística e, portanto, da SGL. Também conforme ALIB (2014) e Cristianini (2007). Com base nesses referenciais, optamos por trabalhar com três faixas etárias para os gêneros feminino e duas faixas etárias para o gênero masculino.

Não foram realizadas entrevistas com menores de 18 anos. As faixas etárias selecionadas foram assim divididas para os sujeitos participantes do gênero feminino: faixa etária I – de 18 a 30 anos; faixa etária II – de 40 a 65 anos; faixa etária III – de 66 a 100 anos.

Como era mais provável encontrarmos apenas duas faixas etárias para sujeitos participantes do gênero masculino, que estivessem disponíveis para a entrevista, estabelecemos apenas duas faixas etárias para esse gênero, pois já havíamos constatado essa dificuldade em sondagem prévia que realizamos nos nossos pontos de pesquisa. Então, foram adotadas apenas duas faixas etárias para os sujeitos participantes do gênero masculino sendo faixa etária I – de 40 a 65 anos e faixa etária II – de 66 a 100 anos.

Justificamos nossa decisão, de adotar apenas duas faixas etárias para o gênero masculino, devido ao fato de que os homens, nas comunidades pomeranas, saem mais cedo das comunidades para procurar emprego ou passam muito tempo nas lavouras. Por isso, em nossas visitas de campo, tornou-se raro ver jovens do gênero masculino falando em pomerano, em relação à presença de jovens do gênero feminino falando em pomerano. Como já esclarecemos, para afirmar isso, baseamo-nos na sondagem prévia que fizemos nas localidades alvo da pesquisa, que foi realizada antes da proposição do projeto ao CEP. Durante a sondagem, não foi possível encontrar nenhum rapaz da faixa etária I – de 18 a 30 anos que falasse o pomerano e estivesse disponível para a entrevista (nem mesmo, posteriormente, quando fomos em Santa Maria Jetibá/ES). Tivemos a confirmação dessas situações quando fomos realizar as entrevistas efetivamente e mantivemos apenas duas faixas etárias para o gênero masculino.

A necessidade de considerar as faixas etárias das gerações mais idosas deveu-se ao fato de que, devido ao processo de perda da cultura escrita em ortografia germânica, existe um grande número de falantes de pomerano que possuem muitos conhecimentos restritos à oralidade, geralmente os da faixa etária II - de 40 a 65 anos e os da faixa etária III - de 66 a 100 anos. Os consideramos importantes para as entrevistas, porque o pomerano está sofrendo uma diminuição no número de falantes, principalmente nas novas gerações, e o maior número de falantes é encontrado na terceira faixa etária, por isso também a faixa intermediária (faixa etária II) necessita de maior investigação, pois são a geração de filhos e netos falantes, que, na maioria das vezes falavam o pomerano na fase infantil e deixam de falar da adolescência pra frente, fato

que observamos durante as visitas nas localidades quando buscávamos candidatos às entrevistas e nos informávamos com os contatos-chave.

4.2.3.4 Escolaridade

Quanto à escolaridade, consideramos desde pomeranos que declararam ter pouca escolaridade, alfabetização e/ou primário, até aqueles que cursaram a graduação. Quando se tratavam de pomeranos com dificuldades de leitura, os termos foram lidos em voz alta, na presença de uma testemunha ou de uma pessoa da família. O critério alfabetizado ou não-alfabetizado não foi motivo de inclusão ou exclusão de sujeitos-participantes das entrevistas, mas a escolaridade dos participantes está informada no QS, para permitir futuras análises.

4.2.4 Dos procedimentos para a realização das entrevistas e da descrição dos riscos

Assim como Takano (2013), realizamos um contato face-a-face prévio com os sujeitos candidatos a serem participantes das entrevistas, a fim de propiciar uma situação favorável para a coleta de dados e quebrar a barreira linguística. Takano (2013) seguiu esse procedimento por adotar a orientação de Labov (1972), o qual recomenda que se quebre a distância entre o pesquisador e o sujeito participante da pesquisa antes de efetuar uma entrevista. Esse procedimento visa "minimizar" o stress no momento da entrevista. Então, decidimos adotar o mesmo procedimento.

4.2.4.1 Medidas para contenção dos riscos

Para diminuir o risco dos sujeitos participantes serem identificados, nosso procedimento foi identificar os sujeitos participantes da seguinte maneira: ponto – gênero – faixa etária, como no exemplo: IT-MG-F-I, ou seja, ponto de pesquisa Itueta/MG, sujeito participante do gênero feminino e faixa etária um. Para isso baseamo-nos no modelo de Cristianini (2007).

4.2.4.2 Preparação dos instrumentos de coleta

Para coletar os dados orais por meio das entrevistas foi necessário um planejamento prévio, no qual estabelecemos um roteiro, preparamos os instrumentos de coleta, adaptamos os questionários adotados e adquirimos o equipamento de gravação.

As questões do Questionário Semântico-Lexical/Lexikalisch-Semantischer Fragebogen (doravante QSL/LSF) adotado têm uma formulação inicial igual ou parecida, a fim de assegurar

uma uniformidade no método, tendo o objetivo de evitar a indução ou sugestão de respostas, pois são questões de orientação onomasiológica.

O QSL, a partir do qual fizemos as adaptações, possuía originalmente 202 questões. Nossa adaptação foi elaborada a partir do QSL anexo ao Atlas Linguístico do Grande ABC (CRISTIANINI, 2007), que já havia sido adaptado a partir da versão original do ALIB – Atlas Linguístico do Brasil (iniciado em 1996 e publicado em 2014). Assim, excluímos algumas questões que não seriam coerentes com a realidade pomerana ou que acreditamos que poderiam sofrer estranhamento por sua cultura, por estarem fora de contexto.

Sobre isso, esclarecemos que o questionário foi previamente adaptado para o contexto da cultura pomerana, a fim de evitar prolongamento desnecessário do tempo de entrevista, com questões que poderiam causar constrangimento ou que fossem muito alheias ao contexto histórico da variedade linguística em questão. Evitamos também questões que pudessem fazer os sujeitos-entrevistados se sentirem muito testados e frustrados, por não conseguirem responder.

Acrescentamos também questões básicas que não faziam parte do QSL original, são elas: os meses do ano, os dias da semana, os números e as cores. Consideramos a importância de manter apenas questões que tivessem um potencial produtivo e, assim, deixamos o questionário mais objetivo, com um total de 148 questões.

Também é parte do método a utilização de gravuras, em algumas questões, a fim de auxiliar na obtenção de dados e no desenvolvimento da interação no momento de entrevista. Foram utilizadas seis gravuras: Canga, forquilha, flor de camomila, bala, grampo de cabelo e pão francês. O procedimento era mostrar a figura e esperar que o sujeito a observasse e dissesse o que era, caso ele não dissesse, perguntávamos se ele sabia o nome do objeto.

Pelo método interativo da SGL é facultado ao sujeito-entrevistador realizar as adequações necessárias, no momento da entrevista. Inúmeras situações surgem durante o diálogo e o sujeito-entrevistador precisa estar preparado para se reorganizar sem perder o elo de cumplicidade estabelecido com o sujeito-entrevistado.

Após a edição e adaptação das questões, realizamos a tradução do QSL para o alemão passando a ser chamado de LSF e entregamos para a revisão por um tradutor juramentado. Nossa tradução foi revisada duas vezes, sendo a segunda revisão para corrigir detalhes e declinações que nos escaparam na primeira versão. Esclarecemos que mantivemos propositalmente a orientação onomasiológica original da formulação das questões, mesmo em alemão.

O nosso questionário foi produzido de forma a ser bilíngue alemão-português, com o intuito de aplicá-lo, prioritariamente, como LSF, ou seja, realizamos as perguntas em alemão, com o objetivo de não influenciar os bilingues em pomerano-português a responderem em português, visto que o nosso objetivo era a coleta de conteúdo em pomerano. Os itens do QSL/LSF se distribuem por treze áreas semânticas, a saber: 1. Acidentes geográficos; 2. Fenômenos atmosféricos; 3. Astros e tempo; 4. Atividades agropastoris; 5. Fauna; 6. Corpo humano; 7. Ciclos da vida; 8. Convívio e comportamento social; 9. Religião e crenças; 10. Jogos e diversões infantis; 11. Habitação; 12. Alimentação e cozinha; e 13. Acessórios.

Esclarecemos que não insistimos muito em questões que fossem estranhas ao contexto pomerano, não fazia sentido perguntar, por exemplo, como se diz "semáforo" em pomerano nas localidades que abrangemos. Portanto, evitamos questões que tiveram alto grau de improdutividade nas entrevistas piloto (no item 4.2.4.3, a seguir, falamos sobre as entrevistas piloto) a fim de não causar stress desnecessário aos sujeitos-entrevistados.

Uma pequena amostra do que é o QSL/LSF (APÊNDICE E), pode ser visualizada abaixo, na Figura 14:

Figura 14: Questionário Semântico-Lexical em alemão, excerto.

QUESTIONÁRO SEMÂNTICO-LEXICAL (QSL)
LEXIKALISCH-SEMANTISCHER FRAGEBOGEN (LSF)

ACIDENTES GEOGRÁFICOS / Geografische Gegebenheiten

1. CÓRREGO / RIACHO / 1. WASSERLAUF / BACH
...um rio pequeno de uns dois metros de largura? / Ein kleiner Bach ... etwa zwei Meter breit? (Wie nennt man das? / Wie heißt man das?)

2. FOZ / 2. MÜNDUNG
...o lugar onde o rio termina ou encontra com outro rio? / ... Der Ort, wo der Fluss endet oder in einen anderen Fluss fließt?

3. REDEMOINHO (DE ÁGUA) / 3. WASSERSTRUDEL
...muitas vezes, num rio, a água começa a girar, formando um buraco, na água, que puxa para baixo. Como se chama isto? / ... Oft beginnt sich ein Fluss zu drehen, wodurch ein Strudel entsteht, der alles nach unten zieht. Wie nennt man das?

4. ONDA DE RIO / CORRENTEZA / 4. FLUSSWELLE / FLUSSSTRÖMUNG
...o movimento da água do rio? *Imitar o balanço das águas.* / Bewegung ... das Wasser des Flusses? Imitieren der Welle des Wassers.

Fonte: Elaboração própria.

Outro instrumento utilizado foi o Questionário Sociolinguístico – QS, em nossa versão adaptada para o contexto pomerano, a partir das versões de Cristianini (2007) e Takano (2013). O questionário possui um total de 55 questões, aplicadas em português, pois todos os sujeitos

participantes são bilíngues em pomerano-português. As questões do QS coleta dados pessoais, como idade, escolaridade, local de nascimento e de residência durante a vida, podendo identificar migrações, ascendência e história breve do participante, da família e da localidade. As questões do QS também buscam identificar costumes e hábitos, contatos com os meios de comunicação de massa, gostos, nível de consciência linguística, etc. O QS pode ser visualizado no APÊNDICE F.

4.2.4.3 Testagem do método: realização de entrevistas piloto

Antes de fazer as entrevistas efetivas, realizamos duas entrevistas piloto para afinar a metodologia que adotamos e testar a eficiência dos questionários. Assim, ao aplicarmos o QS e o QSL/LSF pudemos perceber que os questionários eram eficientes, porém percebemos que era importante adaptar as questões no momento da interação da entrevista, priorizar um ambiente de naturalidade e deixar os sujeitos-entrevistados à vontade.

A realização de entrevistas piloto também permitiu que alinhássemos o questionário ao contexto pomerano. Sobre algumas questões éramos informados que determinados objetos não existiam na Pomerânia e, portanto, seus familiares só vieram a conhecer tais objetos no Brasil, de modo que aprenderam tais palavras somente em português.

O principal ganho do teste foi perceber que nossa metodologia possuía potencial para conseguir coletar conteúdo lexical em pomerano e que, em algumas circunstâncias, era necessário, como última opção, fazer as perguntas em português e pedir a resposta em pomerano. Se após todas as alternativas, o sujeito não sabia ou não se lembrava, de forma alguma, perguntávamos se conheciam a palavra, porém em alemão. Nesse caso, acabavam lembrando-se do item perguntado e respondiam em pomerano.

Os testes dos questionários (duas entrevistas piloto) foram fundamentais como forma de nosso treinamento pessoal e preparação para as entrevistas. Eles também nos ajudaram a descrever para o CEP as características da entrevista que propúnhamos, a fim de obtermos deferimento no processo de permissão para realização das entrevistas. Esclarecemos que as entrevistas pilotos não foram transcritas nem contabilizadas.

4.2.4.4 O método de Wenker

Pretendíamos também utilizar para coleta de dados orais o método de Wenker (1880), conhecido como *Die Wenkersätze* (as frases de Wenker), que consiste em questões desenvolvidas por Georg Wenker (1880) para coletar variações dialetais do alemão e elaborar

atlas linguísticos. Além de ser uma metodologia alemã, permitiria uma comparação ao consultar o acervo da *Marburg Universität* que contém respostas desse questionário. Além disso, permitiria a comparação com os dados dos atlas de falas alemãs (*Deutsche Sprachatlas*, que estão sendo digitalizados pelo projeto DiWA – *Digitaler Wenker Atlas*), com alguns *corpora* do baixo-alemão existentes, por exemplo, o *Corpus of Historical Low German* (*Corpus* do baixo-alemão histórico), e o *Referenzkorpus Mittelniederdeutsch* (*Corpus* de referência do baixo-alemão médio) e também comparar com os exemplos do *Audio-Katalog* do REDE[60].

Essa sugestão nos foi dada por um dialetólogo alemão, porém ao realizar algumas tentativas de aplicação, percebemos que não estava sendo eficiente para a coleta dos dados em nosso contexto. Inferimos que a improdutividade do método pode dever-se ao fato de os sujeitos-entrevistados sentirem-se testados e ter ocorrido um retraimento dos participantes no momento da utilização das questões. Porém, acreditamos que o principal fator possa ter sido o cansaço dos sujeitos, que já haviam sido perguntados antes sobre as questões do QS e do QSL/LSF, no total de 203 questões.

Embora esse método não tenha sido eficiente nas poucas oportunidades que tivemos de testar sua aplicação, pretendemos tentar utilizá-lo novamente, em outras oportunidades de pesquisas. Para conhecer o método de Wenker (1880), ver ANEXO B. Já o QSL na versão alemã LSF mostrou ser eficiente e produtivo para atender aos nossos objetivos, portanto o utilizamos para coletar conteúdo lexical em pomerano e obtivemos resultados satisfatórios.

4.2.4.5 Do equipamento para gravação das entrevistas

O primeiro passo foi pesquisar quais eram os gravadores recomendados para a coleta de dialetos, pois precisávamos de um aparelho eficiente que gravasse o som com qualidade. Realizamos algumas pesquisas na Internet e contatamos pesquisadores que trabalhavam com a coleta de dados orais para elaboração de atlas linguísticos. Fomos orientados por um dialetólogo a utilizar o equipamento *Fostex* FR2-LE, porém ele teria que ser importado e, mesmo usado, tinha um custo muito alto. Por isso, foi necessário que optássemos por um aparelho mais acessível. Porém, a escolha do gravador não poderia ser feita sem critérios, visto que a qualidade baixa de gravação invalidaria a coleta de dados orais e até impossibilitaria as transcrições.

Então, decidimos consultar um especialista da área e obtivemos consultoria de um técnico de som direto e editor de som. Nossas exigências principais eram captar som de

[60]*Regional Sprache*. Site que contém arquivos de áudio das falas regionais da Alemanha. Disponível em: <regionalsprache.de>. Acesso em: 15/05/2016.

qualidade estéreo por meio de um gravador portátil que fosse adequado para a aplicação que pretendíamos, até porque não poderíamos constranger o sujeito participante da entrevista com um gravador grande e/ou que ficasse posicionado perto de sua boca enquanto falasse, como em uma entrevista formal, pois o planejamento visava à preparação de um ambiente de entrevista interativo. Outra exigência era que a gravação fosse satisfatória para posterior transcrição do material coletado, sua apreciação e estudo.

O técnico informou-nos a respeito de um aparelho com microfones externos posicionados em uma figura estéreo XY, mais adequado para a captação de falas; uma fonte sonora central/pontual. Também fomos informados de que deveríamos levar em consideração fatores técnicos importantes, como microfone interno, compatibilidade para a reprodução monofônica, gravação em WAVE e MP3, entrada para microfone externo, conexão USB e uso de pilhas comuns.

Sob esses critérios e orientações, decidimos pelo *Zoom H1 Stereo 2.0*, aparelho pequeno e discreto que possui apenas 44mm (largura) x 136mm (comprimento) x 31mm (altura) de tamanho e pesa 60g, compatível para gravação e reprodução monofônica, pois os diálogos e depoimentos são tradicionalmente reproduzidos em mono. Assim, sentimo-nos seguros ao optar pelo aparelho em questão, que possui qualidade do som muito superior aos minigravadores de 10 anos atrás. As gravações foram feitas em formato WAV devido a qualidade de som desse modo ser superior ao do modo MP3.

4.2.5 Da aplicação dos questionários – entrevistas para coleta dos dados orais

Foram utilizados os termos de cessão e permissões de uso adaptados ao projeto e disponibilizados pela Plataforma Brasil. As entrevistas foram cedidas mediante consentimento livre e esclarecido dos sujeitos participantes voluntários. Todos os sujeitos participantes das entrevistas foram informados, antecipadamente, sobre tudo que envolvia a participação na pesquisa por exemplo: o que é a pesquisa, quais são os objetivos dela e quais os riscos e os benefícios da pesquisa.

Foram utilizados o QSL/LSF (APÊNDICE E) e o QS (APÊNDICE F). Optamos pela aplicação da versão LSF do QSL, em alemão. Embora na prática o questionário não tenha sido seguido *stricto sensu*, ele foi um roteiro, pois houve pequena variação na formulação das perguntas "ao vivo", já que a entrevista funcionava como uma conversação dialógica, não podendo ser mecânica e distanciada dos falantes. Assim, as perguntas foram seguidas, porém

não foram decoradas ou reproduzidas mecanicamente, nem obrigatoriamente seguidas na mesma ordem e exatidão, deixávamos que um assunto atraísse o outro.

Esses questionários foram as principais ferramentas que decidimos utilizar para coletar atos reais da fala e informações sobre a cultura, história e espaço geográfico dos pomeranos, baseados na metodologia da SGL, que preconiza o momento da entrevista como um momento de interação face-a-face, uma relação estabelecida no momento do diálogo com o sujeito-entrevistado.

A aplicação dos questionários foi oferecida nas duas versões paralelas: português, para o caso dos falantes bilíngues, e a versão *Hochdeutsch* – alemão-padrão –, pois os falantes de *Pommersches Plattdüütsch* – o pomerano –, compreendem boa parte do alemão-padrão, conforme constatamos ao realizar as entrevistas e em outros momentos de interação e contato com os pomeranos. Iniciávamos as perguntas em alemão e fazíamos as perguntas em português só em último caso, quando não eram compreendidas totalmente em alemão. As respostas eram quase sempre fornecidas em pomerano.

Essa metodologia nos permitiu verificar em qual forma o sujeito participante sentia-se mais à vontade e o grau de inteligibilidade entre pomerano e alemão. Algumas questões puderam ser verbalmente traduzidas para a pronúncia pomerana no momento da entrevista, pela autora deste trabalho, operando a substituição lexical, com base na fala pomerana presenciada, e também nas consultas realizadas no próprio conjunto dos *corpora* escritos, na época ainda em construção.

A aplicação dos questionários, QSL/LSF e QS, teve o objetivo de determinar a presença e a repetição dos fenômenos linguísticos nos pontos de pesquisa abrangidos em nosso trabalho de campo, para permitir futuras análises e elaboração de cartogramas (CRISTIANINI, 2007, p. 101), principalmente, para que suas transcrições possibilitassem a constituição de um *corpus* oral dialetal do pomerano.

4.2.5.1 Da duração das entrevistas

As entrevistas tiveram duração média de uma hora, podendo variar para um pouco mais ou um pouco menos, respeitando o ritmo de fala e a espontaneidade de cada sujeito participante, bem como seu maior ou menor nível de inibição. Apenas um sujeito participante da entrevista, do gênero feminino, demonstrou maior timidez, que foi logo superada pela descontração de algumas questões. Também percebemos que se tratava de um traço da personalidade da falante; o falar baixo e devagar, o jeito de ser mais comedido, o que pode ser confundido com timidez.

Aqui percebemos o quanto a variável psicológica e a visão da entrevista como interação são importantes.

Houve também o caso de uma entrevista que durou mais de duas horas, pela fluência da falante, que ao ser questionada sobre um item, como, por exemplo, casa, além de respondê-lo, deteve-se em contar toda a história da aquisição de sua casa, em pomerano, sua língua materna.

A duração do tempo de todas as 19 entrevistas somou 31 horas, 1 minuto e 18 segundos, de modo que cada entrevista durou em média 90 minutos, incluindo a aplicação do QS. Mas houve entrevistas que duraram quase 150 minutos. As entrevistas com sujeitos do gênero masculino giraram em torno de 60 minutos. Mas houve uma variação, com uma entrevista de apenas 40 minutos, considerando somente o LSF/QSL. A partir disso, observamos que os sujeitos-entrevistados do gênero feminino interagiam mais conosco do que os do gênero masculino, tornando as entrevistas com as mulheres mais longas. Os sujeitos-entrevistados do gênero masculino eram mais objetivos durante as interações.

4.2.6 Tratamento dos dados orais

A transcrição foi o principal tratamento dos dados orais realizado para constituir o nosso *corpus* oral. Com essa atividade, tornamos os dados aptos para serem lidos pelo computador e pelo WST. Cabe esclarecer que não utilizamos o procedimento de alinhamento som-escrita.

A seguir, descrevemos o tratamento dos dados orais ao falar do método de transcrição, da etiquetagem parcial e da exclusão dos dados antroponímicos.

4.2.6.1 Método de transcrição

O método de transcrição consistiu na audição atenta dos áudios das entrevistas, com o uso de fones de ouvido. As entrevistas foram transcritas de forma digitada manualmente, diretamente em arquivos de extensão doc. As entrevistas foram escutadas mais de uma vez. Íamos pausando, voltando e conferindo os dados digitados durante a transcrição.

Acreditávamos inicialmente que seria necessário filtrar os áudios em programas para extração de ruído, amplificação de som e, até, desaceleramento da fala. Porém, foi possível transcrever as entrevistas sem o auxílio desses recursos. Todavia, reconheçemos que foi um processo difícil, cansativo e longo.

O *corpus* oral constituído pelas 19 entrevistas foi transcrito pelo método ortográfico, seguindo a forma de grafia já convencionada no documento que expomos no tópico 4.1.3.2 (Figura 07), adotando assim a mesma forma de escrita utilizada na conversão e convenção dos

corpora escritos, excetuando-se os casos de variação, que não foram normatizados, justamente para não perdermos o registro das variações.

Quando os sujeitos-entrevistados respondiam "misturando"[61] português e pomerano, transcrevemos fielmente, grafando inclusive as variações do português não-padrão, conforme pronunciado pelos falantes. Fizemos essa opção para ser fiel aos dados e, também, porque pode permitir o estudo da situação do português entre os pomeranos bilíngues em pomerano-português.

A transcrição das 19 entrevistas durou exatos 94 dias e cerca de 2.256 horas de tempo despendidas nas transcrições, sendo os 30 primeiros dias sem dedicação exclusiva, nos quais transcrevemos cinco entrevistas, e os 64 dias restantes com regime de dedicação total e exclusiva, durante todos os dias e quase todas as horas desses dias. A média era de cinco dias diretos para cada entrevista. Inserimos essas informações detalhadas sobre o tempo de transcrição, pois acreditamos que essa informação pode auxiliar quem queira elaborar um cronograma para um projeto de pesquisa que envolva compilação de *corpus* oral e/ou transcrição de falas.

Durante as transcrições dos dados orais consultávamos nos *corpora* escritos a recorrência de escritas de uma mesma forma, tanto manualmente, quanto com o auxílio do WST. Desse modo, quando um item era originalmente escrito nos *corpora* escritos de diferentes maneiras, porém, uma forma era mais frequente, utilizávamos esse critério para grafar itens sobre os quais tínhamos dúvidas em como transcrever o que ouvíamos.

Além de consultar constantemente os textos da literatura pomerana, já mencionados neste trabalho, para nos basear ao transcrever as entrevistas, também consultamos a *Barther Bibel* (mencionada no capítulo 2 deste trabalho), em PDF, e nos baseávamos na escrita dela para decidir sobre a escrita de alguns itens. No entanto, como o formato desta Bíblia não permitia consulta dos itens pela função localizar, por estar em letra gótica, desenvolvemos o seguinte método: consultávamos pela função busca da Bíblia *online* em alemão, versão de Lutero, um versículo que contivesse a palavra desejada de se grafar. Em seguida, ao saber em qual versículo conteria aquele item, íamos para página, capítulo e versículo correspondentes na Bíblia pomerana, localizávamos pela leitura atenta aquele item, dentro do versículo e, assim,

---

[61] Falamos em mistura e não em interferência, com base em Bossmann (1953), Heye (1986), Gumperz (1998), Von Borstel (2011) e Takano (2013). A interferência parece ser unilateral, já a mistura é quando os falantes aparentemente não percebem os empréstimos, sendo para eles um código só, uma fala homogênea, bidirecional e em situação de estabilidade, como um mesmo código, então preferimos falar em mistura. Constatamos que alguns pomeranos misturam pomerano e português sem perceber e já outros comentam que hoje o português e o pomerano está "misturado".

descobríamos como o item foi grafado originalmente no pomerano da *Barther Bibel*. Quando conferia com a pronúncia e era o mesmo item lexical, escrevíamos conforme encontrado na Bíblia pomerana.

Também salvamos os textos em várias versões: uma versão completa em doc, contendo os diálogos entre sujeito-entrevistado e sujeito-entrevistador, com os nomes dos participantes; uma versão em doc com o nome dos participantes já codificados; e uma versão igual a esta última, porém, em extensão txt, e somente com as respotas dos sujeitos-entrevistados para permitir a leitura pelo WST.

Esclarecemos que a transcrição dos dados orais e a escrita dos dados dos *corpora* como um todo – incluindo os dados escritos – carecem de uma validação, a qual poderá ser realizada futuramente por um falante de pomerano que conheça a grafia pomerana.

Outro tratamento dos dados orais foi a etiquetagem parcial em alguns textos, para identificar alguns itens, quando consideramos que era necessário inserir informações, pois a transcrição não carrega consigo as informações sonoras e, por isso, ressalvamos algumas anotações que contribuem para identificação do som e das variações. Além disso, inserimos outras etiquetas, sobre as quais explicaremos melhor no próximo tópico.

4.2.6.2 Etiquetagem parcial

Criamos alguns códigos ou siglas para identificar as informações que inserimos como uma forma de etiquetagem parcial dos dados orais, no que tange a traduções para o português, para o alemão, e também observações sobre identificação de verbos, substantivos, variações dentro do pomerano, recuperação do item lexical no alemão-padrão e outras informações que consideramos úteis.

A etiquetagem parcial foi um recurso útil para identificar entre parêntesis angulares "< >" (ou *tags*) os itens que eram verbos, substantivos e artigos, nos limitamos a eles por questão de tempo e também porque pretendíamos produzir um glossário. Também fazíamos a tradução de algumas palavras menos comuns para o português e, quando percebíamos que era uma variação no uso de um mesmo item que fora pronunciado diferente em outra região, também etiquetávamos. Outra forma de etiquetagem foi indicar a forma dos itens em alemão-padrão, para permitir comparações posteriores durante as análises.

A etiquetagem foi parcial, pois foi feita quase exclusivamente no *corpus* oral e, mesmo assim, somente em alguns arquivos. Não haveria tempo hábil para etiquetar todos os dados dos

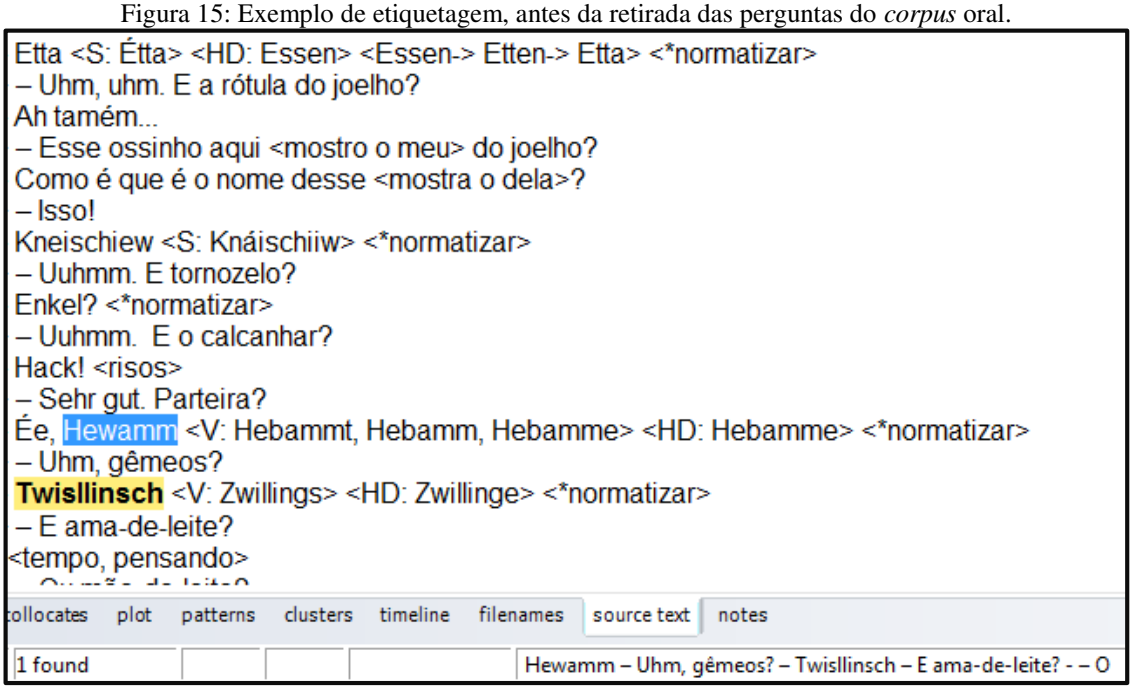
corpora, em apenas dois anos de mestrado, portanto a prioridade foi etiquetar itens que se destacavam pela raridade, variação diatópica ou variação em relação ao alemão-padrão.

Seguem alguns exemplos de etiquetagens realizadas durante o tratamento de dados:

Hochtiedsbirer <HD: Hochzeitsbitter, PT: Convidador para casamento>.
<V: Hochtiedbirar>. Em que, HD é *Hochdeutsch*, PT é português e V é variação (1).

Também criamos algumas etiquetas para identificar a variação de som, para uma forma de escrita que padronizamos, por exemplo: *Fruug* <S: Fruuch>, onde S é som. Ainda outro caso de etiqueta ou anotação entre parêntesis angulares era quando precisávamos fazer alguma observação sobre aquela fala, nossas conjecturas sobre metaplasmos, marcação de itens que gostaríamos de verificar se se tratam de inovações que ocorreram no alto-alemão e não ocorreram no baixo-alemão, etc. Segue, na Figura 15, alguns exemplos da nossa etiquetagem parcial, em um trecho com perguntas realizadas em português.

Figura 15: Exemplo de etiquetagem, antes da retirada das perguntas do *corpus* oral.



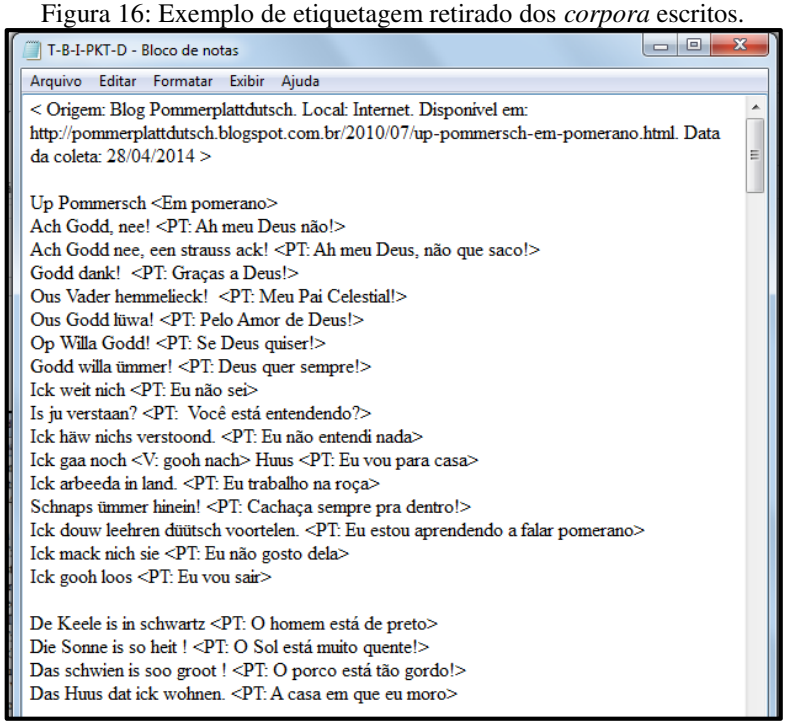
Fonte: Elaboração própria.

Para conseguirmos padronizar a escrita de transcrições que tínhamos feito no início e as que fizemos por último, marcamos os casos que tivemos dúvida sobre a escrita com <*normatizar> para podermos buscar pela etiqueta e normatizar. A figura acima mostra um trecho de interação com perguntas em português, antes da extração das perguntas feitas por nós e do salvamento da última versão.

Reconhecemos que podem haver variações reais na pronúncia que foram etiquetadas com <V: >, mas também podemos ter escrito de maneira diferente, pois o *corpus* ainda carece de mais uma revisão atenta para normatização, o que demanda muito tempo.

Apresentamos, na Figura 16, a seguir, mais um exemplo de etiquetagem, retirado dos *corpora* escritos.

Figura 16: Exemplo de etiquetagem retirado dos *corpora* escritos.



Fonte: Elaboração própria.

Foram criadas as etiquetas: <S:> para representar o som, com o intuito de informar a pronúncia pomerana; <PT:> para inserir a tradução para o português; <HD:> para inserir a forma em alemão-padrão; <HR:> para marcar itens possívelmente hunsriqueanos e fazer sua posterior conferência; <BP> para marcar intes possívelmente provenientes do *Brasilianisch-Pommersch*, ou seja, a variedade brasileira do pomerano; <V:> para identificar variações não normatizadas; <M:> para anotar hipóteses sobre metaplasmos e a etiqueta livre < >, com comentários e análises do pesquisador. Essas etiquetas foram utilizadas tanto para os *corpora* escritos quanto para o *corpus* oral. No caso de algumas escritas pomeranas, que encontramos na forma transliterada, as mesmas foram etiquetadas para manter a forma original paralela à nossa conversão da escrita. Não foram todos os casos e arquivos de dados que foram etiquetados, somente quando houve tempo hábil para fazê-lo. O conjunto dos *corpora* poderá ser futuramente etiquetado de forma completa.

4.2.6.3 Exclusão de dados antroponímicos

Devido à exigência do CEP de não identificar os sujeitos participantes das entrevistas, foi necessário omitir todos os nomes e sobrenomes dos pomeranos na transcrição dos dados. Por isso, criamos dentro dos arquivos uma codificação que se refere à identificação das falas, criando um código para a etiqueta de identificação do sujeito-entrevistado, com as iniciais de seu nome, e um código de identificação do sujeito-entrevistador, com as iniciais de seu nome, sendo estas últimas antes da exclusão das questões nos arquivos.

Adotamos esse procedimento, embora fosse desejo de algumas famílias que seus sobrenomes fossem divulgados, essas manifestaram contentamento por serem participantes de nossa pesquisa. Também consideramos que seria importante expor os sobrenomes nos *corpora*, pois isso permitiria o estudo de antropônimos. No entanto, obedecemos às determinações do CEP e excluímos os nomes e sobrenomes. Permaneceram nos arquivos finais apenas as codificações de identificação mencionadas.

4.3 Organização dos *corpora* escritos e do *corpus* oral como uma unidade

Descrevemos, nesta seção, os procedimentos aplicados para organizar, identificar e armazenar os *corpora* compilados.

Devido à diversidade de fontes de coleta, os dados compilados são provenientes de diversos gêneros, como músicas, dísticos, receitas, etc., de modo que foi necessário organizar todos os textos de uma forma lógica e criteriosa. A união dos *corpora* escritos e do *corpus* oral constitui para nós uma unidade, o conjunto dos *corpora* do pomerano. Por isso, decidimos tratar da organização dos dados escritos e orais de uma só vez.

A seguir, tratamos sobre a codificação que originou a nomenclatura dos arquivos, a identificação interna da origem de cada texto pelos cabeçalhos e o agrupamento, separação e reagrupamento, segundo a classificação dos dados em diamesias, gêneros, domínio discursivo e suportes.

4.3.1 Codificação e nomenclatura dos arquivos

Vamos agora demonstrar como nomeamos os arquivos, elaborando codificações para os seus títulos, a fim de identificá-los dentro do conjunto dos *corpora* escritos e do *corpus* oral do pomerano. A organização também foi realizada separando os arquivos e reagrupando-os. Para isso, os critérios observados foram as diamesias, gêneros, domínio discursivo e suportes identificados em relação aos textos.

Primeiramente, foi preciso criar um padrão de nomenclatura que identificasse os arquivos, bem como nos orientasse em relação ao conteúdo de cada arquivo em relação ao tipo de fonte, origem e gênero dos textos. Por isso, definimos uma sequência alfabética e nomeamos cada arquivo com códigos, de acordo com os critérios: diamesia – escrita ou falada; autoria; localidade de coleta; gênero ou suporte. A partir disso, convencionamos, por meio de quadros descritivos, a nomenclatura do conjunto de *corpora*.

No Quadro 3, a seguir, descrevemos o significado de cada sílaba alfabética que foi utilizada para cada tipo de arquivo. Como exemplo, podemos citar a dupla de letras "AT" – que no conjunto dos dados significa Arroio do Tigre, uma das localidades onde foram coletadas amostras da variedade pomerana. Deste modo, todos os arquivos que contêm "AT" são provenientes de coletas realizadas em Arroio do Tigre, no Vale do Rio Pardo, no Rio Grande do Sul. E assim por diante, todos os conjuntos de letras, aos quais chamamos "códigos", utilizados para identificar os arquivos, estão abaixo discriminados:

Quadro 3: Legendas dos códigos de identificação dos conjuntos dos *corpora* escritos e oral.

Legendas para identificação dos *corpora*
Autoria
FG – Família G
FHT – Família HT
FK – Família K
FP – Família P
KS – Kuhn Silva
T – Tressmann
JW – Jehowas Witness
SA – Sem autoria
Localidade de origem/coleta
AT-RS – Arroio do Tigre no Vale do Rio Pardo no Rio Grande do Sul
C-RS – Canguçu no Rio Grande do Sul
IT-MG – Itueta em Minas Gerais
SLS-RS – São Lourenço do Sul no Rio Grande do Sul
VC-RS – Vera Cruz no Vale do Rio Pardo no Rio Grande do Sul
VN-MG – Vila Neitzel em Minas Gerais
ZRN-MG – Zona Rural Norte (de Itueta) em Minas Gerais
SMJ-ES – Santa Maria de Jetibá no Espírito Santo
I – Internet
Localização no acervo segundo a diamesia: falada ou escrita
PKO – *Pommersch Korpus* Oral
PKE – *Pommersche Korpora Escritos*
Suporte da fonte original
A – Áudios WAV
D – Digitados Doc
DT – Digitalizados em PDF (copiável ou conversível em texto por OCR)
CIP – Cópias de Impressos, coletadas pessoalmente.
M – Manuscritos
P – Pictóricos: fotos, lápides fotografadas, imagens

V – Vídeos WMP
J – Jornal – Folha Pomerana
L – Livros com conteúdos em pomerano ou contendo trechos em pomerano
LC – Lápides de Cemitérios
WS – Wandschoener- inscrições em tecidos, paredes, tábuas e quadros
Gêneros
AA – Artigos Acadêmicos
AI – Artigos de Internet (conteúdos de matérias ou reportagens de blogs e sites)
CP – Cartas Pessoais
DOC – Documentários (conteúdos extraídos das legendas)
ENT – Entrevistas
M – Músicas
PK – Pomerischen Kalendar – calendários pomeranos
PL – Poéticos e literários – poemas, versos, rimas e trovinhas
RC – Receitas Culinárias
RE – Registros Eclesiásticos
RD – Religiosos diversos: orações e lembranças de rituais eclesiásticos como CC – Convites de casamento, LB - Lembranças de batismo, LC – Lembranças de casamento
Observação: Tentamos coletar scripts de Rádios, porém não obtivemos sucesso. Aqui teríamos também programas de rádios em pomerano se tivéssemos obtido êxito.

Fonte: Elaboração própria.

No Quadro 4, a seguir, estabelecemos uma ordem para as sequências alfabéticas descritas no quadro anterior, pois, para fins de organização e clareza na leitura dos dados, foi preciso determinar uma posição para cada sílaba, de acordo com o tipo de informação que cada sílaba contém. Assim, "AA" na posição primeira não significaria o mesmo que "AA" na posição quinta, caso houvesse repetição. Deste modo, fica explícito que a primeira sequência é o tipo de material coletado, a segunda é a autoria (quando houver), a terceira indica o local de coleta, a quarta indica a localização no acervo (diamesia escrita ou oral) e a quinta posição indica o suporte no qual fora coletado originalmente o material antes do tratamento dos dados.

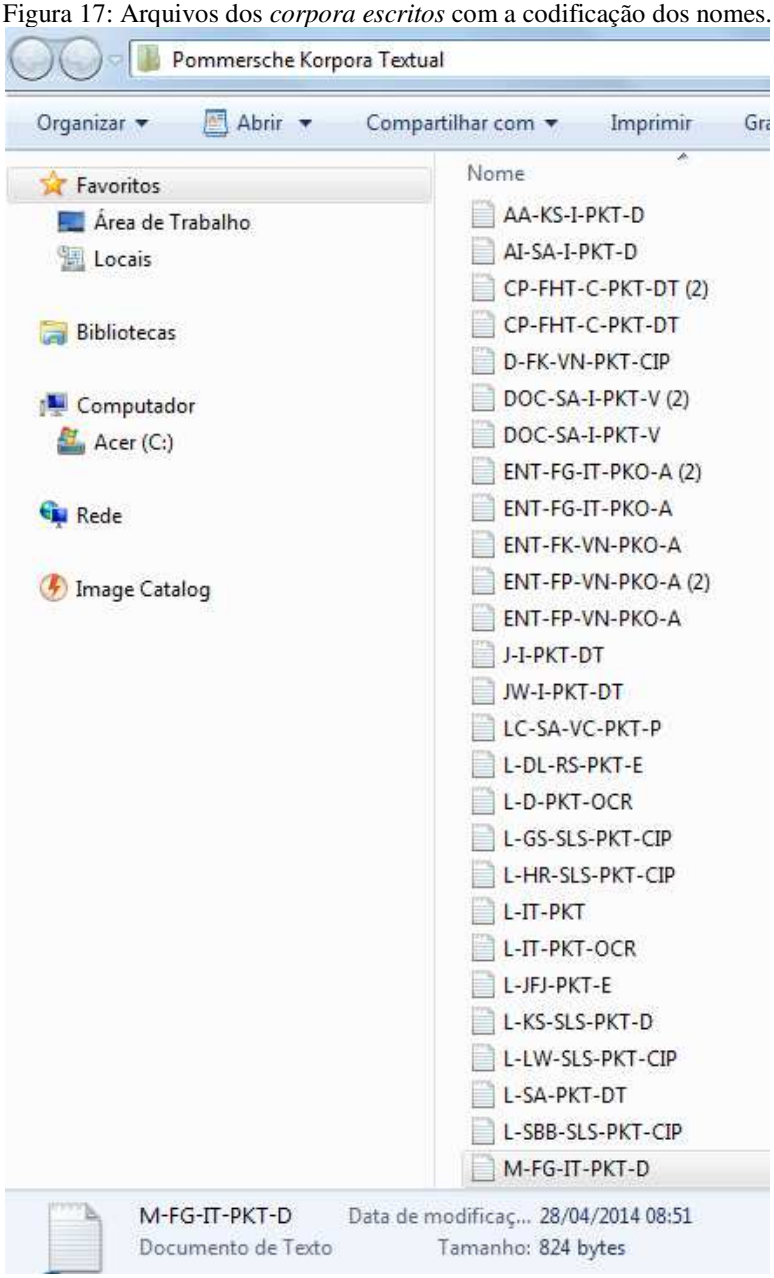
Codificação para nomenclatura dos arquivos dos *corpora* do pomerano*				
1ª Posição	**2ª Posição**	**3ª Posição**	**4ª Posição**	**5ª Posição**
Gêneros	**Autoria**	**Local de coleta**	**Localização no acervo segundo a diamesia**	**Suporte da fonte original**
AA	KS	I	PKE	D
AI	AS	I	PKE	D
CP	FHT	C-RS	PKE	DT
RD	AS	VC-RS	PKE	CIP
DOC	AS	I	PKE	V
D	FK	VN-MG	PKE	CIP
ENT	FG	IT-MG	PKO	A
J	-**	I	PKE	DT

RD	JW	I	PKE	DT
LC	AS	VC-RS	PKE	P
*Foram exemplificados apenas os dez primeiros casos, visto que se trata de um acervo extenso. **Na segunda posição, além de ocorrer fontes sem autoria definida ou localizada, também há variações nas iniciais do nome dos autores.				

Quadro 4: Ordem dos códigos de identificação do conjunto dos *corpora* com exemplos.
Fonte: Elaboração própria.

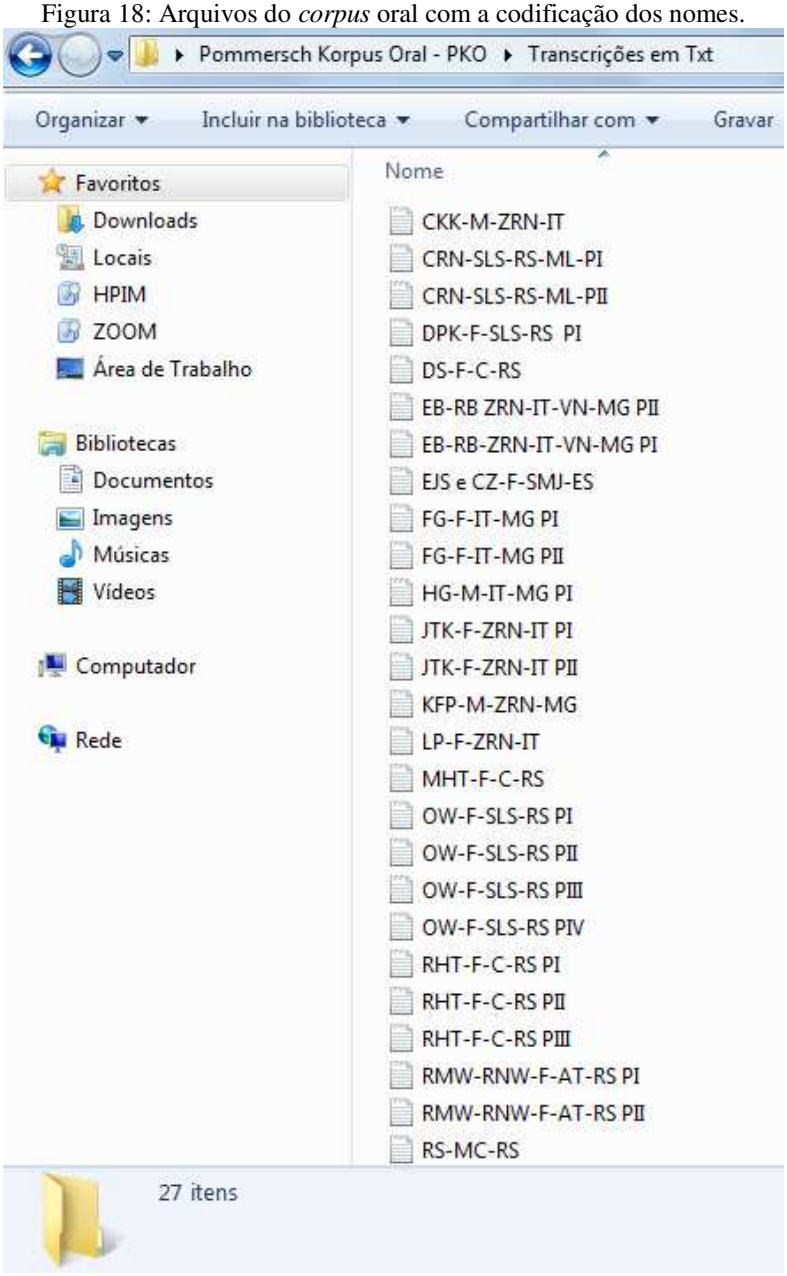
A sequência dos códigos que conferem nomes aos arquivos e os códigos podem ser visualizados por meio das Figuras 17 e 18, inseridas abaixo, que mostram, respectivamente, os arquivos dos *corpora* escritos e os arquivos das transcrições do *corpus* oral, já nomeados e armazenados no computador:

Figura 17: Arquivos dos *corpora escritos* com a codificação dos nomes.



Fonte: Elaboração própria.

Figura 18: Arquivos do *corpus* oral com a codificação dos nomes.



Fonte: Elaboração própria.

Além da identificação externa de cada arquivo de texto do conjunto dos *corpora*, conforme foi detalhado nos dois quadros mencionados (Quadros 3 e 4), também fizemos a identificação interna dos textos, inserindo cabeçalhos, onde foram informados: a origem de cada texto, que poderia ou não coincidir com o local da coleta; a autoria (quando havia); e também, a data da coleta do texto.

Na próxima seção, explicamos como reunimos e organizamos toda a diversidade de dados coletados.

4.3.2 Agrupamento, separação e reagrupamento segundo a classificação em diamesias ou gêneros ou domínio discursivo ou suportes

Nesta seção, descrevemos o desenho do nosso conjunto de *corpora*, pois, para conseguir organizar a diversidade de dados obtidos na compilação, foi necessário apoiar-nos em nossos referenciais (ver Capítulo 3) para, sob seus princípios, fazer uma classificação dos materiais coletados e compilados.

Começamos a separar os textos salvos em txt, depois de agrupá-los em pastas de arquivos, conforme demonstram os Quadros (3 e 4) e as Figuras (17 e 18) citados no item 4.3.1. Nessas pastas constavam todos os textos coletados e também as entrevistas transcritas. Foi necessário observar as diamesias, gêneros, o domínio discursivo e os suportes dos dados, para poder separar os *corpora* e organizá-los em pastas virtuais e reagrupar os arquivos de acordo com a identificação dos mesmos.

Partindo das diamesias de Berruto (1993), realizamos o primeiro passo, que foi a separação dos dados em duas subpastas diferentes: *corpus* oral, diamesia falada; e *corpora* escritos, diamesia escrita. Por exemplo, no caso do *corpus* oral transcrito, ele foi separado fisicamente por ser material proveniente da fala. No caso dos *corpora* escritos, todos os arquivos foram reunidos em uma pasta nomeada de diamesia escrita, porém foram reorganizados em subpastas conforme seus gêneros.

O segundo passo desta etapa foi separar os arquivos dos *corpora* escritos, conforme os gêneros ou suportes, com base em Marcuschi (2008). Porém não seria possível utilizar todas as categorias do autor, pois para classificar os *corpora* em tipo e domínio discursivo (exceto o discurso religioso), teríamos que ler o conteúdo dos textos do acervo e analisá-los. Assim, estando nós apenas no momento de separação e reagrupamento dos textos, fizemos uma categorização superficial, na medida em que atendesse a nossa necessidade de criar unidades, conjuntos para formar os *corpora*.

Contudo, em alguns casos não foi possível separar os *corpora* escritos por gêneros, então, o terceiro passo, foi separar os arquivos que, na impossibilidade de separá-los por gêneros, optamos por separá-los pelos suportes.

Reconhecemos que todos os gêneros coletados possuem um suporte, mas não conseguimos distinguir alguns gêneros e tipos, pois para isso teríamos primeiro que realizar alguns estudos dos conteúdos dos *corpora*. Um exemplo é o *corpus* das inscrições tumulares. Explicamos: as lápides que contêm os textos do *toofel* (ou *tåfel*), que não é um gênero em si,

mas uma prática cultural da simbologia mortuária[62] pomerana – que consiste em colocar no túmulo uma tábua de madeira ou placa de cimento com textos que representam as crenças do falecido e de sua família – não puderam ser separadas e reagrupadas por gênero. Nesse caso, tanto as placas de madeira quanto as lápides em cimento são suportes para os textos e não um gênero em si. Os gêneros existentes nas lápides variam, sendo encontrados versículos bíblicos, mensagens pessoais, adágios, dísticos (séries de até dois versos rimados), *Sprüche* (sentenças, máximas) e fraseologismos diversos. Então, para não ter que deixar uma inscrição em cada arquivo, em cada pasta separada, conforme cada um desses gêneros, optamos por reuni-los sob o critério de classificação: suporte – lápide de túmulo, e assim constituímos um conjunto de arquivos.

Um caso excepcional foi o do diário pessoal, o qual não constituiu um *corpus*, pois, de acordo com o critério que adotamos, exposto na discussão teórica (Capítulo 3), um *corpus* é constituído de dois ou mais textos. No caso, teríamos que ter conseguido pelo menos mais um diário para formar um conjunto.

Para expor o modo como agrupamos os *corpora*, segundo as classificações de suas instâncias comunicativas e fontes de coleta, elaboramos um quadro. O Quadro 5 sistematiza essas informações de maneira compacta:

[62] Conforme Cunha (2011). Temos ciência de que a forma dicionarizada para as práticas é mortuórias. Porém, como nos baseamos em um autor que estudou o assunto, decidimos reproduzir o termo conforme o uso do autor. A expressão "simbologia mortuária pomerana" já ficou conhecida assim.

Quadro 5: Agrupamento do conjunto dos *corpora* do pomerano

Agrupamento dos *corpora* por diamesia / gênero / suporte / domínio discursivo				
***Corpus* de:**	**Diamesia**	**Gênero**	**Domínio discursivo**	**Suporte**
Músicas pomeranas		**X**		
Receitas Culinárias		**X**		
Cartas Pessoais		**X**		
Livros de Registros Eclesiásticos				**X**
Livros impressos em papel, doc ou PDF				**X**
Documentários		**X**		
Artigos acadêmicos		**X**		
Inscrições das lápides dos túmulos				**X**
Corpus do jornal Folha Pomerana				**X**
Literários: poemas, versos, rimas e trovinhas		**X**		
Wandschoener-inscrições em tecidos, paredes, tábuas e quadros				**X**
Entrevistas	**X**			
Textos religiosos			X (discurso religioso)	
Diversos textos de Internet*: artigos, matérias jornalísticas, reportagens.				**X**
Observação: *Entendemos que a página da internet é o suporte onde o texto é fixado, mas também entendemos que a internet é ao mesmo um veículo, pois é por meio da rede que a informação é transmitida. E é também um serviço, pois sem o serviço de internet não temos acesso aos dados. Na internet o lugar onde o texto é "fixado" pode mudar.				

Fonte: Elaboração própria.

Ressalvamos que nossos *corpora* não são balanceados, pois não conseguimos encontrar a mesma quantidade de textos dos mesmos gêneros e tipos. Por meio da experiência que adquirimos com a nossa pesquisa, acreditamos que elaborar um *corpus* de pomerano balanceado não é uma tarefa factível.

Conforme já mencionamos, ampliamos a gama de opções a fim de reunir o maior número possível de textos e lexias[63] (palavras soltas) em pomerano, para o resgate do léxico dessa variedade.

4.4 Recursos

Chamamos de recursos aos *softwares* e utilitários que nos serviram de ferramentas de trabalho, para a compilação, organização e estudo dos dados.

Os recursos adotados neste trabalho foram o *OmniPage*®, o *ArcGis*® e o *WordSmith Tools*®. Também utilizamos pontualmente um extrator de legendas, o *SubExtractor*®, versão de 2013, com o auxílio de um colaborador, para compilar as legendas dos documentários pomeranos.

Para compilar textos de livros, utilizamos o conversor de imagem em texto, ou seja, um programa reconhecedor de caracteres ópticos - OCR (*Optical character recognition*), o *OmniPage Professional*, versão 17.0 de 2009. O programa foi criado em 1995 pela *Nuance Communications*.

O reconhecimento de caracteres ópticos é o processo de extração de texto a partir de imagens. A imagem pode resultar de escaneamento de um documento em papel ou pode ser aberta de um arquivo de imagem eletrônica. O texto se torna editável no computador. De acordo com o manual do programa, o arquivo não precisa ser redigitado manualmente. Porém, observamos que é necessária uma revisão atenciosa, pois acontece a desconfiguração de algumas letras em caracteres, como, por exemplo, o *å* (a-*Ablaut*) e o *ß* (*Eszett*).

Esse procedimento foi utilizado para otimização de tempo na digitação de livros inteiros e também na compilação dos arquivos de registros eclesiásticos. Nem todos os arquivos puderam ser tratados por este programa, porque estavam em caracteres de escrita gótica cursiva. Nesse caso, fizemos os "*prints*" das telas, colamos no documento de texto com formatação (o *Word*), fomos ampliando a visualização, observando e digitando os textos. Para reconhecer algumas letras utilizamos um material didático de paleografia gótica (STEMMER, 2008) como recurso auxiliar.

---

[63] Definimos, para o contexto deste trabalho, lexema segundo Bidermann: "Termos como palavra e vocábulo da linguagem comum se prestam a equívocos e imprecisões. Por essa razão os linguistas cunharam o termo lexema para designar a unidade léxica abstrata em língua. [...] Os lexemas se manifestam, no discurso, através de formas ora fixas, ora variáveis. Essa segunda alternativa é a mais frequente nas línguas flexivas e aglutinantes. [...] A essas formas que aparecem no discurso, daremos o nome de lexia. [...] Dessa forma evitamos ambiguidades e imprecisões inerentes aos termos palavra e vocábulo, [...] Paralelamente, vamos contrastar o termo Léxico (= acervo dos lexemas de uma língua) a Vocabulário (= Conjunto das lexias registradas na obra de um autor, por exemplo)" (BIDERMAN, 2001, p.169-170).

Para realizar o mapeamento das localidades específicas em que existe a presença de descendentes de pomeranos no Brasil, utilizamos o *ArcGis* (2013, versão 10.2), desenvolvido pela ESRI (*Environmental Systems Research Institute*).

O *ArcGis* é um *software* de geoprocessamento, ou melhor, um sistema de informação geográfica que permite a elaboração e manipulação de mapas, de modo que o usuário pode escolher uma base cartográfica e marcar as localidades desejadas. Esse programa foi muito útil para nós, pois a metodologia da SGL implica a identificação dos pontos de pesquisa em mapas (ou cartogramas, na linguagem da SGL). A identificação das localidades é realizada para documentar a variação dos fenômenos linguísticos no espaço geográfico.

Como recurso para a LC, utilizamos o WST (2015, versão 6.0), criado por Scott em 1996, publicado pela *Oxford University Press*. O WST é um conjunto integrado de programas que desempenha várias funções, dentre elas, geração de listas de palavras, concordanciador, etc. É definido por seu criador como um *software* de análise lexical. O WST é muito popular entre os linguistas de *corpus*. Atualmente, já está na sua versão 7.0 e completa 20 anos em 2016.

O WST é uma opção que atende às necessidades da nossa pesquisa. Por meio de suas ferramentas é possível testar os *corpora* e gerar listas de palavras, além de visualizar as linhas de concordâncias do pomerano, com muitos nódulos de uma só vez e em uma forma de visualização vertical.

Neste trabalho, o WST foi utilizado para testar se os *corpora* eram legíveis pelo *software* durante o processo de compilação e transcrição. Também foi adotado como ferramenta de análise, embora não tenhamos nos comprometido em fazer análises, pois o objetivo era a compilação dos *corpora*. Contudo, esboçamos algumas análises a partir dos *corpora* do pomerano carregados no WST. Trataremos sobre isso no capítulo referente à análise dos resultados.

4.5 Listagem

Elaboramos também um arquivo em doc que constitui uma listagem de origens dos dados coletados em diamesia escrita. Sobre os dados escritos, informamos na listagem a origem mais detalhada dos textos do que no cabeçalho dos arquivos. A lista dos textos que compõem o conjunto dos *corpora* escritos está disponível nos apêndices deste trabalho (APÊNDICE D).

Não elaboramos uma lista dos dados orais (coletados em diamesia falada e transcritos), pois poderíamos apenas citar a codificação que criamos para nomenclatura dos arquivos, conforme já exposta nos Quadros (3 e 4) e Figuras (17 e 18) citados no item 4.3.1, visto que as

determinações do CEP impedem a divulgação de quaisquer dados que possam colocar em risco a identificação dos sujeitos-participantes da pesquisa.

4.6 Arquivamento de segurança do conjunto dos *corpora*

Para arquivamento do conjunto dos *corpora* já reunidos e organizados, realizamos o procedimento de fazer *upload* de todos os arquivos e salvá-los em um repositório virtual. Decidimos "subir" os dados *online*, pois o tamanho total do conjunto dos dados, incluindo algumas fotos de fontes originais, somou 53,4 GB e não caberia em um CD-ROM. Esse procedimento também foi adotado como medida de segurança para evitar a perda dos dados.

O conjunto dos *corpora* do pomerano está disponível na rede da Internet e pode ser acessado mediante autorização, pois os dados ainda não foram validados, e mediante o uso de uma senha.

Após realizar a descrição detalhada dos procedimentos metodológicos que desenvolvemos para a compilação dos *corpora* do pomerano, apresentaremos, no próximo capítulo, quais foram os resultados que alcançamos.

5. RESULTADOS

Neste capítulo, apresentamos os resultados que alcançamos, decorrentes da aplicação dos nossos procedimentos metodológicos e de todo o processo da pesquisa. Primeiramente, tratamos do resultado principal do nosso trabalho: a compilação e constituição de *corpora* do pomerano. Em seguida, descrevemos os resultados mais específicos: os *corpora* escritos individualmente e também o *corpus* oral. E por último, avaliamos se alcançamos nosso objetivo geral e nossos dois objetivos específicos.

5.1 Resultado Geral: A constituição de um conjunto de dados a partir dos *corpora* do pomerano

O nosso primeiro resultado é também o mais importante, pois é o nosso principal produto de pesquisa, os *corpora* do pomerano. Conseguimos constituir um conjunto de dados linguísticos autênticos do pomerano no Brasil. Compomos um conjunto de *corpora*. Falamos em conjunto, porque nos referimos aos *corpora* escritos e ao *corpus* oral como uma unidade. Instituímos um nome para o nosso conjunto de *corpora* do pomerano: *Pommersche Korpora* – PK, que é a união dos *corpora* escritos do pomerano, nomeados individualmente como *Pommersche Korpora Escritos* – PKE e do *corpus* oral do pomerano, nomeado individualmente como *Pommersch Korpus Oral – PKO*.

Adotamos o singular por se tratar de um todo, um conjugado, o *Pommersche Korpora* (o conjunto dos *corpora* do pomerano), e grafamos o nome dos *corpora* com K, para remeter à forma alemã, conforme grafado pelo IDS – *Institut für Deutsche Sprache* (Instituto para a língua alemã).

Criamos também um símbolo representativo para identificar o *Pommersche Korpora* (doravante PK), um logotipo, apresentado pela Figura 19.

Figura 19: Logotipo do *Pommersche Korpora.*

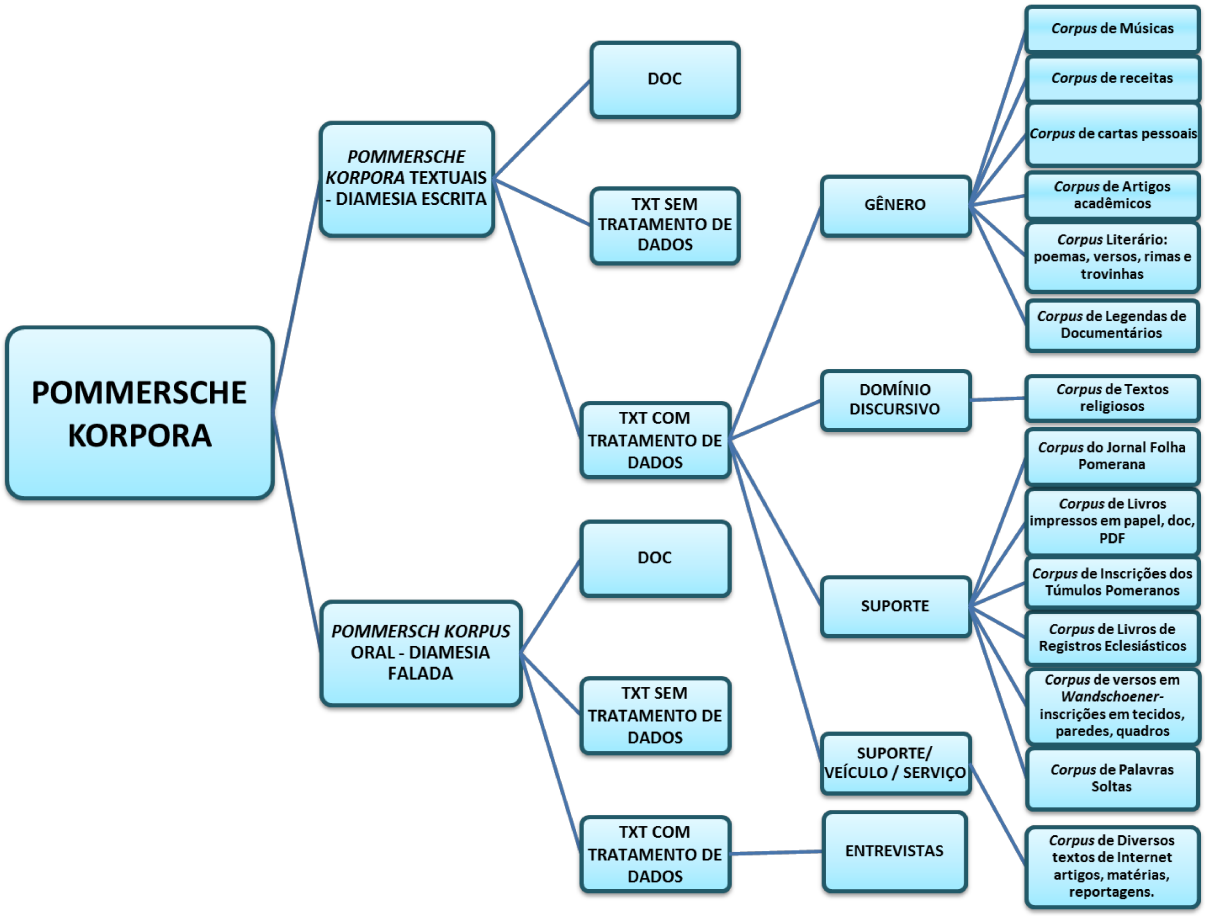


Fonte: Elaboração própria.

O logotipo do *Pommersche Korpora* contém as letras iniciais do seu nome “P” e “K”, a bandeira da Pomerânia e duas formas de escrita da lexia “sim” em pomerano: *Jå* e *Jo*.

Também consideramos necessário obter uma forma de visualização geral dos tipos de arquivos que compõem os nossos *corpora* do pomerano, dada sua diversidade, bem como para representar o modo como sua estrutura está armazenada no computador. Para obtermos essa “visão do todo”, elaboramos um organograma do PK, conforme apresentamos na Figura 20, que se segue.

Figura 20: Organograma com o detalhamento da estrutura do *Pommersche Korpora*.



Fonte: Elaboração própria.

Conforme verificamos acima, o organograma permite observar como os *corpora* foram separados, organizados e desenhados dentro do conjunto, ou seja, a arquitetura "virtual-concreta" do PK.

5.2 Resultado da coleta de diversas fontes de dados pomeranos

De um total de 17 *corpora* diferentes que pretendíamos compilar, conseguimos atingir apenas 15 deles, incluído o *corpus* oral. Sendo que um *corpus* não foi constituído por haver apenas um arquivo/texto do mesmo gênero – o Diário, e o outro porque nem sequer conseguimos obter algum material – os *scripts* de rádio. Segue abaixo a listagem dos *corpora* que conseguimos compilar e constituir, com seus respectivos nomes individuais em português:

1. *Corpus* de Inscrições dos Túmulos Pomeranos – CITP
2. *Corpus* de Livros de Registros Eclesiásticos – CLRE
3. *Corpus* de Cartas Pessoais – CCP
4. *Corpus* de Receitas Culinárias – CRC
5. *Corpus* do Jornal Folha Pomerana – CJFP

6. *Corpus* de Textos Diversos da Internet – CTDI
7. *Corpus* de Legendas de Documentários – CLD
8. *Corpus* de Trabalhos Acadêmicos – CTA
9. *Corpus* de Textos Religiosos – CTR
10. *Corpus* de Músicas Pomeranas – CMP
11. *Corpus* Literário Pomerano (poemas, contos, versos, rimas e trovinhas) – CLP
12. *Corpus* de Livros – CL
13. *Corpus* de *Sprüche* Diversos (*Corpus* de fraseologismos diversos) – CSD
14. *Corpus* de Palavra Soltas – CPS
15. *Corpus* Oral de Entrevistas Interativas – COEI (o PKO)

Cada um desses *corpora* será descrito nos tópicos 5.3 e 5.4 e serão feitas breves observações e análises sobre eles.

5.3 Os *Pommersche Korpora Escritos* como resultados da primeira fase de compilação

Neste tópico, vamos descrever detalhadamente o que é cada *corpus* que compõe os *corpora* escritos do PK; os *Pommersche Korpora Escritos*, doravante PKE.

5.3.1 *Corpus* de Inscrições dos Túmulos Pomeranos – CITP

O *Corpus* de Inscrições dos Túmulos Pomeranos – CITP é um conjunto de arquivos de textos constituídos a partir dos dados coletados e compilados por meio de sessões de fotografias nos pontos de pesquisa Vila Neitzel – VN/MG, Vera Cruz – VC/RS, Arroio do Tigre – AT/RS, Ponte Rio Pardinho – RP/RS (cemitério no Vale do Rio Pardo) e em Santa Maria de Jetibá – SMJ/ES. Todos os cemitérios onde coletamos dados são cemitérios evangélicos (luteranos) de imigrantes alemães, incluindo pomeranos e seus descendentes[64].

[64] Decidimos coletar inscrições somente nos túmulos evangélicos, pois os católicos eram minoria e não justificaria o deslocamento para outros cemitérios fora de nossa rota. Segundo apontam as pesquisas de Seibel (2010), "os católicos eram minorias e terminaram sendo absorvidos pelos pomeranos" (SEIBEL, 2010, p. 113; 167; 197; 205; 249), no sentido de se converterem à religião evangélica e participarem das igrejas luteranas. Como a quantidade de cemitérios católicos pomeranos seria potencialmente menor, com base nas pesquisas do autor mencionado, então avaliamos que não seria muito produtivo o nosso deslocamento em busca de cemitérios católicos, além dos custos financeiros para sairmos de nossa rota e irmos às outras localidades.

Os cemitérios abrangidos foram identificados por siglas e suas respectivas etiquetas são: o cemitério evangélico localizado na zona rural norte de Itueta/MG, próximo à Vila Neitzel <CEVN/MG>; o cemitério evangélico de Vera Cruz/RS <CEVC/RS>; o cemitério evangélico de Arroio do Tigre/RS <CEAT/RS>; o cemitério evangélico de Ponte Rio Pardinho/RS <CERP/RS>; e o Cemitério Evangélico de Santa Maria de Jetibá/ES <CESMJ/ES>.

Esse *corpus* é um conjunto de arquivos que contém inscrições encontradas originalmente no suporte "lápide" ou em outras partes do túmulo e foram captadas por fotografia, transformando-se em material pictórico; consistem em textos curtos da área semântica fúnebre de diversos gêneros, como versículos bíblicos, mensagens pessoais, adágios, dísticos (séries de até dois versos rimados), *Sprüche* (sentenças, máximas) e fraseologismos diversos.

A prática da escolha de frases para deixar gravadas nos túmulos, pelo sujeito a ser futuramente enterrado ou pelos seus familiares, é uma prática sociocultural pomerana relacionada ao modo como representam a morte. Para os pomeranos é importante deixar uma inscrição que remeta ao modo como o sujeito ali enterrado viveu e também que expresse suas crenças na ocorrência de um reencontro com os familiares e na ressurreição. Algumas inscrições representam mais a ideia da continuidade da vida do que a finalidade da morte.

O total de transcrições somam 124 lápides, contendo todas as inscrições dos túmulos. Alguns túmulos possuem inscrições não só nos lados frontais, mas também nos versos e laterais.

No que se refere à linha de tempo que envolve todo o *corpus* de Inscrições dos Túmulos Pomeranos, podemos informar que a lápide mais antiga remonta ao ano de 1815 e a mais recente ao ano de 1990.

Quanto à quantificação desse *corpus*, realizamos a contagem dos itens que somaram 1.844 palavras corridas (*tokens* ou *running words*[65]) e 640 palavras distintas (*types* ou *distinct words*) em contagem realizada pelo WST, conforme é possível visualizar por meio da Figura 21.

---

[65] Consideramos aqui como *running words* todas as palavras corridas usadas para *Wordlist*, mas não todos os *Tokens in text*, pois observamos que a contagem de *running words in text* inclui os números presentes nos textos. Primeiramente, refizemos manualmente o cálculo *da type/token ratio* e observamos que o valor só bate com o cálculo do WST se considerarmos os *Tokens used for wordlist*. Em seguida, testamos o dado representado na aba *statistics* por "#" e percebemos que representam os números no *corpus*, assim, a diferença entre os *Tokens in text* e *os Tokens used for Wordlist* é justamente a quantidade dos números quantificada na linha "#'. Assim, consideramos, para efeito deste trabalho, que as palavras corridas nos nossos dados, chamadas de *Tokens*, são aquelas contabilizadas para a *Wordlist*, excetuando-se os números.

Figura 21: Estatísticas do *Corpus* de Inscrições dos Túmulos Pomeranos.

WordList

File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	ype/toke atio	standard TTR	STTR std.dev.	STTR basis
1	Overall	35.166	2.143	1.844		640	34,71	90,47	10,51	10
2	LC-SA-AT-PKT-P.txt	13.816	783	654		198	30,28	88,46	10,89	10
3	LC-SA-RP-PKT-P.txt	2.794	158	136		95	69,85	88,00	14,83	10
4	LC-SA-SMJ-PKT-P.txt	5.822	363	324		206	63,58	93,33	6,65	10
5	LC-SA-VC-PKT-P.txt	8.136	523	457		212	46,39	91,73	8,12	10
6	LC-SA-VN-PKT-P.txt	4.598	316	273		165	60,44	91,29	6,98	10

Fonte: Elaboração própria.

Na imagem a seguir, Figura 22, podemos observar duas lápides de túmulos pomeranos, uma com o *toofel*, chamado assim quando a inscrição é feita em uma tábua de madeira, e outra com inscrições gravadas em uma lápide de cimento.

Figura 22: Exemplos de inscrições em lápides pomeranas em Itueta/MG e Vera Cruz/RS.



Fonte: Fotos de cemitérios pomeranos. Arquivo do autor.

Inserimos abaixo um exemplo de texto transcrito de um túmulo pomerano, localizada no verso de uma lápide, conforme a Figura 3, acima, no lado direito: "*Nun ruhet Jhr in Frieden, geliebte Eltern Jhr, die Jhr so langgeschieden in Gott es Frieden hier*".

Essa mensagem pessoal tem características de uma sintaxe que parece vir da oralidade, e pode ter sido gravada na lápide conforme a fala, pois seu ordenamento foge às normas do alemão-padrão. Inferimos isso, porque sua constituição torna difícil a detecção do sentido do ponto de vista do HD e também porque contém lexias que são pronunciadas de forma diferente na fala pomerana, conforme nossa experiência e testes que realizamos. Nesse caso, a escrita não precisa ser necessariamente diferente para ser considerada pomerana e confirmamos que

trata-se de um túmulo pomerano. Assim, fizemos uma tradução, a mais aproximada possível, para o português: “Agora descansem em paz, amados pais, vós embora tão longamente separados, tendes aqui em Deus vossa paz”.

Esse *corpus* apresentou traços dialetais como escritas com “J” onde no alemão-padrão seria “i”, por exemplo, *jst* e também palavras escritas juntas que no alemão-padrão seriam escritas separadas, além de ser possível perceber que algumas palavras, embora escritas iguais ao alemão-padrão, possuem pronúncia diferente em pomerano.

À guisa de exemplificação, inserimos abaixo a Figura 23, na qual é possível visualizar alguns itens *Hapax Legomena* (com uma única ocorrência) que representam casos de palavras que em pomerano são escritas juntas e possuem pronúncia diferente, como *unvergeslichengeschieden*.

Figura 23: Itens *Hapax Legomena* do *Corpus* de Inscrições dos Túmulos Pomeranos.

561	UNSEUERN	1	0,05	1	16,67
562	UNSRE	1	0,05	1	16,67
563	UNSTERBLICH	1	0,05	1	16,67
564	UNTER	1	0,05	1	16,67
565	UNVERGESLICHENGESCHIEDEN	1	0,05	1	16,67
566	UNVERGESLICHER	1	0,05	1	16,67
567	UNVERGESSLICHE	1	0,05	1	16,67
568	UNVERGESSLICHER	2	0,09	2	33,33

Fonte: Elaboração própria.

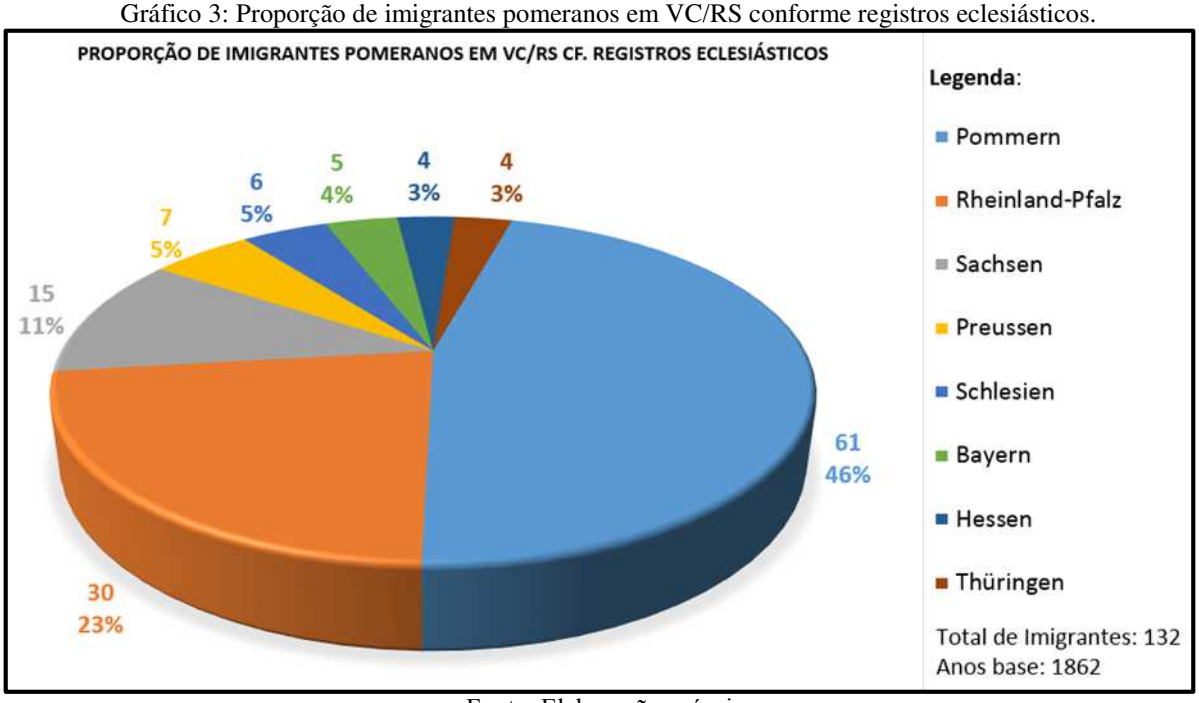
Em geral, os textos presentes no *Corpus* de Inscrições dos Túmulos Pomeranos são simbólicos, muito pessoais e muito representativos para que fossem escritos em uma língua e grafia que não representasse os sujeitos ali enterrados. De modo que, para nós, as palavras escritas iguais ao alemão-padrão não descaracterizam essas inscrições enquanto pomeranas e são, de fato, dados provenientes de túmulos pomeranos.

5.3.2 *Corpus* dos Livros de Registros Eclesiásticos – CLRE

O *Corpus* dos Livros de Registros Eclesiásticos – CLRE é um conjunto de arquivos que contém os dados paroquiais dos imigrantes e seus descendentes, provenientes dos arquivos das igrejas luteranas. É um acervo documental em formato de livros digitalizados que foram compilados. O recorte temporal abrangido nos registros vai de 1854 a 1873.

O material contém dados dos imigrantes recém-chegados à região, informando inclusive a origem de alguns imigrantes, de modo que é possível comprovar a presença maciça de pomeranos e seus descendentes na região de Vera Cruz/RS. Para organizar esses dados, elaboramos um gráfico (Grafico 3) que indica a maioria pomerana na região:

Gráfico 3: Proporção de imigrantes pomeranos em VC/RS conforme registros eclesiásticos.



Fonte: Elaboração própria.

No gráfico acima é possível perceber que os imigrantes pomeranos eram maioria na localidade de Vera Cruz em 1862, no total de 61 pomeranos registrados, a maioria, em 46% dos casos. Observamos que sete imigrantes foram registrados apenas como prussianos, e dentre eles poderiam haver 100% de pomeranos o que resultaria numa maioria de 51% ou 68 imigrantes provenientes da província pomerana da Prússia, calculados sobre uma amostragem total de 132 imigrantes.

Quanto ao conteúdo presente nos registros, podemos encontrar as palavras do cotidiano, pois eles contêm informações sobre batismos, falecimentos, cidades de procedência, paternidade, idade, casamentos, confirmações de fé, causas de óbito, nomes de doenças, nome de famílias, etc. Sobre os dados antroponímicos, como o nome dos noivos e dois pais dos noivos, por exemplo, é possível verificar os casamentos mistos, dentre muitos outros estudos que podem ser realizados.

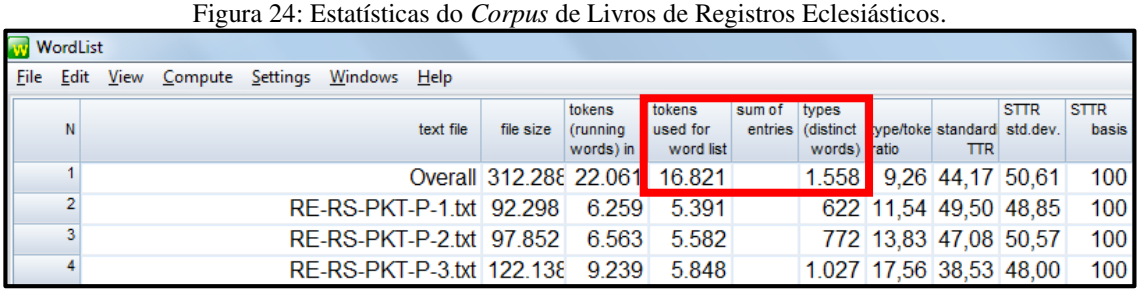
Esclarecemos que consideramos os dados em pomerano, como os nomes das famílias e das localidades da Pomerânia de onde eles vieram, que editamos o material excluíndo ou

etiquetanto (para não contabilizar) dados não-pomeranos e que, ainda assim, consideramos a existência de interferências dialetais no alemão-padrão da época.

Notamos, em uma breve busca pela *Wordlist* do *corpus*, que a família de sobrenome Neitzel de Minas Gerais também é uma família que aparece nos dados do Rio Grande do Sul (posição[66] 168, com 16 ocorrências), fato inferido pela presença do mesmo sobrenome nos registros. Alguns sobrenomes pomeranos comuns que aparecem nesse *corpus* são, por exemplo, Püttelkow (variação de Pitelkow), Raddow, Rakow, Ramelow, etc.

Quanto à quantificação desse *corpus*, a contagem dos itens somou 16.821 palavras corridas e 1.558 palavras distintas em contagem realizada pelo WST, conforme é possível visualizar por meio da Figura 24.

Figura 24: Estatísticas do *Corpus* de Livros de Registros Eclesiásticos.



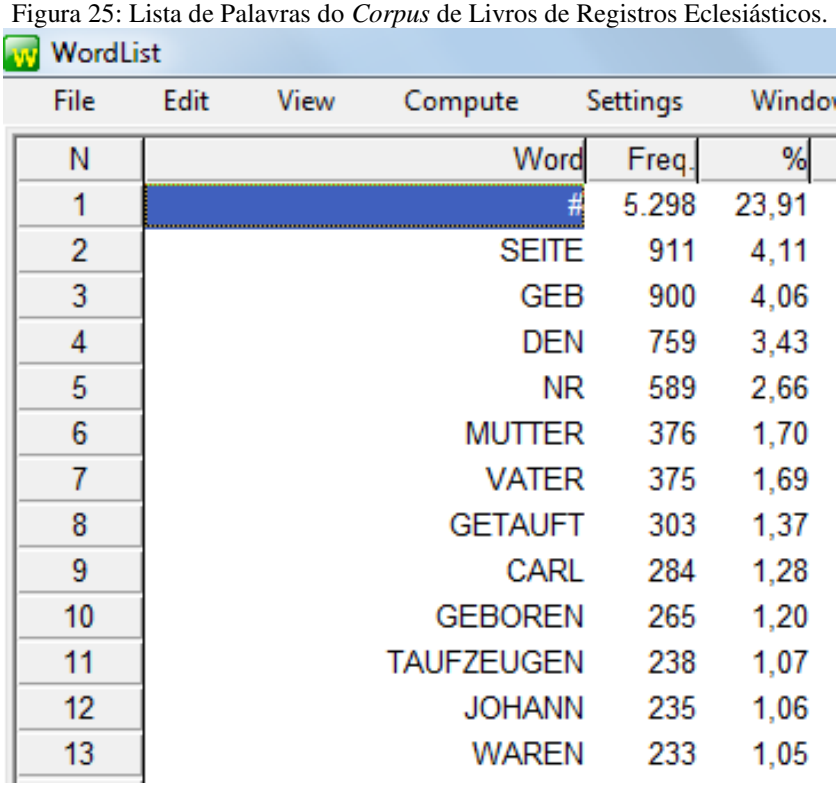
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standardized TTR	STTR std.dev.	STTR basis
1	Overall	312.288	22.061	16.821		1.558	9,26	44,17	50,61	100
2	RE-RS-PKT-P-1.txt	92.298	6.259	5.391		622	11,54	49,50	48,85	100
3	RE-RS-PKT-P-2.txt	97.852	6.563	5.582		772	13,83	47,08	50,57	100
4	RE-RS-PKT-P-3.txt	122.138	9.239	5.848		1.027	17,56	38,53	48,00	100

Fonte: Elaboração própria.

Com um olhar qualitativo superficial sobre os dados desse *corpus*, acreditamos que esse índice é plausível, pois, ao passo que o *corpus* apresenta muitas palavras distintas, como sobrenomes de famílias diferentes e localidades de origens diferentes, também há muita repetição das informações básicas, como a palavra "*Seite*", por exemplo, a mais frequente do *corpus*, com 911 ocorrências; a qual indica a página do livro de registros. E também o verbo "*waren*", verbo muito frequente no *corpus*, com 233 ocorrências. Nesse contexto significa "foram", utilizado para indicar quem foram os padrinhos no caso dos batismos (dados expostos na Figura 25, a seguir). Lembramos que *Waren* também é o nome de uma cidade da *Vorpommern* (Pomerânia anterior).

[66] Posição dentro do *corpus* por ordem de frequência na lista de palavras do WST. Doravante, sempre que indicarmos a posição no *corpus* será no mesmo sentido aqui exposto.

Figura 25: Lista de Palavras do *Corpus* de Livros de Registros Eclesiásticos.



WordList

File Edit View Compute Settings Windo

N	Word	Freq.	%
1	#	5.298	23,91
2	SEITE	911	4,11
3	GEB	900	4,06
4	DEN	759	3,43
5	NR	589	2,66
6	MUTTER	376	1,70
7	VATER	375	1,69
8	GETAUFT	303	1,37
9	CARL	284	1,28
10	GEBOREN	265	1,20
11	TAUFZEUGEN	238	1,07
12	JOHANN	235	1,06
13	WAREN	233	1,05

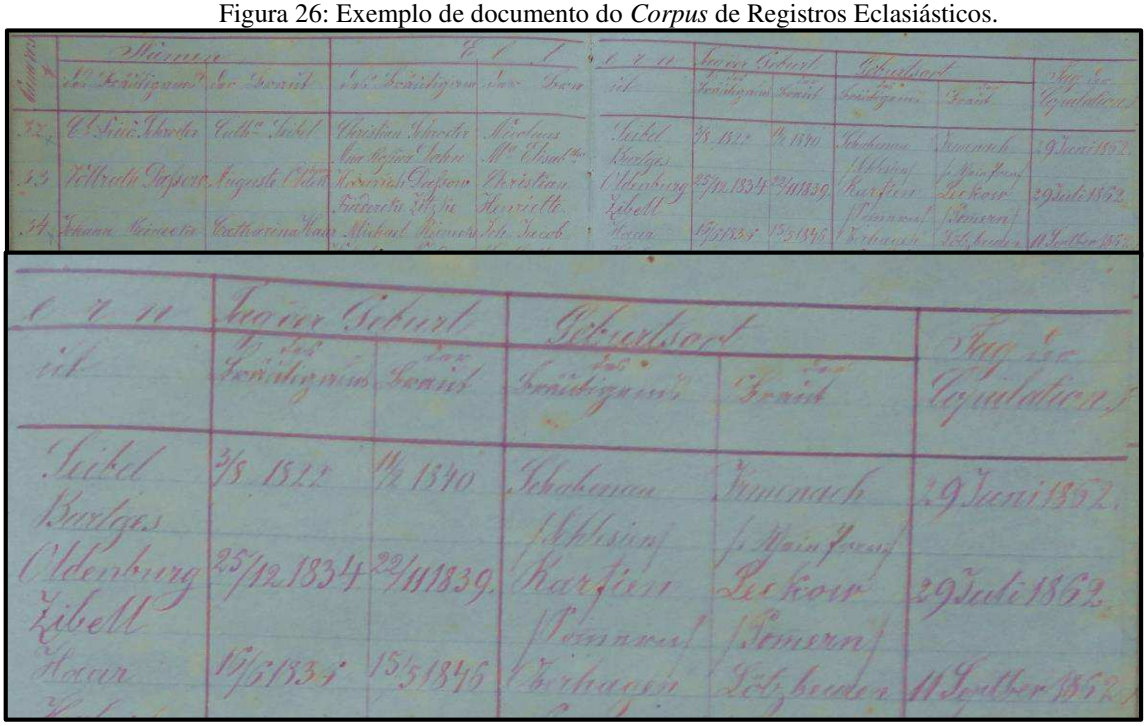
Fonte: Elaboração própria.

Nos dados do *Corpus* de Registros Eclesiásticos, incluindo dados das antigas comunidades de Ferraz e Vila Theresa, é possível obter informações sobre a origem dos imigrantes dentro da Pomerânia, checar se eles eram procedentes da Pomerânia ocidental ou oriental, comparar com a origem dos pomeranos de outras regiões como Vila Neitzel/MG, etc. O *corpus* de registros também possui potencial para permitir a realização de estudos toponímicos dos locais de procedência dos imigrantes, pois na região da ex-Pomerânia oriental, muitas localidades têm hoje nomes poloneses.

Um fato que observamos no *corpus* foi que vieram pomeranos de *Natzmersdorf* para o sul, assim como vieram para o estado capixaba (consulta realizada nos registros de entrada do porto de Santos *online* e no museu de Santa Maria de Jetibá/ES); muitas famílias da mesma localidade de origem são encontradas em diferentes regiões do Brasil, e essas comparações podem ser feitas, por exemplo, utilizando-se dos registros eclesiásticos. Assim, entendemos que pomeranos que partiam de mesmas localidades na Pomerânia, assentavam-se no Brasil em regiões diferentes.

Na imagem a seguir, Figura 26, podemos observar um exemplo do *Corpus* de Registros Eclesiásticos:

Figura 26: Exemplo de documento do *Corpus* de Registros Eclasiásticos.



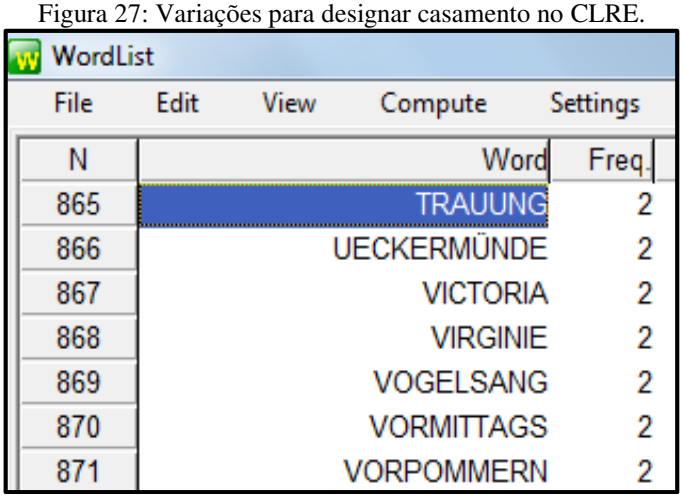
Fonte: Foto do Registro Eclesiástico de VC/RS. Arquivo do autor.

Exemplificaremos a transcrição[67], devido à imagem estar em um tamanho e uma cor de difícil visualização. O registro 33, na segunda linha da imagem sobreposta é: <Registro nº 33> "*Nummer 33. Namen: Des Bräutigams Vollrath Dassow; Der Braut Auguste Oldenburg. Eltern: Des Bräutigams Heinrich Dassow Friederike Zitzke; Der Braut Christian Oldenburg Henriette Zibell. Tag der Geburt: Des Bräutigams 25. Dezbr. 1834; Der Braut 22. Novbr.1839. Geburtsort: Des Bräutigams Rarfin (Kreis Belgard) – Pommern; Der Braut Leckow (Kreis Belgard) – Pommern. Tag der Copulation 29. Juli 1862*".

Outro fato notável é que os substantivos para casamento dicionarizados e comumente referidos em alemão-padrão, como *Heirat* e *Hochzeit*, apresentam variações menos comuns no *corpus*, como *Trauung* (posição 865, com 2 ocorrências) e *Ehrentrau* (posição 1.012, com 1 ocorrência). Na Figura 27, a seguir, podemos observar um exemplo:

[67] Baseamo-nos no trabalho de transcrição realizado por Migotto (2011) dos registros de Vera Cruz/RS, porém refizemos grande parte do trabalho para a nossa pesquisa, devido a algumas reticências e erros encontrados, como, por exemplo, no sobrenome Reddatz, em que o correto é Raddatz.

Figura 27: Variações para designar casamento no CLRE.



N	Word	Freq.
865	TRAUUNG	2
866	UECKERMÜNDE	2
867	VICTORIA	2
868	VIRGINIE	2
869	VOGELSANG	2
870	VORMITTAGS	2
871	VORPOMMERN	2

Fonte: Elaboração própria.

Consideramos que o *Corpus* dos Livros de Registros Eclesiásticos de Vera Cruz contém dados ricos sobre os pomeranos, ainda que tenham sido escritos em alemão-padrão, com presença de exceções, o qual naquela época apresentava traços dialetais que o difere do alemão-padrão de hoje. Ressalvamos, novamente, que há dados toponímicos e antroponímicos pomeranos e em pomerano nesse *corpus*.

5.3.3 *Corpus* de Cartas Pessoais – CCP

O *Corpus* de Cartas Pessoais – CCP é um conjunto formado por 9 arquivos que contém 9 cartas pomeranas, uma em cada arquivo. São missivas pessoais que os pomeranos escreviam para seus familiares, contendo assuntos diversos sobre a vida, a família, as plantações, seus sentimentos em relação ao pomerano e ao Brasil, etc. Um estudo específico sobre o conteúdo desse *corpus* poderia revelar os tipos textuais presentes e predominantes nas cartas como, por exemplo, do tipo narrativo, descritivo, informativo, etc.

As cartas foram coletadas na zona rural de localidades do Rio Grande do Sul e Minas Gerais, como Canguçu e Itueta, por exemplo. Todas as cartas estavam originalmente no formato manuscrito em papel. Aparentemente, seus temas principais giram em torno de assuntos pessoais, familiares e também do cotidiano no campo. No que se refere à linha de tempo que envolve todo o *Corpus* de Cartas Pessoais, a carta mais antiga data de 1923 e a mais recente data de 2014. Não foi possível identificar o recorte cronológico de todas as cartas.

Quanto à quantificação desse *corpus*, a contagem dos itens somaram 1.762 palavras corridas e 763 palavras distintas, conforme é possível visualizar por meio da Figura 28.

Figura 28: Estatísticas do *Corpus* de Cartas Pessoais.

WordList

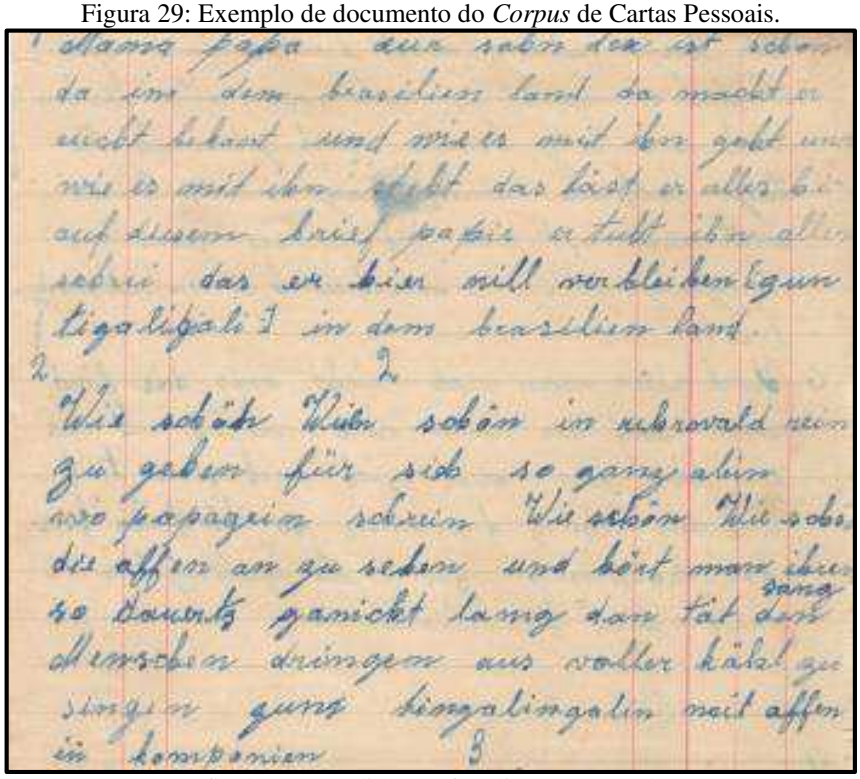
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	ype/toke atio	standard TTR	STTR std.dev.	STTR basis
1	Overall	22.584	1.77	1.762		763	43,30	94,57	5,30	10
2	CP-AWD-LS-PKT-DT.txt	4.312	335	330		174	52,73	95,15	5,48	10
3	CP-FHT-C-PKT-DT 1.txt	4.384	370	366		200	54,64	94,59	5,03	10
4	CP-FHT-C-PKT-DT 2.txt	2.258	178	176		125	71,02	96,47	5,44	10
5	CP-FHT-C-PKT-DT 3.txt	2.658	229	229		111	48,47	90,45	11,27	10
6	CP-FHT-C-PKT-DT 4.txt	940	57	57		44	77,19	96,00	5,27	10
7	CP-FHT-C-PKT-DT 5.txt	2.462	193	193		119	61,66	96,32	3,16	10
8	CP-FHT-C-PKT-DT 6.txt	2.618	202	202		128	63,37	95,00	5,79	10
9	CP-FHT-C-PKT-DT 7.txt	1.254	83	83		73	87,95	97,50	2,66	10
10	CP-MHT-C-PKT-DD.txt	1.698	126	126		91	72,22	91,67	6,73	10

Fonte: Elaboração própria.

Na imagem a seguir, Figura 29, podemos observar um exemplo do *Corpus* de Cartas Pessoais:

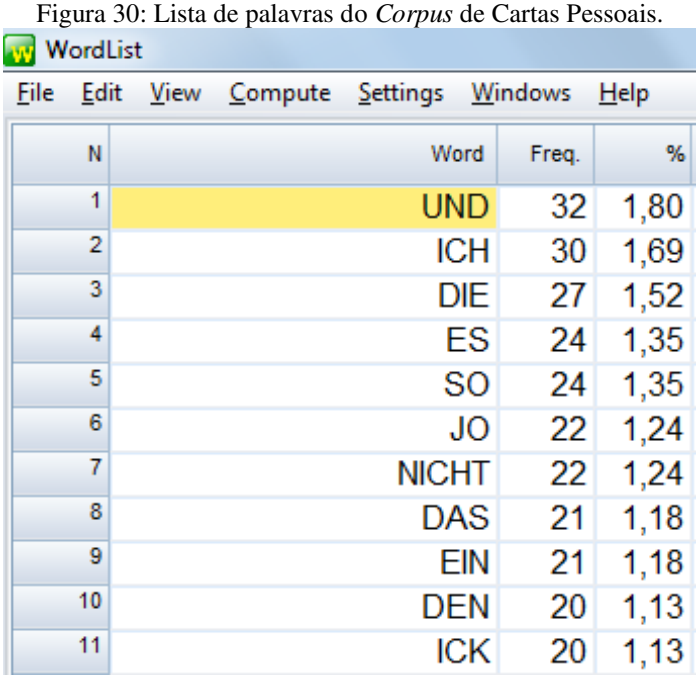
Figura 29: Exemplo de documento do *Corpus* de Cartas Pessoais.



Fonte: Fotografia de carta cedida por família pomerana de Canguçu/RS.

Encontramos casos de variação na escrita que podem representar variações na fala também ou simplesmente uma grafia não-normatizada, como por exemplo, *jetz musz ich* para *jetzt muss ich* (agora eu tenho que) e *biette* para *bitte* (por favor). Podemos visualizar algumas das palavras mais frequentes no CCP, por meio da Figura 30, a seguir:

Figura 30: Lista de palavras do *Corpus* de Cartas Pessoais.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	UND	32	1,80
2	ICH	30	1,69
3	DIE	27	1,52
4	ES	24	1,35
5	SO	24	1,35
6	JO	22	1,24
7	NICHT	22	1,24
8	DAS	21	1,18
9	EIN	21	1,18
10	DEN	20	1,13
11	ICK	20	1,13

Fonte: Elaboração própria.

Os itens mais frequentes são *Ich* e sua variação *Ick* (eu), *Familch* e suas variações *Familie* e *Famillian* (Família), *u* abreviação de *und* ou de *un* (e). Encontramos apenas uma interferência do português nas cartas em "*Vovó Elviras Voter hat Sanfona gespielt, ja? Oder war es eine Koncertina?*" (O pai da vovó Elvira tocava Sanfona, né? Ou aquilo era uma concertina?); contudo, encontramos também algumas interferências do alemão-padrão na escrita das cartas pomeranas.

5.3.4 *Corpus* de Receitas Culinárias – CRC

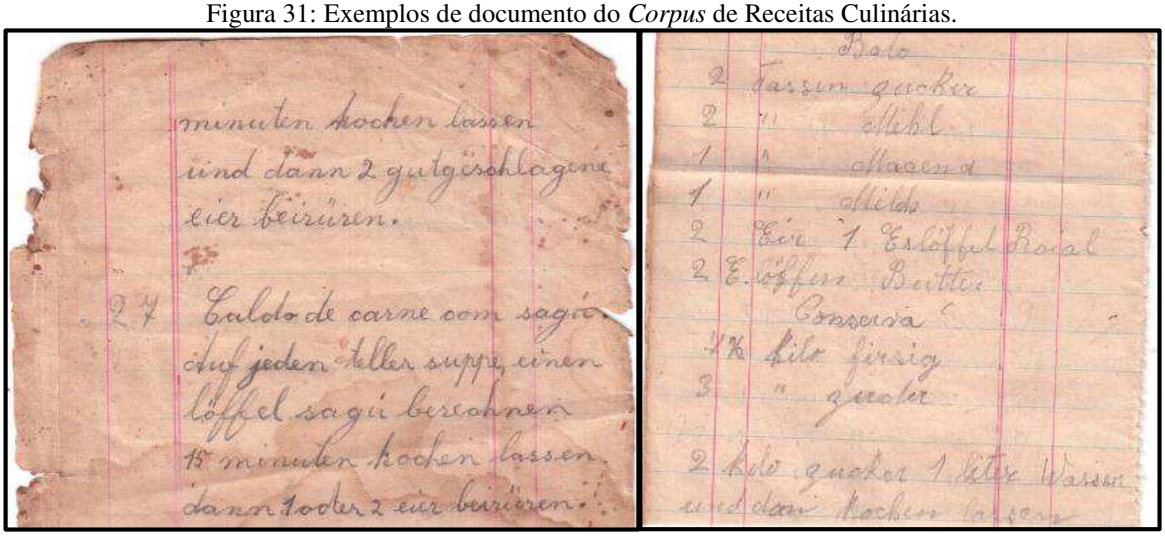
O *Corpus* de Receitas Culinárias – CRC é um pequeno conjunto constituído por 4 arquivos, contendo 9 pequenas receitas em pomerano, que na maioria das vezes indicam apenas a descrição dos ingredientes e quase nunca a descrição do modo de preparo. Foram coletadas no interior do Rio Grande do Sul, em Canguçu, em formato de manuscritos em papel.

As receitas estão divididas em duas receitas de bolo, duas de doce, uma de torta, uma de sobremesa, uma de conserva e uma de prato salgado, sendo que uma receita está incompleta. Tratam-se de receitas da culinária pomerana, alemã e gauchesca.

Não foi possível identificar o recorte cronológico das receitas, pois elas não contêm datas, apenas numeração não-sequencial em algumas delas. As receitas estão em papéis antigos e apresentam desgaste da tinta e das folhas de papel.

Na imagem a seguir, Figura 31, podemos observar um exemplo do *Corpus* de Receitas Culinárias:

Figura 31: Exemplos de documento do *Corpus* de Receitas Culinárias.



Fonte: Fotografias de receitas doadas por família pomerana de Canguçu/RS.

Quanto à quantificação desse *corpus*, realizamos a contagem dos itens que somaram 149 palavras corridas e 81 palavras distintas, conforme é possível visualizar na Figura 32.

Figura 32: Estatísticas do *Corpus* de Receitas Culinárias

WordList

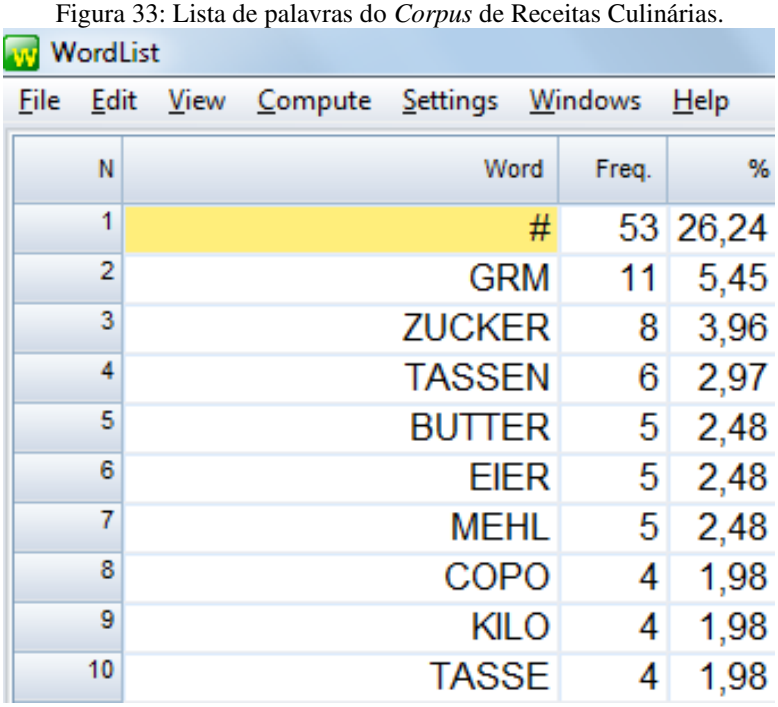
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	3.864	202	149		81	54,36	78,33	22,19	10
2	RC-RS-PKT-E -1.txt	1.160	70	49		36	73,47	75,71	21,37	10
3	RC-RS-PKT-E -2.txt	886	47	35		26	74,29	75,00	24,64	10
4	RC-RS-PKT-E -3.txt	826	47	36		30	83,33	85,00	8,38	10
5	RC-RS-PKT-E -4.txt	992	38	29		27	93,10	80,00	18,93	10

Fonte: Elaboração própria.

Existe variação dos nomes dos ingredientes utilizados para as diferentes receitas que compõem esse *corpus*. As medidas são itens muito frequentes nesse *corpus*, bem como ingredientes mais comuns, como o *Zucker* (açúcar), por exemplo, dentre outros casos possíveis de serem observados na listagem de palavras. Apresentamos a seguir a lista de palavras do CRC, na Figura 33.

Figura 33: Lista de palavras do *Corpus* de Receitas Culinárias.



N	Word	Freq.	%
1	#	53	26,24
2	GRM	11	5,45
3	ZUCKER	8	3,96
4	TASSEN	6	2,97
5	BUTTER	5	2,48
6	EIER	5	2,48
7	MEHL	5	2,48
8	COPO	4	1,98
9	KILO	4	1,98
10	TASSE	4	1,98

Fonte: Elaboração própria.

Encontramos itens com variação na escrita, aparentemente provenientes da interferência linguística, como: *schocolato* para chocolate, *Macena* para maisena, *Dosse* para doce, *Fermente* e *roial* para fermento e também *kocken* que é a forma pomerana para *kochen* (cozinhar). As receitas foram escritas por pomeranas, em pomerano e, algumas vezes, em pomerano-alemão-português numa mesma receita, o que, para nós, é um indício do contato de línguas e da variedade brasileira do pomerano.

5.3.5 *Corpus* do Jornal Folha Pomerana – CJFP

O *Corpus* do Jornal Folha Pomerana – CJFP é um conjunto de arquivos que contém textos de um jornal em formato digital que circula *online* gratuitamente no Brasil. Nele são publicados pequenos textos em pomerano, mas sobretudo matérias jornalísticas diversificadas, em português, sobre os pomeranos, contendo algumas vezes palavras, frases e trechos em pomerano.

Os arquivos com o conteúdo dos jornais estão organizados nesse *corpus* em 3 arquivos em txt, ano I, ano II e ano III, contendo os números que apresentaram algum conteúdo em pomerano. O recorte cronológico abrangido nesse *corpus* é de 19 de agosto de 2013 a 20 de fevereiro de 2016.

Quanto à quantificação desse *corpus*, realizamos a contagem dos itens e somaram 1.969 palavras corridas e 980 palavras distintas, conforme é possível visualizar por meio da Figura

34.

Figura 34: Estatísticas do *Corpus* do Jornal Folha Pomerana.

WordList

File Edit View Compute Settings Windows Help

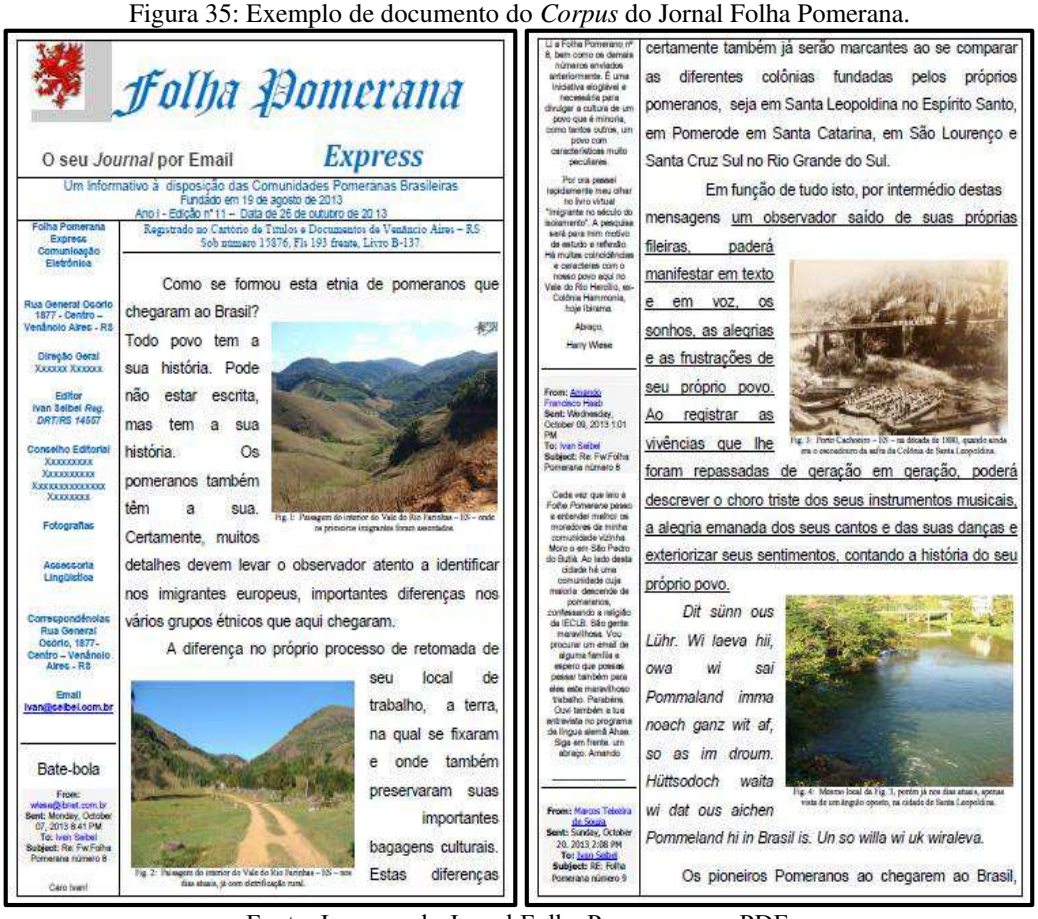
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	ype/toke ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	53.600	2.003	1.969		980	49,77	97,10	2,23	10

Fonte: Elaboração própria.

Trata-se de um *corpus* pequeno, composto de textos curtos inteiros, frases e palavras soltas em pomerano, já que editamos e consideramos apenas o conteúdo escrito em pomerano no jornal.

Na imagem a seguir, Figura 35, podemos observar um exemplo do *Corpus* do Jornal Folha Pomerana, no qual há no lado direito da página, em baixo, um texto em pomerano:

Figura 35: Exemplo de documento do *Corpus* do Jornal Folha Pomerana.



Folha Pomerana

O seu *Journal* por Email **Express**

Um informativo à disposição das Comunidades Pomeranas Brasileiras
Fundado em 19 de agosto de 2013
Ano I - Edição nº 11 – Data de 26 de outubro de 2013

Registrado no Cartorio de Titulos e Documentos de Venâncio Aires – RS
Sob número 15876, Fls 193 frente, Livro B-137

Como se formou esta etnia de pomeranos que chegaram ao Brasil?

Todo povo tem a sua história. Pode não estar escrita, mas tem a sua história. Os pomeranos também têm a sua. Certamente, muitos detalhes devem levar o observador atento a identificar nos imigrantes europeus, importantes diferenças nos vários grupos étnicos que aqui chegaram.

A diferença no próprio processo de retomada de seu local de trabalho, a terra, na qual se fixaram e onde também preservaram suas importantes bagagens culturais. Estas diferenças

certamente também já serão marcantes ao se comparar as diferentes colônias fundadas pelos próprios pomeranos, seja em Santa Leopoldina no Espírito Santo, em Pomerode em Santa Catarina, em São Lourenço e Santa Cruz Sul no Rio Grande do Sul.

Em função de tudo isto, por intermédio destas mensagens um observador saído de suas próprias fileiras, poderá manifestar em texto e em voz, os sonhos, as alegrias e as frustrações de seu próprio povo. Ao registrar as vivências que lhe foram repassadas de geração em geração, poderá descrever o choro triste dos seus instrumentos musicais, a alegria emanada dos seus cantos e das suas danças e exteriorizar seus sentimentos, contando a história do seu próprio povo.

Dit sünn ous Lühr. Wi laeva hii, owa wi sai Pommaland imma noach ganz wit af, so as im droum. Hüttsodoch waita wi dat ous aichen Pommeland hi in Brasil is. Un so willa wi uk wiraleva.

Os pioneiros Pomeranos ao chegarem ao Brasil,

Fonte: Imagem do Jornal Folha Pomerana em PDF.

Podemos observar que as palavras gramaticais são as mais frequentes nesse *corpus*: os artigos *dai* (a ou as, feminino, vem de *die*, *dei*), *De* (o, artigo masculino, vem de *der*) e *dat* (artigo neutro, vem de *das*), estão entre os itens mais frequentes, assim como a preposição *von* (de, também grafada *fon*), conforme podemos observar por meio da Figura 36, a seguir.

Figura 36: Lista de palavras do *Corpus* do J. Folha Pomerana.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	#	34	1,70
2	DAI	28	1,40
3	IN	28	1,40
4	DE	25	1,25
5	A	21	1,05
6	DER	19	0,95
7	DIE	17	0,85
8	DO	16	0,80
9	UND	16	0,80
10	VON	16	0,80

Fonte: Elaboração própria.

Outro item frequente nesse *corpus* é, por exemplo, o substantivo *Lüür* (pessoa, gente), com 10 ocorrências (posição 26) no primeiro ano de publicação da Folha Pomerana. Esse *corpus* merece ser estudado com maior detalhamento, pois constitui uma fonte contemporânea sobre o pomerano.

5.3.6 *Corpus* de Textos Diversos da Internet – CTDI

O *Corpus* de Textos Diversos da Internet – CTDI é um conjunto de 12 arquivos que possui conteúdos de blogs, sites, redes sociais e notícias com textos em pomerano que circulam na rede. Os mesmos foram salvos primeiramente em extensão doc (documento de texto com formatações), editados, tratados e depois armazenados em extensão txt (*plain text* – texto plano).

O recorte cronológico do CTDI abrange de julho de 2009 a agosto de 2015, sendo consideradas as datas das postagens, e quando essas não foram localizadas – consideramos a data de busca das matérias sobre o pomerano na Internet.

O conteúdo vai desde notícias sobre festas pomeranas com o nome das atividades em pomerano até blogs de famílias que postaram seus vocabulários *online*. O conteúdo é bastante diversificado e contém, na maioria das vezes, palavras soltas, frases, expressões e textos curtos em pomerano.

Quanto à quantificação desse *corpus*, os itens somaram 7.927 palavras corridas e 2.942 palavras distintas, conforme é possível visualizar por meio da Figura 37.

Figura 37: Estatísticas do *Corpus* de Textos Diversos da Internet.

WordList

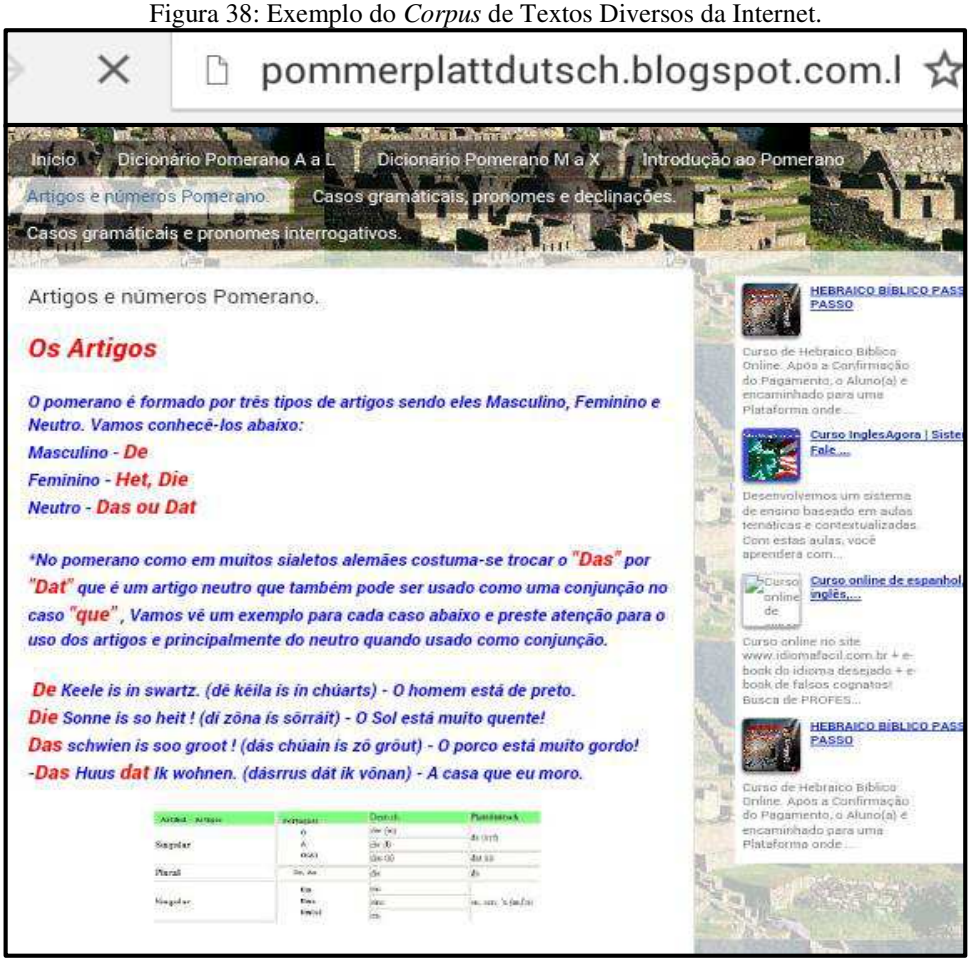
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	159.024	8.136	7.927		2.942	37,11	75,22	26,27	100
2	AI-SA-I-PKT-D.txt	1.225	39	39		35	89,74			100
3	PK-PKT-I-E (2).txt	13.506	258	234		161	68,80	65,00	24,75	100
4	PK-PKT-I-E (3).txt	12.476	454	393		210	53,44	60,25	35,19	100
5	PK-PKT-I-E.txt	6.790	149	124		107	86,29	73,00		100
6	T-B-I-PKT-D.txt	13.640	712	682		539	79,03	84,71	16,92	100
7	T-FB-I-PKT-D.txt	80.570	5.230	5.176		1.692	32,69	73,35	24,14	100
8	T-MK-I-PKT-D.txt	3.277	204	204		186	91,18	91,00	6,36	100
9	T-PKT-I-E (2).txt	3.556	484	484		349	72,11	83,75	11,24	100
10	T-PKT-I-E (4).txt	1.936	66	66		44	66,67			100
11	T-PKT-I-E.txt	1.446	12	12		11	91,67			100
12	TPS-PKT-I-D.txt	1.488	30	29		21	72,41			100
13	WPP-I-PKT-D.txt	19.114	498	484		394	81,40	87,25	9,94	100

Fonte: Elaboração própria.

Esse *corpus* apresentou grande quantidade de itens *Hapax Legomena*, provenientes do vocabulário mais variado encontrado na rede. Na imagem a seguir, Figura 38, podemos observar um exemplo do *Corpus* de Textos Diversos da Internet:

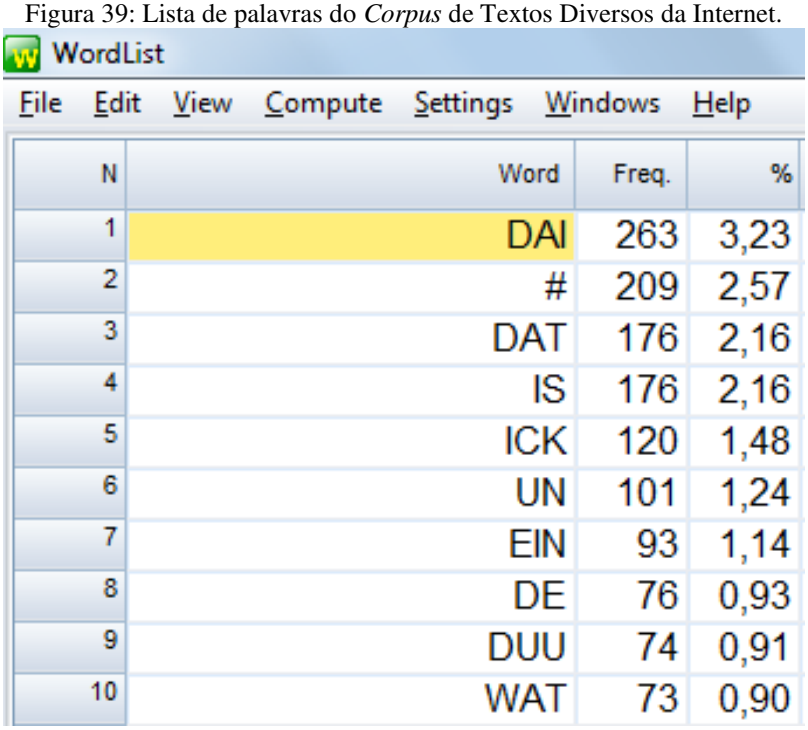
Figura 38: Exemplo do *Corpus* de Textos Diversos da Internet.



Fonte: http://www.pommerplattdutsch.blogspot.com.br

Dentre os itens mais frequentes nesse *corpus*, destacamos os pronomes pessoais *Ick* (eu) e *Duu* (você) e também o verbo "ser" conjugado na terceira pessoa do singular *is*, conforme exposto a seguir, na lista de palavras desse *corpus*, apresentada na Figura 39:

Figura 39: Lista de palavras do *Corpus* de Textos Diversos da Internet.



N	Word	Freq.	%
1	DAI	263	3,23
2	#	209	2,57
3	DAT	176	2,16
4	IS	176	2,16
5	ICK	120	1,48
6	UN	101	1,24
7	EIN	93	1,14
8	DE	76	0,93
9	DUU	74	0,91
10	WAT	73	0,90

Fonte: Elaboração própria.

Também encontramos no *Corpus* de Textos Diversos da Internet itens escritos iguais, mas referindo-se a coisas diferentes. Por exemplo, *wull* para nuvem (vem de *Wullken*, *Wolken*) e *wull* para querer, em *Ick wull* (Eu quero). Enfim, esse *corpus* contém maior quantidade de exemplos de conjugação verbal em comparação aos outros *corpora* mencionados anteriormente, apresentando, em geral, uma grande riqueza lexical a ser explorada.

5.3.7 *Corpus* de Legendas de Documentários – CLD

O Corpus de Legendas de Documentários – CLD é constituído de dois arquivos com todas as legendas de dois documentários sobre os pomeranos com áudio em pomerano e grafia das legendas em alemão-padrão, mas com características dialetais, contendo variações em alguns itens nas legendas. Acreditamos que essas variações na escrita são formas de aproximar a fala para representar os sons do pomerano. Trata-se do documentário *Dai Pomerer: Dai gang fon ainem folk* (Pomeranos: A trajetória de um povo), com direção de Vanildo Kruger, e do documentário *Walachai*, com direção de Rejane Zilles.

Esclarecemos que, nesse caso, a compilação foi feita com base nas legendas extraídas em alemão, mas transcritas adaptando a escrita ao áudio dos documentários, ou seja, considerou-se o conteúdo já escrito em alemão como base, porém, foi feita a conversão das lexias conforme a audição e transcrição da forma pomerana pronunciada nos áudios.

A seguir inserimos um exemplo, retirado do documentário *Dai Pomerer: Dai gang fon ainem folk* (2009):

Legenda: *"Die Eroberung, die Einheit, Stärke und Reichtum eines Volkes, [...]"*.

Transcrição e adaptação conforme o áudio: *"Dai Arbeid, Da Einigkeet, dai Kraf un da Rieckkeet for einem Folk [...]"*.

Na Figura 40 é possível visualizar um excerto do documentário e sua legenda.

Figura 40: Imagem do Documentário *Dai Pomerer: Dai gang fon ainem folk*.



Fonte: KRUGER, 2009.

O recorte cronológico compreende a data de produção dos dois documentários: o primeiro, *Dai Pomerer: Dai gang fon ainem folk*, de 2009, que trata da chegada dos imigrantes no Espírito Santo, do desbravamento das terras para plantio e da história dos pomeranos; o segundo, *Walachai*, de 2011, que trata do cotidiano de Walachai, uma localidade onde os moradores são etnicamente germânicos, da sua história, contada pelos moradores e da perseguição histórica sofrida por falarem outra língua que não o português.

No caso do documentário *Walachai*, são exibidas cenas com diversos falantes de um alemão dialetal. Acreditamos que ali há hunsriqueanos e pomeranos, porque alguns falantes pronunciam palavras como *Oowend* (abend, noite) *Dag* (Tag, dia), *Ick* (eu), que são formas

pomeranas, mas também aparece *Ich* (eu), *schaft* (trabalho em hunsriqueano), dentre outros casos. Procuramos estudos que determinassem a origem do dialeto falado em Walachai, porém não encontramos pesquisas que tenham realizado um censo étnico e linguístico na localidade. Segundo reportagem do IPOL (Instituto de Investigação e Desenvolvimento em Política Linguística), de 2014, sobre Walachai, "as pessoas da localidade falam um dialeto alemão chamado *Hunsrückisch*"[68]. Porém, acreditamos que existam também pomeranos na localidade, até porque, conforme já mencionamos neste trabalho, havia casamentos mistos, por exemplo, entre hunsriqueanos e pomeranos.

No documentário é possível ouvir algumas expressões pomeranas, como, por exemplo, *Wenn de Süünn up is*, *De Süünn up ging* (som de *jing*, com os fonemas do alemão), *und die aneren haw* [...], *Nu düütsch allas (verlernt)*, observadas durante um minuto, dos 04:42 aos 05:42 minutos de exibição do documentário. Também observamos uma variação nas legendas que "fugiam" um pouco do alemão-padrão, por exemplo, *Holzkarren* para o alemão-padrão *Leiterwagen*, que em hunsriqueano seria *Karose* (carroça)[69].

Quanto à quantificação desse *corpus*, realizamos a contagem dos itens e somaram 9.749 palavras corridas e 2.470 palavras distintas, conforme é possível visualizar por meio da Figura 41, abaixo:

Figura 41: Estatísticas do *Corpus* de Legendas de Documentários.

WordList

File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	121.432	9.804	9.749		2.470	25,34	73,49	25,09	100
2	D-TP-ES-PKT-WMP.txt	48.400	3.594	3.574		1.307	36,57	76,34	23,65	100
3	D-W-RS-PKT-WMP.txt	73.032	6.210	6.175		1.492	24,16	71,89	28,61	100

Fonte: Elaboração própria.

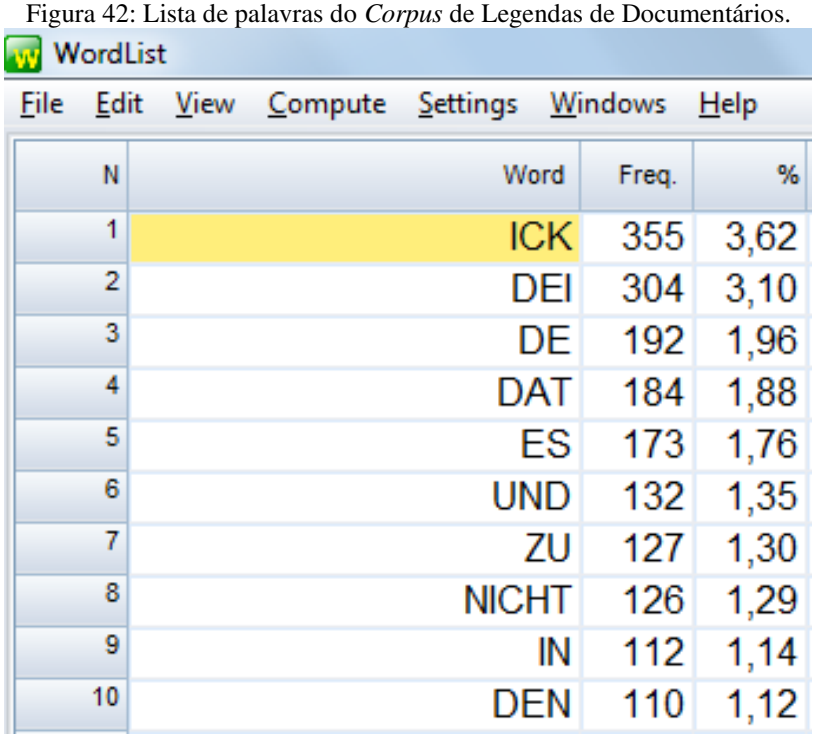
O conteúdo do CLD apresenta alta ocorrência dos pronomes pessoais como *ick* (Eu), artigos como *dei* (*dai*, "a" ou "as"), *de* (*der*, "o") e *dat* (neutro), conjunções, preposições, etc. Podemos observar também a frequência da forma de negação *nicht*, com 126 ocorrências, conforme demonstra a Figura 42.

---

[68] Matéria: O dialeto esquecido. IPOL. Publicada em 06/01/2014. Disponível em: <http://e-ipol.org/o-dialeto-esquecido/>. Acesso em: 18/05/2016.

[69] Para evitar que palavras hunsriqueanas fossem identificadas como pomeranas e contadas no *corpus*, etiquetamos todas as legendas que não identificamos claramente como pomeranas, ou seja, grande parte da transcrição do documentário está entre parêntesis angulares para não haver contagem de itens não-pomeranos.

Figura 42: Lista de palavras do *Corpus* de Legendas de Documentários.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	ICK	355	3,62
2	DEI	304	3,10
3	DE	192	1,96
4	DAT	184	1,88
5	ES	173	1,76
6	UND	132	1,35
7	ZU	127	1,30
8	NICHT	126	1,29
9	IN	112	1,14
10	DEN	110	1,12

Fonte: Elaboração própria.

Como há muitos diálogos falados livremente nos documentários, acreditamos que a sintaxe das orações poderá apresentar formações mais livres. Porém, essa é uma hipótese direcionada pelos dados resultantes da compilação desse *corpus*, fato linguístico que merece ser melhor investigado.

5.3.8 *Corpus* de Trabalhos Acadêmicos – CTA

O *Corpus* de Trabalhos Acadêmicos – CTA é um conjunto constituído por 5 textos com trechos em pomerano retirados de artigos e trabalhos acadêmicos como dissertações e teses que, ao tratar sobre o tema do pomerano dos mais diferentes pontos de vista (antropologia, religião, música, educação), inserem excertos de textos em pomerano, os quais compilamos para compor nosso *corpus*.

Embora envolvam áreas do conhecimento diversificadas, não encontramos muitos trabalhos acadêmicos contendo textos em pomerano, pois os mesmos se limitam a alguns trabalhos que mencionam as práticas culturais pomeranas em pomerano, bem como exemplos de textos utilizados para as análises que cada trabalho propunha. Assim, não conseguimos tantos textos acadêmicos com a quantidade de conteúdo em pomerano como esperávamos inicialmente.

No que se refere à linha de tempo que envolve todo o CTA, o recorte cronológico dos textos coletados vão desde 1995 até o mais recente, em 2015.

Quanto à quantificação desse *corpus*, a contagem somou 15.055 palavras corridas e 3.554 palavras distintas, conforme é possível visualizar por meio da Figura 43.

Figura 43: Estatísticas do *Corpus* de Trabalhos Acadêmicos.

WordList

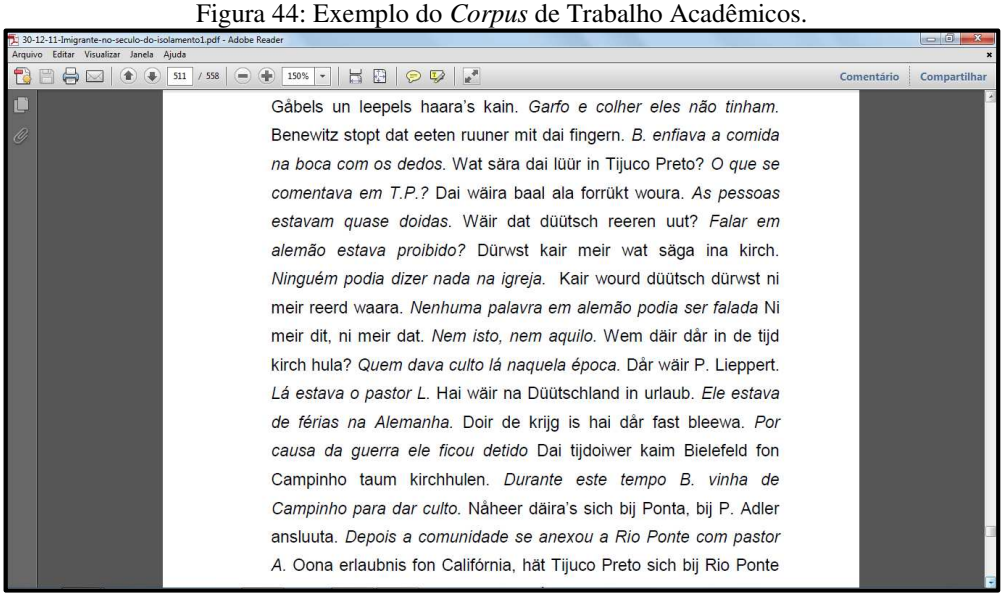
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	93.320	15.09	15.055		3.554	23,61	78,64	22,63	100
2	DM-SBK-CTA-PKT-D.txt	5.812	415	400		207	51,75	68,25	23,76	100
3	TA-DKS-MEH-PKT-D.txt	1.118	137	137		88	64,23	64,00		100
4	TA-DKS-NCP-PKT-D.txt	1.995	266	266		126	47,37	55,00	31,82	100
5	TA-DKS-PP-PKT-D.txt	1.626	227	227		161	70,93	79,00	14,85	100
6	TA-HR-P-PKT-D.txt	3.564	52	52		48	92,31			100
7	TD-CTA-IS-PKT-D.txt	79.205	14.001	13.973		3.216	23,02	79,37	19,85	100

Fonte: Elaboração própria.

Na imagem a seguir, Figura 44, podemos observar um exemplo do *Corpus* de Trabalhos Acadêmicos:

Figura 44: Exemplo do *Corpus* de Trabalho Acadêmicos.

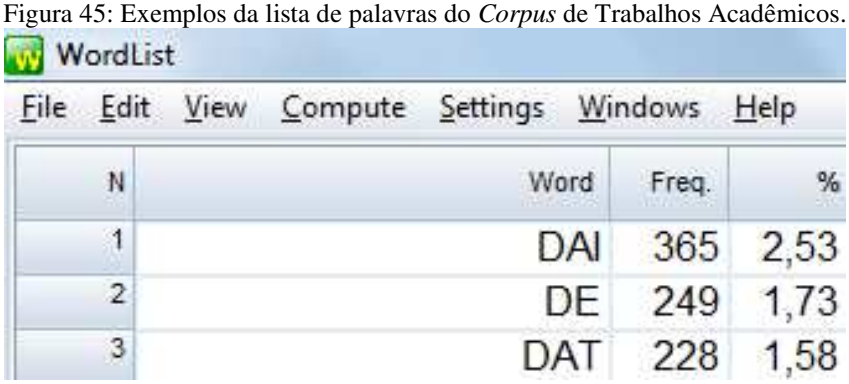


Gåbels un leepels haara's kain. *Garfo e colher eles não tinham.* Benewitz stopt dat eeten ruuner mit dai fingern. *B. enfiava a comida na boca com os dedos.* Wat sära dai lüür in Tijuco Preto? *O que se comentava em T.P.?* Dai wäira baal ala forrükt woura. *As pessoas estavam quase doidas.* Wäir dat düütsch reeren uut? *Falar em alemão estava proibido?* Dürwst kair meir wat säga ina kirch. *Ninguém podia dizer nada na igreja.* Kair wourd düütsch dürwst ni meir reerd waara. *Nenhuma palavra em alemão podia ser falada* Ni meir dit, ni meir dat. *Nem isto, nem aquilo.* Wem däir dår in de tijd kirch hula? *Quem dava culto lá naquela época.* Dår wäir P. Lieppert. *Lá estava o pastor L.* Hai wäir na Düütschland in urlaub. *Ele estava de férias na Alemanha.* Doir de krijg is hai dår fast bleewa. *Por causa da guerra ele ficou detido* Dai tijdoiwer kaim Bielefeld fon Campinho taum kirchhulen. *Durante este tempo B. vinha de Campinho para dar culto.* Nåheer däira's sich bij Ponta, bij P. Adler ansluuta. *Depois a comunidade se anexou a Rio Ponte com pastor A.* Oona erlaubnis fon Califórnia, hät Tijuco Preto sich bij Rio Ponte

Fonte: SEIBEL, 2010.

A frequência de alguns itens do *Corpus* de Trabalhos Acadêmicos pode ser observada por meio de um excerto da sua lista de palavras (*Wordlist*), conforme a Figura 45.

Figura 45: Exemplos da lista de palavras do *Corpus* de Trabalhos Acadêmicos.



Fonte: Elaboração própria.

Pela figura apresentada é possível verificar que os três artigos do pomerano estão nas primeiras posições em ordem de frequência na listagem desse *corpus*, *dai* (“a” ou “as”, artigo feminino, vem de *die*, *dei*), *de* (“o”, artigo masculino, vem de *der*) e *dat* (artigo neutro, vem de *das*). Também observamos, em uma breve busca pelo *corpus*, que o pronome pessoal *hai* (ele) também é muito frequente, com 121 ocorrências (posição 12), assim como o item *door* (“porta”, de *door* ou “lá”, de *dort*) é um dos mais frequentes, com 99 ocorrências (posição 16). A geração de linhas de concordâncias e o estudo dos itens podem esclarecer sobre os usos das formas pomeranas empregadas nesse *corpus*.

5.3.9 *Corpus* de Textos Religiosos – CTR

O *Corpus* de Textos Religiosos – CTR é um conjunto composto por 8 arquivos com diversos conteúdos em pomerano retirados de matérias publicadas pelos *Jehowas Witness* em material de evangelização em pomerano, bem como outros textos com conteúdos religiosos, como a tradução do evangelho de Lucas para o pomerano, trechos de bíblias pomeranas utilizadas pelos pomeranos no Brasil e outros textos escritos em pomerano com conteúdo religioso.

O conteúdo gira em torno de textos cristãos, tanto luteranos quanto das Testemunhas de Jeová. A maioria deles é proveniente do novo testamento, pois o CTR contém, por exemplo, a oração do “Pai Nosso” em pomerano, histórias bíblicas resumidas e em linguagem moderna, etc.

No que se refere ao recorte cronológico do *Corpus* de Textos Religiosos, o material coletado abrange desde 1960 até o mais recente, em 2014.

Quanto à quantificação desse *corpus*, a contagem somou 7.437 palavras corridas e 1.759 palavras distintas, conforme é possível visualizar por meio da Figura 46.

Figura 46: Estatísticas do *Corpus* de Textos Religiosos.

WordList

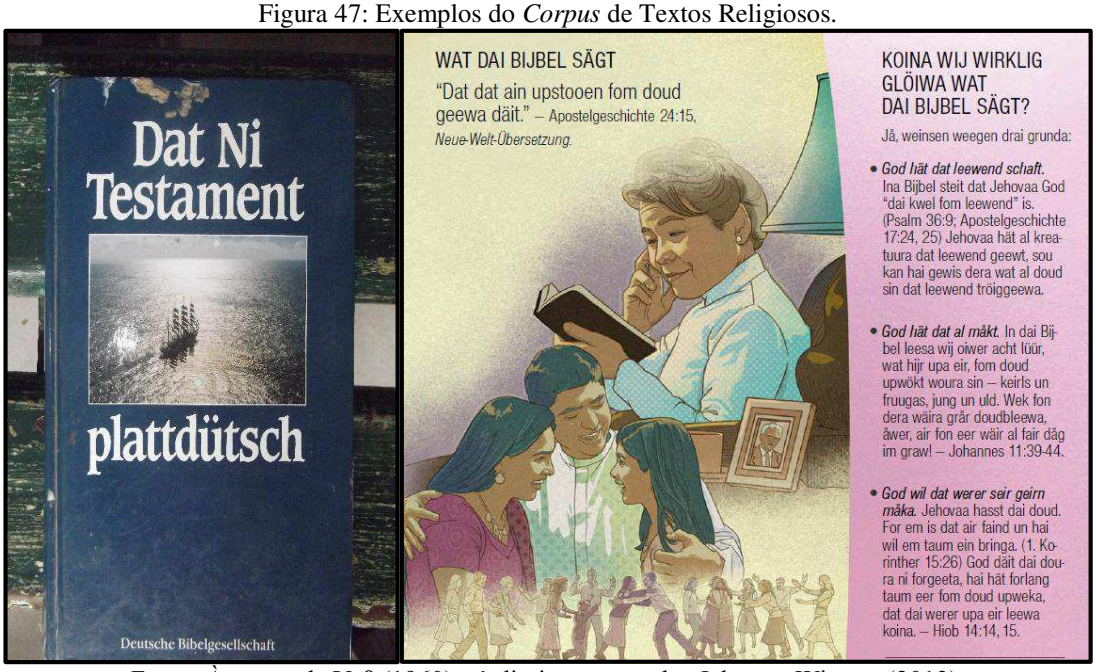
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	72.894	7.859	7.437		1.759	23,65	65,15	33,98	100
2	R-O-PKT-I-V.txt	1.226	80	80		53	66,25			100
3	TBA-SBB-PDF.txt	1.778	277	264		151	57,20	69,50	21,57	100
4	T-UW-I-E.txt	1.830	130	130		87	66,92	73,00		100
5	JW-R-PKT-WMP.txt	4.674	279	279		137	49,10	66,50	23,69	100
6	BPD-EV-FK-ZRN-IT-MG-IP.txt	5.940	550	526		231	43,92	67,80	25,89	100
7	TEL3-KT-SLS-RS.txt	12.724	1.040	1.001		367	36,66	54,50	31,44	100
8	TEL2-KT-SLS-RS.txt	12.924	1.226	1.173		387	32,99	68,67	28,17	100
9	JW-I-PKT-DT.txt	14.590	2.596	2.384		564	23,66	64,92	33,25	100
10	TEL1-KT-SLS-RS.txt	17.208	1.681	1.600		461	28,81	67,50	30,03	100

Fonte: Elaboração própria.

Na imagem a seguir, Figura 47, podemos observar um exemplo do *Corpus* de Textos Religiosos:

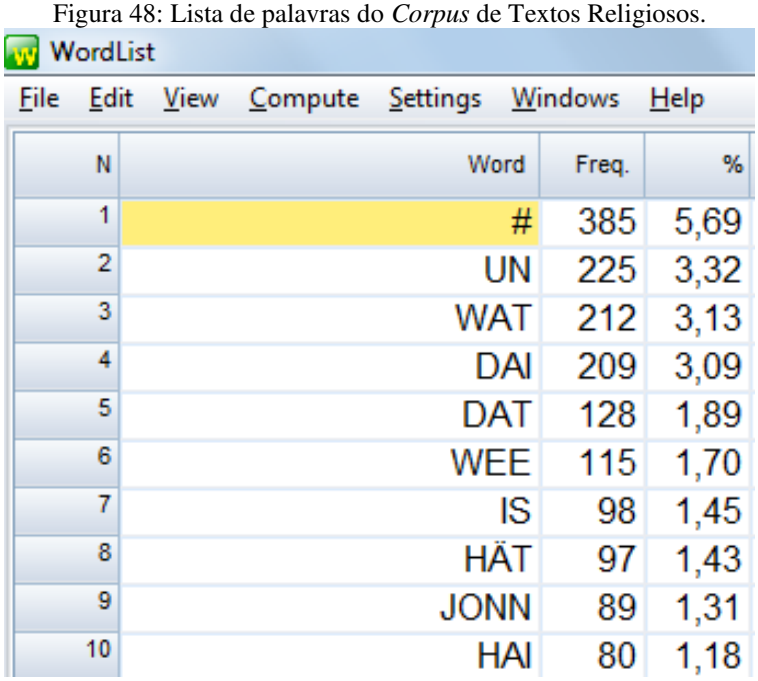
Figura 47: Exemplos do *Corpus* de Textos Religiosos.



Fonte: À esquerda Voß (1960) e à direita encarte dos Jehowas Witness (2013).

Por meio da lista de palavras desse *corpus*, podemos observar a alta frequência de verbos flexionados, como "ser" e "ter", respectivamente, *is* (é), com 98 ocorrências (posição 7) e *hät* ("tem" ou "tinha"), com 97 ocorrências (posição 8). Também podemos observar que o item mais frequente é a conjunção *un* ("e", o mesmo que *und*), conforme exposto na Figura 48.

Figura 48: Lista de palavras do *Corpus* de Textos Religiosos.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	#	385	5,69
2	UN	225	3,32
3	WAT	212	3,13
4	DAI	209	3,09
5	DAT	128	1,89
6	WEE	115	1,70
7	IS	98	1,45
8	HÄT	97	1,43
9	JONN	89	1,31
10	HAI	80	1,18

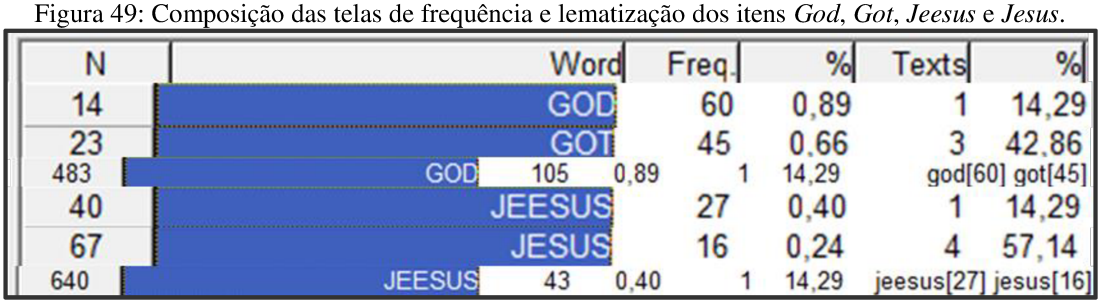
Fonte: Elaboração própria.

Presupúnhamos que, assim como na contagem da *Bíblia Online*[70], o item "Jesus" ocorreria no *Corpus* de Textos Religiosos em uma proporção maior no novo testamento em relação a "Deus" (assim como "Deus" ocorre em uma proporção maior em relação a "Jesus" no Antigo Testamento), já que o CTR é composto de textos evangelísticos e de trechos do novo testamento em pomerano.

Seguindo essa suposição, checamos o número de ocorrências de "Deus" e "Jesus" em pomerano no CTR e constatamos que: *God* ("Deus", 60 ocorrências, posição 14) e sua variação *Got* (45 ocorrências, posição 23), lematizadas, ocorreram 105 vezes no CTR, enquanto *Jeesus* ("Jesus", 27 ocorrências, posição 40) e sua variação *Jesus* (16 ocorrências, posição 67), lematizadas, ocorreram 43 vezes no CTR. Assim, concluímos que, diferentemente do que supúnhamos, a proporção de ocorrências do item "Jesus" é menos da metade (apenas 43,80%) das ocorrências do item "Deus" nos textos religiosos em pomerano, baseados no novo testamento. A Figura 49 ilustra os dados pesquisados, conforme as linhas do WST que "printamos" e editamos.

[70] O site *Bíblia Online*, disponível em várias línguas, oferece atualmente a contagem dos itens pesquisados na bíblia. Na contagem da *Bíblia Online* para os itens de busca "Deus" e "Jesus", "Deus" ocorre 4.969 vezes no antigo testamento e 2.123 no novo testamento, " Jesus" ocorre 35 vezes no antigo testamento e 1.669 no novo testamento, numa proporção de 78,61% em relação às ocorrências de "Deus" no novo testamento.

Figura 49: Composição das telas de frequência e lematização dos itens *God*, *Got*, *Jeesus* e *Jesus*.



N	Word	Freq.	%	Texts	%
14	GOD	60	0,89	1	14,29
23	GOT	45	0,66	3	42,86
483	GOD 105 0,89	1	14,29	god[60] got[45]	
40	JEESUS	27	0,40	1	14,29
67	JESUS	16	0,24	4	57,14
640	JEESUS 43 0,40	1	14,29	jeesus[27] jesus[16]	

Fonte: Elaboração própria.

A análise acima indica que os dados dos *corpora* podem envidenciar fatos diferentes do que pressupomos, podendo negar hipóteses e dirigir as análises para outros rumos. Também acreditávamos que um substantivo muito frequente no CTR seria *glouwe* (fé), porém esse *corpus* também não comprovou esse dado, visto que *glouwe* ocorreu apenas 4 vezes no *Corpus* verificado.

5.3.10 *Corpus* de Músicas Pomeranas – CMP

O *Corpus* de Músicas Pomeranas – CMP é um conjunto composto por 12 arquivos contendo 21 músicas[71]. São textos com letras de músicas em pomerano. As músicas que compõem esse *corpus* são de diversos tipos, por exemplo, canções tradicionais pomeranas, versões pomeranas de canções folclóricas germânicas, canções populares, músicas compostas atualmente por cantores pomeranos e/ou bandas de músicas pomeranas, trovinhas infantis cantadas por crianças em diferentes versões, paródias e traduções de músicas brasileiras para o pomerano.

Algumas músicas foram preservadas por meio da oralidade; embora tenham origem no passado, são ressignificadas (o vocabulário é atualizado, ganha sentidos novos, conforme o contexto atual) quando são reproduzidas em diferentes versões pomeranas nos dias de hoje. No que se refere à linha de tempo que envolve todo o *Corpus* de Músicas Pomeranas, podemos informar que a música mais antiga remonta ao ano de 1807 e a mais recente ao ano de 2014.

[71] Neste caso, o número de músicas não é igual ao número de arquivos, pois decidimos não organizar uma música por arquivo, porque coletamos as músicas em documento de texto digital e já estavam agrupadas por arquivo, assim respeitamos a fonte, conforme a encontramos quando recebemos a doação de algumas músicas, por parte do pesquisador de músicas pomeranas Kuhn Silva, ao visitar São Lourenço do Sul/RS.

Quanto à quantificação do *corpus* de músicas pomeranas, realizamos a contagem dos itens e contabilizamos pelo WST um total de 1.746 palavras corridas e 647 palavras distintas, conforme comprova a Figura 50.

Figura 50: Estatísticas do *Corpus* de Músicas Pomeranas.

WordList

File Edit View Compute Settings Windows Help

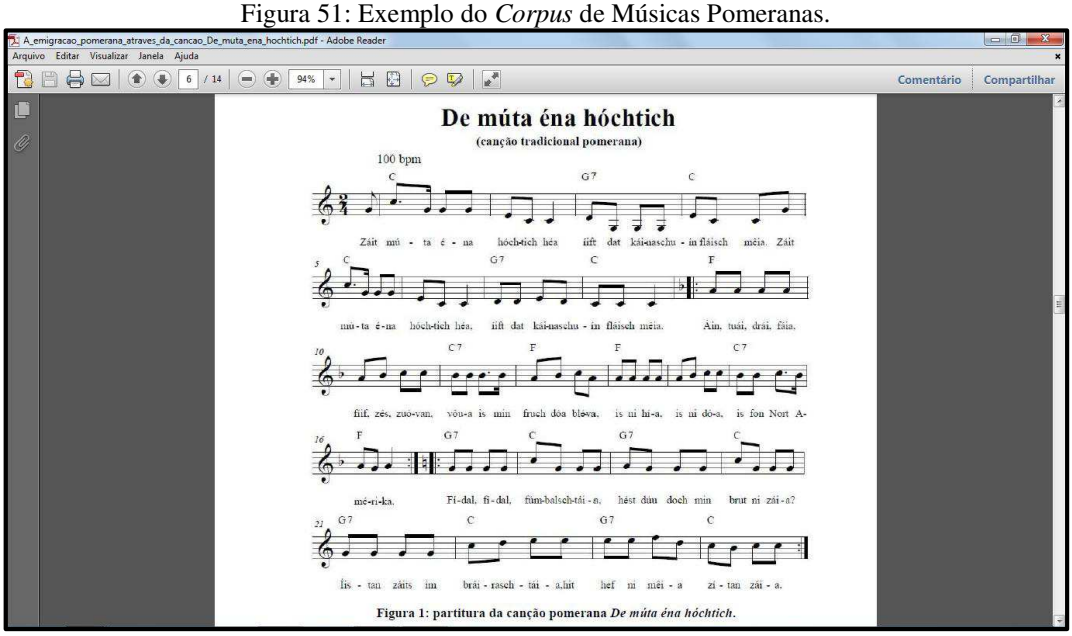
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	12.476	1.750	1.746		647	37,06	88,64	10,93	10
2	M-MHT-C-RS-PKT-E.txt	558	97	97		71	73,20	93,33	6,00	10
3	M-DKS-SLS-RS-DT1.txt	632	97	97		51	52,58	84,44	14,91	10
4	M-EML-SLS-RS-DT2.txt	654	94	94		49	52,13	81,11	13,76	10
5	M-MHT-C-RS-DT2.txt	699	128	128		53	41,41	85,00	15,13	10
6	M-DKS-SLS-RS-DT3.txt	706	93	93		31	33,33	86,67	15,71	10
7	M-FG-IT-MG-PKT-D.txt	894	159	159		82	51,57	95,33	4,09	10
8	M-DKS-SLS-RS-DT2.txt	935	175	173		118	68,21	96,47	3,59	10
9	M-DKS-SLS-RS-PDF2.txt	1.074	135	135		92	68,15	93,08	8,39	10
10	M-EGH-SLS-RS-DT.txt	1.113	188	188		92	48,94	90,56	8,01	10
11	M-DKS-SLS-RS-PDF.txt	1.151	130	130		94	72,31	90,77	12,18	10
12	M-EML-SLS-RS-DT1.txt	1.320	224	224		91	40,63	81,36	16,94	10
13	M-DKS-SLS-DT.txt	2.740	230	228		116	50,88	85,65	13,06	10

Fonte: Elaboração própria.

É um *corpus* pequeno, formado por textos de músicas de curta extensão, que possuem pequenas estrofes. Observamos que também há muita repetição das letras das canções tradicionais germânicas. A canção "*Marie, Mara, Maruschkacka*", por exemplo, possui três diferentes versões, e seu refrão é repetido em todas elas, o que influencia diretamente na repetição de itens lexicais nesse *corpus*. A repetição é uma característica das canções, inclusive pela apresentação de palavras repetidas nos refrães, mesmo dentro de uma única versão, o que limita, pelo seu gênero, a densidade de um *corpus* de músicas.

Na imagem a seguir, Figura 51, podemos observar um exemplo extraído do *Corpus* de Músicas Pomeranas:

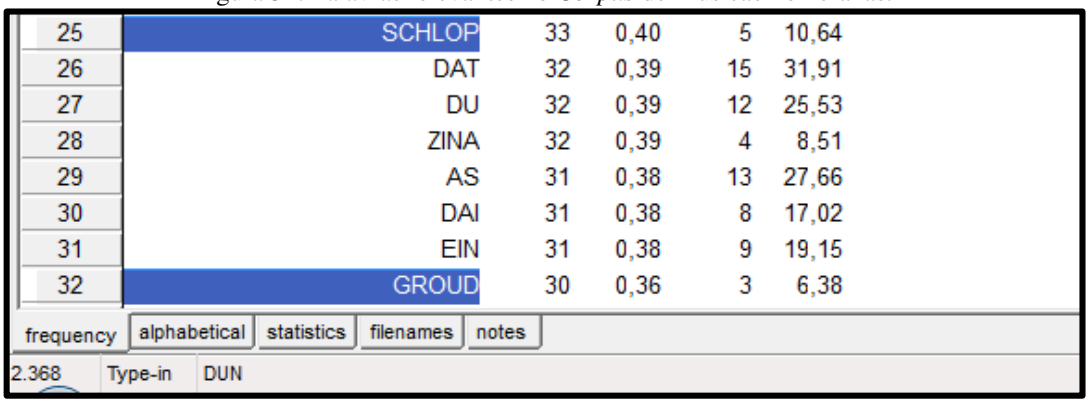
Figura 51: Exemplo do *Corpus* de Músicas Pomeranas.



Fonte: Kuhn Silva (2013).

Na Figura 52, apresentada a seguir, destacamos alguns itens lexicais frequentes e relevantes no *Corpus* de Músicas Pomeranas: *schlop* (durma) e *groud* (grande):

Figura 52: Palavras relevantes no *Corpus* de Músicas Pomeranas.



Fonte: Elaboração própria.

É possível verificar que *schloop*, na forma *schlop*, apresentou variação na escrita, e ocorreu 33 vezes no *corpus* de músicas (sem lematização), assim como *groud*, variação de *grout*, ocorreu 30 vezes. Uma explicação para essa frequência poderia ser que a forma imperativa *schloop* (durma!), repetida muitas vezes nas canções de ninar (por exemplo, "*schloop, kinka schloop...*"); e *grout* (grande), que pode referir-se a *groutvooter* ou *groutmotter* (vovô ou vovó), quando utilizadas como prefixo de *vooter* e *motter*, conforme busca que realizamos no *corpus* por meio do WST. Acreditamos que os exemplos acima referidos são itens comuns em canções de ninar, cantigas de roda e canções folclóricas.

5.3.11 *Corpus* Literário Pomerano – CLP

O *Corpus* Literário Pomerano – CLP é um conjunto de 8 arquivos de textos constituídos a partir dos dados coletados e compilados por meio de visitas de campo e também da internet. Esse conjunto de textos contém poemas, versos, rimas, trovinhas e um conto germânico clássico, que narra uma história infantil em versão pomerana. Esse conto foi coletado em São Lourenço do Sul/RS e doado pelo pesquisador Kuhn Silva. Identificamos que se trata do conto *Dai suowen kleiner seechen*, em uma versão atualizada e transliterada em pomerano do conto *Der Wolf und die sieben jungen Geißlein* (O lobo e os sete cabritinhos), dos Irmãos Grimm, de 1812.

O *Corpus* Literário Pomerano também contém trovinhas (brincadeiras de roda, ritmadas e rimadas contendo historinhas) na versão pomerana. São muito conhecidos popularmente e reproduzidos frequentemente no cotidiano, como *Ludwig-Ludwig*, *Hana, Blintkauh*, etc.

No que se refere à linha de tempo que envolve todo o *Corpus* Literário Pomerano, não conseguimos precisar a datação dos conteúdos, pois alguns textos não contêm datação e tivemos que considerar a data de coleta. Então podemos dizer que esse *corpus* possui um recorte cronológico que passa por 1812 (no caso do conto atualizado em 2013), 1932 (no caso do poema Grout sin wi Pommern) e até poemas produzidos recentemente, em 2014.

Quanto à quantificação desse *corpus*, os dados somaram 972 palavras corridas e 531 palavras distintas, conforme é possível visualizar por meio da Figura 53.

Figura 53: Estatísticas do *Corpus* Literário Pomerano.

WordList

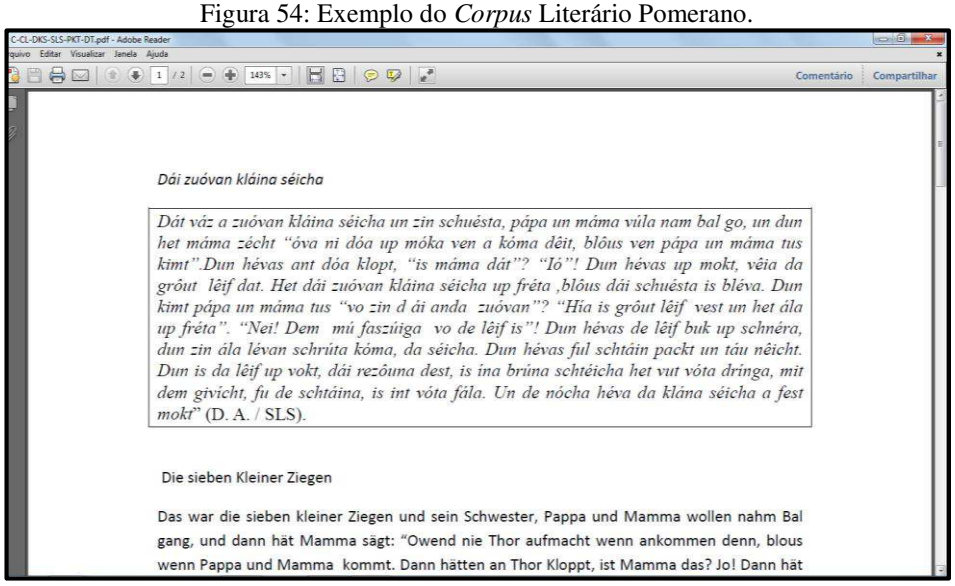
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	14.045	975	972		531	54,63	93,04	9,71	10
2	P-CL-I-PKT-D.txt	355	49	49		38	77,55	90,00	11,45	10
3	P-CL-WB-WS-RS-P.txt	417	73	73		56	76,71	90,00	13,12	10
4	P-SA-IT-MG-PKT-E.txt	654	79	79		51	64,56	87,14	10,09	10
5	P-CL-DKS-SLS-M.txt	898	134	134		50	37,31	87,69	14,39	10
6	P-CL-C-RS-PKT-E.txt	1.054	62	62		45	72,58	98,33	2,45	10
7	VT-CL-I-PKT-D.txt	1.139	163	163		93	57,06	91,88	6,88	10
8	VT-SA-IT-PKT-CIP2.txt	1.204	65	65		55	84,62	96,67	4,91	10
9	VT-SA-IT-PKT-CIP1.txt	1.448	45	42		40	95,24	97,50	2,86	10
10	C-CL-DKS-SLS-DT.txt	2.614	157	157		85	54,14	96,00	6,69	10
11	C-CL-EML-SLS-WMP.txt	4.262	148	148		117	79,05	96,43	3,02	10

Fonte: Elaboração própria.

Na imagem a seguir, Figura 54, podemos observar um exemplo do *Corpus* Literário Pomerano:

Figura 54: Exemplo do *Corpus* Literário Pomerano.



Dái zuóvan kláina séicha

Dát váz a zuóvan kláina séicha un zin schuésta, pápa un máma vúla nam bal go, un dun het máma zécht "óva ni dóa up móka ven a kóma dêit, blôus ven pápa un máma tus kimt".Dun hévas ant dóa klopt, "is máma dát"? "Ió"! Dun hévas up mokt, vêia da grôut lêif dat. Het dái zuóvan kláina séicha up fréta ,blôus dái schuésta is bléva. Dun kimt pápa un máma tus "vo zin d ái anda zuóvan"? "Hía is grôut lêif vest un het ála up fréta". "Nei! Dem mú faszúiga vo de lêif is"! Dun hévas de lêif buk up schnéra, dun zin ála lévan schrúta kóma, da séicha. Dun hévas ful schtáin packt un táu nêicht. Dun is da lêif up vokt, dái rezôuna dest, is ína brúna schtéicha het vut vóta dringa, mit dem givicht, fu de schtáina, is int vóta fála. Un de nócha héva da klána séicha a fest mokt" (D. A. / SLS).

Die sieben Kleiner Ziegen

Das war die sieben kleiner Ziegen und sein Schwester, Pappa und Mamma wollen nahm Bal gang, und dann hät Mamma sägt: "Owend nie Thor aufmacht wenn ankommen denn, blous wenn Pappa und Mamma kommt. Dann hätten an Thor Kloppt, ist Mamma das? Jo! Dann hät

Fonte: Kuhn Silva (2013).

Na lista de palavras do CLP, podemos observar que o item mais frequente, na posição 01, é o verbo "ser", conjugado na terceira pessoa do singular *is* (é), assim como o substantivo "criança", *kinka* (variação do singular *kint* e do plural *kinner* ou *kinder*), também é muito frequente (14 ocorrências, posição 05), conforme podemos observar na Figura 55, a seguir:

Figura 55: Lista de palavras do *Corpus* Literário Pomerano.

WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	IS	20	2,65
2	DAI	18	2,39
3	UN	17	2,25
4	DAT	14	1,86
5	KINKA	14	1,86
6	DU	13	1,72
7	KLAUCK	12	1,59
8	DUU	10	1,33
9	WAT	10	1,33
10	A	9	1,19

Fonte: Elaboração própria.

Também é possível observar na listagem acima que o item 10, "a", é frequente, ele refere-se nos textos a *ain* (variação escrita de *ein*, um); fato também observado em outros

corpora do pomerano compilados. Esse dado sugere uma perda na pronúncia da terminação, da vogal "i" e da consoante "n"; hipótese direcionada pelo PK, que merece um estudo específico.

5.3.12 *Corpus* de Livros – CL

O *Corpus* de Livros – CL é um conjunto de 6 arquivos de livros escritos em pomerano, constituído a partir dos 6 livros compilados. Três dos livros foram tratados por OCR, redigitados nos trechos com distorção, convertidos e revisados para a composição desse *corpus*, e se tratam daqueles que foram totalmente escritos em pomerano. Os outros 3 livros citavam apenas algumas palavras e expressões em pomerano, as quais consideramos para a compilação. O conteúdo deles abrange histórias sobre o cotidiano dos pomeranos, história da Pomerânia e histórias infantis.

No que se refere à linha de tempo que envolve todo o *Corpus* de Livros, o livro mais antigo data de 1925 (Publicado na Alemanha), encontrado na casa de uma pomerana que confirmou ter lido o livro e que está escrito em pomerano, e o mais recente foi publicado no Brasil, em 2012.

Quanto à quantificação desse *corpus*, a contagem dos itens somou 12.216 palavras corridas e 3.861 palavras distintas, conforme é possível visualizar na Figura 56.

Figura 56: Estatísticas do *Corpus* de Livros.

WordList

File Edit View Compute Settings Windows Help

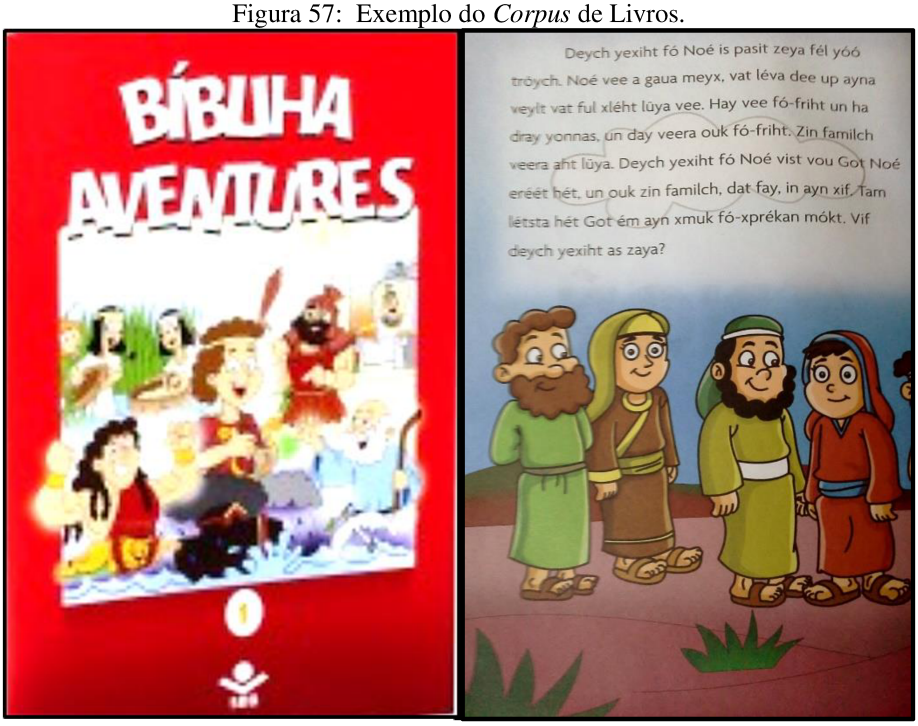
N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	149.509	12.288	12.216		3.861	31,61	77,20	22,44	100
2	L-D-PKT-OCR.txt	1.854	134	132		115	87,12	84,00		100
3	L-DL-RS-PKT-E.txt	11.707	592	592		495	83,61	93,00	3,98	100
4	L-GS-SLS-PKT-CIP.txt	2.636	153	148		123	83,11	81,00		100
5	L-HR-SLS-PKT-CIP.txt	2.626	182	176		116	65,91	70,00		100
6	L-IT-PKT-OCR.txt	17.550	1.336	1.336		404	30,24	66,46	31,28	100
7	L-IT-PKT.txt	854	62	62		37	59,68			100
8	L-JFJ-PKT-E.txt	644	26	23		20	86,96			100
9	L-LW-SLS-PKT-CIP.txt	5.114	297	297		212	71,38	81,50	13,08	100
10	L-SA-PKT-DT.txt	8.656	700	658		443	67,33	80,14	15,64	100
11	L-SBB-SLS-PKT-CIP.txt	97.868	8.806	8.792		2.296	26,11	77,52	22,66	100

Fonte: Elaboração própria.

Para o contexto pomerano, um *corpus* com 12.216 mil itens distintos não pode ser considerado pequeno, tendo em vista que o único dicionário pomerano publicado no Brasil alcança 16 mil itens (verbetes).

Na imagem a seguir, Figura 57, podemos observar um exemplo do *Corpus* de Livros:

Figura 57: Exemplo do *Corpus* de Livros.



Fonte: SBB (2012).

Quanto à lista de palavras desse *corpus*, podemos dizer que além de conter uma alta frequência das palavras gramaticais, também apresenta alta frequência do verbo *hawen* (ter), com 177 ocorrências somando as formas flexionadas econtradas e lematizadas. Na lista de frequências, podemos vizualizar que a flexão *hät* está entre os dez itens mais frequentes nesse *corpus*, na posição 4, com 122 ocorrências, conforme mostra a Figura 58, a seguir:

Figura 58: Lista de Palavras do *Corpus* de Livros.

WordList

File Edit View Compute Settings Windows

N	Word	Freq.
1	DAI	142
2	UN	140
3	DAT	128
4	HÄT	122
5	IS	86
6	WAT	66
7	MIT	56
8	IN	44
9	AS	41
10	HAI	38

Fonte: Elaboração própria.

Um dos itens encontrados como *Hapax Legomen* e que indica uma variação vocabular é o uso do verbo *schlepa* (posição 3.773), como variação à *schloopa*, utilizado para referir-se àquele que dorme, vem de *schloopen* (*schlafen*), dormir. Enfim, o estudo dos dados desse *corpus* pode revelar inúmeros casos de variação lexical e resgatar itens lexicais do pomerano utilizados com menor frequência ou perceber itens provenientes de outras variedades que entraram em contato com o pomerano.

5.3.13 *Corpus* de *Sprüche* Diversos – CSD

O *Corpus* de *Sprüche* Diversos – CSD é um conjunto de 6 arquivos de texto contendo fraseologismos constituídos a partir dos dados coletados na internet e também pessoalmente nas visitas de campo: são ditados populares, versos, expressões, máximas, versículos que já se tornaram adágios e jargões. Encontrados em livros, cadernos e anotações dos pomeranos, fotos de casa (*Wandmoolen*), quadros (*Wandspruch*), tecidos (*Wandschoener*), transcritos da fala e de vídeos, e até trava-línguas coletados com professoras em São Lourenço do Sul/RS, etc.

Falam sobre a sorte, sobre a vida, alguns correspondem a fraseologismos que também existem em português, outros falam sobre hábitos dos pomeranos, etc. Um exemplo, muito conhecido na tradição oral, é *Een pommerschen moogen koinna allas vordreegen* (em alemão-padrão: *Ein pommerscher Magen kann alles vertragen*), que significa *"*Um estômago pomerano pode aguentar tudo*"*. Outro exemplo é *Lüüsa goicka upa mien koopp*, que seria "piolhos coçam sobre minha cabeça", mas com o sentido equivalente ao fraseologismo brasileiro "macacos me mordam".

No que se refere à linha de tempo que envolve todo o *Corpus* de *Sprüche* Diversos, é difícil determinar, pois alguns fraseologismos datam de mais de um século e meio atrás, pois já eram conhecidos na Prússia, antes da imigração pomerana para o Brasil. Assim, em nenhum texto conseguimos identificar a data dos fraseologismos. Se considerarmos as datas de publicação, no caso dos fraseologismos encontrados em livros, e de veiculação, nos casos encontrados na internet, então o recorte cronológico abrange de 1923 a 2015.

Quanto à quantificação desse *corpus*, a contagem dos itens somou 730 palavras corridas e 405 palavras distintas, conforme é possível visualizar por meio da Figura 59.

Figura 59: Estatísticas do *Corpus* de *Sprüche* Diversos.

WordList

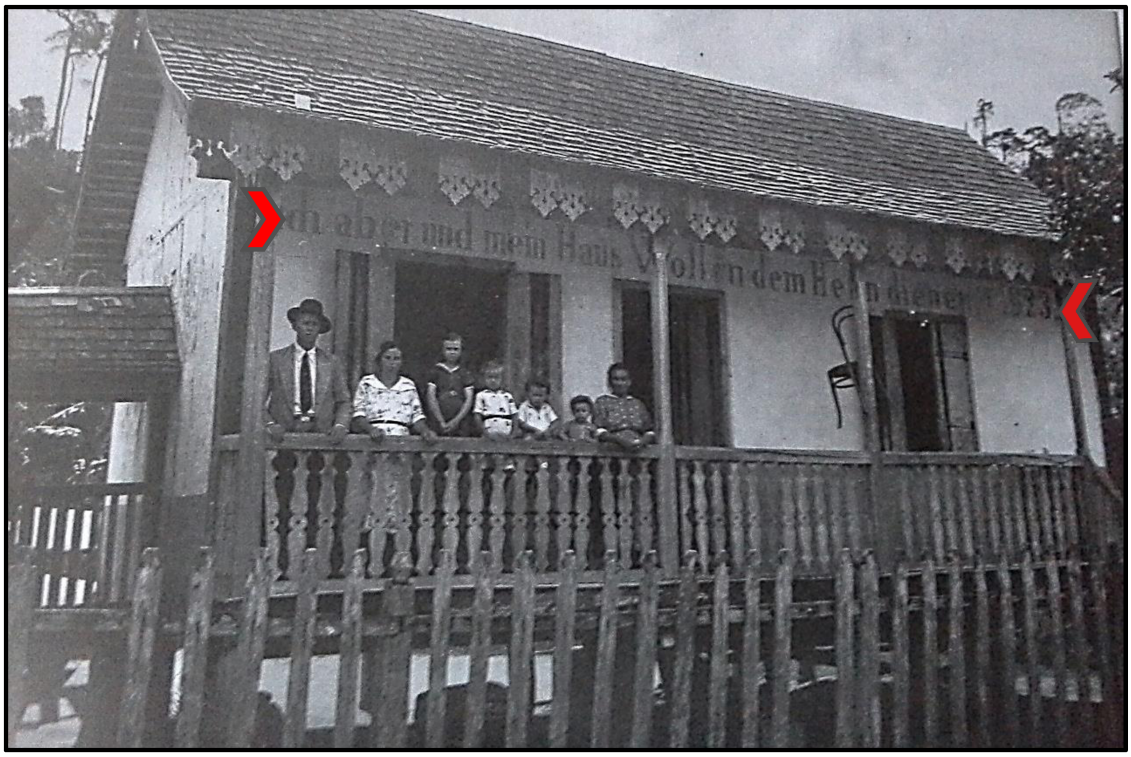
File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	ype/toke atio	standard TTR	STTR std.dev.	STTR basis
1	Overall	14.312	734	730		405	55,48	93,24	6,14	10
2	WS-SL-SA-PKT-F.txt	738	51	49		42	85,71	96,00	3,59	10
3	D-CS-LD-PKT-Pl.txt	1.494	56	56		41	73,21	94,00	4,11	10
4	D-CS-I-FD-PKT-PD.txt	1.814	143	143		105	73,43	92,86	5,60	10
5	D-CS-VC-PKT-PD.txt	1.990	32	32		24	75,00	90,00	9,46	10
6	F-L-I-CS-PKT-PD.txt	3.030	228	228		153	67,11	94,09	8,42	10
7	F-AD-CS-PKT-PD.txt	5.246	224	222		149	67,12	92,27	6,37	10

Fonte: Elaboração própria.

Na imagem a seguir, Figura 60, podemos observar um exemplo de *Wandmoolen*, onde um verso é pintado na parede da casa para abençoá-la, exemplo extraído do *Corpus* de *Sprüche* Diversos:

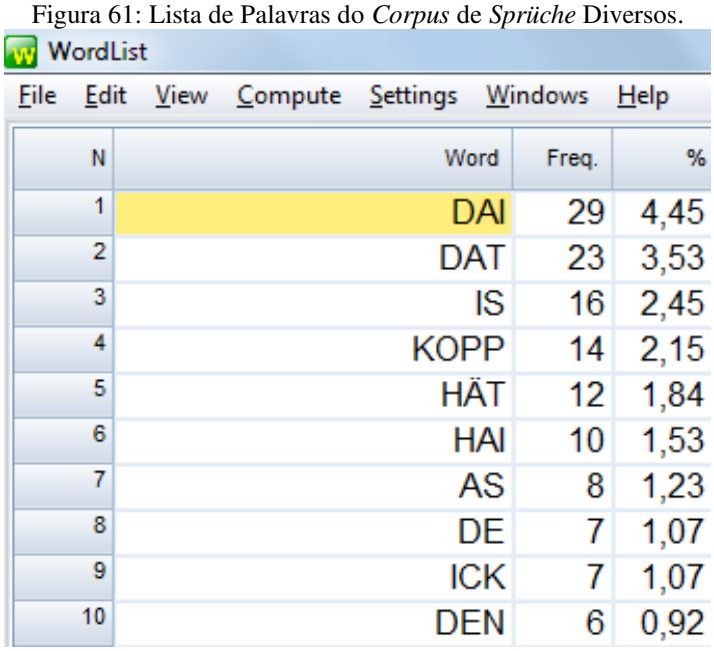
Figura 60: Exemplo do *Corpus* de *Sprüche* Diversos.



Fonte: Wandmoolen. Francisco Seibel, ([20--]).

A seguir, é possível verificar, por meio da lista de palavras do *Corpus* de *Sprüche* Diversos, que os itens mais frequentes foram *dai* (artigo feminino “a” ou “as”) na primeira posição, com 29 ocorrências, *dat* (artigo neutro) na segunda posição, com 23 ocorrências, e o verbo *is* (e) conjugado na terceira pessoa do singular, na terceira posição, com 16 ocorrências, conforme demonstra a Figura 61.

Figura 61: Lista de Palavras do *Corpus* de *Sprüche* Diversos.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	DAI	29	4,45
2	DAT	23	3,53
3	IS	16	2,45
4	KOPP	14	2,15
5	HÄT	12	1,84
6	HAI	10	1,53
7	AS	8	1,23
8	DE	7	1,07
9	ICK	7	1,07
10	DEN	6	0,92

Fonte: Elaboração própria.

Também é possível verificar acima a frequência do substantivo *kopp* (cabeça), 14 ocorrências, na quarta posição da lista de frequências, que é utilizado nos fraseologismos pomeranos, como por exemplo, *hals upa kopp*, para falar de alguém precipitado, que põe o "pescoço acima da cabeça" (de sentido equivalente ao nosso "por a carroça à frente dos bois"), e também *de kopp hänga loota*, para falar de alguém desanimado, que tem "a cabeça pendurada". Enfim, são muitos os comentários e análises possíveis a partir da observação dos dados desse *corpus*.

5.3.14 *Corpus* de Palavras Soltas – CPS

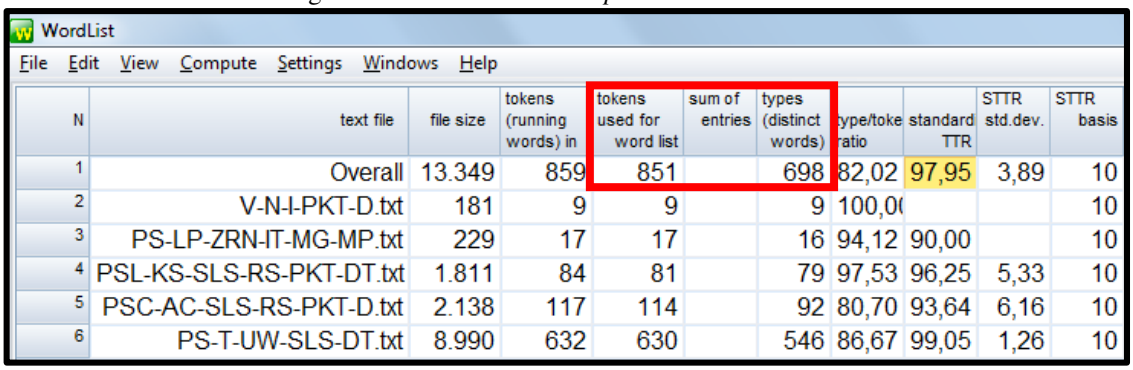
O *Corpus* de Palavras Soltas – CPS é um conjunto de oito arquivos contendo palavras em pomerano, constituídos a partir dos dados coletados pessoalmente em Vila Neitzel/MG, São Lourenço do Sul/RS e também compilados de livros e da internet. Ele contém substantivos, verbos, artigos, preposições (também contém frases curtas e expressões), podendo contribuir para a identificação do nome de frutas, partes do corpo humano, cores e objetos diversos.

No que se refere à linha de tempo que envolve todo o *Corpus* de Palavras Soltas, podemos dizer que são dados contemporâneos e que a datação coincide, nesse caso, com a publicação dos livros e das postagens na internet. Quanto às palavras coletadas (em forma escrita) pessoalmente nas visitas de campo, algumas não possuem datas. Podemos dizer que o recorte cronológico desse *corpus* se encontra entre 2008 a 2014.

Quanto à quantificação desse *corpus*, a contagem dos itens somou 851 palavras corridas

e 698 palavras distintas, conforme é possível visualizar por meio da Figura 62.

Figura 62: Estatísticas do *Corpus* de Palavras Soltas.



WordList

File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	13.349	859	851		698	82,02	97,95	3,89	10
2	V-N-I-PKT-D.txt	181	9	9		9	100,0(			10
3	PS-LP-ZRN-IT-MG-MP.txt	229	17	17		16	94,12	90,00		10
4	PSL-KS-SLS-RS-PKT-DT.txt	1.811	84	81		79	97,53	96,25	5,33	10
5	PSC-AC-SLS-RS-PKT-D.txt	2.138	117	114		92	80,70	93,64	6,16	10
6	PS-T-UW-SLS-DT.txt	8.990	632	630		546	86,67	99,05	1,26	10

Fonte: Elaboração própria.

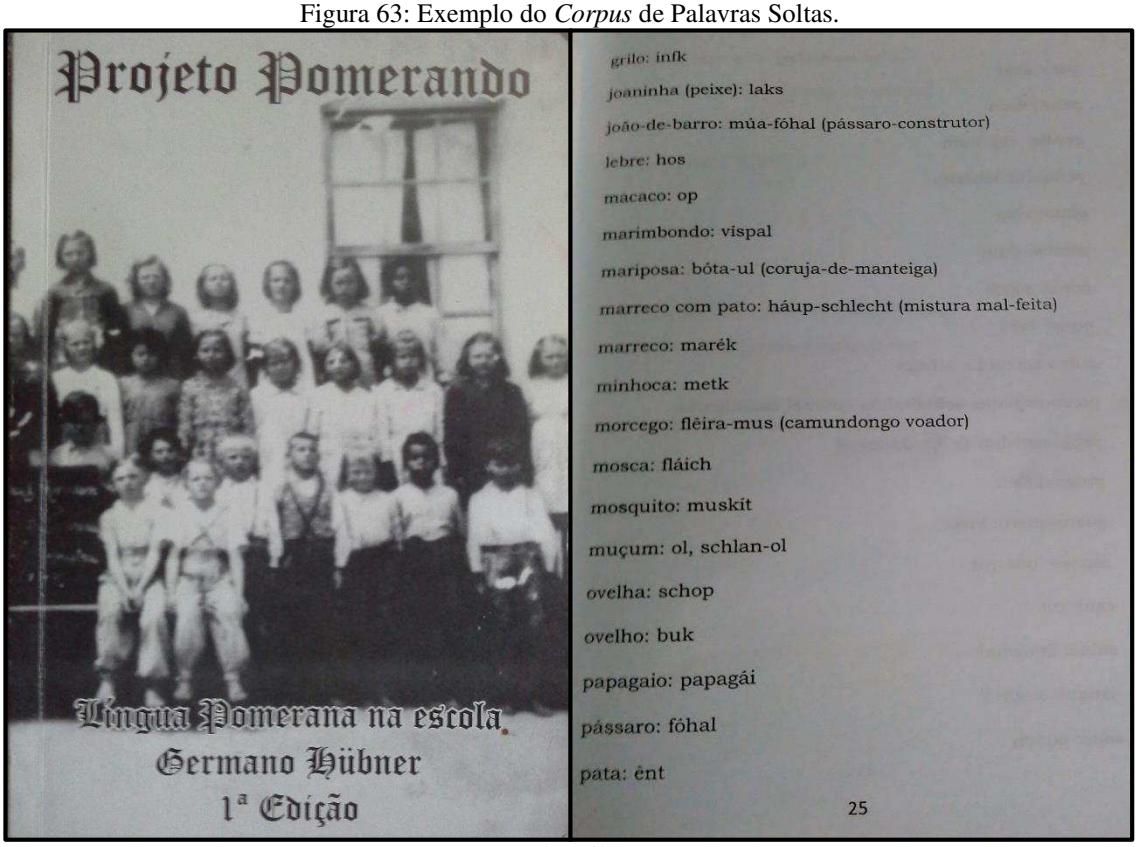
É um conjunto de arquivos com palavras soltas, porque não se trata de textos, são glossários e listas de palavras, fontes dos arquivos desse *corpus*, que possuem origens diferentes e podem coincidir alguns itens entre arquivos.

Optamos por não excluir as palavras repetidas entre as diferentes listas de palavras coletadas e por não juntar tudo em um único arquivo, decisões estas tomadas por acreditarmos que isso seria alterar os dados originais, pois as diferentes listas foram coletadas de fontes diferentes e, por isso, demonstram o vocabulário pomerano utilizado por sujeitos diferentes. As listas são úteis, pois informam a respeito do vocabulário pomerano ativo. Coletamos, por exemplo, o vocabulário utlizado por crianças no Rio Grande do Sul (glossário publicado no livro Projeto Pomerando, KUHN SILVA, 2012) e o vocabulário utilizado pela família Krause, que postou na internet o vocabulário pomerano utilizado em seu cotidiano, também em forma de glossário, com algumas traduções para o PTB.

Por meio do *Corpus de Palavras Soltas*, podemos verificar a variação em alguns itens lexicais, observar variações na grafia dos itens e contrastar as palavras isoladas com as palavras que apareceram no contexto dos *corpora* formados por textos, até mesmo para verificar se o significado da palavra isolada (seja pela tradução do glossário, seja pela tradução das nossas etiquetas) foi utilizado em sentido igual ou diferente nos contextos dos outros *corpora*. Um trabalho que vise à revitalização dialetal não poderia desprezar esse material sem correr o risco do prejuízo da perda em obter palavras raras.

Por meio da imagem a seguir, Figura 63, podemos observar um exemplo do *Corpus* de Palavras Soltas:

Figura 63: Exemplo do *Corpus* de Palavras Soltas.

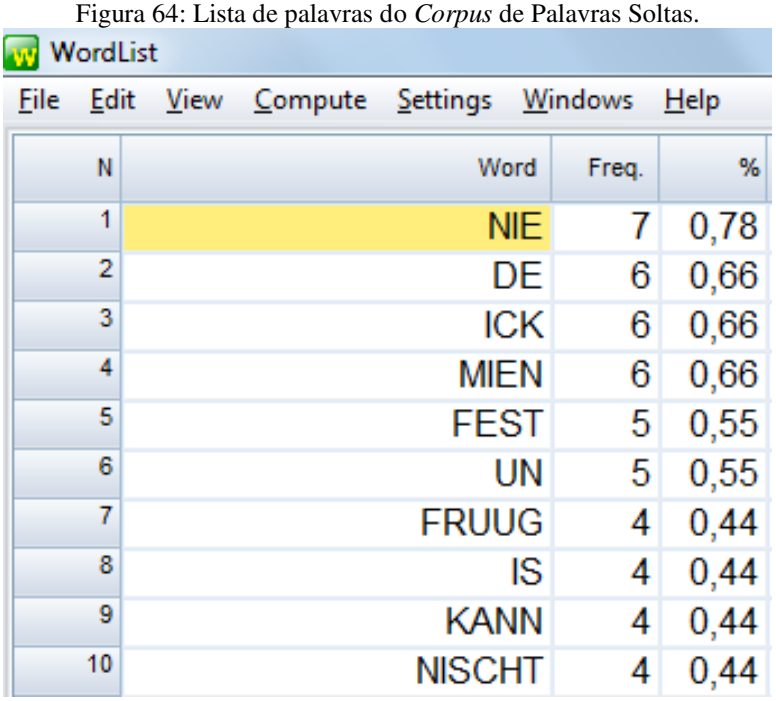


Fonte: Kuhn Silva, 2012.

Embora esse *corpus* não seja constituído de textos como tradicionalmente ocorre na LC, o mesmo é composto de itens lexicais originalmente utilizados pelos pomeranos e pode ser útil para o resgate do dialeto, documentação e até identificação de lexias já desaparecidas na antiga região da Pomerânia.

As palavras soltas reunidas pela lista de palavras do WST também podem ser contrastadas com os outros *corpora* compilados e assim contribuir para a identificação do léxico pomerano ativo no Brasil. A seguir, apresentamos um recorte da lista de palavras do *Corpus* de Palavras Soltas, conforme Figura 64.

Figura 64: Lista de palavras do *Corpus* de Palavras Soltas.



WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
1	NIE	7	0,78
2	DE	6	0,66
3	ICK	6	0,66
4	MIEN	6	0,66
5	FEST	5	0,55
6	UN	5	0,55
7	FRUUG	4	0,44
8	IS	4	0,44
9	KANN	4	0,44
10	NISCHT	4	0,44

Fonte: Elaboração própria.

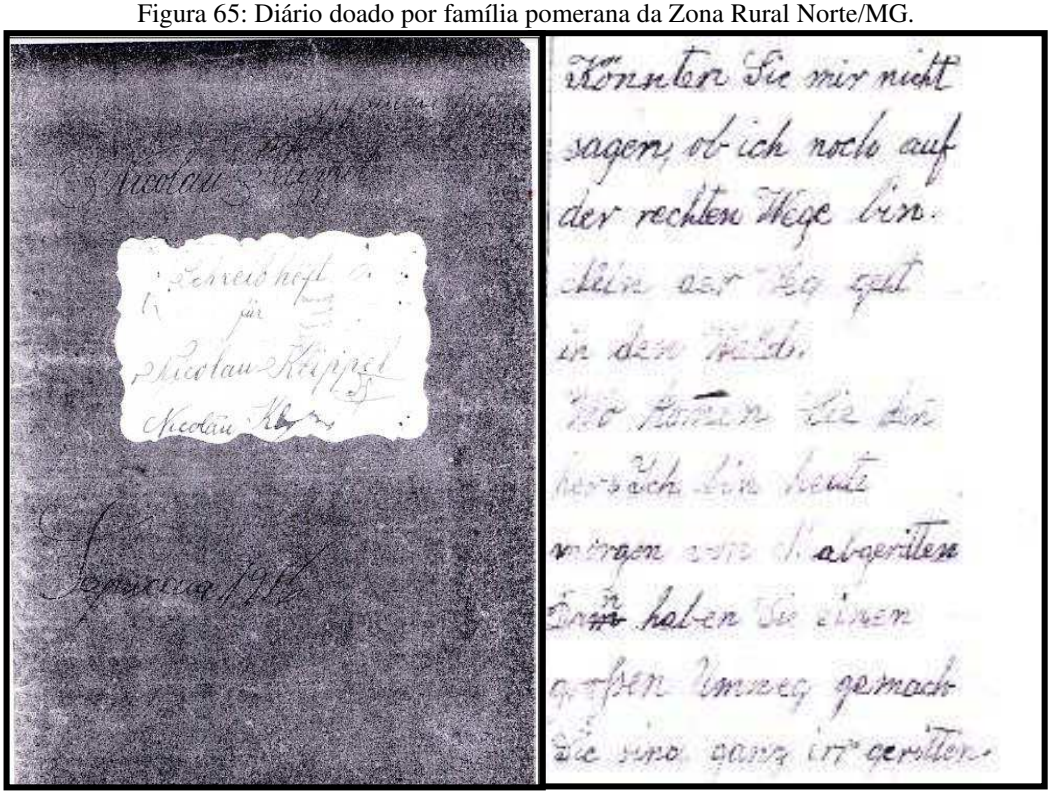
Na listagem acima podemos observar que o item mais frequente é *nie* (não, nunca), com sete ocorrências calculadas entre todos os oito arquivos desse *corpus*. Além dessa forma de negação pomerana, confirmada pelos dados empíricos, também encontramos na sétima posição da lista, o substantivo *fruug* (mulher), com 4 ocorrências, bem como o verbo *kann* (poder), nesse caso igual ao alemão-padrão, sendo que em pomerano existem as variações (variações na conjugação para os pronomes de terceira pessoa, masculino e feminino, no singular e no plural) *koinna* e *kast*, por exemplo, em *dai koinna* (Ela(s) pode(m)) e *hai kast* (Ele pode).

5.3.15 *Corpus* de diários – Resultado não alcançado

Coletamos e compilamos um diário pessoal que possui relevante riqueza lexical e de outras ordens linguísticas, conforme descrição já realizada sobre ele na metodologia deste trabalho. O diário de 1916 não pôde constituir um *corpus*, pois não conseguimos atender ao critério de um *corpus* ser constituído de dois ou mais textos (de gênero igual ou próximo), ou seja, por ser um texto único do gênero, por termos conseguido coletar apenas um diário, ele não constituiu um *corpus* (um conjunto de no mínimo dois textos) e, portanto, não conseguimos alcançar o resultado de obter um *corpus* de diários. Priorizando a noção de conjunto, o Diário não pode ser agrupado pelos critérios adotados (gênero, suporte, domínio do discurso, etc.), pois ficaria isolado sem se agrupar a outro material semelhante. No caso anterior, o do *Corpus*

de Palavra Soltas, obtivemos um conjunto de arquivos reunidos por um traço em comum, o que não ocorreu no caso do diário. Segue na Figura 65, a imagem do referido diário.

Figura 65: Diário doado por família pomerana da Zona Rural Norte/MG.



Fonte: Fotografia da cópia do diario. Arquivo do autor.

Sobre a questão do diário acima estar isolado, informamos que recebemos, posteriormente, no fim da pesquisa, um material do mesmo gênero em folhas soltas, porém já não havia mais tempo hábil para sua compilação e composição de *corpus* de diários. Esses materiais ficam como possibilidade para futuras ampliações dos *corpora*.

5.3.16 *Corpus* de material de rádios – *Scripts* – Resultado não alcançado

Uma de nossas metas era coletar material proveniente de rádios pomeranas e/ou com programação em pomerano, como por exemplo: rascunhos para apresentação dos programas em pomerano; *scripts*; áudios para transcrever; etc. Porém, procuramos algumas rádios por meio de seus sites, e-mails e telefones e não obtivemos sucesso em nenhum dos casos. Não conseguimos coletar nem receber a concessão de nenhum material proveniente de rádios. Na maioria dos casos, os e-mails nem foram respondidos. Por isso, desistimos de coletar material de rádios. Conseguimos apenas realizar um levantamento das rádios que possuem alguma

programação (parcial ou total) em pomerano ou outras variedades baixo-alemãs, como o *Plautdietsch Menonite*, e/ou alemão com traços dialetais.

Elaboramos um quadro (Quadro 6) reunindo as informações coletadas sobre as rádios no Brasil que transmitem conteúdos programáticos contendo variedades germânicas, incluindo o pomerano:

Quadro 6: Levantamento das rádios com programação em pomerano.

RÁDIOS COM PROGRAMAÇÃO EM ALEMÃO E/OU POMERANO:			
Nome	**Localidade**	**Acesso**	**Programação**
Rádio Pomerana FM	Santa Maria de Jetibá/ES	www.pomeranafm.com.br	Programação mista, com músicas em português e em pomerano.
Rádio Geração FM 107.1	Salto do Jacuí/RS	www.radiogeracao.com.br	*Deutsche Stunde* aos domingos das 7 às 9:30
Pomerisch Radio	Santa Maria de Jetibá/ES	www.pommerradio.com.br/	Programação mista, com músicas em português e em pomerano.
Rádio Kultura FM	Santa Maria de Jetibá/ES	http://www.kulturafm.com/	Programação mista, com músicas em português e em pomerano.
Radio Germânica	Blumenau/SC	http://radiogermanica.com/	Programação 99% em alemão e variações em pomerano
Rádio Kerb	Canguçu/RS	www.kerbfm.com.br	Programação mista com alguns programas em Pomerano
Rádio Pomerode	Pomerode/SC	www.radiopomerode.com.br	Programação mista com locução e músicas em pomerano, alemão e português
Radio Plautdietsch Menonite	—	www.igrejamenonita.org.br/radioweb	Programação 100% em *Plattdüütsch*
Radio *Pommerplatt*	—	http://pommerplattdutsch.blogspot.com.br/2010/07/up-pommersch-em-pomerano	Programação 100% em Plattdüütsch
Rádio Alternativa	Agudo	radioalternativa.org.br/	Programação mista com alguns programas em Pomerano
Radio Jaraguá AM	Jaraguá do Sul/SC	www.jaraguaam.com.br	Programação mista com alguns programas em Pomerano
Radio Studio FM	Jaraguá do Sul/SC	www.studiofm.com.br	*Deutsche Schalger* Dom. 06:00-8:00 horas. Programa Alemão de Sandro Schuenke
Radio São Lourenço	São Lourenço do Sul/RS	www.radiosaolourenco.com.br	Músicas em alemão e pomerano entre a programação gaúcha
Radio Litoral Sul FM	São Lourenço do Sul/RS	www.radiolitoralsulfm.com.br	Músicas em alemão e pomerano entre a programação gaúcha

Radio Venâncio Aires	Venâncio Aires/RS	www.radiovenancioaires.com.br	Encontro Alegre. Canções alemãs e germânicas da região. Seg. à Sex. 13:00-15:00
Rádio Vera Cruz	Vera Cruz/RS	http://radioveracruz.com.br	Hora Alemã
Programa Radiofônico AHAI	—	www.brasilalemanha.com.br/novo_site/paginas/programa-ahai	A Hora Alemã Intercomunitária/*Die deutsche Stunde der Gemeinden*

Fonte: Elaboração própria.

Conforme demonstrado pelo Quadro 6 acima, listamos informações sobre as rádios que possuem alguma programação com conteúdo em pomerano e/ou alemão e que são transmitidas no Brasil, todas elas também transmitidas pela Internet. Foram consideradas tanto aquelas que possuem toda a programação em alemão e que apresentam conteúdo em alemão dialetal, como a Rádio Germânica, de Blumenau/SC, como também aquelas que possuem apenas duas horas e trinta minutos semanais de conteúdo em pomerano, como a Hora Germânica ou *Deutsche Stunde*, da Rádio Geração FM de Salto do Jacuí/RS.

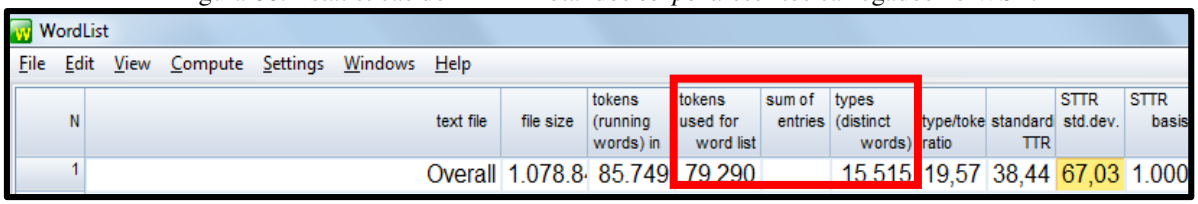
Chegamos a testar os *links* e sintonizar as rádios pela internet e ouvi-las, mas não chegamos a gravar a programação, pois não teríamos tempo hábil para fazer as transcrições, o que poderá ser feito futuramente, a fim de expandir os *corpora* do pomerano e como forma de recuperar material relevante em pomerano a partir dessas rádios que levantamos.

5.4 Resultados quantitativos do PKE em conjunto

Decidimos carregar todos os *corpora* escritos compilados no WST, a fim de verificar os resultados quantitativos que alcançamos, visto que nosso objetivo principal era coletar dados em pomerano e fazer a compilação de *corpora*.

A Figura 66, a seguir, apresenta os dados estatísticos do PKE, o total dos *corpora* escritos carregados no WST.

Figura 66: Estatísticas do PKE – Total dos *corpora* escritos carregados no WST.



Fonte: Elaboração própria.

Conforme apresentado acima, o conjunto de todos os 14 *corpora* escritos do pomerano somaram 79.290 palavras corridas e 15.515 palavras distintas, indicando uma considerável quantidade de vocabulário pomerano presente no PKE.

5.5 O *Pommersch Korpus Oral – PKO* como resultado da segunda fase de compilação

O conjunto das 19 entrevistas transcritas constituiu nosso *Corpus* Oral de Entrevistas Interativas – COEI chamado assim individualmente em português, mas nomeado definitivamente de PKO – *Pommersch Korpus* Oral no conjunto do PK.

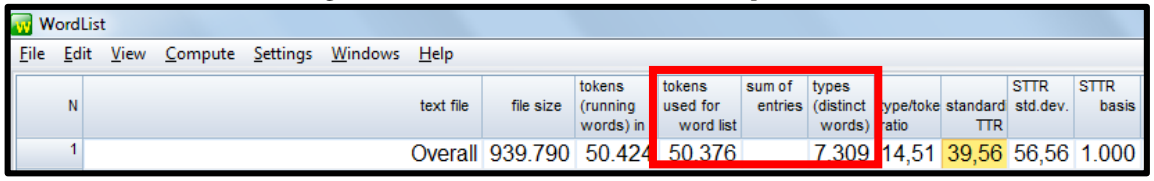
O PKO é constituído das respostas aos LSF/QSL, pois editamos os arquivos de transcrição para a versão final do PKO, deixando apenas as respostas e excluindo as transcrições das nossas falas, proferidas enquanto sujeito-entrevistador, durante as interações nas entrevistas.

Atingimos um total de 25 arquivos de transcrição das falas, pois algumas entrevistas tinham parte um, dois e três, devido a interrupções ou término das baterias. Então cada arquivo de áudio foi transcrito em um arquivo de texto. Os dados contêm entrevistas exclusivamente faladas em alemão e pomerano, como também entrevistas em que os falantes alternavam entre português e pomerano, e ainda, entrevistas em que os pomeranos respondiam sobre o item perguntado em pomerano, mas faziam uma conversação em português.

Notamos que em todas as entrevistas realizadas há itens em comum, iguais, entre alemão e pomerano, tendo sido impossível obter uma entrevista contendo falas em um pomerano totalmente diferente do alemão, em que não houvesse pelo menos alguns poucos itens coincidentes. A maioria dos falantes entrevistados, no entanto, não tem clara percepção de que aqueles itens são falados de forma igual também em alemão, porque para eles é tudo pomerano.

Contabilizamos os dados quantitativos do PKO e todos os arquivos somaram 50.376 palavras corridas e 7.309 palavras distintas, conforme demonstrado por meio da Figura 67, a seguir.

Figura 67: Estatísticas do *Pommersch Korpus* Oral.



Fonte: Elaboração própria.

Ao observar os dados do PKO, percebemos que o nosso *corpus* oral apresenta respostas coincidentes às questões propostas nas entrevistas (realizadas em diferentes regiões do Brasil)

e também uma considerável variação lexical, pois ocorreram variações nas respostas e também variações linguísticas regionais para um mesmo item perguntado, permitindo assim que coletássemos itens singulares em pomerano, lembrando que, para nós, já existe um *Brasilianisch-Pommersch*, fruto do contato linguístico.

Ao fazer uma breve busca pelo PKO, buscando as respostas para "madeira" em pomerano, encontramos tanto *Holt*, com 4 ocorrências, quanto *Holz* (alemão-padrão), com 3 ocorrências. Também encontramos uma variação para *Würmer* (verme, bicho de pau podre), que foi *Holtruupper*, na posição 462, com 2 ocorrências, conforme demonstra a Figura 68:

Figura 68: Lista de palavras do PKO.

WordList

File Edit View Compute Settings Windows Help

N	Word	Freq.	%
453	HINGE	2	0,05
454	HINKED	1	0,02
455	HINKZ	1	0,02
456	HINZ	1	0,02
457	HOCHDEUTSCH	4	0,09
458	HOCHTIED	3	0,07
459	HOCHTIEDBIRER	4	0,09
460	HOCHTIEDFEST	4	0,09
461	HOLT	4	0,09
462	HOLTRUUPPER	2	0,05
463	HOLZ	3	0,07
464	HOOHN	1	0,02

Fonte: Elaboração própria.

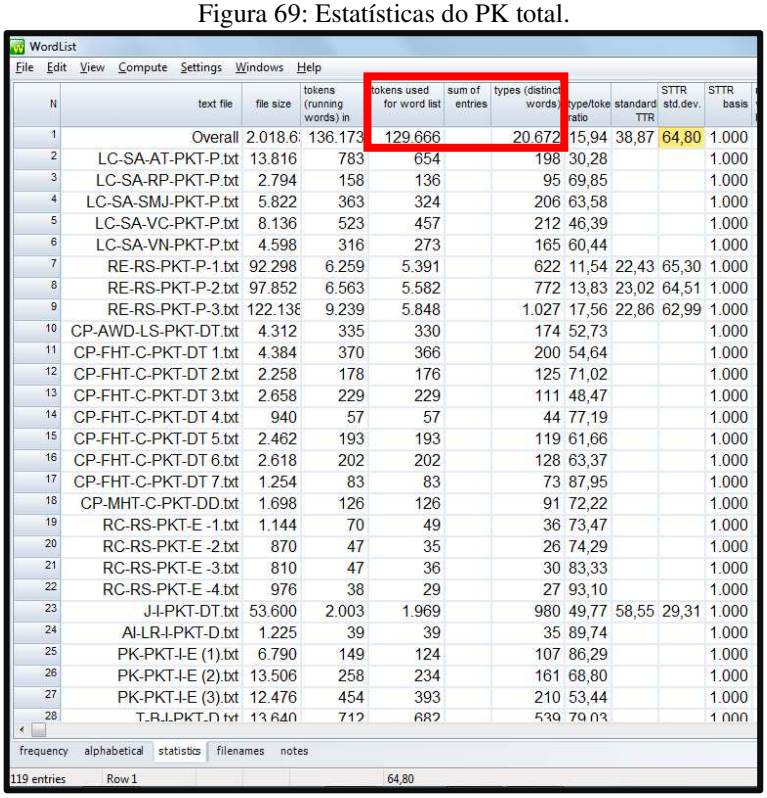
Ao observar a lista de palavras do PKO, em ordem alfabética, percebemos algumas variações, como a frequência do item *aber* (mas), que ocorreu 70 vezes no PKO (posição 115) e que apresentou no mesmo *corpus* a variação *awer* com 42 ocorrências (posição 169) e comparado com o *Corpus de Textos Religiosos*, apresentou ainda outra variação, *owa*, que ocorreu 13 vezes (posição 87). Com base nesse exemplo, acreditamos que muitas análises podem ser realizadas com base nos dados do PK (PKE e PKO), tanto de forma individual, com base em cada *subcorpus*, quanto comparando *corpora* diferentes.

5.6 Resultados quantitativos do PK (PKE e PKO)

Quanto aos dados quantitativos do nosso resultado geral e produto principal do nosso trabalho, o conjunto do *Pommersche Korpora* – PK, constituído pelo PKE e pelo PKO,

apresentamos nesse momento o total de dados obtidos por nosso processo de compilação, incluindo o conjunto formado por todos os 15 *corpora* do pomerano, conforme a Figura 69, a seguir:

Figura 69: Estatísticas do PK total.



WordList

File Edit View Compute Settings Windows Help

N	text file	file size	tokens (running words) in	tokens used for word list	sum of entries	types (distinct words)	type/token ratio	standard TTR	STTR std.dev.	STTR basis
1	Overall	2.018.6	136.173	129.666		20.672	15,94	38,87	64,80	1.000
2	LC-SA-AT-PKT-P.txt	13.816	783	654		198	30,28			1.000
3	LC-SA-RP-PKT-P.txt	2.794	158	136		95	69,85			1.000
4	LC-SA-SMJ-PKT-P.txt	5.822	363	324		206	63,58			1.000
5	LC-SA-VC-PKT-P.txt	8.136	523	457		212	46,39			1.000
6	LC-SA-VN-PKT-P.txt	4.598	316	273		165	60,44			1.000
7	RE-RS-PKT-P-1.txt	92.298	6.259	5.391		622	11,54	22,43	65,30	1.000
8	RE-RS-PKT-P-2.txt	97.852	6.563	5.582		772	13,83	23,02	64,51	1.000
9	RE-RS-PKT-P-3.txt	122.13	9.239	5.848		1.027	17,56	22,86	62,99	1.000
10	CP-AWD-LS-PKT-DT.txt	4.312	335	330		174	52,73			1.000
11	CP-FHT-C-PKT-DT 1.txt	4.384	370	366		200	54,64			1.000
12	CP-FHT-C-PKT-DT 2.txt	2.258	178	176		125	71,02			1.000
13	CP-FHT-C-PKT-DT 3.txt	2.658	229	229		111	48,47			1.000
14	CP-FHT-C-PKT-DT 4.txt	940	57	57		44	77,19			1.000
15	CP-FHT-C-PKT-DT 5.txt	2.462	193	193		119	61,66			1.000
16	CP-FHT-C-PKT-DT 6.txt	2.618	202	202		128	63,37			1.000
17	CP-FHT-C-PKT-DT 7.txt	1.254	83	83		73	87,95			1.000
18	CP-MHT-C-PKT-DD.txt	1.698	126	126		91	72,22			1.000
19	RC-RS-PKT-E -1.txt	1.144	70	49		36	73,47			1.000
20	RC-RS-PKT-E -2.txt	870	47	35		26	74,29			1.000
21	RC-RS-PKT-E -3.txt	810	47	36		30	83,33			1.000
22	RC-RS-PKT-E -4.txt	976	38	29		27	93,10			1.000
23	J-I-PKT-DT.txt	53.600	2.003	1.969		980	49,77	58,55	29,31	1.000
24	AI-LR-I-PKT-D.txt	1.225	39	39		35	89,74			1.000
25	PK-PKT-I-E (1).txt	6.790	149	124		107	86,29			1.000
26	PK-PKT-I-E (2).txt	13.506	258	234		161	68,80			1.000
27	PK-PKT-I-E (3).txt	12.476	454	393		210	53,44			1.000
28	T-B-I-PKT-D.txt	13.640	712	682		539	79,03			1.000

frequency alphabetical statistics filenames notes

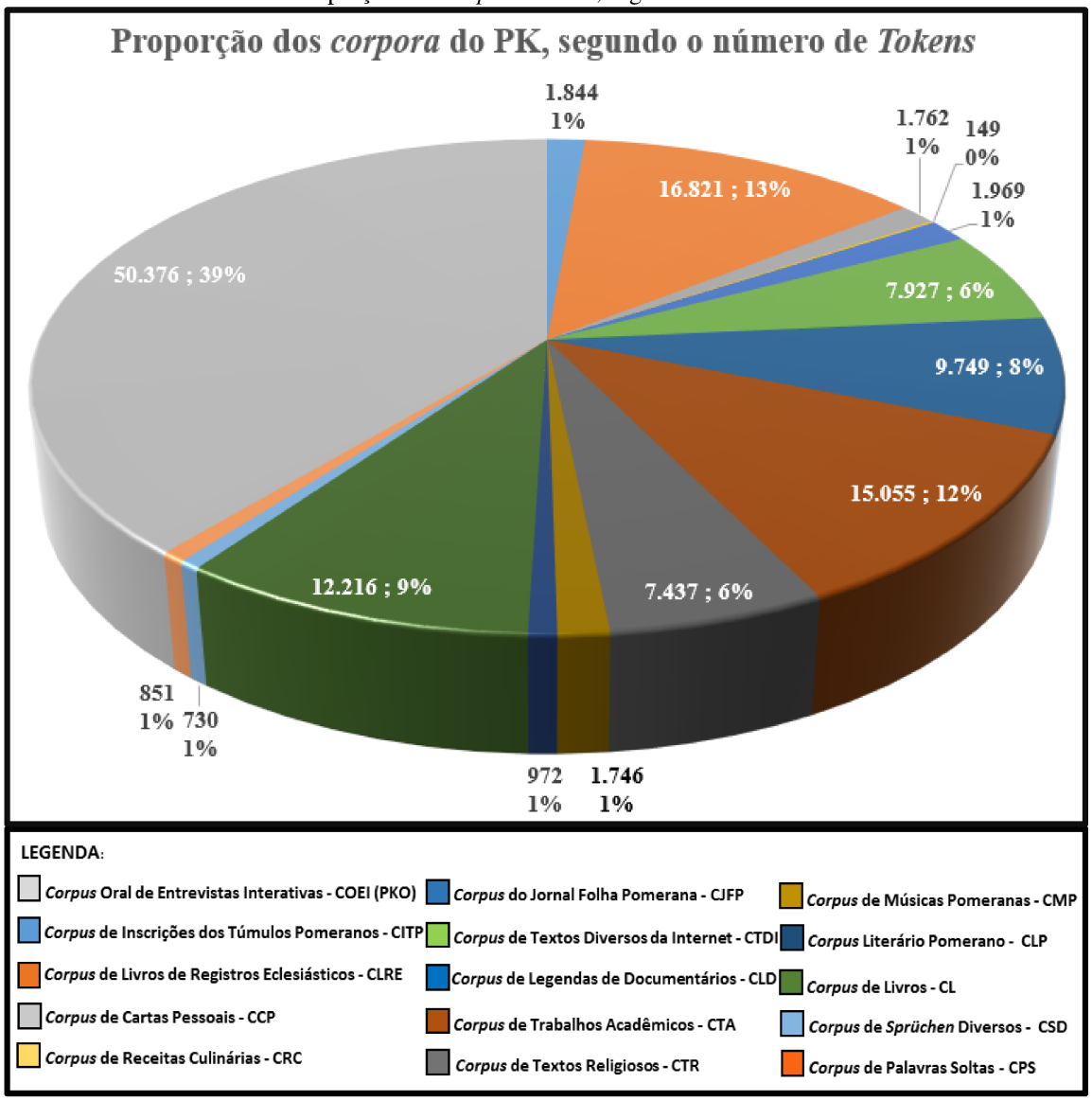
119 entries Row 1 64,80

Fonte: Elaboração própria.

Segundo a contagem realizada pelo WST 6.0, conseguimos compilar e reunir 129.666 palavras corridas e 20.672 palavras distintas, quantidade expressiva, tendo em vista que o tamanho do PK é considerável em relação ao contexto pomerano.

Apresentamos a seguir o Gráfico 4, com a distribuição dos *corpora*, incluindo o nosso *corpus* oral, dentro do conjunto do PK, a fim de permitir uma visão proporcional de todo o conjunto de dados que compilamos.

Gráfico 4: Proporção dos *corpora* do PK, segundo o número de *Tokens*.



Fonte: Elaboração própria.

Também consideramos ser útil a elaboração de uma tabela reunindo todos os dados quantitativos dos *corpora* compilados, além de auxiliar na identificação dos dados do Gráfico 4 acima, no qual cada *corpus* pode ser identificado, pelas cores ou pelo número de *tokens*. Expomos a seguir, a Tabela 3.

Tabela 3: Dados quantitativos parciais e totais do *Pommersche Korpora.*

DADOS QUANTITATIVOS PARCIAIS E TOTAIS DO *POMMERSCHE KORPORA*		
Corpus	***Tokens***	***Types***
1. *Corpus* de Inscrições dos Túmulos Pomeranos - CITP	1.844	640
2. *Corpus* de Livros de Registros Eclesiásticos - CLRE	16.821	1.558
3. *Corpus* de Cartas Pessoais - CCP	1.762	763
4. *Corpus* de Receitas Culinárias - CRC	149	81
5. *Corpus* do Jornal Folha Pomerana - CJFP	1.969	980
6. *Corpus* de Textos Diversos da Internet - CTDI	7.927	2.942
7. *Corpus* de Legendas de Documentários - CLD	9.749	2.470
8. *Corpus* de Trabalhos Acadêmicos - CTA	15.055	3.554
9. *Corpus* de Textos Religiosos - CTR	7.437	1.759
10. *Corpus* de Músicas Pomeranas - CMP	1.746	647
11. *Corpus* Literário Pomerano - CLP	972	531
12. *Corpus* de Livros - CL	12.216	3.861
13. *Corpus* de *Sprüche* Diversos - CSD	730	405
14. *Corpus* de Palavras Soltas - CPS	851	698
15. *Corpus* Oral de Entrevistas Interativas - COEI (PKO)	50.376	7.309
Pommersche Korpora Escritos -PKE	79.290	15.515
Pommersch Korpus Oral - PKO	50.376	7.309
***Pommersche Korpora* - PK**	**129.666**	**20.672**

Fonte: Elaboração própria.

Observamos que os dados do PK listados na tabela acima não se referem à soma dos dados de cada *corpora*, mas referem-se à contagem realizada a partir do carregamento de cada *corpus* no WST, separadamente. Depois fizemos o carregamento de todos os arquivos de todos os *corpora*, incluindo o *corpus* oral. Também ressalvamos que escolhemos citar os *Tokens used for Wordlist* (*Tokens* usados para a *Wordlist*) como *Tokens* (palavras corridas), pois os *tokens* totais (*running words in text*) levam em consideração a presença de números nos *corpora*, representados por "#" (*hashtag*) e nossa intenção era calcular palavras, excetuando-se as numerações presentes nos dados.

5.7 Resultados referentes aos objetivos específicos

Retomamos, a seguir, os dois objetivos específicos da nossa dissertação:

1) Tentar identificar a sobrevivência ou o desaparecimento do pomerano nas regiões abrangidas pela coleta, principalmente nos Vales do Rio Doce/MG e do Rio Pardo/RS;

2) Comparar, brevemente, o pomerano coletado nas duas regiões principais que a pesquisa abrangeu (MG e RS), e apresentar a avaliação de alcance ou não dos mesmos por meio da pesquisa e dos *corpora* compilados, respectivamente.

Quanto ao primeiro objetivo específico, podemos responder que, ao buscar material em pomerano, constatamos que o mesmo ainda sobrevive (está em uso) no Vale do Rio Doce em Minas Gerais, pois encontramos falantes ativos entre as faixas etárias II e III principalmente, meia-idade e terceira idade; encontramos materiais escritos, bem como notamos que um sujeito-entrevistado de 93 anos de idade sabia escrever palavras em pomerano, por exemplo, *Schope* (*Schoope*, em HD: *Schaf*, em PTB: ovelha), sem consulta ao dicionário pomerano.

Embora seja raro encontrar crianças e adolescentes falantes nessa região, notamos crianças que ainda sabem pequenas músicas, "trovinhas" e também algumas palavras, pois circunstanciaram durante as entrevistas algumas interrupções, com manifestações espontâneas de crianças, que não entrevistamos por não termos solicitado essa autorização ao CEP. O sistema de ensino público no leste de Minas Gerais não abrange o pomerano e, portanto, a tendência é que as novas gerações tenham dificuldades em manter a fala pomerana.

Já no Vale do Rio Pardo/RS, não conseguimos encontrar falantes de pomerano para realizar entrevistas com eles e, quanto aos materiais escritos, só encontramos os registros e lápides, que datam de mais de um século atrás. Então, inferimos que o pomerano não está tão preservado nessa região, embora o alemão-padrão e genérico (conforme discutido no capítulo 2) esteja ativo na cidade ao lado, Santa Cruz do Sul, e lembramos que o Gráfico 3 (tópico 5.3.2), já apresentado, demonstra a descendência majoritária de pomeranos em Vera Cruz. Contudo, esclarecemos que nossa avaliação não é conclusiva sobre esse ponto específico, pois precisaríamos voltar à localidade e insistir, inclusive na zona rural, na busca por falantes de pomerano. Assim, quanto a identificar a sobrevivência do pomerano nessa região, não podemos fazer afirmações categóricas como afirmar que essa variedade não tenha sobrevivido em Vera Cruz, apenas constatamos que na zona urbana o pomerano não está muito preservado, pela dificuldade em encontrar evidências atuais do mesmo na referida localidade. O fato de não encontrarmos vestígios recentes de fala e da escrita pomerana na localidade poderia indicar que o pomerano estaria extinto ali, porém não podemos afirmar isso apenas com base em tentativas pontuais de coletar dados, sendo necessário um aprofundamento das investigações na localidade.

Quanto ao segundo objetivo específico, podemos responder, principalmente com base nos dados do PKO que, ao comparar as duas regiões abrangidas (todas as localidades de Minas Gerais e todas as localidades do Rio Grande do Sul pesquisadas), encontramos muitos itens lexicais idênticos, mas também alguns itens lexicais que apresentaram variações.

Como exemplos de itens pomeranos idênticos em Minas Gerais e Rio Grande do Sul, citamos: para a pergunta *Wie nennt man dies körperlich teil*? (Como se chama esta parte do corpo, apontando para a unha do dedo da mão) a resposta em ambas as regiões foi *Finhanoogel* (19 ocorrências, posição 755), e observamos que para a unha do pé o nome seria outro (*Teehnanoogel*). Para a pergunta *Der Blume ... gelb, rund, mit einer Scheibe mit Samen in der Mitte? Wie nennt man das?* (Uma flor... amarela, redonda, com uma rodela de sementes no meio? Como se chama isso?) a resposta em ambas as regiões foi *Süünnblaume* (23 ocorrências, posição 630) para designar o Girassol. Dentre muitos outros casos em que coincidiram as respostas entre as regiões, temos *Komellatee* (para Camomila), *Oossvoogel* (para Urubu), etc.

Como exemplos de itens lexicais com variações entre Minas Gerais e Rio Grande do Sul, citamos: variação, principalmente na pronúncia da sílaba final, com a presença das síladas "*nd"* no Rio Grande do Sul. Para a pergunta *Wie nennt man dies körperlich teil*? (Como se chama esta parte do corpo, apontando para o olho) em Minas Gerais obtivemos a resposta *Oucho* (31 ocorrências, posição 453) e no Rio Grande do Sul obtivemos a resposta *Ouchend* (9 ocorrências, posição 1.602).

Para a pergunta *Die Früchte ... kleiner als Orangen, die mit der Hand geschält werden, und in der Regel einen Duft auf der Hand lassen? Wie nennt man das?* (As frutas pequenas como a laranja, que se descascam com a mão, e, normalmente, deixam um cheiro na mão? Como se chama isso?), correspondentes em português a mexerica ou bergamota, obtivemos o seguinte resultado: em Minas Gerais obtivemos a resposta *Krawe* (12 ocorrências no PK, posição 1.160) e no Rio Grande do Sul obtivemos a resposta *Peechamota* (4 ocorrências, posição 3.665).

Em Minas Gerais notamos alguns pratos da culinária brasileira e mineira que tinham seus nomes em português desconhecidos pelos falantes como curau de milho e mingau de milho, o máximo que conseguimos obter foi *miechpapa* (papa de milho, com 8 ocorrências lematizadas), *miechmelka* (milho com leite, com 1 ocorrência) e *miechtouce* (milho doce, com 2 ocorrências).

Dessa forma, pudemos comparar o pomerano falado nas duas regiões de escopo da nossa pesquisa e avaliamos que existem algumas variações da fala pomerana entre as regiões, mas

também existe uma norma pomerana, pois para a maioria dos itens lexicais perguntados as respostas foram as mesmas.

Assim, expomos neste capítulo, quais foram os resultados alcançados por nossa pesquisa, além de descrevê-los, comentá-los e analisá-los brevemente. A seguir, realizaremos a análise dos resultados, com foco na classificação dos *corpora*, na testagem das nossas hipóteses de pesquisa e em alguns fatos para os quais os *corpora* nos direcionaram.

6. ANÁLISE DOS RESULTADOS

Neste momento, retomaremos nossos resultados principais, apresentados no capítulo anterior, para analisá-los, começando por fazer a classificação dos *corpora* escritos em conjunto – o PKE, e também do *corpus* oral – PKO, que obtivemos como resultados do nosso trabalho.

Também retomaremos as nossas quatro hipóteses de pesquisa, para avaliar se obtivemos respostas aos questionamentos que propomos. E faremos, ainda, outras análises básicas a partir da observação de alguns fatos linguísticos direcionados pelos *corpora*, coletados com metodologia empírica e estudados de maneira descritiva com base no método da LC, a saber, geração e observação de listas de palavras e linhas de concordâncias.

6.1 Classificação e tipologia geral dos *corpora*

Realizaremos, neste tópico, a classificação geral e a tipologia dos *corpora* escritos, o PKE, em conjunto, segundo a taxonomia já desenvolvida na área da LC. Na sequência, realizaremos a classificação e a tipologia a respeito do nosso *corpus* oral, o PKO.

6.1.1 Classificação geral e tipologia dos *corpora* escritos, o PKE

Com base na tipologia de *corpus*, proposta por Berber Sardinha (2004) e posteriormente ampliada por Teixeira (2008), realizamos uma classificação e tipologia do PKE:

a. Tipo: dialetal e dialetal-regional – variedade linguística pomerana proveniente do *Pommersches Plattdüütsch* – baixo-alemão pomerano – das diatopias mineira (leste de Minas Gerais) e sul-rio-grandense;
b. Conteúdo: multilíngue e multivarietal – contém amostras conjuntas, dentro dos mesmos textos, de pomerano, alemão-padrão, interferências de outras variedades alemãs, como HR e também do português, com indicativos de já se constituir como uma variedade brasileira do pomerano proveniente dos contatos de línguas;
c. Modos: escrito, composto de textos escritos impressos e escritos não-impressos, manuscritos, textos digitais, imagens digitais de textos e/ou imagens com textos, textos pictóricos, com gravações entalhadas em cimento, em madeira, e também escritos em tecidos (bordados);
d. Autorias: de língua de "falantes nativos" do pomerano. Consideramos como "pomeranos nativos" aqueles nascidos no Brasil, que adquiriram primeiramente o pomerano no contexto familiar, como língua materna e só depois, em contexto escolar, aprenderam o português. Consideramos também a autodeterminação identitária. Os textos foram produzidos por diversas

autorias, entidades, representantes eclesiásticos, etc. Alguns conteúdos pertencem à tradição oral, e por isso não têm autoria definida, como ditados e trovinhas, outros têm autorias coletivas como poemas e canções tradicionais, ou têm autoria acadêmica de pesquisadores;

e. Seleções: por amostragem (*sample corpus*), composto por porções de textos ou de variedades textuais, planejado para ser uma amostra da linguagem pomerana como um todo, já que é um *corpus* geral de uma variedade de língua;

f. Tamanho e Representatividade: pequeno (PKE: 79.290 *tokens*). Porém, significativo e representativo do contexto atual do pomerano. Seria classificado como "pequeno", se considerássemos os parâmetros de tamanho da LC[72], porém, o tamanho dos *corpora* é tudo que foi factível reunir por uma única pesquisadora em dois anos;

g. Finalidade: de estudo, de registro e de descrição do pomerano, e para futura composição de banco de dados (com gerenciamento e interface de consulta) para estudos em LC;

h. Tempo: histórico e contemporâneo. Os diversos *corpora* escritos contêm textos com diferentes datas de produção e de conteúdo, assim temos em diferentes textos, um espaço de tempo representado, cada um em sua sincronia. Exemplo: *Corpus* de Registros Eclesiásticos – 1859-1863; histórico, pois representa um período de tempo passado e contém textos de valor documental. O *corpus* de músicas contém paródias pomeranas de músicas famosas na atualidade, ou seja, textos contemporâneos. Portanto, o PKE, de modo geral, é histórico e contemporâneo. Os diversos recortes temporais presentes nos *corpora* permitem estudos diacrônicos, com vários períodos de tempo não lineares. Permite estudos sincrônicos (que recortam um período de tempo específico) e estudos diacrônicos (que utilizam textos representativos de diferentes períodos de tempo)[73].

i. Balanceamento: não balanceado – impossibilidade diante do contexto de amostragem definido e do objetivo de resgate de uma variedade dialetal em processo de diminuição considerável do número de falantes. Não é possível distribuir equitativamente os textos de forma que haja um equilíbrio considerável entre eles. Para isso, teríamos que considerar somente textos produzidos em quantidades e gêneros iguais;

---

[72] Segundo Berber Sardinha (2004, p. 26), quanto aos parâmetros de tamanho da LC, até o ano de 2004, um *corpus* de menos de 80 mil palavras seria classificado como pequeno; um *corpus* de 80 a 250 mil palavras seria classificado como pequeno-médio; um *corpus* de 250 mil a 1 milhão de palavras seria classificado como médio; um *corpus* de 1 milhão a 10 milhões de palavras seria classificado como médio-grande e um *corpus* de 10 milhões ou mais de palavras seria classificado como grande.

[73]As definições de *corpus* do tipo contemporâneo, histórico, sincrônico e diacrônico, bem como a classificação dos *corpora* nesses termos foram baseadas nos trabalhos já realizados por Berber Sardinha (2004, p. 20) e Teixeira (2008, p. 162).

j. Integralidade: composto de textos integrais, textos curtos, trechos de livros e alguns *corpora* de fragmentos como frases, versos e até palavras soltas (para resgatar o léxico);

k. Fechamento/Status: estático, composição encerrada no contexto desta pesquisa. Aguardando validação;

l. Disposições internas: comparável multilíngue contatual[74].

m. Nível de codificação: parcialmente etiquetado.

6.1.2 Classificação e tipologia do *corpus* oral, o PKO

Com base nos mesmos critérios do tópico 6.1.1, realizamos uma classificação e tipologia do PKO:

a. Tipo: dialetal e dialetal-regional – variedade linguística pomerana proveniente do *Pommersches Plattdüütsch* – baixo-alemão pomerano – das diatopias mineira (leste de Minas Gerais) e sul-rio-grandense;

b. Conteúdo: bilíngue e/ou multilíngue – contém amostras conjuntas, dentro dos mesmos diálogos, de pomerano, alemão e português, o que é característico da variedade brasileira do pomerano e do fato dos pomeranos serem bilíngues em pomerano-português e/ou trilíngues em pomerano-alemão-português;

c. Modo: falado, não-alinhado em som-escrita, composto pelas transcrições das falas;

d. Autoria: de língua de "falantes nativos" do pomerano. Consideramos como "pomeranos nativos", mesmo aqueles nascidos no Brasil, mas que adquiriram primeiramente o pomerano no contexto familiar, como língua materna e depois aprenderam o português;

e. Seleção: por amostragem (*sample corpus*), composto por porções de fala, constitui uma amostra finita da linguagem pomerana como um todo;

f. Tamanho e Representatividade: pequeno (PKO: 50.376 *tokens*), porém significativo e representativo do contexto atual do pomerano (já discutido no Capítulo 2). O tamanho do PKO

[74] Contatual, neste contexto, refere-se a apresentação de indícios do contato de línguas, pois notamos a presença de dados multilíngues dentro de um mesmo texto e dentro de uma mesma fala nos *corpora* do pomerano. Não estaríamos classificando com precisão apenas como multilíngue, pois poderia levar ao equívoco de se pensar que os *corpora* são formados por textos diferentes em línguas diferentes, mas o fenômeno acontece de uma só vez, em uma mesma fonte, como indício do contato de línguas pomerano-alemão-hunsriqueano-português. Assim como utiliza-se o termo contatual para "Atlas Linguístico-Contatual" e para "Geolinguística Pluridimensional Contatual" (ALTENHOFEN, 2013), inferimos ser possível utilizar contatual para *corpus* contatual e *corpora* contatuais, pois os dados da variedade linguística pomerana não estão isolados nas evidências empíricas.

é considerável em se tratando das peculiaridades de compilação de *corpus* oral, do tempo de transcrição e da quantidade de compiladores (apenas uma, no nosso caso);

g. Finalidade: de estudo;

h. Tempo: Contemporâneo. Permite estudos sincrônicos. Composto de falas gravadas no período em que foram realizadas, na época atual (2015);

i. Balanceamento: não balanceado – foram feitas 148 perguntas pelo método do LSF/QSL, então a variação do número de palavras das respostas seria pequena, permitindo assim um balanceamento entre as 20 entrevistas, sendo 10 em cada região e 5 em cada localidade dividida entre zona rural e zona urbana. Porém, tornou-se não-balanceado, pois foi necessário descartar duas entrevistas e realizar uma entrevista extra, de modo que se efetivaram 19 entrevistas transcritas, sendo algumas mais longas e outras mais curtas devido à espontaneidade da fala;

j. Integralidade: composto de entrevistas integrais. Embora a fala seja segmentada, todas as respostas às questões, falas, diálogos e narrativas dos sujeitos participantes como entrevistados foram transcritas por inteiro, excluídas as falas que poderiam trazer risco de identificação dos mesmos;

k. Fechamento/Status: estático. Composição encerrada no contexto desta pesquisa, contendo 19 entrevistas transcritas. Aguardando validação;

l. Disposições internas: comparável multilíngue contatual.

m. Nível de codificação: parcialmente etiquetado.

Observamos quanto ao tamanho que a classificação dos *corpora* foi realizada de forma individual, por isso, o PKE e o PKO foram considerados de tamanho pequeno, por estarem dentro do parâmetro de até 80 mil palavras. Porém, se considerarmos o tamanho total do cojunto dos *corpora* (PKE + PKO), nosso PK é classificado como pequeno-médio, pois atinge 129.666 *tokens* (conforme parâmetros já explicados na nota 80). Também classificamos o PK, no geral, como um conjunto de *corpora* dialetais multilíngues contatuais.

6.2 Testagem e avaliação das hipóteses de pesquisa

1) Primeira hipótese – O pomerano não é ágrafo – Confirmada

O pomerano não é e não foi ágrafo no Brasil, pois a coleta e compilação de *corpora* demonstrou que existem formas de escrita anteriores à publicação do dicionário pomerano-português no Brasil em 2006 (TRESSMANN, 2006). Os textos pomeranos coletados confirmam a hipótese já baseada na literatura pomerana pré-existente à imigração. Além de

havermos encontrado textos com traços dialetais pomeranos, presentes em cartas e textos antigos, também encontramos bíblias em baixo-alemão pomerano no leste de Minas Gerais, entre descendentes de pomeranos, que ainda as utilizam para sua leitura e, também, existiam outras formas de escritas anteriores.

Não estamos aqui julgando o mérito das formas de escrita, apenas constatando que o fato de termos encontrado textos escritos com traços dialetais pomeranos comprovam que o pomerano não é ou era ágrafo, mas o que aconteceu no Brasil foi um processo de perda da cultura escrita, perda do conhecimento e prática da escrita na forma germânica que os imigrantes trouxeram da sua origem. Processo este influenciado, em parte, pelo fim das escolas coloniais alemãs, a partir da proibição em 1938, pelo Decreto-Lei nº. 383, conforme já relatado no capítulo 2 deste trabalho. E também houve uma desconsideração de fontes primárias produzidas espontâneamente por pomeranos enquanto vestígios de escrita pomerana, por distoarem da nova proposta de escrita do pomerano, dicionarizada em 2006.

A própria constatação da diversidade de escritas pomeranas que encontramos durante o processo de coleta sugere a não-agrafia. Essas formas de escrita vão desde materiais mais antigos que seguem traços da família germânica, próximas ao HD, até outras que se distanciam com elementos da escrita moderna do pomerano, proposta por Tressmann (2006). E, ainda outras vezes, foram encontradas escritas que seguem a interpretação e intuição dos falantes de pomerano, aqueles que não tiveram oportunidade de alfabetização em escrita alemã, mas que grafam o pomerano seguindo as regras ortográficas do português, como fez Kuhn Silva (2012).

Com base nisso, tentamos desmistificar a suposta agrafia e lembrar a necessidade de haver ensino de escrita pomerana em forma germânica, resgatando o conhecimento da cultura que os pomeranos e seus descendentes já tinham internalizado e podendo, inclusive, utilizar as escritas já existentes como recursos paralelos de auxílio no processo de ensino-aprendizagem.

Outro fato que reforça a ideia de que não se trataria de um caso de língua ágrafa é todo um conjunto de produção escrita, existente há vários séculos, se considerarmos, por exemplo, a *Barther Bibel*, publicada desde 1588 e conhecida como a bíblia pomerana, bem como toda a literatura pomerana pré-existente (exemplificada, no Capítulo 2). Assim, acreditamos que os pomeranos que vieram para o Brasil já conheciam algum tipo de escrita, mesmo que não-padrão entre si e mesmo que apresentando traços linguísticos que os diferenciassem da fala do alto-alemão.

Também lembramos, conforme mencionado no item 5.8, que encontramos um sujeito-entrevistado que soube escrever em pomerano sem consulta ao dicionário.

2) Segunda hipótese – Inexistência de falantes de pomerano na zona urbana – Negada

No início da imigração, as comunidades pomeranas estabeleceram-se no campo, em áreas predominantemente rurais. Ainda hoje os pomeranos são considerados como um grupo predominantemente camponês e o censo demográfico das localidades pesquisadas apontaram para a maioria de população rural. Por isso, acreditávamos que os falantes de pomerano seriam encontrados quase exclusivamente na zona rural, podendo ser inexistente, em algumas localidades, a presença de falantes de pomerano na zona urbana, mas essa hipótese foi negada, visto que encontramos falantes de pomerano na zona urbana de Itueta/MG e também na zona urbana de Canguçu/RS. Portanto, consideramos que nossa segunda hipótese não se confirmou, pois não é possível generalizar e nem afirmar que só se encontra falantes de pomerano na zona rural.

Os pomeranos estão atualmente muito ativos nas redes sociais, nos diversos meios de comunicação e muitos são os pomeranos que ingressam no ensino superior, tornam-se mestres e doutores (como Küster, Kuhn Silva, Seibel, Tressmann, etc.). Portanto, acreditamos que a visão do pomerano como camponês tende a mudar, no sentido de que eles não estão mais tão isolados em áreas rurais, onde existe inclusive acesso à internet.

3) Terceira hipótese – Contato do pomerano com outras variedades – Confirmada

Nos pontos de pesquisa selecionados como principais – Vale do Rio Pardo/RS e Vale do Rio Doce/MG – houve, de fato, o contato de pomeranos com outras etnias e os dados do PK apresentam evidências desse contato interdialetal e também evidências de variação lexical na comparação entre essas regiões.

No PKE, encontramos na análise dos registros eclesiásticos de Vera Cruz/RS, dados das origens dos imigrantes, relatando inclusive o nome das localidades onde nasceram na Pomerânia. Os dados do *Corpus* de Registros Eclesiásticos confirmaram que entre os pomeranos havia silesianos, bávaros, renanos, etc, embora em menor quantidade em relação aos pomeranos que eram maioria, mas esse fato já é um indício do contato interétnico e interdialetal.

Outro fato que confirma essa hipótese é a presença de itens lexicais não pomeranos e provavelmente hunsriqueanos[75], presentes no PKO, pois duas irmãs entrevistadas eram filhas

[75]Para afirmar isso, pesquisamos alguns itens em um dicionário de *Hunsrückisch*, o Rheinisches Wörterbuch, disponível *online* e desenvolvido pela Universidade de Trier, na Alemanha.
Para o item *Schaft*, a referência é: <http://www.woerterbuchnetz.de/RhWB?lemma=schaffen_ii>. Para o item *Hinge*, a referência é: <http://www.woerterbuchnetz.de/LothWB?lemma=hinkel> e para o item *Geschpenz*, a referência é: <http://www.woerterbuchnetz.de/DWB?lemma=gespenst>. Acesso em: 02/06/2016.

de pai pomerano e mãe hunsriqueana e ao responderem ao LSF/QSL, apresentaram algumas respostas não-pomeranas, como organizamos no Quadro 7, a seguir:

Quadro 7: Respostas ao LSF/QSL que apresentaram interferência de outras variedades germânicas.

Respostas que apresentaram interferência de outras variedades germânicas				
Para o Português:	**Em pomerano seria:**	**Responderam:**	**Hunsrückisch**	**Alemão-padrão:**
Granizo ou chuva de pedra	Hoogel	Haagela	–	Hagel
Diarista	Dagelohner ou arbeira up Dag	schaft uf dag	schaft	Tagelöhner
Galinha	Huihna	Hinge	Hinkel/Hünkel	Huhn
Galinha sem rabo, cotó	Klitterhuihna	Schotterhingel	–	Ein Huhn ohne Schwanz
Fantasma ou Assombração	Spuicka, Spauck	Geschpenz	Gespenst	Geist
Feitiço ou macumba	Behexa	Das eingetan	–	Hexen, Zauber
Benzedeira	Besprecka	Braucher	–	Besprechen

Fonte: Elaboração própria.

E ainda, um terceiro fato que evidencia a confirmação desta hipótese é que tivemos que descartar uma entrevista realizada na zona rural, no leste de Minas, pois o sujeito-entrevistado falava hunsriqueano e não pomerano.

Por tudo isso, consideramos confirmada nossa hipótese de que houve contato de pomeranos com outras etnias alemãs em algumas regiões do Brasil, de modo que os textos dos *corpora* podem apresentar evidências desses contatos.

Observamos, ainda, que não foi possível encontrar nenhum texto ou fala pomerana, em nenhuma região, que tivesse o léxico totalmente diferenciado do alemão-padrão, visto que sempre havia algumas palavras que coincidiam sendo iguais tanto em pomerano quanto em alemão-padrão, como *Fluss*, *Blitz*, *Licht*, *Blind*, *Karre*, *Schnaps*, etc.

4) Quarta hipótese – Interferência do português e existência de uma Variedade Brasileira do Pomerano – Confirmada

O contato dos descendentes de pomeranos no Brasil com o português e o prestígio do português enquanto língua oficial do país pode ter permitido uma interferência português-

pomerano a ponto de que o pomerano falado atualmente no Brasil já represente uma variedade brasileira do pomerano, na qual o contato linguístico tenha propiciado usos com empréstimos lexicais do português no sistema pomerano: encontramos indícios na formação de frases e na inovação de itens lexicais que demonstram essa relação de "mistura" pomerano-português.

Essa hipótese se confirma ao considerarmos alguns exemplos extraídos do PK, pois nos textos e nas falas, ocorreram casos de "mistura" de pomerano-português ou pomerano-alemão-português, em um mesmo texto ou em uma mesma frase. Demonstramos, por meio dos Quadros 8 e 9 a seguir, alguns exemplos.

Quadro 8: Indícios do *Brasilianisch-Pommersch*, a variedade brasileira do pomerano no PK.

Indícios da Variedade Brasileira do Pomerano, o Brasilianisch-Pommersch no PK		
Forma encontrada	**Português**	**Alemão-padrão:**
Bolaspiela (3 ocorrências) Bolinhaspela (2 ocorrências)	Jogo bolinha-de-gude	Klicker
Klockacht (2 ocorrências) Kockacht (2 ocorrências)	Galinha D'Angola, cocá, cocár, angolista	Angola-huhn, Guinea-huhn
Pomdock (2 ocorrências)	Bodoque ou estilingue	Schleuder, Schlinge
Pomitta (6 ocorrências)	Palmito	Palmetto
Padariabrout, (3 ocorrências) Bäckariabrout, (1 ocorrência)	Pão-francês (Stuteln/stuuta)	Französisches Brot
Moskitt, Moskitte e Moskitta (15 ocorrências, itens lematizados)	Pernilongo	Stelzenläufer
Komella (6 ocorrências) Komellatee (3 ocorrências)	Flor de camomila	Kamille
Sarawä/Sarawee (5 ocorrências)	Sorigué, Saruê[76]	Stinktier

Fonte: Elaboração própria.

Além dos exemplos de substantivos mencionados no Quadro 8, percebemos também a presença de orações nas quais os sujeitos-entrevistados respondiam "misturando" português e pomerano, alguns exemplos podem ser visualizados no Quadro 9, a seguir:

[76] Gambá, de origem tupi. Segundo o dicionário de palavras brasileiras de origem indígena Saruê ou Sariguê é o gambá (CHIARADIA, 2008, p. 525-526). Segundo o dicionário do léxico tupi-português: "Sarig(u)ê é um nome indígena para o "gambá" ou "saruê", ou "sarig(u)eya" ou "sorig(u)é" (çoo = animal + (r)-ig(u)é = saco; entrada), alusão à bolsa onde se criam os filhos (DOMENICO, 2008, p. 895).

Quadro 9: Exemplos de frases do *Brasilianisch-Pommersch* extraído do PKO.

***Brasilianisch-Pommersch* – orações extraídas do PKO**			
Entrevista	**Sujeito-entrevistado**	**Pergunta**	**Resposta**
RHT-C-ZR-RS	<RHT-F-III>	142. Ein kleiner Raum mit einer Theke, wohin Männer in der Regel gehen um Schnaps zu trinken und wo man auch anderes kaufen kann?	„*Een bolische sägt ou wenda, pode ser, kann ouck weend säg*“.
RHT-C-ZR-RS	<RHT-F-III>	136. Wie nennt man das Objekt, das an den Wänden ist und dazu dient, die Lampe einzuschalten?	„*Ah eh dei, lichtshuolter oder lichtansticka nois semp chamemo*“.
RHT-C-ZR-RS	<RHT-F-III>	112. Person..., die Schwierigkeiten hat, Dinge zu lernen?	„*Kümm ni veel, dumm, ou como qui vo dize, tem que ser dumm, nee, pouco inteligente é dumm*“.
RB-VN-ZR-MG	<RB-M-III>	QS-5. (2) Por gentileza, fale um pouco sobre o seu local de nascimento, de infância... Você sempre morou aqui?	„[...] *nunca nóis aah, weerde treffte deet, häwa hier, wära ümmer ick mit mien pappa in da meehr wohna, in Cassadiera deet* [...]“.
HG-M-IT-MG	<HG-M-II>	90. Die Person, die mit der linken Hand isst, und alles mit dieser Hand tut?	„*Canhoto num sei não, dat seria ah mit der linker arm..., ah só assim, det schriewa, eu tenho um irmão assim nee, schriewa mit linker arm, hand*“.
HG-M-IT-MG	<HG-M-II>	80. Die Person ... die keine Zähne mehr hat?	„*Não mais é mãe in up düütsch... der hät keene täihna, dai? Num sei não..., dar hät keener tähna*[...]“.
HG-M-IT-MG	<HG-M-II>	70. Kleine Insekten mit langen Beinen, die in der Nacht um unsere Ohren fliegen.	„[...] *aí eu falo assim na roça, lá eu falo nee moskitte lass uns nichta, keiner schloofa, num deixa dormi, moskitte [...] schloop, schloofe* [...]“.

Fonte: Elaboração própria.

Inferimos que a interferência do português na fala pomerana já esteja em um nível no qual os falantes transitam livremente entre as formas na sua comunicação, produzindo sentidos, nomeando objetos, falando um pomerano-brasileiro, conforme já abordamos na seção

introdutória e no Capítulo 2 deste trabalho, acreditamos que exista um *Brasilianisch-Pommersch*, ou seja, uma variedade brasileira do pomerano (definida no tópico 2.3.1). Acreditamos que as transformações linguísticas ao longo do tempo permitiram um ambiente favorável ao surgimento de uma variedade com características próprias de escrita, pronúncias, usos e semas.

6.3. Análise do resultado das entrevistas interativas para coleta de dados orais

Gostaríamos de tecer algumas considerações sobre a experiência que tivemos com as entrevistas e o que elas nos levam a cogitar sobre o "Paradoxo do observador" (LABOV, 1972). Quando temos uma preocupação sobre a interferência do sujeito-entrevistador no momento da realização da entrevista e na possível interferência negativa da presença deste, no momento dos diálogos, mas acreditamos que há um exagero, comumente atribuído a esse fato.

Não há como se ausentar do ambiente, mesmo porque gravações livres do cotidiano dos falantes, deixando que se esqueçam do gravador instalado em suas residências, seria algo dificilmente aprovado pelo CEP atualmente. Então, não havendo como suprimir o pesquisador do ambiente de pesquisa e conversação, ele tende a fazer todo o possível para minimizar o constrangimento do momento da entrevista.

Percebemos em todas as 23 entrevistas realizadas (incluindo aqui as descartadas e as entrevistas-piloto), que esse constrangimento e algum "travamento", timidez e nervosismo que acontecia de ambas as partes (sujeito participante como entrevistado e sujeito participante como entrevistador), limitava-se aos instantes iniciais das entrevistas. A suposta "não-naturalidade", que na verdade é natural também, pois faz parte das relações humanas e da comunicação, logo era superada. Notamos que os sujeitos começavam a sentirem-se bastante à vontade, interagindo com as questões, fazendo piadas e brincadeiras, fazendo perguntas ao sujeito-entrevistador, narrando por decisão espontânea histórias de infância, de família, etc., estabelecendo diálogos espontâneos em pomerano com outras pessoas presentes no local das entrevistas e, também, percebíamos um prazer em falar sobre sua cultura e identidade, pois sentiam-se valorizados por serem convidados a participar de uma pesquisa.

No início das experiências, ficamos muito preocupados com os meandros que as entrevistas iam tomando ao sair dos roteiros previamente determinados pela sequência dos questionários e também pela recorrência de circunstantes, pessoas que se "intrometem" na entrevista. Porém, fomos percebendo que é justamente a interferência dos circunstantes e o suposto "paradoxo do observador" que permitiram o surgimento de um ambiente de fala

espontânea. Assim, a participação de circunstantes e o não seguimento dos questionários *stricto sensu* foram diretamente proporcionais à autenticidade dos dados, à originalidade e espontaneidade dos conteúdos dialogados durante as entrevistas. Ou seja, quanto maior a interação presente nas entrevistas, maior é a espontaneidade da fala.

Podemos citar alguns elementos que nos levam a pensar isso, como a rapidez da fala, em ritmo e prosódia extremamente naturais, próprios de falantes nativos; a diversidade e frequência de inserção de assuntos não perguntados; construções sintáticas mais livres do que numa fala premeditada e monitorada; retornos; gaguejos; pausas não terminais e retomadas; intercalação de assuntos; lembranças momentâneas que interferiam nos conteúdos das orações; etc.

Essas constatações, a partir de nossa experiência com a realização das entrevistas, nos permitem reforçar nossa ideia de que a entrevista é mais um momento de interação do que de interferência negativa para as metodologias de coletas de dados de fala espontânea, pois é um momento único, que pode começar truncado, mas no qual é possível contornar as dificuldades com algumas técnicas do pesquisador, a fim de "quebrar o gelo" e, desta forma, permitir o surgimento de um momento onde o diálogo interativo é estabelecido. Percebemos que dentro de alguns instantes o falante da variedade pesquisada passa a sentir-se à vontade para falar com naturalidade e consideramos que a entrevista interativa pelo método da SGL, mesmo com todas as variáveis que agem sobre o fenômeno da fala, permite um contato pessoal, próximo, face a face e, por isso, natural e espontâneo.

Portanto, concluímos que a interferência do observador e seus paradoxos são limitados pela força expressiva do falante, pelo prazer em expressar sua identidade e pelo seu instinto em preservar sua fala e língua materna, de forma que, as dificuldades têm uma duração limitada e podem ser superadas durante a entrevista, permitindo sim que os dados coletados sejam autênticos e provenientes de falas espontâneas. Se entrevistar é interferir, então temos que interferir para que muitas variedades linguísticas sejam estudadas, registradas, preservadas. Concluímos que as entrevistas são formas de interação com os falantes e podem ser naturais e fluir com muita espontaneidade.

6.4 Outros resultados e hipóteses direcionados pelos *corpora*: esboços

Nossa pesquisa, a princípio, não estava voltada para a análise de *corpora*, apenas para a sua compilação e as questões nela envolvidas. A partir daqui, esboçamos algumas análises preliminares a respeito de fatos que foram percebidos por meio dos resultados alcançados e que

consideramos importante não deixar de mencioná-los. Análises detalhadas podem ser melhor desenvolvidas em pesquisas futuras.

Embora inicialmente tenhamos formulado hipóteses que limitariam nosso estudo a uma abordagem *corpus-based*, além de responder às hipóteses, os dados nos direcionaram para a observação de outros fatos linguísticos, por isso consideramos nosso estudo como sendo também uma abordagem *corpus-driven*.

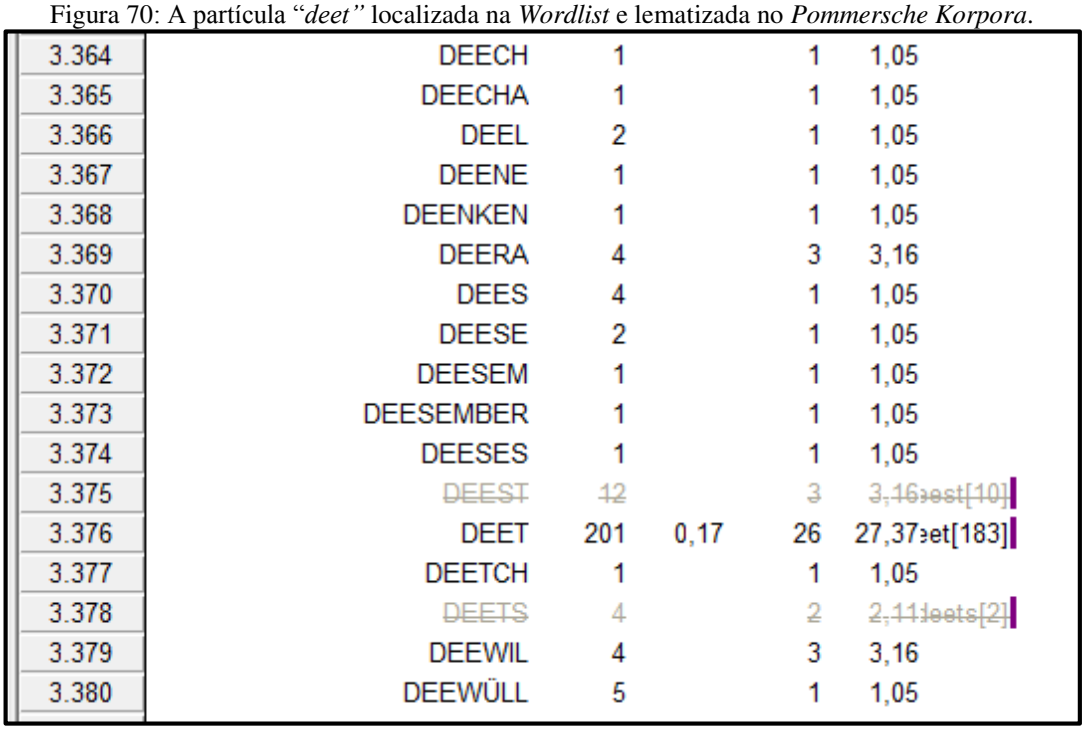
Ao realizarmos um experimento com uma pequena amostra extraída do PK, um estudo piloto do *corpus* de música pomeranas, percebemos alguns padrões e, mesmo no caso de alguns itens que não foram padronizados por se tratarem de variações, fizemos a contagem de suas ocorrências por meio da lematização, procedimento muito útil para a análise. Dessa forma, analisamos as recorrências de uma mesma lexia por meio do agrupamento das várias formas encontradas, além de checar os usos, por meio das linhas de concordâncias. Esse fato ocorreu nos casos de *deit, deyt, dêit, dêita, deet, deist e däit.*

Ao gerar listas de palavras e concordâncias e, ao estudá-las, chegamos a alguns resultados que não faziam parte de nossas hipóteses previamente formuladas, acerca do léxico pomerano evidenciado pelo PK, mas foram fatos linguísticos para os quais os *corpora* nos direcionaram. Apresentamos esses resultados a seguir:

a) Observamos que tanto na fala quanto na escrita pomerana, um dos itens mais frequentes é *deet*; no PK ele obteve 201 ocorrências. Convencionamos sua escrita como *deet*, forma encontrada na literatura. Porém, não foi possível converter todas as formas de escrita, pois houve casos de flexão do *deet*. Assim, consideramos mais produtivo para a análise fazer a lematização. Desse modo, foi possível observar melhor o comportamento do *deet* nas linhas de concordâncias que geramos a partir dele.

A Figura 70, a seguir, comprova a presença do *deet* no PK e demonstra sua lematização:

Figura 70: A partícula "*deet"* localizada na *Wordlist* e lematizada no *Pommersche Korpora*.

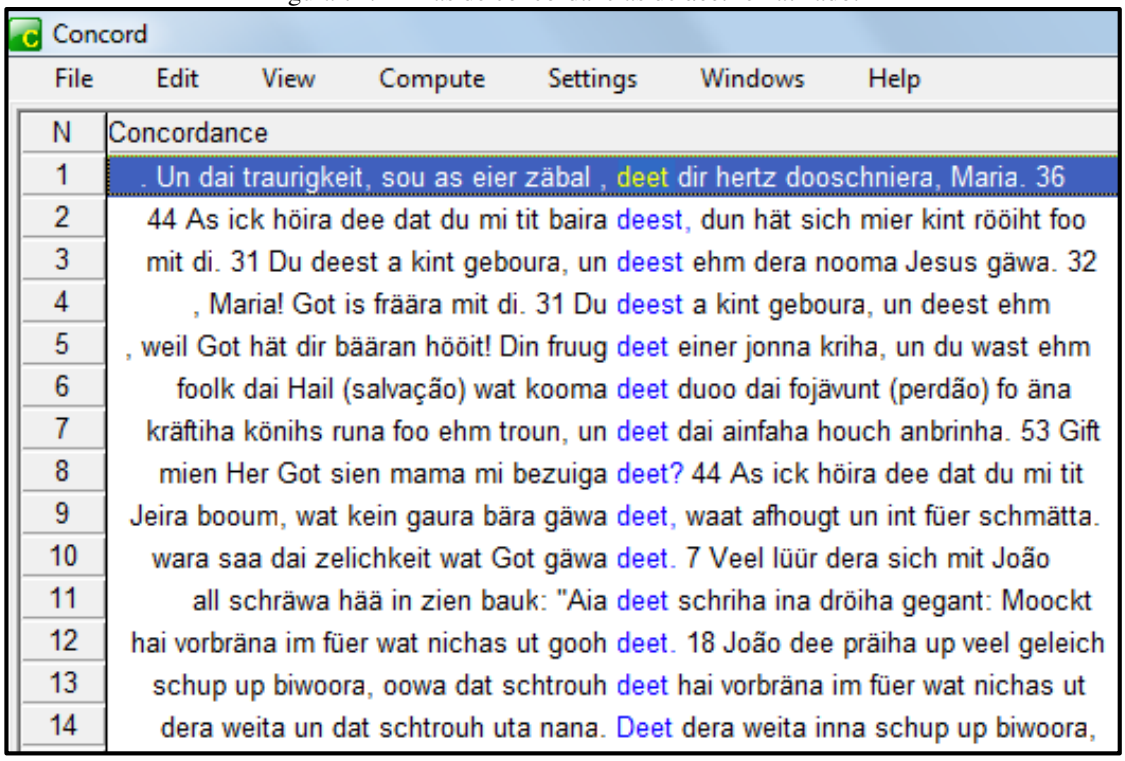


3.364	DEECH	1		1	1,05
3.365	DEECHA	1		1	1,05
3.366	DEEL	2		1	1,05
3.367	DEENE	1		1	1,05
3.368	DEENKEN	1		1	1,05
3.369	DEERA	4		3	3,16
3.370	DEES	4		1	1,05
3.371	DEESE	2		1	1,05
3.372	DEESEM	1		1	1,05
3.373	DEESEMBER	1		1	1,05
3.374	DEESES	1		1	1,05
3.375	~~DEEST~~	~~12~~		~~3~~	~~3,16ɜest[10]~~
3.376	DEET	201	0,17	26	27,37ɜet[183]
3.377	DEETCH	1		1	1,05
3.378	~~DEETS~~	~~4~~		~~2~~	~~2,11ɜets[2]~~
3.379	DEEWIL	4		3	3,16
3.380	DEEWÜLL	5		1	1,05

Fonte: elaboração própria.

A Figura 71 mostra as linhas de concordâncias de *deet* lematizado, ou seja, agrupada em todas as formas de escrita e de variações de flexão em que esse item ocorreu nos *corpora*.

Figura 71: Linhas de concordâncias de *deet* lematizado.



Fonte: elaboração própria.

Observamos, por meio das linhas de concordâncias e dos contextos, que o *deet* parece ser usado com frequência como partícula enfática, uma hipótese para a qual os *corpora* nos direcionaram, sendo nesse momento uma abordagem no estilo *corpus-driven*, pois fomos enxergando fatos sobre o pomerano que não havíamos imaginado antes, quando da elaboração do projeto, da formulação das hipóteses, e nem mesmo poderíamos prever o que encontraríamos ao nos deparar com a tela de leitura vertical do PK. Então os *corpora* direcionaram nossa atenção para outras evidências.

Conjecturamos que o *deet* assemelha-se ao comportamento do *to do* em Inglês e que é utilizado para fazer o passado simples. Inicialmente, pensamos que ele era usado sempre no final das orações, no entanto, ao observar os dados das linhas de concordâncias do PK, percebemos que essa não é uma posição fixa para ele, pois encontramos o *deet* em outras posições dentro das orações e, assim, verificamos que o *deet* não é utilizado de forma tão restrita como imaginávamos *a priori*.

Observamos que ele também é usado na segunda posição das orações, conforme linha 4 do *Concordance*, evidenciada na Figura 72 (linha 4), em "*Du deest a kint geboura*", "*você terá um filho*", onde embora o *deet* realmente funcione como partícula enfática, pois nesse caso não possui sentido sem o *geboura* (nascer), a sua flexão *deest* indica um tempo futuro.

Conforme observamos nas linhas de concordâncias, o *deet* é de fato muito utilizado como auxiliar para enfatizar uma ação ocorrida em um passado recente, como *gäwa deet* (deu), *moocka deet* (fez) e *kooma deet* (veio).

Outro exemplo de uso do *deet* em pomerano é *Dag waara deet*, expressão utilizada para falar que "amanheceu". Essa forma é característica da oralidade, pois se recuperarmos a expressão escrita teremos: *As dat dag waara deet*, que em uma tradução literal significa: "Assim o dia se fez fazer", mas com o sentido de informar que o dia "amanheceu". Nesse caso o *deet*, apenas enfatiza a ação já ocorrida (no tempo verbal passado) realizado pelo verbo *waara*.

b) Observamos também no PK que alguns traços que distinguem o pomerano do alemão-padrão são recorrentes nas mesmas formas e posições, fazendo parecer que o pomerano é uma variação sistemática em relação ao alemão-padrão, pois existem padrões recorrentes de variação (nas mesmas formas e posições), por exemplo, entre os fonemas do alemão-padrão /ei/ e do pomerano /ie/ (*mein* – HD e *mien* – BP), exceto em itens que são iguais em alemão e pomerano.

Outro padrão observado é o grafema /z/: em posição inicial, em pomerano tem o fonema de /t/ (*tied* – BP e *Zeit* – HD) e o grafema /t/ duplo tem fonema de /r/, assim como ocorre no

inglês (*better*, *berer* – em inglês). Deste modo, em *Hochtiedsbirer*, por exemplo, substituindo os grafemas, conforme os casos acima expostos, teremos *Hochzeitsbitter* para a forma em alemão-padrão do mesmo substantivo, que em PTB é "convidador de/para casamentos". Observamos inúmeros casos de repetição desses padrões nos *corpora*; são fatos linguísticos que carecem de um estudo específico e detalhado.

Há também casos em que só a pronúncia pomerana é que varia, e quando são grafados diferentes para tentar captar essa variação de som é que fazem com que aparentem ser mais distantes do alemão do que realmente são. Explicamos, a seguir:

- Toda vez que aparece o grafema /g/, quando está posicionado no início das palavras, ele pode ser identificado como aquele que se comporta como o fonema do /j/ no alemão-padrão. Ou seja, o /g/ inicial em pomerano, teria o fonema de /i/ no português brasileiro. Exemplos: *Geschichte*, *Geld*, *Gelb*, etc. que em pomerano tem som de /i/ no início das palavras.
- Em palavras como *Hoogel, Noogel, Voogel* – o grafema /g/ sempre se comporta como o fonema /ch/ gutural do alemão-padrão, quando estão posicionados no meio da palavra.
- Nas posições finais como em *lustig*, por exemplo, ocorrem variações na pronúncia de /g/, com o som de /k/, ou de /ch/ ou de /sch/. Essas foram as três variações sonoras que identificamos para /g/ nas entrevistas, porém ainda não presenciamos um /g/ nas posições finais com o fonema do /g/ como em PTB.

Nesses casos exemplificados acima, seria o caso de uma escrita diferente para o som ou de uma interpretação sonora diferente para um mesmo sinal gráfico? Um estudo mais aprofundado precisaria ser feito para analisar tais questões, pois foi um fato observado por meio do PKO, porém não é o foco de estudo desta pesquisa.

Acreditamos, porém, que em alguns casos, não é a escrita que deva ser diferente, porque já existe um padrão de pronúncia convencional, ligando um determinado som para um sinal gráfico convencionado e que, comumente nas línguas, variedades, variações e diatopias as interpretações sonoras das grafias também variam. Nossa intenção não é discutir fonética e fonologia, mas sim, levar à reflexão sobre alguns fatos linguísticos que encontramos nos nossos *corpora*, tendo em vista que todos os aspectos da língua são importantes e inseparáveis.

Embora estejam fora de nosso escopo, as percepções registradas acima não poderiam deixar de serem ressaltadas, pois vêm corroborar nossa ideia de que há uma diversidade de estudos que podem ser feitos por meio da abordagem-metodologia da LC. A existência de tais fatos linguísticos sobressalta ao estudioso de *corpus* quando ele realiza uma leitura vertical das linhas de concordâncias em uma abordagem direcionada pelo *corpus*.

6.5 Análise geral dos resultados

Ao compilar os dados para constituir os *corpora* pomeranos, pudemos perceber que é possível desenvolver um processo de aprendizagem por meio da compilação, bem como perceber algumas características linguísticas, durante a transcrição e a organização dos dados.

Realizamos algumas análises ao carregar o PK no WST, e também ao carregar separadamente o PKE e o PKO no programa. De modo que os dados do PK podem ser analisados, tanto em conjunto, quanto podem ser analisados em seus *subcorpora*. Assim, podemos avaliar tanto as características do pomerano nos dados escritos em seus diferentes gêneros textuais, como também as características da fala pomerana presente no *corpus* oral.

Quanto à adequação dos *corpora*, assumimos que o PK pode não ser necessariamente adequado à investigação de qualquer característica linguística (BERBER SARDINHA, 2004), pois consideramos palavras soltas para a compilação, mas o PK é adequado aos objetivos geral e específicos da investigação que delimitamos na nossa proposta. De qualquer forma, o *Pommersche Korpora* é uma fonte de informação linguística sobre o pomerano, pois nosso trabalho ambiciona contribuir para a descrição do léxico pomerano, bem como contribuir para outras pesquisas, para produção de conhecimento sobre a LC e sobre o pomerano.

Avaliamos que os resultados do nosso trabalho, em geral, são satisfatórios, pois conseguimos atingir os nossos objetivos (geral e específicos) e responder às nossas hipóteses, compilamos conteúdo lexical em pomerano, para constituição de um conjunto de *corpora* escritos, realizamos a coleta de dados orais e transcrevemos um *corpus* oral dialetal. Ademais, avançamos positivamente o nosso escopo, pois esboçamos algumas análises preliminares sobre o pomerano e percebemos o surgimento de novas hipóteses a partir das evidências.

No próximo capítulo, vamos tecer nossos comentários finais, avaliar a obra realizada como um todo, fazer breves considerações acerca de possíveis desdobramentos deste trabalho e lançar algumas possibilidades para estudos futuros sobre o pomerano, momento no qual encerraremos o nosso texto.

Un wenn ick wüst dat dai Wirlt mooin uunergüng, den wür ick hüüt nog een appelboum plante[77].

Und wenn ich wüßte, daß morgen die Welt unterginge, würde ich noch heute ein Apfelbäumchen pflanzen.
Martin Luther.

[77] Versão pomerana da célebre frase do teólogo alemão Martinho Lutero. Traduzida com base no método de consulta ao PK e como uma proposta de variação da tradução de WILLE, 2011. Em português: Se eu soubesse que o mundo acabaria amanhã, ainda hoje eu plantaria uma macieira.

7. À GUISA DE CONSIDERAÇÕES FINAIS: DESDOBRAMENTOS FUTUROS E POSSIBILIDADES PARA O *POMMERSCHE KORPORA*

Consideramos que este trabalho permite ampliar a visão a respeito do pomerano, pois nele apresentamos a Linguística de *Corpus* para os pomeranos e o pomerano para os linguistas de *corpus,* oferecemos uma opção para a abordagem do pomerano e para o estudo do pomerano com base em evidências coletadas a partir do uso autêntico. Abrangemos um recorte cronológico histórico e contemporâneo, consideramos tanto documentos históricos quanto a fala espontânea, levantamos dados cartográficos e censitários (no leste de Minas) sobre os pomeranos, trabalhamos algumas lacunas no que se refere aos estudos sobre o tema, apresentamos fontes sobre a definição do pomerano, enfretamos a discussão sobre sua classificação e nos posicionamos a respeito, tratamos do pomerano na Europa, resumimos algumas informações sobre a imigração pomerana; apresentamos algumas motivações que contribuíram para os pomeranos emigrarem e para o Brasil receber esses imigrantes. Também apresentamos algumas pesquisas sobre o pomerano realizadas no país.

Em seguida, construímos nossa fundamentação teórica, tratando pontualmente de alguns conceitos básicos como Lexicologia e léxico, Linguística de *Corpus* e Sociogeolinguística, trouxemos contribuições de autores da LC alemã e elaboramos nossa definição de LC, descrevemos detalhadamente nossa metodologia; quando falamos sobre os processos de compilação e construção dos *corpora*, os problemas encontrados e as soluções adotadas, as fases e instrumentos de coleta, desenvolvemos procedimentos de tratamentos de dados como conversão e convenção da escrita e um método de consulta à bíblia pomerana de 1588. Ampliamos o leque de fontes para a coleta de dados linguísticos e fragmentos dialetais e fizemos transcrições de fontes primárias (inclusive em letra gótica).

Ao final, expomos o conjunto dos *corpora* constituídos, sua codificação e arquitetura, demonstramos o potencial dos *corpora*, apresentamos dados estatísticos, descrevemos detalhadamente cada um dos *corpora* compilados e fizemos análises superficiais de cada um. Classificamos os *corpora* compilados, avaliamos os resultados e comprovamos como os dados reunidos e observados permitem falar em variedade brasileira do pomerano – o que nomeamos como *Brasilianisch-Pommersch* (BP). Também realizamos breves análises gerais, embora o foco fosse a compilação; pudemos testar nossas hipóteses; responder aos objetivos específicos e lançar um olhar sobre outras questões direcionadas pelos *corpora*, dada a riqueza lexical contida nos dados.

Em suma, organizamos diversos *corpora,* considerados em sua totalidade de tamanho pequeno-médio ao reunir 14 *corpora* escritos provientes de diversos gêneros e/ou suportes, e compilamos um *corpus* oral, constituindo um conjunto classificado como: *corpora* dialetais multilíngues contatuais, acervo que compõe a unidade que denominamos *Pommersche Korpora* – PK.

Portanto, podemos afirmar que obtivemos êxito ao cumprir nossa proposta, pois segundo o levantamento estatístico do PK, coletamos 129.666 palavras corridas e 20.672 palavras distintas, de modo que pudemos alcançar uma considerável quantidade de amostras contendo o léxico pomerano.

Enfim, ao retomar todos os feitos realizados, consideramos que conseguimos cumprir todas as nossas propostas e metas, e acreditamos que este trabalho iniciou a descrição lexical por métodos de coleta objetiva de dados empíricos (fontes primárias) que desejamos realizar, ou ver realizada, sobre o pomerano preservado no Brasil.

Avaliamos que a LC é uma abordagem-metodologia atual e eficiente que confere estatuto científico ao trabalho, principalmente por ter caráter descritivo e levantar dados empiricamente comprováveis. A análise dos textos nessa perspectiva é importante e proveitosa e tem contribuído fundamentalmente não só para a criação de um acervo do léxico pomerano, mas também para a própria convenção de um padrão de escrita, baseado nas frequências de uso e recorrência das formas encontradas na realidade linguística.

Assim, a LC se configura também como uma forma de abordagem da língua que permite sua descrição e registro e, por isso, o título do nosso trabalho incluiu a expressão de ser uma proposta metodológica para compilação de *corpora* dialetais, visto que os princípios da LC nos permitiram não só estudar o pomerano e abordar os fenômenos compilados, mas também desenvolver procedimentos metodológicos para possibilitar a constituição do conjunto de nossos *corpora*, o PK.

A LC amplia o nosso olhar sobre fatos e evidências, abrindo inúmeras possibilidades metodológicas e analíticas para os estudos linguísticos. Por isso, vamos lançar luz sobre as possibilidades que este trabalho abre para desdobramentos futuros, conforme tratamos a seguir.

7.1 Possibilidades futuras para os *Pommersche Korpora*

Neste momento, propomo-nos a refletir sobre as possibilidades que nosso trabalho tem de se desdobrar, pois a nossa produção poderá oferecer alguma contribuição para a manutenção da variedade brasileira do pomerano, na medida em que constitui, mesmo que timidamente, um

registro do pomerano e uma descrição lexical do mesmo, com base em *corpora* e poderá, também, fornecer material para o desenvolvimento de outros estudos sobre o pomerano.

Um grande ganho deste trabalho, a nosso ver, é a inspiração para explorar inúmeras possibilidades que o PK tem, como comparar o pomerano com o *COSMAS* (*Korpus* do alemão HD, com 1,7 bilhões de palavras), com os dados dos atlas alemães (*Deutsche Sprachatlas*) digitalizados pelo projeto DiWA (*Digitaler Wenker Atlas*), com os exemplos do *Audio-Katalog* do REDE; explorar o *Brasilianisch-Pommersch* e os neologismos apresentados pelos *corpora*; produzir materiais didáticos e propor um ensino do pomerano na perspectiva do DDL (*Data Driven Learning*, JOHNS, 1991), que consiste no ensino com acesso a dados linguísticos extraídos de linguagem autêntica, o qual permite realizar inferências e levantar hipóteses, descobrir características sobre o funcionamento linguístico, fazer generalizações, observar os usos espontâneos e os padrões recorrentes, etc.

Todas as nossas sugestões exigiriam o trabalho de uma vida inteira dedicada à pesquisa linguística, por isso, listamos algumas possibilidades de desdobramentos, a partir dos dados obtidos neste trabalho, que podem ser efetuadas por outros pesquisadores, público-alvo de nossa pesquisa, por pomeranos e/ou por quem mais se interessar por tal objeto de estudo:

1. Expansão dos *corpora* do PK e/ou transformar o PK em monitor;
2. Etiquetagem completa dos *corpora;*
3. Revisão e validação dos *corpora*;
4. Estudo lexical dos dados do PK, por exemplo: estudo dos verbos e substantivos mais frequentes (para produzir um glossário, comparar com outras línguas), etc.;
5. Transformação do PK em banco de dados com SGBD para que haja um gerenciamento do acervo e permita uma interface de consulta;
6. Análise dos resultados do ponto de vista do contato de línguas, bilinguismo, diglossia, etc.;
7. Produção de um atlas linguístico semântico-lexical parcial do pomerano no Brasil, com base nas respostas obtidas, nesta pesquisa, por meio da aplicação do LSF/QSL;
8. Comparação com outros *corpora*, por exemplo, o *Corpus of Historical Low German* (*Corpus* do baixo-alemão histórico) e o *Referenzkorpus Mittelniederdeutsch* (*Corpus* de referência do baixo-alemão médio);
9. Utilização dos dados como ferramenta auxiliar na tradução pomerano-português, pomerano-alemão-padrão;
10. Levantamento de hipóteses sobre a sintaxe pomerana;

11. Produção de materiais didáticos, utilizando exemplos do uso real e cotidiano do pomerano, evitando a formulação de frases artificias para exemplificações;
12. Produção de artigos sobre o pomerano e divulgação em site de conteúdos, desde que não firam direitos autorais;
13. Estudos de perdas de alguns fonemas e grafemas nas posições finais das palavras (padrão que observamos nos *corpora*);
14. Estudo da variação regional e dos processos de interferências;
15. Auxílio na construção de gramáticas pomeranas;
16. Auxílio na produção de obras lexicográficas;
17. Produção de aplicativos com base no conteúdo dos *corpora* (por exemplo, no estilo dos aplicativos *Bravolol*®, com áreas semânticas, aúdios, exemplos de uso, tradução para o português, etc.);
18. Contribuição para a compilação de outros *corpora* dialetais por meio do exemplo de metodologia;
19. Compilação de um *corpus* de referência do *Pommersches Plattdeutsch* (baixo-alemão pomerano) por meio de portais da Alemanha (disponíveis *online*) para gerar parâmetros de análises;
20. Verificar a possibilidade de gerar palavras-chave do pomerano, com base em algum dos *corpora* do baixo-alemão existentes (alguns deles já mencionados neste trabalho);
21. Análise Sociolinguística com base nas respostas do QS coletadas por este trabalho por meio das 19 entrevistas.

Dentre as possibilidades que vislumbramos para o estudo do pomerano, a partir da abordagem-metodologia da LC e com base nos dados do PK, podemos citar estudos lexicológicos, lexicográficos, morfológicos, sociogeolinguísticos, etc.

A LC foi o grande norte do nosso trabalho e a responsável pelo nosso produto, pois acreditamos que, de outro modo, não teríamos a visão do pomerano e nem a abrangência que pudemos alcançar com a utilização dessa metodologia-abordagem que funcionou como um "microscópio lexical". Os instrumentais da LC (estatísticas, linhas de concordâncias, leitura vertical, colocados, etc) permitem realizar micro e macro-análises, pois além de vermos minúcias em grande escala, podemos enxergar mais longe, como um telescópio, devido a grande quantidade de amostras que reúne e das quais nos aproxima. Em palavras pomeranas, poderíamos reiterar que *"Dai Teleskopa häwa ous joo veel weesa, wat man früüher ni vörher hät"* (Os telescópios nos fizeram ver coisas que antes não percebíamos), metáfora presente

desde o início do nosso texto, pois traduz um dos grandes atributos da LC: ampliar nossa percepção sobre os fatos linguísticos.

Uma conclusão direta a que chegamos é a de que não é possível encontrar dados pomeranos, no nosso contexto de pesquisa, sem alguma interferência do alemão-padrão e que os *corpora* compilados apresentam indícios de contatos linguísticos, além disso, as fontes primárias, enquanto evidências autênticas, recolocam uma série de dados enquanto escritas pomeranas, que haviam sido até então desconsideradas por não diferir radicalmente da tradição ortográfica alemã. Concluímos que os dados são multilíngues e contatuais, pois apresentam, em alguns casos, manifestações de até quatro variedades (pomerano, alemão-padrão, hunsriqueano e português) de uma só vez.

Consideramos, sobre o nosso trabalho como um todo, que a compilação de *corpus*/*corpora* documenta fatos linguísticos e constitui um acervo de registros; fontes que contêm ricas evidências com as quais devemos nos relacionar de maneira ética e com abordagens descritivas, para explorar os fatos oferecidos pelos dados e produzir conhecimento a partir deles, bem como formular hipóteses sobre o funcionamento linguístico.

Além de descobrir muitas direções para onde prosseguir com os estudos linguísticos, há um grande potencial para produção de materiais didáticos, com base no uso autêntico do pomerano e assim, fomentar sua revitalização e manutenção.

Escolhemos finalizar nosso texto com alguns valores autenticamente pomeranos, que foram elementos motivadores para que conseguíssemos obter êxito nesta pesquisa, tais valores são, segundo Granzow (2009, p. 136), virtudes pomeranas: “resistência, perseverança, assiduidade e obstinação”.

REFERÊNCIAS BIBLIOGRÁFICAS

AIJMER, K.; ALTENBERG, B. **English Corpus Linguistics**: Studies in Honour of Jan Svartvik. London: Longman, 1991, 250-258.

ALMEIDA, V. C. **Investigando colocações em um Corpus de aprendiz**. 2014. 164 f. Tese (Dourado em Linguística Aplicada) – Universidade Federal de Minas Gerais. Belo Horizonte, 2014. Disponível em: <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/MGSS-9PHMJ6> Acesso em: 14 mar. 2015.

ALTENHOFEN, C. V. Migrações e contatos linguísticos na perspectiva da geolinguística pluridimensional e contatual. **Revista de Letras Norte@mentos**, v. 6, n. 12, 2013. Disponível em: <http://sinop.unemat.br/projetos/revista/index.php/norteamentos/article/view/1216>. Acesso em: 9 mai. 2016.

ALVES, C. F. A contribuição da análise contrastiva e de um corpus de erros para o desenho de um dicionário de espanhol para aprendizes brasileiros. **Cadernos de Letras**. Rio de Janeiro, n. 28, jul. 2011. Disponível em: <http://www.letras.ufrj.br/anglo_germanicas/cadernos/numeros/072011/textos/cl31072011alves.pdf >. Acesso em: 06 abr. 2016.

ALVES, D. B. Cartas de imigrantes como fonte para o historiador: Rio de Janeiro – Turíngia (1852-1853). Instituto de História Ibérica e Latinoamericana. Universidade de Colônia. **Revista Brasileira de História.** São Paulo, v. 23, n. 45, p. 155-184, jul. 2003.

_____. **Das Brasilienbild der deutschen Auswanderungswerbung im 19. Jahrhundert**. Berlin: WVB, 2000.

ANDRADE, M. M. de. Lexicologia, terminologia**:** definições, finalidades, conceitos operacionais. In: ISQUERDO, A. N.; OLIVEIRA, M. P. P. (Org.). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2. ed. Campo Grande, MS: Ed. UFMS, 2001. p. 191-200.

ANTUNES, D. **Descendentes de etnia germânica vivem isolados em área rural de Minas**. Reportagem online para o Hoje em dia. Itueta, 2011. Disponível em: <http://www.nanademinas.com.br/exibe-cultura.php?id=928>. Acesso em: 07 fev. 2013.

ARAUJO, L. S. A elaboração de um corpus dialetal da língua espanhola falada na Argentina: contribuições dos gêneros discursivos e da análise textual automática. **Revista Estudos Linguísticos**, São Paulo, v. 41, n. 1, p. 246-261, jan./abr 2012.

ARCGIS. Versão 10.2, Environmental Systems Research Institute, 2013 – CD.

BAHIA, J. **"O tiro da bruxa"**: identidade, magia e religião entre camponeses pomeranos do Estado do Espírito Santo. Tese (doutorado em Antropologia Social) – Universidade Federal do Rio de Janeiro. Rio de Janeiro: Museu Nacional, 2000.

BARBOSA, M. A. Lexicologia, lexicografia, terminologia, terminografia: objeto, métodos, campos de atuação e de cooperação. In: SEMINÁRIO DO GEL, 39., 1991, Franca. **Anais** ...

Franca: UNIFRAN, p. 182-189, 1991.

BEAU, A. E. **Langenscheidts Taschenwörterbuch Deutsch-Portugiesisch**. Berlin: Langenscheidts, 1983. (tomo segundo).

BEILKE, N. S. V. Ach Jå! Fraseologismos em pomerano e em alemão. **Domínios de Lingu@gem**, Uberlândia, v. 8, n. 2, p. 178-201, 2014. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem>. Acesso em: 16 jan. 2015.

_____. Ciência da fala e arbitrariedade perceptiva dos sons. In: NOVODVORSKI, A . (Org.). Ensaios em Teorias Linguísticas. Uberlândia: EDUFU, 2016. **E-book**. [No prelo].

_____. Do nativo ao pomerano: as línguas, os dialetos e falares vivos de um Brasil pouco conhecido. **Revista Domínios de Linguagem**, Uberlândia, v. 7, n. 1. p. 264-283. jan./jun. 2013. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem>. Acesso em 16 jan. 2014.

_____. Pomerano: uma variedade germânica em Minas Gerais. In: SIMPÓSIO INTERNACIONAL DE LETRAS E LINGUÍSTICA. v. 3, n.1., 4., 2013, Uberlândia. **Anais**...Uberlândia: EDUFU, 2013. Disponível em: <http://www.ileel2.ufu.br/anaisdosilel>. Acesso em: 24 jun. 2014.

BENINCÁ, L. R. **Dificuldade no Domínio de Fonemas do Português por Crianças Bilíngues de Português e Pomerano**. 2008. 227 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal do Espírito Santo, Vitória, 2008.

BERBER SARDINHA, T. B. **Linguística de *Corpus***. Barueri, SP: Manole, 2004.

_____. Linguística de Corpus: histórico e problemática (Corpus Linguistics: History and Problematization): retrospectiva. **Delta**, São Paulo, v. 16, n. 2, p. 323-367, 2000. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000100005&lng=en&nrm=iso>. Acesso em: 16 abr. 2016.

_____. Linguística forense. In: BERBER SARDINHA, T. **Pesquisa em lingüística de Corpus com *WordSmith* Tools**. Campinas, SP: Mercado das Letras, 2009. p. 69-81.

_____. **Pesquisa em linguística de Corpus com WordSmith Tools**. Campinas, SP: Mercado das Letras, 2009.

BERRUTO, G. Le varieta del repertorio. In: SOBRERO, A. A. (Ed.). **Introduzione all'italiano contemporaneo**: La variazione e gli usi. Roma: Laterza, 1993. p. 3-36.

BEVILACQUA, C. Por que e para que a Linguística de Corpus na Terminologia? In: TAGNIN, S. E. O.; BEVILACQUA, C. (Org.). **Corpora na Terminologia**. São Paulo: Hub Editorial, 2013. p. 11-27.

BIDERMAN, M.T.C. **Dicionário contemporâneo de português**. Petrópolis: Vozes, 1992.

_____. **Teoria linguística**. 2. ed. São Paulo: Martins Fontes, 2001.

_____. As ciências do léxico. In: ISQUERDO, A. N.; OLIVEIRA, M. P. P. (Org). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: Ed. UFMS, 2001. v. 2. p. 13-22.

BOST, B. **Pommersche Sprache erlebt Renaissance in Brasilien. Preußischen: Allgemeinen Zeitung**, 2010. Disponível em <http://www.vda-globus.de/fileadmin/bilder/Kopie_von_Pommersche__Sprache_erlebt_Renaissance_in_Brasilien.pdf> . Acesso em: 09 mar. 2015.

BOSSMANN, R. Zur deutschbrasilianischen, Mischsprache. In: **Letras** I, Curitiba: s.e., 1953.

BÖTTCHER, A. **Pommern Geschichte**, 2005. Disponível em: <http://www.der-familienstammbaum.de/pommern/pommern-geschichte.php>. Acesso em: 02 jan. 2015.

BRUM-DE-PAULA, M. R.; ESPINAR, G. S. Coleta, trancrição e análise de produções orais. In: BRUM-DE-PAULA, M.R.; SCHERER, A.E.; PARAENSE, S.C.L. (Org.). **Letras**, n. 21. Santa Maria: PPGL Editores, 2002.

BUENO, F. da S**. Grande dicionário etimológico-prosódico da língua portuguesa: vocábulos, expressões da língua geral e cientifica-sinônimos contribuições do tupi-guarani**. Vol 8 (T-Z), Brasília, 1974.

CARTER, R. Introduction. 2003. In: SINCLAIR, J. McH. **Trust the Text**: Language, corpus and discourse, London: Routledge, 2004. p. 1-6.

CHIARADIA, C. **Dicionário de palavras brasileiras de origem indígena**. Limiar, 2008.

CORPUS DE REFERÊNCIA DO PORTUGUÊS CONTEMPORÂNEO – CRPC. Centro de Linguística da Universidade de Lisboa. Disponível em: <www.clul.ul.pt/pt/recursos/183-reference-corpus-of-contemporary-portuguese-crpc>. Acesso em: 20 abr. 2016.

CRISTIANINI, A. C. **Atlas semântico-lexical da região do Grande ABC:** um estudo geolinguístico. 2007. 635 f. Tese (Doutorado em Linguística) – Universidade de São Paulo, São Paulo, 2007. Disponível em: <http://www.fflch.usp.br/dlcv/lport/pdf/slp22/02.pdf>. Acesso em: 16 fev. 2015.

CRISTIANINI, A. C. Sociogeolinguística: uma abordagem para o estudo do léxico. In: SANTOS, I. P.; CRISTIANINI, A.C. (Org.). **Sociogeolinguística em questão:** reflexões e análises. São Paulo: Paulistana, 2012. p. 21-32.

CRYSTAL, D. **The Cambridge Encyclopedia of Language**. Cambridge: Cambridge University, 1997.

CUNHA, G. P. A simbologia mortuária pomerana. **Revista Ciências da Religião**. São Paulo, v. 9. n. 2. p. 72-99, dez. 2011. Disponível em:
< http://editorarevistas.mackenzie.br/index.php/cr/article/view/3148>. Acesso em: 02 mar. 2015.

DA CUNHA, A. G. **Dicionário etimológico da língua portuguesa**. Rio de Janeiro, Lexikon, 2007.

Dai Pomerer: dai gang fon ainem folk. Documentário. Vanildo Kruger. Santa Maria de Jetibá, KG Produções e Rede Click: 2009. 1 DVD.

DIETRICH, A. M. **Imigração suíça e protestantismo no século XIX**: reflexões sobre a micro-história de Guilherme Gaensly: imagens de São Paulo: Gaensly no acervo da Light (1899-1925). São Paulo: Light, 2001.

DRÄGER, E. **Dörpwihnachten,** 1878. Disponível em: <www.pommerscher-greif.de>. Acesso em: 20 dez. 2014.

DUMMER, C. et al. **Vera Cruz:** tempo, terra e gente. Vera Cruz: LupaGraf, 2009.

DÜCK, E. S. **Witmarsum uma comunidade trilíngue**: Plautdietsch, Hochdeutsch e português. 2005. Dissertação (mestrado em Linguística) – Universidade Federal do Paraná, Curitiba, 2005.

_____. **O trilinguismo no Colégio Fritz Kliewer de Witmarsum (Paraná).** CELSUL, Anais... Porto Alegre, 2008. Disponível em: < http://www.celsul.org.br/Encontros/08/trilinguismo_col_fritz_kliewer.pdf>. Acesso em 04. Fev. 2015.

ENCARNAÇÃO, M. R. T. **Estudo geolinguístico de aspectos semântico-lexicais nas comunidades tradicionais do município de Ilhabela**. 2005. 167f. Dissertação (Mestrado em Linguística) - Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2005.

FEITOSA, M. P. **Uma proposta de anotação de corpora paralelos com base na Lingüística Sistêmico-Funcional**. 2005. 177 f. Dissertação (Mestrado em Linguística Aplicada) – Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

FERNÁNDEZ PÉREZ, M. El Corpus Koiné de habla infantil. Líneas Maestras. In: FERNÁNDEZ PÉREZ, M. (Coord.). **Linguística de corpus y adquisición de la lengua**. Madrid: Arco Libros, 2011. p. 11-36.

FERREGUETTI, K.; PAGANO, A.; FIGUEREDO, G. Significados existenciais no português brasileiro: um estudo contrastivo em textos traduzidos e não traduzidos. ELC – Encontro de Linguística de Corpus. 10., **Anais**... UFMG, Belo Horizonte, 2012, p. 280-293.

FERREIRA, C.; CARDOSO, S. **A dialetologia no Brasil**. São Paulo: Contexto, 1994.

FILLMORE, C. J. Corpus linguistics or computer-aided armchair linguistics. In: PROCEEDINGS OF NOBEL SYMPOSIUM. 1992. New York. **Directions in corpus linguistics**. New York: [s.n.], 1992. p. 35-60.

FROMM, G. Resenha de *Corpora* no Ensino de Línguas Estrangeiras. VIANA, V.; TAGNIN, S. E. O. (Org.). Corpora no ensino de línguas estrangeiras. São Paulo: HUB Editorial, 2010. p. 357-361. Resenha de livro. **Domínios de Lingu@gem**, v. 4, n. 2, 2° sem., Uberlândia, 2010, p. 196-199. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem>. Acesso em: 17 mai. 2016.

FROMM, G. O Uso de Corpora na análise linguística. **Revista Factus**, São Paulo, v. 1, n. 1, p. 69-76, 2003.

GLÜSING, J. Pommern im Urwald: Wie ein deutscher Dialekt in Brasilien überlebte. **Der Spiegel**. Berlin, Global Village. p. 92, 2013. Disponível em: <spiegel.de/app392013brasilien> . Acesso em: 15 jan. 2014.

GONÇALVES, L. B. Linguística de corpus e análise literária: o que revelam as palavras-chave. In: TAGNIN, S. E. O.; VALE, O. A. (Org.). **Avanços da linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p.387-405.

GONZALEZ, Z. M. G. **Linguística de Corpus na análise do Internetês**. 2007. 123 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – Pontifícia Universidade Católica, São Paulo, 2007.

GRANZOW, K. **Pomeranos**: sob o cruzeiro do sul, colonos alemães no Brasil. Trad. Selma Braum. Edição Comemorativa dos 150 anos da Imigração Pomerana no Espírito Santo. Vitória: Arquivo Público do Estado do Espírito Santo, 2009. v. 10. (Coleção Canaã).

GREIFSWALD. Ernst-Moritz-Arndt-Universität. [Bibliothek von Greifswald Universität]. Disponível em: <http://www.uni-greifswald.de/bibliothek.html>. Acesso em: 10/05/2016.

GUMPERZ, J. J. Convenções de contextualização. In: B. T. Ribeiro, e O. M. Garcez, (Orgs.). **Sociolinguística interacional:** antropologia, linguística e sociologia em análise do discurso. Porto Alegre: AGE Ed., 1998, p. 98-119.

HEINEMANN, J. **Carta**. Cód. 0104/ A2013 Re. Endereçada à Neubiana Silva Veloso Beilke, 2013.

HERRMANN-WINTER, R. ; VOLLMER, M. **Das Pommersche Wörterbuch:** Vorläufer – Ansätze –Konzepte. Tausend Jahre pommersche Geschichte. Köln: [s.n.], 1999, S. 281-306.

HERRMANN-WINTER, R. **Plattdeutsch-hochdeutsches Wörterbuch für den mecklenburgisch-vorpommerschen Sprachraum**. 5. Veränd. Aufl. Rostock: Hinstorff, 2003.

_____. **Sprachatlas für Rügen und die vorpommersche Küste**. Rostock: Greifswald Universität, 2013.

_____. Zur Geschichte der Dialektgeographie in Pommern. In: ASMUS, I. et al. (Hg.): **Geographische und historische Beiträge zur Landeskunde Pommerns**. Eginhard Wegner zum 80. Geburtstag. Greifswald: [s.n.]. 1998, S. 299-304.

HEYE, J. Sociolinguística. In: PAIS, C. T. et al. (Org.). **Manual de linguística**. 2. ed. Petrópolis: Vozes, [1978]. p. 203-237.

HÖHMANN, B; SAVEDRA, M. M. G. Das Pommerische in Espírito Santo. **Pandaemonium Germanicum**, São Paulo, n. 18, Dez. 2011, p. 283-300. Disponível em:< www.fflch.usp.br/dlm/alemao/pandaemoniumgermanicum>. Acesso em: 15 nov. 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISÍTCA – IBGE. Censo Demográfico de 2010. (29 de novembro de 2010). Consulta a lei de acesso à informação em 14 jun. 2013 sobre a população em Vila Neitzel/MG. Disponível em: <http://www.ibge.gov.br> . Acesso em: 15 jun. 2013.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISÍTCA – IBGE. Censo Demográfico de 2010. (29 de novembro de 2010). Consulta a lei de acesso à informação em 11 mai. 2016 sobre a densidade população de várias cidades de Minas Gerais e Rio Grande do Sul. Disponível em: <http://www.ibge.gov.br> . Acesso em: 13 mai. 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISÍTCA – IBGE. Censo Demográfico de 2010. (29 de novembro de 2010). Consulta a lei de acesso à informação em 07 jun. 2016 sobre a densidade demográfica por gênero. Disponível em: <http://www.ibge.gov.br> . Acesso em: 08 jun. 2016.

ISQUERDO, A. N.; OLIVEIRA, M. P. P. (Org). **As ciências do léxico**: lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: Ed. UFMS, 2001.

KAPPLER, A. **Perfil da Alemanha**. Berlin: Departamento de Imprensa e Informação do Governo Federal, 1999.

KORTH, H. F.; SILBERSCHATZ, A. **Sistemas de Banco de Dados**. Makron Books: 1994.

KRONE, E. E. **Comida, memória e patrimônio cultural**: a construção da pomeraneidade no extremo sul do Brasil. 2014. 175 f. Dissertação (Mestrado em Antropologia) – Universidade Federal de Pelotas, Pelotas, 2014. Disponível em: <http://guaiaca.ufpel.edu.br/handle/ri/2675>. Acesso em: 09 set. 2015.

KUHN SILVA, D. **Projeto Pomerando**: língua pomerana na Escola Germano Hübner. São Lourenço do Sul: Danilo Kuhn Silva, 2012.

KÜSTER, S. B. **Cultura e língua pomeranas**: um estudo de caso em uma escola do ensino fundamental no município de Santa Maria de Jetibá – Espírito Santo – Brasil. 2015. 255 f. Dissertação (mestrado em Educação) – Universidade Federal do Espírito Santo, Vitória, 2015. Disponível em: <http://www.educacao.ufes.br/pos-graduacao/PPGE/detalhes-de-pessoal?id=17768> Acesso em: 31 mai. 2016.

LABOV, W. **Padrões sociolinguisticos**. Trad. Marcos Bagno, Mary Scherre. São Paulo: Parábola Editorial, (1972) 2008.

LEMNITZER, L.; ZINSMEISTER, H. **Korpuslinguistik:** eine Einführung. Tübingen: Narr, 2006.

MALTZAHN, G. M. **Família, ritual e ciclos de vida:** estudo etnográfico sobre narrativas pomeranas em Pelotas (RS). 2011. 151 f. Dissertação (Mestrado em Ciências Sociais) – Universidade Federal de Pelotas, Pelotas, 2011. Disponível em: < http://repositorio.ufpel.edu.br:8080/handle/123456789/1563>: Acesso em: 2016.

MARCUSCHI, L. A. **Produção textual, análise de gêneros e compreensão**. São Paulo: Parábola Editorial, 2008.

MARTINS, A. L.; COHEN, I. S. **O Brasil pelo olhar de Thomas Davatz (1856-1858)**. São Paulo: Atual, 2000. (Série: O olhar estrangeiro).

MCENERY, T.; WILSON, A. **Corpus Linguistics an Introduction**. Manchester: Edinburgh University Press, 1996. Digital text – CD.

_____. **Corpus Linguistics.** Manchester: Edinburgh University Press, 2001. Digital text – CD.

MCENERY, T. **Corpus: Some key terms**. Manchester: Edinburgh Lancaster University, 2013. Digital Glossary – CD.

MELLO, H. Os *corpora* orais e o c-oral-brasil. In: RASO, T.; MELLO, H. (Org.). **C-Oral-Brasil**: corpus de referência do português brasileiro falado. Belo Horizonte: Ed. UFMG, 2012. – CD1.

MORRIS, J. **Sociolinguistic variation and regional minority language bilingualism:** an investigation of Welsh-English bilinguals in North Wales. 2013. 363 f. Tese (Doutorado em Filosofia) – University of Manchester, Manchester, 2014. Disponível em: <academia.edu/johnathanmorris>. Acesso em: 02 jul. 2015.

MUJICA, M. M. **Atitude, orientação e identidade linguística dos pomeranos residentes na comunidade de Santa Augusta- São Lourenço do Sul-RS- Brasil**. 2013. 90f. Dissertação (Mestrado em Letras). Universidade Federal de Pelotas. Pelotas, 2013. Disponível em: < http://repositorio.ufpel.edu.br:8080/handle/123456789/2176>. Acesso em: 06 fev. 2015.

MÜLLER, T. L. **Colônia Alemã**: histórias e memórias. Porto Alegre: Escola Superior de Teologia São Lourenço de Brindes, 1981.

MÜNCHOW, A.; VORPAGEL, M.; WENDLER, H. **Bíbliha aventures**: aventuras da Bíblia em pomerano. Barueri: Sociedade Bíblica do Brasil, 2012.

MURAKAMI, D. S.; TAGNIN, S. E. O. Linguística de Corpus e ensino: a compilação de um corpus de especialidade para preparação e implementação de um curso preparatório rápido para exame de proficiência. **Letras & letras**. Uberlândia, v. 30, n. 2, p. 349-366, jul/dez. 2014. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras>. Acesso em: 20 mar. 2015.

NASCIMENTO, A. K. P.; TAGNIN, S. E. O. A música e os ruídos na legendagem francesa para surdos e ensurdecidos. **Letras & letras**. Uberlândia, v. 30, n. 2, p. 244-260, jul/dez. 2014. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras>. Acesso em: 20 mar. 2015.

NEMÉNYI, À. V. N. **Der Slawen-Mythos**: Wie aus Ostgermanen ein Volk der "Slawen" mit fremder Sprache und Mythologie wurde. Norderstedt: Verlag Books on Demand, 2015.

NOVODVORSKI, A. **Estilo das traduções de Sergio Molina de obras de Ernesto Sabato**: um estudo de corpora paralelos Espanhol/Português. 2013. 259 f. Tese (Doutorado em Linguística Aplicada) – Universidade Federal de Minas Gerais, Belo Horizonte: Ed. UFMG, 2013. Disponível em <http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/LETR-96LMU8>. Acesso em: 2016.

_____. **Linguística de Corpus aplicada a pesquisas de base empírica**. PPGEL/UFU, Uberlândia, 2015, Slide 15 (Material utilizado em aula).

NOVODVORSKI A.; FINATTO, M.J.B. Linguística de Corpus no Brasil: uma aventura mais do que adequada. **Letras & letras**. Uberlândia, v. 30, n. 2. p. 7-16, jul/dez. 2014. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras>. Acesso em: 02 mar. 2016.

OMNIPAGE PROFESSIONAL.Versão 17.0, Nuance Communications, 2009 – CD.

PARODI, G. **Linguística de Corpus**: de la teoría a la empiria. Madrid: Iberoamericana, 2010.

PERINI, M. A. **Estudos de gramática descritiva**: as valências verbais. São Paulo: Parábola Editorial, 2008.

_____. **Princípios de linguística descritiva**: introdução ao pensamento gramatical. São Paulo: Parábola Editorial, 2006.

PESSOA, M. do S. **Ontem e hoje**: percurso linguístico dos pomeranos de Espigão D'Oeste-RO. 1995. 242 f. Dissertação (mestrado em Linguística) – Unicamp, Campinas, 1995. Disponível em <http://www.bibliotecadigital.unicamp.br/document/?code=vtls000101925&opt=3>. Acesso em: 07 nov. 2014.

PINHEIRO, G. M. **Redações do ENEM**: estudos dos desvios da norma padrão sob a pespectiva de corpus. 2007. 152 f. Dissertação (Mestrado em Estudos Linguísticos e Literários em Inglês) – Universidade de São Paulo, São Paulo, 2007.

POMPE, G. A. **Das Pommernlied**, 1851. Disponível em: <http://www.westfalenlied.de/westfalenlied3/deutschesreich_pommern.htm> Acesso em: 03 mar. 2016.

RASO, T. MELLO, H. Parâmetros de compilação de um corpus oral: o caso do C-ORAL-BRASI. **Revista Veredas On-Line**: linguística de corpus e computacional. Juiz de Fora, v. 2, p. 20-35, fev, 2009.

RAECK, F. **Pommersche Literatur**. Hamburg: Pommerscher Zentralverband, 1969. Disponível em: <http://www.uni-greifswald.de/bibliothek.html>. Acesso em: 20 dez. 2014.

RHEINISCHES WÖRTERBUCH. Universidade de Trier. Alemanha. Disponível em: <http://www.woerterbuchnetz.de>. Acesso em: 02/06/2016.

RODRIGUES, A. F. **A narrativa de Thomas Davatz:** relato de memória e documento histórico, sentimentos e ressentimentos na História. (1850-1888). 2008. 116 f. Monografia (Graduação em História) – Intituto de História, Universidade Federal de Uberlândia, Uberlândia, 2008.

SÁNCHEZ, A. et al (Org.). **CUMBRE – Corpus Linguistico del Espanol Contemporaneo**: fundamentos, metodologia, y Aplicaciones. Madrid: SGEL, 1995.

SANTOS, J. V.; BRITO, G. S. Fotografia técnica de documentos para formação de *corpora* digitais eletrônicos: o método desenvolvido no Lapelinc. NOVODVORSKI, A. FINATTO, M. J. B. (Org). **Letras & letras**. v. 30, n. 2, p. 421-430, jul/dez. 2014. Disponível em: <http://www.seer.ufu.br/index.php/letraseletras>. Acesso em: 20 mar. 2015.

SANTOS, I. P. Memória Coletiva, Geolinguística e relações textuais-discursivas. In: SANTOS, J. B. C (Org.). **Sujeito e sujetividade**: discursividade contemporânea. Uberlândia, MG: EDUFU, 2009.

SEIBEL, I. **Imigrante no século de isolamento**: 1870 - 1970 São Leopoldo: EST/PPG, 2010. 350 f. Relatório (Pós-Doutorado em Teologia) – Escola Superior em Teologia. São Leopoldo, 2010.

SEYFERTH, G. A conflituosa história da formação da etnicidade teutobrasileira. In.: FIORI, Neide Almeida. **Etnia e educação**: a escola "alemã" do Brasil e estudos congêneres. Florianópolis: Editora da UFSC, 2003. p. 21-61.

SCOTT, M. **Manual WordSmith Tools**: version 6.0, Oxford University Press: 2015 – CD.

SAUSSURE, Ferdinand de. **Curso de Linguística Geral**. Org. por Charles Bally e Albert Sechehaye. Colaboração de Albert Riedlinger. Trad. Antônio Chelini, Losé Paulo Paes e Izidoro Blikstein. São Paulo: Editora Cultrix, 1969.

SCHAEFFER, S. C. B. **Descrição Fonética e Fonológica do Pomerano Falado no Espirito Santo**. 2012, XX f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal do Espírito Santo. Vitória, 2012. Disponível em: <http://repositorio.ufes.br/handle/10/1622>. Acesso em: 08 nov. 2014.

SINCLAIR, J. McH. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

_____. **Trust the Text**: Language, corpus and discourse, London: Routledge, 2004.

SOARES, R. de C. S. **Atlas semântico-lexical da região norte do alto Tietê (ReNAT)**. São Paulo. Tese (Doutorado em Linguística). USP, 2012. 664f. Disponível em: <www.teses.usp.br/teses/disponiveis/8/8139/tde-09012013-113318/>. Acesso em: 20 abr. 2016.

STEMMER, G. H. **Apostila de Paleografia Gótica**: visando a leitura de manuscritos em alemão em letra gótica cursiva. São Leopoldo, 2008.

STUBBS, M. **Texts, Corpora, and Problems of Interpretation:** A Response to Widdowson. Applied Linguistics. Local [S.l]: Oxford University Press, 2001. p.49-172.

STURZ, J. J. **Die Deutschen Auswanderung und die Verschleppung deutscher Auswanderer**. Berlin: Verlag von Fr. Rorttampf, 1868.

_____. **Die Krisis der Deutschen Auswanderung**. Berlin: Sidethier Verlag, 1862.

SUBEXTRACTOR. Versão 1031, 2013. Disponível em: <https://subextractor.codeplex.com/releases/view/100949>. Acesso em: 13 set. 2014.

TAGNIN, S. E. O. **Comet**: um corpus Multilingue para o ensino e tradução. São Paulo: Ed. USP, 2001. p. 1-6.

_____. **Um corpus multilíngue para ensino e tradução**: COMET. **Tradterm**, São Paulo, v. 10, p. 117-141, 2004. Disponível em: < http://www.revistas.usp.br/tradterm/article/view/47050/50773>. Acesso em: 2016.

_____; VIANNA, V. **Corpora no ensino de línguas estrangeiras**. Belo Horizonte: Hub Editorial, 2010.

_____. Glossário de Linguística de Corpus. In: VIANA, V.; TAGNIN, S. E. O. (Org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: HUB Editorial, 2010. p. 357-361.

_____. Linguística de Corpus e Fraseologia. In: Maria Luisa Ortiz Alvarez & Enrique Huelva Unternbäumen. (Org.). **Uma (Re)visão da Teoria e da Pesquisa Fraseológicas**. 1ed. Campinas: Pontes Editores, 2011.

_____. **O jeito que a gente diz**. São Paulo: Disal, 2005.

TAKANO, Y. **Esboço do Atlas do Falar dos Nipo-brasileiros do Distrito Federal**: Aspecto Semântico-Lexical. 2013. 362 f. Tese (Doutorado em Linguística) – Universidade de São Paulo, São Paulo, 2013.

TAYLOR; FRANCIS. **Corpus Stylistics**, Taylor & Francis (Publisher): [20--].

TEIXEIRA, E. D. **A Linguística de Corpus a serviço do tradutor**: proposta de um dicionário de culinária voltado para a produção textual. 2008. 439 f. Tese (Doutorado em Estudos Linguísticos e Literários em Inglês) – Universidade de São Paulo, São Paulo, 2008. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-16022009-141747/pt-br.php>. Acesso em: 02 fev. 2016.

TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam e Philadelphia: John Benjamins, 2001.

TRESSMANN, I. O pomerano: uma língua baixo-saxônica. **Educação, cultura e sociedade**: Revista da Farese. Santa Maria de Jetiba, v. 1. p. 10-21. 2008. Disponível em: <http://www.farese.edu.br/pages/biblioteca/revistafarese.php>. Acesso em 06 ago. 2015.

_____. **A Classificação da Língua Pomerana**. Santa Maria de Jetibá, ES, 2010. Disponível em: <http://www.pomerano.com/videos/fatos-hitoricos-da-imigracao.html>. Acesso em: 07 fev. 2013.

_____. **A Língua Pomerana.** Santa Maria de Jetibá, 2010. (vídeo). Disponível em: <http://www.pomerano.com/videos/fatos-hitoricos-da-imigracao.html>. Acesso em: 07 fev. 2013.

_____. Bilinguismo no Brasil: o caso da comunidade pomerana de Laranja da Terra. **Revista da Associação de Estudos da Linguagem do Rio de Janeiro**, Rio de Janeiro: UFRJ, 1998. Disponível em: <https://www.google.com.br/url?sa=t&source=web&rct=j&url=http://www.farese.edu.br/pages/artigos/pdf/ismael/O%2520caso%2520da%2520Com.%2520pomer.%2520de%2520LT.%25201998.pdf&ved=0ahUKEwjopOnE3cXNAhXGIpAKHegzCwwQFggkMAM&usg=AFQjCNFARK3OWeQckkNgF2CMI3N0mjbwOw >. Acesso em: 06 ago. 2015.

_____. **Dicionário Enciclopédico Pomerano-Português.** Santa Maria de Jetibá: Ed. Farese, 2006.

TRESSPACH, R. **As relações históricas entre Brasil, Alemanha e as populações de língua alemã**. Artigo publicado em My Heritage Brasil, 2013. Disponível em: <http://www.rodrigotrespach.com/imigracao-alema-no-brasil/>. Acesso em: 10/06/2016.

VIANA, V. Linguística de *corpus*: conceitos, técnicas e análises. In: VIANA, V.; TAGNIN, S. (Orgs.). ***Corpora*** **no ensino de línguas estrangeiras**. São Paulo: HUB Editorial, 2010. p. 25-96

VOGT, O. P. Germanismo e Nacionalização em Santa Cruz do Sul, RS**.** In: **História Política de Santa Cruz do Sul**. UNISC. Santa Cruz do Sul, v.7, n.2, p. 49-92, jul/dez. 2001.

VOLLMER, M. **Zur pommerschen Dialektlexikographie. Kosegartens Wörterbuch der niederdeutschen Sprache älterer und neuerer Zeit**: Jahrbuch des Vereins für niederdeutsche Sprachforschung. Greifswald: [s.n.], 2008.

VON BORSTEL, C. N. **A linguagem sociocultural do Brasildeutsch**. São Carlos: Pedro & Paulo Editores, 2011.

_____. Sociolinguística: teoria, método e objeto em pesquisas in loco. In: **Web-Revista Sociodialeto**, Campo Grande, v. 4, nº 12, mai. 2014, p. 504-524. Disponível em: <www.sociodialeto.com.br>. Acesso em: 09 fev. 2016.

WALACHAI. Documentário. Rejane Zilles. Walachai, Okna Produções e Zilles Produções Culturais, Traquitana Filmes e Artesanato Digital: 2011. 1 DVD.

WALTER, H. **A Aventura das Línguas no Ocidente**: origem, história e geografia. Trad. Sérgio Cunha dos Santos. São Paulo: Mandarim, 1997.

WANGERINER-TREFFEN, 2015. Disponível em: <file:///E:/17-05-2016/Heimatverein%20Wangerin%20-%20Wangeriner%20Treffen%20in%20Bispingen%20-%20Fahrten%20nach%20Wangerin%20_%20Wegorzyno_arquivos/left_heimatverein.htm>. Acesso em: 03 mar. 2016.

WIESEMANN, U. (Org). **Manual Pomerano**. São Leopoldo: [s.n.], 2008. (Material não publicado).

WILLE, Leopoldo. **Pomeranos no sul do Rio Grande do Sul**: trajetória, mitos, cultura. Canoas, 2011.

WITTER, J. S. **A Revolta dos Parceiros**. São Paulo: Brasiliense, 1986. (Série Tudo é História).

WORDSMITH TOOLS. Versão 6.0, Stroud: Lexical Analysis Software, Oxford University Press, 2015. – CD.

ZAVAGLIA, A.; WELKER, H. O que é Léxico. [2008] 2013. **GTLEX. ANPOLL**. Disponível em: < www.letras.ufmg.br/gtlexnovo > Acesso em: 01 abr. 2015.

ZILLY, B. **Grande Sertão Veredas, em Alemão**. Palestra realizada em 05 abr. 2013 na Faculdade de Letras – FFLCH-USP. Convênio entre a Freie Universität Berlin e Universidade de e a Universidade Federal de Santa Catarina – UFSC. Disponível em: <https://www.youtube.com/watch?v=wr7im4vl-AU>. Acesso em: 19 mai. 2016.

APÊNDICES

APÊNDICE A - Levantamento de algumas famílias pomeranas residentes e falantes de pomerano no leste de Minas Gerais (censo informal)

FAMÍLIAS POMERANAS RESIDENTES EM ITUETA/MG E FALANTES OU NÃO DE POMERANO:			
Sobrenomes	Fala	Não Fala	Total
ACKER	1		
AGNER	1		
AHNERT	1		
BATKE	1		
BAUTZ		1	
BERGER	1		
BESSERT	1		
BICH		1	
BEILKE	1		
BILKE	1		
BINOW	1		
BORCHARDT	1		
BRANDEMBURG	1		
BRODEL	1		
BUTZKE	1		
DANN	1		
DUMMER	1		
EGGERT	1		
ELLER		1	
EMERICH		1	
FANCKEN		1	
FEHLBERG	1		
FELZ		1	
FICK	1		
FINCK		1	
FOESTER	1		
FRIEDERICH	1		
GABRECHT	1		
GAEDE	1		
GUMS	1		
HACKER	1		
HENCKER	1		
HEUDEMANN	1		
IANCKER	1		
JANN	1		
JANUTTE	1		
KAMKE	1		
KAMPIN	1		
KAPICHE		1	
KEFFIER	1		
KILL	1		
KLEMZ	1		
KLIPER	1		
KLIPPEL	1		
KLUG	1		
KOLLER	1		
KRAUSE	1		
KRUGEL	1		
KRÜGER	1		
KUSTER	1		
LEMCKE	1		
LOOSE	1		
LORETE		1	
LUKAS		1	
LUXINGER	1		
MARGUADT	1		
MUTZ	1		
NEITZEL	1		
NEUMANN	1		
OHNESORGE		1	
PERCILIUS	1		
PETER	1		
PIEPER	1		
PISKE	1		
PITELKOW	1		
POLACK		1	
PONATH	1		
RAASCH	1		
RATZKE	1		
RECKEL	1		
ROSSOW	1		
RÜSTER	1		
RUTSATZ	1		
SCHEIBER	1		
SCHEIDER		1	
SCHIMMELPFENNIG	1		
SCHNEIDER	1		
SCHULTZ	1		
SEIBEL	1		
SEIBERT	1		
SEIDLER	1		
SHWENKE		1	
SITLER	1		
STEFFEN			
STEIN		1	
STHUTTZ	1		
STRELOW	1		
STUHR	1		
TESCH	1		
TETZNER	1		
TONES	1		
UHLIG	1		
WAGNER			
WELLMANN	1		
WELMER	1		
WENDT	1		
WILL		1	
WITT	1		
ZUMACK	1		
Quantidade	39	16	55
Porcentagem	70,91%	29,09%	100%

APÊNDICE B - Levantamento de algumas famílias pomeranas que emigraram do leste de Minas Gerais (censo informal)

LEVANTAMENTO DE FAMÍLIAS POMERANAS QUE MIGRARAM DO LESTE DE MINAS E FALANTES DE POMERANO

SOBRENOMES	FALAVAM	NÃO FALAVAM	NÃO-IDENTIFICADAS	SOBRENOMES	FALAVAM	NÃO FALAVAM	NÃO-IDENTIFICADAS
ABEL			1	KREIDLOW	1		
ALBERT			1	KÜHL			1
BEILKE	1			KUSANKE			1
BELING	1			LUFT			1
BELZ	1			MANTAICH			1
BÖHLING	1			MEIER			1
BOLDT	1			MÜLLER			1
BRANDT	1			NEUBAUER		1	
BRAUN	1			NIENKE	1		
BRUNOW			1	NIMMER	1		
BUGUER			1	NITRE			1
BUSS	1			NÖMEG	1		
EBERT	1			PISOLER		1	
ERBERT	1			PRÖDEL	1		
ERDMANN	1			RAMLOW	1		
FIEGEL			1	RETZ			1
FISCHER			1	ROMAIS	1		
FLEISCHMANN			1	ROOS	1		
GAMIEL			1	ROSSMANN	1		
GÖRING			1	SASSAENBURG	1		
GÖRL			1	SCHEENK			1
GÖSE	1			SCHEINER			1
GRAUNKE			1	SCHÖROCK	1		
GRULKE	1			SCHRÖDER	1		
HAASE			1	SCHWANZ	1		
HACKBART			1	SEEFEIDT**		1	
HAMMER			1	SCHMIDT	1		
HERBS			1	SOLIMIDA			1
HING			1	STIFELBEIN			1
HINZE			1	TIMM	1		
HOFMANN	1			TOM	1		
HOLZ	1			VERDIN	1		
ICKERT			1	WALKINIR			1
JAHNUT	1			WELKE			1
JESKE			1	WERNECK		1	
KADRATZ			1	WERNTAL			1
KANITZ			1	WESTPHAL	1		
KAUN*			1	WILKE			1
KLABUNDE			1	WOLFGRAM	1		
KÖPP	1			WRUCK	1		
				WUTKE	1		
				ZEMKE			1
				ZIMMERMANN	1		

Total 83	39	4	40	83
100%	47%	5%	48%	100%

Total de Famílias redidentes e ausentes do leste de Minas Gerais pesquisadas: 83 + 97 = 180 famílias pomeranas.

APÊNDICE C – Procedimento metodológico para transcrições e compilações: Documento de conversão e convenção da escrita dos *corpora* do pomerano

1) /å/ = /oo/ – Justificativa: encontramos na literatura pomerana o duplo "o" quando o som representa um "ó" longo. Além disso, é uma escrita simplificada, que não dá problemas de distorção no WST.
2) /ó/ = /oo/ – Justificativa: trata-se apenas de uma variação de escrita para o mesmo som do item 1, já era grafado com duplo "o" = "oo". E também porque os dobrados são característicos do pomerano.
3) /g/ com som de /ch/ = /g/ – Quando estiver no meio ou no final das palavras. Justificativa: a interpretação sonora do "g" é que muda. Outra justificativa é que grafar "ch" pode causar confusões, por exemplo: Dach = telhado e Dag = dia
4) /g/ com som de /j/ = /g/ – Quando está no início das palavras, o "g" comporta se como o "J" no padrão germânico, essa característica é recorrente na fala pomerana. Exemplos: Geld - Jeld, Genau - Jenau, Geschichte - Jeschichte, etc.
5) /g/ com som de /g/ = /g/ – Ocorre, em posições finais, mas existem variações de pronúncia, por exemplo, o som de "k" ou de "ch" no final, assim, o "g" será preservado como "g", pois trata-se de uma interpretação sonora diferente para uma mesma grafia. Alterar muito a escrita pode fazer o pomerano parecer mais distante do alemão, do que realmente é.
6) /ö/ – /ö/ – Na primeira versão da conversão e convenção havíamos adotado "oe", porque o programa WST não estava conseguindo ler os caracteres, porém, conseguimos resolver o problema configurando também o computador e os arquivos de texto para o alemão. E, não somente a configuração do WST, como anteriormente, assim, o problema foi resolvido e não ocorreu mais a distorção das vogais com trema. Então, decidimos manter os tremas, pois fazem parte das grafias germânicas originalmente encontradas nas fontes.
7) /ä/ – /ä/ - justificativa: Idem.
8) /ü/ – /ü/ – justificativa: Ibidem.
9) Em glouwe, a pronúncia varia para glauwe (HD: glaube), manteremos "ou" quando escrito "ou" e "au" quando escrito "au", pois trata-se de variação e faremos as observações ao lado da transcrição em etiquetas.
10) /Das/ – /dat/ – justificativa: trata-se da variação do artigo neutro com o som do "t", além de já ser frequentemente grafado assim.
11) /es/ – /et/ - justificativa: Idem. Respeitando a variação, quando for pronunciado "es" será grafado fielmente como "es".
12) deit, dêit, deyt, deet = deet - justificativa: Trata-se da mesma palavra e já era escrito dessa forma desde 1588.
13) mij/mii – mie = justificativa: o som do "i" longo foi convencionado no padrão germânico (Duden, 1901) como "ie" em que prevalece o som do "i".
14) /j/ – Manteremos o "j" em casos em que eram originalmente escritos com "j", porém nos casos em que o "j" é usado excessivamente no lugar do "e" depois do "i" ("ie"), manteremos a ortografia "e" para não fazer parecer que a escrita do pomerano tem uma estética ou aparência eslava, pois não foram encontrados elementos eslavos no pomerano atual. Exemplos: mitglijd, mitkrijga, dij, mijcha, mijn (casos dicionarizados por Tressmann, 2006), escrevemos: mitglied, mitkriega, die, miecha, mien. Mas em casos como jeden, jetzt, etc., permanecem a escrita com "j".
15) schlafen, schlópa, xlópa, slåpa = schloopa - justificativa: É a forma de escrita encontrada na Barther Bibel, além disso reproduz o som necessário, pois trata-se de uma variação linguística e tem os dobrados característicos do pomerano em sua origem.
16) Casos escritos como Sward, srijwen, etc., em que foram retirados da grafia o "ch" em "sch" que ainda são pronunciados, recolocamos o "ch", ou seja, grafamos *schwartz* ou *schwart*, quando

não há a pronúncia do “z”. E, em srijwen, grafamos “schriewen”. Nos contatos com os pomeranos constatamos a pronuncia “sch” e não “s” (Mas reconhecemos que podem haver variações diatópicas que não presenciamos).
17) Manter os doppels – justificativa: as letras dobradas eram características presente na literatura germânica, encontrada até hoje na biblioteca de Greifswald.
18) Não diferenciar maiúsculas de minúsculas nos substantivos – justificativa: Nas variedades de falas germânicas não existia a grafia dos substantivos com a primeira letra maiúscula, essa forma veio a ser convencionada depois, por volta de 1901 (Duden) com a padronização do HD e não era uma característica pomerana. Em todas as grafias do pomerano que encontramos as iniciais dos substantivos eram minúsculas.

Observação: Questões de conversão e convenção da escrita não abrangidas aqui, serão oportunamente identificadas pelas etiquetas nas transcrições e compilações dos *corpora*. As variações linguísticas e diatópicas do pomerano, do alemão e do português foram respeitadas nas transcrições. Padronizamos as decisões de conversão e convenção da escrita como recurso metodológico para leitura dos *corpora* pelo WST. Ficam aqui documentadas as nossas opções metodológicas.

APÊNDICE D - Lista de textos coletados e que compõem o PKE - *Pommersche Korpora Escritos*

1. LC-SA-AT-PKE-P – Textos (Inscrições) de Lápides de cemitérios, sem autoria, origem cemitério evangélico de Arroio do Tigre/RS.
2. LC-SA-RP-PKE-P – Textos (Inscrições) de Lápides de cemitérios, sem autoria, origem cemitério evangélico do Rio Pardo/RS.
3. LC-SA-SMJ-PKE-P – Textos (Inscrições) de Lápides de cemitérios, sem autoria, origem cemitério evangélico de Santa Maria de Jetibá/ES.
4. LC-SA-VC-PKE-P – Textos (Inscrições) de Lápides de cemitérios, sem autoria, origem cemitério evangélico de Vera Cruz/RS.
5. LC-SA-VN-PKE-P – Textos (Inscrições) de Lápides de cemitérios, sem autoria, origem cemitério evangélico do entorno de Vila Neitzel, Zona Rural Norte de Itueta/MG.
6. RE-RS-PKE-P-1 – Textos dos Livros de Registros eclesiásticos da IECLB de Vera Cruz/RS antiga Villa Theresa. Autoria Pastor Wolfram, dados baseados em Elisandro José Migotto. Parte I dos Taufregister. Registros 1 a 85. 1861.
7. RE-RS-PKE-P-2 – Textos dos Livros de Registros eclesiásticos da IECLB de Vera Cruz/RS antiga Villa Theresa. Autoria Pastor Wolfram, dados baseados em Elisandro José Migotto. Colônias e Parte II dos Taufregister. Registros 86 a 267. 1862.
8. RE-RS-PKE-P-3 – Textos dos Livros de Registros eclesiásticos da IECLB de Vera Cruz/RS antiga Villa Theresa. Autoria Pastor Wolfram, dados baseados em Elisandro José Migotto. Parte III – Copulationsregister e origens dos imigrantes.
9. CP-AWD-LS-PKE-DT – Carta pessoal. Doação da família WP da ZRN/IT/MG- autor: ÁJWD- Doada pela destinatária LP. Compilada do original manuscrito em papel.
10. CP-FHT-C-PKE-DT 1 – Carta pessoal. Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
11. CP-FHT-C-PKE-DT 2 – Carta Pessoal. Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
12. CP-FHT-C-PKE-DT 3 – Carta Pessoal (falta trechos). Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
13. CP-FHT-C-PKE-DT 4 – Carta Pessoal (incompleta). Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
14. CP-FHT-C-PKE-DT 5 – Carta Pessoal. Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
15. CP-FHT-C-PKE-DT 6 – Carta Pessoal. Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
16. CP-FHT-C-PKE-DT 7 – Carta Pessoal (parece incompleta ou uma carta curta). Origem: Família HT. Doação de MHT em Canguçu/RS. Manuscrito em papel.
17. CP-MHT-C-PKE-DD – Carta Pessoal da família HT. Origem: Canguçu/RS. Doada por MHT. Compilada a partir do original em manuscrito em papel.
18. RC-RS-PKE-E -1 – Receita culinária. Autoria: RT. Zona rural de Canguçu/RS. Manuscrito em papel.
19. RC-RS-PKE-E -2 – Receita culinária. Autoria: RT. Zona rural de Canguçu/RS. Manuscrito em papel.
20. RC-RS-PKE-E -3 – Receita culinária. Autoria: RT. Zona rural de Canguçu/RS. Manuscrito em papel.
21. RC-RS-PKE-E -4 – Receita culinária. Autoria: RT. Zona rural de Canguçu/RS. Manuscrito em papel.

22. J-I-PKE-DT-1 – Jornal. Folha Pomerana Express. Informativo das Comunidades Pomeranas Brasileiras - Fundado em 19 de agosto de 2013. Ano I de 2013 a 2014. Venâncio Aires/RS.
23. J-I-PKE-DT-2 – Jornal. Folha Pomerana Express. Informativo das Comunidades Pomeranas Brasileiras - Fundado em 19 de agosto de 2013. Ano II de 2014 a 2015. Venâncio Aires/RS.
24. J-I-PKE-DT-3 – Jornal. Folha Pomerana Express. Informativo das Comunidades Pomeranas Brasileiras - Fundado em 19 de agosto de 2013. Ano III de 2015 a 2016. Venâncio Aires/RS.
25. AI-LR-I-PKE-D – Textos Diversos da Internet. Turismo e Cultura Pomerana em Santa Maria do Jetibá/ES Ana Paula de Abreu e Lima, Reinaldo Dias.
26. PK-PKE-I-E (1) – Textos Diversos da Internet. O Calendário Pomerano: Datas e Épocas Importantes.
27. PK-PKE-I-E (2) – Textos Diversos da Internet. Pommerisch Koonelar. Calendário Pomerano. Blog do Charles Ledebuhr.
28. PK-PKE-I-E (3) – Textos Diversos da Internet. Calendário Pomerano 2012. Pomerisch Koolener 2012. Autor: Ismael Tressmann.
29. T-B-I-PKE-D– Textos Diversos da Internet. Up Pommersch. Em Pomerano. Blog do Pommerplattdütsch.
30. T-FB-I-PKE-D – Textos Diversos da Internet. Post diversos em pomerano. Facebook do grupo pomeranos (Transcrição e conversão de 392 posts).
31. T-MK-I-PKE-D – Textos Diversos da Internet. Palavras do Cotidiano Krause - Alto Graminha. Dicionário Pomerano dos Netos. Blog da Família Krause (Glossário, frases e textos curtos).
32. T-PKE-I-E (1) – Textos Diversos da Internet. Ous Hubschigkeit, ous Kultur - Nossa beleza, nossa cultura. Rio Ponte, Domingos Martins-ES. Blog Rio Pontes. E texto Páscoa Pomerana. Site Montanhas Capixabas.
33. T-PKE-I-E (2) – Textos Diversos da Internet. Up Pommersch. Blog Pommerplattdütsch. Versos, frases, palavras e outras postagens.
34. T-PKE-I-E (3) – Textos Diversos da Internet. Nüüg Pommerland is hier! Blog Pomeranos Pomer.
35. TPS-PKE-I-D – Textos Diversos da Internet. 25ª Pomerisch Fest. Pomeranos fazem a festa em Santa Maria de Jetibá. Blog Montanhas Capixabas.
36. WPP-I-PKE-D – Textos Diversos da Internet. Lista de Palavras do Wiktionary Pomerano.
37. D-TP-ES-PKE-WMP – Documentário *Dai Pomerer: dai gang fon ainem folk*. Pomeranos: A trajetória de um povo. Direção de Vanildo Kruger, 2009.
38. D-W-RS-PKE-WMP – Documentário *Walachai* (Trechos de Legendas transcritas e convertidas das cenas com fala pomerana). Direção de Rejane Zilles, 2011.
39. TA-DKS-MEH-PKE-D – Trabalho Acadêmico. Artigo sobre a emigração pomerana através da canção de motter einer Hochtied. Autor: Danilo Kuhn Silva. Origem: São Lourenço do Sul/RS.
40. TA-DKS-NCP-PKE-D – Trabalho Acadêmico. Artigo sobre "A música pomerana como narrativa da memória cultural". Autor: Danilo Kuhn Silva. Origem: São Lourenço do Sul/RS.
41. TA-DKS-PP-PKE-D – Trabalho Acadêmico. Artigo sobre "Projeto Pomerando: Língua Pomerana na escola Germano Hübner". Autor: Danilo Kuhn Silva. Origem: São Lourenço do Sul/RS.
42. TA-HR-P-PKE-D – Trabalho Acadêmico. Artigo com trechos do Livro Pomeranos de Helmar Rölke.
43. TD-CTA-IS-PKE-D – Trabalho Acadêmico. Textos anexos em pomerano. Relatório de pós-doutorado. Imigrante no século de isolamento. Autor: Ivan Seibel. São Leopoldo.

44. BPD-EV-FK-ZRN-IT-MG-IP – Textos Religiosos. Dat Ni Testament Plattdütsch. Dat Evangelium von Johannes - 6. Kapitel. Tradução de Ernst Voß, 1960.
45. JW-I-PKE-DT – Textos Religiosos. Hörg God un Leew up ümmer. Encarte dos Testemunhas de Jeová para evangelismo dos pomeranos. Jehowas Witness. Editoras Watch Tower Bible and Tract Society of New York, Inc. Brooklyn, New York, U.S.A. 2013 Watch Tower Bible and Tract Society of Pennsylvania.
46. JW-R-PKE-WMP – Textos Religiosos. Biebels bläärer e Privatsiteit Politik. Encarte dos Testemunhas de Jeová para evangelismo dos pomeranos. Jehowas Witness. Editoras Watch Tower Bible and Tract Society of New York, Inc. Brooklyn, New York, U.S.A. 2013 Watch Tower Bible and Tract Society of Pennsylvania.
47. TBA-SBB-PDF – Textos Religiosos. Jesus geeht rinna in Jerusalem. Texto Bíblico do Advento. SBB. Evangelho do Primeiro Domingo no Advento. Baseado no livro de Lucas 19.28-40.
48. TEL1-KT-SLS-RS – Textos Religiosos. Tradução do evangelho de Lucas, capítulo 1 para o pomerano. SBB. Canguçu/RS.
49. TEL2-KT-SLS-RS – Textos Religiosos. Tradução do evangelho de Lucas, capítulo 2 para o pomerano. SBB. Canguçu/RS.
50. TEL3-KT-SLS-RS – Textos Religiosos. Tradução do evangelho de Lucas, capítulo 3 para o pomerano. SBB. Canguçu/RS.
51. T-UW-I-E– Textos Religiosos. Dai hochtied in Caná (O casamento em Caná). Autoria coletiva de pomeranos. Organizadora: Ursula Wiesemann. São Lourenço do Sul/RS.
52. M-DKS-SLS-DT – Músicas Pomeranas. “Siegel, siegel, sächer bot”,“Schloop kint, schloop”, “Schloop, kinga schloop”, “Ringel, ringel roo”, “Hoopa, hoopa mähla”. Autoria: Conhecimento popular coletado por Danilo Kuhn Silva em São Lourenço do Sul/RS.
53. M-DKS-SLS-RS-DT1 – Músicas Pomeranas. Outra versão de “Schloop kinka schloop” doada por entrevistada RW.
54. M-DKS-SLS-RS-DT2 – Músicas Pomeranas. “Groute bouhnen leed”. Autoria: Conhecimento popular doada por entrevistada RW.
55. M-DKS-SLS-RS-DT3 – Músicas Pomeranas. “Mäcken wust du frühn”. Autoria: Conhecimento popular doada por entrevistada RW.
56. M-DKS-SLS-RS-PDF – Músicas Pomeranas. “De mutter einer hochtied”. Autoria: Conhecimento popular (cantada pelas bandas locais) coletado por Danilo Kuhn Silva em São Lourenço do Sul/RS.
57. M-DKS-SLS-RS-PDF2 – Músicas Pomeranas. “De fest”. Autoria: Conhecimento popular (cantada pelas bandas locais) coletado por Danilo Kuhn Silva em São Lourenço do Sul/RS.
58. M-EGH-SLS-RS-DT – Músicas Pomeranas. “Marie, Mara, Maruschkaka”. Autoria: Conhecimento popular (cantada em atividades da escola Germano Hübner). Coletada pessoalmente com professoras da zona rural de São Lourenço do Sul/RS.
59. M-EML-SLS-RS-DT1 – Músicas Pomeranas. Nossa transcrição de outra versão de “Marie, Mara, Maruschkaka”. Autoria: Conhecimento popular (cantada em atividades da escola Martinho Lutero). Coletada pessoalmente com professoras da zona rural de São Lourenço do Sul/RS.
60. M-EML-SLS-RS-DT2 – Músicas Pomeranas. Outra versão de “Marie, Mara, Maruschkaka”, mais curta. Autoria: Conhecimento popular doada por entrevistada RW.
61. M-HG-IT-MG-PKE-D – Músicas Pomeranas. Música “ich geehn nach Blumenau”, música hunsriqueana na versão pomerana-brasileira. Autoria: Conhecimento popular. Versão HG de Itueta/MG.
62. M-MHT-C-RS-DT2 – “Hopp hopp hopp in schwien galopp”. Autoria: Conhecimento popular retirado da tradição oral. Coletado pessoal com MHT em Canguçu/RS.

63. M-MHT-C-RS-PKE-E – Músicas Pomeranas. “Eina, meina, miena, mäcka”. Autoria: Conhecimento popular. Coletado pessoal com MHT em Canguçu/RS.
64. C-CL-DKS-SLS-DT – Texto literário. “Dai suowen kleiner seechen” (Versão transliterada de Der Wolf und die sieben jungen Geißlein. Grimm, 1812). Conto infantil em pomerano: O lobo e os sete cabritinhos. Doado por Danilo Kuhn Silva. Coletado em São Lourenço do Sul/RS.
65. C-CL-EML-SLS-WMP – Texto literário. Outra versão do conto “Dai suowen kleiner seechen”, os sete cabritinhos. Nossa transcrição da versão coletada pessoalmente na zona rural de São Lourenço do Sul/RS. Doada por professoras da Escola Martinho Lutero.
66. P-CL-C-RS-PKE-E – Texto literário. Poesia “ein, swei, drei”. Cedido por MHT em Canguçu/RS. Manuscrito em papel. Autor: V. K. Shneid. Escola: M. E. F. Júlio de Castilhos.
67. P-CL-DKS-SLS-M – Texto literário. Poema “Wat wilst du inner Walt?”. Autor: Danilo Kuhn Silva. Origem: São Lourenço do Sul/RS.
68. P-CL-I-PKE-D – Texto literário. Poema “Mien Herz”. Origem: Blog Pommerplattdütsch.
69. P-SA-IT-MG-PKE-E – Texto literário. “Ick bin klein” e “Wist duu wurst”. Versos do conhecimento popular coletados pessoalmente em IT/MG, com família KK que foi entrevistada e doou.
70. P-CL-WB-WS-RS-P – Texto literário. Poema “Groow sin wi Pomern”. Autor: Wilhelm Bornemann. Poema tradicional pomerano, versão com variações do original “Grout sien wier Pommern”. Origem: Wandspruch em mármore, em localidade próxima a São Lourenço do Sul/RS.
71. VT-CL-I-PKE-D – Texto literário. Versinhos, trovinhas, brincadeiras de roda, como “Hana, span an”, “Ludwig, Ludwig, duu büst besoopen” etc.
72. VT-SA-IT-PKE-CIP1 – Texto literário. Versos escritos em um Tauferinerrung. Coletado pessoalmente em visita de campo na zona rural norte de Itueta/MG. Doado por um pomerano.
73. VT-SA-IT-PKE-CIP2 – Texto literário. Versos escritos em um Tauferinerrung. Coletado pessoalmente em visita de campo na zona rural norte de Itueta/MG. Doado por um pomerano.
74. L-DL-RS-PKE-E – Livro. Projeto Pomerando. Autor: Danilo Kuhn Silva. Origem: São Lourenço do Sul/RS. 2012.
75. L-GS-SLS-PKE-CIP – Livro Os Pomeranos. Valores Culturais da Família de Origem Pomerana no Rio Grande do Sul – Pelotas e São Lourenço do Sul. Autores: SALAMONI; ACEVEDO; ESTRELA. 1995.
76. L-HR-SLS-PKE-CIP – Livro Aspectos Históricos e Geográficos da Emigração Pomerana para o Brasil. Autor: Helmar Rölke, 1996.
77. L-JFJ-PKE-E – Livro A Pomerânia Brasileira uma eterna imigração. Autor: Jorcy Foerste Jacob. Vila Pavão, 2010.
78. L-LW-SLS-PKE-CIP – Livro WILLE, Leopoldo. Pomeranos no sul do Rio Grande do Sul: trajetória, mitos, cultura. Canoas, 2011.
79. L-SBB-SLS-PKE-CIP – Livro Bíbliha Aventuras. Aventuras Bíblicas em Pomerano. SBB, 2012.
80. D-CS-I-FD-PKE-PD – Sprüche. Fraseologismos e expressões. Diversas fontes, incluindo Rölke, 1996. Tressmann, 2006. Wille, 2011. Pessoalmente em SLS/RS e Internet.
81. D-CS-LD-PKE-PI – Sprüche. Fraseologismos e expressões. Fonte: Artigo da Secretaria de Cultura e Turismo de Santa Maria do Jetibá/ES. Autores: Lima e Dias.
82. D-CS-VC-PKE-PD – Sprüche. Fraseologismos e expressões. Retirados de uma versão do Pomerisch Koolener impresso. Autor: Ismael Tressmann, 2012. Origem: Santa Maria de Jetibá. Coletado em São Lourenço do Sul/RS.
83. F-AD-CS-PKE-PD – Sprüche. Fraseologismos e expressões. Trava-línguas e ditados divrsos. Fonte e Origem: Doados pela professora Gisela Wienke Wachholz em reunião dos professores da ICK-Pommer, associação para trabalhar o pomerano nas escolas de São Lourenço do Sul/RS e outros enviados por ela por email em 04 de Setembro de 2014.

84. F-L-I-CS-PKE-PD – Sprüche. Fraseologismos e expressões. Fonte: Folhas impressas dos autores (excertos de seus livros). Autores: Rölke, 1996 e Tressmann, 2006.
85. WS-SL-SA-PKE-F– Sprüche. Fraseologismos e expressões. Wandmoolen. Livro da Exposição "Francisco Seibel Um Fotógrafo Rural no Espírito Santo". S/D. Data Estimada 2010. Livro impresso doado pelo museu de Santa Maria de Jetibá/RS.
86. AA-KS-I-PKE-D – Palavras Soltas. Fonte: Projeto Língua Pomerana na escola Germano Hübner. Livro Impresso e Também em Documento de Texto. Origem/Autor: Danilo Kuhn Silva. Data e local da coleta: 16/08/2014, SLS - cedido pelo autor>
87. PSC-AC-SLS-RS-PKE-D – Palavras Soltas em Cartazes fotografados e folhas de livro fotografado. Cartazes e atividades de turmas da escola Germano Hübner com as cores e o nome dos animais em pomerano. Zona Rural de São Lourenço do Sul/RS. Distrito de Santa Tereza e de Santa Augusta. E trechos da tese de Bahia (2000).
88. PSL-KS-SLS-RS-PKE-DT – Palavras Soltas. Origem: Projeto Língua Pomerana na escola Germano Hübner. Livro Impresso e também em documento de texto. Autor: Danilo Kuhn Silva (Organizador) Produção coletiva entre professores e alunos na escola, 2012. São Lourenço do Sul/RS.
89. PS-LP-ZRN-IT-MG-MP – Palavras Soltas. Anotações de LP. Durante a entrevista na zona rural norte de Itueta/MG recebemos a doação de um pequeno manuscrito em papel.
90. PS-T-UW-SLS-DT – Palavras Soltas. Manual Pomerano. Vocabulário de pomeranos escrito por eles na zona rural de Canguçu/RS, organizado por Ursula Wiesemann em 2008. Doado por um pastor pomerano luterano em folhas impressas.
91. D-FK-VN-PKE-CIP – Diário – Doado por Família pomerana KK em visita de campo na zona rural norte de IT/MG. Não constituiu parte dos *corpora* escritos.

APÊNDICE E – Questionáro Semântico-Lexical (QSL) / Lexikalisch-Semantischer Fragebogen (LSF)

ACIDENTES GEOGRÁFICOS / GEOGRAFISCHE GEGEBENHEITEN

1. CÓRREGO / RIACHO / 1. WASSERLAUF / BACH
...um rio pequeno de uns dois metros de largura? / Ein kleiner Bach ... etwa zwei Meter breit? (Wie nennt man das? / Wie heißt man das?)

2. FOZ / 2. MÜNDUNG
...o lugar onde o rio termina ou encontra com outro rio? / ... Der Ort, wo der Fluss endet oder in einen anderen Fluss fließt?

3. REDEMOINHO (DE ÁGUA) / 3. WASSERSTRUDEL
...muitas vezes, num rio, a água começa a girar, formando um buraco, na água, que puxa para baixo. Como se chama isto? / ... Oft beginnt sich ein Fluss zu drehen, wodurch ein Strudel entsteht, der alles nach unten zieht. Wie nennt man das?

4. ONDA DE RIO / CORRENTEZA / 4. FLUSSWELLE / FLUSSSTRÖMUNG
...o movimento da água do rio? *Imitar o balanço das águas.* / Bewegung ... das Wasser des Flusses? Imitieren der Welle des Wassers.

FENÔMENOS ATMOSFÉRICOS / ATMOSPHÄRISCHE ERSCHEINUNGEN

5. REDEMOINHO (DO VENTO) / 5. WIRBELWIND
...o vento que vai girando em roda e levanta poeira, folhas e outras coisas leves? / Der Wind beginnt sich zu drehen und bildet einen Wirbel, der Blätter, Staub und leichte Gegenstände mit sich reißt

6. RELÂMPAGO / RAIO / 6. BLITZ
...um clarão que surge no céu em dias de chuva? / Ein starkes Licht, das am Himmel an Regentagen entsteht?

7. TROVÃO / 7.DONNER
...barulho forte que se escuta logo depois de um _____ (cf. item 7)? //...Lautes Geräusch, das man direkt nach einem _____ (siehe Punkt 7) hören kann?

8. TEMPORAL / TEMPESTADE / VENDAVAL / 8. STURM
...uma chuva com vento forte que vem de repente? / Starker Regen mit starkem Wind, der plötzlich kommt?

9. NOMES ESPECÍFICOS PARA TEMPORAL / 9. BESONDERE NAMEN FÜR STÜRME
...Existem outros nomes para _____ (cf. item 9)? / Es gibt noch andere Namen für ... _____ (vgl. Artikel 9)?

10. TROMBA D'ÁGUA / 10. STURZREGEN
...uma chuva de pouca duração, muito forte e pesada? / Kurzer aber sehr starker Regenguss (Wie nennt man das?)

11. CHUVA FORTE / 11. SCHWERER REGEN
...uma chuva forte e contínua? / Schwerer Regen ... und ständig?

12. CHUVA DE PEDRA / 12. HAGEL
...Durante uma chuva, podem cair bolinhas de gelo. Como chamam essa chuva? / Während eines Regens können Eiskörner fallen, die man Hagel nennt

13. ESTIAR / COMPOR O TEMPO / 13. ES KLÄRT AUF/ DAS WETTER FESTIGT SICH
Como dizem aqui quando termina a chuva e o sol começa a aparecer? / Wie sagt man hier, wenn der Regen aufhört und die Sonne zu scheinen beginnt?

14.ARCO-ÍRIS / 14. REGENBOGEN
Quase sempre, depois de uma chuva, aparece no céu uma faixa com listras coloridas e curvas (mímica). Que nomes dão a essa faixa? / Fast immer, nach einem Regen, erscheint ein farbiger Bogen am Himmel mit bunten Streifen und Kurven (Mime). Welche Namen gibt man diesem Bogen.

15. GAROA / 15. NIESELREGEN
... uma chuva bem fininha? / Ein Regen ... sehr schwach?

16. TERRA UMEDECIDA PELA CHUVA / 16. REGENFEUCHTER BODEN
Depois de uma chuva bem fininha, quando a terra não fica nem seca, nem molhada, como é que se diz que a terra fica? / Nach einem sehr dünnen Regen, wenn der Boden noch nicht trocken, aber auch nicht richtig ist,, wie sagt man da, dass die Erde ist?

17. ORVALHO / SERENO / 17. TAU / MORGENFRISCHE
De manhã cedo, a grama geralmente está molhada. Como chamam aquilo que molha a grama? / Am frühen Morgen ist das Gras in der Regel nass. Wie nennt man das, was das Gras nass macht?

18. NEVOEIRO / CERRAÇÃO / NEBLINA / 18. NEBEL
Muitas vezes, principalmente de manhã cedo, quase não se pode enxergar por causa de uma coisa parecida com fumaça, que cobre tudo. Como chamam isso? / Oft kann man, vor allem in den frühen Morgenstunden, kaum etwas sehen, weil alles weiß verschwommen ist, wie im Rauch, der alles abdeckt. Wie nennt man das?

ASTROS E TEMPO / STERNE UND ZEIT

19. AMANHECER / 19. TAGESBEGINN
...a parte do dia quando começa a clarear? / Teil des Tages an dem es anfängt hell zu werden?

20. NASCER (DO SOL) / 20. SONNENAUFGANG
O que é que acontece no céu de manhã cedo quando começa a clarear? / Was passiert am Himmel in den frühen Morgenstunden, wenn es beginnt hell zu werden?

21. POR DO SOL / 21. SONNENUNTERGANG
E o que acontece no céu no final da tarde? / Und was passiert am Himmel am späten Nachmittag?

22. ANOITECER / 22. ABEND
...o começo da noite? / Der Beginn der Nacht ...?

23. ESTRELA MATUTINA / VÊNUS / ESTRELA DA MANHÃ / ESTRELA D'ALVA / 23. MORGENSTERN
De manhã cedo, uma estrela brilha mais e é a última a desaparecer. Como chamam esta estrela? / Am frühen Morgen ..., glänzt ein Stern, der als letzter verschwindet. Wie nennt man diesen Stern?

24. ESTRELA VESPERTINA / VÊNUS / ESTRELA DA TARDE / 24. ABENDSTERN / VENUS
De tardezinha, uma estrela aparece antes das outras, perto do horizonte, e brilha mais. Como chamam esta estrela? / Am späten Nachmittag, erscheint ein Stern vor den anderen, in der Nähe des Horizonts und glänzt. Wie nennt man diesen Stern?

25. MESES DO ANO / 25. MONATE DES JAHRES

Quais são os meses do ano? / Was sind die Monate des Jahres?

26. DIAS DA SEMANA / 26. TAGE DER WOCHE

27. CONTAR DE 1 A 10 / 27. ZÄHLEN VON 1 BIS 10

28. ONTEM / 28. GESTERN
Hoje é segunda-feira. E Domingo que dia foi? / Heute ist Montag. Wann war Sonntag?

29. ANTEONTEM / 29. VORGESTERN
...o dia que foi antes desse dia? (E um dia pra trás). / Der Tag vor diesem Tag? (Und noch ein Tages davor).

ATIVIDADES AGROPASTORIS / LAND- UND VIEHWIRTSCHAFTLICHE TÄTIGKEITEN

30. TANGERINA/ MEXERICA / 30. MANDARINE / TANGERINE
...as frutas menores que a laranja, que se descascam com a mão, e, normalmente, deixam um cheiro na mão? Como elas são? / Die Früchte ... kleiner als Orangen, die mit der Hand geschält werden, und in der Regel einen Duft auf der Hand lassen? Wie heißen sie?

31. AMENDOIM / 31. ERDNÜSSE
...o grão coberto por uma casquinha dura, que se come assado, cozido, torrado, ou moído? / Ein Korn mit einer harten Schale, das man gebacken, geröstet, gekocht oder gemahlen essen kann?

32. CAMOMILA / 32. KAMILLE
...umas florezinhas brancas com miolo amarelinho, ou florezinhas secas que se compram na farmácia ou no supermercado e servem para fazer um chá amarelinho, cheiroso, bom para dor de barriga de nenê/bebê e até de adulto e também para acalmar? *Mostrar.*
Ein weißes Blümchen mit gelber Mitte, oder kleine getrocknete Blumen, die man in der Apotheke oder im Supermarkt kaufen kann und mit denen man Tee machen kann, gut gegen Bauchschmerzen bei Kleinkindern und auch Erwachsenen. Dient auch zur Beruhigung? Zeigen.

33. PENCA DE BANANA e CACHO DE BANANA / 33. BANANENBÜNDEL und BANANENSTAUDE
...cada parte que se corta do cacho da bananeira para pôr para madurar/amadurecer e a bananeira inteira? / ...Teil, der Bananenstaude, den man abschneidet und reifen lässt? Und die ganze Staude von Bananen...?

34. BANANA DUPLA / FELIPE /GÊMEAS/ 34.DOPPELBANANE/ZWILLINGSBANANE
...duas bananas que nascem grudadas? / Zwei Bananen ... die zusammengewachsen sind?

35. PARTE TERMINAL DA INFLORESCENCIA DA BANANEIRA / UMBIGO / CORAÇÃO / 35. ENDTEIL DER BANANENBLÜTE / NABEL / HERZ
...a ponta roxa no cacho da banana? / ... die lilafarbene Spitze der Bananestaude?

36. ESPIGA / 36. KOLBEN
Quando se vai colher o milho, o que é que se tira o pé? (Quando se vai à feira comprar milho, compra-se o quê?) / Wenn der Mais geernet wird, was erntet man da? (Wenn Sie auf den Markt gehen, um Mais zu kaufen, was kaufen?)

37. SABUGO / 37. STRUNK
Quando se tira da _____ (cf. item 36) todos os grãos do milho, o que sobra? / Wenn man die Maiskörner abzieht, was bleibt da über? _____ nehmen (vgl. Artikel 45) Alle Maiskörner, was ist links?

38. GIRASSOL / 38. SONNENBLUMEN

...flor grande, amarela, redonda, com uma rodela de sementes no meio? / Blume ... gelb, rund, mit einer Scheibe mit Samen in der Mitte?

39. VAGEM DO FEIJÃO / BAINHA / 39. STANGENBOHNE

Onde é que ficam os grãos do feijão, no pé, antes de serem colhidos? / Wo sind die Bohnen, bevor wir sie ernten?

40. MANDIOCA / AIPIM / 40. MANIOK
...aquela raiz branca por dentro, coberta por uma casca marrom, que se cozinha / Die Wurzel, die innen weiss ist, mit einer braunen Schale und die man kochen kann.

41. CARRINHO DE MÃO / CARRIOLA / 41. SCHUBKARRE / WÄGELCHEN
...um veículo de uma roda, empurrado por uma pessoa, para pequenas cargas em trechos curtos? / Ein Fahrzeug mit einem Rad, geschoben von einer Person, für kleine Lasten auf kurzen Strecken?

42. CANGALHA / FORQUILHA / 42. JOCH / GABEL
...a armação de madeira, que se coloca no pescoço de animais (porco, terneiro, bezerro, carneiro, vaca) para não atravessarem a cerca? / ... Der Holzrahmen, der den Tieren über den Hals gelegt wird (Schwein, Kalb, Kalb, Schaf, Kuh) damit sie nicht durch den Zaun brechen?

43. CANGA / 43. JOCH
...a peça de madeira que vai no pescoço do boi, para puxar o carro ou arado? *Mostrar*

gravura. / Die ... Stück Holz, das über den Hals des Ochsen gelegt wird, wenn er den Pflug, oder einen Wagen ziehen soll? Bild zeigen.

44. JACÁ / BALAIO / 44. KORB
...aqueles objetos de vime, de taquara, de cipós trançados, para levar batatas (mandioca, macaxeira, aipim, etc) no lombo do cavalo ou do burro? / Körbe aus Weide oder Bambus um Kartoffeln (Maniok, Cassava, etc.) auf dem Rücken des Pferdes oder Esel zu transportieren?

45. BORREGO / CORDEIRO (Do nascer até...) / 45. LAMM / LAMM (Von der Geburt bis ...)
...a cria da ovelha logo que nasce? E até que idade se dá esse nome? / Neugeborene Schafe? Und bis zu welchem Alter wird dieser Name gegeben?

46. TRABALHADOR DE ENXADA EM ROÇA ALHEIA / LAVRADOR / DIARISTA / 46. Landarbeiter auf fremdem Feld / TAGELÖHNER
...o homem que é contratado para trabalhar na roça de outro, que recebe por dia de trabalho? / Mann ... der angestellt wird, um im Feld eines anderen zu arbeiten Er wird nach Arbeitstagen entlohnt?

47. PICADA / ATALHO ESTREITO/ 47. WALDWEG / SCHNEISSE
O que é que se abre com o facão, a foice para passar por um mato fechado? / Weg, der mit der Machete (Buschmesser) oder mit der Sense in die Wildnis geschlagen wird?

48. TRILHO / CAMINHO / VEREDA / TRILHA / 48. PFAD / TRAMPELPFAD / WEG
...o caminho, no pasto, onde não cresce mais grama, de tanto o animal ou o homem passarem por ali? / Wege auf der Weide, wo kein Gras mehr wächst, weil ständig Tiere und Menschen darüber gehen

FAUNA / FAUNA

49. URUBU / 49. AASGEIER
...a ave preta que come animal morto, podre? / ... der Schwarze Vogel, der tote Tiere und Faules frisst?

50. COLIBRI / BEIJA-FLOR / 50. KOLIBRI
...o passarinho bem pequeno, que bate muito rápido as asas, tem o bico comprido e fica parado no ar? / Der kleine Vogel, der sehr schnell mit den Flügel schlägt, einen langen Schnabel hat und in der Luft stehen bleiben kann?

51. JOÃO-DE-BARRO / 51. ROSTTÖPFER
...a ave que faz a casa com terra, nos postes, nas árvores e até nos cantos da casa / Der Vogel ... der aus Lehm sein Haus auf Bäumen und Masten und sogar an Hausecken macht.

52. GALINHA D'ANGOLA / GUINÉ / COCAR / 52. ANGOLA-HUHN / GUINEA-HUHN
...a ave de criação parecida com a galinha, de penas pretas com pintinhas brancas? / Ein Huhn, so ähnlich wie die normalen Hühner, nur mit schwarzen Federn und weißen Flecken?

53. PAPAGAIO / 53. PAPAGEI
...a ave de penas coloridas que, quando presa, pode aprender a falar? / Der Vogel mit den bunten Federn, der, wenn er gefangen wird, sprechen lernen kann

54. SURA / 54. SURA
...uma galinha sem rabo? / Ein Huhn ohne Schwanz ...?

55. COTÓ / PITOCO / 55. STUMPF
...um cachorro de rabo cortado? / Ein Hund mit abgeschnittenem Schwanz?

56. GAMBÁ / 56. STINKTIER
...um bicho que solta um cheiro ruim quando se sente ameaçado? / Ein Tier, das einen schlechten Geruch abgibt, wenn es sich bedroht fühlt?

57. CRINA (DO PESCOÇO) / 57. MÄHNE (AM HALS)
...o cabelo em cima do pescoço do cavalo? / Die Haare auf dem Hals des Pferdes?

58. CRINA DE CAUDA / RABO DE CAVALHO / 58. PFERDESCHWANZ
...o cabelo comprido na traseira do cavalo? / Die langen Haare des Schwanzes des Pferdes?

59. LOMBO / DORSO / 59. RÜCKEN / KUPPE
...a parte do cavalo onde vai a sela? / Teil des Pferdes worauf der Sattel liegt?

60. ANCA / GARUPA / CADEIRA / COLA / 60. RÜCKEN / KUPPE
... a parte larga atrás do _____ (cf. item 59)? / ... der breite Teil hinter dem _____ (vgl. Artikel 59)?

61. CHIFRE / 61. HORN
O que o boi tem na cabeça? / Was der Stier auf der Stirn hat?

62. ÚBERE / UBRE / 62. EUTER
Em que parte da vaca fica o leite? / Das Teil, wo die Kuh gemolken wird?

63. RABO / 63. SCHWANZ
...a parte com que o boi espanta as moscas? / Teil mit dem der Ochse die Mücken verscheucht?

64. MANCO / 64. LAHM
...o animal que tem uma perna mais curta e que puxa de uma perna? / Tiere die ein Bein nachziehen?

65. MOSCA VAREJEIRA / 65. SCHMEIßFLIGE
... um tipo de mosca grande, esverdeada, que faz um barulho quando voa? Eine große grüne Mücke, die laute Geräusche macht, wenn sie fliegt

66. SANGUESSUGA / 66. STECHMÜCKE
...um bichinho que se gruda nas pernas das pessoas quando elas entram num córrego ou banhado (cf. item 1)? / Eine kleine Mücke, die oft an Armen und Beinen klebt und sticht, wenn man in der Nähe vom Wasser ist. (vgl. Punkt 1) ?

67. LIBÉLULA / 67. LIBELLE
...o inseto de corpo comprido e fino, com quatro asas bem transparentes, que voa e bate a parte traseira na água? / Ein schlankes Insekt mit vier durchsichtigen Flügeln das, wenn es über das Wasser fliegt, mit dem Hinterteil das Wasser berührt?

68. BICHO DE FRUTA / 68. OBSTMADEN
...aquele bichinho branco, enrrugadinho, que dá em goiaba, em coco? / Kleine weisse Tiere in den Früchten (Guaven und Kokos)?

69. CORÓ / BICHO DE PAU-PODRE / 69. HOLZWÜRMER
...aquele bicho que dá em esterco, em pau-podre? Kleine Tiere die im faulen Holz und in der Gülle leben?

70. PERNILONGO / 70. STELZENLÄUFER
...aquele inseto pequeno, de perninhas compridas, que canta no ouvido das pessoas, de noite? Imitar o zumbido. / Kleine Insekten mit langen Beinen, die in der Nacht um unsere Ohren fliegen.

CORPO HUMANO / DER MENSCHLICHE KÖRPER

71. PÁLPEBRA / CAPELA DOS OLHOS / 71. LID
...esta parte que cobre o olho? *Mostrar*. / Das Teil, das die Augen abdeckt? Zeigen.

72. CISCO / 72. SPLITTER IM AUGE
...alguma coisinha que cai no olho e fica incomodando? Etwas, was in unsere Augen fällt und sehr stört?

73. CEGO DE UM OLHO / CAOLHO / 73. BLIND AUF EINEM AUGE
...a pessoa que só enxerga com um olho? /...Die Person, die nur mit einem Auge sieht?

74. VESGO / 74. SCHIELEN
...a pessoa que tem os olhos voltados para direções diferentes? Completar com um gesto dos dedos. / Wenn die Augen in unterschiedliche Richtungen sehen? Zeigen mit einer Geste der Finger.

75. TERÇOL / VIÚVA / BONITINHA / 75. LEBERFLECK
...a bolinha que nasce na _____ (cf. item 89.), fica vermelha e incha? / (p*opularmente as pessoas falam que as grávidas é que "jogam" no olho da pessoa) /* Der Ball ... in _____ (vgl. Artikel 89). Bei der Geburt ist er rot und schwillt an? (Im Volksmund sagen die Leute, dass schwangere Frauen, es mit dem bösen Blick auf Menschen werfen)

76. CONJUTIVITE / DOR NOS OLHOS / 76. BINDEHAUTENTZÜNDUNG / SCHMERZ IN DEN AUGEN
...a inflamação no olho que faz com que o olho fique vermelho e amanheça grudado? / ... Entzündung im Auge, die das Auge rot und morgens verklebt macht.

77. CATARATA / 77. STAR
...aquela pele branca no olho que dá em pessoa mais idosa? / Das ... weiße Häutchen im Auge, dass bei älteren Menschen auftritt?

78. DENTES CANINOS / PRESAS / 78. ECKZÄHNE
...esses dois dentes pontudos? *Apontar*. / Diese beiden spitzen Zähnen ...? zeigen

79. DENTES DO SISO / DO JUÍZO / 79. WEISHEITSZÄHNE

...os últimos dentes, que nascem depois de todos os outros, em geral quando a pessoa já é adulta? / ... Die letzten Zähne, die nach all den anderen kommen, in der Regel, wenn die Person bereits erwachsen ist?

80. DESDENTADO / BANGUELA / 80. ZAHNLOS
...a pessoa que não tem dentes? / Die Person ... die keine Zähne mehr hat?

81. FANHOSO / FANHO / 81. NÄSELN
...a pessoa que parece falar pelo nariz? *Imitar.* / Die Person ... die anscheinend durch die Nase spricht? Imitieren.

82. MELECA / TATU / 82. ROTZ
...a sujeirinha dura que se tirado nariz com o dedo? / Harter Schmutz ... der mit dem Finger aus der Nase geholt wird?

83. SOLUÇO / 83. SCHLUCKAUF
...este barulhinho que se faz? *Soluçar.* / Dieses kleine Geräusch? Schluckauf imitieren.

84. NUCA / 84. NACKEN
...isto? *Apontar.* / hier..... zeigen

85. POMO DE ADÃO / GOGÓ / 85. ADAMSAPFEL
...esta parte alta do pescoço do homem? *Apontar.* Der spitze Teil des Halses beim Mann? zeigen

86. CLAVÍCULA / 86. SCHLÜSSELBEIN
...o osso que vai do pescoço até o ombro? *Apontar.* / ... Der Knochen, der vom Hals bis zur Schulter läuft? Zeigen.

87. CORCUNDA / 87. BUCKEL
...a pessoa que tem um calombo grande nas costas e fica assim (*mímica*)? / Die Person, die ... eine große Beule auf dem Rücken hat und so aussieht (imitieren)?

88. AXILA / 88. ACHSEL
...esta parte aqui? *Apontar.* / Dieser Teil hier ...? Zeigen.

89. CHEIRO NA AXILA / 89. GERUCH IN DEN ACHSELHÖHLEN
...o mau cheiro embaixo dos braços? / Der Gestank ... unter den Armen?

90. CANHOTO / 90. LINKHÄNDER
...a pessoa que come com a mão esquerda, faz tudo com essa mão? *Gesticular.* / ... Die Person, die mit der linken Hand isst, und alles mit dieser Hand tut? Gestikulieren.

91. VOMITAR / 91. ERBRECHEN
...Se uma pessoa come muito e sente que vai pôr/botar para fora o que comeu, se diz que vai o quê? / Wenn ... ein Mensch zu viel isst und das Gefühl hat, dass er alles wieder ausspucken muss, was er gegessen hat. Was sagt man dazu?

92. ÚTERO / BARRIGA / 92. GEBÄHRMUTTER / BAUCH
...a parte do corpo da mãe onde fica o nenê / bebê antes de nascer? / Der Teil des Körpers, wo die Mutter vor der Geburt das Kind trägt?

93. PERNETA / 93. EINBEIN
...a pessoa que não tem uma perna? /...Die Person, die nur ein Bein hat?

94. MANCO / 94. LAHMER
...a pessoa que puxa de uma perna? / arrasta uma das pernas? /... Die Person, die ein Bein nachzieht?

95. RÓTULA / PATACA / 95. KNIESCHEIBE
...o osso redondo que fica na frente do joelho? / Der runde Knochen vor dem Knie?

96. TORNOZELO / 96. FESSEL
...isto? *Apontar*. / Diese ...? Zeigen.
(Junto com as perguntas 95 a 97 perguntar o nome de outras partes do corpo, ir apontando e perguntando.).

97. CALCANHAR / 97. FERSE
...isto? *Apontar*. / / Diese ...? Zeigen.

98. CÓCEGAS / 98. KITZELN
Que sente uma criança quando se passa o dedo na sola do pé? *Mímica*. / Was ein Kind fühlt, wenn man mit dem Finger über die Fußsohle streicht? Imitieren.

99. PARTEIRA / 99. HEBAMME
...a mulher que ajuda a criança a nascer? / Die Frau ... die bei der Geburt hilft?

100. DAR À LUZ / 100. GEBÄREN
Chama-se a _____ (cf. item 99) quando a mulher está para _____. / Wie nennt man das, wenn die Frau _____ (vgl. Artikel 99)

101. GÊMEOS / 101. ZWILLINGE
...duas crianças que nasceram no mesmo parto? / Zwei Kinder ... die zusammen auf die Welt kommen?

102. ABORTO / 102. FEHLGEBURT
Quando a mulher grávida perde o filho, se diz que ela teve um _____. / Wenn eine schwangere Frau ihr Kind verliert, wird gesagt, dass sie eine _____ hatte.

103. ABORTAR / 103. FEHLGEBURT HABEN / ABTREIBUNG MACHEN
Quando a mulher fica grávida e, por algum motivo, não chega a ter a criança, se diz que ela vai _____. / Wenn eine Frau schwanger wird, und aus irgendeinem Grund, nicht das Kind gebiert, sagt man, dass sie _____.

104. AMA-DE-LEITE / MÃE-DE-LEITE / 104. AMME
Quando a mãe não tem leite e outra mulher amamenta a criança, como chamam essa mulher? Wenn die Mutter keine Milch hat und eine andere Frau das Kind stillt, wie nennt man diese Frau?

105. FILHO ADOTIVO / POSTIÇO / 105. ADOPTIVKIND

...a criança que não é filho verdadeiro do casal, mas que é criada por ele como se fosse? / Das Kind ... das ist nicht das wahre Kind ist, aber von dem Paar versorgt wird, als wäre es das eigene Kind?

106. FILHO MAIS MOÇO / CAÇULA / 106. JÜNGSTE SOHN / JÜNGSTE
...o filho que nasceu por último? / Sohn ... der als letzter geboren wurde?

107. MENINO / GURI / PIÁ / 107. JUNGE
Criança pequenininha, a gente diz que é bebê. E quando ela tem de 5 a 10 anos, do sexo masculino? / Ein kleines Kind nennen wir Säugling, oder Baby. Und wenn es 5-10 Jahre alt und männlich ist?

108. MENINA / GURIA / 108. MÄDCHEN
E se for do sexo feminino, como se chama? / Und wenn es weiblich ist, wie nennen wir es?

109. MADRASTA / 109. STIEFMUTTER
Quando um homem fica viúvo e casa de novo, o que a segunda mulher é dos filhos que ele já tinha? / Wenn ein Mann Witwer wird und wieder heiratet, was ist dann die zweite Frau für die Kinder, die er schon hatte?

110. FINADO / FALECIDO / 110. VERSTORBENE
Numa conversa, para falar de uma pessoa que já morreu, geralmente as pessoas não a tratam pelo nome que tinha em vida. Como é que se referem a ela? / In einem Gespräch über eine Person, die schon gestorben ist, nennt man sie oft nicht mit ihrem Namen, den sie hatte, als sie noch lebte. Wie nennt man sie?

CONVÍVIO E COMPORTAMENTO SOCIAL / LEBENSGEMEINSCHAFT UND SOZIALVERHALTEN

111. PESSOA TAGARELA / 111. SCHWÄTZER
... a pessoa que fala demais? /...Die Person, die zu viel redet?

112. PESSOA POUCO INTELIGENTE / 112. DUMMKOPF (vielleicht entfernen)
...pessoa que tem dificuldade de aprender as coisas? / Person ..., die Schwierigkeiten hat, Dinge zu lernen?

113. PESSOA SOVINA / MÃO-DE-VACA / UNHA-DE-FOME / 113. GEIZIGE PERSON / GEIZHALS
...a pessoa que não gosta de gastar seu dinheiro e, às vezes, passa até dificuldades para não gastar? / Die Person ... die nicht gerne ihr Geld ausgibt und manchmal sogar in Schwierigkeiten kommt, nur weil sie nichts ausgeben will?

114. MAU PAGADOR / CALOTEIRO / 114. SCHLECHTER ZAHLER / PRELLER
...a pessoa que deixa de pagar suas contas penduradas? / ... Die Person, die ihre offenen Rechnungen nicht bezahlt?

115. BÊBADO / 115. BESOFFENER
Que nome dão a uma pessoa que bebeu demais? / Welchen Namen gibt man einer Person, die zu viel getrunken hat?

RELIGIÃO E CRENÇAS / RELIGION (BEKENNTNIS UND GLAUBEN)

116. DIABO / 116. TEUFEL
Deus está no céu e no inferno está o _____ / Gott ist im Himmel und in der Hölle ist der _____.

117. FANTASMA / 117. GEIST
O que algumas pessoas dizem já ter visto, à noite, em cemitérios ou em casas, que se diz que é de outro mundo? / Was einige Leute glauben, nachts auf Friedhöfen oder in Häuser gesehen zu haben, und was die angeblich aus einer anderen Welt sein soll?

118. FEITIÇO / MACUMBA / 118. ZAUBER / MACUMBA
O que certas pessoas fazem para prejudicar alguém e botam, por exemplo, nas encruzilhadas? / Was einige Leute tun, um jemanden zu schaden und, zum Beispiel, an der Kreuzungen legen?

119. AMOLETO / 119. AMULETT
...o objeto que algumas pessoas usam para dar sorte ou afastar males? / Objekt ... das einige Leute benutzen um Glück anzuziehen oder das Böse abzuwehren?

120. BENZEDEIRA / 120. BESPRECHEN
...uma mulher que tira o mau-olhado com rezas, geralmente com galho de planta? / Eine Frau ... die mit Gebeten und Sprüchen den bösen Blick wegnimmt, oft auch mit Pflanzenzweigen?

121. CURANDEIRO / RAÍZEIRO / 121**.** HEILER
...a pessoa que trata de doenças através de ervas e plantas? / Die Person, die Krankheiten durch Pflanzen und Kräuter behandelt?

122. MEDALHA / 122. MEDAILLE
...a chapinha de metal com um desenho de santo que as pessoas usam geralmente no pescoço, presa numa corrente? Ein Metallblättchen mit einem Heiligenbild, das die Leute an einem Kettchen um den Hals tragen?

123. PRESÉPIO / 123. KRIPPE
No Natal, monta-se um grupo de figuras representando o nascimento do Menino Jesus. Como chamam isso? / An Weihnachten baut man die Figuren auf, die die Geburt des Jesuskindes darstellen. Wie nennt man das?

JOGOS E DIVERSÕES INFANTIS / DIE SPIELE DER KINDER UND KULTUR

124. CAMBALHOTA / 124. PURZELBAUM
...a brincadeira em que se gira o corpo sobre a cabeça e acaba sentado? *Mímica.* /Das Spiel,... wenn du den Körper über den Kopf abrollst und im Sitz landest? Vorspielen

125. BOLINHA DE GUDE / 125. KLICKER
...as coisinhas redondas de vidro com que os meninos gostam de brincar? / Die kleinen runden Kugeln aus Glas mit denen Jungs gerne spielen?

126. ESTILINGUE / SETRA / BODOQUE / 126. SCHLEUDER / PFEIL / SCHLINGE

...o brinquedo feito de uma forquilha e duas tiras de borracha (*mímica*), que os meninos usam para matar passarinho? / Spielzeug aus ... einer Gabel und zwei Gummi-Streifen (Imitieren), die die Jungs benutzen um Vögel zu jagen?

127. PAPAGAIO DE PAPEL / PIPA / 127. DRACHEN
...o brinquedo feito de varetas cobertas de papel que se empina no vento por meio de uma linha? / Spielzeug aus Holzstäbchen, mit Papier bespannt, die man dann an einer Leine im Wind fliegen lassen kann?

128. ESCONDE-ESCONDE / 128. VERSTECK-DICH
...a brincadeira em que uma criança fecha os olhos, enquanto as outras correm para um lugar onde não são vistas e depois essa criança que fechou os olhos vai procurar as outras? / Ein Spiel, bei dem ein Kind seine Augen schließt, während die anderen an einen Ort rennen, wo sie nicht gesehen werden können. Das Kind, das die Augen geschlossen hatte muss dann die anderen Kinder suchen?

129. CABRA-CEGA / COBRA-CEGA / 129. BLINDE KUH
...a brincadeira em que uma criança, com os olhos vendados, tenta pegar as outras? / Ein Spiel, bei dem ein Knd mit verbundenen Augen versucht die andern Kinder zu fangen?

130. PEGA-PEGA / PIQUE-PEGA / 130. FANGSPIEL
...uma brincadeira em que uma criança corre atrás das outras para tocar numa delas antes que alcance um ponto combinado? / Ein Spiel, wo ein Kind hinter den anderen herrennt um es zu berühren, bevor das andere Kind ein bestimmtes Ziel erreicht?

131. GANGORRA / 131. WIPPE / WIPPSCHAUKEL
...uma tábua apoiada no meio, em cujas pontas sentam duas crianças e quando uma sobe, a outra desce? *Mímica*. / ... Ein Brett, in der Mitte abgestützt, und an dessen Enden jeweils ein Kind sitzt und die dann auf und ab schaukeln? Imitieren

132. BALANÇO / 132. SCHAUKEL
...uma tábua, pendurada por meio de cordas, onde uma criança se senta e se move para frente e para trás? *Mímica*. / ... Ein Brett, an Seilen aufgehängt, worauf ein Kind sitzt und sich vorwärts und rückwärts schaukelt? Imitieren

133. AMARELINHA / 133. HIMMEL UND HÖLLE
...a brincadeira em que as crianças riscam uma figura o chão, formada por quadrados numerados, jogam uma pedrinha (mímica) e vão pulando com uma perna só? (Solicitar descrição detalhada). / Ein Spiel, wo Kinder auf dem Boden eine Figur mit quadratischen Feldern zeichnen, in die dann Steine geworfen werden. Dann muss mit einem Bein auf die Felder gesprungen werden. (Imitieren)? (Detaillierte Beschreibung Ask).

HABITAÇÃO / BEHAUSUNG

134. TRAMELA / 134. RIEGEL
...aquela pecinha de madeira, que gira ao redor de um prego, para fechar a porta, a janela? / Dieses kleine Stück Holz, das sich um einen Nagel dreht, um die Tür, das Fenster zu schließen?

135. FULIGEM / 135. RUSS

...aquilo, preto, que se forma na chaminé, na parede ou no teto da cozinha, acima do fogão a lenha? / Das Schwarze am Schonstein, an der Wand und der Decke über dem Herd?

136. INTERRUPTOR DE LUZ / 136. LICHTSCHALTER
Como se chama o objeto que fica nas paredes e serve para acender a lâmpada? / Wie nennt man das Objekt, das an den Wänden ist und dazu dient, die Lampe einzuschalten?

(Neste momento pedir para descrever outras partes da casa, apontando para elas, por exemplo: casa, parede, teto, janela, telhado, porta, chão).

ALIMENTAÇÃO E COZINHA / ERNÄHRUNG UND KÜCHE

137. CAFÉ DA MANHÃ / 137. FRÜHSTÜCK
...a primeira refeição do dia, feita pela manhã? /...Die erste Mahlzeit des Tages, morgens?

138. GELÉIA / 138. GELEE
... uma pasta feita de frutas para passar no pão, no biscoito? /...Eine Paste aus Früchten hergestellt, die man aufs Brot oder Plätzchen schmiert?

139. CARNE MOÍDA / 139. HACKFLEISCH
...a carne depois de triturada na máquina? / Fleisch in der Maschine gemahlen? /

140. CURAU / CANJICA / 140. MAISBREI
...uma papa cremosa feita com coco e milho verde ralado, polvilhada com canela? / Ein Brei aus Mais und Kokosnuss gemacht und mit Zimt bestreut?

141. ÁGUA ARDENTE / PINGA / 141. SCHNAPS
...a bebida alcoólica feita de cana-de-açúcar? /...Alkoholisches Getränk aus Zuckerrohr?

142. BODEGA / BAR / BOTECO / 142. BAR
...um lugar pequeno, com um balcão, onde os homens costumam ir beber aguardente / pinga e onde também se pode comprar alguma outra coisa? / Ein kleiner Raum mit einer Theke, wohin Männer in der Regel gehen um Schnaps zu trinken und wo man auch anderes kaufen kann? /

143. EMPANTURRADO / CHEIO / ESTUFADO / SATISFEITO / 143. ÜBERSATT, VOLL, ZUFRIEDEN
Quando uma pessoa acha que comeu demais, ela diz: Comi tanto que estou ____. / Wenn eine Person denkt, dass Sie zu viel gegessen haben, sagt sie: Ich hab so viel gegessen, dass ich ____.

144. GLUTÃO / GULOSA / 144. VÖLLER
...uma pessoa que normalmente come demais? / Eine Person, die ... in der Regel zu viel isst?

145. BALA / CONFEITO / BOMBOM / DOCE / PIRULITO / 145. BONBON / KEKSE / PRALINEN / SÜßIGKEITEN / LUTSCHER
...aquilo embrulhado em papel colorido que se chupa ou se come e as crianças adoram? *Mostrar.* PEDIR PARA DESCREVER. / Was ... in buntem Papier eingewickelt ist und was die Kinder gerne lutschen oder essen? Zeigen. Um zu beschreiben.

146. PÃO FRANCES / 146. FRANZÖSISCHES BROT
...isto? *Mostrar*. / Diese ...? Anzeigen.

ACESSÓRIOS / ZUBEHÖR

147. GRAMPO (DE CABELO) /RAMONA/MISSE/ 147. HAARKLAMMER/ HAARSPANGE / MISSE
...um objeto fino de metal, para prender o cabelo? *Mostrar*. / Ein Objekt aus dünnem Metall, um das Haar zu halten? Zeigen.

148. DIADEMA / ARCO / TIARA / 148. DIADEM / HAARBOGEN / KRONE
...o objeto de metal ou plástico que pega de um lado ao outro da cabeça e serve para prender os cabelos? *Mímica*. / Ein Objekt aus Metall oder Kunststoff, das dazu dient das Haar zu halten? Mime.

APÊNDICE F – Questionário Sociolinguístico – QS

Questionário Sociolinguístico dos Pontos de Pesquisa
Data da entrevista: ___/___/_____ Horário de início: ____:____Término: ___:___
Sujeito-entrevistador: _____ Localidade: _____

Orientações ao sujeito-entrevistador

Fazer contato prévio.

Antes de começar a entrevista esclarecer o sujeito sobre a pesquisa.

Os comentários depois de 2 barras (//) são exclusivamente para o entrevistador, não havendo necessidade de serem lidos para o entrevistado.

Este questionário servirá de base para fazermos um levantamento sócio-histórico, cultural e linguístico da comunidade pomerana.

Questionário

1. Gênero

(a) Masculino (b) Feminino

2. Data de nascimento: _____/_____/_____ (idade _____)

3. Você é?

(1) Pomerano (2) Alemão (3) brasileiro (4) descendente de pomerano (5) descendente de alemão (6) Mestiço (7) Não-descendente (8) nenhuma das alternativas (9) Outra etnia: _____

4. Qual a sua nacionalidade:

(1) Alemã (2) Brasileira (3) dupla nacionalidade (4) naturalizado brasileiro (5) outra: _____

5. Onde você nasceu?

(1) País_____ Cidade: _____ Bairro: _____ UF:

_____

(2) Por gentileza, fale um pouco sobre o seu local de nascimento, de infância... Você sempre morou aqui?

(3) Você conhece a história desta localidade (cidade/comunidade?)

6. Complete o quadro

Lugar de nascimento	Período de permanência:
Zona rural	
Zona urbana	
Residência atual	

7. Qual seu estado civil?

(1) Solteiro(a) (2) Casado(a) (3) Divorciado(a) (4) Viúvo(a) (5) Outros _____

A – Qual é/era a descendência do cônjuge?
(1) Pomerano (2) Alemão (3) brasileiro (4) descendente de pomerano (5) descendente de alemão (6) Mestiço (7) Não-descendente (8) nenhuma das alternativas (9) Outra etnia:

_____

8. Qual é a sua profissão? _____

A – Até que idade você morou neste local (nascimento)? (_____anos)

9. (Para nascidos fora do Brasil) Quando você veio ao Brasil? Quantos anos você tinha?

(Em _____ Anos _____)
10. Com quem você veio para o Brasil?

11. Quanto tempo pretendia ficar no Brasil?

12. Onde foi o primeiro lugar em que morou?

Cidade: _____Bairro:_____ UF: _____
13. Antes de vir ao Brasil, estudou a língua portuguesa?

(1) Não (2) Sim (Onde? _____)

14. Estudou português no Brasil?

(1) Não (2) Sim (Idade_____, onde? _____)

A) Por quanto tempo você estudou português? Qual a sua formação?

_____
B) Quais foram os motivos que levaram você a estudar português?

_____
15. Você frequentou a escola no seu país de origem?

(1) Não (2) Sim

Caso a resposta seja "sim":
( ) ensino fundamental de 1ª a 4ª série ( ) ensino médio
( ) ensino fundamental de 5ª a 8ª série ( ) ensino superior _____

16. Você frequentou a escola no Brasil? (a partir deste item é para todos) [78]
(1) Não (2) Sim

Caso a resposta seja "sim":
( ) ensino fundamental de 1ª a 4ª série ( ) ensino médio
( ) ensino fundamental de 5ª a 8ª série ( ) ensino superior _____
17. Havia pomeranos no lugar onde nasceu?
(1) Sim (2) Não
18. (1) Sim (2) Não
A) Quantas famílias pomeranas moravam nesse lugar?
(_____ famílias)
B) Mantinha contato com brasileiros?
(1) Não (2) Sim (Em que oportunidade? _____)

19. Quando e de onde você veio à esta localidade? Por quê?

Ano: _____ Bairro:_____ Cidade:_____ UF:____
_____
_____
20. Quando você chegou à esta localidade, existia algum tipo de associação (ou comunidade eclesiástica) pomerana? / Se não existia, hoje existe?
(1) Sim (2) Não / (1) Sim (2) Não
A) Você frequenta/frequentou essa associação?
(1) Não (2) Sim (Qual? _____)
21. Quando você veio à esta localidade, você mantinha contato com os brasileiros?
(1) Não (2) Sim (Em que oportunidade? _____)
22. Na sua infância, qual era a língua falada pela sua família?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

23. Você já frequentou alguma escola de língua pomerana / alemã?
(1) Não (2) Sim (Qual? _____)

A) Por quanto tempo você frequentou a escola de língua pomerana / alemã? _____
B) O que você estudou? (disciplinas)
_____
_____
_____

C) Que tipo de atividades culturais eram realizadas na escola?
(1) Danças pomeranas
(2) Teatro

[78] A partir do item 16 é para todos os sujeitos da pesquisa.

(3) Artesanato
(4) Wandschoner
(5) Informações adicionais: _____

24. Com quem você mora atualmente?
(1) Família (2) Sozinho(a) (3) Outros _____
Utilize a legenda para responder os quadros abaixo

25. Com relação às pessoas da sua família, qual a língua que você usa/usava para conversar?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

	tem	não tem	
Marido/esposa	( )	( )	123456 descendente( ) não descendente( )
Filho(a)	( )	( )	123456 descendente( ) não descendente( )
Neto(a)	( )	( )	123456 descendente( ) não descendente( )
Pai	( )	( )	123456 descendente( ) não descendente( )
Mãe	( )	( )	123456 descendente( ) não descendente( )
Genro/nora	( )	( )	123456 descendente( ) não descendente( )
Irmãos(as)	( )	( )	123456 descendente( ) não descendente( )
Avô	( )	( )	123456 descendente( ) não descendente( )
Avó	( )	( )	123456 descendente( ) não descendente( )
Outros	( )	( )	123456 descendente( ) não descendente( )

26. Com relação aos seus familiares, qual a língua que eles usam para conversar com você?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

	tem	não tem	
Marido/esposa	( )	( )	123456 descendente( ) não descendente( )
Filho(a)	( )	( )	123456 descendente( ) não descendente( )
Neto(a)[79]	( )	( )	123456 descendente( ) não descendente( )
Pai	( )	( )	123456 descendente( ) não descendente( )
Mãe	( )	( )	123456 descendente( ) não descendente( )
Genro/nora	( )	( )	123456 descendente( ) não descendente( )

[79] Mais de um filho ou neto acrescentar o número ao lado.

Irmãos(as)	( )	( )	123456 descendente( ) não descendente( )
Avô	( )	( )	123456 descendente( ) não descendente( )
Avó	( )	( )	123456 descendente( ) não descendente( )
Outros	( )	( )	123456 descendente( ) não descendente( )

27. Na sua casa, costuma-se reunir (pais, filhos, avôs, parentes) durante as refeições?
(1) Sim (2) Não
A) Nessa caso qual a língua usada?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

MÍDIA
28. Que tipo de leitura é feita por você?

TIPO	LÍNGUA		
Material impresso	Pomerano ( )	Alemão ( )	Português ( )
Jornal	Pomerano ( )	Alemão ( )	Português ( )
Revista	Pomerano ( )	Alemão ( )	Português ( )
Livro	Pomerano ( )	Alemão ( )	Português ( )
Publicações locais da igreja ou associação	Pomerano ( )	Alemão ( )	Português ( )
Outros: _____	Pomerano ( )	Alemão ( )	Português ( )

29. Com que frequência você lê material impresso em Pomerano/Alemão (Düütsch)?
(1) Leio muito
(2) Leio às vezes
(3) Praticamente não leio
(4) Não leio de forma alguma
(5) Não existe/não conheço.

30. Com que frequência você lê material impresso em Português?
(1) Leio muito
(2) Leio às vezes
(3) Praticamente não leio
(4) Não leio de forma alguma

31. Com que frequência você assiste TV Alemã (TV a cabo) ou programas pomeranos (de Rádio)?
(1) Diariamente
(2) Assisto muito
(3) Praticamente não assisto

(4) Não assisto nada (segue para a questão 33)

32. Você assiste programas (filmes, documentários, novelas, etc) em Pomerano?
(1) Assisto muito
(2) Assisto às vezes
(3) Praticamente não assisto
(4) Não assisto de forma alguma

A) Que tipo de programa pomerano você mais gosta?
(1) Filmes históricos
(2) Novela
(3) Notícias
(4) Documentário
(5) Programas turísticos
(6) Programa de música
(7) Outros:_____

B) Que tipo de atividade pomerana você mais gosta?
(1) Festas
(2) Apresentações de dança
(3) Assistir a apresentação de bandas que cantam em pomerano
(4) Programas de rádio na língua
(5) Artesanato / Wandschoner
(6) Encenação do casamento pomerano
(7) Recitação de poesias em pomerano
(8) Outros: _____

33. Você assiste programa de TV brasileira?
(1) Assisto muito
(2) Assisto às vezes
(3) Praticamente não assisto
(4) Não assisto de forma alguma

34. Você ouve rádio pomerana/alemã ou alguma rádio local com programas em pomerano?
(1) Ouço muito
(2) Ouço às vezes
(3) Praticamente não ouço
(4) Não ouço de forma alguma

35. Com que frequência você acessa a internet? Pesquisa material em pomerano/alemão (Düütsch)?
(1) Muito
(2) Às vezes
(3) Praticamente não
(4) Não acesso de forma alguma

USO DA LINGUAGEM

36. Você usa a língua pomerana no seu trabalho?
(1) Sim (Em que oportunidade? _____)

(2) Não
37. Você participa das reuniões ou eventos das associações pomeranas da sua comunidade?
(1) Sim (2) Não

A) Qual a língua utilizada nessas reuniões?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

38. que língua você usa ao falar com seus amigos pomeranos/alemães?
(1) Somente a língua pomerana
(2) Mais a língua pomerana do que a língua portuguesa
(3) Meio a meio
(4) Mais a língua portuguesa do que a língua pomerana
(5) Somente a língua portuguesa
(6) Alemão
(7) Outra língua. Qual? _____

// Utilize a legenda para completar o quadro

39. Qual a sua fluência em pomerano?
(1) Falo bem
(2) Falo mais ou menos
(3) Falo só um pouco
(4) Não falo nada
(5) Não falo mais nada/esqueci/desaprendi
(6) Falo misturado com o português

Em conversas sobre política e/ou economia	1 2 3 4 5 6
Em conversa do trabalho	1 2 3 4 5 6
Em conversa do cotidiano	1 2 3 4 5 6
Em conversa entre família	1 2 3 4 5 6
Cumprimentos e palavras fáceis	1 2 3 4 5 6

40. Como é sua leitura em Pomerano?
(1) Leio bem
(2) Leio razoavelmente
(3) Leio só um pouco
(4) Leio com auxílio de alguém
(5) Não leio nada em pomerano
(6) Só leio em Português
(7) Não leio nem em português/ Não sou alfabetizado

Jornais e revista	1 2 3 4 5 6 7
Documentos de trabalho	1 2 3 4 5 6 7
Revistas em quadrinhos	1 2 3 4 5 6 7
Cartas de amigos e parentes	1 2 3 4 5 6 7
Diário	1 2 3 4 5 6 7

Anotações simples	1 2 3 4 5 6 7

41. Você escreve em Pomerano?
(1) Escrevo bem
(2) Escrevo razoavelmente
(3) Escrevo só um pouco
(4) Não consigo escrever nada
(5) Não sei escrever nem em português
(6) Desconheço se existe forma escrita dessa língua que falamos
(7) Outra situação: _____

Cartas de amigos e parentes	1 2 3 4 5 6 7
Diário	1 2 3 4 5 6 7
Anotações simples	1 2 3 4 5 6 7

42. Como você compreende o pomerano?
(1) Compreendo bem
(2) Compreendo razoavelmente
(3) Compreendo só um pouco
(4) Não consigo compreender nada
(5) Compreendo um pouco de cada língua (pomerano, alemão e português)
(6) Outra situação: _____

Em conversas sobre política e/ou economia	1 2 3 4 5 6
Em conversa do trabalho	1 2 3 4 5 6
Em conversa do cotidiano	1 2 3 4 5 6
Em conversa entre família	1 2 3 4 5 6
Cumprimentos e palavras fáceis	1 2 3 4 5 6

43. Qual a sua fluência em Português?
(1) Falo bem
(2) Falo mais ou menos
(3) Falo só um pouco
(4) Não falo nada
(5) Nunca aprendi nenhuma palavra ou nunca tive contato
(6) Falo misturado com o alemão
(7) Falo misturado com o pomerano

Em conversas sobre política e/ou economia	1 2 3 4 5 6 7
Em conversa do trabalho	1 2 3 4 5 6 7
Em conversa do cotidiano	1 2 3 4 5 6 7
Em conversa entre família	1 2 3 4 5 6 7
Cumprimentos e palavras fáceis	1 2 3 4 5 6 7

44. Como você compreende a língua portuguesa?
(1) Compreendo bem
(2) Compreendo razoavelmente
(3) Compreendo só um pouco
(4) Não consigo compreender nada
(5) Compreendo um pouco de cada língua (pomerano, alemão e português)
(6) Outra situação: _____

Em noticiário do rádio	1 2 3 4 5 6

Em noticiário da TV	1 2 3 4 5 6
Em conversa do cotidiano	1 2 3 4 5 6
Em conversa entre família	1 2 3 4 5 6
Cumprimentos e palavras fáceis	1 2 3 4 5 6

45. Como é a sua leitura em português?
(1) Leio bem
(2) Leio razoavelmente
(3) Leio só um pouco
(4) Leio com auxílio de alguém
(5) Não leio nada em português
(6) Só leio em pomerano
(7) Só leio alemão
(8) Não leio nem em português/ Não sou alfabetizado

Jornais e revista	1 2 3 4 5 6 7 8
Documentos de trabalho	1 2 3 4 5 6 7 8
Revistas em quadrinhos	1 2 3 4 5 6 7 8
Cartas de amigos e parentes	1 2 3 4 5 6 7 8
Diário	1 2 3 4 5 6 7 8
Anotações simples	1 2 3 4 5 6 7 8

46. Você escreve em português?
(1) Escrevo bem
(2) Escrevo razoavelmente
(3) Escrevo só um pouco
(4) Só escrevo em pomerano
(5) Só escrevo em alemão
(7) Não consigo escrever nada
(8) Não sei escrever em nenhuma língua
(9) Outra situação: _____

Documentos de trabalho	1 2 3 4 5 6 7 8 9
Cartas de amigos e parentes	1 2 3 4 5 6 7 8 9
Diário	1 2 3 4 5 6 7 8 9
Anotações simples	1 2 3 4 5 6 7 8 9

47. Você incentivou para que seu(s) filho(s) aprendesse(m) a língua pomerana?
(1) Sim (2) Não

A) Por quê?
(1) Por motivos profissionais
(2) Porque é possível ter as ultimas novidades da Alemanha e do mundo
(3) Porque quero que se interesse pela cultura pomerana
(4) outro motivo: _____
(4) outras respostas:_____
_____

48. Você acha necessário que as novas gerações de descendentes de pomeranos aprendam a língua pomerana?
(1) Sim

(2) Se houver interesse do pomerano
(3) Não vejo necessidade (Por quê?_____
_____)

49. Você se expressa melhor falando em pomerano, conforme usado no Brasil ou no alemão, ou ainda no Português usado aqui?
Por quê? _____
_____
_____

50. Você já teve a experiência de ir na Alemanha / Pomerânia (o que restou do local de origem[80])?
(1) Sim (2) Não

A) Com qual objetivo você foi à Alemanha?
(1) Turismo (quando _____, onde _____)
(2) Visitar parentes (quando _____, onde _____)
(3) A trabalho/negócios(quando _____, onde _____)
(4) Estudar (quando _____, onde _____)
(5) Outros: _____
_____

B) O seu pomerano / alemão era compreendido na Alemanha?
(1)Sim, suficientemente
(2)Mais ou menos
(3)Só um pouco
(4)Não compreendiam nada

C) Você passou alguma situação constrangedora por causa da diferença da língua falada no Brasil e na Alemanha? Conte a sua experiência?
(1) Sim _____
_____
_____

(2) Não

51. Você considera que "mistura" a língua portuguesa com o pomerano quando está conversando com familiares e/ou com amigos?
(1) Sim. (Em que momento? _____)
(2) Não.
52. Você "mistura" a língua portuguesa com a língua alemã quando está conversando com familiares e/ou com amigos?
(1) Sim. (Em que momento? _____)
(2) Não.

[80] O território onde existia a Pomerânia foi, em sua parte oriental, perdido para a Polônia, território de onde vieram a maioria dos pomeranos. O que resta hoje é a pomerânia ocidental, *Mecklenburg Vorpommern* pertencente à Alemanha, da qual vieram menor número de imigrantes, segundo Granzow (2009).

53. Você “mistura” o pomerano e o alemão quando está conversando com familiares e/ou com amigos?
(1) Sim. (Em que momento? _____)
(2) Não.
54. Você (acredita ou percebe se) mistura português, pomerano e alemão quando está conversando com familiares e/ou com amigos?
(1) Sim. (Em que momento? _____)
(2) Não.
55. Qual a sua opinião em relação à mistura da língua portuguesa com o pomerano?
_____
_____
_____
_____

(Agradecimentos. Esclarecimentos. Dar liberdade se caso o sujeito-entrevistado quiser tecer algum comentário. Verificar se ele permite o uso dos dados na pesquisa e se tem dúvidas sobre a pesquisa).

ANEXOS

ANEXO A - Parecer consubstanciado do CEP - nº.1145346

UFU Comitê de Ética em Pesquisa — UNIVERSIDADE FEDERAL DE UBERLÂNDIA/MG — Plataforma Brasil

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: A VARIEDADE BRASILEIRA DO POMERANO: UM ESTUDO COM FOCO NOS VALES DO RIO DOCE/MG E DO RIO PARDO/RS

Pesquisador: Guilherme Fromm

Área Temática:

Versão: 1

CAAE: 45410015.6.0000.5152

Instituição Proponente: Instituto de Letras e Linguística

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.145.346

Data da Relatoria: 03/07/2015

Apresentação do Projeto:

Segundo o protocolo, a pesquisa que se pretende desenvolver foca como tema o pomerano, "uma variedade linguística germânica falada no Brasil e praticamente extinta no local de origem – Alemanha e Polônia".

Objetivo da Pesquisa:

O protocolo tem como objetivo geral: "realizar um estudo em linguística para verificar a permanência da fala pomerana no Brasil, observando se os traços linguísticos ainda podem ser perceptíveis e utilizados até os dias de hoje".

E como objetivos específicos: "identificar a sobrevivência ou desaparecimento da fala pomerana nas regiões delimitadas pelo projeto; coletar, compilar, organizar e classificar os corpora do pomerano".

Avaliação dos Riscos e Benefícios:

Na lógica argumentativa do protocolo, os riscos são mínimos aos participantes envolvidos com a pesquisa, mas haverá o procedimento previsto na Resolução 466/12.

E os benefícios contribuirão com a descrição e preservação linguística do pomerano. E continua a

Endereço: Av. João Naves de Ávila 2121- Bloco "1A", sala 224 - Campus Sta. Mônica
Bairro: Santa Mônica **CEP:** 38.408-144
UF: MG **Município:** UBERLANDIA
Telefone: (34)3239-4131 **Fax:** (34)3239-4335 **E-mail:** cep@propp.ufu.br

listagem: "a confecção de um Glossário, produto final que será
disponibilizado gratuitamente na Internet e permitirá que outros estudos sejam desenvolvidos".

Comentários e Considerações sobre a Pesquisa:

A proposta metodológica adotada pelo protocolo no resgate do corpus linguístico do pomerano, nas suas manifestações textuais e oral, em áreas distintas: "Vila Neitzel (Itueta) /MG, Itueta/MG, Vera Cruz/RS e Arroio do Tigre/RS". Para a execução do processo investigativo, a entrevista será usada como técnica de coleta de informações. Apresenta o plano de recrutamento dos participantes da pesquisa, com os seus critérios de inclusão e exclusão vinculados aos objetivos da pesquisa. Demonstra a amostra, segundo dois critérios: gênero e localidade geográfica.

Considerações sobre os Termos de apresentação obrigatória:

Foram apresentados.

Recomendações:

Conclusões ou Pendências e Lista de Inadequações:

De acordo com as atribuições definidas na Resolução CNS 466/12, o CEP manifesta-se pela aprovação do protocolo de pesquisa proposto.

O protocolo não apresenta problemas de ética nas condutas de pesquisa com seres humanos, nos limites da redação e da metodologia apresentadas.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Data para entrega de Relatório Final ao CEP/UFU: julho/2016

OBS.: O CEP/UFU LEMBRA QUE QUALQUER MUDANÇA NO PROTOCOLO DEVE SER INFORMADA IMEDIATAMENTE AO CEP PARA FINS DE ANÁLISE E APROVAÇÃO DA MESMA.

O CEP/UFU lembra que:

a- segundo a Resolução 466/12, o pesquisador deverá arquivar por 5 anos o relatório da pesquisa

e os Termos de Consentimento Livre e Esclarecido, assinados pelo sujeito de pesquisa.
b- poderá, por escolha aleatória, visitar o pesquisador para conferência do relatório e documentação pertinente ao projeto.
c- a aprovação do protocolo de pesquisa pelo CEP/UFU dá-se em decorrência do atendimento a Resolução CNS 466/12, não implicando na qualidade científica do mesmo.

Orientações ao pesquisador :
• O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Res. CNS 466/12) e deve receber uma via original do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado.
• O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Res. CNS 466/12), aguardando seu parecer, exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade de regime oferecido a um dos grupos da pesquisa que requeiram ação imediata.
• O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Res. CNS 466/12). É papel de o pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.
• Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. Em caso de projetos do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma, junto com o parecer aprobatório do CEP, para serem juntadas ao protocolo inicial (Res.251/97, item III.2.e).

UBERLANDIA, 10 de Julho de 2015

Assinado por:
Sandra Terezinha de Farias Furtado
(Coordenador)

ANEXO B - Die 40 (endgültigen) Sätze Georg Wenkers (1880)

(Varianten in Schrägschrift, nach '/' oder in ( ))

1. Im Winter fliegen die trockenen Blätter in der Luft herum.
2. Es hört gleich auf zu schneien, dann wird das Wetter wieder besser.
3. Tu Kohlen in den Ofen, damit die Milch bald zu kochen anfängt.
4. Der gute alte Mann ist mit dem Pferd(e) auf dem Eis eingebrochen und in das kalte Wasser gefallen.
5. Er ist vor vier oder sechs Wochen gestorben.
6. Das Feuer war zu heiß, die Kuchen sind ja unten ganz schwarz gebrannt.
7. Er ißt die Eier immer ohne Salz und Pfeffer.
8. Die Füße tun mir (so sehr) weh, ich glaube, ich habe sie (mir) durchgelaufen.
9. Ich bin selber bei der Frau gewesen und habe es ihr gesagt, und sie sagte, sie wolle es auch ihrer Tochter sagen.
10. Ich will es auch nicht mehr wieder tun/machen.
11. Ich schlage dich gleich mit dem Kochlöffel um die Ohren, du Affe.
12. Wo gehst du (denn) hin? Sollen wir mitgehen (mit dir gehen)?
13. Das /es sind schlechte Zeiten.
14. Mein liebes Kind, bleib hier unten stehen, die bösen Gänse beißen dich tot.
15. Du hast heute am meisten gelernt und bist artig gewesen, du darfst früher nach Hause gehen als die anderen.
16. Du bist noch nicht groß genug, um eine Flasche Wein allein auszutrinken, du mußt erst noch wachsen und größer werden.
17. Geh, sei so gut und sag deiner Schwester, sie soll die Kleider für eure Mutter fertig nähen und mit der Bürste rein machen.
18. Hättest du ihn gekannt! Dann wäre es anders gekommen, und es täte besser um ihn stehen.
19. Wer hat mir meinen Korb mit Fleisch gestohlen?
20. Er tat so, als hätten sie ihn zum Dreschen bestellt (; sie haben es aber selbst getan).
21. Wem hat er (denn) die neue Geschichte erzählt?
22. Man muß laut schreien, sonst versteht er uns nicht.
23. Wir sind müde und haben Durst.
24. Als wir gestern abend heim/zurück kamen, da lagen die anderen schon im Bett und waren fest eingeschlafen/am schlafen.
25. Der Schnee ist diese Nacht liegen geblieben, aber heute morgen ist er geschmolzen.
26. Hinter unserem Hause stehen *drei schöne Apfelbäume* /drei Apfelbäumchen mit roten *Äpfeln*/Äpfelchen.
27. Könnt ihr nicht noch *einen Augenblick* /ein Augenblickchen auf uns warten? Dann gehen wir mit (euch).
28. Ihr dürft nicht solche Kindereien treiben.
29. Unsere Berge sind nicht so (sehr) hoch, die euren sind viel höher.
30. Wieviel Pfund Wurst und wieviel Brot wollt ihr haben?
31. Ich verstehe euch nicht, ihr müßt ein bißchen lauter sprechen.
32. Habt ihr kein Stückchen weiße Seife auf meinem Tisch(e) gefunden?
33. Sein Bruder will sich zwei schöne neue Häuser in eurem Garten bauen.
34. Das Wort kam ihm von Herzen.
35. Das war recht von ihnen!
36. Was sitzen da für Vögelchen oben auf dem Mäuerchen?

37. Die Bauern hatten (fünf) Ochsen und (neun) Kühe und (zwölf) Schäfchen vor das Dorf gebracht, die wollten sie verkaufen.
38. Die Leute sind heute alle draußen auf dem Feld(e) und mähen.
39. Geh nur, der braune Hund tut dir nichts.
40. Ich bin mit den Leuten da hinten über die Wiese ins Korn gefahren.

Disponível em : http://staff-www.uni-marburg.de/~naeser/wenker.htm. Acesso em: 11 nov. 2013.

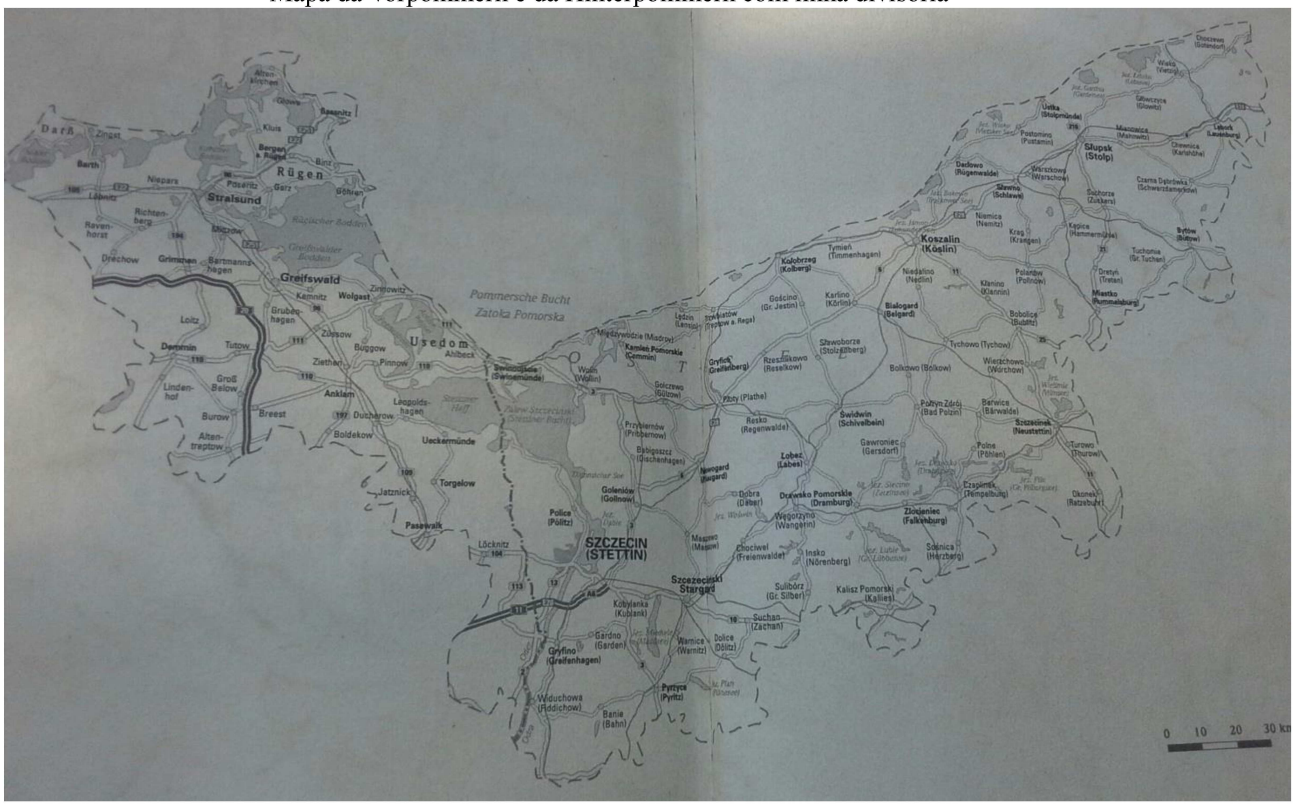
ANEXO C – Mapas da Pomerânia

Mapa da Vorpommern e da Hinterpommern



Fonte: http://stolp.de/landkarten_pommern.html

Mapa da Vorpommern e da Hinterpommern com linha divisória



Fonte: Komet, 2007.