

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS

SANDRA ESPÍNDOLA

**ANÁLISE DE UM *CORPUS* DE PRODUÇÃO ESCRITA EM PORTUGUÊS POR  
CRIANÇAS E ADULTOS INDÍGENAS BILÍNGUES/MONOLÍNGUES DE  
DOURADOS/MS A PARTIR DA LINGUÍSTICA DE CORPUS**

PORTO ALEGRE  
2014

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS  
ÁREA: ESTUDOS DA LINGUAGEM  
LINHA DE PESQUISA: LINGUÍSTICA APLICADA

**ANÁLISE DE UM *CORPUS* DE PRODUÇÃO ESCRITA EM PORTUGUÊS POR  
CRIANÇAS E ADULTOS INDÍGENAS BILÍNGUES/MONOLÍNGUES DE  
DOURADOS/MS A PARTIR DA LINGUÍSTICA DE CORPUS**

SANDRA ESPÍNDOLA

ORIENTADORA: Profª. Dra. INGRID FINGER

COORIENTADORA: Profª. Dra. SIMONE SARMENTO

Tese de Doutorado em Linguística Aplicada,  
apresentado como requisito para obtenção do  
título de Doutor em Letras pelo Programa de  
Pós-Graduação em Letras pela Universidade  
Federal do Rio Grande do Sul

PORTO ALEGRE  
2014

CIP - Catalogação na Publicação

Espíndola, Sandra  
Análise d um corpus de produção escrita em  
português por crianças e adultos indígenas  
bilingües/monolíngües de Dourados/MS / Sandra  
Espíndola. -- 2014.  
167 f.

Orientador: Ingrid Finger.  
Coorientador: Simone Sarmento.

Tese (Doutorado) -- Universidade Federal do Rio  
Grande do Sul, Instituto de Letras, Programa de Pós-  
Graduação em Letras, Porto Alegre, BR-RS, 2014.

1. Ensino de língua. 2. Educação indígena. 3.  
Linguística de corpus. I. Finger, Ingrid, orient. II.  
Sarmiento, Simone, coorient. III. Título.

### **Agradeço...**

Em primeiro lugar, a Deus, por ter guiado meus passos e ter colocado em meu caminho pessoas com as quais eu pude contar nessa difícil caminhada.

À minha família, meu marido Marco Aurélio e meus filhos, Bruna e Odilon pela paciência e tolerância nos momentos de distância e de nervos exacerbados.

À Capes por proporcionar este Dinter, que possibilitou minha qualificação e crescimento profissional.

À Uems na figura das pessoas que não mediram esforços para a realização deste Dinter e acreditaram em seus resultados positivos.

À Ufrgs pela parceria.

À professora Doutora Ingrid Finger por aceitar ser minha orientadora.

À professora Doutora Simone Sarmento por aceitar ser minha coorientadora.

À Larissa Goulart pelo auxílio com o Programa *WordSmith Tools*.

Ao meu grande amigo Odilon Dutra da Silva pela leitura, revisão e formatação deste trabalho.

E, principalmente, ao meu colega de trabalho e amigo Luiz Otávio Batista, que não mediu esforços para me orientar e com suas sugestões me dar ânimo e me fazer acreditar que daria certo nos momentos difíceis em que nem eu acreditava que o trabalho alcançaria o êxito que eu desejava (*In memorian*).

A linguagem, portanto, é uma **nebulosa**, no sentido astronômico do termo, *onde léxico, gramática, discurso e semântica* circulam, colidem, se fundem, fazem surgir novas estrelas e planetas, onde cometas brilham repentinamente para logo desaparecer, nos ininterruptos processos de criação, destruição e recriação de mundos que é nosso universo mental, que só ganha forma e conteúdo nas trocas incessantes com outros universos mentais, com outras nebulosas. Afinal, como escreveu o poeta inglês John Donne (1572-1631), “nenhum homem é uma ilha”, e é precisamente nessa condição inescapável de *arquipélago* que nós nos situamos no oceano da existência em sociedade.

Marcos Bagno (2011)

## RESUMO

Com a finalidade de entender a origem das dificuldades apresentadas por crianças e adultos indígenas na produção de textos em português, surgiu a presente investigação. a partir da Linguística de *Corpus*. Para tanto, foi construído um *corpus* de 483 textos de crianças e 349 textos de adultos escritos em língua portuguesa produzidos por crianças e adultos indígenas e não indígenas. A amostra do grupo das crianças contou um total de 175 crianças, sendo 111 indígenas (71 bilíngues Guarani/Kaiowá e 40 Terena monolíngues) e 64 não indígenas, falantes monolíngues de português, alunos do 4º e do 5º ano do Ensino Fundamental. O grupo de adultos foi formado por um total de 118 adultos, sendo 74 indígenas (36 bilíngues Guarani/Kaiowá e 38 Terena monolíngues) e 44 não indígenas, falantes monolíngues de português, do 1º e do último ano do Ensino Superior. Os objetivos específicos da pesquisa foram: (a) verificar se existem diferenças entre o tipo de dificuldades reveladas pelos indígenas monolíngues e bilíngues de diferentes etnias – Kaiowá/Guarani e Terena – em comparação com os monolíngues não indígenas na produção de textos narrativos em português; (b) na comparação entre os dois grupos etários, crianças e adultos, observar em que medida o caminho percorrido do ensino básico à formação acadêmica interferiu no desenvolvimento da habilidade de escrita de textos; e (c) no caso dos grupos de participantes adultos, investigar se o tempo de permanência no curso de graduação (alunos que estão no primeiro e no quarto ano de curso) interfere no nível de dificuldade na produção de textos. Os dados foram analisados através da ferramenta *AntConc*, a partir do viés teórico da Linguística de *Corpus*. A partir dessa proposta de pesquisa espera-se contribuir para que os professores, tanto os que atendem os acadêmicos quanto os que atendem as crianças, compreendam como a escrita desses dois grupos indígenas se estrutura. Essas informações são essenciais para futuras orientações nos trabalhos de leitura e escritas propostos pela escola e pelos cursos universitários que recebem acadêmicos indígenas.

Palavras-chave: Ensino de Língua. Ensino Indígena. Linguística de Corpus.

## ABSTRACT

In order to understand the origin of the difficulties faced by indigenous children and adults in the production of texts in Portuguese, this research emerged, from Corpus Linguistics. To that end, was built a corpus of 483 children and 349 adults texts of texts written in Portuguese produced by children and indigenous and non-indigenous adults. The sample of children group counted a total of 175 children, with 111 indigenous (71 bilingual Guarani / Kaiowá and Terena 40 monolingual) and 64 non-indigenous, monolingual speakers of Portuguese, students of the 4th and 5th year of elementary school. The adult group consisted of a total of 118 adults, with 74 indigenous (36 bilingual Guarani / Kaiowá and Terena 38 monolingual) and 44 non-indigenous, monolingual speakers of Portuguese, the first and last years of higher education. The specific objectives of the research were: (a) determine whether there are differences between the kinds of problems revealed by monolingual and bilingual indigenous ethnic groups - Kaiowá / Guarani and Terena - compared to non-indigenous monolingual in the production of narrative texts in Portuguese; (b) the comparison between the two age groups, children and adults, to observe to what extent the traveled way of basic education to academic interfered in the development of written texts skill; (c) in the case of adults participating groups, to investigate whether the time spent in the undergraduate course (students who are the first and fourth year of course) interferes with the level of difficulty in producing texts. Data were analyzed by AntConc tool from the theoretical bias of Corpus Linguistics. From this research proposal is expected to contribute to teachers, both those who meet the academic and the attending children, understand how the writing of these two indigenous groups structure. This information is essential for future guidance in reading and written work proposed by schools and university courses receiving indigenous academics.

Keywords: Language Education. Teaching Indigenous. Corpus Linguistics.

## LISTA DE ABREVIATURAS E SIGLAS

RFHB -	Reserva Indígena Francisco Horta Barbosa
RCNEI -	Referenciais Curriculares Para a Escola Indígena
MEC -	Ministério da Educação e do Desporto
Secadi -	Secretaria de Educação Continuada, Alfabetização, Diversidade e Inclusão
CGEEI -	Coordenação Geral de Educação Escolar Indígena
Funai -	Fundação Nacional do Índio
SPI -	Serviço de Proteção ao Índio
Proind -	Programa de Apoio à Formação Superior e Licenciaturas Indígenas
SN -	Sintagma Nominal
UEMS -	Universidade Estadual de Mato Grosso do Sul
UFGD -	Universidade Federal de Mato Grosso do Sul
UFMG -	Universidade Federal de Minas Gerais
SEU -	<i>Survey of English Usage</i>
COLT -	<i>Bergen Corpus of London Teenager Language</i>
MICASE -	<i>Michigan Corpus of Academic Spoken English</i>
BASE -	<i>British Academic Spoken English</i>
LLC -	<i>London-Lund Corpus of Spoken English</i>
COBUILD -	<i>Collins e Birmingham University International Language Database</i>
BNC -	<i>British National Corpus</i>
CIC -	<i>Cambridge International Corpus</i>
ICE -	<i>International Corpus of English</i>
ANC -	<i>American National Corpus</i>
COCA -	<i>Corpus of Contemporary American English</i>
CoMAprend -	Corpus Multilíngue de Aprendizes
CorTrad -	Corpus de Tradução
LOCNESS -	<i>Louvain Corpus of Native Essays</i>
ICLE -	<i>International Corpus of Learner English</i>
VOICE -	<i>Vienna-Oxford International Corpus of English</i>
ELFA -	<i>Academic Settings</i>
Br-ICLE -	<i>The Brazilian Sub-corpus of the International Corpus of Learner English</i>
GELC -	Grupo de Estudos de Linguística de Corpus
TALC -	<i>Introducing concordances to teach Spanish to Brazilian students</i>

## LISTA DE QUADROS

QUADRO 1 - Fábula <i>A tartaruga e os peixes</i> .....	41
QUADRO 2 - Regra Descritiva do Sintagma Nominal .....	69
QUADRO 3 - As posições de preenchimento do Sintagma Verbal. (BAGNO, 2011, p.514).....	74
QUADRO 4 - Texto de Aluno do 4º Ano do Ensino Fundamental do Grupo Infantil Terena Monolíngue com Problemas de Concordância.....	77
QUADRO 5 - Exemplo de Hipo e Hipersegmentação em Texto de Aluno do 4o Ano do Ensino Fundamental – Guarani Bilíngue.....	84
QUADRO 6 - Exemplo de Hipo e Hipersegmentacao em Texto de Aluno do 5o Ano do Ensino Fundamental – Guarani Bilíngue.....	85
QUADRO 7 - Exemplo do critério concordância nominal de gênero em texto do grupo infantil.....	108
QUADRO 8 - Exemplo de alguns critérios: erro de tempo verbal, ausência de determinante, hipossegmentação, ausência de concordância de número e pessoal, ausência de concordância nominal de gênero em texto do grupo infantil.....	108
QUADRO 9 - Exemplo do critério hipossegmentação em texto do grupo infantil .....	109
QUADRO 10 - Exemplo do critério hipersegmentação em texto do grupo infantil.....	109
QUADRO 11 - Exemplo de texto com problemas de ortografia do 4o ano do ensino fundamental Guarani/Kaiowá bilíngue.....	134
QUADRO 12 - Exemplo de texto com problemas de ortografia do 4o ano do ensino fundamental Terena Monolíngue.....	134
QUADRO 13 - Exemplo de texto com problemas de ortografia do 5o ano do ensino fundamental Guarani/Kaiowá bilíngue.....	135
QUADRO 14 - Exemplo de texto com problemas de ortografia do 5o ano do ensino fundamental Guarani/Kaiowá bilíngue.....	136
QUADRO 15 - Exemplo de texto com problemas de ortografia e acentuação do 4o ano do ensino fundamental Guarani/Kaiowá bilíngue .....	136
QUADRO 16 - Exemplo do emprego do gênero no texto de tema 03 do 5o ano do ensino fundamental Guarani/Kaiowá bilíngue.....	137
QUADRO 17 - Exemplo de texto do 4o ano do ensino fundamental, do subgrupo não indígena com problemas citados anteriormente .....	138

QUADRO 18 - Exemplo de texto do 4o ano do ensino fundamental, do subgrupo não indígena com poucos desvios da normatização da escrita em língua portuguesa.....	139
QUADRO 19 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue.....	140
QUADRO 20 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue .....	141
QUADRO 21 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue.....	141
QUADRO 22 - Exemplo de emprego de concordância nominal de gênero num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue .....	143
QUADRO 23 - Exemplo de emprego de concordância nominal de gênero num texto do 5o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue .....	144
QUADRO 24 - Exemplo de emprego de concordância nominal de gênero num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue ....	145
QUADRO 25 - Exemplo de emprego de concordância nominal de gênero num texto do 5o ano do ensino fundamental, do subgrupo indígena Terena monolíngue ....	145
QUADRO 26 - Exemplo de emprego de concordância verbal num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue.....	146
QUADRO 27 - Exemplo de emprego de concordância verbal num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue.....	147
QUADRO 28 - Exemplo de emprego de concordância verbal num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue.....	147
QUADRO 29 - Exemplo de hipossegmentação num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue .....	148
QUADRO 30 - Exemplo de hipossegmentação num texto do 5o ano do ensino fundamental, do subgrupo indígena Terena monolíngue .....	148
QUADRO 31 - Exemplo de hipo/hipersegmentação num texto do 4o ano do ensino fundamental, do subgrupo não indígena do grupo 1 .....	148
QUADRO 32 - Exemplo de hipo/hipersegmentação num texto do 5o ano do ensino fundamental, do subgrupo não indígena do grupo 1 .....	149

QUADRO 33 - Fragmento do texto de acadêmico do primeiro ano do ensino superior do subgrupo Guarani/Kiaiowá bilíngue, mostreo o emprego de verbos .....	150
QUADRO 34 - Exemplo de emprego de verbos e ausência de concordâncias num texto do 1o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2.....	151
QUADRO 35 - Exemplo de ausência de concordâncias e hipersegmentação num texto do 1o ano do ensino superior, do subgrupo indígena Terena monolíngue do grupo 2.....	152
QUADRO 36 - Exemplo de ausência de concordância de gênero e número num texto do 5o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2.....	152
QUADRO 37 - Exemplo de ausência de concordância de número num texto do 4o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2 .....	153

## LISTA DE TABELAS

Tabela 1.	Total de informantes por grupo considerado idade e experiência linguística	23
Tabela 2.	Número de textos coletados (e palavras totais produzidas) em cada grupo de participantes: .....	24
Tabela 3.	Lista de Palavras .....	60
Tabela 4.	Lista de palavras .....	61
Tabela 5.	Relação de palavras hipo e hipersegmentadas do 4o ano do ensino fundamental do grupo indígena bilíngue guarani .....	84
Tabela 6.	Relação de palavras hipo e hipersegmentadas do 5o ano do ensino fundamental do grupo indígena bilíngue guarani .....	85
Tabela 7.	Casos de hipossegmentação do grupo indígena Guarani/Kaiowá do 4o ano do ensino fundamental .....	86
Tabela 8.	Casos de hipersegmentação do grupo indígena Guarani/Kaiowá do 4o ano do ensino fundamental .....	86
Tabela 9.	Casos de hipo e hipersegmentação como serão analisados neste trabalho...	87
Tabela 10.	Subgrupo infantil indígena dividido por etnia .....	97
Tabela 11.	Subgrupo Infantil Não Indígena .....	97
Tabela 12.	Subgrupo Adulto Indígena subdividido por etnia .....	98
Tabela 13.	Subgrupo Infantil Não Indígena .....	98
Tabela 14.	Temas solicitados para os textos .....	99
Tabela 15.	Códigos utilizados para identificação dos textos do grupo infantil .....	101
Tabela 16.	Códigos utilizados para identificação dos textos do grupo dos acadêmicos	101
Tabela 17.	Relação dos sujeitos e de palavras produzidos pelo grupo infantil .....	103
Tabela 18.	Relação do número de sujeitos e de palavras produzidos pelo grupo adulto	104
Tabela 19.	Códigos para a etiquetagem .....	105
Tabela 20.	Quantificação dos dados do grupo de crianças Guarani/Kaiowá do 4o ano do ensino fundamental .....	112
Tabela 21.	Quantificação dos dados das crianças Terena monolíngues de português do 4º ano do ensino fundamental .....	113
Tabela 22.	Quantificação dos dados das crianças não indígenas do 4º ano do ensino fundamental .....	114

Tabela 23.	Quantificação dos dados das crianças Guarani/Kaiowá do 5º ano do ensino fundamental .....	114
Tabela 24.	Quantificação dos dados das crianças Terena do 5º ano do ensino fundamental .....	115
Tabela 25.	Quantificação dos dados das crianças não indígenas do 5º ano do ensino fundamental .....	116
Tabela 26.	Quantificação dos dados dos adultos indígenas Guarani/Kaiowá bilíngues do 1º ano do ensino superior .....	116
Tabela 27.	Quantificação dos dados dos adultos indígenas Terena monolíngues do 1º ano do ensino superior .....	117
Tabela 28.	Quantificação dos dados dos adultos não indígenas do 1º ano do ensino superior .....	118
Tabela 29.	Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Guarani/Kaiowá bilíngue .....	118
Tabela 30.	Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Terena monolíngue.....	119
Tabela 31.	Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Não Indígena .....	120
Tabela 32.	Os tipos de dificuldades averiguadas no corpus sob forma de etiquetas....	121
Tabela 33.	Dados gerais, por grupo etário e étnico/de língua, em cada uma das estruturas sintáticas.....	121
Tabela 34.	Comparação entre crianças indígenas Guarani/Kaiowá e crianças indígenas Terena – grupo 1 .....	123
Tabela 35.	Comparação entre crianças indígenas Guarani/Kaiowá e crianças não indígenas – grupo 1 .....	124
Tabela 36.	Comparação entre crianças indígenas Terena monolíngues e crianças não indígenas – grupo 1 .....	124
Tabela 37.	Comparação entre Adultos indígenas Guarani/Kaiowá bilíngues e Adultos indígenas Terena monolíngues – grupo 2.....	125
Tabela 38.	Comparação entre Adultos indígenas Guarani/Kaiowá bilíngues e Adultos Não Indígenas – grupo 2.....	125
Tabela 39.	Comparação entre Adultos indígenas Terena monolíngue e Adultos Não Indígenas – grupo 2 .....	126

Tabela 40.	Comparação entre Grupo 1 Guarani/Kaiowá bilíngues e Grupo 2 Guarani/Kaiowá bilíngues – grupo 2.....	127
Tabela 41.	Comparação entre Grupo 1 Terena monolíngue e Grupo 2 Terena monolíngue – grupo 2.....	128
Tabela 42.	Comparação entre Grupo 1 Não Indígena monolíngue e Grupo 2 Não Indígena monolíngue – grupo 2.....	129
Tabela 43.	Comparação entre Primeiro Ano do Ensino Superior Guarani/Kaiowá bilíngue e Último ano do Ensino Superior - Guarani/Kaiowá – grupo 2.....	130
Tabela 44.	Comparação entre Primeiro Ano do Ensino Superior Terena Monolíngue e Último ano do Ensino Superior - Terena Monolíngue – Grupo 2.....	130
Tabela 45.	Comparação entre Primeiro Ano do Ensino Superior Não Indígena e Último ano do Ensino Superior Não Indígena – Grupo 2.....	131
Tabela 46.	Total de textos produzidos pelo grupo 1, dividido em subgrupos.....	133
Tabela 47.	Total de textos produzidos pelo grupo 1, dividido em subgrupos.....	150

## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>19</b>
<b>2. O INDÍGENA E AS CULTURAS LETRADAS X CULTURAS NÃO LETRADAS</b>	<b>27</b>
2.1 A População Indígena no Brasil .....	27
2.1.1 Os Modelos de Educação Indígena no Decorrer do Tempo.....	28
2.1.2 As Políticas Públicas para a Educação Indígena .....	30
2.1.3 A Escola Indígena na reserva Francisco Horta Barbosa (RFHB) .....	33
2.2 Sociedades Letradas, Sociedades Não Letradas e a Teoria da Grande Divisão .....	38
2.2.1 Sociedades Letradas e Sociedades Não Letradas .....	38
2.2.2 A Teoria da Grande Divisão e a População Indígena Hoje .....	43
<b>3 A LINGÜÍSTICA DE CORPUS .....</b>	<b>45</b>
3.1 Fundamentos da Linguística de <i>Corpus</i> .....	45
3.1.1 Percurso Histórico e Conceitos.....	45
3.1.2 Algumas Considerações acerca do Termo <i>Corpus</i> .....	47
3.1.3 Análise Quantitativa e Qualitativa, Representatividade e Tamanho de um Corpus..	49
3.1.4 Tipos de Corpus/Corpora .....	52
3.1.5 <i>Corpus</i> de Aprendiz.....	55
3.1.5.1 Implicações dos <i>Corpora</i> de Aprendiz.....	56
3.1.6 Termos Técnicos da Linguística de <i>Corpus</i> .....	1
3.1.6.1 A Palavra na LdC .....	59
3.1.6.2 Combinações de Palavras em <i>Corpora</i> .....	63
3.1.6.3 Técnicas e Análises em <i>Corpora</i> .....	64
<b>4. A ANÁLISE GRAMATICAL .....</b>	<b>65</b>
4.1. As Investigações Linguísticas .....	66
4.2. A Gramática – Algumas Considerações a partir de Castilho (2010; 2011).....	67
4.3. Unidades da Gramática Descritiva - O Sintagma.....	68

4.3.1 A Gramática - Algumas Considerações a partir de Bagno.....	70
4.3.2 O Verbo .....	73
4.3.2.1 O Sintagma Verbal .....	74
4.3.2.2 Atributos dos Verbos .....	75
4.3.3 A Concordância .....	76
4.3.3.1 A Concordância Nominal de Gênero e Número.....	78
4.3.3.2 Concordância Verbal .....	79
4.3.3.3 Concordância em Guarani - Breves Considerações .....	80
4.3.4 Hipossegmentação e Hipersegmentação .....	82
4.3.4.1 A Oralidade e a Escrita.....	82
<b>5 METODOLOGIA.....</b>	<b>88</b>
5.1. Objetivos.....	88
5.2. A Escola Indígena na RFHB hoje .....	89
5.3. O Contexto da Pesquisa.....	92
5.4. A Construção do <i>Corpus</i> e a Análise de <i>Corpora</i> .....	94
5.4.1 Sistema de Anotação de partes do discurso.....	94
5.4.2 Anotação de Partes do Discurso para este Trabalho ( <i>part-of- speech annotation</i> ) ...	95
5.4.3 Etiquetagem de Erros .....	95
5.5 A Coleta dos Dados .....	96
5.5.1 A Seleção dos participantes.....	96
5.5.2 O Grupo Infantil .....	96
5.5.3 O Grupo Adulto .....	97
5.6 Procedimentos para a coleta .....	98
5.6.1 Os textos .....	99
5.6.2 O Questionário.....	100
5.6.3 Os Códigos Utilizados para Identificar os Textos.....	100
5.6.4 Os Percalços Para a Coleta dos Textos.....	101

5.6.5 O <i>Corpus</i> Geral Indígena de Dourados - Grupo Infantil.....	102
5.6.6 O <i>Corpus</i> Geral Indígena de Dourados e Região - Grupo Adulto.....	103
5.6.7 Os Critérios e os procedimentos para a descrição dos dados .....	104
5.6.8 O Material Pesquisado.....	107
5.6.8.1 A Seleção do material.....	107
5.6.8.2.As Técnicas para a Análise dos Dados Conforme as Ferramentas Computacionais do Programa <i>AntConc</i> .....	110
5.6.8.3 O Concordanciador.....	110
<b>6. ANÁLISE QUANTITATIVA DOS DADOS.....</b>	<b>112</b>
6.1 Análise Descritiva.....	112
6.1.1 Análise do grupo de crianças do 4º ano do ensino fundamental.....	112
6.1.2 Análise do grupo de crianças do 5º ano do ensino fundamental.....	114
6.1.3 Análise do grupo de adultos do 1º ano do ensino superior .....	116
6.1.4 Análise do grupo de adultos do último ano do ensino superior .....	118
6.2 Análise Estatística dos Dados.....	120
<b>7. ANÁLISE QUALITATIVA DOS DADOS .....</b>	<b>133</b>
7.1 Descrição Geral .....	133
7.1.1 Do Subgrupo Indígena Infantil.....	133
7.1.2 A Ortografia.....	133
7.1.3 O Gênero Textual .....	137
7.1.4 O Emprego do Verbo na Língua Portuguesa.....	139
7.1.5 A Concordância Nominal de Gênero.....	142
7.1.5.1 A Concordância de Gênero na Língua Portuguesa.....	142
7.1.6 A Concordância Verbal .....	146
7.1.6.1 Concordância Verbal na Língua Portuguesa .....	146
7.1.7 A Hipossegmentação e a Hipersegmentação.....	148
7.1.8 Do Ensino Superior – Primeiro e Último Ano da Graduação .....	149

7.1.8.1 Emprego dos Verbos .....	149
CONSIDERAÇÕES FINAIS .....	155
REFERÊNCIAS BIBLIOGRÁFICAS .....	159
ANEXOS .....	162

## 1. INTRODUÇÃO

A Universidade Estadual de Mato Grosso do Sul - UEMS recebe um grande número de estudantes indígenas, pois é adepta do sistema de cotas, não apenas no *campus* sede em Dourados, mas também nas outras 15 unidades universitárias instaladas no MS. A maioria deles fez o curso médio, do tipo EJA - Educação de Jovens e Adultos; outros estudaram nas escolas dentro de suas aldeias e outros ainda fizeram o ENEM do governo federal<sup>1</sup>. Há ainda um curso de nível médio oferecido pelo Estado aos Guarani/Kaiowá denominado *Ará Verá*, o qual trabalha dentro de um viés cultural indígena.

Ultimamente, nos cursos mais concorridos para o ingresso na UEMS, como Enfermagem ou Direito, eles se preparam por meio de cursinhos pré-vestibulares gratuitos, oferecidos pelo governo do Estado. Em outros cursos de graduação, entretanto, chegam sem nenhum preparo. Além desses, recentemente, os indígenas das etnias Kaiowá e Guarani da reserva participam de um curso de Licenciatura Indígena Intercultural, oferecido pela Universidade Federal da Grande Dourados (doravante UFGD), cujo objetivo é formar professores indígenas para atuar em sua própria comunidade, respondendo ao proposto nos *Referenciais Curriculares Nacionais para as Escolas Indígenas* – RCNEI (1998), que determina que o ensino para essas comunidades seja diferenciado, bilíngue e intercultural.

Apesar do amplo acesso à universidade, esses grupos nem sempre conseguem acompanhar os cursos e muitos abenam disciplinas para cursá-las posteriormente, devido às dificuldades financeiras como, por exemplo, falta de recurso para se deslocar até o local de suas aulas, para se alimentar no *campus* e para xerocopiar os textos solicitados pelos professores. Além disso, afora as questões de ordem econômica, há dificuldades de ordem emocional, pois grande parte dos alunos indígenas sofre a pressão das diferenças e da

---

<sup>1</sup> A prova que o governo utiliza atualmente para a obtenção da conclusão do ensino médio é o ENEM – Exame Nacional do Ensino Médio.

exclusão social. Como se vê, trata-se de um conjunto de fatores que, somados, resultam em um grande peso para o aluno indígena que sai de sua terra, de sua zona de conforto para adentrar num mundo no qual ele normalmente não é *bem-vindo*.

Além dos problemas relatados acima com relação à dificuldade apresentada pelo acadêmico indígena na UEMS, no contato com as escolas situadas na reserva indígena Francisco Horta Barbosa (doravante RFHB), de Dourados/MS, de onde se origina grande parte dos alunos indígenas que ingressam na UEMS, percebe-se que os relatos dos professores com relação à dificuldade na escrita por parte dos alunos assemelham-se às queixas dos professores da universidade. Os professores das escolas da reserva, a maior parte deles indígenas, apontam que pouquíssimos de seus alunos chegam completamente alfabetizados ao quinto ano do ensino fundamental. Claramente, o sistema escolar pelo qual atravessam crianças e adultos indígenas na região de Dourados parece apresentar fragilidades que comprometem de forma determinante a possibilidade de sucesso dessa população, principalmente no que se refere ao desenvolvimento da habilidade de produção escrita, pelo menos tendo como parâmetro o que se espera de um texto escrito dentro da normatização da língua portuguesa.

Nos dois contextos citados acima, a saber, cursos de graduação na UEMS e escolas indígenas localizadas na região de Dourados/MS, as prováveis hipóteses comumente formuladas informalmente para explicar a origem da dificuldade da população indígena em termos de produção escrita se assemelham, sendo que a mais divulgada é a suposição de que a língua indígena seria um grande obstáculo à aprendizagem da língua portuguesa. Ou seja, é comum ouvirmos frases como “Se pelo menos eles falassem menos guarani e mais português, aprenderiam a escrever português direito...”. Como se esse fosse o principal motivo para os alunos indígenas cometerem os desvios da norma ortográfica do português.

Essa realidade revela que a dificuldade na produção escrita em português tanto do aluno do ensino fundamental quanto do acadêmico indígena merece uma atenção e uma reflexão bem mais séria e mais profunda. Além disso, à universidade também cabe se preocupar com a formação do profissional indígena que vai atuar na reserva indígena, haja vista que esse profissional vem reivindicar uma atenção mais apurada em sua formação, dado ao índice de insucesso na alfabetização de seus alunos indígenas.

Assim nasceu o interesse pela pesquisa relatada na presente tese. De um lado, o acadêmico indígena com suas dificuldades de escrita em português em seu curso de

graduação e, de outro, a dificuldade constatada no processo de alfabetização e letramento realizado pelos professores indígenas que atuam nas escolas bilíngues da reserva. Duas frentes de trabalho se abriram a partir dessas duas perspectivas. A primeira, com relação à necessidade de compreender melhor as dificuldades de escrita dos acadêmicos e, a segunda, com relação às dificuldades de escrita que aparecem nos textos formulados pelas crianças do ensino fundamental que estudam na reserva. As duas frentes se interligam na formação dos acadêmicos indígenas.

Nesse sentido, a fim de compreender a origem de tantas críticas à escrita dos acadêmicos indígenas (por parte da própria universidade, sobretudo de professores) e dos problemas comuns de alfabetização e letramento que ocorrem no caso das crianças indígenas (reivindicação dos próprios professores da reserva), foi desenhada a investigação que é relatada nesta tese. A partir da pesquisa proposta, espera-se contribuir para uma melhor compreensão, por parte de professores que lecionam na reserva, bem como dos professores dos cursos de graduação, sobre como a escrita em língua portuguesa é organizada pelos alunos indígenas do ensino fundamental e pelos acadêmicos dos cursos de graduação. Essas informações poderão, futuramente, orientar o ensino da leitura e da escrita no contexto da reserva indígena, bem como fornecer subsídios para a formação de professores e elaboração de material didático a ser utilizado nesse contexto.

Assim, a presente investigação tem como objetivo geral descrever dificuldades na produção de textos narrativos escritos em português por crianças e adultos indígenas de três etnias – Kaiowá, Guarani e Terena – em comparação com crianças e adultos falantes monolíngues de português, a partir da Linguística de *Corpus*. Os objetivos específicos da pesquisa são:

- (a) Verificar se existem diferenças entre o tipo de dificuldades reveladas pelos indígenas monolíngues e bilíngues de diferentes etnias – Kaiowá/Guarani e Terena – em comparação com os monolíngues não indígenas na produção de textos narrativos em português.
- (b) Na comparação entre os dois grupos etários, crianças e adultos, observar em que medida o caminho percorrido do ensino básico à formação acadêmica interferiu no desenvolvimento da habilidade de escrita de textos.

- (c) No caso dos grupos de participantes adultos, investigar se o tempo de permanência no curso de graduação (alunos que estão no primeiro e no quarto ano de curso) interfere no nível de dificuldade na produção de textos.

A fim de atingir os objetivos da pesquisa, foram selecionados os grupos de participantes, a partir dos critérios faixa etária e língua/etnia, a saber:

- **faixa etária** → dois grupos, formados por crianças e por acadêmicos adultos. O **Grupo 1**, formado pelas crianças, foi subdividido em alunos matriculados no 4º ano do ensino fundamental e alunos matriculados no 5º ano do ensino fundamental. O **Grupo 2**, composto por informantes adultos, foi dividido em dois subgrupos, sendo um formado por acadêmicos que estão em início do curso de graduação e o outro composto por acadêmicos do último ano de graduação;
- **língua e etnia** → três grupos, formados por (a) falantes bilíngues de português e guarani/kaiowá/indígenas, (b) falantes monolíngues de português/ indígenas Terena; (c) falantes monolíngues de português/ não indígenas.

Os informantes do grupo das crianças indígenas foram alunos que frequentam as escolas situadas na reserva. O grupo de crianças não indígenas foi selecionado em uma escola pública estadual. O grupo das crianças indígenas foi dividido em bilíngues Guarani/Kaiowá e Terena monolíngues. A junção das crianças Guarani e Kaiowá num único grupo, denominado de bilíngues Guarani/Kaiowá, surgiu porque ambas as etnias são extremamente semelhantes nos costumes e na língua e ambas são falantes de sua língua materna. Com relação aos Terena monolíngues, a escolha se deu porque não se encontra na reserva número suficiente de falantes do idioma terena, pois os indígenas dessa etnia, nessa reserva, costumam se comunicar essencialmente em português. A amostra do grupo das crianças contou um total de 175 crianças, sendo 111 indígenas (71 bilíngues Guarani/Kaiowá e 40 Terena monolíngues) e 64 não indígenas, falantes monolíngues de português.

O grupo de adultos bilíngues e monolíngues, por sua vez, foi composto por um total de 118 acadêmicos da UEMS e de Universidades/Centro Universitários de Dourados e

região<sup>2</sup>. A decisão sobre a inclusão na pesquisa do grupo de participantes de acadêmicos se deu por dois motivos: primeiro, por eles terem sido os motivadores do trabalho; e segundo, por ser a outra ponta do ensino. No início, está a criança em processo de alfabetização e, no final, o acadêmico graduado, ou seja, vê-se aqui o início da vida escolar formal e sua finalização. Optou-se, então, pela inclusão de acadêmicos do 1º e do último ano do ensino superior de qualquer curso de graduação. Assim, o grupo de adultos foi formado por um total de 118 adultos, sendo 74 indígenas (36 bilíngues Guarani/Kaiowá e 38 Terena monolíngues) e 44 não indígenas, falantes monolíngues de português.

A Tabela abaixo mostra a divisão dos grupos:

Tabela 1. Total de informantes por grupo considerado idade e experiência linguística

Língua/etnia	Grupo 1: Crianças		Grupo 2: Adultos	
	4º ano do ensino fundamental	5º ano do ensino fundamental	1º ano do curso de graduação	último ano do curso de graduação
Falantes bilíngues de português e guarani/kaiowá/ indígenas ( <i>n</i> = 107)	36	35	18	18
Falantes monolíngues de português/ indígenas Terena ( <i>n</i> = 78)	19	21	19	19
Falantes monolíngues de português/ não indígenas ( <i>n</i> = 108)	34	30	20	24
TOTAL	89	86	57	61

A coleta de dados se deu da mesma forma para todos os grupos de informantes. Todos os participantes da pesquisa foram solicitados a escrever 03 (três) textos a partir de temas propostos pela pesquisadora<sup>3</sup>. O tema de número 1 foi “Um conto de minha infância”, a partir do qual os participantes foram instruídos a contar uma história que ouviram na infância, sendo que a intenção era oportunizar que os informantes narrassem uma história que fizesse parte de sua cultura. O tema de número 02 versava sobre a família: “Minha família e eu”, pensado com o objetivo de proporcionar que os informantes pudessem descrever as relações familiares. Por fim, o último tema denominava-se “O que é ser índio para mim”, sendo que aqui os indivíduos foram solicitados a se posicionar sobre a questão da identidade indígena. No caso dos grupos compostos por não indígenas, o tema 03 foi

<sup>2</sup> Não foi possível encontrar participantes em número suficiente somente em uma das instituições de ensino superior, pois os índices de evasão dos indígenas são bastante grandes nos vários cursos de graduação.

<sup>3</sup> Uma descrição mais detalhada sobre a pesquisa é apresentada no capítulo 3.

levemente modificado: “Como deve ser ‘ser índio’ para você”. Ainda, além de serem solicitados a produzir três textos, os participantes da pesquisa também preencheram um questionário de histórico da linguagem e informações pessoais que têm como objetivo auxiliar a pesquisadora a analisar questões culturais importantes para a compreensão das questões linguísticas apresentadas nos textos.

Ao todo, foram coletados 542 textos de crianças e 354 textos de adultos, como mostra a tabela abaixo:

Tabela 2. Número de textos coletados (e palavras totais produzidas) em cada grupo de participantes:

Língua/etnia	Grupo 1: Crianças		Grupo 2: Adultos	
	4º ano do ensino fundamental	5º ano do ensino fundamental	1º ano do curso de graduação	último ano do curso de graduação
Falantes bilíngues de português e guarani/kaiowá/indígenas	94 textos (8.204 palavras)	91 textos (8.816 palavras)	52 textos (9.641 palavras)	54 textos (9.642 palavras)
Falantes monolíngues de português/indígenas Terena	57 textos (4.542 palavras)	61 textos (6.466 palavras)	56 textos (7.809 palavras)	57 textos (11.569 palavras)
Falantes monolíngues de português/ não indígenas	97 textos (10.308 palavras)	83 textos (5.729 palavras)	59 textos (6.777 palavras)	71 textos (10.656 palavras)
TOTAL de textos produzidos TOTAL de número de palavras produzidas	248 textos (23.054 palavras)	235 textos (21.011 palavras)	167 textos (24.227 palavras)	182 textos (31.867 palavras)

Como suporte teórico e metodológico, optou-se por duas perspectivas teóricas que se coadunam: os estudos gramaticais contemporâneos e a Linguística de *Corpus* (doravante LdC), que serão discutidos a seguir. A primeira justifica-se pela necessidade de se caracterizar a escrita dos informantes da pesquisa. Nesse sentido, foram consultadas algumas

gramáticas descritivas do português contemporâneo e escolheu-se a *Gramática Pedagógica do Português Brasileiro*, de Marcos Bagno (2011), como principal suporte para direcionar a descrição e análise dos dados. A escolha ocorreu por essa gramática apresentar uma perspectiva contemporânea da língua portuguesa, a partir de exemplos retirados da língua em uso.

A escolha da Linguística de *Corpus* (LdC) se deu devido à necessidade de organizar e quantificar a descrição obtida nos textos produzidos. O *corpora* foi organizado através do programa *AntConc*<sup>4</sup>, com o objetivo de obter indícios de como se processa a escrita dos grupos participantes da pesquisa, com vistas a conhecer os mecanismos de escrita utilizados por eles. A utilização da Linguística de *Corpus* tornou-se pertinente, na medida em que, por meio de seus pressupostos se conseguirá uma análise mais precisa dos dados, já que essa área trabalha com *corpus* eletrônico. A Linguística de *Corpus* oferece, também, uma visão ampla e minuciosa do *corpus* pesquisado, e, a partir dessa visão poderá se pensar em produção de material didático adequado a cada situação de ensino de língua.

O escopo da Linguística de *Corpus* abrange, então, uma investigação linguística empírica, a partir da exploração nada assistemática de um *corpus*, que pode ser coletado apenas para essa finalidade (isso se aplica a esta pesquisa) ou pode ser acessado eletronicamente, ou seja, de algum *corpus* digitalizado e aberto ao público. Dessa forma, a LdC se revela de grande valia, pois os dados não podem ser inventados ou negligenciados. Ao contrário, obrigatoriamente o pesquisador terá que desenvolver os trabalhos, tendo em vista o que estiver registrado pelo *corpus*, escrito ou oral, não se admitindo manipulações desse ou daquele fato linguístico. Seu trabalho tem o foco no sistema de probabilidade linguística, o qual “prioriza o estudo de padrões de usos em contextos linguísticos específicos, por parte de um dado grupo de pessoas, em gêneros/registros particulares, entre outras possibilidades” (VIANA, 2011, p. 34).

A LdC, então, mostrou-se a metodologia mais adequada para o estudo em questão, pois o *corpus* coletado é de uma região indígena e de grupos indígenas que possuem algo em comum: sua forma de escrever. Como é uma população culturalmente diferenciada em seus costumes e língua materna, que recebe a cultura do não indígena por meio do letramento – via instituição escolar - faz-se imprescindível que seja estudado como ocorre esse processo para que, dessa forma, a instituição escolar indígena e não indígena saiba como agir com

---

<sup>4</sup><http://www.laurenceanthony.net/software.html>

relação ao ensino/aprendizagem, tanto com respeito ao método, quanto com relação às concepções de língua e ensino que deverão ser inseridos nessas comunidades.

A partir do levantamento de padrões linguísticos presentes na superfície dos textos, próprios desses grupos se discutirão as questões linguísticas e então poderão ser pensadas adequações de material didático mais apropriado para esses grupos, como já ocorre em outras comunidades indígenas em várias regiões do país, em que as comunidades escolares indígenas confeccionam seu próprio material didático. Assim, se vê que compreenderam o funcionamento do ensino/aprendizagem nestas localidades, cumprindo aquilo que o RCNEI (Referenciais Curriculares Nacionais para a Escola Indígena) apregoa: a educação escolar indígena deve ser intercultural e bilíngue, respeite sua cultura e sua língua materna.

Esta tese está estruturada em quatro capítulos, além da Introdução. No Capítulo 2, é feita uma explanação acerca das diferenças entre sociedades orais e sociedades letradas, cuja finalidade é mostrar o caminho traçado entre as sociedades orais e letradas, as concepções que nortearam seus argumentos, até então, e entender o porquê das sociedades indígenas ainda carregarem esse estigma de *atrasadas*. Na sequência, no capítulo 3, são apresentados os fundamentos e objetivos da Linguística de *Corpus* e uma breve apresentação do programa *AntConc*.

No Capítulo 4, são discutidos os estudos gramaticais contemporâneos da língua portuguesa, a fim de situar a perspectiva do ensino. A partir desses estudos os critérios, selecionados para a descrição serão discutidos e comentados.

O Capítulo 5 traz o percurso metodológico utilizado na pesquisa, o como e o porquê do interesse por esse objeto de investigação, a coleta de dados, bem como os princípios norteadores a partir da Linguística de *Corpus*. Suas ferramentas serão apresentadas, assim como a nomenclatura própria utilizada para direcionar de forma clara as descrições, a partir do programa *AntConc*.

A descrição e análise quantitativa dos dados, realizada a partir da retomada dos objetivos da pesquisa, está no Capítulo 6. No capítulo 7, é feita a análise qualitativa dos dados. Finalmente, seguem as Considerações Finais.

## **2. O INDÍGENA E AS CULTURAS LETRADAS X CULTURAS NÃO LETRADAS**

Neste capítulo, são apresentados alguns fatos sobre a população indígena no Brasil, bem como sobre as políticas públicas para a educação indígena e seu percurso educacional do passado até o momento. Além disso, faz-se uma descrição da Reserva Indígena Francisco Horta Barbosa, de suas escolas e da realidade educacional local. Na sequência, abordam-se, a partir da teoria da grande divisão, as diferenças entre culturas letradas e não letradas e suas consequências para os povos indígenas da atualidade.

### **2.1 A População Indígena no Brasil**

Uma volta ao passado mostra que a relação do povo brasileiro não indígena com o povo indígena nunca foi de harmonia. Desde o início, que os portugueses aportaram em terras brasileiras, houve confrontos, uma vez que vieram para cá tomar posse da terra, da qual os índios eram legítimos proprietários. Então, os portugueses alternavam entre tentar enganar os indígenas, comprá-los com bugigangas ou escravizá-los e, que nada disso dava certo, matavam-nos. No ano de 1500, cerca de cinco milhões de índios pertenciam à nação brasileira. De lá para cá, apenas 10% desse povo sobreviveu. Segundo estudo do indigenista Darcy Ribeiro (apud GOMES, 2012), nos meados dos anos 50, esse povo chegou a totalizar 100.000 pessoas. Cerca de 500 povos de diferentes etnias e culturas foram dizimados e, com eles, valores e conhecimentos nunca mais serão recuperados (GOMES, 2012).

Atualmente, segundo Gomes (2012), existem 530 mil índios no território brasileiro, ou seja, cerca de 0,47% da população brasileira, sendo que destes, por volta de 360 mil estão nas cidades, de acordo com o Censo 2010 do IBGE. Para De Paula e Vianna, em seu livro

*Mapeo Políticas Públicas para os Povos Indígenas*, publicado em 2011, existem 220 povos indígenas, que vão de etnias formadas por menos de 10 pessoas (AKUNTSU e JUMA de Rondônia com 05 pessoas) até as etnias com 50 mil integrantes (GUARANI - com 53.000 indivíduos). Conforme informam esses estudiosos, trata-se não apenas de habitantes de terras demarcadas, mas também daqueles que estão ocupando outros espaços que não são reconhecidos como indígenas.

Em reportagem recente da Revista *Isto É!* lê-se que somente 13% dos índios brasileiros usufruem do território nacional. Ainda segundo essa mesma reportagem:

112.983.625 hectares do território brasileiro são ocupados por índios. Isso representa 0,008 índio para cada oito mil hectares. 98% da extensão das terras indígenas estão na Amazônia Legal. Os outros 2% de terras indígenas se espalham em Mato Grosso do Sul e nas regiões Nordeste, Sudeste e Sul (*Revista Isto É!* 06/2013, p. 39).

Apesar de a maioria dos indígenas se concentrarem na zona rural, cada vez mais parte desse grupo migra para áreas urbanas. Sem mencionar aqueles cujas aldeias já estão praticamente urbanizadas, por estarem ao lado ou praticamente no interior das cidades e outras, por se constituírem de vilas apenas para indígenas dentro das cidades, denominadas ‘aldeias urbanas’ (DE PAULA e VIANNA, 2011). Há aqueles, entretanto, que ainda estão isolados, afastados dos centros urbanos, vivendo um cotidiano primitivo, nos moldes antigos da caça e da pesca. Alguns recebem mais assistência do governo, outros menos, chego a “situações de muito maior abeono, vulnerabilidade e quase ‘invisibilidade’ social” (DE PAULA e VIANNA, 2011, p. 8). A seguir, será apresentado um breve retrospecto dos modelos de educação indígena que vigoraram no Brasil e o modelo atual.

### 2.1.1 Os Modelos de Educação Indígena no Decorrer do Tempo

A realidade retratada na seção anterior a respeito da redução do povo indígena, assim como a diversificação devido a questões demográficas, tem se delineado paulatinamente no Brasil. Nesse sentido, é interessante notar que a educação indígena, apesar das mais complexas circunstâncias a que esses povos estiveram submetidos, sempre existiu. É necessário, entretanto, esclarecer o conceito de *Educação Indígena*, que não significa o mesmo que *Educação Escolar Indígena*, segundo Bartolomeu Meliá (1979).

*Educação Indígena* é o processo natural dos conhecimentos tradicionais de cada povo e, para isso, os mais velhos são tutores naturais, segundo MAHER (2006). Na cultura indígena, o ensino e a aprendizagem não ocorrem em lugares e horários pré-determinados; ocorrem no momento em que a oportunidade ou a necessidade requeiram determinado conhecimento, sem pressão e sem cobranças. Já a *Educação Escolar Indígena*, segundo Meliá (1979), estabeleceu-se pela necessidade da convivência e da interação com os não indígenas que vivem no entorno das aldeias e conseqüentemente para a inserção desses povos na sociedade brasileira, haja vista estarem à margem dela. A Reserva Francisco Horta Barbosa (RFHB) em Dourados/MS mostra claramente essa necessidade, dado que precisa da educação escolar para garantir a sobrevivência de suas aldeias, que estão sendo *tragadas* pela área urbana.

O conceito de Educação Escolar Indígena, todavia, desde a sua estruturação vem sofrendo alterações. Inicialmente, o objetivo centrava-se na assimilação total do índio aos costumes e à língua portuguesa, negeo totalmente suas origens – o chamado *Modelo Assimilacionista de Submersão* (MAHER, 2006). Como esse movimento não surtiu o efeito esperado, surgiu o *Modelo Assimilacionista de Transição*, responsável por criar escolas dentro das aldeias nas quais as crianças eram alfabetizadas na língua materna e paulatinamente inseria-se a língua portuguesa até que língua indígena fosse totalmente substituída (MAHER, 2006). Esses modelos comungavam o mesmo objetivo: a substituição dos valores e crenças tradicionais indígenas pelos valores e crenças dos não indígenas sendo que o modelo cristão-ocidental era almejado como o ideal. Nessa concepção, a única forma possível de interação entre os povos era que os povos ditos primitivos aceitassem os valores da sociedade não indígena, abdicasse de suas crenças, de seus costumes e de sua língua.

Por fim, Maher (2006, p.22), acrescenta que, nos últimos anos, vem se delineando um novo paradigma, o chamado *Emancipatório*:

E sob seus princípios que é construído o *Modelo de Enriquecimento Cultural e Linguístico*. Nele, o que se quer promover é um bilinguismo aditivo: pretende-se que o aluno indígena adicione a língua portuguesa ao seu repertório linguístico, mas pretende-se também que ele continue cada vez mais proficiente na língua de seus ancestrais. Para tanto, insiste-se na importância de que a língua de instrução seja a língua indígena ao longo de todo o processo de escolarização e não apenas nas séries iniciais. Além disso, esse modelo busca promover o respeito às crenças, aos saberes e às práticas culturais indígenas (MAHER, 2006, p. 22).

No entanto, apesar de esse novo modelo se insurgir muito timidamente, pois o anterior ainda prevalece, as escolas indígenas estão se empenhando em reforçar o bilinguismo aditivo, superando as adversidades que se apresentam. Trata-se de um processo lento e gradual, que se instaura conforme as forças culturais e políticas possibilitarem. Pelo fato de os indígenas terem um tempo diferenciado de aprendizagem surge o emprego de adjetivos como *incapazes* ou *lentos* na aprendizagem, que os indígenas recebem da sociedade não indígena. A nosso ver, um dos motivos que levam a essa caracterização preconceituosa é o fato de que os indígenas estão inseridos em uma sociedade ágrafa e com uma língua materna que não é o português.

Entretanto, um dado que não se pode perder de vista é o de que as sociedades indígenas são multilíngues. O Brasil é um país de várias línguas, no qual as línguas indígenas chegam a 180, e onde são faladas mais de 30 línguas de imigração. Então, o mito do país monolíngue não se confirma. Conforme a Organização das Nações Unidas (ONU), há uma média de 5.000 línguas distribuídas entre 193 países, confirmando que a maior parte da população mundial é bilíngue ou multilíngue.

Na subseção seguinte, tratemos das políticas públicas relativamente recentes para a educação indígena. Esse novo modelo de educação indígena ao qual cada comunidade indígena terá que se adaptar e trazer para seu contexto ainda está em construção<sup>5</sup>.

### 2.1.2As Políticas Públicas para a Educação Indígena

Há uma grande diversidade na população indígena que vive no país. Há índios aldeados e não aldeados, índios que vivem nas cidades ou em seus arredores, já adaptados à cultura não indígena; há também índios que vivem completamente alheios a outras culturas indígenas e não indígenas e há aqueles que estão em processo de adaptação à sociedade não indígena. Essa diversificação de culturas indígenas resulta em diferentes perspectivas por parte de cada povo que se trata de questões pertinentes à sua etnia, seus costumes, e também à sua concepção de escola. De Paula e Vianna (2011) mais uma vez contribuem ao afirmarem que há os povos que concebem a escola como um lugar de valorização da cultura e da identidade e outros que extraem dela apenas o imprescindível para aprender a lidar

---

<sup>5</sup> O termo *construção* é usado no sentido de que cada escola indígena deve se inteirar dos documentos que apontam perspectivas de ensino para as escolas indígenas e adaptar conteúdos, currículos e metodologias à sua realidade, haja vista que as comunidades indígenas possuem características próprias.

melhor com a sociedade não indígena e usufruir de algum de seus benefícios (DE PAULA e VIANNA, 2012). Para esses estudiosos:

os contextos reais e específicos, nos quais cada um dos povos indígenas está imerso, apontarão sempre para a diversidade e para as complexidades de situações, estejamos nós ocupados em observar as políticas públicas voltadas para a regularização fundiária, atendimento à saúde, fiscalização e vigilância territorial, geração de renda, gestão ambiental, valorização do patrimônio cultural, etc. (DE PAULA e VIANNA, 2012, p. 9).

Nesse sentido, qualquer tipo de política pública voltada para os indígenas terá que levar em conta esse quadro complexo de interesses dos diferentes grupos sociais, sobretudo no que tange à educação, pois cada realidade indígena demerita uma política específica (DE PAULA e VIANNA, 2012). Desde 1996, já figura no cenário legal nacional, por meio da Lei de Diretrizes e Bases da Educação, nos artigos nºs 78 e 79 da Constituição Federal (Lei nº 9.394), a normatização para a Educação Escolar Indígena postuleio um ensino diferenciado, intercultural e bilíngue.

A Resolução 03/99 do Conselho Nacional de Educação, em seu artigo 1º reza o seguinte:

Estabelecer, no âmbito da educação básica, a estrutura e o funcionamento das Escolas Indígenas, reconhecendo-lhes a condição de escolas com normas e ordenamento jurídicos próprios, e fixar as diretrizes curriculares do ensino intercultural e bilíngüe, visando à valorização plena das culturas dos povos indígenas e à afirmação e manutenção de sua diversidade étnica.  
(<http://portal.mec.gov.br/cne/arquivos/pdf/CEB0399.pdf> Resolução03/99)

O Parecer 14/99, que subsidia a Resolução, declara em seu artigo, 2º que “O Ensino Fundamental regular será ministrado em língua portuguesa, assegurada às comunidades indígenas também a utilização de suas línguas maternas e processos próprios de aprendizagem”. Desde 1998, o Ministério da Educação e do Desporto (MEC) por meio da Secretaria de Educação Fundamental, disponibilizou para os professores indígenas o *Referencial Curricular Nacional para as Escolas Indígenas* (RCNEI). O documento se divide em duas partes: a primeira se volta para os técnicos e funcionários das secretarias municipais e estaduais de educação; a segunda oferece referências para a prática pedagógica dos professores índios e não índios que atuam nas reservas e aldeias, advertindo que, dada a grande pluralidade e diversidade dos povos indígenas espalhados por todo o país, o RCNEI é um documento que se propõe a subsidiar os professores, não se constituindo de um receituário pronto para ser utilizado diretamente na sala de aula. A expectativa era de que esse documento auxiliasse a escola de cada comunidade indígena a adequar os

conhecimentos necessários a cada contexto individualmente, refletindo e orientando o desenvolvimento de suas práticas pedagógicas, a partir de suas experiências e saberes locais (RCNEI, 1998). O objetivo específico do documento é:

oferecer subsídios para: a) a elaboração e implementação de programas de educação escolar que melhor atendam aos anseios e interesses das comunidades indígenas, b) a formação de educadores capazes de assumir essas tarefas e de técnicos aptos a apoiá-las e viabilizá-las (RCNEI, 1998, p. 13).

A construção desse documento nasceu de necessidades da realidade educacional indígena, cujas diferenças com relação à educação não indígena mostravam-se por meio da ineficácia e do desinteresse dos alunos indígenas pelos conteúdos curriculares aplicados, bem como pela não compreensão desses alunos do que os professores, a maioria não indígenas, proporcionavam a eles, tendo em vista que esses conteúdos se distanciavam de suas realidades. Além do mais, esse documento surgiu no momento em que as discussões sobre educação escolar indígena tomavam corpo, devido, talvez, à inclusão das escolas indígenas nos sistemas educacionais em momento relativamente recente da história da Educação Brasileira (RCNEI, 1998).

A educação escolar indígena diferenciada, todavia, instituiu no cenário educacional brasileiro algumas reformas, adaptações e ações que precisaram ser garantidas para que houvesse funcionalidade. Uma delas foi a criação de um setor no Ministério da Educação e Cultura (MEC) responsável por gerir as questões educacionais voltadas para a diversidade e inclusão, a *Secretaria de Educação Continuada, Alfabetização, Diversidade e Inclusão – SECADI*, dentre elas a educação indígena. A SECADI, através da *Coordenação Geral de Educação Escolar Indígena (CGEEI)*, que é o órgão subordinado a ela para atuar incisivamente nas questões relacionadas à educação indígena, possui dentre suas principais atribuições formação inicial e continuada de professores em nível médio (magistério indígena) e superior (licenciaturas interculturais), além da produção de material didático próprio a cada comunidade linguística pelos professores indígenas, sempre com o apoio financeiro do MEC.

Ainda com relação às funções da *Coordenação Geral de Educação Escolar Indígena*, sob os auspícios da SECADI, citadas anteriormente, está a de estabelecer relações mais diretas com a *Fundação Nacional do Índio (Funai)*. A ideia é viabilizar aos professores e lideranças indígenas o conhecimento de seus direitos para que eles mesmos possam

exercer o controle social de seus programas de apoio à educação, bem como das ações necessárias para garantir sua realização.

Entretanto, surgiram críticas originadas a partir da implementação daquilo que os documentos apregoavam. Uma crítica forte, relacionada à conexão que deveria haver entre as três instâncias governamentais, a saber, municipal, estadual e federal, é de que a regulamentação entre os três poderes é tênue com relação à educação escolar indígena:

Muitas vezes, os complicados mecanismos de repasse de verbas implicam dificuldades para que insumos básicos do funcionamento escolar – como a merenda e o material didático – cheguem até a ponta inferior do sistema (...) Isso para não falar na falta de transparência com que os recursos do setor são gastos, faciliteo as suspeitas de desvios e corrupção (DE PAULA e VIANNA, 2012, p.75).

Para a Educação Superior foi criado em 2005 o *Programa de Apoio à Formação Superior e Licenciaturas Indígenas* – PROLIND, ofertado em sua grande maioria pelas Universidades Federais. Trata-se de cursos com habilitação em Ciências da Vida e da Natureza, Múltiplas Linguagens, Matemática e Ciências Sociais e Humanidades (DE PAULA e VIANNA, 2012). Inclusive, em Dourados/MS, local da pesquisa que deu origem a esta tese, há um curso de Licenciatura Intercultural Indígena na Universidade Federal da Grande Dourados, com o apoio do PROLIND. Esse curso é direcionado para indígenas das etnias Kaiowá e Guarani sendo que, além do requisito de pertencer a uma dessas etnias, exige-se também para o aceite no curso que o indígena seja falante da língua guarani.

Afora o PROLIND, há também a opção das vagas suplementares, em qualquer curso de qualquer universidade, em seus cursos regulares, nos quais os indígenas podem se inscrever e concorrer, inclusive pelo sistema de cotas.

### 2.1.3 A Escola Indígena na reserva Francisco Horta Barbosa (RFHB)

A fim de caracterizar as escolas indígenas existentes na Reserva Francisco Horta Barbosa (doravante RFHB) de Dourados/MS, local onde vive a maior parte dos informantes desta pesquisa, é necessário apresentar uma breve descrição da própria reserva, que foi criada pelo Decreto nº401, em 1915, pelo então Governo do Estado do Mato Grosso, que doou um lote de terras de 3600 hectares. Em 1925, com o título definitivo de propriedade restavam apenas 3539 hectares, ou seja, queo as terras foram legalizadas definitivamente em

1985, 61 hectares haviam sido perdidos. A reserva, desde então, passou a ser responsabilidade do Serviço de Proteção ao Índio (SPI), hoje FUNAI (AYLWIN, 2009).

Inicialmente, foram instalados na reserva os indígenas da etnia Guarani-Kaiowá<sup>6</sup>, seguidos pelos Guarani-Ñeeva - do tronco tupi. Na década de 30, vieram os Terena, do tronco Aruak. A vinda dos Terena teve como objetivo a inserção da população indígena à sociedade não indígena, por meio do trabalho com a agricultura, pois os Terena já utilizavam a língua portuguesa, além de utilizarem o manejo agrícola de forma semelhante ao não indígena.

Aylwin (2009) destaca que esses grupos, no entanto, vieram de culturas e tradições completamente diferentes. Os Guarani se originaram de um mesmo tronco linguístico e cultura semelhantes e viviam em famílias extensas, agrupando-se de maneira espontânea. Os Terena, por sua vez, se agrupavam em vilas, uma casa ao lado da outra, e possuíam uma organização política baseada na hierarquia, com um líder que surgia da escolha do povo.

Com o passar do tempo, a cidade de Dourados cresceu bastante, assim como o número de nascimento de crianças indígenas. Com isso, a reserva tornou-se pequena para abrigar os povos que ali habitam. Atualmente, convivem na reserva os Guarani, os Kaiowá, os Terena, os mestiços (resultado de casamentos interétnicos ou de indígenas e não indígenas), além de alguns poucos não indígenas que se casaram com indígenas.

Segundo o censo do IBGE (Instituto Brasileiro de Geografia e Estatística) de 2010, em Dourados convivem cerca de 15.000 índios de três etnias no interior de duas aldeias. A primeira denomina-se Bororó e a segunda, Jaguapiru. Na primeira, estão reunidos os Guarani- Kaiowá e os Guarani- Ñeeva, na segunda, estão os Terena, mais Guarani-Ñeeva e alguns poucos Guarani-Kaiowá, além dos mestiços e dos não indígenas (casados (as) com indígenas). Há, ainda nesta mesma reserva, a Aldeia Panambizinho, que fica um pouco mais distante. Nela concentram- se os Guarani-Kaiowá.

Os indígenas da família linguística Tupi-Guarani se dividem entre três subgrupos: Kaiowá, Mbyá e Ñeeva e somam cerca de 53.000 índios espalhados nos Estados de Mato Grosso do Sul, Pará, Tocantins, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (GOMES, 2012). Na Reserva de Dourados estão os Guarani-

---

<sup>6</sup>Nesta tese, será utilizado o nome da etnia com a inicial em maiúscula e no singular, à maneira dos antropólogos. O nome do idioma é escrito com a inicial minúscula.

Ñeeva e os Guarani-Kaiowá, que se autodenominam Guarani (Ñeeva) e Kaiowá. Na realidade, o grupo Kaiowá é o que mais preserva a língua guarani/kaiowá (para a maioria a denominação é guarani, para aqueles que falam o kaiowá a língua é o kaiowá - não há consenso quanto a isso). Esses dois grupos - Guarani e Kaiowá - pela convivência estreitaram seus hábitos e costumes e sua língua praticamente se fundiu. Eles carregam consigo os costumes, as tradições e a língua guarani/kaiowá.

Os Terena, por sua vez, pertencem ao tronco linguístico Arauak. Somam um total de 28.845 distribuídos entre dois Estados: Mato Grosso do Sul (doravante MS) e Mato Grosso (MT), sendo que no MS, a maioria ocupa o entorno das cidades de Aquidauana, Jardim e Sidrolândia, distribuídos em várias aldeias. Lá, eles preservam seus costumes e sua língua. Em Dourados, no entanto, são minoria frente aos Guarani<sup>7</sup> e poucos falam o idioma terena.

Após quase um século de convivência na reserva de Dourados, as famílias Guarani, Kaiowá e Terena aprenderam a se respeitar e a interagir de forma pacífica e harmoniosa, cada uma dentro de suas tradições, respeitando os costumes e a língua uns dos outros. Entretanto, conflitos, sobretudo de interesses políticos obviamente existem, sendo que um grupo sempre está disputando espaço com outro, como é natural em qualquer comunidade. Que se trata, porém, da defesa de questões indígenas importantes para o povo, eles se unem para se fortalecer. Uma das questões que mais gera conflitos entre indígenas e não indígenas na região é a demarcação das terras indígenas, devido ao pequeno tamanho do espaço de terra conferido a cada família, que se torna cada vez menor à medida que cresce a população indígena, mas o espaço territorial designado pelo Governo permanece o mesmo. Assim, em momentos de luta pela terra, todas as etnias se unem.

O espaço das aldeias Bororó e Jaguapiru são espaços bastante divergentes. Por se tratar de três etnias dividindo esse espaço, cada etnia se organizou de forma a ficar próximo a seus parentes consanguíneos e preservar sua tradição. Aqueles indígenas que quase não saem de sua aldeia - os Kaiowá, que moram na aldeia Bororó, por exemplo – vivem à moda dos habitantes da zona rural. Todavia são indígenas, e como não há mais espaço para caça e pesca fazem pequenas roças para sobreviver e moram em casas precárias, com banheiro externo e quase não possuem mobiliário, nem aparelhos eletroeletrônicos, como os demais que moram mais próximos à rodovia. Muitos deles falam apenas guarani, mas se esforçam

---

<sup>7</sup> Vamos nos referir aos Guarani e Kaiowá ora como Guarani/Kaiowá ou simplesmente Guarani, entendendo que o nome Guarani abrange os dois grupos: Guarani e Kaiowá.

para que as crianças aprendam a língua portuguesa como uma forma de ascensão social, por meio da escola.

Os Guarani e os Terena, residentes na Aldeia Jaguapiru, têm uma convivência mais próxima, sendo que alguns deles interagem diariamente com os não indígenas, devido à proximidade com a rodovia que dá acesso à área urbana. Dessa forma, os bens de consumo a que os não indígenas estão habituados, passaram, há muito tempo, a ser objeto de consumo desses indígenas. Por isso, muitos deles já trabalham na cidade, estudam e realizam suas atividades de lazer.<sup>8</sup>

Na RFHB em Dourados, há quatro línguas: o guarani, que é a língua corrente nas aldeias, praticamente todos a utilizam; o kaiowá, que é uma variação do guarani, falada apenas pelos mais antigos; o terena, falado por alguns Terena e o português, que junto com o guarani é a língua mais falada na reserva. Isso se reflete no ensino, pois há escolas em que a maioria das crianças fala guarani; há aquelas em que todos falam guarani; há aquelas nas quais a maioria fala português; e há aquelas que o professor não fala guarani e assume uma sala de crianças que ou a maioria ou todos falam guarani. Há, aquelas em que tanto o professor quanto as crianças falam guarani.

Esse novo modelo de educação voltado para o índio está mais próximo da realidade do povo indígena. Os primeiros documentos a que se tem acesso sobre a educação escolar indígena na reserva, que datam do século passado, informam que o primeiro arremedo de escola que foi implantado chegou por meio de missionários americanos que vieram para essa região com a finalidade de converter os índios ao cristianismo. Apoiado, pelo então, SPI (Serviço de Proteção ao Índio), hoje FUNAI (Fundação Nacional do Índio), implantaram fora dos limites da reserva, a primeira escola para os indígenas, em 1938. Dentro do espaço da *Missão Evangélica Caiuá*. Inicialmente, o objetivo era o de atender as crianças do orfanato *Nheeroga*, o qual abrigava crianças órfãs, que perderam os pais devido a uma epidemia de febre amarela. Conforme registros históricos, essa foi a primeira das escolas

---

<sup>8</sup>Há outros grupos indígenas na região da Grande Dourados. Em cidades com mais ou menos 80 quilômetros de distância, localizam-se duas outras aldeias. Em Caarapó, localizada cerca de 50 quilômetros de Dourados, encontra-se a aldeia Te'kwe. Já em Amambai, que se situa a cerca de 80 quilômetros de Dourados, encontra-se a Aldeia Limão Verde, todas da etnia Guarani, da vertente Guarani-Mbyá, como os indígenas do Rio Grande do Sul. Alguns desses indígenas moram em Dourados para estudar nas universidades, onde há um centro universitário, uma franquia de uma universidade particular paulista, uma universidade estadual e uma universidade federal. Todas recebem estudantes de vários municípios adjacentes e até mais distantes, além de estudantes de outros estados do Brasil. Há, inclusive, indígenas do Amazonas estudeo na Universidade Estadual de Mato Grosso do Sul (doravante UEMS).

indígenas a alçar o *status* de *Escola Indígena* dentre as cinco demais presentes na Reserva. Denominada *Escola Municipal Francisco Meireles*, em atividade desde 1938, antes era conveniada com a *Missão Evangélica Caiuá*. Com o tempo, essa escola foi municipalizada e atualmente atende do primeiro ao nono ano um número expressivo de crianças Terena, Kaiowá, Guarani e mestiços, além de alguns poucos não indígenas, e está sob a tutela da prefeitura local, como as demais, com exceção da Escola Estadual Intercultural *Guateka* – Marçal de Souza, de responsabilidade do Estado, que é a única que se dedica ao Ensino Médio.

É importante ressaltar que os primeiros anos de ensino na reserva eram totalmente voltados para a conversão do indígena à fé cristã, mais especificamente à fé evangélica, uma vez que a *Missão* era protestante. O trabalho era desenvolvido em conjunto com o Serviço de Proteção ao Índio (SPI), e introduziam-se noções de higiene, saúde, agricultura, com vistas a encaminhar o indígena para a sociedade dita *civilizada*, torneio-os “úteis à nação, que vivia a lutar contra a falta de braços para a lavoura com tantos filhos selvagens a serem instruídos. À Igreja cabia, através de seus missionários guiá-los, com muita paciência e obstinação”<sup>9</sup>

A reserva indígena de Dourados é um caldeirão multicultural, para onde vários povos com costumes, tradições e línguas diferentes convergiram em períodos espaçados e objetivos diferenciados. Os Guarani e Kaiowá chegaram primeiro e os Terena vieram depois e cada povo tem sua cultura, costumes e língua. Hoje, professores Terena, Guarani e Kaiowá disputam vagas entre eles. Essa realidade anteriormente não acontecia, pois os indígenas mal terminavam a primeira fase do ensino fundamental. Hoje, os indígenas estão nos cursos universitários da cidade, estudeo, se aperfeiçoao e deo retorno dessa formação para suas comunidades, como profissionais da saúde ou da educação. Na seção seguinte, será abordada o tema da *Grande Divisão*, que por muito tempo separou as sociedades em letradas/modernas e não letradas/atrasadas. O objetivo dessa seção é mostrar que essa divisão ainda persiste queo se trata da questão indígena, reforço a imagem de um indígena lento, preguiçoso e com dificuldades de aprendizagem.

---

<sup>9</sup>Sem autoria - Documento Intitulado “A implantação do regime tutelar e atividades escolares na Reserva de Dourados e na Aldeia Panambizinho”, Maio/Junho 2007.

## 2.2 Sociedades Letradas, Sociedades Não Letradas e a Teoria da Grande Divisão

A comunidade indígena não herdou nenhum modelo escrito em sua língua materna. Ao contrário, a preocupação com a língua escrita é recente, sendo que são poucos os manuais de língua que já foram desenvolvidos e os problemas com essa modalidade linguística aparecem de forma bastante contundente nas produções escritas dos indígenas. Herdeiros da tradição oral, poucas são as experiências escritas na comunidade. No entanto, elas ocorrem, pois eventos de letramentos surgem, seja nos cultos das igrejas, num curso sobre a confecção de sabão caseiro, ou numa reunião a respeito dos problemas da luta pela terra. A questão do letramento na sociedade será abordada nas próximas seções.

### 2.2.1 Sociedades Letradas e Sociedades Não Letradas

Em 1963, em um artigo intitulado *As consequências do letramento*, o antropólogo Jack Goody e o crítico literário e especialista em história da literatura Ian Watt (1917-1999) principiaram uma grande discussão que atravessou os meios acadêmicos e instigam uma gama imensa de pesquisas sobre esse tema até os dias de hoje: *as diferenças entre as sociedades orais e as sociedades escritas*. Os autores iniciam o texto demonstrando que, entre os itens envolvidos na herança cultural de uma geração a outra, estão as *palavras*, os elementos mais significativos das línguas. Evidentemente, as palavras não apenas incluem “o que pensamos normalmente como algo corriqueiro, mas também aqueles itens e ideias de tempo, espaço, objetivos e aspirações gerais” (GOODY e WATT, 2006, p. 14), que vão garantir às gerações subsequentes uma continuidade dos conhecimentos acumulados, assim como das experiências vivenciadas pelo grupo.

Nessa perspectiva, Goody e Watt (2006) caracterizam as sociedades orais como dependentes exclusivamente da memória humana: “todas as crenças e valores, todas as formas de conhecimento são comunicadas entre os indivíduos, no contato, face a face, ao contrário do conteúdo material da tradição cultural, como a pintura em cavernas, por exemplo” (idem, p. 14). Goody (1977), em *A domesticação da mente selvagem*, afirma que a escrita potencializou o conhecimento cumulativo enquanto a oralidade é uma forma semipermanente. Segundo Goody (2010), a escrita faz desaparecer o problema da armazenagem na memória, pois doravante se pode estudar o texto estático. A partir dessa dicotomia – sociedades orais e sociedades letradas –, Goody e Watt estabeleceram também

suas inferências sobre o processo de pensamento, competências cognitivas e as conquistas sociais dos dois tipos de sociedades.

As formas especializadas da escrita, conforme os pressupostos dos autores, garantem às sociedades modernas letradas o progresso, o cientificismo, a racionalidade do conhecimento e da experiência empírica. Elas acontecem desde a Grécia Clássica, que se tornou o ponto de referência de sociedade letrada em sua época e influenciou taxativamente o mundo atual. Na visão antropológica de Goody (1977, 2010), o surgimento da escrita foi um marco divisor entre as culturas antiga e moderna, entre aquelas sociedades que mantinham a tradição oral e aquelas que passaram a utilizar a escrita, pois essas últimas evoluíram de forma estrondosa mediante as anteriores. O sistema de escrita, a partir do alfabeto, foi uma grande descoberta e cindiu o mundo entre dois polos, ficeo, de um lado, o atraso, o pensamento primitivo, desprovido de intelectualidade, de abstração, e do outro, o progresso, sustentado pelo pensamento lógico, racional, científico. Evidentemente, esse último é o que tem significância para o crescimento intelectual e conseqüentemente econômico do mundo (GOODY, 1977; GOODY e WATT, 1963).

Muitos estudiosos seguiram as ideias de Goody e Watt e também se debruçaram sobre o tema da grande divisão. Dentre eles, destaca-se Patrícia Grandenfield (1972, *apud* Street, 1984), por exemplo, que realizou uma pesquisa com um grupo de crianças escolarizadas e um grupo de crianças não escolarizadas entre os Wolof do Senegal. Em seu estudo, a autora conclui que “diferenças significativas entre os grupos investigados ocorreram devido às diferenças entre a escrita e a oralidade” (p.20). Os Wolof escolarizados passaram a conhecer a grafia a partir da escola e os Wolof não escolarizados não a conheciam, pois sua língua é ágrafa. Para Grandenfield, os falantes da língua ágrafa confiavam muito mais no contexto para a transmissão de conteúdos verbais, o que levaria a implicações no processo cognitivo. Segundo ela, a fala dependente do contexto estava vinculada ao pensamento dependente do contexto, que é justamente o oposto do pensamento abstrato, sendo abstração na visão dela entendida como “a separação mental de um elemento da situação contextual na qual ela está inserida” (idem, p. 21).

Street (1984), no entanto, esclarece que a pesquisadora confunde as questões, uma vez que se sabe que a fala envolve interação face a face e é menos generalizada do que a escrita, que, por sua vez, se insere na maioria dos grupos culturais e não possui um referencial comum compartilhado por todos os seus usuários. O modelo que Grandenfield

(1972) apresenta para justificar suas conjecturas é baseado na educação ocidental, cujo foco centra-se no ensino puramente teórico, desvinculado da prática, na qual a criança adquire usos abstratos de pensamento ouvindo a fala do professor (STREET, 1984, p. 21).

Toda essa discussão só chegou até esse ponto devido, primeiro, ao fato de que, relacionada ao advento do surgimento da escrita, em algumas sociedades do passado, surgiu a ideia de que os povos que possuíam um sistema de escrita estavam cognitivamente à frente dos povos que não o possuíam. Em outras palavras, isso significava assumir que povos ágrafos não poderiam adquirir tecnologia, pois eram desprovidos de pensamento formal, que somente pode ser alcançado mediante o acesso à escrita. E, segundo, com o impulso do advento da imprensa no século XVI, houve a propagação da escrita de forma avassaladora mundo afora. Kleiman (1995), em um dos primeiros textos sobre o oral e o escrito no Brasil, propõe algumas discussões a esse respeito. A autora defende que a crença de que a escrita possibilita operações mentais abstratas, num nível de processamento mental mais elaborado, caracteriza essa confiança extrema no poder da escrita como transformadora. Nesse sentido, Kleiman (1995) afirma que

os processos mentais orais seriam mais simples, subjetivos, tradicionais, voltados para a exterioridade, para os aspectos vitais da condição humana, enquanto que o pensamento dos povos que têm a escrita seria mais complexo, objetivo, inovador, voltado para a vida psicológica interna (KLEIMAN, 1995, p. 32).

Essa concepção fez com que “as consequências cognitivas da aquisição da escrita” (op.cit., p23) fossem associadas àqueles que obtiveram a capacidade de utilizar a língua de forma a solucionar problemas relativos à classificação, categorização, memorização, etc. A nosso ver, no entanto, isso ocorreu, não devido ao poder da escrita em si, mas como resultado da escolarização formal, uma vez que esse tipo de habilidade é resultado de uma conquista, tendo em vista os objetivos que se deseja.

Para Street (1984), esse tipo de estudo reforçou a teoria da grande divisão entre povos lógicos/pré-lógicos, primitivos/modernos e agora letrados/não letrados, como mais uma forma de discriminar culturas diferentes e não que tivesse algum valor científico. O autor justifica que a perpetuação da ideia da grande divisão se deve aos trabalhos rasos, ou seja, com poucos dados e tendenciosos que foram realizados, como os de Grandenfield (1972).

A visão europeia prevalecente, resultado de trabalhos científicos superficiais e de testagens inadequadas, gerou conclusões equivocadas, tendo em vista que, queo se observa

mais atentamente as sociedades tidas como primitivas, ditas sem pensamento lógico, imediatamente percebe-se que o que se fazia era classificar esses grupos pelo que eles não tinham com relação à sociedade europeia. Nesse sentido, não se levou em consideração a sua forma distinta de pensar a partir de uma realidade não europeia, mas que nem por isso pode ser considerada desprovida de lógica ou de cientificidade (STREET, 1984).

Conforme Street (1984), o que estava em jogo eram convenções de sociedades letradas de uma classe dominante, e não uma lógica universal. A antropologia avançou a passos largos ao lembrar que o pensamento científico e não científico está presente em todas as sociedades e em todos os indivíduos de qualquer cultura (Idem, p.26). O próprio Street, convidado para uma palestra em 2007 na Universidade Federal de Minas Gerais (UFMG), deixou mais clara a questão de como são vistas as diferenças culturais entre os povos, ao narrar uma fábula budista colhida da *internet*, a qual traz com muita pertinência a noção de etnografia.

A fábula chama-se *A tartaruga e os peixes*. Eis sua reprodução em uma forma livre:

#### QUADRO 1 - Fábula *A tartaruga e os peixes*

“Era uma vez uma tartaruga que vivia num lago com um grupo de peixes. Certa feita ela resolveu sair do lago e dar uma volta em terra seca. Queo voltou os peixes a questionaram acerca do que tinha na terra seca: essa terra seca é molhada? Lá é fresco? Tem ondas? Lá se pode nadar? A cada pergunta a resposta da tartaruga era uma negativa. Então os peixes disseram: se não há nada disso então essa terra não existe, é fruto da sua imaginação. Então os peixes resolveram conhecer a terra seca. Ao voltarem, os peixes disseram à tartaruga: não nos diga o que não é, nos diga o que é. Ao que a tartaruga respondeu: Não consigo. Não tenho uma língua para descrever isso” (STREET, 2007, p.41).

Essa fábula nos revela o que a cultura ocidental construiu como modelo de verdade: a sua própria cultura. Tudo que foge àquilo que a cultura dominante instituiu como sendo o paradigma verdadeiro não existe. Em outras palavras, não se concebe a existência de um sistema religioso, político, educacional distinto do paradigma instituído pela cultura ocidental. Fica fácil, nessa perspectiva, afirmar que povos sem escrita, por exemplo, são desprovidos de pensamento lógico, pois eles não o têm conforme o que a sociedade ocidental estipulou como pensamento lógico e daí, a nosso ver, se originam rótulos discriminatórios como *atrasados*, *ignorantes*, com *déficits de aprendizagem*

Street (1984) cita o trabalho do sociolinguísta Labov (1973) para confirmar sua linha de raciocínio. A investigação desse pesquisador a respeito da língua e da sociedade considero a fala dos jovens de classe baixa de um gueto nova-iorquino mostra claramente que a forma de falar desses jovens, longe de demonstrar deficiência na língua inglesa, apontava para um novo viés que os trabalhos científicos linguísticos deveriam tomar doravante. Pelo contrário, Labov (1973) defendeu que a suposta ausência de gramaticalidade em suas frases não evidenciava deficiência cognitiva nem intervinha em seu desempenho cognitivo. A forma destoante com que os jovens utilizavam a gramática da língua, modifício e alterno alguns usos, como troca de pronome sujeito por pronome objeto ou a ausência de cópula, por exemplo, também exigia uma regra de uso dentro de sua variação da língua, pois se tratava de um padrão, ou seja, não era um uso aleatório, sem sentido.

Além disso, o fato de se utilizarem de uma variedade linguística distinta não tinha nenhuma relação com ausência ou presença de pensamento lógico, concluiu Street, a partir dos dados de Labov, asseguro que tanto a abstração quanto a lógica são empregadas no conjunto total da língua o tempo todo e em qualquer língua. Nesse momento, Labov já desenvolve uma pesquisa etnográfica de forma a não focar sobre o que a comunidade dos jovens do gueto nova-iorquino não possui linguisticamente com relação à comunidade linguística padrão. Ao contrário, ele mostra que a construção linguística dessa comunidade apresenta um sistema lógico e provido de regras tanto quanto o inglês da sociedade dominante e não é por ser diferente que é menor ou *errado*, conforme eram rotulados usualmente pela academia.

A herança da teoria da grande divisão ainda está presente em muitas comunidades das sociedades atuais. Considero-se o contexto da comunidade indígena de Dourados/MS, por exemplo, a ideia da divisão entre povos modernos e atrasados estarem vinculados a povos letrados e não letrados facilita a compreensão da visão que o não indígena e até o próprio indígena têm de sua relação com a escola ou universidade, enfim, com o ensino formal, de um modo geral.

É comum a queixa de professores, tanto das escolas da reserva quanto da universidade de que os alunos indígenas são *fracos*, possuem *dificuldades de aprendizagem*, não conseguem aprender a língua portuguesa. Conjecturas baseadas no senso comum e as justificativas mais comumente ouvidas como resposta a isso são que eles não aprendem porque falam outra língua materna, ou por alguma questão cultural: sua cultura é ágrafa, não

mantém o costume (como o nosso) de guardar coisas, a concepção de vida deles é muito diferente da nossa, etc. Com isso, se resolve a questão e não se precisa pensar mais sobre o assunto nem tentar entender como e por que agem dessa ou daquela forma com relação à escrita.

Os rótulos colocados nos indígenas – de que não aprendem por que são indolentes e sem interesse pelo ensino é comum nas instituições de ensino. Assim, todos aqueles que presumivelmente se preocupam com a questão educacional no país, nos estados ou nos municípios, e lutam incansavelmente para que todos tenham direito à educação, se isentam de responsabilidade, afinal, o ônus recai sobre o indígena, dado que ele proveio de uma sociedade não letrada.

### 2.2.2 A Teoria da Grande Divisão e a População Indígena Hoje

A teoria conhecida como a Teoria da Grande Divisão, introduzida por Goody e Watt (1963), responsável por induzir um grande número de estudiosos da linguagem a acreditar que a capacidade de raciocínio lógico-formal seria possível apenas aos povos que dominavam a escrita, sobretudo a escrita alfabética, não encontra mais respaldo nem em seus precursores. A divisão dos povos entre letrados e não letrados não se justificou no passado, e nem se justifica no presente, pois estudos comprovam que qualquer pessoa tem condições de aprender, não importa sua proveniência.

Queo se convive com sociedades indígenas, entretanto, os pressupostos da Teoria da Grande Divisão estão sempre presentes e são muitas vezes usados para justificar atitudes dos indígenas que sejam diferentes daquelas apresentadas por não indígenas, seja com relação ao processo de ensino/aprendizagem seja queo ele está reivindicando direito à terra. É possível que isso ocorra devido à tradição histórica brasileira que herdamos, através de um sistema educacional que exclui as minorias. Maher (2006) afirma, com relação aos indígenas, que

Fomos educados no interior de um sistema de educação construído a partir de um posicionamento ideológico que procura diluir as identidades indígenas com o intuito de torná-las menos visíveis aos olhos da nação brasileira. Para tanto, vimos, desde os primórdios da nossa História, procureo firmar essa noção de *índio genérico* para desidentificar os povos indígenas: uma estratégia eficaz queo se quer dominar alguém é destituí-lo de qualquer singularidade, é emprestar-lhe invisibilidade (MAHER, 2006, p. 15).

A relação dos indígenas com a comunidade local na região de Dourados e do Mato Grosso do Sul todo é extremamente complexa. Vários são os exemplos de aversão aos índios, devido à questão de disputa de terras. A teoria da grande divisão se faz clara nesses momentos: o indígena é visto como uma cultura menor, sendo que em muitos dos textos que compõem a amostra das crianças não indígenas para esta pesquisa aparece a expressão “tenho dó dos índios”. Esses relatos revelam a visão de uma população ora violenta, usurpadora de terras alheias, ora digna de pena grassa entre a população dessa região.

A divisão entre letrados x não letrados, que resulta na dicotomia sociedades atrasadas x sociedades avançadas, precisa ser desfeita, para que o indígena passe a ser considerado parte integrante da sociedade, sem olhares oblíquos. Para isso, é necessário um trabalho de inserção desses grupos minoritários na sociedade majoritária. Esses grupos precisam se sentir pertencentes à sociedade brasileira. Este trabalho pretende dar uma contribuição nessa direção e, para esse intento, optou pela Linguística de *Corpus* como colaboradora. No capítulo a seguir, serão destacados, sua história, seus pressupostos, suas delimitações, enfim tudo que abrange essa área relativamente nova nos estudos linguísticos.

## 3 A LINGUÍSTICA DE CORPUS

### 3.1 Fundamentos da Linguística de *Corpus*

Essa seção trata dos fundamentos, contexto histórico e principais conceitos da Linguística de *Corpus*.

#### 3.1.1 Percurso Histórico e Conceitos

Para McEnery e Wilson (1996), a Linguística de *Corpus* (doravante LdC) renasceu após muitos anos no ostracismo. Ressurgiu e trouxe consigo uma ampliação de seu escopo. Esses autores descrevem a LdC como “o estudo da língua baseada nos exemplos da língua usada na vida real” (p.01) e é anterior a Chomsky. Segundo eles, vários pesquisadores já vinham utilizando essa abordagem metodológica, embora ela ainda não fosse assim nomeada. Os estudos longitudinais, conhecidos desde 1957 são exemplos de tipo de pesquisa que utilizava coleção de textos para seus trabalhos. Há trabalhos anteriores a década de 50 que envolvem vários campos: aquisição de língua, convenções ortográficas, pedagogia da língua, linguística comparativa e sintaxe e semântica.

Para Sardinha (2004) é a “LdC se ocupa da coleta e da exploração de *corpora*, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” (p. 03). Qualquer que seja o pesquisador que se aventurar a discorrer sobre a LdC vai reiterar a mesma ideia, a de que o trabalho com *corpus* prioriza a língua em uso, a língua falada por pessoas de determinada localidade, numa dada circunstância e num dado momento histórico. Não é uma língua inventada por um pesquisador, use o sua intuição de falante nativo. Se essa língua destoar da língua padrão, da normatização linguística instituída pela ortografia oficial, por exemplo,

não importa, o que importa é que essa língua existe, pois pelo menos dentro de um determinado grupo, ela é verbalizada daquela forma.

Sardinha também concorda que trabalhos de pesquisa com *corpora*<sup>10</sup> existem desde o século XX, evidentemente não eram eletrônicos, mas manuais e esses trabalhos objetivavam exclusivamente o ensino de línguas. Estudiosos como o educador Thorndike e os linguistas Boas e Fries utilizaram-se de *corpus* em suas investigações. Evidentemente que as perspectivas naquela época eram outras. Primeiro, o *corpus* era organizado e mantido manualmente, já que não havia computador. Depois a ênfase era nos trabalhos de ensino de língua (pedagogia) e não na descrição da linguagem como ocorre hoje, em que as evidências empíricas são captadas eletronicamente.

Muito se discutiu e se discute com relação à validade dos trabalhos com *corpus*. A mais antiga discussão vem de Noam Chomsky que discorda que os trabalhos com *corpus* sejam de fato úteis para a pesquisa linguística, isso porque desacredita no foco sobre o desempenho linguístico. McEnery e Wilson, (1996, p. 06) declaram que Chomsky descrevia a “competência como nosso conhecimento internalizado, tácito da língua, e performance como a evidência externa da competência da língua e que seria usada em ocasiões particulares, que o outros fatores além de nossa competência linguística possam afetar sua forma”.

Segundo, McEnery e Wilson, para Chomsky, em sua visão racionalista, o linguista deveria se interessar em saber como funciona a língua, ou seja, é a competência do falante e não seu desempenho que deve ser objeto da investigação linguística. Sendo o *corpus* uma coleção de sentenças externalizadas e, portanto, influenciadas pelo ambiente, tanto externo quanto mental. Os dados não teriam a acurácia necessária para se alcançar um modelo de língua.

Chomsky defende que os exemplos da língua devem ser retirados da mente do falante nativo que conhece seus mecanismos de funcionamento, ou seja, pela introspecção ou intuição apenas. Sarmiento (2008) destaca que foi no auge das teorias chomskianas que surgiu a primeira compilação eletrônica de dados, através de Nelson Francis e Henry Kucera. Era um corpus sincrônico que contava com aproximadamente um milhão de palavras do inglês escrito, publicado nos Estados Unidos, denominava-se *Brown Corpus*, e

---

<sup>10</sup>*Corpora*(pronuncia-se córpora) – palavra de origem latina utilizada como plural de *corpus*. *Corpora* é um conjunto de *corpus* (conjunto de documentos).

sua compilação ocorreu entre os anos de 1961 a 1964 e foi rapidamente disponibilizado para uso (apud SARMENTO, 2008)

O *Brown Corpus*, no entanto, só se concretizou efetivamente devido a um *corpus* não computadorizado que deu origem aos *corpora* de hoje, o SEU (*Survey of English Usage*) compilado manualmente pelo londrino Reolf Quirk e sua equipe, a partir de 1959. O *corpus* contava com número fixo de 200 textos e com 5.000 palavras para cada texto. Sardinha expõe o seguinte a respeito do *Survey*:

O *Survey* foi organizado em fichas de papel, cada uma contendo uma palavra do *corpus* inserida em 17 linhas do texto. As palavras foram analisadas gramaticalmente, com cada ficha recebendo uma categoria gramatical. O conjunto de categorias resultante serviu de base para os etiquetadores computadorizados contemporâneos que fazem a identificação de traços gramaticais automaticamente” (SARDINHA, 2004, p. 04).

A alteração do *Survey* para *corpus* eletrônico falado (*London Lund Corpus*) ocorreu antes do escrito.

No Brasil, apenas no final da década de 90, do século passado, a LdC começou a se fazer presente nos trabalhos de pesquisa dos estudiosos que investigam as questões linguísticas. Apesar de a interface com a educação ser ainda incipiente (VIANA, 2010), hoje ela está um pouco mais disseminada. Os trabalhos enfatizam o processamento de Linguagem Natural, a Lexicografia e a Linguística Computacional. Percebe-se muito interesse do setor comercial/empresarial nas pesquisas envolvendo a LdC (SARDINHA, 2004).

A relação empresa/universidade tem como pioneiro o *Cobuild*, parceria entre a Universidade de Birmingham da Grã-Bretanha e da Editora Collins, que resultou na publicação de vários dicionários, gramáticas e livros didáticos para o ensino do inglês. Após o *Cobuild*, as parcerias nunca mais deixaram de existir.

### 3.1.2 Algumas Considerações acerca do Termo *Corpus*

Para compreender a LdC é fundamental que se comece pelos conceitos de *corpus* e *corpora*, pois são os conceitos responsáveis pelo direcionamento desta investigação linguística. Sardinha (2004) chama a atenção para o amplo número de definições de *corpus* que existem na literatura de referência, dentre eles: “*corpus* é uma coletânea de textos

naturais (*Naturally Occurring*), escolhidos para caracterizar um estado ou variedade de linguagem” (p. 16).

Textos naturais se referem a textos que existem espontaneamente, sem terem outra finalidade que a interação entre os interlocutores, tanto na fala quanto na escrita. Sardinha também ressalta que é natural para diferenciar os textos produzidos por seres humanos dos produzidos artificialmente, por máquinas. Para uma definição mais completa Sardinha (2004, p. 17) redefine *corpus* como “um corpo de linguagem natural (autêntica) que pode ser usado como base para pesquisa linguística”. Dessa forma, fica claro que o objetivo do *corpus* é a investigação linguística. Apesar dos textos serem espontâneos, naturais, eles são compilados com o intuito de servirem para pesquisas e, portanto, tornam-se artificiais, travestindo-se em artefatos de pesquisa.

Sardinha utiliza a expressão “porções de linguagem” em preferência a “textos”, pois entende que não há consenso com relação à definição do conceito do que vem realmente a ser um texto. *Corpus*, então, é: “uma coletânea de porções de linguagem que são selecionadas e organizadas de acordo com critérios linguísticos explícitos, a fim de serem usadas como amostras da linguagem”. (SARDINHA, 2004, p. 16)

Para Viana (2008, p.31), o termo *corpus* se define como:

uma compilação eletrônica e criteriosa de (amostras de) textos que ocorrem naturalmente com o objetivo de representar uma dada língua ou algum de seus aspectos mais pontuais de forma a possibilitar uma análise linguística previamente delineada. (VIANA, 2008, p.31).

Para este autor, *corpus* pode ser definido como um conjunto de textos orais ou escritos que tenham sido produzidos por falantes de uma determinada língua e que de algum modo representa de forma geral ou específica, características próprias da língua daquele contexto. Além disso, pode ser compilado e transformado em um formato eletrônico para que qualquer investigador possa acessá-lo e utilizá-lo com propósitos próprios.

Isso não significa que qualquer conjunto de textos em formato eletrônico seja um *corpus*, apenas aqueles organizados e planejados minuciosamente para esse fim se constitui como tal. Desse modo, qualquer tipo de texto, tanto escrito quanto falado pode vir a ser um *corpus*, se for cuidadosamente selecionado e organizado para isso. Flowerdew baseado em McCarthy (2001, p. 63) assinala que “*corpora* são artefatos sociais cuja investigação pode desvendar comportamentos discursivos de determinada comunidade”. Por isso, a escolha

dessa metodologia para este trabalho, uma vez que a intenção é mostrar o comportamento linguístico escrito das crianças da escola indígena de Dourados e a escrita dos universitários indígenas que frequentam as universidades de Dourados e região.

Viana (2010) argumenta que um *corpus* deve:

“i) compreender um conjunto de textos; ii) contemplar textos orais ou escritos que tenham sido efetivamente produzidos por falantes de determinada língua; iii) consistir numa forma de representar empiricamente o uso que se faz de uma língua em seu sentido geral ou específico; iv) reproduzir a produção linguística de toda a população que se quer investigar ou uma amostra representativa dessa população, com base em princípios claros e bem definidos; v) assumir a forma eletrônica com vistas a ser investigada pelo computador; vi) objetivar a possibilidade da realização de uma pesquisa linguística”(VIANA, 2008, p. 27).

A noção de *corpus*, portanto, é essencial para se compreender sua aplicabilidade e compreender os resultados que surgirem a partir de seus dados em pesquisas. Com amostras de textos autênticos, extraídas de um contexto real de uso e interação a fidelidade da análise será mais confiável.

### 3.1.3 Análise Quantitativa e Qualitativa, Representatividade e Tamanho de um Corpus

Um corpus, todavia, para ser confiável deve ter um grau de representatividade. McEnerye Wilson (2004) contribuíram grandemente na discussão desse ponto ao colocarem que o trabalho com dados de amostras devem ser extremamente cuidadosos, pois, às vezes, os dados podem ser enganosos. Dados que não ocorreriam comumente na língua como um todo, poderiam aparecer mais na amostra e dados que são comuns na língua teriam uma rápida aparição e isso mascararia os resultados.

Isso chama outra discussão, a de que a LdC privilegia os dados quantitativos e despreza os dados qualitativos, pois para os autores as diferenças entre ambas as abordagens é que na pesquisa quantitativa as características linguísticas que aparecem são classificadas, contadas e modelos estatísticos mais complexos são construídos na tentativa de explicar o que é observado (p. 76). Já no modelo de pesquisa qualitativa, “os dados seriam usados apenas como base para identificar e descrever aspectos usados na língua e para fornecer exemplos retirados da ‘vida real’ de um fenômeno particular” (McEnerye Wilson, 2004, p. 76).

Outros trabalhos científicos, no entanto, esclarecem que ambas as abordagens são fundamentais numa pesquisa baseada em *corpus*. McEnery e Wilson assinalam o trabalho de Schiemed (1993), que defende que tanto uma abordagem quanto a outra são necessárias para se trabalhar com dados. A qualitativa permite uma perspectiva mais rica e detalhada, contemple uma descrição de determinado termo em toda a sua amplitude e a quantitativa poderia ser estendida à comunidade linguística toda (SCHIEMED, 1993 apud MCENERY e WILSON, 1996)

A análise qualitativa pode fornecer uma maior riqueza e precisão dos dados, enquanto a quantitativa pode oferecer dados estatisticamente confiáveis e resultados mais amplos. McEnery e Wilson acrescentam ainda que há, atualmente, uma tendência de pesquisa que veio das Ciências Sociais que privilegiam as abordagens que utilizam multimétodos em suas pesquisas, dessa forma, não há um fechamento da abordagem metodológica, mas sua ampliação para receber benefícios de outras áreas.

A LdC, então, constitui-se numa metodologia que trabalha com perspectivas quantitativas e qualitativas, a primeira, devido à utilização de ferramentas computacionais realizam uma análise formal, quantificam os dados. A segunda, naturalmente vai se estendendo para a função do objeto linguístico a ser estudado, pois parte-se do que o programa de computador extraiu primeiro. Assim, a análise qualitativa será melhor evidenciada para o observador que terá que usar de toda a sua argúcia e perspicácia para apreender os dados que surgirem do *corpus*. Dessa forma:

a abordagem qualitativa se revela fundamental para decodificar os padrões evidenciados pela ferramenta computacional, contextualizar os resultados em termos tanto da população investigada como da macro área em que uma pesquisa se insere, explicar o que os números revelam e interpretar os achados de uma forma em geral (VIANA, 2010, p. 35).

A representatividade está estreitamente ligada à noção de probabilidade. Para Sardinha (2004), a linguagem é um sistema de probabilidades, ou seja, alguns traços são mais recorrentes que outros. Segundo seus estudos, no âmbito léxico-semântico pode-se separar as palavras entre aquelas de maior e de menor frequência, sendo que algumas palavras são de baixa frequência, necessitam de um *corpus* maior para que elas sejam representadas na comunidade linguística que se quer estudar. Para melhor sistematizar isso, é necessário separar forma de sentido, pois as palavras podem se repetir várias vezes - mesma forma - mas, com sentidos diferentes. Nas palavras de Sardinha: “para que seja

representativo, um *corpus* deve conter o maior número possível de cada forma” (2004, p. 24).

Stubbs (2001, apud SARMENTO, 2008) acrescenta que, por ser um sistema probabilístico, que leva em conta as probabilidades do aparecimento de uma palavra e não de outra, por exemplo, muitas características e combinações são possíveis de ocorrer na língua, porém nem todas são prováveis de ocorrer. Sardinha (2004) destaca que o mais importante com relação à frequência das palavras (ou formas) é não serem aleatórias. Elas sempre ocorrem por uma escolha daquela comunidade linguística em detrimento de outras. Existe uma padronização dentro de uma variação sistemática em contextos linguísticos específicos e a variação desses traços específicos provê essa padronização à língua (SARDINHA, 2004).

Sardinha acrescenta que a extensão do corpus comporta três dimensões: a primeira, relacionada ao número de palavras, quanto maior o *corpus* maior será a probabilidade de emergirem palavras de baixa frequência, que são as que formam a maioria das palavras de uma língua. A segunda está ligada ao número de textos, que o se trabalha com uma língua específica, pois quanto maior o número de textos maior será a representatividade com relação aos gêneros/tipos de textos presentes na comunidade linguística em questão. Por fim, a última, que o o trabalho se aplica a *corpora* variados, o maior número possível de gêneros, registros ou tipos textuais presentes naquele contexto linguístico melhor representará a comunidade como um todo.(SARDINHA, 2004).

Viana (2010) colabora também com essa grande discussão sempre recorrente, que o se trabalha com *corpus*, e provavelmente, a maior delas, a concepção de representatividade, haja vista que não há regras claras e precisas para defini-la. Viana (2010), entretanto, argumenta que apesar não haver regras rígidas para definir a representatividade de um *corpus*:

alguns critérios são úteis na compilação de *corpora* representativos. Um deles corresponde à noção de diversidade: um *corpus* que objetiva representar a totalidade de uma língua precisa abarcar uma ampla gama de gêneros discursivos, contextos de produção, participantes (de diversas faixas etárias, origens geográficas, sexo, classes sociais etc), entre outros. Ao mesmo tempo, a diversidade deve ser temperada com a concepção de equilíbrio. Por exemplo, uma coletânea composta somente por textos escritos não representa uma língua, da mesma forma que outra com textos disponíveis em sites de jornais não é possível de ser compreendida com uma amostra da linguagem jornalística. Em ambos os casos, o que se tem é uma compilação mais específica, a saber, um corpus escrito e outro de linguagem jornalística *on-line* (VIANA, 2010, p. 28).

Conclui o autor, a partir da proposta de Bowker e Pearson (2002, p.45-46 apud VIANA, 2010) que, às vezes, consegue-se extrair informações valiosas para seu trabalho num *corpus* menor, pois será muito mais representativo para aquilo que se busca, do que de um *corpus* maior que não foi elaborado para seus propósitos.

Outra questão polêmica está relacionada ao tamanho do *corpus*. Para Bowker e Pearson (2002, p.45-46, apud VIANA, 2010, p.29), nesse caso, o objetivo que se quer alcançar é que vai definir o tamanho ideal de um *corpus*, isto é, vai depender das exigências de seu projeto, da disponibilidade de dados que se alcançou, bem como o tempo disponível para se trabalhar com esses dados.

As técnicas de amostras reômicas utilizadas para construir *corpus* para pesquisa de LdC são as mesmas técnicas usadas por outras ciências que trabalham com *corpus*. Essas técnicas são padrão, todavia, algumas questões são pertinentes de serem lembradas: deixar o mais claro possível os limites da população que se almeja estudar antes de definir os procedimentos de amostragem especificamente para esse objetivo, é uma delas (Biber, 1993). O tamanho da amostra e o número de textos desta amostra que seriam ideais para representar a escrita da comunidade linguística estudada também é um aspecto de grande relevância neste tipo de pesquisa.

#### 3.1.4 Tipos de Corpus/Corpora

Segundo Lin e Adolphs (2011, p. 597), “corpora são projetados para representar uma variedade de determinada língua” e há diferenças entre corpora especializados e gerais. O primeiro está relacionado a linguagens específicas, como por exemplo, a linguagem acadêmica, a linguagem jornalística, etc. Exemplo de um *corpora* de língua especializada é o *Bergen Corpus of London Teenager Language* (COLT), que representa um modo particular de discurso, geralmente, presente no meio acadêmico, como o *Michigan Corpus of Academic Spoken English* (MICASE) e o *British Academic Spoken English* (BASE) - versão britânica do MICASE.

O segundo é mais amplo e inclui todo tipo de texto, orais e escritos e seu objetivo é servir de fonte de pesquisa, para posterior elaboração de materiais de vários tipos, como dicionários, por exemplo. Por vezes, é também denominado *Corpora* de Referência, pois

serve de contraste em relação ao especializado. O mais conhecido é de língua inglesa, o *British Nacional Corpus* (BNC).

Os *Corpora Monitor* são semelhantes ao Histórico - que objetivam apontar as mudanças que ocorrem na língua no decorrer do tempo, como o *Tycho Brahe* que contém textos em português de autores nascidos entre os séculos XIV e XIX. A diferença entre o Histórico e o Monitor incide no fato de que esse último aponta mudanças simultâneas, atuais e, por isso, são alimentados em períodos curtos de tempo a fim de estarem sempre atualizados.

Dois dos mais importantes *corpora* elaborados nos anos 80 e 90 foram o *Collins e Birmigham University International Language Database* (COBUILD) e o *Bank of English*, que fornecem amostras da língua escrita, sobretudo a britânica, bem como a transcrição da fala, por meio de entrevistas e da conversação usada no cotidiano. Esses *corpora* são alimentados constantemente e já atingiram o índice de 450 milhões de palavras. Seu principal objetivo é oferecer suporte a pesquisadores envolvidos com pesquisas lexicográficas e com organização de dicionários. Muitas editoras também desenvolvem seus próprios *corpora* para servir a seus autores. Exemplos são o *Cambridge International Corpus* (CIC), o *Longman Corpus Network* e o *Oxford English* (LIN e ADOLPHS, 2011).

Recentemente, surgiram dois importantes *corpora* do inglês americano: o *American National Corpus* (ANC) e o *Corpus of Contemporary American English* (COCA). Até 2009, o ANC colheu cerca de 22 milhões de palavras do inglês americano falado e escrito produzido desde o início dos anos 90. Já o COCA possui mais de 400 milhões de palavras do inglês americano, sendo que destes, 21% é de fala e 79% de material escrito. A cada ano são adicionadas 20 milhões de palavras ao COCA, ele também pode ser classificado como *Corpus Monitor* (LIN e ADOLPHS, 2011).

Além desses, há o COMPARA, destinado a registrar o uso das línguas portuguesa e inglesa em textos literários traduzidos ou originais, baseado em Portugal, o COMPARA é exemplo de *Corpus* paralelo (VIANA, 2010, p. 30). Como esses, há um número cada vez maior de *corpus* linguísticos disponíveis na *internet* para todo tipo de pesquisa que se queira realizar.

Os *corpora* podem também ser classificados, conforme seu emprego em determinada pesquisa. Segundo Viana (2010, p. 33): “caracteriza-se um *corpus* como **de estudo** que o se

deseja observar, descrever e explicar o uso da linguagem registrado nos textos nele contidos. Já um *corpus de referência* representa uma coletânea de textos que provê o padrão de comparação para o *corpus* de estudo”.

Segundo Viana (2010), os *corpora* podem ainda ser diferenciados pelo número de línguas que contêm. *Corpus* monolíngue, como o *Corpus Lessico di Frequenza dell’Italiano Scritto Contemporaneo*, que coletou amostras apenas da língua italiana, ao contrário do *Corpus* Multilíngue de Aprendizes (CoMAprend), classificado como *Corpus* Multilíngue, cujas amostras abrangem textos de alunos das línguas alemã, espanhola, francesa, inglesa e italiana.

Esses *corpora* podem ser também considerados como paralelos. É o caso do *Corpus* de Tradução (CorTrad), desenvolvido pela Universidade de São Paulo, a partir de textos jornalísticos, literários e técnico-científico, produzidos em português e traduzidos para ser comparados, uma vez que a coleta dos textos foi realizada com a finalidade de observar características semelhantes em ambas as línguas ou nas variedades de uma delas (VIANA, 2010).

Que o *corpus* é produzido na língua materna do falante denomina-se *corpus* de primeira língua (L1), língua nativa ou materna, como é o caso do *Louvain Corpus of Native Essays* (LOCNESS) que é composto por produções em inglês americano ou britânico. Ao contrário do *International Corpus of Learner English* (ICLE), que consta de produções de aprendizes de uma língua estrangeira, e é mais conhecido como *Corpus* de Aprendiz. Além desses, o *Cambridge Learner Corpus the Longman Learners Corpus* e os que exploram o uso do inglês como língua franca, o *Vienna-Oxford Internacional Corpus of English* (VOICE) e o *corpus* que trabalha com o inglês como língua franca em ambientes acadêmicos, o *English as a Lingua Franca in Academic Settings* (ELFA). Muitos deles possuem suas próprias ferramentas disponíveis na *internet* (LIN e ADOLPHS, 2011).

Na presente pesquisa, trabalha-se com um *corpus* de estudo coletado manualmente<sup>11</sup> nas escolas e nas residências da Reserva Indígena Francisco Horta Barbosa. Outra parte do *corpus* de estudo foi coletada manualmente nas universidades em que estudavam alunos indígenas. O *corpus* de referência foi coletado em uma escola pública estadual do centro da cidade de Dourados na Universidade Estadual de Mato Grosso do Sul (UEMS), como forma

---

<sup>11</sup> Os textos que compõem o *corpus* do presente estudo foram coletados através de redações escritas a mão pelos participantes do estudo. Por isso, o caráter manual de coleta do *corpus*.

de comparação, ou como um grupo de controle. Tendo como referência a escrita não indígena, ou seja, a escrita não indígena será o paradigma adotado para os termos comparativos com os textos indígenas. Poderia se escolher um *corpus* eletrônico não indígena com um número maior de palavras, como grupo de controle, no entanto, o fato de o *corpus* de referência estar dentro da mesma localidade oferece maior credibilidade aos dados, pois se supõe que a escrita desses indígenas obrigatoriamente seja influenciada pelas adjacências educacionais de sua região.

Os textos coletados são de aprendizes de língua portuguesa numa comunidade linguística de fala guarani/terena, tal abordagem se encaixa na denominação *Corpus de Aprendiz*, que será destacada na seção a seguir.

### 3.1.5 *Corpus* de Aprendiz

Os *corpora* de aprendizes são uma categoria que, inicialmente, abrangia textos escritos ou falados de alunos que estavam em processo de aquisição de língua estrangeira. Esses textos foram digitalizados e disponibilizados eletronicamente. Com esses textos, pôde-se perceber as áreas nas quais os aprendizes possuíam mais dificuldades e poderiam ser melhor ajudados. Num segundo momento, esse *corpus* de aprendiz eletrônico centrou-se no ensino de língua materna e o grande benefício foi o de se compreender as variedades que surgiam nas comunicações e como a escrita tinha características próprias em cada comunidade escolar. (LIN e ADOLPHS, 2011).

Amaya Mendiola, Susana Murcia Bielsa e Paul Rollinson da *Universidad autónoma de Madrid* publicaram um trabalho, fosse o a Análise Contrastiva e Análise de Erros, dentro da abordagem do ensino de língua baseada em *corpus*. Nesse trabalho, distinguem duas formas pelas quais o trabalho com *corpus* pode ser um grande aliado para auxiliar no processo ensino/aprendizagem. O primeiro seria identificar as características próprias da língua em estudo e desenvolver materiais direcionados, especificamente, para esse contexto. O Projeto INTELeNG, esse é seu nome, pretendia construir um banco de dados de erros extraídos de um *corpus* para identificar os vários tipos de problemas que surgissem, avaliar o nível de dificuldade encontrada e desenvolver matérias didáticos adequados à situação (ROLLINSON *et al.*, 2010).

O Projeto INTELeNG mostra como um pequeno *corpus* de aprendiz pode auxiliar na elaboração de materiais pedagógicos que sejam mais eficazes, uma vez que resolvem problemas pontuais, de determinado grupo de aprendizes, pois se cada indivíduo é único, cada grupo também o é. O conteúdo recebido através dos livros didáticos, pelas escolas, nem sempre surte efeito de aprendizagem para todos os grupos de aprendizes uniformemente, uma vez que são gerais e não visam a nenhum grupo especificamente. Ao detectar o problema de uma determinada comunidade linguística escrita, a elaboração de material didático direcionado para essa comunidade pode resultar em solução para aquele problema que é específico daquela comunidade. Segundo Rollinson *et al.* (2010), “Nossa hipótese é que ao compreender as dificuldades dos aprendizes, professores e materiais de ensino possam ajudar os estudantes a melhorarem” (p. 181).

#### 3.1.5.1 Implicações dos *Corpora* de Aprendiz

Ao abordar o texto a partir da perspectiva do aprendiz fica claro que a concepção que cada grupo linguístico tem do ensino de língua nem sempre coincide com a que a agência de letramento principal, que é a escola, espera dele. Na reserva indígena, foco deste trabalho, o modelo de ensino de língua é o mesmo que o da escola não indígena, não obstante o contexto linguístico e cultural serem diferentes. Os textos dos indígenas são uma tentativa de se encaixar nos moldes das exigências do ensino formal não indígena. O desafio é entender essa escrita para propor uma concepção e uma metodologia que preencha as lacunas deixadas pela escolarização formal.

O fato de a maioria desses indígenas terem como língua materna o guarani e, outros em menor número, o terena, pode explicar boa parte da forma como se relacionam com a língua portuguesa escrita (suas análises estarão nos capítulos 6 e 7 deste trabalho). Para Lin e Adolphs (2011), os critérios para elaborar um *corpus* de aprendiz obedecem a critérios um pouco diferentes dos estabelecidos para elaborar um *corpus* de falante nativo, pois leva-se em conta o nível de consistência dos recursos que ele utiliza, ou seja, de conhecimento que o falante tem da língua, seu nível de proficiência, bem como seu nível de conhecimento da língua materna.

Para Sardinha (2004, p. 255) o uso de *corpora* no ensino se resume a quatro áreas de atuação:

- a) Descrição da língua materna.
- b) Descrição da língua do aprendiz.
- c) Transposição de metodologia de pesquisa acadêmica para a sala de aula.
- d) Desenvolvimento de abordagens de ensino e material didático.

A primeira área desses estudos trata da descrição da língua materna. Entretanto, os professores que atuam em sala de aula estão bastante distantes das universidades e, nem sempre têm acesso a essas pesquisas. Portanto, não utilizam a descrição da linguagem baseada em *corpus* em suas atividades de ensino. Essa abordagem, entretanto, seria bastante produtiva para seus alunos, pois construiriam outra perspectiva do ensino/aprendizagem da língua e da variedade de língua que utilizam.

A segunda área de atuação desses estudos restringe-se ao ensino de língua estrangeira. Trata-se de uma vertente também bastante restrita ao ambiente acadêmico ainda não disponível com clareza para ser empregada nas salas de aulas. No entanto, muito benéfica, uma vez que podem apontar alternativas de ensino para o aprendiz de uma segunda língua e, dessa forma, tornar a aprendizagem mais próxima da língua alvo real.

A terceira, e, em seguida a quarta área de atuação da Linguística de *Corpus* diz respeito a aplicabilidade das ferramentas computacionais na sala de aula, como o concordanciador. Esta está estreitamente ligada ao objetivo deste trabalho, que é o de após constatar as questões elencadas para análise nos textos, pensar em possibilidades de produção de materiais didático-pedagógicos apropriados para este contexto. Para Sardinha (2004), essa área pode auxiliar na:

criação de metodologias ou abordagens de ensino inspiradas na exploração de corpora ou em conceitos da Linguística de *Corpus*. Nesse sentido, há pelo menos três propostas de utilização no ensino: o Currículo Lexical, a Abordagem Lexical e o Ensino Movido a Dados (DDL).

Na perspectiva de Sardinha (2004), um *corpus* de aprendiz pode ser utilizado com dois propósitos fundamentais: o primeiro, para a extração de partes de textos ou de textos inteiros com finalidade de orientar o ensino/aprendizagem em contextos de ensino de sala de aula, bem como na elaboração de materiais didático-pedagógicos que sirvam de suporte para esses contextos. O segundo, como uma tentativa de reaver os contextos como um todo para descrever a língua real desses ambientes, bem como contribuir no desenvolvimento de atividades de ensino.

Granger (2002) foi a primeira pesquisadora a relacionar a pesquisa em *corpus* de aprendiz com aquisição de segunda língua e ensino de língua estrangeira, destacou a importância de se estabelecer critérios rigorosos para esse tipo de investigação linguística. A partir disso, surgiram as primeiras discussões acerca da utilização de *corpora* eletrônicos nessas pesquisas. A comparação entre a língua do nativo e a língua dos aprendizes de segunda língua, assim como entre os diferentes tipos de aprendizes de uma segunda língua, que é essencial nos dias de hoje para avançar na melhoria do ensino, haja vista que, os possíveis “erros”<sup>12</sup> cometidos pelos aprendizes podem ser compartilhados, promovendo um novo olhar e, conseqüentemente, uma nova forma de ensinar, são devido aos estudos de Granger. Além disso, ela introduziu os diferentes tipos de análises linguísticas, os quais podem ser utilizados para conseguir realizar essas comparações com eficiência e precisão.

Na defesa dos trabalhos com Linguística de *Corpus*, Granger (2002) defende que, embora os trabalhos com LdC sejam apenas uma fonte de evidência entre tantas outras, ela é a única fonte confiável que se trata de detectar características como frequência, pois se tem muito pouca consciência intuitiva sobre essa questão. As frequências demonstram não apenas do que é possível na língua, mas do que é provável de ocorrer (Granger, 2002, p. 04).

Um dos argumentos mais fortes para o uso de *corpora* eletrônico é a autenticidade. Sinclair (1996, apud Granger) esclarece que todo o material para a pesquisa é coletado de comunicações ocorridas espontaneamente, em momentos em que as pessoas conversam sem se preocupar em cuidar a linguagem. Granger, porém, chama a atenção para o fato de que na questão do ensino de língua estrangeira, não seria possível estabelecer uma equidade entre a espontaneidade de um falante não nativo e um nativo. Ela continua essa ideia afirmando que numa situação de sala de aula, por exemplo, pode-se até criar uma espontaneidade na tarefa de escrita ao solicitar aos alunos que escrevam sobre o que quiserem, ou seja, um tema livre. Todavia, será uma situação artificialmente criada pelo pesquisador com uma intenção própria.

Os *corpora* de aprendizes, então, operam com diferentes graus de autenticidade, desde uma conversa informal sobre questões envolvendo o ambiente de trabalho do informante quanto uma atividade solicitada pelo professor em sala de aula. Na coleta deste trabalho optou-se por solicitar aos alunos (crianças e universitários) três textos, nos quais expusessem livremente suas ideias acerca de família e cultura, divididos em três temas. Os

---

<sup>12</sup> A palavra “erros” é utilizada como desvio da norma padrão, sinônimo da norma ensinada pela escola.

textos foram extraídos da realidade desta comunidade indígena, representem a escrita como esse grupo a utiliza. Assim, na apreensão das porções de linguagem *reais* se poderá compreender melhor como funciona sua linguagem escrita proporcione condições de visualizar futuras opções de trabalho em sala de aula ou até com professores.

Outra questão importante a ser lembrada com relação a *corpora* de aprendiz é que sua elaboração deve partir de trechos contínuos de textos e não de palavras isoladas. Também deve privilegiar trechos escritos ou falados dentro dos padrões gramaticais estipulados como *certo*<sup>13</sup> e *errado*, ou seja, padrões gramaticais aceitos e não aceitos pela norma escolar (Granger, 2002). A ideia é trabalhar com a língua utilizada pelo usuário dessa língua, não importa se ela está dentro das normas escolares ou não. O que importa são os objetivos de cada pesquisa. Os *corpora* eletrônicos podem ser utilizados e serem muito úteis, mas para trabalhos específicos, como é o caso deste, a coleta do próprio material se torna bastante interessante.

### 3.1.6 Termos Técnicos da Linguística de *Corpus*

#### 3.1.6.1 A Palavra na LdC

Não seria pertinente trazer o termo Linguística de Corpus à discussão sem considerar o termo “palavra”, ou seja, “o reconhecimento dessa unidade como uma sequência de letras precedida e seguida por um espaço” (VIANA, 2011) nem sempre é uma leitura isenta de problemas. O conceito de palavra é essencial para essa perspectiva dos estudos linguísticos. A palavra, como o computador a lê, pode oferecer alguns contratempos, como, por exemplo, a palavra hifenizada (VIANA, 2011). O pesquisador precisa estar bastante atento na hora de classificar essa palavra, definir que papel ela tem quando colocada junto a outra, pois conforme sua posição e sua companhia, poderá mudar o sentido, a função ou a classe gramatical.

Na LdC, a palavra possui dois conceitos mais específicos. Vejamos o exemplo mais claramente no texto abaixo, retirado de um total de mais de 800 textos coletados na Reserva Indígena Francisco Horta Barbosa, em Dourados/MS. O texto é de um aluno do 5º ano do

---

<sup>13</sup> A palavra *certo* está sendo utilizado como antônimo de *errado*, no sentido de desvio da norma utilizada pela escola. O *certo*, portanto é tudo aquilo que está estipulado pela normatização da instituição escolar e não as demais variedades da língua utilizadas pelos falantes nos vários contextos nos quais emprega a língua.

ensino fundamental da Escola Municipal Indígena Ramão Martins. Foram solicitados três textos, sendo o de número 1 na tabela abaixo:

Tabela 3. Lista de Palavras

**Tokens:** 120

**Types:** 73

Número	Frequência	Palavra
1	1	A
2	1	aldeia
3	1	alguns
4	1	amado
5	4	arvore
6	1	As
7	1	chefe
8	1	chegan-la
9	1	choram
10	1	colocaram
11	1	com
12	1	contou
13	1	crianças
14	3	Da
15	1	De
16	1	depois
17	1	desobedeceu
18	1	desseram
19	1	Dia
20	1	dias

Na LdC as palavras são denominadas *item* (*token* em inglês) e *formas* (do inglês *types*). As duas tabelas apresentadas aqui foram geradas a partir do programa eletrônico TEXTQUIM/UFES. A primeira, Tabela 3, está disposta em ordem alfabética e a segunda, Tabela 4 (abaixo), está disposta por ordem de frequência. Essas opções são dadas pelo próprio programa.

Tabela 4. Lista de palavras

**Tokens: 120 Types: 73**

<b>Número</b>	<b>Frequência</b>	<b>Palavra</b>
1	8	e
2	7	o
3	6	menino
4	5	os
5	5	ele
6	4	arvore
7	3	do
8	3	da
9	3	pais
10	3	então
11	2	numa
12	2	era
13	2	foi
14	2	guaraná
15	2	foram
16	2	que
17	2	um
18	2	índios
19	2	uma
20	2	viram

Nessa tabela, há 120 itens (*tokens*), sequência de letras separadas por espaço ou por algum sinal de pontuação, ou seja, é cada uma das ocorrências no texto acima e 73 formas (*types*), que são as palavras diferentes que aparecem no *corpus*. Desses 120 itens, 20 são palavras que ocorrem mais de uma vez. Na tabela 3, os itens aparecem por ordem alfabética. Nesse caso, o pesquisador tem mais trabalho para procurar a frequência, o que não ocorre na tabela 4.

Na tabela 4, nota-se que 20 palavras se repetem com frequências diferentes. Segundo Viana (2011, p.38), para se contabilizar as formas não se considera a frequência bruta das palavras. Apesar de totalizarem 67 instâncias, elas são contadas como apenas 20 para o total de formas. Resumindo, o total de itens corresponde ao total de palavras encontradas num texto/*corpus* e o total de formas corresponde à quantidade de palavras distintas que ocorrem neles. As palavras que ocorrem apenas uma única vez são denominadas de *hapax*. No texto há 53 *hapax*.

Outros conceitos como lema e lematização também são característicos da LdC. Sarmiento (2008) corrobora com outros pesquisadores na definição de lema, exemplificam com as formas verbais *is* e *be*. Para ambos, *is* e *be* são partes da mesma palavra, sendo que ambos são apenas conjugações diferentes do mesmo verbo. Nesse sentido, pertencem ao mesmo lema. Isso não ocorre quando as palavras pertencem a classes gramaticais diferentes, como *vestir* e *vestido*, por exemplo. A lematização torna-se relevante para a LdC, segundo Sarmiento, apenas quando se quer observar substantivos, excluindo adjetivos e verbos. Para apreender a lematização, o *corpus* deve estar anotado, ou seja, ao lado de cada palavra que se quer analisar deve aparecer a etiqueta substantivo, adjetivo, verbo, etc.

A anotação consiste na etiquetagem manual ou pelo computador de um *corpus* a fim de acrescentar informações às palavras desse corpus. A anotação pode ser morfossintática (*Part-of-speech tagging* ou *POS Tagging*) ou sintática (*parser*). Na etiquetagem morfossintática, é anexado um código criado especificamente pelo pesquisador para cada palavra, indica sua classe gramatical. Já na etiquetagem sintática, estabelece-se a análise de um texto, a partir da identificação de sua estrutura sintática, como sintagma nominal, verbal, etc. No presente trabalho utiliza-se a etiquetagem sintática, ou seja, os códigos junto às palavras indicam de que tipo de sintagma se trata<sup>14</sup>.

A importância da etiquetagem, no entanto, não é partilhada por todos os pesquisadores que trabalham com pesquisas *corpus-driven*. Para Sinclair, (2004, apud SARMENTO, 2008), por exemplo, um dos teóricos que se coloca contra o uso de anotações, ela só tem utilidade de fato quando se objetiva realizar traduções automáticas, ou compor dicionários eletrônicos. Já, Houston (2002, apud SARMENTO, 2008) questiona se os próprios pesquisadores que utilizam as anotações têm a noção exata das suposições teóricas que sustentam as formas distintas de anotação. O consenso é de que uma versão original do *corpus* sem anotação nenhuma deve ser disponibilizada ao público, além do fato de que os pesquisadores devem estar cientes dos princípios teóricos que regem as anotações (SARMENTO, 2008).

---

<sup>14</sup> As etiquetas empregadas na análise serão apresentadas no Capítulo da Metodologia.

### 3.1.6.2 Combinações de Palavras em *Corpora*

Viana (2011) afirma que não é apenas com palavras que a LdC se ocupa, mas com suas combinações em *corpora* representativos de textos orais e escritos. O *modus operari* da LdC, nas palavras de Shepherd (2012), desde suas primeiras aparições no cenário acadêmico, combina método de investigação com habilidade de ler linhas de concordância e listas de palavras que são geradas por programas eletrônicos desenvolvidos, especificamente, para essa finalidade. Essa metodologia de trabalho inaugura uma nova forma de se fazer pesquisa linguística, devido ao fato de não separar léxico de gramática, direcionando para perspectivas diferenciadas de estudos da linguagem como a colocação, a coligação, a prosódia semântica e a preferência semântica. Segue abaixo, a descrição de cada uma delas.

#### Fraseologia

Para a LdC, o importante não é a descrição de combinações possíveis de se concretizarem, mas sim, ao contrário, interessa à LdC analisar as combinações ou padrões mais prováveis de ocorrerem. Isso significa que a língua é aqui concebida como probabilidade. Nesse sentido, segundo Sinclair (1991, apud SARMENTO, 2008), a utilização de uma língua possui dois princípios norteadores: o *princípio da livre escolha* (*open-choice principle*) e o *princípio idiomático* (*idiom principle*). Afirma ainda que não há distinção entre o padrão sintático e o semântico. É necessário, assim, lançar mão desses dois princípios distintos para auxiliar na interpretação de um texto.

O *princípio da livre escolha* determina que cada falante/ouvinte escolhe livremente e de forma sequencial as palavras que quer utilizar. A única restrição desse princípio é a gramaticalidade, isto é, a necessidade de que haja concordância com as regras gramaticais vigentes na língua empregada.

O princípio idiomático respalda na ideia de que o usuário da língua tem disponível um grande número de frases pré-construídas que se constituem de uma única escolha, ou seja, a reiteração e a repetição são uma constante, não havendo inovação. Graças ao princípio idiomático, surgiu o campo da Fraseologia, que faculta a observação dos padrões presentes na língua, subestimado até esse momento. Conformes Tagnin (2011), a Fraseologia constitui no estudo “de qualquer tipo de ocorrência fraseológica de uma língua como, por exemplo, as colocações, os binômios, as expressões idiomáticas, etc.” (p. 359).

### 3.1.6.3 Técnicas e Análises em *Corpora*

Os trabalhos com *corpora* não seriam possíveis, na magnitude em que são feitos hoje, sem o auxílio do computador. Isso por que a quantidade de dados é exorbitante para que mãos humanas dessem conta de trabalho tão hercúleo. Programas computacionais próprios para esse fim surgiram no decorrer do desenvolvimento das pesquisas com *corpus* e hoje são vários. Alguns *corpora*, por exemplo, possuem programas próprios, mas há também programas independentes que podem ser baixados ou comprados via *web* e serem utilizados pelo pesquisador para quantificar e interpretar seus dados. Os dados também podem ser colhidos de *corpora* eletrônicos ou serem coletados e depois digitados e organizados num *corpus* para o próprio pesquisador. É o caso desta tese, em que os textos foram coletados e depois digitados no computador.

Nesta pesquisa, optou-se pela utilização do programa *AntConc*. O programa foi adquirido via *web* e instalado no computador para uso doméstico. A ferramenta utilizada foi o concordanciador, que será descrita no capítulo da metodologia.

No próximo capítulo, será apresentada a discussão sobre os estudos gramaticais que foram utilizados como base para a construção da presente pesquisa.

#### 4. A ANÁLISE GRAMATICAL

Na busca de suporte teórico para esta pesquisa, muitos caminhos foram pensados. A grande questão que se fazia sempre presente, porém, era a do ensino oferecido para os alunos/acadêmicos indígenas. Desde 1996, pesquiso e realizo projetos de extensão na reserva juntamente com alunos indígenas em formação no curso de Letras. Durante esse período, pude observar a prioridade da tradição gramatical no ensino de língua portuguesa. Apesar de já existir desde a Constituição de 1988 e posteriormente, pelos *Referenciais Curriculares para as Escolas Indígenas* (RCNEI), uma orientação para o ensino diferenciado intercultural e bilíngue para as etnias indígenas, o ensino atual das aldeias ainda está se adequando e se reestruturando para atender às novas demandas exigidas por essas comunidades.

Pensando na realidade escolar da Reserva Indígena Francisco Horta Barbosa (RFHB), não há outro caminho teórico que possa dar conta de uma pesquisa realizada a partir de uma configuração real de ensino que não sejam os estudos gramaticais. Optou-se, então, por escolher estudos gramaticais que tivessem um viés contemporâneo e que privilegiassem a língua em uso, aquela que ocorre em situações reais de fala e escrita. Por isso, a escolha pelos estudiosos da língua em uso, principalmente Bagno (2011) e Castilho (2010; 2011), cujos trabalhos descrevem o português brasileiro contemporâneo divorciado das questões gramaticais do português europeu. Nesse sentido, o português brasileiro é descrito a partir de suas variações diatópicas ou diastráticas, com sua história própria, com exemplos extraídos da norma urbana culta do povo brasileiro (Projeto NURC, Castilho 2010). Além disso, os critérios adotados para a análise dos dados obtidos na presente pesquisa partem dessa visão do português falado/escrito culto, ou norma urbana culta (NURC), que foram submetidos à ferramenta *Concord* do programa *AntConc*.

#### 4.1. As Investigações Linguísticas

Desde os primórdios dos estudos linguísticos, tem sido possível perceber que a língua varia e, às vezes, dessa variação resulta mudança. Pesquisas na área da sociolinguística, assim como o maior rigor científico que os estudos variacionistas receberam a partir dos anos 60 do século passado, mostram cada vez mais que a língua se encontra num processo contínuo de variação, que não necessariamente acarretam mudança. Entretanto, sabe-se que toda mudança ocorre a partir de uma variação. Essas ideias formam a base fundamental para uma virada na visão que se tinha do ensino até então. Questões como preconceito linguístico e as noções de *certo* e *errado* na língua se relativizaram e passaram a ser cada vez mais discutidas. Não basta mais à escola ensinar regras do *bem falar* e do *bem escrever*. Com essa nova visão, outras discussões surgiram como, por exemplo, quais seriam os parâmetros para estipular uma padronização para o ensino da língua. O modelo não poderia mais ser o português de Portugal, ou a escrita dos escritores consagrados do século XIX.

Não tardou a se perceber que mais importante que ensinar questões relacionadas diretamente à gramática normativa da língua, a escola teria que promover o letramento de seu alunado. Para Bagno (2011, p.20), “para essa promoção do letramento, as atividades fundamentais são a *leitura* e a *escrita*, com foco na diversidade de gêneros textuais que circulam na sociedade”. Além disso, deve-se também ter um foco epilinguístico, nas palavras de Bagno, “a partir de textos autênticos, falados e escritos, dos quais se possa depreender o funcionamento da língua na construção dos sentidos” (p.20). O autor defende que o aspecto a ser valorizado deve ser o semântico-pragmático-discursivo, valorizando as reflexões sobre a gramática. Assim, haverá respaldo para mostrar ao aprendiz que os recursos que são ativados pela língua produzem os sentidos e a interação social. Logo, “é do uso que se depreende a gramática” (p. 20). Essa convicção também está presente em Castilho (2011), segundo o qual:

Quando escrevemos ou quando falamos, uma intensa atividade é desencadeada em nossas mentes. Isso ocorre com enorme rapidez, acionando quatro sistemas linguísticos, cada um deles configurado por um elenco de categorias: o léxico, a semântica, o discurso e a gramática. Esses sistemas são articulados pelos princípios sociocognitivos que regem a conversação (CASTILHO, 2011, p. 32).

Segundo o Castilho (2011), uma boa gramática é aquela capaz de revelar esse conhecimento linguístico que está armazenado na mente de cada usuário da língua, do analfabeto ao mais erudito dos escritores. Dessa forma, percebe-se que é do discurso que se

chega às regularidades da língua, dado que discurso e gramática se imbricam num processo de codependência, estando um inserido no outro. Ao separá-las, o ensino cai na taxionomia pura e simples, sem sentido para o usuário da língua (BAGNO, 2011, p. 21).

Longe de afirmar que não se deve ensinar gramática na escola, deve-se mostrar que existe gramática nos usos da língua, na sociointeração. É necessário, contudo, mostrar como essa gramática se desvela independentemente da nomenclatura gramatical tradicional. Por isso esta pesquisa vai prescindir dos estudos gramaticais como base para sua análise.

#### 4.2. A Gramática – Algumas Considerações a partir de Castilho (2010; 2011)

O conceito de gramática exposto por Castilho, no prefácio de sua *Nova Gramática do Português Brasileiro* (2010), afirma que tanto o gramático quanto o linguista “operam com um objeto guardado em sua mente e na mente dos indivíduos de sua comunidade, lidando com uma propriedade interna a ele, não evidente no mundo real” (p.41), assim como fazem os antropólogos, por exemplo. Isso significa que antes de iniciar qualquer trabalho de ordem linguística é estabelecida uma teoria, e o pesquisador disponibiliza de antemão uma visão prévia sobre as línguas e as gramáticas. Por isso, há tantos conceitos e definições para língua e para a gramática. Segundo ele, cada pesquisador aborda o assunto, a partir de sua própria perspectiva, não significando que uma exclua a outra. Ao contrário, comenta, pois cada um aborda um aspecto que faz parte de um todo. O autor divide as teorias linguísticas em quatro grandes blocos (CASTILHO, 2010, p. 42):

1. Língua como um conjunto de produtos estruturados;
2. Língua como um conjunto de processos estruturantes;
3. Língua como um conjunto de produtos e processos em mudança;
4. Língua como um conjunto de usos bons.

Interessa para a descrição linguística proposta nesta investigação a proposição de número 1, uma vez que apresenta a língua do ponto de vista do produto, interpretando a língua como um conjunto de enunciados hierarquicamente dispostos. A análise realizada parte de textos escritos, ou seja, do produto final. Para o momento da análise, o foco recairá sobre o sistema linguístico (morfologia, sintaxe), pois nesse momento ele tem papel relevante, ou seja, o contexto não é levado em consideração. No entanto, um dos postulados dessa concepção denominada gramática descritiva da língua, a saber, a concretude das

unidades fonológicas, morfológicas e sintáticas configuradas na sentença, depende da variação de uso, tendo em vista fatores linguísticos e extralinguísticos.

Na apresentação de sua *Gramática de usos do português*, a pesquisadora Moura Neves declara que “(...) é no uso que os diferentes itens assumem seu significado e definem sua função, e de que as entidades da língua têm de ser avaliadas em conformidade com o nível em que ocorrem, definindo-se, afinal na sua relação com o texto” (2011, p. 13). Trata-se, na verdade, de uma das metas de sua gramática, que também serve aqui para respaldar teoricamente a discussão dos resultados desta tese, uma vez que os textos selecionados cumprem a função de permitir que seus itens sejam apreendidos em uso real. Assim, tem-se uma língua viva, em pleno funcionamento, “exibindo todas as possibilidades de composição que estão sendo aproveitadas pelos usuários para obtenção do sentido desejado em cada instância” (idem, p. 13).

Não se trata de uma abordagem de ordem prescritiva. Ao contrário, trata-se de uma abordagem descritiva da língua que aparece nesses textos, nesse contexto histórico. Castilho (2010, p.47) argumenta que “a descrição linguística implica na identificação das unidades de uma dada língua, mediante a segmentação e a comutação dos enunciados recolhidos no *corpus*, designando-os por uma nomenclatura própria”. A segmentação ou contraste ocorre no eixo sintagmático da língua e a comutação ou oposição se dá no eixo paradigmático, utilizando uma linguagem técnica estruturalista. Na subseção abaixo será detalhado o eixo sintagmático, por ser a parte visível na estrutura superficial, ou seja, as palavras, frases, períodos. Na discussão dos resultados, o eixo paradigmático vai surgir, pois as escolhas dos alunos e dos acadêmicos por uma palavra ou uma expressão pode ter alguma relevância para a sintaxe desse texto.

#### 4.3. Unidades da Gramática Descritiva - O Sintagma

Na perspectiva de Castilho (2010), há quatro unidades lexicais passíveis de análise: o fonema, o morfema, a palavra e o sintagma. Interessa para esta pesquisa o sintagma que, nas palavras do próprio Castilho,

(..) provém da terminologia militar grega, em que designava um esquadrão, ou seja, um número fixo de soldados, distribuídos de forma também regular aos quais eram atribuídas funções próprias. Os linguistas se apropriaram desse termo, que parecia talhado para indicar o modo como o substantivo, o verbo, o adjetivo, o advérbio e a preposição costumam agregar outras classes de palavras. (...) O estruturalismo especializou o termo, restringindo-o à designação dos grupos de palavras que

formam uma unidade sintática hierarquizada maior que uma palavra, pois resulta de uma associação de palavras, e menor que a sentença, de que é um constituinte (CASTILHO, 2010, p. 56).

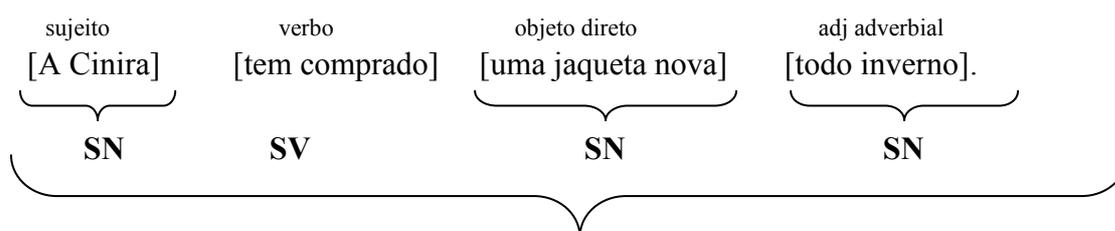
Para Castilho (2010), a regra descritiva do sintagma nominal é a seguinte:

#### QUADRO 2 - Regra Descritiva do Sintagma Nominal

SN → (Especificadores) + Núcleo + (Complementadores)
--

O sintagma pode ser nominal, verbal, adjetival, adverbial ou preposicional. No sintagma nominal, a construção sintática tem por núcleo um substantivo ou um pronome, em que o primeiro é uma classe basicamente designadora, e o segundo uma classe dêitica/fórica/substituidora (CASTILHO, 2010, p. 453). Os núcleos são representados pelos substantivos, pronomes pessoais, pronomes neutros e pronomes adverbiais. Os especificadores são os artigos, demonstrativos, possessivos, e por quantificadores os sintagmas adjetivais e preposicionais.

Bagno (2011) define sintagma como um constituinte menor do que a sentença, possuidor de um núcleo em torno do qual se situam elementos à esquerda e à direita. O núcleo pode ser preenchido por verbo, substantivo, adjetivo, advérbio ou preposição. Veja o exemplo abaixo, retirado de Bagno (2011, p. 449):



No sintagma verbal, temos um núcleo em torno do qual se agrupam três sintagmas nominais, sendo que um está à esquerda, outro à direita e o terceiro funciona como seu complementador (uma jaqueta nova). Cada sintagma nominal, por sua vez, possui seu próprio núcleo: no da esquerda *Cinira* é núcleo, no da direita, *inverno* e *jaqueta* são complementadores, todos cercados pelos respectivos determinantes e complementadores.

O estudo dos sintagmas é importante, por exemplo, para se compreender bem como funcionam as regras de concordância, pois mesmo que os constituintes estejam afastados na

sentença os sintagmas serão reconhecidos como nominal, verbal, etc, pelos seus constituintes. Para Bagno, a língua portuguesa ou o português brasileiro, como ele prefere denominá-la, está cada vez mais está se caracterizando como uma língua sintagmática, isto é, uma língua que emprega várias palavras para desempenhar uma mesma função sintática.

Tanto nas variedades linguísticas estigmatizadas quanto nas variedades urbanas de prestígio mais espontâneas a concordância de número e gênero se dá no plano sintagmático, uma vez que a flexão do substantivo acaba por ser redundante.

#### 4.3.1 A Gramática - Algumas Considerações a partir de Bagno

O pensamento ocidental deve à filosofia platônica o pensamento dicotômico. A partir da dualidade do mundo sensível x mundo inteligível criou-se toda uma tradição que perpassou todos os tempos e se ancora no momento atual. Saussure e Chomsky também beberam nessa fonte, pois basearam seus estudos sobre verdades dicotômicas. É afirmando isso que Bagno começa seu capítulo sobre a gramática em sua *Gramática Pedagógica do português Brasileiro* (2011).

Em sua visão, a gramática tradicional também adotou as dicotomias. Bagno cita a visão bipartida que os filólogos alexerinos concebiam da língua, separando-a em fala e escrita, valorizando a primeira como língua de fato, pela sua suposta homogeneidade e por isso, deveria servir como exemplo, como regra para se falar e escrever bem. A fala era o lugar dos erros e dos equívocos e não deveria servir de modelo. Para Bagno, essa ideia é um grande equívoco, pois não se pode comparar duas modalidades tão diferentes da língua.

Nossa tradição escolar está toda baseada nessas dicotomias. O resultado mais pernicioso foram as noções de *certo* e *errado* que grassa por todo o sistema escolar até hoje. Mais adiante, Bagno lembra que a palavra “gramática” (*grammatiké*) tem origem grega derivada de *grámma* (em latim *littera*) que inicialmente significava “letra” e mais tarde, os filósofos a descreviam como “uma entidade com três propriedades: nome, forma (ou figura) e som (ou valor)”. No entanto, para Bagno (2011, p.403),

No atual estágio dos estudos linguísticos, é fundamental distinguir os sons da fala de suas representações na escrita, as letras. Um estudante de linguística, assim que começa seu curso, é (ou deveria ser) instruído a fazer claramente essa distinção e a não se deixar levar pela tradicional confusão entre as duas entidades. E o mesmo vale, é claro, para quem vai ensinar qualquer língua e principalmente a língua materna (BAGNO, 2011, p. 403).

Bagno esclarece, porém, que os filósofos gregos não pretenderam criar nenhum tipo de regra para a escrita ou mesmo para a fala. Seus estudos eram filosóficos, questionadores, por isso Platão distinguiu na proposição daquele que fala duas partes principais: o argumento e a predicação que incide sobre ele. Para Platão a atividade comunicativa se resumia à predicação, ou seja, “falar de” e “dizer que” (BAGNO, 2011, p. 406). Essa divisão (*ónoma e rhēma*) não distinguia entre categorias lexicais e funções sintáticas, interessava-se pela natureza das palavras e não pela sua forma.

Quando se estuda a gramática da perspectiva filosófica:

Só podemos reprovar a prática tradicional de ensino, que primeiro apresenta as classes de palavras, isoladamente para, mais adiante, às vezes só no ano escolar seguinte, tratar das funções que elas podem exercer no enunciado. Essa prática é frequentíssima nos livros didáticos e só serve para transformar o estudo da gramática num amontoado de conceitos desconexos, de definições isoladas umas das outras, de gavetas autossuficientes, onde o que está dentro de uma parece passar muito bem sem o que está nas outras... (BAGNO, 2011, p. 403).

Segundo Bagno (2011, p. 22), dados do INAF/BRASIL (Índice Nacional de Alfabetismo Funcional) revelam que 75% da população brasileira entre 14 e 64 anos não consegue ler e compreender um texto com grau de dificuldade média e apenas 38% dos que completaram o ensino médio estão plenamente alfabetizados. Isso mostra o quanto o modelo de ensino formal adotado pela escola brasileira vem exigindo mudanças.

O ensino gramatical tão desdenhado e criticado nos últimos tempos recebe agora uma nova feição. Segundo Bagno (2011, p.23), “a gramática tradicional tem que ser muito bem conhecida por aqueles que, profissionalmente, serão confrontados a ela – cobrados para que a ensinem e desafiados a dizer por que não a ensinam”. Apesar de seu teor altamente prescritivo, a gramática tradicional é o repositório de relevantes reflexões filosóficas e filológicas que inegavelmente contribuíram para o conhecimento e aprofundamento da linguagem humana, bem como das línguas.

Então, faz-se necessário ao profissional que trabalha com a língua conhecer essa gramática, para compreender suas regras ou refutá-las ou, ainda, reorganizá-las conforme novos pressupostos que venham a surgir. Segundo o mesmo autor, por trás da palavra língua “se esconde um universo multidimensional de saberes, práticas, crenças, histórias, ritos, instrumentos, mecanismos de poder, papéis sociais, tradições culturais” (p. 28). Afirma ainda que boa parte desses saberes são aprendidos em casa, com a família, outra parte aprendemos na escola, a principal agenciadora de ensino até o momento. Por isso, a escola

deve ser o lugar de interação e troca de experiências e saberes de natureza diversa, dentre eles linguísticos, culturais e históricos que cada um traz consigo de seu lugar de origem.

Bagno reitera também que a escola é a agência responsável pelo letramento escolar, uma vez que o próprio aprendizado da língua materna desde a primeira infância constitui um processo de letramento, a aquisição da gramática da língua materna, a formação de uma intuição linguística, a capacidade de reconhecer uma manifestação linguística como bem ou malformada, como pertencente ou não à própria língua – tudo isso também é letramento. E mesmo que se refira a letramento como a aprendizagem da leitura e da escrita, há pessoas que não precisam frequentar escola para aprender a ler e escrever. No entanto, o número dessas pessoas é muito pequeno quando comparado com a multidão daquelas que só vão fazer essa aprendizagem no ambiente escolar (BAGNO, 2011, p. 28).

Uma grande confusão que se faz com relação ao ensino da língua nas escolas, segundo Bagno, é a de “ensino da norma-padrão” que, para ele,:

(...) é um modelo de língua descrito-prescrito pela tradição gramatical, uma língua extremamente idealizada, construída com base nos usos de um grupo não muito amplo de escritores, e, mesmo assim, não de todos esses usos, mas só daqueles que o próprio gramático considera exemplares ou recomendáveis (BAGNO, 2011, p. 31).

Essa norma, argumenta Bagno, é arcaica, baseada apenas na escrita, não em qualquer escrita, mas na escrita literária, mais ainda, escrita literária do passado, de outro momento histórico, sincronicamente marcada, ou seja, que hoje não se usa mais. Além disso, é uma língua na qual o escritor pode usar de sua licença poética para criar, recriar e brincar com a linguagem dentro de um mundo imaginário, irreal. Trata-se de uma língua idealizada e não de uma língua correntemente utilizada por seus usuários.

Para esse autor, o ensino da língua materna não deve se restringir a uma única modalidade. Ao contrário, “educar em língua materna é permitir o acesso dos aprendizes ao maior número possível de modalidades faladas e escritas de sua língua” (p.31). Essa língua deve ser mostrada dentro dos gêneros textuais vivos, presentes nas várias mídias, faladas ou escritas, pois dessa forma, os aprendizes terão acesso à multiplicidade de seus possíveis usos (p. 32). À escola cabe o papel de privilegiar o ensino dos gêneros escritos mais monitorados para mostrar o funcionamento real da gramática da língua, pois a partir deles o aluno será levado a conhecer, reconhecer, dominar e empregar a língua portuguesa com eficácia e precisão exigidas pela prescrição gramatical oficial, além do que são esses gêneros que serão cobrados dele em situações formais.

A seção subsequente vai dispor de cada um dos critérios de descrição da análise que foram selecionados para nortear a discussão dos resultados, a começar pelo verbo, seu conceito, características e funcionalidades na língua portuguesa, lembrando que serão tratadas apenas as categorias que interessam a este trabalho de pesquisa.

#### 4.3.2 O Verbo

A definição que Bagno (2011) tem de verbo não destoa muito de outros estudiosos da língua: trata-se de uma palavra nucleica que agrega outras ao seu redor. O verbo é tão importante para a língua que o conceito de sentença ou oração só se constitui como tal devido a sua presença. Na tradição gramatical brasileira, verbo é definido como uma palavra que indica ação, estado ou processo. Tanto para Bagno (2011) como para Castilho (2010), no entanto, seria mais adequado classificá-lo, a partir de seus caracteres sintáticos, semânticos e pragmáticos.

Na definição trazida por Bagno (2011), morfossintaticamente o verbo se constituiria de um radical e de sufixos. Os sufixos se compõem de vogal temática + sufixo modo-temporal + sufixo número pessoal. O verbo *escrever* conjugado na 1ª pessoa do plural do pretérito imperfeito do subjuntivo, por exemplo, *escr + e + vê + sse + mos*, onde: *escr* é o radical; *e*, a vogal temática indicando que o verbo pertence a 2ª conjugação; *vê*, sufixo modo-temporal, que indica o modo (se indicativo, subjuntivo ou imperativo) e o tempo (se presente, pretérito ou futuro); *mos*, sufixo indicador de número (se singular ou plural) e pessoa (quem fala).

Semanticamente, o verbo se constitui como a expressão das ações, dos estados e dos eventos que são necessários para que a comunicação falada ou escrita passe a ter existência (BAGNO, 2011). Já a definição de verbo em seu caráter discursivo coloca o texto em foco: o papel que o verbo exerce no discurso (BAGNO, 2011; CASTILHO, 2010).

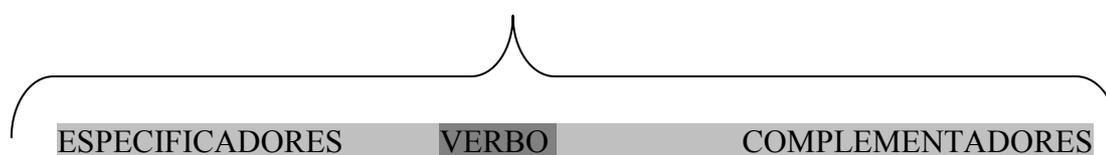
Bagno (2011) toma do estudo sobre os verbos, que Ilari e Basso (2008) chamaram de “funções dos verbos”, a ideia de que o verbo cria uma matriz ou molde em que dependendo da exigência desse verbo podem ser criadas várias opções de preenchimento. Dessa forma, seria então o verbo o responsável por selecionar seus complementos. Além disso, o verbo oferece perspectivas, a partir das quais os eventos se organizam, bem como a presença das pessoas necessárias para essa realização. Também é função do verbo permitir o olhar sobre

as pessoas do discurso, possibilitando sua identificação, bem como os papéis que desempenham.

Outra função essencial do verbo, contido em sua morfologia, é a noção de tempo que ele traz, não o tempo físico, mas o tempo ancorado no momento da fala, no discurso e fora da realidade empírica. Junto ao tempo a noção de aspecto, ou seja, dentro da categoria tempo pode haver vários aspectos desse tempo. A noção de modalização também é própria da categoria dos verbos (Bagno, 2011, *apud* Ilari e Basso, 2008).

#### 4.3.2.1 O Sintagma Verbal

O sintagma verbal é constituído de um núcleo, de especificadores e complementadores. Veja a representação, abaixo, retirada de Bagno (2011, p. 513).



Têm-se três posições de preenchimento. O núcleo, que será sempre um verbo, como posição central do sintagma, sendo um verbo principal, do ponto de vista sintático. À esquerda do núcleo estão os especificadores, que podem ser inclusive outros verbos, nesse caso verbos auxiliares. À direita, a posição será preenchida pelos complementadores, que podem ser objeto direto, objeto indireto e complemento relativo (ou oblíquo) (BAGNO, 2011, p. 513). Veja nos exemplos:

QUADRO 3 - As posições de preenchimento do Sintagma Verbal. (BAGNO, 2011, p.514)

	<b>CHOVIA</b>			
PODE ESTAR	<b>CHOVENDO</b>			
TEM	<b>CHOVIDO</b>	GRANIZO		
POSSO	<b>LEVAR</b>	ISABEL		
VOU	<b>LEVAR</b>	UNS DOCES	PRA ISABEL	
PRECISO	<b>LEVAR</b>	ISABEL	COMIGO	
PRETENDO	<b>LEVAR</b>	ISABEL	DA ESCOLA	PRA CASA

Por este exemplo, percebe-se que o verbo é como um sol. Ao redor dele se organizam todos os outros elementos da sentença. O verbo projeta sua luz sobre os demais de acordo com a necessidade. A diferença entre a sentença e o sintagma verbal “é que a sentença inclui o sujeito, que é sempre um sintagma nominal. Se o verbo é impessoal, o sintagma verbal e a sentença coincidem. Isso prova a importância central do verbo na língua” (BAGNO, 2011, p. 514).

#### 4.3.2.2 Atributos dos Verbos

No padrão falado ou escrito estipulado para a língua portuguesa, o verbo se divide em modo, tempo e voz<sup>15</sup>, que junto com o aspecto, se constitui como uma das categorias semânticas do verbo. Bagno (2011) explica que:

Em muitas línguas o modo está devidamente gramaticalizado em cada um dos *paradigmas morfológicos* que permitem ao falante expressar sua atitude com relação ao estado de coisas. Esse estado de coisas pode ser apresentado como real (*indicativo*), irreal (*subjuntivo*), desejável (*optativo*), sujeito a condições particulares (*condicional*), exigido ou solicitado a outrem (*imperativo e causativo*). Em algumas línguas, é possível distinguir morfológicamente o modo *interrogativo* e o modo *negativo* porque elas apresentam marcas próprias para expressá-los<sup>16</sup> (BAGNO, 2011, p. 556).

Inseridos nos modos estão os tempos, que, por sua vez, trazem aspectos. A diferença entre modo e aspecto é que enquanto o modo comporta um conjunto de tempos, o aspecto mostra uma visão específica do falante dentro de uma única perspectiva temporal. Castilho (2010, p. 418) corrobora a assertiva acima ao afirmar que “trata-se de um ponto de vista sobre o estado de coisas”, e acrescenta ainda, que apesar de o tempo e o aspecto poderem ser vistos no domínio da predicação, o tempo deve remeter a situação de fala para que se apreenda a anterioridade, a simultaneidade e a posterioridade dos fatos encadeados na sentença. Nesse sentido, defende o autor, “o tempo pressupõe o aspecto, mas este não pressupõe aquele” ( p. 417).

Aprende-se na escola que no português há três modos verbais: indicativo, subjuntivo e imperativo. No indicativo, temos os tempos presente, pretérito perfeito simples, pretérito imperfeito, pretérito mais-que-perfeito, futuro do presente, futuro do pretérito, além dos tempos compostos: pretérito perfeito composto, pretérito mais-que-perfeito composto, futuro

<sup>15</sup>Com relação a voz do verbo não será tratada aqui, uma vez que não haverá enfoque nessa categoria nesse momento. Optou-se pelas categorias de modo e tempo (aspecto).

<sup>16</sup>Grifos do autor.

do presente e futuro do pretérito compostos. No subjuntivo, temos: presente, pretérito imperfeito e futuro, além dos tempos compostos: pretérito perfeito, pretérito mais-que-perfeito, e futuro composto do subjuntivo. O modo imperativo conta de apenas o afirmativo e o negativo. Há ainda as formas nominais do verbo, que são infinitivo, gerúndio e particípio, que ora funcionam como verbos ora como nomes.

Sabe-se, entretanto, que nem todas as línguas possuem a categoria tempo; algumas possuem apenas o aspecto para dar a ideia de temporalidade. A confusão entre tempo gramatical e tempo natural vem desde os primórdios dos estudos gramaticais. Bagno (2010) exemplifica isso mostrando que o tempo presente, por exemplo, é não-marcado (neutro) e por isso, pode ser utilizado sem o menor problema para expressar eventos que já ocorreram, que ainda vão ocorrer ou realidades atemporais ou eternas, como os ditados populares. Uma questão interessante com relação ao emprego do tempo futuro, por exemplo, é de que eles têm um valor de uso modal maior do que de temporalidade propriamente dita. O futuro é sempre uma certeza, então a condicionalidade, muitas vezes, representa melhor a ideia de algo que poderá ocorrer.

#### 4.3.3 A Concordância

A concordância tanto nominal quanto verbal sempre foi o maior fator de discriminação entre falantes do vernáculo urbano de prestígio e os falantes das variantes linguísticas estigmatizadas. O fato de os falantes mais cultos utilizarem um número maior de casos de concordância na língua e os falantes menos escolarizados utilizarem menos não significa que seja preguiça ou ignorância desses últimos, haja vista haver línguas cujas regras de concordância coincidem com as falas desse português brasileiro menos monitorado, além de haver línguas sem nenhuma concordância (BAGNO, 2011).

A língua portuguesa é diferente do inglês, do francês e até de línguas crioulas que se distinguem pela concordância não redundante; é uma língua em que a concordância é extremamente redundante ou tautológica, ou seja, é dispensável e excessiva. Num exemplo, retirado do *corpus* coletado para esta pesquisa, um aluno do 4º ano do ensino fundamental indígena terena, monolíngue em português escreveu:

QUADRO 4 - Texto de Aluno do 4º Ano do Ensino Fundamental do Grupo Infantil Terena Monolíngue com Problemas de Concordância

OEMTMPCERM4FTXT2<sup>17</sup>  
 “Minha família  
 Eu gosto da minha  
 família começamos  
 bico de joga *bola na  
 campo fumo embora*  
 para casa chegamos,  
*fumo vamo* pecha  
 os animais  
 ele gal...!!

A concordância no texto acima destoa do exigido pela norma tradicional da língua portuguesa, em que as principais marcas de concordância incidem sobre o número e o gênero das palavras. O verbo está na 1ª pessoa do plural “começamos” e o sujeito está oculto *nós*, que se refere provavelmente à família, ou seja *eu e minha família = nós*. É um tipo de concordância possível na língua portuguesa. É uma exceção ao padrão da língua chamada *silepse*. Trata-se de uma concordância ideológica, em que o mais importante é o sentido das palavras e não sua forma. A *silepse* pode ser de pessoa, de gênero e de número. No exemplo acima, a concordância leva em conta o sujeito que fala e os outros membros que compõem sua família para fazer a concordância. É uma *silepse* de número.

Há outras questões relativas à ausência de concordância. A omissão da letra *s*, que é a letra que marca o plural de forma geral na língua portuguesa, nos finais nas formas verbais *fomos (fumo)*, *vamos (vamo)*, por exemplo, não compromete a compreensão do texto, mas foge ao padrão. Na verdade, trata-se de um problema ortográfico, que também ocorre com o apagamento do *r* que marca o infinitivo, *pegar (grafado pecha)* e *bico (brincar)*. Essas duas ocorrências são próprias da oralidade da língua portuguesa e se repetem na escrita. Com relação à concordância de gênero também é destoante da norma da língua portuguesa: “*na campo*” na segunda linha do texto o determinante está no gênero feminino e o núcleo do sintagma está no masculino.

E compreensível na língua portuguesa, entretanto, que a marca de gênero ou de número apenas num dos especificadores dificilmente será empecilho para a compreensão da sentença. Bagno (2011) explica que a redundância da concordância existe devido a pessoa e o número serem marcados no plural duas vezes, nos índices de pessoa e de não pessoa além

<sup>17</sup>OEMTMPCERM4FTXT2: em que OEM correspondem às iniciais do nome do aluno; TM: corresponde a etnia e grupo linguístico; PC: Professora C; ERM: iniciais do nome da escola; 4f: 4º ano do ensino fundamental e TXT2: texto de nº 2.

de também ser marcada na terminação verbal. Intuitivamente o falante, dentro do processo da economia linguística, sabe que tudo que é demais deve ser cortado, elimina as marcas de redundância marcando o número geralmente apenas no primeiro determinante.

#### 4.3.3.1A Concordância Nominal de Gênero e Número

O fenômeno da concordância referenciado acima foi estudado por Taralo (1999). Ele descreve que em todas as comunidades linguísticas há formas em variação (variantes)<sup>18</sup> e pode-se dizer a mesma coisa em um mesmo contexto com um mesmo valor de verdade. Na língua portuguesa, a marcação de número no sintagma nominal está em variação há bastante tempo. Para Taralo (1999, p. 8), “a essa variável corresponde duas variantes linguísticas, as adversárias do campo de batalha da variação: a variante 1 é a presença do segmento fônico /s/, a contrapartida é a ausência desse segmento, ou seja, a forma “zero”.

Há então três formas de marcação do sintagma nominal em português, segundo Taralo (1999, p.9):

1. aS meninaS bonitaS
2. aS meninaS bonita $\phi$
3. aS menina $\phi$  bonita $\phi$

Na primeira, o falante retém a marca de plural ao longo do sintagma nominal (SN), denotando a presença da norma padrão da língua portuguesa. No segundo exemplo, o falante retém a variante /s/ na posição de determinante e lança a variante / $\phi$ / para a posição de modificador. No terceiro exemplo, o falante utiliza-se da variante não- padrão / $\phi$ / nas duas posições finais do SN, retendo a marca de plural apenas na posição inicial. Esse tipo de concordância ocorre tanto da fala dos usuários cultos que usam as variedades urbanas de prestígio quanto daqueles que se utilizam das variantes linguísticas estigmatizadas (usando a nomenclatura de Bagno).

Taralo (1999) afirma que as variantes de uma comunidade estão sempre em concorrência uma com a outra. São três os tipos de concorrência estipuladas por ele: padrão x não padrão, conservadora x inovadora e de prestígio x estigmatizadas. A variante padrão, além de ser conservadora é também a que mantém o maior prestígio social, prestígio que se define a partir das classes médias e altas da sociedade. As variantes inovadoras são não-padrão e estigmatizadas por essas classes.

<sup>18</sup>A um conjunto de variantes dá-se o nome de “variável linguística” (TARALO, 1999).

A concordância nominal ou verbal, por ser redundante, vem abandonando ao longo do tempo as muitas regras de concordância sem prejudicar a compreensão da comunicação. Um dos motivos, já mencionado por Bagno (2011, p.705), foi o da ação do princípio cognitivo da economia linguística. O autor afirma que:

No entanto, como é importante preservar a informação do *número*, as marcas morfológicas de plural permanecem no elemento que encabeça o sintagma nominal, isto é, nos *determinantes*. Ou, mais raramente, apenas no substantivo, como no inglês, em que os determinantes são invariáveis (com a única exceção de *this/these* e *that/those*). Esse processo não é exclusivo do português brasileiro, pois é exatamente o que ocorre em línguas como o francês, o africâner, o cabo-verdiano, o haitiano, entre muitas outras línguas do mundo. O mesmo se pode dizer do gênero dos adjetivos: como exprimem qualidades/propriedades que são, de fato, neutras, é comum a manutenção dos adjetivos no masculino que é, como já sabemos, a forma não marcada em nossa língua. Isso ocorre sobretudo quando o adjetivo está um pouco afastado do núcleo do sintagma nominal com o qual deveria concordar (BAGNO,2011, p. 707).

Então, a tendência do português brasileiro é a favor da não realização de concordância do sujeito com o verbo posposto, bem como da não concordância de número e gênero com o adjetivo que ocorre antes do substantivo no sintagma nominal. Ainda, segundo Bagno, de modo mais geral, pode-se concluir que “a posição à esquerda do núcleo do sintagma desfavorece a concordância tanto nominal quanto verbal” (p. 707). Isso provavelmente ocorre devido “à gramaticalização da ordem sujeito, verbo, complemento do português brasileiro e, no caso dos adjetivos, à colocação predominante desses vocábulos depois do substantivo” (BAGNO, 2011, p. 707).

#### 4.3.3.2Concordância Verbal

No latim clássico, o índice de número e pessoa já vinha agregado à forma verbal, pois essa era uma língua com morfologia verbal abundante. Trazia também agregados o modo, o tempo, o aspecto e a voz. Assim também é o espanhol e o italiano, que são chamados de línguas paradigmáticas, pois recorrem mais à morfologia do que a sintaxe. Outras línguas, porém, sofreram redução em sua morfologia verbal e, por isso, para que possam ser fazer entender recorreram a explicitação do sujeito, através do índice de pessoa e não pessoa. São línguas sintagmáticas, analíticas (BAGNO, 2011, p. 645).

O português, no decorrer do tempo, sofreu muitas alterações. Uma das que mais impactaram a concordância verbal foi a generalização do índice de pessoa *você*, que abandonou a função de pronome de tratamento e vem substituindo o índice de 2ª pessoa, *tu*.

Esse pronome, entretanto, apesar de semântica e pragmaticamente utilizar o lugar de 2ª pessoa, morfologicamente se emprega como de 3ª, ou de não pessoa, como também sucedeu com a coocorrência de nós e a gente, ambos causando uma redução no número nas conjugações verbais do português brasileiro (BAGNO, 2011, p. 645).

Para o ensino da escola, no entanto, a concordância é apresentada e requerida a partir de regras gerais e especiais. A velha máxima conhecida por todos aqueles que frequentaram os bancos escolares de que “o verbo concorda com o sujeito” ainda se instaura como uma verdade absoluta. No entanto, sabe-se que é justamente o contrário, é o sujeito que concorda com o verbo, tendo em vista que o verbo é que projeta seus valores semânticos sobre os demais elementos da sentença, agindo como seu núcleo. Dessa forma, é o sujeito que tem seu número e pessoa atribuídos pela valência do verbo (BAGNO, 2011).

#### 4.3.3.3 Concordância em Guarani - Breves Considerações

Outro fator importante na mudança da concordância foram as gramaticalizações da língua. Segundo Bagno (2011), as terminações do latim clássico nada mais são que antigos índices de pessoa e não pessoa que se aglutinaram ao verbo. Depois desapareceram e retornaram mais tarde. Na língua guarani, por exemplo, as desinências verbais são prefixos. Assim como nas línguas do grupo banto estão aglutinados ao verbo, veja o exemplo:

78

<b>Ake</b>	eu durmo	<b>Roke</b>	nós (exclusivo) dormimos
<b>Reke</b>	you dorme	<b>Peke</b>	you dormem
<b>Oke</b>	ele/ela dorme	<b>Oke</b>	eles/elas dormem
<b>Jake</b>	nós (inclusivo) dormimos		

Conforme P. Pablo Scott, em seu *Curso Breve Del Idioma Guarani* (1981), na conjugação dos verbos em guarani, às raízes dos verbos se agregam partículas para completar seu sentido. Na conjugação acima a primeira letra antes de cada raiz indica a pessoa. Note que há duas formas de 1ª pessoa do plural, o inclusivo e o exclusivo **Jake** inclui quem fala e quem escuta e **Roke** inclui apenas os que falam, excluindo os que escutam. Observe também que a 3ª pessoa do singular e do plural possuem a mesma forma: **oke**. Para desfazer equívocos, ao usar apenas o verbo sem mencionar o sujeito acrescenta-se a partícula *hikuai*.

A gramaticalização está também no modo e no tempo. O primeiro modo estudado em guarani é o volitivo – aquele que expressa um desejo – e vem acoplado ao verbo por meio da partícula *se*, por exemplo, “eu quero dormir” – *Che akese*. Onde *Che* é pronome pessoal de 1ª pessoa, *a* índice de 1ª pessoa, prefixado ao verbo; *ke* a raiz do verbo e *se* a partícula que expressa o volitivo. No português isso também ocorre no futuro do pretérito do indicativo:

eu *leria* – eu ia ler;  
 tu *lerias* - tu ias ler; você ia ler;  
 ele *leria* – ele ia ler;  
 nós *leríamos* – nós íamos ler; a gente ia ler;  
 vós *leríeis* – vós íeis ler; vocês iam ler;  
 eles *leriam* – eles iam ler.

*Leria*, por sua vez, é resultado da aglutinação anterior do latim clássico *ler + ia*, tal como *lerei* é resultado de *ler + hei*.

A marca de futuro do guarani também se dá através de partícula acoplada ao verbo ou ao adjetivo, por exemplo, “Carlos trabalhará” – Carlos omba’apota – em que mba’apo = trabalhar + ta indica futuridade e em “Carlos trabalhará bem” - Carlos omba’apo pōrāta, a partícula indicadora de futuro se agrega ao advérbio.

Essas questões estão sendo colocadas com o objetivo de mostrar algumas similitudes e diferenças entre o sistema verbal do português e de outras línguas. Os exemplos com a língua guarani direciona para o trabalho, haja vista que os textos selecionados para este trabalho provêm de grupos ou que falam o guarani ou que tem contato auditivo frequente com esse idioma. Percebe-se que o idioma guarani pertence à categoria de língua que não se utiliza da concordância, a 3ª pessoa *ha’e*, por exemplo, é pluralizada *ha’ekuera*. A partícula *kuera* é a única forma de pluralizar substantivos e pronomes, por exemplo, casas = *oga+kuera* = *ogakuera*.

No idioma guarani não há gênero gramatical, ou seja, não há desinências especiais para diferenciar masculino de feminino. Portanto, não há concordância entre substantivos e adjetivos: *kūna kyra* = mulher gorda; *kūnakuera kyra* = mulheres gordas. Se houver a necessidade de diferenciar os sexos se utilizará *kuimba’e* para o masculino e *kūna* para o feminino, *mitã* = criatura; *mitãkuimba’e* = o bebê; *mitākūna* = a bebê; *membykuimba’e* = filho homem da mulher e *membykūna* = filha da mulher.

Por esse viés, pode-se dizer que é errônea a afirmação de que não há concordância na língua portuguesa em: os aluno conversa/ vocês fala/ nós fala, etc. (BAGNO, 2011, p. 647), pois a concordância se dá no plano sintagmático, ela apenas não está redundando, haja vista que o SN “os aluno” já traz a marca de plural para a sentença como um todo. Para Bagno (2011, p. 647), “a concordância é um conjunto de fatos muito mais complexos do que as noções (exclusivamente socioculturais) de certo e errado, que empobrecem drasticamente a boa compreensão do fenômeno”.

Com isso, ao olharmos os textos das crianças indígenas escritos em português relativizamos as questões de certo e errado, conforme o padrão do português. As convenções da escrita de cada língua são apreendidas, entre outros aspectos, a partir do *input* recebido, tanto escrito quanto oral. Nessa visão compreendem-se melhor os desvios da norma padrão da língua portuguesa presentes nos textos desses alunos, muitos desses desvios também presentes nos textos de alunos não indígenas.

#### 4.3.4 Hipossegmentação e Hipersegmentação

##### 4.3.4.1A Oralidade e a Escrita

A influência das variantes sociolinguísticas nas produções escritas são motivos de grande preocupação no meio educacional, sobretudo nas séries iniciais. Juntando a isso os problemas dos *erros* de ortografia parece que todos os problemas que ocorrem nos primeiros anos de alfabetização se resumem a essas duas questões. O tão aclamado *erro* já vem a bastante tempo sendo investigado. Marta Scherre e Anthony Naro (2007) estão entre os pesquisadores brasileiros que mais contribuíram para os estudos nessa área.

A noção de *erro*, nos estudos mais recentes da língua tomou uma perspectiva diferente com relação ao ensino da norma escrita. A ideia de *erro* como um problema se transformou na concepção de que o *erro*, sobretudo o ortográfico, é nada mais que uma tentativa de alcançar a escrita sugerida pela escola, o que nem sempre ocorre. Passou a ser um indicador da direção que o próprio sistema educacional tomou a fim de entender quais são as hipóteses que o aluno construiu até aquele momento e buscar soluções. Ferreiro e Teberoski (1999) chamam de *erro construtivo*, pois por meio desses *erros* o professor pode captar em que fase do processo de aprendizagem de aquisição da escrita o aluno se encontra,

além de perceber quais suas maiores dificuldades que, às vezes, transforma-se em dificuldade permanente.

A escrita das crianças e dos acadêmicos indígenas passa muito por isso, sobretudo o dos alunos do fundamental. Muitos problemas de ortografia em seus textos – considerando esses como desvios do padrão instituído pela ortografia nacional brasileira que rege as normas de como se deve escrever. As hipóteses que se podem levantar são: o contato escasso com textos escritos em língua portuguesa; pouco treino na ortografia da língua portuguesa; pouca importância para essa questão por parte da instituição escolar, dentre outros. Segundo Cunha (2004), no início da alfabetização a criança entende a palavra como um enunciado e não como uma unidade gramatical ou semântica, por isso ela tende à hipossegmentação e hipersegmentação. Com relação à ortografia desse grupo, há grande ocorrência de hipo e hipersegmentação das quais passaremos a tratar a partir de agora.

Para Ferreiro e Pontecorvo (1996, apud Cunha, 2004), quando a criança não tem clara a concepção de palavra e de seus limites, ela vai segmentar essa palavra unindo-a a um clítico ou separando-a por que percebe que há estrutura semelhante dentro de uma palavra. Em seu trabalho Cunha (2004) separou as palavras em fonológicas e gramaticais. Conforme a autora, palavras gramaticais são aquelas que envolvem segmentos com sentido gramatical, como os clíticos, por exemplo. No português, há casos em que os clíticos passam a compor com a palavra de conteúdo adjacente uma mesma unidade, ou seja, uma palavra fonológica (BISOL, 1992). Para essa autora o clítico é uma palavra independente que se une a uma palavra lexical formando um só grupo (idem, 2005). Nessa perspectiva, Cunha (2004) declara que quando a criança segmenta a palavra *de pois*, por exemplo, ela reconhece duas palavras lexicais (*de + pois*), da mesma forma, quando ela segmenta o pronome *com tigo* ou *com migo*.

Em sua dissertação de mestrado, Garcia (2010) faz uma análise da influência da oralidade na escrita a partir dos erros ortográficos dos alunos de séries iniciais. A autora cita vários estudiosos que investigam e trazem à reflexão a ortografia e sua aquisição, dentre eles: Lemle (1982), Carraher (1986), Faraco (1992), Zorzi (1998), Cagliari (2003), Moraes (2003) e Mirea et al. (2005) que trabalham numa perspectiva sociolinguística e sociolinguística educacional. Além de Scliar-Cabral (2003) que enfatiza que a escrita exige uma sistematização mais severa enquanto a fala é espontânea. Traz, ainda, Simões (2006) que declara que a criança ao entrar para a escola obriga-se a aprender a distinguir os sons de

sua representação gráfica o que gera uma série de confusões iniciais entre o som ouvido e a letra escolhida para representar aquele som, pois nem sempre há uma representação biunívoca.

Veçossi e Ferreira-Gonçalves (2011) realizaram um trabalho em que a ênfase recai na influência que a oralidade mantém sobre a escrita, testando hipóteses ortográficas - não apenas os traços segmentais, mas prosódicos. Para eles, isso ocorre devido às tênues fronteiras entre as duas modalidades. Na escrita, por exemplo, as palavras tem seu marco divisório definido pelos espaços em branco na folha escrita, acontecendo os casos de juntura e/ou segmentação inadequada das palavras. A isso Silva (1989, 1994), pioneiro nesse assunto, chamou de *hipossegmentação* quando há junção indevida da palavra e de *hipersegmentação* – quando a palavra é dividida indevidamente, como no texto abaixo, retirado do corpus para este trabalho, de um aluno indígena do subgrupo Guarani bilíngue do 4º ano do ensino fundamental:

QUADRO 5 - Exemplo de Hipo e Hipersegmentação em Texto de Aluno do 4o Ano do Ensino Fundamental – Guarani Bilíngue

ADDGBPLERM4FT01

A minha vo conto **um ma** histo de um cachorro estava indo ni um mato e viu um gato chegou perto deli e o cachorro foi **em bora** e o gato **ssi guiu** ele até **nacassa** para de chove o cchoro e o gato foi **em borra nomesmo** mato que o chorro acho no outro dia o cachorro foi até la Para ve o gato o gato sail **domato e Liz esta vam** Brincando **na queli** mato é o cahorro correu porque o **do mo** estava chamando ele?

Não obstante os problemas de ortografia, a presença da hipo e da hipersegmentação é bastante visível nesse texto de 4º ano do ensino fundamental do grupo indígena, neste vê caso, se destaca a hipersegmentação. Veja melhor na tabela:

Tabela 5. Relação de palavras hipo e hipersegmentadas do 4o ano do ensino fundamental do grupo indígena bilíngue guarani

HIPOSEGMENTAÇÃO	HIPERSEGMENTAÇÃO
<i>Nacassa</i>	<i>um ma</i>
<i>Nomesmo</i>	<i>em bora</i>
<i>Domato</i>	<i>ssi guiu</i>
	<i>em borra</i>
	<i>e Liz</i>
	<i>esta vam</i>
	<i>na queli</i>
	<i>do mo</i>

O próximo exemplo, extraído do grupo indígena do 5º ano do ensino fundamental se destaca a hipossegmentação, mas aparece também a hipersegmentação:

QUADRO 6 - Exemplo de Hipo e Hipersegmentação em Texto de Aluno do 5o Ano do Ensino Fundamental – Guarani Bilíngue

AFSKBPRETM5FT01

A minha infancia era legal gueo Eu era criança bricava de carrio. a minha mãe **milevava** Para a Escola queo eu hia Para a Escola com meu irmal Eu esdudava. na escola Tengatui Marangatu. eu chegava na minha casa cominha o Jantar eu **asisti a TV** gueo eu não inha Pra escola eu estava doente a ninha mãe **milevava** Para o Postinho **mideram** remédio eu melhorei. **nasala** Pasava atividade teo que eu fiquei cansado de tanto de escrever. na escola eu aPrendi muita coisa **Praser**legah Para Colega e Para mais velhos vou **a sinque** eu crese esa e a ninha infância

Em destaque, na tabela abaixo, as palavras hipo e hipersegmentadas do 5º ano:

Tabela 6. Relação de palavras hipo e hipersegmentadas do 5o ano do ensino fundamental do grupo indígena bilíngue guarani

HIPOSEGMENTAÇÃO	HIPERSEGMENTAÇÃO
<i>Milevava</i>	<i>asisti a</i>
<i>Mideram</i>	<i>a sinque (assim que)</i>
<i>Nasala</i>	
<i>Praser</i>	
<i>a sinque (assim que)</i>	

Tomamos como exemplo de trabalho com hipossegmentação e hipersegmentação a dissertação de mestrado de Cunha (2004). Nesse trabalho, a autora estabeleceu quatro categorias de análise para estudar melhor esses dois fenômenos: a) palavra gramatical + palavra fonológica; b) palavra fonológica + palavra gramatical; c) palavra gramatical + palavra gramatical; d) palavra fonológica + palavra fonológica. Como palavra fonológica, entende-se a palavra que possui apenas um só acento primário (Bisol, 2005, ApudVEÇOSSO e FERREIRA-GONÇALVES, 2011), *guarda-chuva*, por exemplo, é uma palavra morfológica, mas tem duas palavras fonológicas, já que cada uma das palavras tem seu acento primário. Cunha (2004) montou um quadro, o qual reproduzimos, a partir de Veçossi e Ferreira-Gonçalves, (2011), pois se adéqua a nossos propósito de pesquisa. Nesse quadro, a autora mostra como analisou o processo de hipo e hipersegmentação em 74 textos de 10 alunos de 1ª a 4ª séries entre 6 a 11 anos de idade, em duas escolas, uma pública e uma particular da cidade de

Pelotas/RS entre os anos de 2001 a 2004. Abaixo a reprodução de seu quadro já com exemplos do presente trabalho:

Tabela 7. Casos de hipossegmentação do grupo indígena Guarani/Kaiowá do 4o ano do ensino fundamental

CASOS DE HIPOSEGMENTAÇÃO		
TIPO	DESCRIÇÃO	EXEMPLOS
Palavra gramatical + palavra fonológica	artigos, preposições, pronomes e conjunções.	“OCAVALO”; “IMEU”; “uotro”; “meopai”; “mifeis”; “poresepo” (por exemplo)
Palavra fonológica + palavra gramatical	substantivo + clítico; substantivo + conjunção;	“mãemi” falou (mãe me falou); “potugesi”(português e); “índiaela”
Palavra gramatical + palavra gramatical	Pouquíssimos casos	“devez <i>enqueo</i> ” (de vez em <i>queo</i> )
Palavra fonológica + palavra fonológica	Formação de frases fonológicas, sendo o cabeça um nome, um verbo ou um adjetivo.	“VAIFEZER”(vai fazer), “elegau” (é legal); “focota” (vou contar); “iaprender”; “tomatera”(toma terra)
Pal gramat + pal fonol + pal gramat + pal gramat <sup>19</sup>	Formação de frases fonológicas intercalando palavras fonológicas e gramaticais	OCAVALOIELI(o cavalo e ele); IVIUCÍOCAVALO(e viu se o cavalo); GOSTAHISTÓRIAPRAPONAJISTÓRIA; vacideita (vá se deitar); PARAAAMIAVÁ;(para a minha avó).

Com relação a hipersegmentação ( também baseado no trabalho de Cunha):

Tabela 8. Casos de hipersegmentação do grupo indígena Guarani/Kaiowá do 4o ano do ensino fundamental

CASOS DE HIPERSEGMENTAÇÃO		
TIPO	DESCRIÇÃO	EXEMPLOS
Palavra gramatical + palavra fonológica	Reconhecimento, na primeira sílaba, de uma palavra gramatical, restando, à direita, uma palavra fonológica, lexical ou não.	EN FASÍA, (infância); naris;
Palavra fonológica + palavra gramatical	Reconhecimento de palavras existentes na língua, à esquerda.	vo teí (voltei); vi éráo (vieram); ÉL-A;(se)EM GANOU; SAI-O (saiu); JAOU VI; OLOB-O(o

<sup>19</sup>Esse último quadro foi acrescentado, pois essa categorização não existia em Cunha. Em grande parte dos textos infantis dos grupos indígenas eles realizam uma hipossegmentação longa.

		lobo)
Palavra gramatical + palavra gramatical	Um único caso: 'por que' (porque). Segmentação viável, visto que ambas as grafias existem na língua.	pó que(por que);
Palavra fonológica + palavra fonológica	Apresentam três configurações, conforme as palavras envolvidas sejam lexicais ou não.	DERRE-PETI;

Para este trabalho, no entanto, vamos propor uma leitura mais geral desse fenômeno. Dividiremos em hipossegmentação e hipersegmentação apenas, sem apontar particularidades e agruparemos os casos de hipo e hipersegmentação em dois grupos apenas. Faremos da seguinte forma:

Tabela 9. Casos de hipo e hipersegmentação como serão analisados neste trabalho

HIPOSEGMENTAÇÃO	HIPERSEGMENTAÇÃO
Mifeis; mãemi; <i>enqueo</i> ; focota;	EN FASÍA; vi éráo; SAI-O; DERRE-PETI
OCAVALOIELI; vacideita	

Faremos, como proposto, inicialmente, um levantamento do que foi encontrado dentro das categorias selecionadas para apreendermos o que aparece, de forma geral, na escrita desses grupos.

## 5 METODOLOGIA

Neste capítulo, será apresentada a metodologia empregada na realização da pesquisa que deu origem a esta tese de doutorado. Inicialmente, será feita a retomada dos objetivos da pesquisa e uma descrição da realidade da reserva onde foram coletados os dados que compõem a amostra. Após isso, será disposto no decorrer do texto: a construção do *Corpus* e a seleção dos informantes, a coleta dos dados, a etiquetagem, a divisão dos grupos infantil e adulto, o procedimento para a coleta, a apresentação dos textos e dos temas, o questionário, os códigos de identificação dos textos, além das dificuldades para a realização de trabalho. Finalizado o capítulo, os critérios e procedimentos para a descrição da análise.

### 5.1. Objetivos

A presente investigação teve como objetivo geral descrever dificuldades na produção de textos narrativos escritos em português por crianças e adultos indígenas de três etnias – Kaiowá, Guarani e Terena – em comparação com crianças e adultos falantes monolíngues de português, a partir da Linguística de *Corpus*. Os objetivos específicos da pesquisa são:

- (a) Verificar se existem diferenças entre o tipo de dificuldades reveladas pelos indígenas monolíngues e bilíngues de diferentes etnias – Kaiowá/Guarani e Terena – em comparação com os monolíngues não indígenas na produção de textos narrativos em português.
- (b) Na comparação entre os dois grupos etários, crianças e adultos, observar em que medida o caminho percorrido do ensino básico à formação acadêmica interferiu no desenvolvimento da habilidade de escrita de textos.

- (c) No caso dos grupos de participantes adultos, investigar se o tempo de permanência no curso de graduação (alunos que estão no primeiro e no quarto ano de curso) interfere no nível de dificuldade na produção de textos.

A fim de atingir os objetivos da pesquisa, foram selecionados os grupos de participantes a partir dos critérios faixa etária e língua/etnia, a saber:

- **faixa etária** → dois grupos, formados por crianças e por acadêmicos adultos. O **Grupo 1**, formado pelas crianças, foi subdividido em alunos matriculados no 4º ano do ensino fundamental e alunos matriculados no 5º ano do ensino fundamental. O **Grupo 2**, composto por informantes adultos, foi dividido em dois subgrupos, sendo um formado por acadêmicos que estão em início do curso de graduação e o outro composto por acadêmicos do último ano de graduação;
- **língua e etnia** → três grupos, formados por (a) falantes bilíngues de português e guarani/kaiowá/ indígenas, (b) falantes monolíngues de português/ indígenas Terena; (c) falantes monolíngues de português/ não indígenas.

## 5.2. A Escola Indígena na RFHB hoje

A educação indígena na reserva hoje conta com oito escolas<sup>20</sup>: Escola F (essa escola é uma extensão de uma escola maior, no entanto, funciona de forma independente), Escola C, Escola D, Escola A, Escola B, Escola G, e a Escola H. Além dessas há a principal delas, antigamente denominada escola polo, é a E. Essa escola atende do primeiro ao nono ano do ensino fundamental e alfabetiza em português. As línguas indígenas guarani e terena são oferecidas do quinto ao nono ano. Como recebe um número grande de alunos Terena (alguns falantes do terena, outros não) e de alunos mestiços (alguns falam o guarani, outros não), o ensino não é bilíngue.

Os professores dessa escola são, em sua maioria, indígenas da etnia Terena, sendo poucos os professores não indígenas. É importante notar que poucos falam o idioma indígena, logo, as aulas de língua terena e guarani são ministradas por professores das

---

<sup>20</sup> Estão sendo computadas aqui apenas as escolas das aldeias Jaguapiru e Bororó. A aldeia Panambizinho, que também faz parte da reserva, está inserida na pesquisa, por isso não será citada.

respectivas etnias, contratados especificamente para ministrar aulas de línguas indígenas. Nessa escola, uma única professora recebeu formação para ministrar aula de língua adicional, uma vez que possui formação em Letras/Inglês. Os demais não possuem qualquer formação que os habilite a trabalhar com língua indígena, exceto o fato de serem falantes nativos.

Com relação à escola E<sup>21</sup>, ela possui uma extensão - uma sala multisseriada de alfabetização - que se localiza aproximadamente 4 km da escola polo, chamada escola F, na Aldeia Jaguapiru. Localizada num terreno grande com outras casas ao redor, a escola se constitui de um único cômodo de madeira. É uma construção antiga, com péssima iluminação e acomodações desconfortáveis, sem nenhum recurso didático apropriado para alcançar sucesso num processo de alfabetização. Somando-se a isso, seu isolamento afasta os alunos do convívio com outras crianças, o que seria produtivo para uma aprendizagem mais concreta beneficiada pela interação. Nessa escola, a alfabetização se processa completamente em guarani. A professora, inclusive, só interage nessa língua com as crianças. Essa professora está concluindo um curso de graduação oferecido pela Universidade Federal da Grande Dourados – UFGD, que recebe professores das etnias Guarani e Kaiowá falantes da língua e sua formação é totalmente voltada para a realidade da Reserva.

Trata-se de uma licenciatura indígena de nome *Teko Areu*, nos moldes instituídos pelo PROLIND, criado em 2005. Nesse curso, eles recebem uma formação geral e depois escolhem a área em que se especializam: linguagens, matemática ou ciências sociais. É um projeto inovador que incentiva os Guarani e os Kaiowá a revitalizarem a língua e trabalharem dentro de sua cultura.

Há número grande de professores que frequentam o curso *Teko Areu* que atuam nas escolas da reserva, sobretudo naquelas escolas em que o número de alunos Guarani e Kaiowá é maior. A escola A, por exemplo, é uma das que recebe em seu quadro, um número bastante grande deles. Essa escola se situa a aproximadamente cinco quilômetros da escola E, na Aldeia Bororó. Nessa escola a língua corrente é a língua guarani e um detalhe interessante: a diretora não é indígena, como nas outras.

---

<sup>21</sup> Para preservar a identidade das escolas, optou-se por colocar um pseudônimo ao invés do nome da escola.

Na escola B, ocorre o mesmo fato, a língua guarani é predominante. No entanto, por ela estar situada próxima ao asfalto, mas pelo outro lado da Reserva, há professores não indígenas e da etnia Terena, que falam apenas o português<sup>22</sup>. Percebe-se claramente que esses professores acabam isolados, pois a maioria conversa em guarani entre eles e também com as crianças. O atendimento nessa escola é do primeiro ao nono ano e um fato interessante, a diretora é da etnia Terena, numa escola praticamente Guarani.

A outra escola, também predominantemente Guarani é a Escola C, a mais recente delas. É uma escola pequena, mas bem construída, com pintura e materiais novos. A maioria dos professores não é Guarani, apesar da diretora e dos coordenadores pertencerem a esta etnia e falarem o tempo todo em guarani, inclusive com as crianças. As crianças, por sua vez, são majoritariamente Guarani e Kaiowá falantes. Há poucas crianças de etnia Terena e mestiças. Uma professora do primeiro ano (na época) confessou que em sua sala havia três alunos que não falavam guarani e dificultava seu trabalho, pois ela tinha que ficar explicando em português para essas crianças e para as outras era muito mais fácil, pois entendiam perfeitamente queo ela falava.

A professora citada acima faz o curso de licenciatura indígena. Já outra professora, na mesma escola revelou que sentia bastante dificuldade em suas aulas, pois era Terena e não falava guarani. Por isso, recorria às outras crianças para ajudá-la em suas aulas. Estava bastante desanimada, pois ouvira boatos de que suas aulas de Língua Portuguesa não estavam rendendo como deveriam para o quarto ano que estava e ela mesma acreditava que isso ocorria por ela não ser falante do guarani. Outra professora, também Terena, confidenciou suas angústias em sala de aula, esta na escola A, também de maioria guarani, pois as crianças estavam com dificuldades e ela não sabia como resolver essas questões, pois eram falantes do guarani e a grande parte não compreendia bem o português. Vale lembrar que as duas últimas professoras citadas são recém-formadas no curso de pedagogia e declararam não saber como lidar numa situação de bilinguismo, a primeira é falante do terena e a segunda, mesmo sendo da etnia terena não é falante.

Finalmente, a Escola D, se situa no fundo da Aldeia Bororó, nos limites com as terras não indígenas. Possui cinco salas de aula e volta-se ao ensino das séries iniciais. Os professores são Guarani e Kaiowá - falantes da língua. O diretor é Kaiowá. Há apenas um

---

<sup>22</sup>Estes dados correspondem ao ano de 2012, época da coleta de dados. O quadro dos professores e do ensino hoje (2014) pode ter-se modificado.

professor não indígena, que trabalha com educação física. Há poucos indígenas formados nessa área na reserva. Com relação a escola G, ela atende o ensino fundamental e fica nos arredores da reserva e apesar de municipal, esta sob os cuidados de uma instituição religiosa.

As escolas, apesar de conhecerem e elaborarem os projetos pedagógicos baseados na educação escolar indígena, proposto pela constituição de 1998 e pelos *Referenciais Curriculares para a Escola Indígena* (RCNEI), para uma educação voltada para a língua, a cultura e a tradição do povo indígena, estão submetidas às normas do ensino proposto pela Secretaria de Educação do Município de Dourados e seguem o mesmo currículo das escolas municipais não indígenas. Com as novas propostas de ensino, entretanto, procuram inserir algumas adequações dos parâmetros da educação indígena postulados no RCNEI em seus currículos, adaptando o que a legislação federal sugere às suas necessidades locais.

### 5.3. O Contexto da Pesquisa

Muito se ouve no meio acadêmico, quando se trabalha com alunos indígenas de suas dificuldades com a língua portuguesa escrita. Juntando-se aos acadêmicos, estão os alunos do 4<sup>o</sup> e do 5<sup>o</sup> ano do ensino fundamental da reserva, pois muitos professores mostravam-se muito preocupados por seus desempenhos nessas séries estarem aquém do necessário exigido deles. A maioria não estava sequer alfabetizado, segundo os professores. Comentários e conjecturas sobre os motivos para tais dificuldades vão desde que eles não dominam o português porque possuem outra língua materna até de que não aprendem por serem indígenas. A curiosidade científica impulsionou esta investigação e a busca de entender o porquê dos alunos indígenas terem problemas com a escrita em língua portuguesa transformou-se num grande questionamento que exigia uma resposta concreta, baseada em investigação e não em suposições apenas.

Para procurar entender os porquês se faz necessário levantar algumas hipóteses e procurar comprová-las ou refutá-las. A partir disso, o primeiro passo foi coletar os textos escritos desse grupo linguístico e buscar na sua forma de escrever, as respostas. Primeiramente, já se sabe pelo senso comum que o ensino nas escolas públicas deixa muito a desejar, apesar de algumas se destacarem no cenário nacional, chama a atenção justamente por isso: são exceções. Nas aldeias da RFHB, as escolas estão ainda passeio pelo processo de

transição da concepção antiga de ensino para os pressupostos indicados pelos *Referenciais Curriculares Nacionais para a Escola Indígena*<sup>23</sup>(RCNEI).

O ensino bilíngue e multicultural está ainda sendo implantando muito lentamente, como costuma acontecer em processos de transição. Os professores ainda estão aprendendo como sair de um processo de ensino que era voltado para o não indígena e começar outro processo de ensino, valorizando sua cultura, seus costumes e, sobretudo, sua língua. Isso, empiricamente, norteia algumas hipóteses, a primeira, é de que o sistema de ensino em si no contexto da reserva está procurando o caminho que o RCNEI aponta, pois cada contexto indígena é diferenciado de outro, e, portanto, cada um tem deverá encontrar sua via de acesso.

Depois, os professores indígenas ainda não são maioria nas aldeias, isso também pode ser causa de desafios com a aprendizagem das crianças. Outra hipótese também plausível pode ser a de que as condições econômico-sociais inibam o desenvolvimento na escola, uma vez que a estória de pobreza e violência na reserva é de conhecimento de todos. Tudo isso somado, de fato, pode vir a interferir no processo ensino/aprendizagem nas escolas indígenas e desaguar no ensino universitário. Há que se lembrar de que a única escola que oferece ensino médio na reserva é recente, anteriormente os alunos indígenas saíam da reserva para frequentar as escolas do ensino médio da cidade.

Após a coleta do material, observou-se que muitos seriam os olhares possíveis para se construir uma análise dos textos. Optou-se, então, por traçar um perfil da escrita encontrada em cada grupo indígena, o grupo infantil e o grupo adulto, comentando brevemente, sem aprofundar ou tentar explicar os problemas encontrados e em seguida olhar mais atentamente para algumas questões pontuais, as quais forma divididas em 3 blocos. O primeiro foi o dos sintagmas verbais - ausência ou erro de concordância verbal de número e pessoa, ausência ou erro de concordância verbal de gênero, erro de tempo verbal – uso de um tempo no lugar de outro e a ausência de cópula. O segundo, foi o dos sintagmas nominais, a saber, ausência de determinantes, ausência ou erro de concordância de gênero e ausência ou erro de concordância de número. No último, ficou a hipossegmentação e a hipersegmentação.

---

<sup>23</sup>Maiores detalhes sobre a educação indígena no capítulo 2 desta tese.

#### 5.4. A Construção do *Corpus* e a Análise de *Corpora*

Pacheco (2010), em sua tese de doutorado, traz uma boa exposição acerca das vantagens e problemas que envolvem a compilação e a análise de *corpora* da qual me aproprio para compor esta seção pela sua objetividade. Segundo a pesquisadora, a confecção de um *corpus* exige um trabalho rigoroso com relação à coleta, ao armazenamento e ao processamento dos dados obtidos. São técnicas que vão desde a seleção dos fatores linguísticos aos não linguísticos, ou seja, os contextuais que envolvem os informantes, como faixa etária, grau de escolaridade e no caso deste trabalho, o fato de serem monolíngues ou bilíngues também é um critério de análise importante.

Outras variáveis como os tipos/gêneros textuais, se os textos são falados ou escritos, que tipos de temas são abordados, se foram escritos de forma espontânea ou tiveram auxílio de outro material escrito ou falado, quem aplicou os textos e acompanhou sua escrita (se o professor da sala ou o pesquisador que solicitou o material), tudo isso se justifica como relevante na hora de contabilizar os dados e pode fazer diferença na hora da análise.

##### 5.4.1 Sistema de Anotação de partes do discurso

Uma das áreas que a LdC atua é na leitura dos dados, a partir dos programas eletrônicos próprios. Para isso, entretanto, para que a máquina leia esses dados é preciso alguma intervenção manual, assim a leitura desses dados terá maior clareza. Uma das intervenções é a anotação de partes do discurso, que pode ser manual ou eletrônica. A etiquetagem manual se torna mais indicada, conforme os pressupostos que o trabalho requer e consiste em colocar junto às porções de linguagem as legendas para que a máquina quantifique o número de vezes que vai ocorrer determinada colocação e assim, se estimar o quanto ela pode ser um padrão ou não de determinada comunidade linguística.

Para McEnery e Wilson (1996), o fato de um *corpus* ser anotado facilita o trabalho do pesquisador, pois uma grande frequência de palavras num *corpus* anotado faz diferença no material disponível tanto no nível lexical quanto sintático. A eficácia de um *corpus* anotado vai depender dos objetivos do trabalho. Com relação aos colocados, por exemplo, utiliza-se a informação mútua (*mutual information*), que é uma relação de associação para determinar ou medir o grau de frequência que as expressões que ocorrem juntas com a proporção de número de vezes que elas ocorrem no *corpus* como um todo.

Segundo esses autores, o cálculo leva em conta todos os pares de palavras que coocorrem num *corpus* e atribui uma pontuação abaixo de zero para colocações que julgue aleatórias e pontuação acima de um (01) para coocorrências vistas como não aleatórias. Se a pontuação for abaixo de zero, é provável que a coocorrência seja aleatória e se for mais alta que zero, provavelmente, essa colocação não seja aleatória, ou seja, haverá uma relação mais estreita entre essas palavras e há um motivo para que ocorram sempre juntas.

#### 5.4.2 Anotação de Partes do Discurso para este Trabalho (*part-of-speech annotation*)

Como dito anteriormente, a coleta e o armazenamento de dados devem servir aos propósitos de cada investigação. Para Granger (2002), o pesquisador pode objetivar a confirmação de uma teoria ou refutá-la, a partir dos novos dados coletados. Ou ainda, melhorar ferramentas e métodos de determinado campo de trabalho. A autora apresenta duas formas de se trabalhar com *corpora* de aprendizes. A primeira, *Contrastive Interlanguage Analysis*, que consiste na comparação entre os dados das línguas em estudo.

A segunda forma é a *Computer-Aided Error Analysis*, ou Análise de Erros, que se adéqua a esta pesquisa, uma vez que abarca um grau de padronização alto e utiliza ferramentas computacionais para ajudar na captação desses dados. Os dados são codificados, conforme o objetivo determinado, se morfológica ou sintaticamente e, posteriormente, lido pelo programa. Há dois tipos de anotação: o *Part-of Speeching* (POS) *tagging* e o *error tagging*. No primeiro, a anotação é apenas de partes do discurso que se pretende analisar. Já o segundo, é a etiquetagem de erros, em qualquer parte do texto. (McENERY e WILSON, 1996; BERBER SARDINHA, 2004).

#### 5.4.3 Etiquetagem de Erros

Vários são os programas que permitem uma contagem de erros eletronicamente. Cada vez mais, esses programas vão se aprimorando para que se tornem mais e mais eficazes. A etiquetagem manual, no entanto, é recomendada como a mais adequada para determinados trabalhos, como este, por exemplo.

Como o objetivo é apreender a linguagem utilizada pelos alunos/acadêmicos indígenas de forma natural, com o mínimo possível de intenção do pesquisador, as opções

por escolher os critérios de análise, bem como os dados que passarão por essa análise exigem uma anotação manual.

Seguindo o modelo utilizado por Pacheco (2010), a etiquetagem deste trabalho ocorrerá de duas formas: uma versão natural, sem anotações e a versão anotada, conforme os critérios estabelecidos, anteriormente, para isso. No item 5.6.7 estão os critérios e os procedimentos utilizados para a etiquetagem dos textos.

## 5.5 A Coleta dos Dados

### 5.5.1 A Seleção dos participantes

A escolha dos participantes obedeceu a critérios que respondem a necessidades das questões investigadas: por que o grupo indígena tem tanto problema de escrita em língua portuguesa? Se os problemas de escrita são fortes nos textos dos cursos universitários, como será nas séries iniciais, mais precisamente quando as crianças estão concluindo o processo de alfabetização? A necessidade de unir as duas pontas do ensino pareceu ser necessária para que a investigação tivesse contundência em seus pressupostos. Os acadêmicos também passaram pelo sistema de ensino da escola indígena de suas aldeias. Se os problemas de escrita existem hoje, poderia ser porque o caminho percorrido até aqui pode oferecer respostas, apesar das diferenças ocorridos devido ao fator tempo.

Os alunos também aprendem o inglês como língua adicional. Há aulas de língua terena e guarani. O processo de alfabetização, que deveria ser bilíngue em guarani/português, conforme o RCNEI postula, na prática não ocorre, apesar do interesse demonstrado pelos professores, na sua maioria indígenas, em desenvolver seu trabalho a partir de uma perspectiva diferenciada e intercultural. Percebe-se um grande esforço por parte de todo o corpo docente e da escola como um todo para que a educação indígena ocorra de forma intercultural e bilíngue.

### 5.5.2 O Grupo Infantil

O grupo infantil abrangeu indígenas e não indígenas. Para o subgrupo indígena optou-se por selecionar crianças de 4<sup>o</sup> e 5<sup>o</sup> anos do ensino fundamental de escolas indígenas

da RFHB em final de processo de alfabetização. Inicialmente, as crianças foram escolhidas de apenas uma escola, entretanto, devido a problemas de várias ordens, crianças de outra escola também foram selecionadas para fechar esse grupo. A ideia inicial era selecionar 20 crianças/alunos Guarani/Kaiowá Bilíngues e 20 crianças/alunos Terena Monolíngues da 4ª e da 5ª série, ou seja, somando os alunos das duas séries, o grupo indígena teria 60 crianças. No entanto, isso não foi possível.

No total, conseguiu-se o que segue:

Tabela 10. Subgrupo infantil indígena dividido por etnia

Subgrupo Guarani/Kaiowá Bilíngue	Nº de participantes <sup>24</sup>
4º ano do ensino fundamental	36
5º ano do ensino fundamental	35
Subgrupo Terena Monolíngue	Nº de participantes
4º ano do ensino fundamental	19
5º ano do ensino fundamental	21

O subgrupo infantil indígena não contabilizou o mesmo número de textos, como se pretendia por uma série de motivos. O primeiro, foi a ausência de alguns alunos no dia da aplicação dos textos, que foram distribuídos em datas diferentes. Isso ocorreu também com o preenchimento do questionário. Para o grupo de controle foram selecionados textos de alunos não indígenas da 4ª e 5ª série de uma escola pública estadual. Desses alunos conseguiu-se mais de 20 textos de cada série. Veja na tabela:

Tabela 11. Subgrupo Infantil Não Indígena

Subgrupo Não indígena	Nº de participantes
4º ano do ensino fundamental	34
5º ano do ensino fundamental	30

### 5.5.3 O Grupo Adulto

Dentro do grupo dois, o grupo dos acadêmicos indígenas e não indígenas, foram selecionados textos da primeira e da última série de qualquer curso de graduação de uma das

<sup>24</sup> A denominação escolhida para os produtores dos textos foi *participantes*, mas pode ocorrer também de aparecer a denominação *informantes* ou *sujeitos da pesquisa*.

universidades/Centro Universitário de Dourados e região<sup>25</sup>. Do subgrupo indígena adulto também se conseguiu menos do que o proposto para o trabalho, ou seja, menos de 20 informantes:

Tabela 12. Subgrupo Adulto Indígena subdividido por etnia

<b>Subgrupo Guarani/Kaiowá Bilíngue</b>	<b>Nº de participantes</b>
1º ano da graduação	18 alunos
Último ano da graduação	18 alunos
<b>Subgrupo Terena Monolíngue</b>	<b>Nº de participantes</b>
1º ano da graduação	19 alunos
Último ano da graduação	19 alunos

O subgrupo adulto não indígena ficou assim:

Tabela 13. Subgrupo Infantil Não Indígena

Subgrupo Não Indígena	Nº de participantes
Primeiro ano do ensino superior	20
Último ano do ensino superior	24

## 5.6 Procedimentos para a coleta

A produção dos textos pelos participantes da pesquisa demandou bastante tempo. Foram meses de trabalho até se conseguir um número aproximado do que se pretendia. Cada sujeito da pesquisa recebia um *kit* contendo: 3 folhas com cabeçalho, indicando tema 1, tema 2 e tema 3; um questionário com questões sobre a língua utilizada por cada um deles em casa, na escola, na igreja etc., além do Termo de Consentimento Livre e Esclarecido.

Na escola, com as crianças, a pesquisadora orientou os alunos na confecção dos textos e do questionário. Com o grupo adulto, no entanto, o *kit* foi distribuído e depois devolvido. Houve apenas um caso em que a pesquisadora orientou a escrita dos textos para acadêmicos. Para as crianças, na escola, cada texto foi escrito em dia separado e um dia também foi escolhido só para preencher o questionário.

<sup>25</sup>Isso ocorreu porque não se conseguia fechar os grupos em uma ou duas instituições de ensino. Nas primeiras séries não havia tanto problema, todavia nas séries finais não se conseguia fechar grupos de 20. A evasão do indígena é bastante grande nos vários cursos de graduação.

## 5.6.1 Os textos

Os textos, denominado *corpus* pela LdC, foram selecionado manualmente, como já relatado de forma a se conseguir a escrita o mais espontânea possível a fim de observar a escrita dessa comunidade linguística. A escolha por estes temas surgiu de discussões acerca do que seria relevante para a comunidade indígena. Os temas deveriam servir às crianças e aos acadêmicos. Pensou-se então em escolher temas relacionados à sua cultura, sua família e sua identidade. Veja na tabela abaixo:

Tabela 14. Temas solicitados para os textos

TEMA 01	Um conto da minha infância
TEMA 02	Minha família e eu
TEMA 03	O que é ser índio para mim.

Nessa fase, os alunos, crianças e adultos receberam três folhas de caderno pautadas, grampeadas juntas com um cabeçalho no qual constava, primeiro, as folhas para a confecção dos textos infantis, e em seguida, dos adultos:

DINTER – UFRGS – UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL/UEMS – UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL. PESQUISADORA: PROFESSORA SERA ESPÍNDOLA – ORIENTADORA: PROFESSORA DOUTORA INGRID FINGER  
 DATA DA COLETA: \_\_\_\_/\_\_\_\_/\_\_\_\_ 4<sup>o</sup> ANO  
 ESCOLA \_\_\_\_\_  
 ALUNO: \_\_\_\_\_  
 IDADE \_\_\_\_\_ ETNIA \_\_\_\_\_ ETNIA:PAI \_\_\_\_\_ ETNIA:MÃE \_\_\_\_\_

TEXTO 1 – TEMA: UM CONTO DA MINHA INFÂNCIA

DINTER – UFRGS – UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL/UEMS – UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL. PESQUISADORA: PROFESSORA SERA ESPÍNDOLA – ORIENTADORA: PROFESSORA DOUTORA INGRID FINGER  
 DATA DA COLETA: \_\_\_\_/\_\_\_\_/\_\_\_\_  
 ALUNO: \_\_\_\_\_  
 IDADE \_\_\_\_\_ ETNIA \_\_\_\_\_ ETNIA:PAI \_\_\_\_\_ ETNIA:MÃE \_\_\_\_\_  
 CURSO: \_\_\_\_\_/\_\_\_\_\_ ( )1<sup>o</sup> ANO ( )ÚLTIMOANO ( ) MAT ( ) VESP ( ) NOT

TEXTO 1 – TEMA: UM CONTO DA MINHA INFÂNCIA

Com relação ao tema 01 – *Um conto da minha infância*, alguns alunos entenderam que deveriam escrever qualquer estória ouvida na infância. A intenção do trabalho objetivava que a escrita fosse de estórias de suas tribos, estórias culturais. Isso resultou numa mistura de estórias ouvidas na escola (contos de fadas/ folclóricos tradicionais) e estórias de fato culturais. O segundo tema não ofereceu maiores dificuldades, alguns falaram

de suas famílias, outros da família indígena. O terceiro tema se caracterizava como um texto mais opinativo. Eles entenderam ser um texto que apresentava uma relação das qualidades do seu povo. Então muitos textos começaram com “ser índio é...” seguia-se uma enumeração do que seria “ser índio” para aquele escrevente. Isso ocorreu com o grupo infantil.

O grupo adulto compreendeu melhor o que se pedia apesar de também ter alguns equívocos como ocorreu com as crianças.

Não foi solicitado tamanho para os textos. Cada um escrevia o tanto que quisesse. Por isso a maioria dos textos das crianças são textos curtos, com uma média de 50 palavras e dos adultos já são um pouco maiores, a maioria com mais de 100 palavras.

### 5.6.2 O Questionário

O questionário foi aplicado com a finalidade de obter dados adicionais que auxiliassem a compreensão da cultura subjacente à língua. Uma tentativa de se entender melhor o contexto sociocultural dos alunos/universitários com a escola/universidade. O questionário, composto por duas páginas, com um cabeçalho, traziam questões que privilegiavam dados pessoais e de uso da língua pelos participantes<sup>26</sup>.

Questões como a idade que a criança começou seus estudos, se esse início escolar foi nas escolas da reserva ou fora dela, assim como saber em que situações se falava a língua indígena e em que situações se falava o português eram importantes para se entender melhor o contexto cultural e linguístico no qual os indígenas convivem. Além disso, conhecer a opinião dos participantes acerca da própria língua e da língua do não indígena auxilia na compreensão do contexto que se está pesquisando. Nesse sentido, o questionário foi um excelente auxiliar e será utilizado para complementar a análise e sua discussão.

### 5.6.3 Os Códigos Utilizados para Identificar os Textos

Como forma de localização dos textos, foram criados códigos que facilitam sua identificação. Esses códigos estão no centro superior da página de cada texto. O código para

---

<sup>26</sup> No anexo há os modelos de questionário respondidos pelos grupos infantil e adulto.

o grupo infantil está assim constituído: ADDGBPLERM4FTXT01, e está explicado na tabela abaixo.

Tabela 15. Códigos utilizados para identificação dos textos do grupo infantil<sup>27</sup>

ADD	Iniciais do nome do sujeito da pesquisa
GB	Grupo étnico e linguístico ao qual o sujeito pertence (Guarani Bilíngue)
PL <sup>28</sup>	Professora X (nome da professora da turma em que os textos e o questionário foram aplicados)
ERM <sup>29</sup>	Escola XX ( nome da escola de onde foi recolhido o material)
4F	Indicação da série que o aluno estuda
TXTO01 TXTO02 TXTO03	Número correspondente ao tema do texto (01, 02 ou 03)

A codificação dos textos dos acadêmicos ficou um pouco diferente. Nesses textos não constam o nome da professora e nem da universidade do acadêmico. Veja no quadro:

CEATMULTSUPTXT003

Tabela 16. Códigos utilizados para identificação dos textos do grupo dos acadêmicos

CEA	Iniciais do nome do sujeito da pesquisa
TM	Grupo étnico e linguístico ao qual o sujeito pertence (Terena Monolíngue)
ULTSUP	Indicação da série que o aluno estuda (se primeiro ou último ano de um curso de graduação)
TXTO01 TXTO02 TXTO03	Número correspondente ao tema do texto (01, 02 ou 03)

#### 5.6.4 Os Percalços Para a Coleta dos Textos

Qualquer pesquisador que se disponha a realizar uma pesquisa de campo sabe dos obstáculos que aqui serão narrados. Necessário se faz falar disso, para apresentar aos que não tiveram essa experiência ou àqueles que tenham a pretensão de adquiri-la terem uma noção do que se enfrenta quando se realiza um trabalho desse porte. Para obter esses textos foi despendido muito esforço e diplomacia, uma vez que a população indígena dessa região tem verdadeira ojeriza a pesquisadores.

<sup>27</sup> A decodificação total está no anexo.

<sup>28</sup> Apenas para os textos do grupo infantil.

<sup>29</sup> Também ocorre apenas com o grupo infantil.

Houve a negação em assinar o termo de consentimento, por parte dos pais de alunos que disseram aos filhos para não escreverem. Adultos contra a ajudante da pesquisadora, que era indígena, por estar auxiliando uma pessoa branca que não iria trazer nenhum benefício para a comunidade indígena.

Problemas de acesso às escolas da reserva no período de chuvas, devido aos atoleiros e, no período de seca, por causa dos buracos e da poeira. Em razão dos bloqueios na rodovia por vários dias, em que não era permitido a não indígenas entrar na reserva. A resistência em escrever, por parte dos acadêmicos, por receio à exposição ou por não verem importância no trabalho, ou por falta de tempo mesmo. Atraso no calendário escolar da escola e falta de salas de aula para os alunos e com isso a pesquisadora teve que esperar a situação se normalizar para orientar a escrita dos textos e do questionário. Houve, além disso, uma semana inteira sem aulas, devido a cursos oferecidos aos professores, ou às semanas comemorativas, como a semana do índio, por exemplo.

Às vezes, também o professor da sala faltava ou aplicava prova no dia marcado para a realização dos textos ou do questionário. Com relação aos acadêmicos, também os problemas foram diversos. Como não se conseguiu número suficiente em Dourados, a saída para cidades próximas foram necessárias. Os textos eram, então, entregues aos acadêmicos em suas casas e depois eram recolhidos, logo muitos textos não foram entregues. Acadêmicos das Unidades da UEMS de Dourados, Maracaju e Campo Grande, UFGD, um campus universitário particular de Dourados e uma universidade particular de Campo Grande participaram da pesquisa.

Todos esses entraves ocorrem quando se trabalha com pesquisa de campo. Nem sempre o que foi planejado ocorre conforme se projetou e aprender a lidar com os obstáculos acaba por fazer parte da pesquisa. Apesar disso se conseguiu coletar um *corpus* significativo para a pesquisa.

#### 5.6.5 O *Corpus* Geral Indígena de Dourados - Grupo Infantil

A tabela abaixo demonstra o número de sujeitos indígenas participantes da pesquisa, de textos e número de palavras do grupo infantil. Lembrando que o subgrupo infantil não indígena é grupo de controle.

Tabela 17. Relação dos sujeitos e de palavras produzidos pelo grupo infantil

	TEXTOS	SUJEITOS <sup>30</sup>	PALAVRAS
4ª SÉRIE G/K BIL	TEXTO 1	30	8.204
	TEXTO 2	31	
	TEXTO 3	33	
4ª SÉRIE T MONOLIN	TEXTO 1	19	4.542
	TEXTO 2	19	
	TEXTO 3	19	
5ª SÉRIE G/K BIL	TEXTO 1	30	8.816
	TEXTO 2	31	
	TEXTO 3	30	
5ª SÉRIE T MONOLIN	TEXTO 1	22 <sup>31</sup>	6.466
	TEXTO 2	20	
	TEXTO 3	19	
4ª SÉRIE NÃO INDÍG	TEXTO 1	31	10.308
	TEXTO 2	34	
	TEXTO 3	32	
5ª SÉRIE NÃO INDÍG	TEXTO 1	26	5.729
	TEXTO 2	28	
	TEXTO 3	29	
TOTAL:		483	44.065

#### 5.6.6 O *Corpus* Geral Indígena de Dourados e Região - Grupo Adulto

A tabela seguinte mostra o número de sujeitos da pesquisa e o total de palavras do grupo adulto. Lembrando que o grupo adulto é composto pelos acadêmicos indígenas e não indígenas, tendo este último a função de servir como grupo de controle.

<sup>30</sup>O número de sujeitos que escreveram cada texto difere do número de sujeitos no cômputo total, por exemplo, a 4ª série do fundamental do grupo Guarani/Kaiowá bilingue totaliza 35 alunos que escreveram textos, no entanto, desses 35 apenas 29 escreveram o texto 01; 30 escreveram o texto 02 e 33 escreveram o texto 03. Assim ocorreu com todos os grupos.

<sup>31</sup>Ocorreu também de uma aluna escrever dois textos com o tema 01. Nesse caso, o total de alunos que escreveram somaram 21 e de textos 22 ( se todos os três foram escritos)

Tabela 18. Relação do número de sujeitos e de palavras produzidos pelo grupo adulto

	TEXTOS	SUJEITOS	PALAVRAS
1ª SUP G/K BIL	TEXTO 1	18	9.641
	TEXTO 2	17	
	TEXTO 3	17	
1ª SUP TER MONOLING	TEXTO 1	18	7.809
	TEXTO 2	19	
	TEXTO 3	19	
ÚLTIMO ANO SUP G/K BIL	TEXTO 1	18	9.642
	TEXTO 2	18	
	TEXTO 3	18	
ÚLTIMO ANO SUP TER MONOL	TEXTO 1	19	11.569
	TEXTO 2	19	
	TEXTO 3	19	
1ª SUP NÃO INDÍG	TEXTO 1	20	6.777
	TEXTO 2	20	
	TEXTO 3	19	
ÚLTIMO ANO SUP NÃO INDÍG	TEXTO 1	24	10.656
	TEXTO 2	24	
	TEXTO 3	23	
TOTAL:		349	56.094

### 5.6.7 Os Critérios e os procedimentos para a descrição dos dados

Na descrição dos dados os grupos infantil e adulto serão vistos separadamente. Primeiro será dada uma atenção geral à escrita de cada grupo, ressaltando aspectos que chamarem mais a atenção. Na sequência, primeiro o grupo infantil e depois o grupo adulto terá seus dados transferidos para o programa *AntConc*, mais especificamente para a ferramenta *Concord*. Após a passagem pelo programa, os dados serão descritos, conforme os critérios previamente estabelecidos abaixo.

Foram elencados nove critérios para o programa verificar a partir da etiquetagem dos textos. Após a leitura dos dados se verificará:

1. A presença da hipossegmentação das palavras, ou seja, palavras que deveriam estar grafadas separadas aparecem justapostas (*um m a*).

2. A presença da hipersegmentação das palavras, ou seja, palavras que deveriam estar grafadas juntas aparecem separadas (*nacassa*).
3. Sintagma nominal: concordância nominal – ausência de determinante, (ela não sabia estrada)
4. Ausência ou erro de concordância de gênero (norio negra)
5. Ausência ou erro de concordância de número (as gente).
6. Sintagma verbal : concordância verbal - ausência ou erro de concordância verbal de número e pessoa (ele foi procuramos).
7. Ausência ou erro de concordância verbal de gênero (a dificuldade erra imenso)
8. Erro de tempo verbal: emprego do tempo presente pelo pretérito ou futuro (eles forao pescava).
9. Ausência de cópula verbo *ser*: ela muito chata.

Os critérios acima foram transformados nas seguintes etiquetas a serem lidas pelo programa *Antconc*:

Tabela 19. Códigos para a etiquetagem

Etiqueta	Nome
HIPO	Hiposegmentação
HIPER	Hipersegmentação
<b>Sintagma nominal – concordância nominal</b>	
SND	Ausência de determinante
SNG	Ausência ou erro de concordância de gênero
SNN	Ausência ou erro de concordância de número.
<b>Sintagma verbal – concordância verbal</b>	
SVNP	Ausência ou erro de concordância verbal de número e pessoa.
SVG	Ausência ou erro de concordância verbal de gênero
SVT	Erro de tempo verbal: emprego do tempo presente pelo pretérito.
SVC	Ausência de cópula (verbo ser)

Os textos coletados foram digitados e transformados em .txt para que o *AntConc* – que é um conjunto de programas integrados destinado à análise linguística, pudesse lê-los. Após isso, foram etiquetados manualmente, um a um. Isso porque, o objetivo geral do trabalho é investigar os textos escritos por crianças e acadêmicos indígenas Guarani/Kaiowá e Terena, comparando-os com os respectivos textos correspondentes aos não indígenas, a fim de perceber se, especificamente há diferenças nos tipos de dificuldades enfrentadas entre os grupos indígenas e não indígenas; se as dificuldades das crianças permanecem as mesmas

ao chegar a um curso de graduação. Por último, com relação ao grupo adulto, investigar se o tempo de permanência na universidade fez alguma diferença no nível de dificuldades enfrentadas em sua escrita.

Para alcançar os objetivos propostos, foram seguidos os seguintes passos: 1<sup>o</sup> o programa *AntConc* contou quantas palavras cada grupo continha, através da ferramenta *Concord*, foram listadas e contadas cada etiqueta colocada nos textos; 2<sup>o</sup> foi feita a comparação entre os textos indígenas e não indígenas do grupo 1 e do grupo 2 para verificar estatisticamente a relevância desses dados. Se cada ocorrência de hipossegmentação, por exemplo, tinha significância ou não.

Como os *corpora* se constituem de tamanhos diferenciados, foi necessário o processo de normalização, para que os corpora pudessem ser comparados entre si. O valor normalizado, ou seja, o número de ocorrência de hipossegmentação, por exemplo, deve ser dividido a cada mil palavras pelo número total de ocorrências de *tokens* do *corpus*, multiplicando-se o resultado por mil (SARMENTO, 2008).

Após a normalização dos valores, foram realizadas as comparações entre

1. Grupo infantil, isto é, 4<sup>o</sup> e 5<sup>o</sup> anos do fundamental agrupados Guarani/Kaiowá bilíngue e Terena monolíngue; Guarani/Kaiowá bilíngue e não indígena; Terena Monolíngue e não indígena; no grupo adulto: 1<sup>o</sup> e último ano do curso de graduação agrupados Guarani/Kaiowá bilíngue e Terena monolíngue; 1<sup>o</sup> e último ano do curso de graduação agrupados Guarani/Kaiowá bilíngue e não indígena; 1<sup>o</sup> e último ano do curso de graduação agrupados terena monolíngue e não indígena.
2. Grupo 1 comparado a grupo 2 -4<sup>o</sup> e 5<sup>o</sup>(agrupados) anos do fundamental Guarani/Kaiowá bilíngue x 1<sup>o</sup> e último ano agrupados Guarani/Kaiowá bilíngue ; 4<sup>o</sup> e 5<sup>o</sup>anos do fundamental Terena monolíngues agrupadosx1<sup>o</sup> e último ano agrupados Terena monolíngues;4<sup>o</sup> e 5<sup>o</sup>anos do fundamental agrupados Não Indígena x 1<sup>o</sup> e último ano agrupados Não Indígenas.
3. Grupo 2 , 1<sup>o</sup> ano da graduação dos Guarani/Kaiowá bilíngue x último ano da graduação dos Guarani/Kaiowá bilíngue; 1<sup>o</sup> ano da graduação dos Terena monolíngue x último ano da graduação dos Terena monolíngues; 1<sup>o</sup> ano da graduação dos Não Indígenas x último ano da graduação dos Não Indígenas.

O objetivo de contrastar esses grupos é averiguar se a presença dos itens solicitados, através das etiquetas, é suficiente para caracterizá-los como um fator de relevância na escrita dos grupos indígenas. Para verificar se a diferença normalizada é de fato, significativa e não aleatória foi utilizado o teste estatístico *Likelihood* (LL). Esse teste, conforme Rayson (2002, apud SARMENTO, 2008), tem se revelado o mais eficiente para a comparação de frequência de palavras ou expressões entre dois corpora.

Se o resultado após a aplicação do teste for de 6,63 ou mais, a probabilidade da diferença entre os dois corpora ter acontecido de forma aleatória é de menos de 1%, então o resultado tem significância. O resultado vem expresso como  $p < 0.01$ . O endereço eletrônico do *Log Likelihood Calculator* é: <http://ucrel.lancs.ac.uk/llwizard.html>.

#### 5.6.8 O Material Pesquisado

Os textos coletados para este trabalho de uma forma geral oferecem variados problemas de escrita, conforme a tradição escrita exigida pela instituição escolar preconiza como textos ‘bem escritos’, ou seja, que obedeçam as normas gramaticais da língua portuguesa. Os problemas de escrita permeiam todos os textos ora de forma semelhante ora diferente. Uma leitura geral será realizada no capítulo referente às categorias de descrição pontuando os desvios das normas do português. Devido ao espaço e ao tempo não seria possível observar todas as questões linguísticas presentes nos textos que seriam interessantes para esta pesquisa.

##### 5.6.8.1 A Seleção do material

Após a coleta, a digitalização e a transferência para o *Bloco de notas*, uma vez que o programa AntConc não lê os dados do *office word*, os textos foram etiquetados, conforme a subseção 5.6.7. O grupo infantil foi colocado em pastas diferentes do grupo adulto. A descrição dos dados se deu, conforme dito, inicialmente de forma a destacar questões mais saliente nos textos, como a forma que era empregada a ortografia, por exemplo, e só após essa descrição geral, os dados foram passados pelo programa de computador, obedecendo aos nove critérios estabelecidos. Um dos critérios está relacionado à concordância nominal de gênero pelos alunos indígenas. Abaixo, estão exemplos de alguns desses critérios nos textos dos alunos do grupo infantil desse grupo:

QUADRO 7 - Exemplo do critério concordância nominal de gênero em texto do grupo infantil

ALCKBPJERM5FTXT01

**A minha familia e muito boma minha familia tem muito as pessoas**Que moro no **minha casa** minha mae e legao meu Pai e legão todas **as minha erma e legão** poriso Que Eu gosto da minha mae meu Pai **gosto da minha família nos diverti tem o melhor erma e de Bringao gogo Bola vole** e muito bom avori também voreis

A concordância de gênero não ocorre, conforme o predisposto pelas normas de escrita do português culto exigido pela escola.

Outra questão peculiar nos textos é a forma como os verbos são utilizados. Além do emprego dos verbos, outros critérios chamam a atenção. Veja alguns exemplos no texto abaixo<sup>32</sup>:

QUADRO 8 - Exemplo de alguns critérios: erro de tempo verbal, ausência de determinante, hipossegmentação, ausência de concordância de número e pessoal, ausência de concordância nominal de gênero em texto do grupo infantil

DSOKBPEETM4FTXT01

- 1 quando Era crianca Eu Brincava.
- 2 **-etamen** Estudava muito, mesmo.
- 3-itanbem **Eu brincava** muito mesmo jogava bola brincava, de buneca.
- 4-itamem **Eu vou** com meu pai **naçidade** fazer compra.
- 5-**vou passia** na minha tia, **vou na casa** da minha vô.
- 6-quando **Era** criança Eu **andava** de bicicleta **andava caia** Leva ntava. quando Era creanca,
- 7 a menha amega mi batia **mimordia**, **quando meu amego foi buscar cavalo** **Eu foi com Ele** Eu
- 8 vi Ele subiu **encema**, do cavalo Eu vi Ele caiu di ma do cavalo. etamem, quando **nois foi** na
- 9 engreja nos fomo, **di carroço** nos não vimos, buraco grandão enos caemo, lã Eu machoguei
- 10meu pes, çoço é menha mãe machuco, braço dela o meu pai machuco, cabeça dele.
- 11- **nos sararam.**

Os problemas ortográficos são de longe o que mais chama a atenção nos textos. No entanto, nesse momento serão comentados de forma breve, ficando para o capítulo da discussão dos resultados. O problema com o emprego dos verbos é o que interessa nesse texto. As passagens negritadas mostram que o verbo no pretérito imperfeito – próprio para narrativas junto com o pretérito perfeito – aparece no começo do texto para em seguida dar lugar ao verbo no presente para continuar elencando as ações ocorridas no passado: eu **brincava** muito (linha03); eu **vou** com meu pai(linha04). Há uma ruptura no tempo, do passado para o presente e na sequência da narrativa. Novamente, há ligeira confusão com a pessoa verbal: quando **meuamego foi** buscar cavalo **Eu foi** com Ele(linha 07). Finalizado,

<sup>32</sup>Foi colocado número de linhas para facilitar a leitura dos dados.

na linha 11: nos **sararam** (nós saramos), outro equívoco, que pode ser também de pessoa (eles sararam).

Além desses, exemplificados, a presença de hipossegmentação (palavras que deveriam ser escritas separadas e são escritas juntas) e hipersegmentação (palavras que deveriam ser escritas juntas e são escritas separadas), também serão alvo da nossa investigação. Seguem os exemplos, primeiro de hipossegmentação:

QUADRO 9 - Exemplo do critério hipossegmentação em texto do grupo infantil

JSVBKPRETM5FT01

- 1 Quando **euera** crianca **eugostava** De brincar mecher ali mechela eu era baguncera
- 2 **umdia** a mia mãe levou eu **nuParque** aí eu comecei a brincar Depois cabou o Parque aí Amia
- 3 mae levou eu nurestaura-nte aí quando eustava comendo pucheí a tualha Da mesa e caiu Tudo
- 4 os copo De vidro quebrou tudinho aí eutevi que pasar o maior vergonha quandocheguei
- 5 emcasa elamíbateu

Na linha 01 dois momentos de hipossegmentação: *euera, eugostava*; na linha 02: *umdia, nuParque*; na última linha: *elamíbateu*. Vários exemplos de hipossegmentação nesse texto. Ainda neste mesmo textos há um exemplo de ausência de concordância verbal no final da linha 3 e começa da linha 04: *Da mesa e caiu Tudo os copo De vidro quebrou tudinho*.

Por fim, um exemplo da presença da hipersegmentação:

QUADRO 10 - Exemplo do critério hipersegmentação em texto do grupo infantil

ADDGBPLERM4FT3

e Bom ser indío a gente vai pra escola pra estudar eu Brinco cominhas **a migas** agente São guarani agente Brinca sedo nos linpa a casa linpa a cosinha onze Hora agente Toma banho pra ir Pra escola Pra estudar

Os critérios que foram selecionados para a etiquetagem foram divididas em nove categorias de análises, porém a presença de outras questões linguísticas não serão descartadas, serão brevemente apresentadas. No entanto, a ênfase recairá nessas nove, por isso foram etiquetadas e lidas pelo programa *AntConc*, pela ferramenta *Concord* (concordanciador). Com relação ao papel do programa *AntConc* a seção seguinte trará mais informações.

#### 5.6.8.2. As Técnicas para a Análise dos Dados Conforme as Ferramentas Computacionais do Programa *AntConc*

A opção por analisar os dados deste trabalho por meio das ferramentas da LdC já foi explicada no início desta seção. Passa-se, agora, à descrição de como funciona a ferramenta disponível para isso. Como se trata de um trabalho em uma comunidade específica, de uma determinada região não havia como acessar os dados via *corpora* eletrônico. Dessa forma, os textos foram coletados manualmente e, posteriormente, foram digitados e transformados em *corpora* digital pela própria pesquisadora. Primeiro, a digitação no *Word for Windows* em pastas separadas por cada série escolar e depois se criou outra pasta para cada texto receber a etiquetagem manual para que as ferramentas do programa identificassem os critérios anteriormente estabelecidos.

Por se objetivar uma análise diagnóstica dos dados, ou seja, para se obter um panorama de como ocorre a escrita das crianças indígenas e dos acadêmicos indígenas dessa comunidade especificamente, um olhar sobre os padrões que porventura apareçam em seus textos é essencial. As inferências a cerca da escrita dessa comunidade, começa a partir desse ponto para, talvez poder oferecer opções de materiais didáticos condizentes com a realidade da escrita, ou dos “problemas” de escrita, se confrontados com a língua dita padrão.

Após a etiquetagem, os dados serão submetidos ao programa computacional *AntConc*, apesar de haver outros, esse foi escolhido por ser o mais utilizado. Ele oferece três ferramentas que auxiliam a análise linguística do *corpus*: gerador de listas de palavras, extrator de palavras-chave e concordanciador. No caso deste trabalho o concordanciador se mostra mais eficaz para alcançar o intento de analisar o sintagma, pois se observará uma palavra central e as periféricas do seu lado direito e esquerdo, até para ver se há algum padrão próprio na escrita desses grupos.

#### 5.6.8.3 O Concordanciador

Na LdC, a ferramenta do *AntConc* responsável pela análise de uma palavra em seu ambiente linguístico com o objetivo de empreender uma observação qualitativa dos dados obtidos (co-texto) chama-se *Concord* (concordanciador). O co-texto de uma palavra constitui-se como os itens que estão dispostos à direita e à esquerda da palavra de busca. Os resultados apresentados pelo concordanciador estão disponibilizados em linhas de concordância, pois são fragmentos curtos, em uma linha apenas (VIANA, 2010).

No próximo capítulo, quantificaremos os dados dos dois grupos, a partir do programa *AntConc*. Com as ferramentas *Lista de Palavras (Wordlist)* e *Concordanciador (Concord)*, será possível a leitura das etiquetas dos textos e teremos um panorama do que ocorre na escrita de cada texto.

## 6. ANÁLISE QUANTITATIVA DOS DADOS

### 6.1 Análise Descritiva

#### 6.1.1 Análise do grupo de crianças do 4º ano do ensino fundamental

A amostra total dos textos do grupo de crianças totalizou 483 textos, resultado da produção escrita de 175 crianças, sendo 71 bilíngues Guarani/Kaiowá, 40 Terena monolíngues e 64 não indígenas. Esses textos foram submetidos à ferramenta *concord* do programa *AntConc*, que mostrou os seguintes resultados:

Tabela 20. Quantificação dos dados do grupo de crianças Guarani/Kaiowá do 4º ano do ensino fundamental

<b>4º Ensino Fundamental – Guarani/Kaiowá bilíngue</b>	<b>Tamanho do <i>corpus</i> em <i>Tokens</i>: 8.204</b>	<b>%</b>
HIPER	130	15,84
HIPO	527	64,23
SND	33	4,02
SNG	73	8,89
SNN	83	10,11
SVNP	114	13,89
SVG	10	1,21
SVT	29	3,53
SVC	15	1,82

Num total de 8.204 palavras, o subgrupo indígena Guarani/Kaiowá bilíngue do 4º ano do EF mostrou que a hipossegmentação foi o item mais recorrente nos textos, ocorrendo 527 vezes, perfazendo um percentual de 64,23% da amostra. O segundo item mais recorrente foi a hipersegmentação, com 130 ocorrências, perfazendo 15,84% do total. A

ausência ou erro da concordância verbal de número e pessoal – SVNP –, foi a 3<sup>a</sup> que mais apareceu neste subgrupo, 114 vezes, perfazendo um total de 13,89%. Por fim, a ausência ou erro da concordância de número – SNN –, foi a 4<sup>a</sup> que mais apareceu neste subgrupo, 83 vezes, ou seja, 10,11%.

Na tabela subsequente, que apresenta os dados do subgrupo Terena monolíngue – também pertencente ao 4<sup>o</sup> ano do EF, as ocorrências de hiposegmentação e de hipersegmentação se mantiveram em primeiro lugar, mas as outras se modificaram um pouco:

Tabela 21. Quantificação dos dados das crianças Terena monolíngues de português do 4<sup>o</sup> ano do ensino fundamental

<b>4<sup>o</sup> Ensino Fundamental – Terena monolíngue</b>	<b>Tamanho do <i>corpus</i> em <i>Tokens</i>: 4.542</b>	<b>%</b>
HIPER	149	32,80
HIPO	300	66,05
SND	4	0,88
SNG	59	12,98
SNN	26	5,72
SVNP	37	8,14
SVG	3	0,66
SVT	10	2,20
SVC	2	0,44

Num universo textual de 4.542 palavras, a hiposegmentação ocupou 66,05% da amostra, ocorrendo 300 vezes. A hipersegmentação a acompanhou com 149 aparições, ou seja, 32,80%. Já o terceiro lugar foi ocupado pela ausência ou erro de concordância de gênero – SNG –, que se repetiu por 59 vezes, gerando um percentual de 12,98% das ocorrências. Em quarto lugar, novamente apareceu o SVNP com 37 ocorrências e 8,14% do total. Uma análise superficial dos dados do 4<sup>o</sup> ano do ensino fundamental dos dois subgrupos indígenas em comparação com o grupo de crianças não indígenas, podemos afirmar que a hiposegmentação, de fato ocorre com certa frequência nas séries iniciais e nos subgrupos indígenas sua incidência parece ser maior. Uma possível significância estatística nessa comparação, no entanto, será verificada mais adiante. Veja as ocorrências encontradas nas amostras do subgrupo não indígena na tabela abaixo:

Tabela 22. Quantificação dos dados das crianças não indígenas do 4º ano do ensino fundamental

<b>4º Ensino Fundamental - Não indígena Escola Castro Alves</b>	<b>Tamanho do <i>corpus</i> em <i>Tokens</i>: 10.308</b>	<b>%</b>
HIPER	26	2,52
HIPO	72	6,98
SND	17	1,64
SNG	47	4,55
SNN	54	5,23
SVNP	54	5,23
SVG	2	0,19
SVT	24	2,32
SVC	1	0,09

Claramente se percebe que, num universo de 10.308 textos, a incidência desses dois itens é pequena, sendo que a hipersegmentação aparece em quarto lugar nas recorrências. O segundo item que mais ocorre, empatados, estão ausência ou erro de concordância de número – SNN – e ausência ou erro de concordância verbal de número e pessoa – SVNP. No terceiro lugar está a ausência ou erro de concordância de gênero – SNG –, seguido pela hipersegmentação.

#### 6.1.2 Análise do grupo de crianças do 5º ano do ensino fundamental

Do subgrupo indígena Guarani/ Kaiowá bilíngue do 5º ano do ensino fundamental, podemos observar o que se segue:

Tabela 23. Quantificação dos dados das crianças Guarani/Kaiowá do 5º ano do ensino fundamental

<b>5º Ensino Fundamental – Guarani/Kaiowá bilíngue</b>	<b>Tamanho do <i>corpus</i> em <i>Tokens</i>: 8.816</b>	<b>%</b>
HIPER	93	10,54
HIPO	242	27,45
SND	19	2,15
SNG	101	11,45
SNN	78	8,84
SVNP	71	8,05
SVG	2	0,22
SVT	36	4,01
SVC	7	0,79

A hipossegmentação continua a se apresentar como o item de maior ocorrência, porém com um percentual bastante reduzido, 27,45%, ou seja, 242 ocorrências num total de 8.816 textos. Na sequência surge a ausência ou erro de concordância de gênero com 101 ocorrências, isto é, 11,45%. Esse item chama a atenção, pois este foi o único subgrupo em que a ausência ou erro de gênero ocorreu com bastante frequência. O terceiro item que mais aparece neste subgrupo é a hipersegmentação, com 10,54% (93 vezes), e a ausência ou erro de concordância de número, com 78 aparições, totalizando 8,84% das ocorrências.

Tabela 24. Quantificação dos dados das crianças Terena do 5º ano do ensino fundamental

<b>5º Ensino Fundamental – Terena monolíngue</b>	<b>Tamanho do corpus em Tokens: 6.466</b>	<b>%</b>
HIPER	70	10,82
HIPO	144	22,27
SND	13	2,01
SNG	34	5,25
SNN	58	8,96
SVNP	58	8,96
SVG	4	0,61
SVT	19	2,93
SVC	0	-

No 5º ano do EF, formado pelas crianças Terena monolíngue, a hipossegmentação também é o elemento linguístico que mais ocorre: 144 vezes, ou seja, 22,27%, e a hipersegmentação ocupa o segundo lugar, com 70 ocorrências e 10,82% respectivamente. Em terceiro lugar, estão empatados a ausência ou erro de concordância de número – SNN – e a ausência ou erro de concordância verbal de número e pessoa – SVNP –, com 58 ocorrências, totalizando um percentual de 8,96 cada.

No subgrupo de crianças não indígenas, podemos perceber que, numa amostra de 5.729 palavras, a hipossegmentação continua ocupando o primeiro lugar de repetições, no entanto, em número extremamente menor que nos textos anteriores. São apenas 33 ocorrências, chegando a um percentual de 5,76% dos textos. Em segundo lugar, aparece a ausência ou erro de concordância verbal de número e pessoa – SVNP – com 31 ocorrências, ou seja, 5,41%. Em terceiro lugar, aparece a ausência ou erro de concordância de número – SNN – com 21 aparições e percentual de 3,66%. Enfim, com 13 ocorrências, correspondendo a 2,26%, está a ausência ou erro de concordância de gênero – SNG. Quase não há problemas de hipersegmentação neste grupo.

Tabela 25. Quantificação dos dados das crianças não indígenas do 5º ano do ensino fundamental

<b>5º Ensino Fundamental - Não indígena</b>	<b>Tamanho do corpus em Tokens: 5.729</b>	<b>%</b>
HIPER	5	0,87
HIPO	33	5,76
SND	1	0,17
SNG	13	2,26
SNN	21	3,66
SVNP	31	5,41
SVG	1	0,17
SVT	8	1,39
SVC	2	0,34

## 6.1.3 Análise do grupo de adultos do 1º ano do ensino superior

Tabela 26. Quantificação dos dados dos adultos indígenas Guarani/Kaiowá bilíngues do 1º ano do ensino superior

<b>1º Ensino Superior - Guarani/Kaiowá bilíngue</b>	<b>Tamanho do corpus em Tokens: 9.641</b>	<b>%</b>
HIPER	8	0,82
HIPO	20	2,07
SND	15	1,55
SNG	40	4,14
SNN	88	9,12
SVNP	68	7,05
SVG	3	0,31
SVT	44	4,56
SVC	8	0,82

No grupo 2, formado pelos adultos, o 1º ano do ensino superior indígena Guarani/Kaiowá bilíngue apresenta a ausência ou erro de concordância nominal de número - SNN - como preponderante com 88 ocorrências, o que corresponde a 9,12%. Em seguida, aparece a ausência de concordância verbal de número e pessoa - SVNP - com 68 ocorrências, perfazendo 7,05%. Na sequência, com 4,56% de ocorrências, 44 vezes, está o erro de tempo verbal - SVT . Foi a primeira vez que essa ocorrência apareceu entre os quatro primeiros critérios de análise. Por último, com 4,14% e 40 aparições, SNG - ausência

ou erro de concordância de gênero. Problemas de hipo e hipersegmentação não foram significativos nesse grupo.

Na tabela seguinte, do grupo 2, 1<sup>o</sup> ano do ensino superior Terena monolíngue, foram encontradas como as mais relevantes ocorrências também em primeiro e em segundo lugar o SNN e o SVNP – ausência ou erro de concordância de número e ausência ou erro de concordância de número e pessoa, respectivamente com 4,86% (38vezes) o primeiro e 3,45 (27vezes) o segundo num universo de 7.809 *tokens*. Aparece em terceiro lugar, a hipossegmentação com 23 ocorrências (2,94%). Finalizando, com 2,17%, ou 17 vezes aparece a ausência ou erro de concordância de gênero – SVG – assim como ocorreu com o 1<sup>o</sup> ano do ensino superior do grupo Guarani/Kaiowá bilíngue, a única diferença é que no subgrupo anterior ocorreu ao erro de tempo verbal – SVT - nesse, substituído pela hipossegmentação.

Tabela 27. Quantificação dos dados dos adultos indígenas Terena monolíngues do 1<sup>o</sup> ano do ensino superior

<b>1<sup>o</sup> Ensino Superior – Terena monolíngue</b>	<b>Tamanho do corpus em Tokens: 7.809</b>	<b>%</b>
HIPER	7	0,89
HIPO	23	2,94
SND	8	1,02
SNG	17	2,17
SNN	38	4,86
SVNP	27	3,45
SVG	2	0,25
SVT	11	1,40
SVC	0	-

Na tabela 28, abaixo, o subgrupo não indígena apresentou os seguintes resultados:

Tabela 28. Quantificação dos dados dos adultos não indígenas do 1º ano do ensino superior

<b>1º Ensino Superior – não indígena</b>	<b>Tamanho do corpus em Tokens: 6.777</b>	<b>%</b>
HIPER	2	0,29
HIPO	4	0,59
SND	4	0,59
SNG	5	0,73
SNN	15	2,21
SVNP	14	2,06
SVG	1	0,14
SVT	13	1,91
SVC	1	0,14

Dentro desse subgrupo formado por acadêmicos de 1º ano de ensino superior, o subgrupo não indígena repetiu praticamente o mesmo comportamento, a saber, SNN, SVNP, SVT e SNG. Nesse subgrupo especificamente, a ausência ou erro de concordância de número (SNN) ocorreu 15 vezes (2,21%); a ausência ou erro de concordância verbal de número e pessoa (SVNP), 14 vezes (2,06%); o erro de tempo verbal (SVT), 13 vezes (1,91%) e a ausência ou erro de concordância de gênero (SNG), 05 vezes, ou 0,73%. A grande diferença ficou por conta do subgrupo Terena monolíngue que trouxe a hipossegmentação e os demais não o fizeram, mas trouxeram o erro de tempo verbal - SVT - o que o subgrupo Terena monolíngue não fez.

#### 6.1.4 Análise do grupo de adultos do último ano do ensino superior

Tabela 29. Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Guarani/Kaiowá bilíngue

<b>Último Ensino Superior – Guarani/Kaiowá bilíngue</b>	<b>Tamanho do corpus em Tokens: 9.642</b>	<b>%</b>
HIPER	5	0,51
HIPO	11	1,14
SND	16	1,65
SNG	42	4,35
SNN	58	6,01
SVNP	75	7,77
SVG	5	0,51
SVT	24	2,48
SVC	7	0,72

No subgrupo indígena Guarani/kaiowá bilíngue do último ano do curso superior do grupo 2, ocorreu praticamente o mesmo que no 1<sup>o</sup> ano. A ausência ou erro de concordância verbal de número e pessoa (SVNP), 75 vezes (7,77%); a ausência ou erro de concordância de número (SNN) ocorreu 58 vezes (6,01%); a ausência ou erro de concordância de gênero (SNG), 42 vezes, ou 4,35%. Finalmente, em quarto lugar, o erro de tempo verbal (SVT), 24 vezes, resultando em 2,48%. No último ano desse grupo, as ocorrências de SNN aparecem em primeiro lugar nos primeiros anos e em segundo, aparece SVNP. No último ano desse subgrupo isso se inverte, primeiro ocorrem os SVNPs e em segundo, os SNNs.

Na tabela 33, na sequência, do grupo 2, subgrupo Terena monolíngue as ocorrências praticamente se repetem:

Tabela 30. Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Terena monolíngue

<b>Último Ensino Superior – Terena Monolingue</b>	<b>Tamanho do corpus em Tokens: 11.569</b>	<b>%</b>
HIPER	2	0,17
HIPO	5	0,43
SND	13	1,12
SNG	22	1,90
SNN	55	4,75
SVNP	50	4,32
SVG	2	0,17
SVT	9	0,77
SVC	3	0,29

Com o último ano do subgrupo Terena monolíngue, se invertem o primeiro e o segundo lugar com relação ao subgrupo dos Guarani/Kaiowá bilíngues, mostrados na tabela anterior: a ausência ou erro de concordância de número (SNN) se fez 58 vezes, ou seja, 4,75%; já a ausência ou erro de concordância verbal de número e pessoa (SVNP), se repetiu por 50 vezes, num percentual de 4,32%; igual na tabela anterior, o terceiro lugar foi ocupado pela ausência ou erro de concordância de gênero (SNG), que se realizou 22 vezes, ou 1,90%. Por fim, em quarto e último lugar, e pela primeira vez nos dados, aparece a ausência de determinante – SND – 13 ocorrência num percentual de 1,12%.

Por fim, o último ano do ensino superior do subgrupo não indígena do grupo 2, foram obtidos os seguintes dados:

Tabela 31. Quantificação dos dados dos adultos indígenas do último ano do ensino superior, grupo 2 – Não Indígena

Último Ensino Superior – Não Indígena	Tamanho do corpus em Tokens: 10.656	%
HIPER	0	-
HIPO	1	0,09
SND	3	0,28
SNG	4	0,37
SNN	15	1,40
SVNP	13	1,21
SVG	1	0,09
SVT	15	1,40
SVC	1	0,09

Nesse subgrupo também não foi muito diferente. As ocorrências morfossintáticas mais relevantes se deram prevalecendo a ausência ou erro de concordância de número (SNN) se fez 15 vezes, ou seja, 1,40%; em segundo, aparece empatado com o primeiro lugar, o erro de tempo verbal (SVT) com 15 ocorrências, 1,40%. Em terceiro lugar, a ausência ou erro de concordância verbal de número e pessoa (SVNP), se repetiu por 13 vezes, num percentual de 1,21% e a ausência ou erro de concordância de gênero (SNG), que se realizou 04 vezes, ou 0,37%.

No grupo 2, pode-se perceber que as repetições ocorridas foram muito parecidas. Lembrando que a única ocorrência significativa de hipossegmentação, apareceu apenas no subgrupo Terena monolíngue.

## 6.2 Análise Estatística dos Dados

Como visto no Capítulo da Metodologia, a análise quantitativa dos dados foi realizada através do teste estatístico *Log Likelihood* (LL). Os resultados apresentados e discutidos abaixo foram analisados considerando um nível de significância de  $p < 0,01$ . A análise estatística levou em conta as variáveis faixa etária e língua/etnia. Na variável faixa



Crianças		Crianças										
		4º ano EF										
Adultos	4º ano EF	Bilíngues Kaiová/Guarani	8.204	130 (15,84%)	527 (64,23%)	33 (4,02%)	73 (8,89%)	83 (10,11%)	114 (13,89%)	10 (1,21%)	29 (3,53%)	15 (1,82%)
		Monolíngues Terena	4.542	149(32,80%)	300(66,05%)	04(0,88%)	59(12,98%)	26(5,72%)	37(8,14%)	3 (0,66%)	10(2,20%)	2(0,44%)
		Monolíngues Portugêses	10.308	26 (2,52%)	72 (6,98)	17 (1,64%)	47 (1,64%)	54 (5,23%)	54 (5,23%)	2 (0,19%)	24 (2,32%)	1 (0,09%)
	5º ano EF	Bilíngues Kaiová/Guarani	8.816	93 (10,54%)	242(27,45%)	19(2,15%)	101(11,45%)	78(8,84%)	71(8,05%)	2(0,22%)	36(4,01%)	7(0,79%)
		Monolíngues Terena	6.466	70(10,82)	144(22,27%)	13(2,01%)	34 (5,25%)	58(8,96%)	58(8,96%)	4(0,61%)	19(2,93%)	0(0%)
		Monolíngues Portugêses	5.729	5(0,87%)	33(5,76%)	1(0,17%)	13(2,26%)	21(3,66)	31(5,41%)	1(0,17%)	8(1,39%)	2(0,34%)
	1º ano ES	Bilíngues Kaiová	9.641	8(0,82%)	20(2,07%)	15(1,55%)	40(4,14%)	88(9,12%)	68(7,05%)	3(0,31%)	44(4,56%)	8(0,82%)
		Monolíngues Terena	7.809	7(0,89%)	23(2,94%)	8(1,02%)	17(2,17%)	38(4,86%)	27(3,45%)	2(0,25%)	11(1,40%)	0(0%)
		Monolíngues Portugêses	6.777	2(0,29%)	4(0,59%)	4(0,59%)	5(0,73%)	15(2,21%)	14(2,06%)	1(0,14%)	13(1,91%)	1(0,14%)
Bilíngues Kaiová/Guarani		9.642	5(051%)	11(1,14%)	16(1,65%)	42(4,35%)	58(6,01%)	75(7,77%)	5(051%)	24(2,48%)	7(0,72%)	
Monolíngues Terena		11.569	2(0,17%)	5(0,43%)	13(1,12%)	22(1,90%)	55(4,75%)	50(4,32%)	2(0,17%)	9(0,77%)	3(0,29%)	
Monolíngues Portugêses		10.656	0(0%)	1(0,09%)	3(0,28%)	4(0,37%)	15(1,40%)	13(1,21%)	1(0,09%)	15(1,40%)	1(0,09%)	
Último ano ES	Bilíngues Kaiová/Guarani	9.642	5(051%)	11(1,14%)	16(1,65%)	42(4,35%)	58(6,01%)	75(7,77%)	5(051%)	24(2,48%)	7(0,72%)	
	Monolíngues Terena	11.569	2(0,17%)	5(0,43%)	13(1,12%)	22(1,90%)	55(4,75%)	50(4,32%)	2(0,17%)	9(0,77%)	3(0,29%)	
	Monolíngues Portugêses	10.656	0(0%)	1(0,09%)	3(0,28%)	4(0,37%)	15(1,40%)	13(1,21%)	1(0,09%)	15(1,40%)	1(0,09%)	

A análise estatística foi realizada com base nos objetivos específicos da pesquisa, retomados no decorrer do texto a seguir. O primeiro objetivo era verificar se existem diferenças entre os tipos de dificuldades reveladas pelos indígenas bilíngues e monolíngues de diferentes etnias – Kaiowá/Guarani e Terena – em comparação com os monolíngues não indígenas na produção de textos narrativos em português. Ou seja, visava-se verificar em que medida o processo de letramento indígena apresenta peculiaridades, que podem porventura se estender até a vida acadêmica. Para isso, foi calculado o *Log Likelihood* a partir da comparação entre os grupos de informantes indígenas com os não indígenas e também comparando-se os grupos indígenas entre si, tanto de crianças quanto o de adultos. Foram levados em consideração o número total de *tokens* produzidos pelas crianças de 4º e de 5º anos do Ensino Fundamental agrupados bem como o dos adultos de 1º e último ano do Ensino Superior agrupados.

A análise revelou que, na comparação entre as crianças indígenas Guarani/Kaiowá e as crianças indígenas Terena, houve diferença significativa entre os grupos somente no número de casos de Hipersegmentação e de Ausência de Cópula, sendo que as crianças Terena produziram um número maior de casos de Hipersegmentação (0,019%) do que as crianças Guarani/Kaiowá (0,013%) e as crianças Guarani/Kaiowá produziram um número maior de casos de Ausência de Cópula (0,013%) do que as crianças Terena (0,001%).

Tabela 34. Comparação entre crianças indígenas Guarani/Kaiowá e crianças indígenas Terena – grupo 1

	Crianças indígenas Guarani/Kaiowá (17.020 <i>tokens</i> )		Crianças indígenas Terena (11.008 <i>tokens</i> )		<i>LL</i>
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	223	0,013	219	0,019	-19,11
HIPO	769	0,045	444	0,040	+3,66
SND	52	0,030	17	0,015	+6,60
SNG	174	0,010	93	0,084	+ 2,24
SNN	161	0,094	84	0,076	+2,60
SVNP	185	0,011	95	0,086	+3,42
SVG	12	0,007	7	0,006	+0,05
SVT	65	0,038	29	0,026	+2,89
SVC	22	0,013	2	0,001	+11,92
Total	1663	0,097	990	0,089	

Na análise da comparação entre as crianças indígenas Guarani/Kaiowá e as crianças não indígenas, falantes monolíngues de português, houve ocorrências de casos de quase

todas as estruturas estudadas, com exceção de problemas de marcação de gênero no sintagma verbal, sendo que as crianças indígenas apresentaram um número significativamente maior de ocorrências do que as crianças não indígenas, como mostra a tabela abaixo.

Comparação entre crianças indígenas Guarani/Kaiowá e crianças não indígenas, falantes monolíngues de português

Tabela 35. Comparação entre crianças indígenas Guarani/Kaiowá e crianças não indígenas – grupo 1

	Crianças indígenas Guarani/Kaiowá (17.020 tokens)		Crianças não indígenas (16.037 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	223	1,310	31	0,193	+ 152,46
HIPO	769	4,518	105	0,654	+ 531,03
SND	52	0,305	18	0,112	+ 15,27
SNG	174	1,022	60	0,374	+ 51,40
SNN	161	0,945	75	0,467	+ 27,16
SVNP	185	1,086	85	0,530	+32,22
SVG	12	0,070	3	0,018	+ 5,26
SVT	65	0,381	32	0,199	+ 9,58
SVC	22	0,129	3	0,018	+ 15,20
Total	1663	9,770	412	2,569	

A análise comparando as crianças indígenas Terena e as crianças não indígenas, falantes monolíngues de português, revelou que as crianças Terena apresentam um número significativamente maior de ocorrências de Hipersegmentação, Hipossegmentação, problemas de marcação de gênero e de número no sintagma nominal, além de dificuldades de marcação de número e pessoa no sintagma verbal, como mostra a tabela abaixo.

Tabela 36. Comparação entre crianças indígenas Terena monolíngues e crianças não indígenas – grupo 1

	Crianças indígenas Terena (11.008 tokens)		Crianças não indígenas (16.037 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	219	1,989	31	0,193	+ 238,70
HIPO	444	4,033	105	0,654	+ 372,08
SND	17	0,154	18	0,112	+ 0,88
SNG	93	0,844	60	0,374	+ 24,97
SNN	84	0,763	75	0,467	+ 9,49
SVNP	95	0,863	85	0,530	+ 10,65
SVG	7	0,063	3	0,018	+ 3,50
SVT	29	0,263	32	0,199	+ 1,17
SVC	2	0,018	3	0,018	- 0,00
Total	1663	0,977	412	2,569	

Com relação ao grupo adulto, grupo 2, na comparação entre o primeiro e o último ano do ensino superior, Guarani/Kaiowá bilíngue e Terena Monolíngues em português e percebeu-se o que se segue:

Tabela 37. Comparação entre Adultos indígenas Guarani/Kaiowá bilíngues e Adultos indígenas Terena monolíngues – grupo 2

	Adultos indígenas Guarani/Kaiowá (19.283 tokens)		Adultos indígenas Terena (19.378 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	13	0,67	9	0,46	+ 0,75
HIPO	32	1,65	28	1,44	+ 0,29
SND	31	1,60	21	1,08	+ 1,98
SNG	82	4,25	39	2,01	+ 15,83
SNN	146	7,57	93	4,79	+ 12,11
SVNP	143	7,41	77	3,97	+ 20,43
SVG	8	0,41	4	0,20	+ 1,38
SVT	64	3,31	20	1,03	+ 24,45
SVC	15	0,77	3	0,15	+ 8,79
Total	534	27,69	294	15,17	

As ocorrências significativas na comparação desse grupo de ensino superior mostraram que os Guarani/Kaiowá produziram quantidade significativa de ausência ou erro de concordâncias, tanto nominais (SNG/SNN), quanto verbais (SVNP/SVT/SVC). A hipo e a hipersegmentação quase não apareceram nesses textos. No subgrupo Guarani/Kaiowá em comparação com português monolíngue aconteceu o seguinte:

Tabela 38. Comparação entre Adultos indígenas Guarani/Kaiowá bilíngues e Adultos Não Indígenas – grupo 2

	Adultos indígenas Guarani/Kaiowá (19.283 tokens)		Adultos não indígenas (17.433 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	13	0,67	2	0,11	+ 7,94
HIPO	32	1,65	5	0,28	+ 19,36
SND	31	1,60	7	0,40	+ 14,05
SNG	82	4,25	9	0,51	+ 60,30
SNN	146	7,57	30	1,72	+ 72,01
SVNP	143	7,41	27	1,54	+ 75,58
SVG	8	0,41	2	0,11	+ 3,28
SVT	64	3,31	28	1,60	+ 11,07
SVC	15	0,77	2	0,11	+ 9,98
Total	534	27,69	135	7,74	

As diferenças nessa comparação são mais evidentes. Todas as ocorrências estão presentes em número maior nos indígenas Guarani/Kaiowá do que nos monolíngues

português. A hipo e a hipersegmentação, por exemplo que não teve muita significância na comparação entre os subgrupos indígenas, neste subgrupo ganha relevância. O subgrupo Guarani/Kaiowá chama atenção no item ausência de cópula (SVC) com 15 ocorrências contra apenas 2 ocorrências no outro subgrupo.

No contraste entre Terena Monolíngue em português e não indígenas ocorreu o seguinte:

Tabela 39. Comparação entre Adultos indígenas Terena monolíngue e Adultos Não Indígenas – grupo 2

	Adultos indígenas Terena monolíngue (19.378 tokens)		Adultos não indígenas (17433 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	9	0,46	2	0,11	+ 4,11
HIPO	28	1,44	5	0,28	+ 15,34
SND	21	1,08	7	0,40	+ 5,92
SNG	39	2,01	9	0,51	+ 17,18
SNN	93	4,79	30	1,72	+ 27,53
SVNP	77	3,97	27	1,54	+ 20,06
SVG	4	0,20	2	0,11	+ 0,48
SVT	20	1,03	28	1,60	- 2,32
SVC	3	0,15	2	0,11	+ 0,11
Total	294	15,17	135	7,74	

A ausência de cópula (SVC) não obteve relevância nesses subgrupos, assim como a hipersegmentação (HIPER), a ausência de determinante (SND), a ausência ou erro de concordância verbal de gênero (SVG) e erro de tempo verbal (SVT). Mas houve diferenças significativas também nas ausências ou erros de concordâncias nominais (SNG/SNN) e verbais (SVNP), não tão evidentes quanto da comparação entre os subgrupos Guarani/Kaiowa e não indígenas, na tabela anterior, se percebe que nesse subgrupo aparece forte também a hipossegmentação, já a hipersegmentação não é significativa como no subgrupo na tabela anterior.

Em resposta ao primeiro objetivo específico que era verificar se existem diferenças entre os tipos de dificuldades reveladas pelos indígenas bilíngues e monolíngues de diferentes etnias – Kaiowá/Guarani e Terena – em comparação com os monolíngues não indígenas na produção de textos narrativos escritos em português, os dados mostraram algumas questões interessantes. Com relação às crianças indígenas do subgrupo

Guarani/Kaiowá e Terena não se percebeu diferenças que chamassem a atenção. Todas possuem dificuldades parecidas no trato da língua portuguesa.

Quando se olha o subgrupo não indígena, percebe-se que eles também possuem dificuldades, o que muda é que com relação à hipossegmentação e à hipersegmentação, por exemplo, os indígenas extrapolam, e isso é algo a ser pensado. As crianças indígenas terena na comparação com as crianças não indígena trazem o *score* de ausência de determinante (SND) praticamente iguais, assim como o *score* de erro de tempo verbal (SVT) e ausência de cópula (SVC). Então não quer dizer que isso seja problema apenas do subgrupo indígena.

Com o grupo adulto, ocorrem as mesmas dificuldades. Continua acontecendo o fenômeno da hipo e da hipersegmentação, embora no subgrupo indígena um pouco mais, no entanto, as questões voltadas para as concordâncias todos os grupos possuem, os indígenas um pouco mais que os não indígenas.

O segundo objetivo específico era observar em que medida o caminho percorrido do ensino básico à formação acadêmica interferiu no desenvolvimento da habilidade de escrita de textos, na comparação entre os dois grupos etários, crianças e adultos. Nessa análise, foram comparadas as ocorrências de dificuldades nas estruturas morfossintáticas dos grupos de informantes crianças (4º e 5º anos do EF agrupados) e de adultos (1º e último ano do ES agrupados) em cada um dos grupos étnicos/ de língua.

Tabela 40. Comparação entre Grupo 1 Guarani/Kaiowá bilíngues e Grupo 2 Guarani/Kaiowá bilíngues – grupo 2

	Grupo 1 Guarani/Kaiowá bilíngues (17.020 tokens)		Grupo 2 Guarani/Kaiowá bilíngues (19.283 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	223	13,10	13	0,67	+ 253,66
HIPO	769	45,18	32	1,65	+ 936,75
SND	52	3,05	31	1,60	+ 8,32
SNG	174	10,22	82	4,25	+ 46,30
SNN	161	9,45	146	7,57	+ 3,80
SVNP	185	10,86	143	7,41	+ 11,91
SVG	12	0,70	8	0,41	+ 1,38
SVT	65	3,81	64	3,31	+ 0,63
SVC	22	1,29	15	0,77	+ 2,35
Total	1663	97,70	534	27,69	

No percurso do ensino fundamental ao nível superior, algumas ocorrências demonstram que o curso universitário não fez tanta diferença nas dificuldades de escrita

desse subgrupo indígena, basta observar que a maioria dos problemas de escrita que as crianças demonstram aparece nos textos dos adultos. Os mais evidentes são os relacionados à ausência ou erro de concordância nominal e verbal (SNN/SVNP). A hipo e a hipersegmentação que seriam inadmissíveis no contexto educacional adulto também aparecem. Evidentemente que há melhora, os casos de desvio da norma padrão diminuem, no entanto, ainda se manifestam, ainda que em número reduzido.

Abaixo, a tabela que mostra as dificuldades do grupo 1 – crianças terena monolíngues e grupo 2 – adultos terena monolíngues.

Tabela 41. Comparação entre Grupo 1 Terena monolíngue e Grupo 2 Terena monolíngue – grupo 2

	Grupo 1 Terena monolíngue (11.008 <i>tokens</i> )		Grupo 2 Terena monolíngue (19.378 <i>tokens</i> )		<i>LL</i>
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	219	19,89	9	0,46	+ 377,01
HIPO	444	40,33	28	1,44	+ 714,34
SND	17	1,54	21	1,08	+ 1,16
SNG	93	8,44	39	2,01	+ 63,71
SNN	84	7,63	93	4,79	+ 9,34
SVNP	95	8,63	77	3,97	+ 25,64
SVG	7	0,63	4	0,20	+ 3,39
SVT	29	2,63	20	1,03	+ 10,62
SVC	2	0,18	3	0,15	+ 0,03
Total	1663	97,70	294	15,17	

Nesses dois grupos, as dificuldades com hipo e hipersegmentação também deveriam ter sido sanadas no decorrer do processo educacional, embora o número de ocorrências se reduzissem bastante. Entretanto, essas dificuldades continuam aparecendo, o que significa dizer que o percurso até o ensino acadêmico, bem como o próprio ensino acadêmico em si não foi suficientemente competente para resolvê-los. As outras ocorrências também apontam para esta falha do ensino, haja vista que outras dificuldades que estão presentes nos textos infantis continuam aparecendo nos textos dos adultos. As questões referentes à ausência ou erro de concordância nominal ou verbal ainda são insistentes nesse nível do ensino, o que não poderia mais acontecer. Lembrando que as ocorrências desviantes diminuem com o decorrer da escolaridade, mas não desaparecem, como seria o esperado.

O resultado da comparação entre esse último grupo demonstra que o percurso até a universidade não foi satisfatório a ponto de reduzir totalmente as dificuldades de produção

escrita desse grupo indígena. Com isso, pode-se conjecturar que as dificuldades não estão na cultura ou na língua, uma vez que na comparação dos grupos indígenas que falam a língua materna, ou trazem-na consigo desde o berço esses desvios do padrão exigido na escrita de língua portuguesa ocorrem tanto quanto no grupo indígena que é monolíngue em português.

Tabela 42. Comparação entre Grupo 1 Não Indígena monolíngue e Grupo 2 Não Indígena monolíngue – grupo 2

	Grupo 1 não indígena monolíngue (16.037 tokens)		Grupo 2 não indígena monolíngue (17.433 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	31	1,93	2	0,11	+ 33,14
HIPO	105	6,54	5	0,28	+ 120,35
SND	18	1,12	7	0,40	+ 5,97
SNG	60	3,74	9	0,51	+ 46,60
SNN	75	4,67	30	1,72	+ 23,86
SVNP	85	5,30	27	1,54	+ 36,58
SVG	3	0,18	2	0,11	+ 0,29
SVT	32	1,99	28	1,60	+ 0,71
SVC	3	0,18	2	0,11	+ 0,29
Total	412	25,69	135	7,74	

O grupo 1 e 2 não indígena demonstrou, em sua comparação que os desvios da norma gramatical na escrita da língua portuguesa desde a infância também não foram sanados, pois ocorrências de hipo e hipersegmentação continuam ocorrendo de forma significativa. Confusões no emprego de tempos verbais (SVT), apesar de não vigorarem como fator estatístico de significância, praticamente permaneceram os mesmos no decorrer do processo de ensino, bem como a ausência ou erro de concordância nominal (SNN) e verbal (SVNP).

Finalmente, o último objetivo específico era investigar se o tempo de permanência no curso de graduação (alunos que estão no primeiro e no último ano de curso) interfere no nível de dificuldade na produção de textos. Para isso, foram comparadas as ocorrências de dificuldades morfossintáticas dos alunos do 1º ano do ES com as ocorrências demonstradas pelos alunos do último ano do ES de cada grupo étnico/ de língua.

Tabela 43. Comparação entre Primeiro Ano do Ensino Superior Guarani/Kaiowá bilíngue e Último ano do Ensino Superior - Guarani/Kaiowá – grupo 2

	1º ES - Guarani/Kaiowá (9.641 <i>tokens</i> )		Último ano ES - Guarani/Kaiowá (9.642 <i>tokens</i> )		<i>LL</i>
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	8	0,82	5	0,51	+0,70
HIPO	20	2,07	11	1,14	+2,65
SND	15	1,55	16	1,65	- 0,03
SNG	40	4,14	42	4,35	- 0,05
SNN	88	9,12	58	6,01	+ 6,21
SVNP	68	7,05	75	7,77	- 0,34
SVG	3	0,31	5	0,51	- 0,51
SVT	44	4,56	24	2,48	+ 5,97
SVC	8	0,82	7	0,72	+ 0,07
Total	294	30,49	243	25,20	

Num universo de mais de 9.000 *tokens*, foi possível observar que as dificuldades de escrita com as quais os acadêmicos adentraram o curso universitário não sofreram grandes alterações nesse subgrupo Guarani/Kaiowá do grupo 2. Não houve nenhuma diferença estatística, conforme o *Log Likelihood*, entretanto, ao olhar a tabela, as evidências se mostram, havendo dificuldades na escrita que aumentaram, como é o caso da ausência de determinante (SND), ausência ou erro de concordância de gênero (SNG), ausência ou erro de concordância verbal de número e pessoa (SVNP) e ausência ou erro de concordância verbal de gênero (SVG). Claro que há que se levar em conta outros fatores – como o próprio processo de letramento de cada indivíduo, no entanto, os dados apontam que houve equívocos gramaticais que apareceram no final da formação e isso, teoricamente não deveria ocorrer.

Tabela 44. Comparação entre Primeiro Ano do Ensino Superior Terena Monolíngue e Último ano do Ensino Superior - Terena Monolíngue – Grupo 2

	1º ES - Terena (7.809 <i>tokens</i> )		Último ano ES - Terena (11.569 <i>tokens</i> )		<i>LL</i>
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	7	0,89	2	0,17	+ 5,25
HIPO	23	2,94	5	0,43	+ 20,69
SND	8	1,02	13	1,12	- 0,04
SNG	17	2,17	22	1,90	+ 0,17
SNN	38	4,86	55	4,75	+ 0,01
SVNP	27	3,45	50	4,32	- 0,89
SVG	2	0,25	2	0,17	+ 0,15
SVT	11	1,40	9	0,77	+ 1,75
SVC	0	-	3	0,29	- 3,09
Total	133	17,03	161	13,91	

Se no subgrupo anterior as dificuldades não foram sanadas pelo curso universitário e alguns problemas aumentaram na escrita dos acadêmicos no grupo mostrado na tabela acima, não foi muito diferente. A diferença estatística relevante foi a presença da hipossegmentação. Como no grupo anterior, este grupo também teve algumas ocorrências que não apareciam quando entraram no curso, mas saíram com elas. A ausência de determinante (SND), a ausência ou erro de concordância de gênero (SNG), a ausência ou erro de concordância de número (SNN), a ausência ou erro de concordância verbal de número e de pessoa (SVNP) e a ausência de cópula (SVC) não faziam parte das dificuldades dos alunos que adentraram o curso, mas essas dificuldades aparecem no último ano.

Não se pode esquecer que se se tratasse de uma pesquisa longitudinal os resultados poderiam ser diferentes, mas mesmo assim não se pode ignorar esses resultados, uma vez que o que se espera é que o acadêmico termine um curso de graduação com um domínio razoável das normas linguísticas do português para a escrita.

Por fim, na tabela abaixo, a comparação entre os 1<sup>os</sup> anos do ensino superior do Grupo 2 – subgrupo não indígena e último ano. Veja os resultados:

Tabela 45. Comparação entre Primeiro Ano do Ensino Superior Não Indígena e Último ano do Ensino Superior Não Indígena – Grupo 2

	1º ES – não indígena (6.777 tokens)		Último ano ES – não indígena(10.656 tokens)		LL
	Número de ocorrências	%	Número de ocorrências	%	
HIPER	2	0,29	0	-	+ 3,78
HIPO	4	0,59	1	0,09	+ 3,54
SND	4	0,59	3	0,28	+ 0,95
SNG	5	0,73	4	0,37	+ 1,02
SNN	15	2,21	15	1,40	+ 1,52
SVNP	14	2,06	13	1,21	+ 1,86
SVG	1	0,14	1	0,09	+ 0,10
SVT	13	1,91	15	1,40	+ 0,66
SVC	1	0,14	1	0,09	+ 0,10
Total	59	8,70	53	7,13	

Não aparece nenhum dado que tenha significância estatística, mas as dificuldades com a escrita que o aluno tinha quando começou o curso permaneceram quase as mesmas. Uma delas até aumentou: SVT – erro de emprego de tempo verbal. As demais ou não foram sanadas ou praticamente se conservaram.

O que se esperava no grupo adulto é que as dificuldades que havia no primeiro ano do curso universitário desaparecessem com a finalização do curso. Isso não ocorreu com

nenhum dos subgrupos adultos, nem com os Guarani/Kaiowa bilíngues, nem com os Terena monolíngues e nem com os não indígenas. Segundo o *Log Likelihood* a significância dos problemas de escrita dos subgrupos indígenas quando comparados uns com os outros não mostrou justificativa para afirmar que as dificuldades ocorram devido à presença/ausência de língua materna.

## 7. ANÁLISE QUALITATIVA DOS DADOS

Este capítulo apresenta a análise qualitativa dos dados. Serão abordadas algumas características gerais observadas nas produções das crianças e dos adultos separadamente. Em seguida, serão discutidos os critérios de análise de estrutura morfossintática selecionados para esta pesquisa, a saber, hipersegmentação (HIPER) e hiposegmentação (HIPO), ausência de determinante (SND), ausência ou erro de concordância de gênero (SNG), ausência ou erro de concordância de número (SNN), ausência ou erro de concordância verbal de número e pessoa (SVNP), ausência ou erro de concordância verbal de gênero (SVG), erro de tempo verbal (SVT) e ausência de cópula (SVC).

### 7.1 Descrição Geral

#### 7.1.1 Do Subgrupo Indígena Infantil

#### 7.1.2 A Ortografia

Esta seção traz uma amostra da escrita das crianças indígenas que frequentam as escolas da reserva de Dourados. Para isso, foram selecionados 483 textos do Grupo 1, subgrupo infantil não indígena e subgrupo Guarani/Kaiowá bilíngue e Terena monolíngue. O Grupo 1, totalizou 483 textos, distribuídos na tabela abaixo:

Tabela 46. Total de textos produzidos pelo grupo 1, dividido em subgrupos

4ª e 5ª série Guarani/Kaiowá Bilíngue	185 textos
4ª e 5ª série Terena Monolíngue	118 textos
Total do Subgrupo Indígena Infantil	303textos
4ª e 5ª série subgrupo Não Indígena Infantil	180 textos

Antes de quantificar os dados, no entanto, é necessário fazer um preâmbulo desses textos. Chamar a atenção para algo muito comum na escrita do subgrupo indígena infantil, tanto Guarani/Kaiowá quanto Terena que é a questão da ortografia. Problemas de desvio da ortografia oficial brasileira são comuns nas séries iniciais, todavia nos textos coletados desses subgrupos indígenas essa questão se sobressai.

Bagno (2011) chama a atenção para a diferença entre *erro de português* e *erro de ortografia*, que a escola trata da mesma forma. Bagno esclarece que errar a ortografia não é errar a língua, pois ortografia é resultado de um ato político respaldado em lei. Existe uma lei ortográfica e todos devem segui-la nos momentos de usar a língua escrita. Logo, *erro de ortografia* nada mais é que um desvio dessa lei que rege a escrita oficial do país. Diz o autor:

Por isso, seria pedagogicamente mais proveitoso substituir a noção de “erro” pela de tentativa de acerto. Embora possa soar como simples eufemismo do tipo “politicamente correto”, na verdade é possível obter um efeito significativo ao trocar um termo de conteúdo negativo (erro) por outro conteúdo positivo (acerto). Afinal, a língua escrita é uma análise da língua falada, e essa análise será feita, pelo usuário da escrita no momento de grafar seu texto, em sintonia com seu perfil sociolinguístico (BAGNO: 2011, p. 353).

Seguindo essa linha de raciocínio, Bagno afirma que quem escreve *chingar* em vez de *xingar* não fez isso porque quis errar, mas porque agiu seguindo as leis da analogia, afinal essa é a ferramenta de que dispõe, haja vista ter pouca prática da leitura e da escrita, ou seja, pouco letramento. Provavelmente esse usuário da língua levou em conta a suposta equivalente biunívoca entre o som e a letra que nem sempre existe.

Veja nos dois exemplos dos subgrupos Guarani/Kaiowá e Terena:

QUADRO 11 - Exemplo de texto com problemas de ortografia do 4o ano do ensino fundamental Guarani/Kaiowá bilíngue

MIFKBPLERM4FTXT1

**Revão** um Chapeusinho Vermelho foi leva **toce** a sua vovo. e quando **icondo** o **lobisome coreu** a te a sua vovo **mai** quando **ehdro dreto** a vovo ela tinha **caído**

QUADRO 12 - Exemplo de texto com problemas de ortografia do 4o ano do ensino fundamental Terena Monolíngue

IOMTMPEETM4FTXT1

um dia estava na casa da **ninha** tia Eu estava **cricandoem baxo** da arvore e a minha **ermão** estava **cricando** Lana grama o cavalo tava a **narado** no **penando** na grama **mais** a com-tece um coisa com a **ninha ermão** o cavalo **xuto** na **ninha ermão** a minha tia que viu o cavalo **xuto** na **ninha ermão** Eu fui **corredo** fala **praninha** não fui **chana** o **neu** pai já que o **neu** pai é

chato. Ele fico bem chato. Ele corre pra Leva **ponegico** eea ta la cara um dia Eu estava com a **ninha** nãe La **noneio** da **rãona** com **ninha** prima .Eu **ninha** prima estava com **bibicleta** .Eu estava **pulãodo** na bicicleta  
 Eu pulei maxu que a **ninha** perna e a **ninha nãe** ele **vono** postinho **neu** fico brabo com **ninhanãe** ele **ondou** Eu lava cozinha Eu fui com **aninha** perna **naxucadonais** doia muito **neu** pai **nee** batia muito que **neu** batia por que Eu não queria lava cozinha **nais tivi** que lava por que **neu** batia

Desvios da ortografia oficial em textos de crianças em processo inicial de alfabetização, em que a oralidade é presença forte é algo relativamente comum, porém no universo dos textos selecionados para esta tese, os desvios encontrados nas produções da crianças indígenas aponta para uma questão que merece a atenção dos educadores. Olhar essa questão de forma mais atenta, pensando em métodos e materiais didáticos que sejam direcionados especificamente para o objetivo da escrita ortográfica, por exemplo, seria um bom começo. Os dois textos citados abaixo, por exemplo, pertencem a crianças de 12 anos de idade, a primeira iniciou seus estudos com 6 anos de idade e a segunda com 5 anos<sup>33</sup>. É comum, nas escolas da reserva, adolescentes de 14 a 17 anos estarem frequentando as séries iniciais.

Os problemas ortográficos, não obstante chamarem bastante atenção, haja vista serem problemas que ou não deveriam mais ocorrer dessa forma ou deveriam ter uma ocorrência irrisória, são frequentes na grande maioria dos textos. Nos 5<sup>os</sup> anos isso também ocorre, veja nos exemplos abaixo:

QUADRO 13 - Exemplo de texto com problemas de ortografia do 5o ano do ensino fundamental Guarani/Kaiowá bilíngue

HEEKBPCERM5TXT1

Conto do mim **pae**

Meu pai conto aqui na aldeia **egisto** o **lobizome Augueí Rapaes encotrou** Bem **grei** ele **vil**  
 Depois o **lobizome vil** ela **correl correl** foi no casa foi **atrais** Dele a mulher Dele **usfusfoi**  
 ouvi um **lobizome** meu tio **correl correl atraí** de o **lobizome** pra di **Atirol** o **lobizome morrel**.  
 FIM

Esse aluno tem 13 anos, pertence à etnia Kiowá, é bilíngue, ingressou na escola com 5 anos de idade, fala o guarani em casa, com vizinhos e parentes e o português na escola ou quando vai à cidade. É bom lembrar que na escola eles intercalam entre a língua portuguesa e a língua guarani quase que o tempo todo.

<sup>33</sup>Para as descrições serão utilizados também os dados retirados dos questionários

QUADRO 14 - Exemplo de texto com problemas de ortografia do 5o ano do ensino fundamental Guarani/Kaiowá bilíngue

ERRTMPJERM5TXT1

o me pai **cazocanha meã** e velha a **miha esão** e da **pões** eu mase façam **pani** para **erão** ou pai compro uma motilho de **anisrisoro e troso que miha meã e tara** grávida na casa **donelha** vá sofrida. fin

O aluno do texto acima pertence à etnia Terena. Não fala a língua terena, apenas o português, tem 11 anos. Como não respondeu o questionário, não temos mais informações a respeito dele. Independentemente da série ou da etnia, os problemas ortográficos não são muito diferentes. Outros dados relacionados à ortografia como a acentuação e o emprego de letra maiúscula e minúscula destoante do padrão escrito da língua portuguesa também são bastante aparentes. A ausência de adentramento de parágrafo, ausência de pontuação, letra maiúscula no meio da palavra sem nenhuma razão ocorreram em vários textos, assim como a hipo e a hipersegmentação também são questões muito pertinentes, das quais apenas estas últimas terão um olhar mais aprofundado neste trabalho.

O Texto abaixo traz vários exemplos de desvios da ortografia, de acentuação, além de outros problemas que serão destacados a seu tempo:

QUADRO 15 - Exemplo de texto com problemas de ortografia e acentuação do 4o ano do ensino fundamental Guarani/Kaiowá bilíngue

JOSKBPCERM5FTXT1

1Era uma **véz**Eu tava jogava, **Bolas**, quanto um **Cachoro**, da **vizicho** ção da vizinha  
 2**Escapou**. Eu foi mordito, por cão mais a quele cão moreu mais, ivelimedí, to muito  
 3bém  
 4mais a quele cão **Era** vacínato por merdicom  
 5mais a quele cão **Era** muito muito **pravozpidia** memo morocodiapó mais um dono  
 6pacou pramí sou purisou não moreu cão dele mais foi po ospintao do vida mais a quele  
 7cachoro moreu  
 8Eu dei feneno  
 9praele  
 10cachoro muito  
 11muito pravoz

Acentuação indevida (*véz*, *pidia*), emprego de maiúscula no meio da frase (*Bolas*, *Cachoro*), troca de letras como no exemplo, *vizicho* em que as palatais *nh* são permutadas pelas *ch*. Ainda nessa sentença há troca de sonora por surda, *mordito*, (fenômeno conhecido como de sonorização, no caso d/t). Às vezes há o contrário troca de som surdo por sonoro como em *pravoz*, em que o *b* sonoro é substituído pelo *p* surdo, além da presença da sonora *z*, ao invés da surda *s* (na tentativa de se fazer plural, o que não se justifica nesse caso em que a palavra deve estar no singular). Também é frequente no texto, o emprego de apenas

uma letra que ocorre um dígrafo como em *cachorro* e *morreu*. Presença de sinal de pontuação de forma irregular, problemas na concordância de gênero (*da vizinho* ao invés de *do vizinho*).

### 7.1.3 O Gênero Textual

Foram escolhidos para os três temas, dois temas que utilizavam o gênero conto, ou seja, uma narração, e o de número três um texto opinativo, no qual o informante tinha que se posicionar acerca da questão de como ele se sentia sendo índio. Percebe-se no texto 01, cujo tema versava sobre contos que se ouvira na infância, a apropriação do modelo dos contos de fadas tradicionais contados pelo professor em sala de aula. Isso aparece através de fórmulas como *era uma vez, felizes para sempre, viveram felizes, Fim*. Houve uma tentativa de encaixar seus textos dentro desse modelo já conhecido.

No terceiro texto, o texto opinativo ocorreu o seguinte:

QUADRO 16 - Exemplo do emprego do gênero no texto de tema 03 do 5o ano do ensino fundamental Guarani/Kaiowá bilíngüe

AMSMTPEETM4FTXT3
<p>Ele gou seidio  <b>poQue</b> vose ganha          sexta <b>poQue</b> a juda          nois <b>poQue</b> nos          samo poBri <b>poQue</b>          nois samo idio  <b>poque</b> nois temo          Casa panhero          sexta nois tem tudo a gora          eu sou veeis a qui na aldeia</p>

Muitas crianças escreveram o terceiro texto no formato de uma lista, como o texto acima, como se quisessem elencar as ideias. Isso pode ter ocorrido devido a algum modelo de escrita que talvez tenham recebido, já que houve várias ocorrências desse tipo. É patente que a estrutura textual geral, aquela que determina a forma de escrita para qualquer texto, independentemente do gênero a que se aplica não é de domínio das crianças. Conhecimentos, tais como: a necessidade do texto ser dividido em parágrafos, de ser iniciado por letra maiúscula, aliás, o emprego da letra maiúscula aparece de forma bastante confusa nos textos, não estão fixados de forma contundente na escrita dos textos. Não há domínio sobre essas questões formais do texto por parte dos alunos, por isso aparecem as formas: *nois poQue, poBri; encotrou Bem grei...* Tanto no meio da palavra quanto no meio da sentença.

O emprego da pontuação também é bastante confuso nos textos desse subgrupo. Predomina o ponto final apenas, porém, muitas vezes, inadequadamente. Há vários textos sem nenhum sinal de pontuação e queo há ponto final, o início da sentença seguinte está em minúscula. São questões estruturais, aparentemente sem maiores complicações para se adequar às normas de escrita da língua portuguesa, de certa forma, vista como partes do processo de alfabetização, o que chama a atenção, entretanto, é a generalização, a imensa maioria possui esses problemas.

Ao contrastar os textos dos alunos dos subgrupos indígenas com os do subgrupo não indígena da mesma série isso fica ainda mais patente. Há textos do subgrupo não indígena com problemas de pontuação, presença de oralidade, questões ortográficas, porém em um universo de 27 textos do tema 01 do 4º ano do ensino fundamental não indígena, apenas 02 textos estão sem nenhum sinal de pontuação. A maioria obedece às normas de escrita aceitáveis pela normatização para a escrita da língua portuguesa. Abaixo, selecionamos dois textos do subgrupo não indígena de um 4º ano do ensino fundamental, o primeiro, com os problemas já citados e o outro com poucos desvios de normatização:

QUADRO 17 - Exemplo de texto do 4o ano do ensino fundamental, do subgrupo não indígena com problemas citados anteriormente

RUCPMPCECA4FTXT1

O Peter pam

Era uma menino que via na floresta ele vivia como um macaco ele pulava em arvores ele pulava muito em todas as arvores ele tambem subia até o topo ela insima dava pra ver uma cachoera muito bonita dava praver o arcoires mas ele tinha um problema muito serio ele não sabia quem era os pais dele ele sabe que os pais de largaram ele numa gavetas de ropa e esegurila ele fala e falou oque acontese e aonde eles tava eles tavam ensima da montanha ele foi la e salvou eles e ai eles ficou felizes para sempre

Também há problemas de ordem estrutural nesse texto, por exemplo, ausência de letra maiúscula em nome próprio (pam), além de ausência de adentramento do parágrafo para iniciar o texto, troca de letras (vevia), repetição exagerada do referente “*ele*”, falta de acentuação (tambem), hipo e hipersegmentação (“ela insima” – “e lá em cima”), concordância inadequada (“eles ficou felizes para sempre). Prevalece a fórmula dos contos de fadas.

Os não indígenas também recontaram um conto de fadas que ouviram na escola ao invés de contar algo que ocorreu com eles na infância, assim como fizeram grande parte do subgrupo indígena. O texto abaixo é um exemplo que corresponde a maioria dos textos desse subgrupo. Há vários sinais de pontuação, início do texto em maiúscula, tentativa de

discurso direto e indireto, ainda não há domínio total da normatização da escrita em língua portuguesa, mas há indicações de que esse subgrupo possui um domínio maior das regras de escrita, como emprego de letras, acentuação e pontuação:

QUADRO 18 - Exemplo de texto do 4o ano do ensino fundamental, do subgrupo não indígena com poucos desvios da normatização da escrita em língua portuguesa

ENNPMPECA4FTXT1

Chapeuzinho vermelho

**Um** dia a mãe de chapeuzinho vermelho estava preparando bolinhos para a vovó de chapeuzinho porque ela estava doente. **E Mãe de chapeuzinho: diz os bolinhos estão prontos.**

**E** ela foi levar os bolinhos ela chegou na casa da vovó

**chapeuzinho: diz**

que nariz grande que orelhas grandes que olhos enormes e que boca enorme

**o lobo: diz**

**É para comer você!**

**Chapeuzinho: diz**

não me coma lobo **socorro!**

**E** o caçador escuta e foi até a casa da vovó salvar a menina e a vovó

**O caçador: diz**

**Mão ao alto lobo!**

**E** a vovó e chapeuzinho ficam saúvas.

**Fim**

Os dois pontos devem aparecer após o verbo discendi - aquele que anuncia a fala seguinte, porém, no texto acima ocorre antes do verbo. Mas já é uma tentativa de usá-lo, denotando conhecimento por parte do escrevente dessa norma. O aluno tem noção de que para colocar uma fala precisa do verbo e dos dois pontos para anunciá-la. A ausência de maiúscula no início da sentença também ocorre, como nos textos dos indígenas, mas em número restrito e intercalado com a maiúscula. Não há ocorrência de maiúscula no meio da palavra ou no meio da sentença.

Essa introdução da coleta dos textos infantis mostra questões mais abrangentes relacionadas aos padrões de escrita da língua portuguesa, não interessando, neste momento as possíveis causas, mas são questões que não poderiam ser ignoradas, uma vez que sua presença nos textos é bastante significativa.

#### 7.1.4 O Emprego do Verbo na Língua Portuguesa

As questões de emprego peculiar das formas verbais, aparece muito nos textos do subgrupo indígena Guarani/Kaiowá bilíngue. Tanto no grupo 1 quanto no grupo 2. São

verbos flexionados de forma a confundir o modo e às vezes o tempo na escrita desses subgrupos.

Segundo Bagno (2011), o verbo deve ser visto de várias perspectivas: sintática, semântica e pragmática, pois ele é um todo que carrega essas características, além de ser o núcleo do qual vão surgir os complementos. Nos textos infantis dos subgrupos indígenas, a presença do verbo ocorre de forma, muitas vezes, inusitada. Veja no exemplo:

QUADRO 19 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue

AMSKBPLERM4FTXT2

eu tenho 2 irmão. tenho um irmão que ce chama lisciano i Outro que ci chama flaviana que tem 3 zanio uotro tem 5 anos **eu gosto** de gogar Bola comi meu irmão mais gradiu. **gosto** di gogar Bola no dia de ducação fizeca e **gosto** di pescar **gosto** di bricar com Bobinha **gosto** di pacia sozinho **gosta** di pescar com meu pai **gosta** di ficar sosinho na casa meu pai **gosta** di ficar dia inteiro em sábado e domingo **gosto** di tomar baio no rrio.

Neste exemplo percebe-se o emprego repetitivo do verbo gostar na 1ª pessoa do singular (gosto), ou seja, o narrador do texto. Mas no encaminhamento para o final da estória, na penúltima linha do texto, essa forma verbal é trocada pelo índice de não pessoa (gosta) sem nenhum motivo aparente. Trata-se de um adolescente de 13 anos que ingressou na escola com 9 anos e tem a língua guarani como língua materna, porém utiliza somente o português em situações nas quais é necessário utilizá-la, quando vai à cidade por exemplo. Na aldeia, ele utiliza o guarani, inclusive na escola utiliza mais o guarani do que o português.

Como a maioria dos professores ministra aulas em português (com algumas exceções), é nesse momento que ele a utiliza e também para falar com uma ou outra pessoa, já que nessa escola, a maioria fala a língua indígena<sup>34</sup>. No questionário, esse aluno elencou de 01 a 05 na ordem de importância as línguas que seriam mais importantes de aprender: 1ª guarani; 2ª kaiowá; 3ª inglês; 4ª português; 5ª espanhol; 6ª terena.

O texto abaixo traz duas questões interessantes com relação ao emprego do verbo *ser* nos textos desse subgrupo: a ausência de acento gráfico na 3ª pessoa do singular, do modo indicativo do presente do verbo ser: a forma *é*. A ausência do acento nessa forma verbal confunde o verbo com a conjunção *e*. Outra ocorrência bastante comum nesse contexto é a

<sup>34</sup>Dados retirados do questionário.

ausência de cópula, como se vê na 1ª linha: “a minha família é legal” (onde a forma verbal é está substituída pelo zero) Observe:

QUADRO 20 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Terena monolíngue

ANOTMPEETM4FTXT2

**Aníha Família** ☐

**legau** níha Fanília nucaBrígou **niha Família**

**E muito** Felis niha Fanília gosta de cos Partíha as coisas **níha Faníla e nuíto sertto**

Esse aluno tem 14 anos, também entrou na escola com 9 anos de idade, como o aluno do exemplo anterior, sempre estudou nas escolas da reserva indígena. Não fala nenhuma outra língua, apenas o português. É um aluno terena de pais da etnia terena. No questionário respondeu que a língua que ele acredita ser a mais importante para aprender a falar é a língua kaiowá (guarani) em 1º lugar, em 2º lugar, a língua terena, em 3º, a língua guarani, em 4º, a portuguesa, em 5º inglês, e por último, o espanhol.

O uso do verbo por esses dois grupos indígenas mereceu destaque, pois são muitas as ocorrências que o envolve. Observe este outro exemplo:

QUADRO 21 - Exemplo de emprego de verbos num texto do 4o ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue

JOSKBPCERM5FTXT1

1**Era uma véz**Eu tava jogava, Bolas, quanto um Cachoro, da vizicho ção da vizinha  
 2Escapou. **Eu foi mordido**, por cão mais a quele cão moreu mais, ivelimedí, to muito  
 3bém  
 4mais a quele cão **Era** vacínato por merdicom  
 5mais a quele cão **Era** muito muito pravo **pidía memo morocodi**apó mais um dono  
 6pacou pramá sou **purisou não moreu cão dele mais foi po ospintao** do vida mais a quele  
 7cachoro moreu  
 8Eu dei feneno  
 9praele  
 10cachoro muito  
 11muito pravo

A presença da locução verbal na primeira linha *estava jogando* fica *tavajogava* Na segunda linha, a expressão *eu fui mordido* aparece como *eu foi mordido*. A primeira pessoa do discurso, aquela que fala tem o verbo conjugado como terceira pessoa. Em seguida, na quinta linha *pedia mesmo para morrer* está *pidía memo* moro. Na linha 6, ocorre uma passagem ambígua em que inicialmente se pensa que quem foi para o Hospital da Vida foi o cachorro, no entanto, a conjugação do verbo *ser* na 3ª pessoa (*foi*) ao invés da 1ª (*fui*) gera

essa ambiguidade. Outra confusão que ele faz com as palavras é a troca do advérbio felizmente por infelizmente (*ivelimedi*) linha 2- causando estranhamento de sentido.

Na verdade, o menino mordido pelo cachorro é que foi para o hospital e teve suas despesas pagas pelo dono do cachorro e, por isso, o cachorro não teria morrido o que é desmentido em seguida quando ele conta que envenenou o cachorro.

Outras questões destoantes da escrita oficial da língua portuguesa também surgem no texto, sobretudo ortográficas, a forma verbal *era* do pretérito imperfeito, usado na fórmula consagrada pela literatura infantil para iniciar os contos de fadas na expressão *era uma vez* está sempre em maiúscula, não importa em que posição da sílaba ela ocorra (linhas 1, 1, 2, 4 e 5), como se essa forma verbal só pudesse aparecer dessa forma.

Esse aluno de 12 anos, filho de pais kaiowá, é falante da língua em casa, na vizinhança ou com parentes, é falante do português apenas nos momentos em que essa língua lhe é exigida, um pouco na escola, quando ouve rádio ou assiste à programação da TV, ou nos momentos em que vai à cidade. No questionário, respondeu com relação à língua que ele considera mais importante de ser aprendida, elenca em 1º, o guarani, em 2º, o português, em 3º, o kaiowá, em 4º, o inglês, em 5º, o terena e, por último, o espanhol. Já com relação às culturas com as quais mais se identifica, ele coloca em 1º lugar a cultura guarani, em 2º, a cultura terena, em 3º, os mestiços em 4º, a kaiowá e, em último, os não indígena. Observe que a identificação com os não indígena ocupa a última opção. No entanto, quando se trata de aprender a língua dessa cultura, ela salta para o 2º lugar. Isso mostra a necessidade que a aprendizagem da língua portuguesa ocupa nessa comunidade que precisa dela para se relacionar com a comunidade não indígena.

### 7.1.5 A Concordância Nominal de Gênero

#### 7.1.5.1 A Concordância de Gênero na Língua Portuguesa

Segundo Bagno (2011), há uma solidariedade semântica entre o nome e o verbo quando se refere à concordância. Como nome, pode-se tranquilamente mencionar os substantivos, os adjetivos, os pronomes e os determinantes, que desempenham papel de adjetivos. Conforme Bagno (2011), o fenômeno da concordância é bastante variável nas línguas, havendo momentos em que é abundante e outros em que é escasso.

Tanto a concordância verbal quanto a nominal é redundante ou tautológica. Por isso, que as regras de concordância sofrem alterações no decorrer do tempo, sendo marcadas, muitas vezes, apenas no determinante ou apenas no substantivo, como no inglês, por exemplo, já que os determinantes dessa língua são invariáveis (à exceção de *this/these; that/those*). Outras línguas como o francês, o africâner, o cabo-verdiano, o haitiano, e o português nas variedades não padrão se somam a esse paradigma. (BAGNO, 2011).

Com o gênero do adjetivo ocorre o mesmo processo. Como são palavras que expressam qualidades/propriedades neutras, o adjetivo está comumente no masculino, que é a forma não-marcada da língua portuguesa. A maior ocorrência se dá quando o adjetivo está mais afastado do núcleo do sintagma nominal ao qual pertence e deveria concordar.

Nos textos dos subgrupos indígenas do grupo 01, o fenômeno da concordância nominal de gênero mereceu destaque por suas peculiaridades. A opção em usar o fonema /a/ ou o fonema /o/ no final das palavras, causa estranhamento, uma vez que o fonema /o/ em final de palavras é a forma não marcada para o masculino para a língua portuguesa, e o fonema /a/ é marcação do feminino. Percebe-se nos textos a ausência de noção de como utilizar a marcação de gênero na língua portuguesa.

Veja o texto abaixo, do Grupo 01, do subgrupo Guarani/Kaiowá 4º ano:

QUADRO 22 - Exemplo de emprego de concordância nominal de gênero num texto do 4º ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue

VFBGBPLERM4FTXT3
1EU UMR togos fodnã
2 Eu quis tudo <b>pro mim</b> mais
3sido mostro parr <b>minquinado</b>
4daria certo para mim sonho
5tudo parese são melhor que no
6real mais portulai Eu intavoce
7tudo que as pessoa estafaleo mudoor
8pra melhor sei que
9Eu ti quero, Eu <b>tiquera</b>
10 <b>paro</b>
11ficar ComiCo

Trata-se de uma aluna de 15 anos de etnia kaiowá, filha de pai e mãe kaiowá que fala quase que somente o guarani, reservando a língua portuguesa apenas para momentos em que não há nenhuma opção de falar sua língua materna. Essa aluna troca o grafema **a** pelo **o** em **pro mim** (linha 2). Há confusão no uso dessas duas vogais, pois na linha 3, ela escreve: **parr(para) minquinado(mim que nada) daria certo**. Na linha 9 aparece: eu ti **quera** ao

invés de **quero** e na linha 11 **paro ficar comiCo**, ao invés de **para ficar comigo**. Percebe-se que não há intimidade nenhuma com a ortografia da língua portuguesa, sobretudo com questões relacionadas à utilização da morfologia que designa o gênero nessa língua.

Segundo o questionário, a aluna entrou na escola da reserva com 9 anos e sempre estudou em escolas indígenas. Enumerou, em ordem de importância, as línguas que ela acredita que deveria aprender a falar: 1º kaiowá, 2º guarani, 3º terena, 4º português, seguidos de inglês e espanhol. Com relação à cultura com a qual mais se identifica em ordem de importância, enumerou: 1º guarani, 2º kaiowá, 3º terena, 4º não indígena e, por último, mestiços (indígenas que casaram com não-indígenas). Com relação ao tema *O que é ser índio para você* não foi realmente sequer tocado. Como se trata de uma adolescente, ela fez um texto intimista, faleo de seus sentimentos.

O exemplo seguinte mostra o mesmo problema da aluna anterior quanto ao emprego dos fonemas /a/ e /o/ na marcação de gênero. Essa aluna é do 5º ano, também da etnia Guarani/Kaiowá bilíngue:

QUADRO 23 - Exemplo de emprego de concordância nominal de gênero num texto do 5º ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue

SPSGBPCERM5FTXT3  
 1Ser india para mim é tudo na **mim vida** quero ser **tudo Otina** também para mim familia quer 2ser Braca **quer muito coisa** que Eu não gosta da quela coisa que não é da india ma mi familia é quær é 3tudo Braco não muito **mim mãe** é do Paraguai mais naci qui na **ondía Borara** Eu não gosto do meu 4irmão porque é Braco Eu não quero ser Braca mim vida mais **um coisa** Eu quero míniten uma coisa ser 5Braco meói da fanilia

A concordância de gênero neste texto é apenas um dos vários questionamentos que se pode levantar com relação ao exigido pela norma escrita da língua portuguesa. Essa aluna tem 12 anos e entrou na escola da reserva aos 5, filha de mãe paraguaia e pai guarani é falante do guarani. No questionário, pontuou, como os demais, identificação com a cultura indígena (guarnai/kaiowá/terena) e por último com os não indígenas e mestiços e línguas mais importantes: guarani/ kaiowá/terena/português e as demais. Observa-se que há influencia da língua da mãe<sup>35</sup> (paraguaia) “mi familia”.

Na linha 1, já se percebe a ausência do conhecimento de como funciona a marcação de gênero em português queo a aluna escreve **mim vida** e não **minha vida**, em seguida, ela

<sup>35</sup>No Paraguai há duas línguas oficiais: o espanhol e o guarani. Muitas vezes, há interferência de uma na outra, então surgem construções do espanhol no meio de uma sentença em guarani. Isso ocorre também em conversas inteiras em que as duas línguas realmente se misturam a esse fenômeno eles dão o nome de *jopara*. Vez ou outra se depara com um ou outro indígena guarani que vem do Paraguai para a reserva de Dourados, uma vez que Dourados situa-se a 120 km da divisa com o Paraguai.

escreve: quero ser **tudoOtina**, por desconhecer a sintaxe da língua portuguesa, provavelmente ela quis dizer **quero ser ótima em tudo**. Na linha 3, **quer muito coisa**, ao invés de **quero muita coisa**. Na segunda ocorrência de ausência de concordância, ou confusão na concordância de gênero, na linha 3, ela escreve **ondía Borara** por **Aldeia Bororó**. Mais uma vez, o equívoco ao usar as vogais **a** e **o**, que na língua portuguesa são as marcas de gerais de gênero feminino e masculino. O mesmo ocorre na penúltima linha do texto com **um coisa**, por **uma coisa**.

Nos textos do subgrupo terena do grupo 1, também ocorre o fenômeno de não concordância de gênero. Veja no exemplo de um aluno do 4º ano:

QUADRO 24 - Exemplo de emprego de concordância nominal de gênero num texto do 4º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

OEMTMPCERM4FTXT3

ser india e muito bom que casa **na mato** capivara panata e pescar paxe mo açudi fui iporop**aroc**asa fida o paxe para faníla come paxe.

Aluno de 12 anos, filho de pais terena, não fala a língua terena, os pais também não falam a língua materna da tribo. No entanto, o menino ouve guarani na escola e nos ambientes que frequenta dentro da reserva. Na primeira linha do texto, ele escreveu: **casa na mato**, ao invés de caça **no** mato. Entrou na escola aos 4 anos e não iniciou seus estudos na escola indígena. No questionário, respondeu que entende o guarani queo ouve. A primeira língua que ele pensa ser a mais importante para aprender é a portuguesa, seguida pela terena/guarani/kaiowá/inglês/espanhol e as culturas que mais se identifica: terena/mestiços/não-indígenas/guarani/kaiowá.

O exemplo abaixo é do mesmo subgrupo, mas do 5º ano:

QUADRO 25 - Exemplo de emprego de concordância nominal de gênero num texto do 5º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

MFATMPJERM5FTXT3

1Ser indio para mim e muito bom por que a gente vai na casa dos primos e primas e para mim e 2bom ser indio a gente faz muita coisa na rezerva os índios guaranis dançam, tem **varios danças** 3que eles faz para eles dançarem, e para mim e legal as danças deles, eles pintam, o rosto, os 4braços, e as pernas, para eles e bom ser índios, tem **varias coisa bom** na rezerva, e os indios 5são muito legal.

Esse aluno de 12 anos, filho de pais terena fala apenas o português, mas ouve a língua guarani na escola. Os pais também não falam o idioma terena. Com relação à ausência de concordância de gênero, vemos na linha 2: **vários danças**, em que o

quantificador adverbial ou pronome indefinido – **vários** - deveria estar no feminino para concordar com o substantivo **danças**. Na linha 4, a expressão **varias coisa bom** também traz problemas de concordância, dessa vez a concordância deveria ocorrer com o adjetivo **boas**, mas em seu lugar aparece o masculino **bom**.

Interessante observar que há concordâncias de verbos, por exemplo, que correspondem às normas da escrita de língua portuguesa: na linha 2 **os índios guaranis dançam**, mas também ocorre a ausência dela, em seguida, na linha 3 **que eles faz para eles dançarem**. Parece que há um princípio de entendimento que a língua portuguesa exige concordância verbo-nominal e de gênero em todos os elementos da sentença, mas isso ainda não está bem internalizado. Outra questão, como ocorre em muitos outros textos, e nesse também, é a ausência de acento gráfico para distinguir a conjunção aditiva **e** da 3ª pessoa do presente do indicativo **é**. Geralmente, esse grafema aparece grafado como conjunção e está empregado como verbo.

#### 7.1.6 A Concordância Verbal

##### 7.1.6.1 Concordância Verbal na Língua Portuguesa

Assim como a concordância nominal, a concordância verbal é fator de muitas controvérsias no ensino da língua portuguesa. Controvérsias, no sentido de que nas relações comunicacionais informais ela quase não ocorre, ou melhor, é marcada, preferencialmente, no determinante. Nos textos das crianças indígenas, não diferentemente dos textos das crianças não indígenas, o processo de concordância segue também o mesmo padrão difuso. Veja o primeiro exemplo retirado do 4º ano do fundamental do subgrupo Guarani/Kaiowá:

QUADRO 26 - Exemplo de emprego de concordância verbal num texto do 4º ano do ensino fundamental, do subgrupo indígena Guarani/Kaiowá bilíngue

BIBKBPLERM4FTXT2

Eu focota Aoti *Eu foi* *Eu foi* masi dade para vove o nau ido *Eu ajuda* a inai ai  
 O mo Pai taba nasaara ono foi resebiba nasa ara ona i não taicosa lifapara.  
 O JiEufití para tas abara é bali Eu falei í ta onau mauto deji da icanla icola ebai para a  
 pada Eu cero para defiamo Eu cero pasa desino anor Eu tai pino Eu cota dago ficato Eu gasa da  
 nobi Ono fiabo.

*Eu core* ti toda Eu voni saso para pasa daficlo mo mão cera fica para tai timo Fo pori  
 Eu tanifo fosada Para nipasa de mofó Eu acho foca bola tabo Eu taEuaboi Eusabou ticolero  
 monai cafo Eu tomoucapa Eucota todea gida ainaEuasada aniana é o mopai solce ajuda anaiana  
 i Eu cota *Eucota tipasa* casa do casa da tia.

O uso da 1ª pessoa com o verbo empregado na 3ª é característico desse texto: **eufoi; eu ajuda; eu corre; eu gosta**. Evidentemente, a questão da oralidade é muito patente no texto, juntamente com a questão da ortografia ser totalmente resultado da fala. Claramente, o aluno não tem domínio da escrita ortográfica da língua portuguesa.

Essa aluna tem 12 anos, entrou na escola aos 6 anos e sempre estudou na escola da reserva. Falante do guarani/kaiowá, só utiliza a língua portuguesa em casos em que essa língua é exigida. Língua que colocou em 3º lugar como a que deveria aprender a falar, depois do guarani e do kaiowá. Com relação à cultura que mais se identifica elencou: a guarani, a kaiowá e a terena como as primeiras, mestiços e não indígenas por último.

Abaixo, o primeiro texto é do 4º ano do fundamental do subgrupo Terena:

QUADRO 27 - Exemplo de emprego de concordância verbal num texto do 4º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

<p>ALSTMPEETM4FTXT2  minha família gosta de passear ir para sídade mas um dia esse dia fa sexta feira o meu primo foi na igreja com nos <b>nos não acheno</b> ele mas meu tíu ele foi conPra pinga acho o meu príno morto e o nome do meu prímo é ríudo e minha mae ficou sabendo e começou a cuidar da minha vo porque era o neto dela mas <b>mataro</b> ele com desesseis facada</p>
--

Não obstante, o texto ter seu início com letra minúscula e problemas de acentuação, letra maiúscula no meio da palavra e ausência de pontuação. A concordância não se estabelece, conforme os padrões necessários para a escrita. Na linha 2, aparece **nós não acheno**, o verbo no subjuntivo, tomando lugar do verbo do indicativo, isso também é fator de não concordância, além da ausência de marca tradicional de plural.

O texto seguinte, também do subgrupo terena monolíngue, do 4º ano traz o seguinte com relação à ausência de concordância:

QUADRO 28 - Exemplo de emprego de concordância verbal num texto do 4º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

<p>AMSTMPEETM4FTXT2  Nome daminha mãe e cirlene e neu Pai Ordelio ieu írmar írma  Eu Brico com meu irmar e Irma eu jogo Bola  Com me irmar  E meu pai taPe e minha mãe tanPen meu Pai trapalha de pedreiro na sidade  E La cuta da casa  Noi vivePen</p>
--

A concordância não é apenas verbal. A conexão entre as palavras e as sentenças também alteram essa concordância, uma vez que a ausência de artigo, a ausência de acento na forma **é** do verbo **ser** na 3ª pessoa do singular, na linha 1, da preposição, a ausência de

cópula também comprometem a escrita do texto, conforme o exigido pelos padrões da língua portuguesa. Trata-se de um aluno de 9 anos, da etnia terena e falante do guarani e do português. Entrou na escola da reserva aos 5 anos e no questionário, optou pela língua guarani como a que deve ser aprendida em 1º lugar pelos indígenas da reserva, seguida pelo kaiowá, terena, português, inglês e espanhol.

### 7.1.7 A Hipossegmentação e a Hipersegmentação

Na seção 4.3.4 constam exemplos de hipossegmentação e hipersegmentação dos textos dos alunos do subgrupo indígena Guarani/Kaiowá bilíngues do 4º e 5º anos, então vamos trazer exemplos dos alunos de 4º e 5º anos do subgrupo terena monolíngues e não indígenas. O exemplo abaixo é de um aluno Terena monolíngue do 4º ano do ensino fundamental, nesse caso a presença da hipossegmentação:

QUADRO 29 - Exemplo de hipossegmentação num texto do 4º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

RORTMPEETM4FTXT2  
meu pai e quarani minha mãe terno eu **corlodomuito** da minha familinha e **dosmeusvos**.

Abaixo, exemplo de texto com presença de hipossegmentação do subgrupo Terena monolíngue do grupo 1, do 5º ano:

QUADRO 30 - Exemplo de hipossegmentação num texto do 5º ano do ensino fundamental, do subgrupo indígena Terena monolíngue

ANRTMPRETM5FTXT2  
Na minha nois somos oito o meu pai que se chama Erade ele trabalha na vassoura ele corta, repica, linpa, rapa, e amarra é isso que meu pai faz, e a minha mãe chama Luciene ela ajuda meu pai nessas coisa que eu falei e eim o meus irmãos a ereina que e mais grande ela fica em casa até 9:30h e depois eu vou pra casa e ela vem e o meu irmão alder estuda **atardecom** o meu tio Junior e cinco hora da tarde e depois os nossos pais chega para jantar.

Abaixo, exemplo de texto de aluno não indígena do grupo 1:

QUADRO 31 - Exemplo de hipo/hipersegmentação num texto do 4º ano do ensino fundamental, do subgrupo não indígena do grupo 1

BEBPMPCECA4FTXT2  
Eu e minha família  
Eu tenho uma família eu agradeço a deus pela saudí deles eles e muito saudaveu e fortes de saudé a minha mãe e muito legal bonsinha e meu patastro também e legal e a minha tinha ela e muito legal **de mas ...**  
Eu gosto muito da minha família todos são legau **midarem** as coisas e não e só pro priso que eu gosto deles é também!  
Eu amo **de mas** a mínha família.  
FIM...

O exemplo a seguir é do mesmo subgrupo, mas do 5º ano:

QUADRO 32 - Exemplo de hipo/hipersegmentação num texto do 5o ano do ensino fundamental, do subgrupo não indígena do grupo 1

<p>AMCPMPEECA5FTXT2          Minha família          Bom eu vou começar falando do meu pai, ele mora em outra cidade mas ele <b>mebuscat</b> todas férias eu gosto muito dele e ele de mim.          Agora de minha mãe, ela sempre me da presentes eu também gosto muito muito dela.          Vou falar de meus avós. Eles são muito importantes para mim.</p>
--

Pode-se observar que os problemas de hipossegmentação e hipersegmentação ocorridos nos textos dos 4<sup>os</sup> e 5<sup>os</sup> anos do grupo Terena monolíngue e do grupo não indígena são bastante semelhantes. Ao contrário do grupo Guarani/Kaiowá no qual esse fenômeno é mais contundente. Nos textos do grupo não indígena, tanto a hipossegmentação quanto a hipersegmentação são comuns de ocorrerem, pois se trata de um processo natural na aquisição da forma escrita da língua. No grupo indígena Guarani/Kaiowá bilíngue, no entanto, esse processo ocorre de forma mais incisiva do que nos outros grupos.

Na seção seguinte, vamos observar os textos do grupo2 – o grupo dos acadêmicos. Começamos com os textos dos primeiros anos do curso de graduação de cada subgrupo. Primeiro, com relação ao emprego de verbos, em seguida, as ocorrências de concordância de gênero e concordância verbal e, por fim, a hipo e hipersegmentação. A metodologia adotada para esse grupo será diferente. Não haverá divisões, segue-se a cada exemplo um apanhado de todos os critérios que foram sugeridos para a descrição.

## 7.1.8 Do Ensino Superior – Primeiro e Último Ano da Graduação

### 7.1.8.1 Emprego dos Verbos

Os textos dos subgrupos indígenas do grupo 2, chamam a atenção pela forma como empregam as formas verbais, por isso trago uma breve introdução desse assunto antes de quantificar esses dados.

Desse grupo todo - grupo 2 - foram coletados os seguintes números de textos:

Tabela 47. Total de textos produzidos pelo grupo 1, dividido em subgrupos

Primeiro e último ano Guarani/Kaiowá Bilíngue	106 textos
Primeiro e último ano Terena Monolíngue	113 textos
Total do Subgrupo Indígena Adulto	219 textos
Primeiro e último ano subgrupo Não Indígena Adulto	130 textos

Diferentemente do subgrupo indígena infantil, os acadêmicos indígenas do subgrupo Guarani/Kaiowá bilíngue apresentaram uma forma peculiar de empregar os verbos. Veja nos textos do primeiro ano do subgrupo Guarani/Kaiowá bilíngue:

QUADRO 33 - Fragmento do texto de acadêmico do primeiro ano do ensino superior do subgrupo Guarani/Kaiowá bilíngue, mostrando o emprego de verbos

EDBGBISUPTXT01

Meu nome XX **quer me expressar** a minha historia real da minha infância como passei a dificuldade e porque é qual experiência que eu adotei durante nessa vida.

Começando com o meu avo e a minha avó, eles são família que foram perseguidos por político do Portugal e Espanhol e por isso eles vieram para cá do Paraguai que lugar que se chama cháco. E nesta região eles trabalharam nos ervais ate onde Rondon os pegaram para demarcar a terra para os indígenas, por isso meu avô foi um homem escolhido para comandar na aldeia no município de Caarapó onde Rondon chamou como capitão da reserva.

No subgrupo indígena Terena monolíngue essas ocorrências não se apresentam. Os textos desse subgrupo são semelhantes aos do subgrupo não indígena.

O texto abaixo é de um acadêmico Kaiowá que está inserido no grupo bilíngue Guarani/Kaiowá. Ele veio de uma cidade próxima (Amambaí) para cursar o curso de Letras na universidade estadual em Dourados e divide aluguel com outro estudante. Observe que logo no início do texto há problemas com relação ao emprego de verbos em português **quando estive completando... já sonhei em estar** numa clara tentativa de escrever dentro dos padrões de escrita da língua portuguesa.

O que se vê é o emprego inadequado dos verbos gerando uma concordância nada adequada na língua portuguesa. Há uma confusão entre tempos e modos verbais. Na sequência, a concordância também não corresponde ao que se espera na língua portuguesa em **adoro verem os outros ficarem alegre,...**, em que a tentativa é de concordar o verbo com o objeto e não com o sujeito da oração, ou seja, o acadêmico sabe que tem que haver concordância só não sabe como fazê-la. Na sentença: **Quando a professora ou a diretora da escola disserem que não haverá aula, (gostava) ficava muito feliz, por que já sei que**

**vou me divertir**. Nessa construção, a confusão com os tempos e modos verbais fica mais intensa, pois há equívoco com relação ao modo subjuntivo e indicativo e pretérito imperfeito e futuro do pretérito. Esses modos e tempos não foram utilizados para fornecer a ideia que se queria passar, em consequência disso, a parte final também fica comprometida, pois há uma confusão e mistura de tempos nos quais os fatos ocorreram e suas prováveis consequências futuras. Na linha final, o aluno confunde o verbo **estar** com **ter** quando escreve: **só 4 ou 5 falta estive** .

Veja na transcrição do texto:

QUADRO 34 - Exemplo de emprego de verbos e ausência de concordâncias num texto do 1o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2

AGNKB1STXT1

Sou eu, XXX<sup>36</sup>, tenho 17 anos, **quando estive completando 10 anos, já sonhei** em estar cursando Letras. Entrei na escola forçado. Mas a minha mãe sempre falou que um dia eu ia gostar de estudar.

Sempre gostei de jogar bola, ir para Igreja, gosto de ajudar o meu pai, **adoro verem os outros ficarem alegre**, dançar e jogar basquete. *Quando a professora ou a diretora da escola disserem que não haverá aula, (gostava) ficava muito feliz, por que já sei que vou me divertir*, não (adorava ter) **gostava ter uma inimiga**, se arranjar, não consigo viver sem pedir desculpa ou perdão.

A escola que (estudava) estudei ficava pertinho da minha casa, e em um ano só 4 ou 5 falta **estive**. Agora estou e tenho certeza que vou continuar assim.

No texto abaixo, de um acadêmico de 1º ano do ensino superior, há alguns desvios da norma padrão de escrita da língua portuguesa em que ocorre a troca de concordância de feminino para o masculino. O adjetivo **bom** deveria concordar com **infância**, ou seja, **a infância foi muito boa**, ao contrário do que ocorre duas vezes “**a infância foi muito bome** no final do texto **a minha infância foi muito bom**. Em seguida há uma hipersegmentação de **estudávamos**. Na sequência, há uma expressão cristalizada já pelo uso na língua portuguesa **era tão gostoso**, referindo ao tempo que passou, no entanto está grafado **era tão gostosa**, expressão que poderia concordar com a festa de formatura ou com a vida naquela época. O que chama a atenção é que essa expressão não é utilizada dessa forma na língua portuguesa.

Outra expressão desse tipo está em **na aldeia é diferente de cidade**, em português provavelmente seria a opção **na aldeia é diferente da cidade**, para estabelecer a concordância da preposição no feminino com a cidade também no feminino. Na terceira linha, **onde comemorava**, deveria estar escrito: **onde comemorávamos**, ou a opção **onde se**

<sup>36</sup>Substituição do nome do acadêmico.

**comemorava**, outro fator de discordância verbal. Nas últimas linhas **brincar, correr, divertir é um direito da Criança e ela tem vive-la**. A expressão verbal **vivê-la** se refere à infância ou ao direito da criança de brincar, correr, divertir-se? O texto mostra um número significativo de ausência de concordância nominal de gênero e de número, bem como caso de hipersegmentação.

Veja a reprodução no texto na íntegra:

QUADRO 35 - Exemplo de ausência de concordâncias e hipersegmentação num texto do 1o ano do ensino superior, do subgrupo indígena Terena monolíngue do grupo 2

AAATMISTXT1

Bom **a infância foi muito bom** tive a oportunidade de brincar bastante até aos 4 anos de idade, após isso aos 5 anos fui para a pré-escola e lá tudo era tão bom. **Estudava-mos** em um salão aberto **onde comemorava** a festa de São Sebastião padroeiro da nossa aldeia. E fiquei ali até aos 6 anos e então só depois fui para a 1ª série. Quando terminei o pré tivemos uma festa de formatura e usamos até **beca era tão gostosa** por que na Aldeia **é diferente de cidadena** Aldeia você tem a liberdade de brincar, correr e sair na casa dos vizinho sem se preocupar. somos bem dizer uma família dentro da Aldeia onde cada um cuida o outro para não acontecer nada de ruim.

Mas, então a **minha infância foi muito bom** não tenho que reclamar só a agradecer por ter a oportunidade de te-la como um papel inportantissimo na vida de uma criança seja ela quem for e brincar, correr, divertir **é um direito da Criança e ela tem vive-la**.

Esse texto foi escrito por uma acadêmica de 20 anos, estudante de Pedagogia, filha de pais Terena, que vive numa reserva Terena. Sua convivência é quase que totalmente com seus compatriotas, pois nessa reserva, que fica numa cidade próxima à região de Dourados e é composta por 9 aldeias e fica na zona rural. Essa acadêmica vai de ônibus da reserva até uma universidade estadual que estuda a qual se localiza na cidade de Maracaju. Não é falante da língua terena, uma vez que só os mais velhos e alguns jovens adultos usam a língua em sua aldeia.

Os exemplo seguintes, obedecem aos critérios estabelecidos para a descrição dos dados.

Comecemos com o texto abaixo do acadêmico do subgrupo Guarani/Kaiowá.

QUADRO 36 - Exemplo de ausência de concordância de gênero e número num texto do 5o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2

AMRBGULTSUPTXT01

Minha infância foi marcada por **varios momentos, uns bom outros ruins**. **Mas a qual** vou relatar a seguir na verdade foi uma tragedia que aconteceu na familia da minha mãe, queo eu tinha 12 Anos Aproximadamente.

Meu avô era um homem muito especial, e muito querido por todos os que conhecia, principalmente pelo fato de ele ter uma grande sabedoria em remédios tradicionais, além de plantar e tirar desses sua renda, por esse fato de viver da roça, ele tinha todo tipo de veneno em

casa que eram guardado num lugar onde só ele encontrá-ria. Já que tinha um grande medo de alguém ou de seus próprios **filhos pegarem e cometer alguma** “loucura”.

Essa acadêmica do 5<sup>o</sup> ano do curso de História de uma universidade pública. Tem 24 anos, é falante do idioma guarani em casa com pais e parentes, no questionário se identifica em primeiro lugar com a etnia kaiowá e em último com os não indígenas. No entanto, que se trata das línguas mais importantes para se aprender na reserva, por ordem ela assinalou o kaiowá(guarani) e em seguida o português. Na linha 1 do texto já surgem dois problemas de desvios na norma da escrita da língua portuguesa. A primeira, natural nos textos dos não indígenas também, é a marcação de plural apenas no determinante: **uns bom outros ruins**. Em seguida, há ausência de marcação de concordância de gênero: **mas a qual vou relatar**, se referindo aos momentos, mas que provavelmente ela estava se referindo à infância. O que não resolveria o problema da concordância, uma vez que não se relata a infância e sim o que aconteceu nela. Na última linha do texto, o desvio da concordância verbal oficial da língua portuguesa mais uma vez aparece: **filhos pegarem e cometer alguma**. Nesse caso, a ideia poderia ser de que algum de seus filhos pegasse o veneno e cometesse alguma coisa.

No texto abaixo,

QUADRO 37 - Exemplo de ausência de concordância de número num texto do 4o ano do ensino superior, do subgrupo indígena Guarani/Kaiowá bilíngue do grupo 2

EDMMTUSTXT01

Eu me lembro que nós iam para a casa de meus avós e lá naquela época não tinha energia era só lamparina, **meus avós ficava** na frente de sua casa **sentados** tomando mate **anoite** quando a lua era cheia porque clariava, lá fora.

**Eles contava varias historia** e eu gostava de escutar, sinto muita falta de minha avó que já foi morar com Deus.

Mas **quando eles começava** a falar na língua materna era assunto de adulto as **crianças não podia** escutar **porque elas não entendia** a língua terena.

Esta aluna está no 4<sup>o</sup> ano de Educação Física de um centro universitário. Filha de pais Terena, é falante do idioma terena em casa com os pais e parentes. No questionário, apontou que a cultura com a qual mais se identifica é a cultura não indígena, em segundo lugar, com a cultura dos mestiços e, em terceiro, com sua cultura. Com relação aos idiomas na ordem de importância assinalou: primeiro, a língua portuguesa, em segundo a língua terena, em terceiro, o idioma guarani, os últimos, kaiowá, espanhol e inglês. Veja que ela intercala entre a presença e a ausência da concordância. Na segunda linha do texto, isso se destaca, **meus avós ficava na frente de sua casa sentados**. No final: **as criança não podia**

**escutar porque elas não entendia**, mostra a concordância feita apenas nos determinantes, como costuma ocorrer nos textos dos não indígenas.

No grupo 2, a hipo e a hipersegmentação ocorrem em número bem menor. Ainda assim, ocorrem nos primeiros anos da graduação. O maior desvio com relação à normatização é de concordância de número. A ausência de concordância de gênero em consonância com a língua portuguesa também é pouca.

## CONSIDERAÇÕES FINAIS

Este trabalho teve por objetivo investigar as dificuldades na produção escrita de crianças e adultos indígenas da etnia Guarani/Kaiowá e Terena em comparação com crianças e adultos não indígenas. O motivo para esta investigação surgiu em dois momentos. O primeiro nas salas e corredores da universidade ao ouvir por anos os professores recriminando a escrita dos alunos indígenas. O segundo, na própria reserva indígena, por meio de professores, ex alunos dos cursos de graduação, também insatisfeitos com a produção dos textos escritos de suas crianças. Nesses momentos algo comum ocorria, a questão cultura e da língua materna sempre surgiam como possíveis fatores para esta dificuldade. Pensando em unir as duas pontas do ensino, mais especificamente, o fim do processo de alfabetização e o curso de graduação, que geralmente aponta o caminho para a profissionalização na vida adulta, a resolução em investigar as verdadeiras razões, razões baseadas em fatos concretos e não no preconceito e no “achismo” foi proposta esta tese.

Como objetivos específicos, almejou-se verificar se há diferenças significativas entre o tipo de dificuldade revelada pelos textos dos alunos indígenas Guarani/Kaiowá bilíngues em guarani e português e Terena monolíngues em português em comparação com os não indígenas monolíngues em português. Como segundo objetivo específico, observar em que medida o caminho percorrido do fundamental à universidade interferiu nas habilidades de escrita dos textos desses alunos, nesse momento os grupos serão divididos por faixa etária – crianças- grupo 1 e adultos – grupo 2. Por último, voltando o olhar apenas para o grupo 2, investigar se o tempo de permanência no curso de graduação interfere na capacidade de produzir textos escritos conforme os critérios estabelecidos como aceitáveis pela língua portuguesa.

Para conseguir alcançar esses objetivos foi organizado um *corpus*, a partir de textos coletados dos alunos/acadêmicos no decorrer do ano de 2011. Em 2012 esse corpus foi digitalizado, organizado em pastas, etiquetado e transformado em .txt para ser lido pelo programa *AntConc*. Das três ferramentas do programa, foi utilizado o concordanciador

(*concord*), que trabalha com blocos de palavras, no caso sintagmas. Essa metodologia está inserida nas perspectivas da Linguística de Corpus, uma área da Linguística que ainda ensaia seus primeiros passos na ciência linguística. Como auxiliar para a discussão morfossintática dos dados empregou-se a “Gramática Pedagógica do Português Brasileiro” de Marcos Bagno.

Para ler os dados e analisá-los foram feitos 3 cruzamentos no *Antconc*: no primeiro, agrupou-se:

4<sup>o</sup> e 5<sup>o</sup> ano do EF Guarani/Kaiowá bilíngue x o 4<sup>o</sup> e o 5<sup>o</sup> ano do EF Terena monolíngue;

4<sup>o</sup> e 5<sup>o</sup> ano do EF Guarani/Kaiowá bilíngue x o 4<sup>o</sup> e o 5<sup>o</sup> ano do EF não indígena;

4<sup>o</sup> e 5<sup>o</sup> ano do EF Terena monolíngue x 4<sup>o</sup> e 5<sup>o</sup> ano do EF não indígena. Nessa fase, os dados dos grupos 1 e o grupo 2 dialogaram entre si.

No grupo 2, agruparam-se os dados do 1<sup>o</sup> e do último ano do ES Guarani/Kaiowá bilíngue x 1<sup>o</sup> e último ano do ES Terena monolíngue;

1<sup>o</sup> e último ano do ES Guarani/Kaiowá bilíngue x 1<sup>o</sup> e último ano do ES Não Indígena;

1<sup>o</sup> e último ano do ES Terena monolíngue x 1<sup>o</sup> e último ano do ES Não Indígena.

Nesse primeiro agrupamento, visou-se responder a questão se havia diferenças significativas entre os tipos de dificuldades na escrita dos grupos indígenas em comparação com a dos não indígenas. Com relação a isso, apurou-se que na comparação entre os textos das crianças indígenas eles possuem praticamente as mesmas dificuldades e comparando-os com os textos dos não indígenas, percebeu-se que eles também trazem bastante dificuldade na escrita. A questão é que os textos das crianças indígenas apontam um índice bastante alto de hipo e hipersegmentação e os não indígenas trazem essa dificuldade num score menor. O subgrupos adulto apresentou praticamente as mesmas dificuldades: a hipo e a hipersegmentação continuam ocorrendo (de forma menos acentuada no grupo não indígena), bem como as dificuldades de concordância tanto nominal quanto verbal.

Dessa forma, com relação ao primeiro objetivo que era verificar se os tipos de dificuldades eram as mesmas pode-se dizer que sim, muito embora nos textos dos não indígenas do grupo 1 e 2 as ocorrências pontuem menos, mas as dificuldades existem neles também. Isso mostra que a questão não é linguística e nem cultural, haja vista que na

comparação entre os indígenas os Guarani/Kaiowá bilíngues tiveram menos dificuldades em alguns itens do que os Terena que são falantes apenas do português. O grupo de crianças não indígena foi selecionado de uma escola tida como “padrão” de ensino na cidade, ou seja, é uma escola localizada na zona central e sua clientela é composta pela classe média. Se, porventura fosse escolhida uma escola de periferia, provavelmente o resultado entre o cruzamento de dados de crianças indígenas e não indígenas daria um resultado muito mais próximo.

Com relação ao segundo objetivo, em que foram comparadas se os casos de dificuldades nas estruturas morfossintáticas permaneceram ou foram sanadas no decorrer do processo educacional, na comparação do grupo 1 (4º e 5º anos do EF agrupados) e o grupo 2 (1º e último ano do ES agrupados) em cada um dos grupos étnicos/ de língua, verificou-se que as dificuldades nessas estruturas linguísticas não foram suficientemente trabalhadas pela escola. Isso pode ser dito verificando os dados estatísticos presentes nas tabelas. As dificuldades de hipo e hipersegmentação continuam aparecendo no ensino superior, mesmo nos textos dos não indígenas, somados às dificuldades de concordância nominal e verbal, sobretudo ausência ou erro de concordância nominal de número (SNN) e ausência ou erro de concordância verbal de número e pessoa (SVNP). Logo, o caminho percorrido do ensino básico rumo à formação universitária não interferiu de forma suficiente no desenvolvimento das habilidades escritas do grupo estudado.

Por fim, o último objetivo específico restringiu-se ao grupo 2 – adultos (1º e último ano do curso de graduação). Intentou-se investigar se o tempo de permanência no curso universitário interferiu no nível de dificuldades morfossintáticas elencadas para este trabalho. Para isso, foram comparadas as ocorrências de dificuldades morfossintáticas dos alunos do 1º ano do ES com as ocorrências demonstradas pelos alunos do último ano do ES de cada grupo étnico/ de língua. O que se verificou, após a leitura dos dados foi que não houve mudança significativa no decorrer desses anos universitários. O tempo de permanência no curso não sanou as dificuldades que os acadêmicos indígenas e não indígenas traziam consigo.

Isso leva a alguns questionamentos. Primeiro, qual o papel que a universidade demanda quando se trata do ensino da escrita? A escrita, tem de fato recebido a relevância necessária dos vários cursos de graduação? Olhando para a escrita dos acadêmicos indígenas

pode-se afirmar que suas dificuldades de escrita em língua portuguesa se deve exclusivamente à questão da primeira língua ou da cultura?

Ficou constatado que os acadêmicos não indígenas também saem dos cursos de graduação permanecendo com as mesmas dificuldades na escrita ou até acrescentando mais problemas do que quando entraram.

O que se deduz com esses dados, de forma geral, é que dificuldades de escrita exigem um olhar e investigações mais aprofundadas e pontuais em cada situação. A ideia que sempre perpassou no meio acadêmico e até fora dele, pelo menos na cidade de Dourados e região era de que os indígenas não conseguiam escrever de forma satisfatória em língua portuguesa devido a sua materna ser a indígena. Também por causa de sua cultura, não obstante o que os dados apontam é que não apenas isso não é verdade, como mostra a visão deturpada e preconceituosa, que a sociedade construiu a respeito do indígena. Percebe-se o ressurgimento da concepção separatista de sociedades letradas – mais avançadas, mais modernas, de sociedades não letradas vinculadas ao atraso, à ineficiência a qual os indígenas insistentemente são atrelados.

Nessa tese ficou comprovado que as dificuldades dos indígenas são praticamente as mesmas que a dos não indígenas, o que se destacou como significativo é a quantidade nas quais elas ocorrem. A presença da hipossegmentação e da hipersegmentação, por exemplo, é um caso bastante acentuado no grupo indígena, por isso se deve prestar atenção. Se as dificuldades ocorrem, as soluções devem aparecer sob forma de material didático direcionado para essa questão específica, para o problema desse grupo. Cursos de capacitação de professores devem visar também essas dificuldades pontuais, locais. Afinal, da perspectiva dos novos letramentos deve-se trabalhar a partir do local e se estender ao global.

## REFERÊNCIAS

AYLWIN, J. *Os Direitos dos Povos Indígenas em Mato Grosso do Sul, Brasil – Confinamento e Tutela no Século XXI*. Informe IWGIA 2009-3. Grupo Internacional de Trabalho sobre Assuntos Indígenas (IWGIA) e Faculdade de Medicina da USP (FMUSP).Copenhague, Dinamarca. Impressão: Rettec Artes Gráficas, Brasil, 2009.

BAGNO, Marcos. *Gramática Pedagógica do Português Brasileiro*. São Paulo: Parábola Editorial, 2011.

BERBER SARDINHA, Tony. *Linguística de Corpus*.Barueri:Manole, 2004.

BISOL, Leda. O acento e o pé binário. *Cadernos de Estudos Linguísticos*. Campinas, janeiro a junho 1992.

<http://revistas.iel.unicamp.br/index.php/cel/article/view/2753/2184>

BRASIL. *Ministério da Educação e do Desporto*. Referencial Curricular Nacional para as Escolas Indígenas. *Brasília:MEC-SEF*, 1998

CASTILHO, Ataliba T. de. *Nova Gramática do Português Brasileiro*. São Paulo:Editora Contexto, 2010.

CUNHA, A. P. N. da. *A Hipo e a Hipersegmentação nos dados de Aquisição da Escrita: Um estudo sobre a influência da prosódia*, 2004. Dissertação (Mestrado em Educação) - Faculdade de Educação, UFPel, Pelotas, 2004.

DE PAULA, Luís Roberto e VIANNA, Ferne de Luiz Brito. *Mapeo políticas públicas para povos indígenas*.Rio de Janeiro:contra Capa Livraria; LACED/Museu Nacional/UFRJ, 2011.

FERREIRO, Emilia, TEBEROSKY, Ana. *Psicogênese da língua escrita*. Porto Alegre: Artes Médicas Sul, 1999.

FREITAS, Ezequias Matos de. *Curso breve Del idioma guarani*. 2ª Ed. P. Pablo Scott SVD: Assuncion, 1981.

GARCIA , Daiani de Jesus. *A Influência da Oralidade na Escrita das Séries Iniciais: Uma Análise a Partir de Erros Ortográficos* . Dissertação de Mestrado em Educação/UFPel.Pelotas/RS,2010.

[http://www.livrosgratis.com.br/arquivos\\_livros/cp147199.pdf](http://www.livrosgratis.com.br/arquivos_livros/cp147199.pdf)

GOMES, Mércio Pereira. *Os índios e o Brasil: passado, presente e futuro*. São Paulo: Contexto, 2012.

GOODY, Jack. *A domesticação da mente selvagem*. Trad. Vera Joscelyne. Petrópolis, RJ: Vozes, 2012.

GOODY, Jack e WATT, Ian. *As consequências do letramento*. Trad. Waldemar Ferreira

GRANGER, Silviane; LUNG, Joseph e PETCH-TYSON, Stephanie (Orgs.) *Computer Learner Corpora, Second Language Acquisition e Foreign Language Teaching*, 2002  
[http://books.google.com.br/books?id=mWoss1TqW6ICeprintsec=frontcoverhl=pt-BResource=gbs\\_ge\\_summary\\_recad=0#v=onepageeqef=false](http://books.google.com.br/books?id=mWoss1TqW6ICeprintsec=frontcoverhl=pt-BResource=gbs_ge_summary_recad=0#v=onepageeqef=false)

KLEIMAN, Angela. (Org.) *Modelos de letramento e as práticas de alfabetização na escola*. In: *Os significados do letramento*. Uma nova perspectiva sobre a prática social da escrita. Campinas: Mercado de Letras, 1995.

LIN, Phoebe M.S. e ADOLPHS, Svenja. *Corpus Linguistics*. In *The Routledge Handbook of Applied Linguistics*. Ed. James Simpson by Routledge 2 Park Square. Milton Park, Abingdon. Taylor e Francis e-Library, 2011.

MAHER, Terezinha Machado. *Formação de professores indígenas: uma discussão introdutória*. In GRUPIONI, Luis Donisete Benzi. (Org). *Formação de professores indígenas: repenseo trajetórias*. Brasília: Ministério da Educação, Secretaria de Educação Continuada, Alfabetização e Diversidade, 2006.

McENERY, Tony e WILSON, Erew. *Corpus Linguistics*. 2ª ed. Edinburgh: Edinburgh University Press Ltd, 1996.

MELIÁ, Bartolomeu. *Educação indígena e alfabetização*. São Paulo: Edições Loyola, 1979.

MOURA NEVES, Maria Helena de. *A gramática passada a limpo – conceitos, análises e parâmetros*. São Paulo: Parábola Editoria, 2012.

PACHECO, Aline. *A aquisição de morfemas em inglês como L2: uma análise dos padrões evolutivos através do BELC*. Tese de doutorado, Porto Alegre/RS, 2010.

Rayson (2002, apud SARMENTO, 2008), in: *O uso dos verbos modais em manuais de aviação em inglês: um estudo baseado em corpus*. Tese de Doutorado- Universidade Federal do Rio Grande do Sul, Departamento de Letras: Porto Alegre, 2008.

ROLLINSON, Paul; MENDIOETXEA, Amaya e BIELSA, Susana Murcia. *Focus on Errors: Learner Corpora as Pedagogical Tools*. In *Corpus-Based Approaches to English Language Teaching*. Mari Carmen Campoy, Begona Belles-Fortunato e Ma Lluïsa Gea-Valor (Org.), Continuum International Publishing Group, 2010.

SCOTT, P. Pablo. *Curso del Idioma Guaraní*. 1981. Assunção. Py.

STREET, Brian V. *Literacy in theory e practice*. Cambridge, UK. Cambridge University Press, 1984.

\_\_\_\_\_. Os novos estudos sobre o letramento: histórico e perspectivas. In: MARINHO, Marildes e CARVALHO, Gilcinei Teodoro de. (Orgs.). *Cultura escrita e letramento*. Belo Horizonte: Editora UFMG, 2010.

SARMENTO, Simone. O uso dos verbos modais em manuais de aviação em inglês: um estudo baseado em *corpus*. Tese de Doutorado- Universidade Federal do Rio Grande do Sul, Departamento de Letras: Porto Alegre, 2008.

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

SILVA, A. *A relação entre a fala e a segmentação na escrita espontânea de crianças da primeira série do primeiro grau*. Dissertação (Mestrado em Linguística). Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1989.

TARALLO, Ferneo. A pesquisa Socolinlinguística. 2 edição. São Paulo: Editora Ática, 1999.

TORRES, Isabelle. O colapso da política indígena. In: *REVISTA ISTO É!* Nº 2273. São Paulo: Editora Três. 12/06/2013. P.36-41.

VEÇOSSO, Cristiano Egger e FERREIRA-GONÇALVES, Giovana A Hipo e a Hipersegmentação em Dados de Escrita de Alunos da 8ª Série: Influência Exclusiva dos Constituintes Prosódicos? *Revista Verba Volant*. Volume 2 – Número 1 – jan/jun 2011 – ISSN 2178-4736

VIANA, Veer e TAGNIN, Stella, E. O. (Orgs.) *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010.

VIANA, Veer. Verbos modais em contraste: análise de *corpus* da escrita de universitários em inglês. 2008. 230 f. Dissertação (Mestrado em Letras) – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, RJ. *Apud* VIANA, Veer e TAGNIN, Stella, E. O. (Orgs.) *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010.

<http://portal.mec.gov.br/sesu/arquivos/pdf/leis2.pdf> Parecer 14/99

<http://portal.mec.gov.br/cne/arquivos/pdf/CEB0399.pdf> Resolução 03/99

[http://www.ufpe.br/remdipe/index.php?option=com\\_contentview=articleid=306&Itemid=238](http://www.ufpe.br/remdipe/index.php?option=com_contentview=articleid=306&Itemid=238) Rede de monitoramento dos povos indígenas do Pernambuco.

(Sem autoria - Documento Intitulado “A implantação do regime tutelar e atividades escolares na Reserva de Dourados e na Aldeia Panambizinho”, Maio/Junho 2007).

## ANEXOS



ADDGBPLERM4FT TXT1.txt

A minha vo conto (HIPER) um ma histo de um cachorro estava indo ni um mato e viu um gato chegou perto deli e o cachorro foi (HIPER)em bora e o gato (HIPER)ssi guiu ele até (HIPO)nacassa (SVT)para de chove o cchoro e o gato foi (HIPER)em borra (HIPO)nomesmo mato que o chorro acho no outro dia o cachorro foi até la Para ve o gato o gato sail (HIPO)domato (HIPER)e Liz (HIPER)esta vam Brinceo (HIPER)na queli mato é o cahorro correu porque o (HIPER)do mo estava chameo ele?



ADDGBPLERM4FT TXT1.txt

A minha vo conto um ma histo de um cachorro estava indo ni um mato e viu um gato chegou perto deli e o cachorro foi em bora e o gato ssi guiu ele até nacassa para de chove o cchoro e o gato foi em borra nomesmo mato que o chorro acho no outro dia o cachorro foi até la Para ve o gato o gato sail domato e Liz esta vam Brinceo na queli mato é o cahorro correu porque o do mo estava chameo ele?



ACSGB1SUP TXT01.txt

A muitos anos atraz queo meus pais eram vivos. Eles sempre-se assentavam em volta da fogueira para nos conta uma história de nossos antepassados. E uma delas é essa que estarei relateo um pouco para vocês, que é sobre o sol e a lua.

A muitos anos atraz uma linda moça fico gravida é seu pai fico muito bravo e toco ela de casa pois ela nem sabia de quem tinha engravidado. Na verdade ela tinha ficado gravida sem ter ficado com (HIPER)nem um homem. Queo o pai dela toco ela de casa, entao ela foi mora no mato. Lá ela fez uma casa para se abrigar da chuva e do frio. Toda noite vinha uma onça para tentar come ela, mas ela muito esperta subia em uma árvore fina reta ate que clareasse o dia. E assim foram passeio os meses. ate que um dia ela sentil dores e precisava de ajuda. Então ela fico na casa dela e a onça ja tava pra chega de novo. Ela fez força e deu a luz a dois meninos (SNN)lindo. Assim que ela ganhou os bebes, (SVNP)sai de casa correndo para o meio do mato, la ela encontrou o pai do mato. Então ela pediu que ele cuidace dos meninos, pois a onça iria (HIPO)comela.

Então o pai do mato disse a ela que iria salvar os meninos, mas iria (HIPO)separalos para que sempre fossem visto por todos. Queo o pai do mato sail a onça pego ela e devorol. Nesse instante o pai do mato chego na casa e vil os dois bebes.

Um chorava muito e o outro tava quitinho. Então o pai do mato fez um ritual e disse que o mas quieto iria virar a lua e aquele que estava bravo seria o sol. Dai endiante(HIPO)nunca mais teve escuridão, apenas o sol (SVT)escureci queo os dois si (SVT)encontram. Que e chamada de eclipse. O sol e muito bravo (HIPO)poriço e quente e a lua calma e fria.



ACSGB1SUP TXT01.txt

A muitos anos atraz queo meus pais eram vivos. Eles sempre-se assentavam em volta da fogueira para nos conta uma história de nossos antepassados. E uma delas é essa que estarei relateo um pouco para vocês, que é sobre o sol e a lua.

A muitos anos atraz uma linda moça fico gravida e seu pai fico muito bravo e toco ela de casa pois ela nem sabia de quem tinha engravidado. Na verdade ela tinha ficado gravida sem ter ficado com nem um homem. Queo o pai dela toco ela de casa, entao ela foi mora no mato. Lá ela fez uma casa para se abrigar da chuva e do frio. Toda noite vinha uma onça para tentar come ela, mas ela muito esperta subia em uma árvore fina reta ate que clareasse o dia. E assim foram passeio os meses. ate que um dia ela sentil dores e precisava de ajuda. Então ela fico na casa dela e a onça ja tava pra chega de novo. Ela fez força e deu a luz a dois meninos lindo. Assim que ela ganhou os bebes, sai de casa correndo para o meio do mato, la ela encontrou o pai do mato. Então ela pediu que ele cuidace dos meninos, pois a onça iria comela.

Então o pai do mato disse a ela que iria salvar os meninos, mas iria separalos para que sempre fossem visto por todos. Queo o pai do mato sail a onça pego ela e devorol. Nesse instante o pai do mato chego na casa e vil os dois bebes.

Um chorava muito e o outro tava quitinho. Então o pai do mato fez um ritual e disse que o mas quieto iria virar a lua e aquele que estava bravo seria o sol. Dai endiantenunca mais teve escuridão, apenas o sol escureci queo os dois si encontram. Que e chamada de eclipse. O sol e muito bravo poriço e quente e a lua calma e fria.



DINTER – UFRGS – UNIVERSIDADE FEDERAL DE MATO GROSSO DO  
SUL/UEMS-UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL.  
PESQUISADORA: SERA ESPÍNDOLA  
ORIENTADORA: INGRID FINGER  
QUESTIONÁRIO PARA AS CRIANÇAS DO ( ) 4<sup>o</sup> OU 5<sup>o</sup> ( ) ANO DO  
ENSINO FUNDAMENTAL INDÍGENA.

Participante n° \_\_\_\_\_

Prezado participante do estudo, Por favor, preencha as lacunas abaixo e responda às perguntas:

(1) Nome Completo \_\_\_\_\_

(2) Idade: \_\_\_\_\_ (3) Sexo: ( ) F ( ) M

(4) Com que idade ingressou na escola? \_\_\_\_\_

(5) Iniciou os estudos na escola da Reserva? \_\_\_\_\_

(6) Você fala outras línguas em casa além do português? ( ) SIM ( ) NÃO

(7) Quais (caso tenha respondido SIM na questão anterior) \_\_\_\_\_

(8) Por favor, liste as línguas que você entende que ouve:

1 \_\_\_\_\_ 2 \_\_\_\_\_ 3 \_\_\_\_\_ 4 \_\_\_\_\_

(9) Escreva as línguas que você ouve mais

em casa \_\_\_\_\_ na escola \_\_\_\_\_ na cidade \_\_\_\_\_

(10) Ligue de 1 a 5, ***por ordem de importância*** as culturas com as quais você mais se identifica:

Guarani 5

Kaiowá 4

Terena 2

Mestiços 3

Não-indígenas 1

(11) Ligue de 1 a 6, ***por ordem de importância*** as línguas que você acha que são importantes aprender a falar:

guarani 5

Kaiowá 4

Terena 2

português 6

Inglês 3

espanhol 1

(12) Que língua você fala com as seguintes pessoas:

	Língua
Pai	
Mãe	
Irmãos	
Avós	
Tios	
Primos	
Amigos da escola	
Vizinhos	
Igreja	

(13) Que língua utiliza para fazer as seguintes coisas:

Atividade	Língua
Brincar	
Fazer as refeições	
Rezar ( em casa)	
Ajudar nas tarefas de casa	
Contar histórias	
Rezar (na Igreja)	

(14) Com quem aprendeu as línguas que fala:

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_

Obs.: Não foi perguntado em que língua assistem TV, ouvem rádio ou leem porque essas coisas só são possíveis na língua portuguesa, uma vez que a língua indígena terena, kaiowá ou guarani não têm material suficiente disponível para leitura ou para veiculação nos meios de comunicação.

MUITO OBRIGADA POR PARTICIPAR DO EXPERIMENTO!



DINTER – UFRGS – UNIVERSIDADE FEDERAL DE MATO GROSSO DO  
SUL/UEMS-UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL.

PESQUISADORA: SANDRA ESPÍNDOLA

ORIENTADORA: INGRID FINGER

QUESTIONÁRIO PARA AS CRIANÇAS DO 5<sup>o</sup> ANO DO ENSINO  
FUNDAMENTAL INDÍGENA.

Participante n° \_\_\_\_\_

Prezado participante do estudo, Por favor, preencha as lacunas abaixo e responda às perguntas:

(1) Nome Completo \_\_\_\_\_

(2) Idade: \_\_\_\_\_ (3) Sexo: ( ) F ( ) M

(4) Com que idade ingressou na escola? \_\_\_\_\_

(5) Iniciou os estudos na escola da Reserva? \_\_\_\_\_

(6) Você fala outras línguas em casa além do português? ( ) SIM ( ) NÃO

(7) Quais (caso tenha respondido SIM na questão anterior) \_\_\_\_\_

(8) Por favor, liste as línguas que você entende ou ouve:

1 \_\_\_\_\_ 2 \_\_\_\_\_ 3 \_\_\_\_\_ 4 \_\_\_\_\_

(9) Escreva as línguas que você ouve mais

em casa \_\_\_\_\_ na escola \_\_\_\_\_ na cidade \_\_\_\_\_

(10) Numere de 1 a 5 as culturas com as quais você mais se identifica:

( ) guarani

( ) kaiowá

( ) terena

( ) mestiços

( ) não-indígenas

(11) Numere de 1 a 6, por ordem de importância as línguas que você acha que são importantes aprender a falar:

( ) guarani

( ) kaiowá

( ) terena

( ) português

( ) inglês

( ) espanhol

(12) Que língua você fala com as seguintes pessoas:

	Língua
Pai	
Mãe	
Irmãos	
Avós	
Tios	
Primos	
Amigos da escola	
Vizinhos	
Igreja	

(15) Que língua utiliza para fazer as seguintes coisas:

Atividade	Língua
Brincar	
Assistir TV(se tiver)	
Ouvir o rádio (se tiver)	
Fazer as refeições	
Rezar ( em casa)	
Ajudar nas tarefas de casa	
Ler	
Contar histórias	
Rezar (na Igreja)	

(16) Com quem aprendeu as línguas que fala:

Língua	Pessoas (pai, mãe, avós, irmãos, etc.)

MUITO OBRIGADO POR PARTICIPAR DO EXPERIMENTO!!



DO

DINTER – UFRGS – UNIVERSIDADE FEDERAL DE MATO GROSSO

SUL/UEMS-UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL.

PESQUISADORA: SERA ESPÍNDOLA

ORIENTADORA: INGRID FINGER

QUESTIONÁRIO PARA ACADÊMICOS UNIVERSITÁRIOS INDÍGENAS

Participante nº \_\_\_\_\_

Prezado participante do estudo, por favor, preencha as lacunas abaixo e responda às perguntas:

- (1) Nome Completo \_\_\_\_\_
- (2) Idade: \_\_\_\_\_ (3) Sexo: ( ) F ( ) M
- (4) Assinale sua escolaridade: ( ) Primeiro ano do curso de \_\_\_\_\_ ( ) Último ano do curso de \_\_\_\_\_ Universidade \_\_\_\_\_
- (5) Com que idade ingressou na escola? \_\_\_\_\_
- (6) Iniciou na escola da Reserva? ( ) sim ( ) não
- (7) Com que idade terminou o ensino médio? \_\_\_\_\_
- (8) Fez ensino médio na escola da Reserva? ( ) sim ( ) não
- (8) com que idade iniciou o curso universitário? \_\_\_\_\_
- (9) Você fala outras línguas em casa além do português? ( ) Sim ( ) Não
- (10) Quais? (caso tenha respondido SIM na questão anterior) \_\_\_\_\_
- (11) Por favor, liste as línguas que você entende que ouve:
1. \_\_\_\_\_ 2. \_\_\_\_\_ 3. \_\_\_\_\_ 4. \_\_\_\_\_
- (12) Escreva as línguas que você ouve mais em casa \_\_\_\_\_ na universidade \_\_\_\_\_ na cidade \_\_\_\_\_
- (13) Ligue de 1 a 5, ***por ordem de importância*** as culturas com as quais você mais se identifica:

Guarani	5
Kaiowá	4
Terena	2
Mestiços	3
Não-indígenas	1

(14) Ligue de 1 a6, ***por ordem de importância*** as línguas que você acha que são importantes aprender a falar:

guarani	5
Kaiowá	4
Terena	2
português	6
Inglês	3
espanhol	1

(15) Que língua você fala com as seguintes pessoas:

	Língua
Pai/mãe	
Filhos	
Irmãos	
Avós	
Tios/primos	
Amigos da universidade	
Vizinhos	
Esposo(a)	

(15) Que língua utiliza para fazer as seguintes coisas:

Atividade	Língua
Fazer as refeições	
Rezar	
Ajudar nas tarefas de	
Contar histórias	
Ler a Bíblia	

(16) Com quem aprendeu as línguas que fala:

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_

Língua: \_\_\_\_\_ com quem aprendeu: \_\_\_\_\_