

LÍNGUA DA TABATINGA: INVESTIGAÇÃO À LUZ DA LINGUÍSTICA DE *CORPUS*

Language of Tabatinga: An Exploration Guided by Corpus Linguistics

DOI: 10.14393/LL63-v38-2022-28

Roberta Gê-Acaiaba*

Ariel Novodvorski**

RESUMO: O presente estudo, guiado por *corpus* (*corpus-driven*) e amparado pela Sociolinguística Variacionista, apresenta a Língua da Tabatinga (doravante LT), uma realidade linguística que se presentifica na cidade de Bom Despacho - MG. As origens dessa língua remontam ao passado escravocrata do Brasil, que levou aos mais diversos cantos da nação a barbárie dos homens brancos contra pessoas negras, e resultou na formação de uma língua verdadeiramente “mestiça”, formada a partir de contribuições linguístico-culturais africanas de origem bantu e de variedades da língua portuguesa brasileira falada no interior de Minas Gerais. Assim, parte-se de uma descrição do processo histórico de formação da LT; segue-se com a identificação de sua presença nas manifestações linguísticas da sociedade bondespachense; e almejam-se a identificação e descrição do processo de formação lexical e morfossintática e da abrangência semântico-pragmática da LT, com subsídios da Linguística de *Corpus*.

PALAVRAS-CHAVE: Língua da Tabatinga. Linguística de *Corpus*. Variação Linguística. Léxico.

ABSTRACT: The current corpus-driven study draws on Variationist Sociolinguistics to present the Language of Tabatinga, a linguistic reality present in the City of Bom Despacho, state of Minas Gerais, in Brazil. The origin of this language traces back to the slavery past of Brazil, which took to the most diverse places of the nation the barbarity of white men against black people and resulted in a truly ‘mixed’ language, formed from African linguistic-cultural contributions of Bantu origin and varieties of Brazilian Portuguese spoken in the Minas Gerais countryside. This study starts from a description of the historical process of this language formation, identifies its presence in the linguistic manifestations of the society in Bom Despacho, and describes the process of lexical and morphosyntactic formation as well as semantic-pragmatic coverage of the Language of Tabatinga, with subsidies from Corpus Linguistics.

KEYWORDS: Language of Tabatinga. *Corpus* Linguistics. Linguistic Variation. Lexicon.

* Doutoranda em Linguística pelo Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) no âmbito do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU). ORCID: 0000-0001-6707-0933. E-mail: robertage(AT)ufu.br

** Doutor em Estudos Linguísticos pela Universidade Federal de Minas Gerais (UFMG), com estágio pós-doutoral pela Universidade Federal do Rio Grande do Sul (UFRGS). Professor associado do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU). ORCID: 0000-0003-1370-8334. E-mail: arivorski(AT)ufu.br

1 Introdução

A sociedade brasileira é formada por um encontro étnico. Um encontro a partir do qual surgiram diferentes formas de representar a realidade em que vivemos. Assim, a identidade de nosso país, tecida hoje frente aos nossos olhos, é pluriétnica, o que se escancara não apenas na formação do nosso povo, mas também e muitas vezes de modo mais marcante, nas práticas sociais que realizamos em nossas interações diárias, conforme afirma Ribeiro (1995).

Matrizes raciais díspares, tradições culturais distintas, formações sociais defasadas se enfrentam e se fundem para dar lugar a um povo novo, num novo modelo de estruturação societária. Novo porque surge como uma etnia nacional, diferenciada culturalmente de suas matrizes formadoras, fortemente mestiçada, dinamizada por uma cultura sincrética e singularizada pela redefinição de traços culturais delas oriundos. (RIBEIRO, 1995, p. 19)

Diante dessas práticas sociais heterogêneas que se formaram e ainda se formam em nossa sociedade, a exemplo do acolhimento de refugiados, presentifica-se na cidade de Bom Despacho - MG a Língua da Tabatinga. Língua essa formada a partir de contribuições linguístico-culturais africanas de origem bantu e de variedades da língua portuguesa brasileira falada no interior de Minas Gerais.

A importância da investigação proposta neste artigo reside na noção elementar, conforme Labov (2008), de que não há comunidades linguisticamente homogêneas, e que a heterogeneidade na formação das línguas é algo natural. Logo, o que se observa hoje em Bom Despacho faz parte da dinâmica das línguas, organismos vivos que se moldam ao mesmo tempo em que moldam a sociedade na qual se desenvolvem. Cabe, portanto, evidenciar que a chave do estudo proposto está na possibilidade de compreender cada vez mais sobre o processo de formação das línguas, avaliando as transformações no momento exato em que elas ocorrem, vivenciando concreta e diacronicamente o fenômeno da constituição de uma nova língua.

Assim, como membro constituinte da sociedade bondespachense, a partir de um mínimo conhecimento prévio da LT e tendo a Língua Portuguesa Brasileira (LPB) como língua materna, minhas hipóteses que surgem para tal estudo são que a LT, de acordo com seu processo de formação, concede à LPB da cidade de Bom Despacho uma expansão lexical, a partir da inserção de vocábulos próprios e específicos do léxico Tabatinga. Assim, por influência histórica e social, a variedade linguística da LPB, existente na cidade de Bom Despacho, agrega elementos típicos da LT, o que a torna única em nosso país.

Além disso, há também a hipótese de que a inserção da LT nas práticas de linguagem da sociedade bondespachense funciona como uma política de autoafirmação e resistência, garantindo a redução do preconceito linguístico e a redução do preconceito social para com a comunidade afrodescendente, a partir da qual a LT se originou.

2 Pressupostos teóricos

Foi no início do século XX que começaram a germinar os estudos que viriam posteriormente – depois de cerca de meio século de domínio de correntes estruturalistas – a florescer e dar frutos no terreno fecundo da área de estudos da linguagem, que veio a ser conhecida como Sociolinguística.

Tendo como expoente William Labov e contrariando o Estruturalismo e o Gerativismo, em 1960 surge a Sociolinguística, que busca tratar do valor social das formas linguísticas e do estudo empírico das mudanças. Logo, falar sobre Sociolinguística é falar sobre regras variáveis, e a determinação dessas regras variáveis é o ponto de partida para o estudo sociolinguístico, que se dá com a presença do componente social.

Em poucas palavras, portanto, o objetivo da Sociolinguística é descobrir os mecanismos que regulam a variação e como essa variação interage com outros elementos do sistema linguístico e com a matriz social em que ocorre, a fim de ocasionar uma mudança linguística. “Com efeito, a Sociolinguística se ocupa da relação entre língua e sociedade, e do estudo da estrutura e da evolução da linguagem dentro do contexto social da comunidade de fala” (COELHO *et al.* 2010, p. 22).

Assim, interessa para o presente estudo a Sociolinguística, visto que toda língua (assim como a LT) apresenta variações decorrentes da heterogeneidade presente nos fenômenos linguísticos, as quais são identificadas e analisadas por meio de registro, descrição e análise sistemática, relacionando essas variações com fatores sociais, numa tentativa de identificar qual fator ou grupo de fatores é o responsável por determinada variação.

Dessa forma, ao reconhecer a heterogeneidade sistemática como inerente a qualquer língua, o estudo reforça a noção elementar de que estudar uma língua é estudar a sociedade heterogênea em que ela se constrói. Assim, para tecer um estudo acerca da LT, nada mais relevante que estudar a comunidade desde o mesmo interior em que se faz presente.

Localizada na periferia da cidade de Bom Despacho, a Tabatinga é um bairro onde se abrigaram os escravos após sua libertação, já que a maioria dos negros vieram para Bom Despacho na condição de cativos juntamente com seus senhores. Assim, após obterem a tão esperada liberdade, os negros se viram, além de livres, desabrigados.

Dessa forma, grande número de indivíduos negros resolveu fixar residência neste local, que posteriormente viria a ser o bairro Tabatinga. Os primeiros moradores do bairro eram escravos da lavoura e se instalaram no local a fim de trabalhar em um garimpo que ali, outrora, existiu. Descobrimo que não havia nenhum metal precioso, cavaram a terra branca do garimpo (tabatinga) e barrearam suas moradias, passando em seguida a tecerem cobertas e cortar capim para vender à confecção de colchões.

Hoje ainda reside na Tabatinga grande número de descendentes dos negros africanos que neste local se instalaram no passado, deixando não somente suas lembranças, mas também sua cultura e sua presença por meio de diversas manifestações afro-brasileiras, que ainda sobrevivem no bairro. Entre as manifestações mais expressivas da cultura negra que ainda sobrevivem no bairro, podem ser citadas as religiões africanas, notadas no grande número de benzedeadas e rezadeiras do bairro, o congado que ainda move números de adeptos a dançar por Nossa Senhora do Rosário e, principalmente, a língua, que faz uso de grande número de vocábulos e expressões da língua africana.

Justifica-se, portanto, a necessidade da abordagem da Lexicologia neste presente estudo, visto que o léxico da LT não é um bloco monolítico, mas um conjunto heterogêneo e dinâmico, e que possui regularidades constitutivas e comportamentais que podem ser investigadas através dos estudos lexicológicos. Assim, a Lexicologia pode oferecer bases para a descrição linguística estruturada no funcionamento discursivo das palavras, na dimensão semântica e nas contextualizações pragmático-comunicativas, o que encontra sustentação nos objetivos do estudo em questão. Krieger (2006, p.160), afirma que “o léxico é um componente de muitas facetas e que ocupa um lugar central nas línguas, tonando-se, em consequência, um ponto de cruzamento dos estudos linguísticos”. A diversidade de campos gramaticais, linguísticos e discursivos que se inter-relacionam com o léxico permite múltiplas abordagens, interligando-o a fatores linguísticos e extralinguísticos, o que corrobora um lugar de destaque aos itens lexicais nos estudos linguísticos.

Também os estudos fraseológicos possuem relevância no estudo em questão, visto que, compreendida a LT como um sistema gramatical pertencente a um grupo de indivíduos, intrinsecamente ligado à estrutura social, quando os fraseologismos aparecem nas gravações orais da LT, são culturalmente marcados, antes de tudo, pela escolha, consciente ou não, de unidades lexicais que permitem chegar a um referente extralinguístico.

Ademais, o estudo em questão ampara-se na Linguística de *Corpus* (LC), uma vez que propicia a possibilidade de uma análise empírica da LT. Ao se falar em LC, ponderamos a possibilidade de construção de estudos mais dinamizados e centrados em dados reais/naturais e análises cada vez mais confiáveis. Logo, o novo paradigma trazido pela LC aos estudos da linguagem (empíria), em nosso caso ao estudo da LT, a nosso ver, constrói um viés teórico inalienável, que garante coleta de dados e rigor metodológico nas análises dos dados coletados.

3 Metodologia

A pesquisa em questão é recorte de uma tese de doutoramento, em curso, que está sendo realizada a partir de uma abordagem quali-quantitativa. A escolha de tal abordagem metodológica reside no fato de possibilitar múltiplas formas de coleta de dados. Nesse sentido, os fatos observados não são previamente determinados; como já indicam Alves-Mazzotti e Gewandszajder (1999, p. 21), a metodologia quali-quantitativa oferece a oportunidade de

descrever e compreender o que está ocorrendo em uma dada situação, sem nos preocuparmos com paradigmas ou modelos predeterminados, favorecendo assim à observação do fenômeno em seu local de ocorrência, com posteriores combinações de métodos, tais como análise de documentos, entrevistas, entre outros, que culminará em um rico material para orientar o pesquisador nas tarefas de registro, análise e apresentação de dados. (ALVES-MAZZOTTI; GEWANDSZAJDER, 1999, p. 25)

Logo, para a realização deste estudo, está sendo feita uma pesquisa sociolinguística, com suporte da LC, para compreender como ocorre o processo de formação, constituição e permanência da LT na cidade de Bom Despacho.

Assim, o primeiro passo se dá a partir da coleta de dados bibliográficos que versam sobre as contribuições africanas para formação sociocultural, linguística e étnica do Brasil.

Em sequência, está sendo feito um reconhecimento e descrição da comunidade linguística analisada: comunidade da Tabatinga. É necessário, portanto, uma pesquisa de campo, a fim de obter os dados necessários à identificação da comunidade.

O próximo passo é a observação e catalogação das manifestações da LT presentes nas práticas de linguagem, não só no interior do bairro da Tabatinga, bem como em toda a cidade de Bom Despacho (MG).

Portanto, para a coleta de dados é necessária a realização de entrevistas com pessoas pertencentes à comunidade do bairro Tabatinga e, também, daquelas que, mesmo fora do bairro, enquanto moradoras de Bom Despacho, fazem uso da LT em suas práticas de linguagem.

Para tanto, buscamos participantes que se encaixam nos critérios de inclusão e, depois de identificados, entramos em contato formalmente, a fim de apresentar a pesquisa, sua finalidade e detalhar como ocorrerá a participação, caso aceite participar. Sendo a proposta aceita, deixamos que o participante escolha a data, o horário e o local para realização da entrevista, que é norteadada pelos axiomas metodológicos propostos por Labov (2008), tendo em vista permitir que a LT vernácula apareça.

Estimamos que serão necessários entre trinta e sessenta participantes nesta pesquisa, de modo a oportunizar uma maior ocorrência e recorrência de itens lexicais da LT e, desse modo, cobrir a necessidade de analisar e descrever aspectos próprios da língua em estudo. Distribuimos, assim, os participantes por faixa etária (crianças e/ou adolescentes, adultos e idosos) e local de moradia no momento da entrevista (falante da língua, morador do bairro Tabatinga ou falante da LT que mora em Bom Despacho, mas não necessariamente no bairro em estudo). Desse modo, propomos a seguinte divisão: de 15 a 30 moradores do bairro (5 a 10 crianças; 5 a 10 adultos; 5 a 10 idosos) e de 15 a 30 moradores de Bom Despacho, mas externos ao bairro de ocorrência da LT (5 a 10 crianças; 5 a 10 adultos; 5 a 10 idosos).

No que diz respeito ao tempo de duração das entrevistas, podem variar de acordo com cada participante e sua relação com a proposta de entrevista que, levando em consideração Labov (2008), deverá solicitar a cada participante que fale sobre situações e/ou assuntos que recriem fortes emoções vivenciadas, a fim de favorecer a espontaneidade. O local em que as entrevistas acontecem é estabelecido mediante a escolha de cada um dos participantes, podendo ser realizadas, inclusive, em suas residências ou via plataforma de videoconferência, como Google Meet e/ou Zoom.

Conforme apresentado, é indispensável a participação de seres humanos na pesquisa e, por isso, houve a necessidade de sua submissão para a apreciação pelo Comitê de Ética em

Pesquisas com Seres Humanos para, então, obtermos legalidade no processo de coleta de dados. Tal submissão foi realizada no dia 24 de maio de 2021, com parecer de aprovação em 30 de setembro de 2021, sob o CAAE 247473721.6.0000.5152, parecer n.º 5.010.584.

Tendo como objetivo compreender como ocorre o processo de formação, constituição, inserção e permanência da LT na cidade de Bom Despacho, a coleta de dados se realiza por meio de entrevistas semiestruturadas com pessoas moradoras do bairro Tabatinga, que fazem uso da LT e pessoas externas ao bairro Tabatinga, mas moradoras da cidade de Bom Despacho, que também fazem uso da LT.

As entrevistas estão sendo gravadas em áudio no formato digital Mp3, com posterior transcrição, adotando as normas de transcrição de dados feita com base nos parâmetros estabelecidos pelo Projeto NURC-SP/USP.

Após a transcrição dos áudios das entrevistas, para processamento com ferramentas características da LC, é necessário converter cada um dos textos obtidos para o formato TXT, com a codificação ANSI, a fim de preparar o *corpus* para indagação por meio do programa de análises lexicais WordSmith Tools, versão 06 (SCOTT, 2012), utilizado ao longo de todo o estudo.

As entrevistas seguem um roteiro previamente estabelecido, montado a fim de obter o vernáculo dos entrevistados. Nesse sentido, é solicitado que os participantes falem de fatos da infância, situações familiares marcantes, sobre como é a vida na cidade, situações engraçadas ou de risco. Logo, a entrevista sociolinguística direcionada pretende envolver afetivamente, durante a entrevista, os participantes, no intuito de que, ao falar de suas experiências, diminua o monitoramento da fala, resultando em produções mais espontâneas e naturais. Há, portanto, uma redução dos efeitos do paradoxo do observador, já que as perguntas funcionam como um gatilho e direcionam a produção de sequências textuais que resultam num todo heterogêneo.

4 Resultados parciais

O *corpus* de estudo, ainda em fase de compilação, totaliza até o momento 5.348 tokens (itens/palavras corridas) e 1.112 types (formas/palavras diferentes), com uma densidade vocabular (type/token ratio) de 20,79%, conforme podemos verificar na figura 1.

Figura 1 – Tokens X Types

Word list completa.lst						
File Edit View Compute Settings Windows Help						
N	text file	file size	tokens (running words) in	tokens used for word list	types si (distinct) of	type/token ratio (TTR)
1	Overall	33.207	5.583	5.348	1.112	20,79
2	ENT01C- 16-10-21.txt	4.191	665	589	244	41,43
3	ENT02C-16-10-21.txt	2.863	471	450	224	49,78
4	ENT03C- 17-10-21.txt	2.643	410	389	205	52,70
5	ENT04C - 17-10-21.txt	4.457	727	706	302	42,78
6	ENT05C - 20-10-21.txt	4.980	864	847	339	40,02
7	ENT06C - 23-10-21.txt	3.536	606	591	253	42,81
8	ENT07C - 03-11-21.txt	3.395	614	596	253	42,45
9	ENT08C - 11-11-21.txt	2.787	475	459	211	45,97
10	ENT09C - 30-11-21.txt	1.505	257	242	143	59,09
11	ENT10C - 06-12-21.txt	2.850	494	479	230	48,02

Fonte: ferramenta *WordList*.

O *corpus* em análise está formado por um recorte de dez entrevistas realizadas dentre aquelas que constituem o *corpus* geral das entrevistas que irão compor o *corpus* do estudo final. É constituído das respostas dos participantes, pois editamos os arquivos de transcrição, deixando apenas as respostas e excluindo as transcrições de nossas falas, proferidas enquanto sujeito-entrevistador, durante as interações nas entrevistas.

Os entrevistados foram 12 falantes nativos, que possuíam, no momento de coleta, entre 12 e 74 anos, sendo nove deles do sexo masculino e dois do sexo feminino. Ou seja, foram realizadas dez entrevistas, nas quais houve 12 participantes, sendo que a entrevista 01 teve três participantes simultâneos.

Tabela 1 – Duração das entrevistas sociolinguísticas

Entrevista	Duração
ENT01	12'27"
ENT02	12'13"
ENT03	5'06"
ENT04	9'03"
ENT05	13'54"
ENT06	10'07"
ENT07	11'32"
ENT08	10'13"
ENT09	4'17"
ENT10	14'12"
TOTAL	1h 43'04"

Fonte: elaborada pelos autores a partir dos dados da pesquisa.

A duração das entrevistas é divergente, já que ao ser utilizado o método da entrevista sociolinguística, houve liberdade na expressão dos falantes. Cada entrevista possui uma duração específica e, juntas, totalizam 1h 43'04". Apresentamos, na Tabela 1, os dados referentes a duração de cada uma delas.

A média das palavras que compõe cada entrevista, 534,8, é uma representação do universo de uso da LT, a qual entendemos ser suficiente para garantir que nelas estejam as características da língua que tentamos descrever. Já que, de acordo com Blanche-Benveniste *et al.* (1990), um texto deve ao mesmo tempo poder representar determinadas propriedades sintáticas e determinadas estruturas dialógicas.

Assim, para dar início à análise, foi gerada uma lista de palavras por meio da ferramenta *WordList*. A partir da análise dessa lista, organizada por ordem decrescente de ocorrência, foi possível identificar as palavras lexicais da LT mais frequentes no *corpus* de estudo. Focalizaremos nossas análises preliminares nas duas mais frequentes: *cuete* (68 ocorrências) e *conjolo* (59 ocorrências). Para ilustrar essa primeira aproximação ao *corpus*, apresentamos na Figura 2 um recorte da lista de palavras, em que é possível identificar os itens *cuete* (homem) e *conjolo* (casa) nas linhas 14 e 15, respectivamente.

Após identificar os dois itens lexicais mais frequentes, utilizamos a ferramenta Concord para gerar listas de concordância para esses itens lexicais. Analisando as linhas de concordância de *cuete* e *conjolo*, foi possível notar que uma parcela considerável dos itens lexicais utilizados durante as entrevistas pertence à LT, ao passo que as palavras gramaticais (tais como conjunções, artigos, preposições) são pertencentes exclusivamente à variedade da LPB falada na cidade de Bom Despacho. À guisa de ilustração, apresentamos, nas Figuras 3 e 4, excertos extraídos do *corpus* de estudo.

Realizando uma leitura vertical, permitida pelas duas listas de concordância criadas, observamos como a utilização de substantivos próprios ocorre com baixa frequência na variedade da LT. Foram identificados apenas treze substantivos próprios, sendo seis deles referentes a nomes de pessoas: "Fiota", "Haroldo", "Célio Luquini", "Sebastiana" "Thiago", "Tassinho". Outros quatro para indicação de localidades da cidade de Bom Despacho; "Irmã Maria", "São Vicente", "Quenta Sol", "Santa Marta"; dois nomes de marcas: "Netflix"

(plataforma de streaming) e “Centro- Oeste Cap” (título de capitalização). E o último, nome de um vírus, “Corona”.

Figura 2 – Recorte da lista de palavras

N	Word	Freq.	%	Texts	%
1	#	235	4,21	10	100,00
2	E	163	2,92	10	100,00
3	EU	153	2,74	10	100,00
4	QUE	153	2,74	10	100,00
5	DE	129	2,31	10	100,00
6	O	99	1,77	10	100,00
7	MAIS	98	1,76	10	100,00
8	NÃO	87	1,56	10	100,00
9	É	86	1,54	10	100,00
10	A	81	1,45	10	100,00
11	NUM	78	1,40	10	100,00
12	NO	76	1,36	10	100,00
13	TEIM	70	1,25	10	100,00
14	CUETE	68	1,22	9	90,00
15	CONJOLO	59	1,06	10	100,00
16	AQUI	56	1,00	10	100,00
17	DO	51	0,91	10	100,00
18	NA	49	0,88	10	100,00
19	RISOS	49	0,88	8	80,00
20	CUM	45	0,81	9	90,00
21	UM	45	0,81	9	90,00
22	ÁÍ	42	0,75	9	90,00
23	MINHA	40	0,72	9	90,00
24	NOIS	40	0,72	8	80,00
25	TÁ	40	0,72	8	80,00

Fonte: Ferramenta WordList.

Figura 3 – Recorte das linhas de concordância com a palavra *cuete*

N	Concordance	Word #	Ser	Ser	Par	Par	Hez	Hez	Sec	Sec	File
1	cuete ? P2M45= tipura af cuete ... oh u cuete ocora .	173	5	3	0	2			0	2	ENT01C- 2022/j
2	.. eu tenho quatro camonim cuete ... famia avura	10	0	2	0	1			0	1	ENT04C - 2022/j
3	= minha famia é catita ... cuete aqui teim ... : :	7	0	1	0	1			0	1	ENT01C- 2022/j
4	= foi aqui memu ... cuns cuete aqui memu ... eu	265	5	4	0	5			0	5	ENT03C- 2022/j
5	de cuete cafuvira ... de cuete sem ingura ... aí	355	1	4	0	3			0	3	ENT05C - 2022/j
6	vergonha de sê língua de cuete cafuvira ... de cuete	351	1	2	0	3			0	3	ENT05C - 2022/j
7	.. P1M54= cambada de cuete ... fica queto ((risos).	68	3	6	0	1			0	1	ENT01C- 2022/j
8	quando tá uma rodinha de cuete amigo ... aí nois faiz	156	7	6	0	2			0	2	ENT08C - 2022/j
9	= cê tipura a ocainha do cuete ... avura ? P3M52=	195	6	8	0	2			0	2	ENT01C- 2022/j
10	.. P2M45= ingura avura do cuete avura caixa no (...)	351	2	6	0	5			0	5	ENT01C- 2022/j
11	né? ... cê é cumba do cuete ...	989	2	1	0	1			0	1	ENT05C - 2022/j
12	... P5M64= pertu dus cuete que num tipura a	217	5	4	0	4			0	4	ENT03C- 2022/j
13	e ligô para tudo que era cuete do conjoalo ao lado	368	6	9	0	6			0	6	ENT10C - 2022/j
14	cumeço a gritá que ele era cuete -ocaia e aí ele choro	585	1	7	0	8			0	8	ENT07C - 2022/j
15	tranquila ... agora esses cuete que caixa marcanjo .	623	1	5	0	9			0	9	ENT06C - 2022/j

Fonte: Ferramenta WordList.

Figura 4 – Recorte das linhas de concordância com a palavra *conjolo*

N	Concordance	Word #	%
1	= Eu curimbó construino conjolo ... P7M39= uai ...	123	12%
2	de uruma cum o cuete do conjolo ao lado ... P10M16	136	25%
3	... e tava longe ainda do conjolo da omana ...	281	52%
4	tiro ... mais aí o dono do conjolo de matuaba ispaiô	209	72%
5	baraio com os cuete do conjolo do ladu P1M54=	100	14%
6	né? P2M45= oh cuete do conjolo de matuaba...teim	549	79%
7	tudo que era cuete do conjolo ao lado do meu ...	370	67%
8	mais cum as janela do conjolo aberta ... igual era	662	95%
9	... cum us as ocaia do conjolo ao lado ... a famia .	351	50%
10	... impurô pra parede do conjolo de matuaba e	601	61%
11	s'oque nois tava perto do conjolo dos viriango e aí	290	54%
12	o telefone tocô ... e era do conjolo de covera ... ês	795	80%
13	setembro ... perto du ... du conjolo du cravinguero	104	20%
14	fiz um puxado aqui no meu conjolo ...alí na frente tá	211	30%
15	uns vinti ano P5M64= Meu conjolo fica longe do	160	31%
16	longe do curimbo ... meu conjolo é aquele lá oh...1	167	33%
17	mora aqui mesmo no meu conjolo a mãe deles	34	7%
18	memu ... foi lá no meu conjolo cunversá cum a	415	76%
19	vez inquando vem pro meu conjolo ...vem as ocaia	120	25%
20	cuete rumo aqui no meu conjolo esse parelhinho e	133	19%
21	época num tinha muito conjolo de cortá orangê	98	14%
22	1 25 ano () só aqui nesse conjolo onde eu vendo a	106	22%
23	nada P4F71= aqui nesse conjolo ? () 1 25 ano () só	96	21%
24	famia é avura ... nois lá no conjolo é treis cuete e	9	1%
25	os cuetinho ... só caxa no conjolo de tipurá P12M40=	26	5%

Fonte: Ferramenta *WordList*.

Partindo desta constatação, notamos que, em poucos momentos durante as entrevistas realizadas, o sujeito falante fez uso de nomes próprios para se referir ao outro (3ª pessoa do discurso), mesmo quando esse era personagem de sua narrativa. Em substituição aos nomes próprios, foram utilizados os substantivos simples da variedade da LT, como *cuete*, *ocaia* e outros, com comum recorrência à perífrase, que serve, no contexto analisado, para nominalizar e caracterizar o ser a quem o locutor se refere, facilitando a identificação do referente. A seguir, apresentamos alguns excertos extraídos das linhas de concordância que evidenciam o uso da perífrase.

P12M40= uai teim muita coisa ((risos)) ... uai pera aí xô vê aqui ... ah : : ... teve uma vez qui tava oteque e eu fui curiá no conjolo na minha omana e ... a minha ocaia tava cum camoninho no jequê ... aí nois resolveu pegá uma camba ... só que nu meio du camonho minha ocaia istorô a bolsa e nois teve que descê dipressa memu ... alí memu onde nois tava ... e tava longe ainda do conjolo da omana ... s'oque nois tava perto do conjolo dos viriango e aí nois cabô de chega lá de tinhame memu e pediu ês para levá minha ocaia pro **conjolo de covera**¹ ... ês pegô o uruma de viriango e levô nois ... o pobrema foi qui um cuete meu viu eu no uruma do viriango e chegô no conjolo da minha omana e falô pra ela que eu tinha é sido preso ((risos)) ... ela ficô ruim

¹ *Conjolo de covera*= 3ª pessoa do discurso, o ser do qual se fala. Perífrase utilizada para se referir a “hospital”.

do tué e ligô para tudo que era **cuete do conjolo ao lado**² do meu ... sô que ninguém tipurava nada né? : : ... aí só depois que meu camoninho nasceu que eu fui ligá pra contá ... porque eu fiquei mei nervoso ... aí ela me xingô até manda pará ((risos)) ... disse que eu matava ela do coração ... esses treim ... mais hoje em dia ela ri da histora tameim ((risos))... (ENT10- P12M40- Extraído do *corpus* de estudo- grifo nosso)

P11M71= um dia ... eu tava no **conjolo de matuaba**³ ... um viriango buscô um muque de undaro pra me dá um tiro ... mais aí o dono do conjolo de matuaba ispaiô cum ele ... falô cum ele que ia chamá o irmão dele que era viriango avura e ele foi bora ... achu que o cuete num era bão do tué não. ENT09- P11M71- Extraído do *corpus* de estudo- grifo nosso)

P7M39= momentu feliz teim muito ... mais o qui a genti mais lembra mesmu é dos momento triste ... pode sê um triste? Intão ... achu que foi no início desse ano que minha mãe pegô esse Corona ... esse treim esquisito ... e ficô internada no conjolo de covera sozinha ... ninguém pudia visita ... eu ia lá todo dia depois do curimbo pra sabê nutiça ... mais os **cuete de covera**⁴ num explicava nada direitu ... falava só qui ela tava no balão e tinha que isperá ... aí um dia de noite ... o telefone tocô ... e era do **conjolo de covera** ... ês tava falanu que minha mãe foi pro isquife eterno ... cumba do céu ... eu chorei que nem camoninho ... gosto neim de lembra muito ... ês dexô eu vê minha mãe só atrais de um vidro ... e nem dexô leva pro conjolo do granjão pra rezá ... passô direto pro buraco do cimetero ... nem na funerara num foi ... disse que podia contaminá os outro ... só colocô num saco e jogô no buraco ... nossinhora ... foi ruim demais ... credu ... quer fala mais não ...

Ao longo das análises, percebemos que seria relevante realizar uma limpeza na lista de palavras inicial, gerando, portanto, uma segunda lista que possibilitaria identificar a quantidade exata de palavras da LT presentes no *corpus* de estudo. Esse procedimento nos forneceria um panorama probabilístico do percentual de palavras da LT, que os membros da comunidade analisada utilizaram durante as entrevistas. Portanto, a limpeza manual, a partir da utilização da ferramenta *Word List*, consistiu na exclusão de todos os vocábulos pertencentes ao português brasileiro local da lista inicial, resultando em uma lista de itens lexicais formada apenas pelo léxico da LT.

A lista obtida após o procedimento mencionado mostrou, com relação à frequência, que os vocábulos na LT estão em número de 476, em um universo de 5.348 *tokens*, ou seja, aproximadamente 8,9% dos itens utilizados pelos participantes da LT pertencem à LT. Ainda

² *Cuete do conjolo ao lado*= 3ª pessoa do discurso, o ser de quem se fala. Perífrase utilizada para se referir a “vizinho”.

³ *Conjolo de matuaba*= 3ª pessoa do discurso, o ser do qual se fala. Perífrase utilizada para se referir a “bar”.

⁴ *Cuete de covera*= 3ª pessoa do discurso, o ser de quem se fala. Perífrase utilizada para se referir a “médico”.

que se trate de um recorte do *corpus*, cuja extensão é pequena, haja vista a características do processo de coleta, cabe ressaltar o potencial presente tanto na ocorrência quanto na recorrência dos itens lexicais, uma vez que possibilitarão seu estudo e descrição em maior amplitude. Ademais, é preciso notar que 113 itens individuais utilizados durante as entrevistas pertencem à LT. A partir do confronto entre o número de *types* (1.112), em que constam todas as formas individuais, sem contar as repetições, utilizadas nas entrevistas, e o número de vocábulos identificados como pertencentes apenas à LT (113), observamos que as do léxico da Tabatinga possui uma ocorrência, dentro da realidade linguística observada, de 10,16% do total de *types*.

Na Figura 5, apresentamos um recorte da segunda lista de palavras, gerada apenas com itens pertencentes à LT.

Figura 5 – Lista de palavras do léxico da LT

N	Word	Freq.	%	Texts	%
1	CUETE	68	1,22	9	90,0
2	CONJOLO	59	1,06	10	100,0
3	OCAIA	37	0,66	8	80,0
4	AVURA	32	0,57	9	90,0
5	OCORA	23	0,41	5	50,0
6	MATUABA	21	0,38	6	60,0
7	CUMBARA	16	0,29	6	60,0
8	INGURA	16	0,29	7	70,0
9	TIPURÁ	15	0,27	5	50,0
10	CAXÁ	14	0,25	5	50,0
11	CURIMBO	14	0,25	9	90,0
12	TIPURA	14	0,25	7	70,0
13	CATITA	13	0,23	5	50,0
14	CURIMBA	13	0,23	8	80,0
15	SENGUE	12	0,21	7	70,0
16	CAMONINHO	11	0,20	6	60,0
17	ORANGÊ	11	0,20	2	20,0
18	CAMONIM	10	0,18	4	40,0
19	TINHAME	10	0,18	7	70,0
20	URUMA	10	0,18	5	50,0
21	CUMBA	9	0,16	4	40,0
22	TIPURANO	9	0,16	5	50,0
23	VIRIANGO	9	0,16	4	40,0
24	CAXA	8	0,14	5	50,0
25	MARCANJO	8	0,14	4	40,0
26	COVERA	7	0,13	4	40,0
27	TUÉ	7	0,13	6	60,0

Fonte: *WordList*.

Tal constatação apresenta a relevância das palavras pertencentes à LT no discurso dos indivíduos da comunidade linguística analisada, visto que mais de 10% das palavras por eles selecionadas para a realização de suas comunicações orais é decorrente do léxico da Tabatinga. Ademais, esses dados justificam a relevância da pesquisa, em virtude dos processos de variação em curso, se considerada a pesquisa anterior e sua inserção atual no contexto da comunidade.

É relevante ainda mencionar que, após uma análise detalhada dessa lista de palavras, observamos que, das 113 palavras, 74 pertencem à classe gramatical dos substantivos. Ao longo da análise de linhas de concordância geradas para outros itens lexicais da LT, outros pontos também nos atraíram a atenção, como a polissemia que acompanha os verbos. Com o intuito de exemplificarmos tal fenômeno linguístico, escolhemos trabalhar com os verbos *tipurar* e *caxar* e algumas de suas acepções em seus contextos de ocorrência.

A seguir, apresentamos alguns exemplos extraídos das linhas de concordância obtidas a partir dos dois verbos mais frequentes na LT, na extensão do *corpus* analisada até o presente momento: *tipurar* e *caxar* após lematização, bem como seus significados na LPB.

- P2M45= *todo cuete catito tem ocaia avura...tipurô? ((risos))*

No exemplo acima, o verbo *tipurar* possui o valor semântico de “compreender”, “entender”.

- P9M12= *eu num quiria tipurá ... mais minha mãe num dexa eu pará de tipurá ... o resto quiria mudá não ...*

Nas linhas acima, o verbo *tipurar* adquire no contexto de uso o sentido de “estudar”.

- P7M39= *qualqué hora nois tipura na gira ... mais teim que sê cum quem sabi tipurá tameim ... senão num teim graça ficá tipurano sozinho ...*

Nas três ocorrências do verbo *tipurar* do excerto acima (*tipura*, *tipurá*, *tipurano*) notamos que o sentido construído com o uso do verbo é “falar”.

Com o verbo *caxar* foi possível identificar a constituição recorrente de unidades fraseológicas. Veja:

- P1M54= *eu gosto é de caxá orufino*

No excerto supramencionado, o verbo *caxar*, em conjunto com o substantivo *orufino* consitui um fraseologismo com o valor semântico de “pescar”.

- *cuete ... cumé que vai? vamú caxá matuaba ”? P11M71= com ês*

Já na ocorrência acima, percebemos a formação do fraseologismo *caxá matuaba*, “beber/ingerir” qualquer bebida alcoólica.

- ((risos)... P2M45= *num pode é caxá camonim...né? se caxá trapaia ...*

Mais uma vez, comprova-se a formação de uma unidade fraseológica a partir do uso do verbo *caxar*, *caxar camonim*, tendo neste contexto evidenciado, o valor semântico de “engravidar”.

A partir da observação dessa formação de unidades fraseológicas com o verbo *caxar*, identificamos como necessária uma análise mais detalhada da ocorrência de fraseologismos na LT, bem como suas funções na língua.

Compreendida a língua como um sistema gramatical pertencente a um grupo de indivíduos, intrinsecamente ligado à estrutura social e, por isso mesmo, extremamente maleável e diversificada, reforçamos o pressuposto teórico de que os fraseologismos refletem a cultura da qual são forma e produto. Dessa maneira, quando os fraseologismos aparecem nas gravações orais da LT, são culturalmente marcados, antes de tudo, pela escolha, consciente ou não, de unidades lexicais que permitem chegar a um referente extralinguístico. Assim, as unidades fraseológicas representam, na LT, traços culturais extralinguísticos. Por essa razão, passamos a inventariar os (3) substantivos, (3) verbos e (3) adjetivos mais frequentes na LT e os fraseologismos com eles produzidos.

Os substantivos que entram na construção de um fraseologismo na LT ocorrem, predominantemente, em seu sentido primeiro, usual. Uma explicação possível a esse fato é a de que um falante da comunidade Tabatinga recorre com facilidade a palavras designando realidades de seu ambiente imediato para formar fraseologismos, facilitando assim sua interpretação, tratando-se de um ambiente compartilhado. A realidade extralinguística está, nesses exemplos, numa relação bastante estreita com o código linguístico. Veja:

- *e ligô para tudo que era **cuete do conjolo ao lado do meu**= vizinho (sentido usual)*
- *cumeçô a gritá que ele era **cuete ocaia** e aí ele chorô mais= homossexual/gay (sentido usual)*
- *um tiro ... mais aí o dono do **conjolo de matuaba** ispaiô cum...= bar (sentido usual)*

- *o meu irmão taméim foi pro conjolo do pé junto ...* = morreu (sentido figurado ativado pela relação entre os vocábulos *conjolo de pé junto* = cemitério= local onde ficam os mortos).

Quanto aos três verbos analisados é necessário evidenciar que eles servem estritamente para exprimir ações, sendo constante a produção de fraseologismos que são atualizados em situação de discurso, já que os verbos na LT se apresentam com sentidos muito amplos. Veja:

- *ele gosta de tipurá ocaia* ((risos)) = paquerar/flertar
- *a genti sempre ouvia os ocora tipurá na língua ... minha mãe= falar*
- *os cuetinho ... só caxa no conjolo de tipurá* P12M= estudar
- *vontade de fazê é pará de caxá o marcanjo ...* = fumar
- *eu gosto é de caxá orufino*= pescar
- *num pode é caxá camonim...* = engravidar
- *...já caxô muita matuaba no tué ...* = beber qualquer bebida alcoólica
- *caxo matuaba ...caxo inbanjeco ... esses treim* = tocar instrumento musical
- *o cue/o cuete vai tê que caxá o cureio e ...cê sabe né.* = comprar o alimento
- *hoje cedu tava caxano omenha ... tava tipequerano* = chovendo

Logo, o que se observa é que os verbos formam com seus complementos fraseologismos, enfatizando o maior grau de culturalidade do fraseologismo em comparação com uma construção linguística livre qualquer.

Quanto aos fraseologismos que possuem adjetivos, é possível notar que eles são poucos e se formam com base em apenas três adjetivos: *avura*, *ocora* e *catito*.

Tais unidades fraseológicas são formadas a partir de aspectos avaliativos positivos e negativos, bem como de construções em sentido figurado. Observe:

- *todo cuete catito tem ocaia avura...tipur* = homem feio/mulher bonita (sentido usual).
- *o marcanjo ... esses marcanjo avura sabe?* = baseado (sentido figurado pelo uso do adjetivo *avura*, que pode representar aquilo que é bom, bonito, grande, forte, valioso associado ao coletivo cultural de que “o cigarro de maconha é algo mais “forte”, em comparação com cigarro de tabaco tradicional”.

Durante as análises, foi observada a necessidade de considerar o apagamento do /-r/ (apócope) nas formas verbais infinitivas da LT. A seguir, alguns exemplos extraídos do *corpus* de estudo.

- 1) cassucará= cassucarar= casar
- 2) caxá orufino= caxar orufino= pescar
- 3) caxá o marcanjo= caxar o marcanjo= fumar
- 4) caxá matuaba=caxar matuaba= beber bebida alcoólica
- 5) caxá o cureio= caxar o cureio= comprar o alimento
- 6) caxá camonim= caxar camonim= engravidar
- 7) curiá= curiar= almoçar
- 8) curimbá= curimbar= trabalhar
- 9) injirá pepita= injirar pepita= jogar bola
- 10) tipurá= tipurar= estudar, entender, saber, ouvir, compreender

Segundo Callou e Lopes (2003), no português do Brasil, o cancelamento do /-r/ em final de verbos no infinitivo é um fenômeno que foi, inicialmente, associado à estratificação social e racial. Chegou a ser considerado uma característica dos chamados falares incultos, sendo usado nas peças de Gil Vicente, no século XVI, para simbolizar o linguajar dos escravos, já que a apócope do -r era observada na estrutura fonético-fonológica das línguas africanas, sendo utilizada para estigmatização de um grupo étnico.

Atualmente, é uma tendência observada na quase totalidade das variedades da LPB brasileira falada, devido à dinamicidade, mutabilidade e a heterogeneidade que a língua possui. Não nos esquecendo de que, com o uso massivo das redes sociais, a língua portuguesa escrita passa pela influência da fala e já se pode observar o apagamento do /-r/ final das formas verbais em textos escritos da LPB.

Na LT, não poderia ser diferente. Seguindo o caráter natural e multifacetado das línguas, a partir dos dados reunidos, foi possível observar a apócope do /-r/ final nas formas verbais da LT, que a nosso ver pode ser explicada por duas razões: herança direta das línguas do grupo bantu (fato similar ao que ocorreu com o português brasileiro), e/ou influência da variedade regional do português brasileiro, na qual também se presentifica esse fenômeno.

Se observarmos com acuidade os sintagmas nominais em uso na LT, nas ocorrências a seguir, teremos em vista que a tendência é a marcação de /-s/ apenas no determinante dos sintagmas nominais. Tal fato também ocorre em variantes linguísticas de uso oral do cotidiano do português brasileiro e se opõe à forma padrão do português brasileiro que, segundo Tarallo (1990), é marcado redundantemente ao longo do sintagma nominal: no determinante, no nome (núcleo) e nos modificadores.

- 1) *foi aqui memu ... cunS cueteØ aqui memu ... eu curimba*
- 2) *quando tá uma rodinha de cueteØ amigoØ ... aí noiS faiz*
- 3) *pertu duS cueteØ que num tipura*
- 4) *... agora esseS cueteØ que caxa marcanjo ...*
- 5) *noiS cunhece todo mundo ... oS cueteØ ... aS ocaiaØ ...*

A seguir, traçamos algumas considerações relacionadas aos resultados aqui expostos.

5 Considerações parciais

O acesso a dados linguísticos da LT, altamente sistematizados e computadorizados, assim como às ferramentas computacionais e estatísticas disponíveis para seu tratamento, tornam as hipóteses sobre a língua passíveis de testagem efetiva e redefinição, com altos níveis de representatividade e confiabilidade.

Sendo assim, a iniciativa apresentada, por meio desta breve análise e descrição da LT, concedeu-nos bases para que pudéssemos identificar algumas das características inerentes à língua. Isso evidencia que a continuidade de nossa pesquisa se faz necessária, com vistas ao reconhecimento e à valorização linguística e cultural da LT, que se desenvolveu em solo nacional, em decorrência do encontro entre a língua portuguesa e línguas de origem bantu. Este último ponto, como um dos resultados que conseguimos alcançar, denota a relevância da continuidade dos trabalhos nessa linha de raciocínio.

Ademais, como resultado de pesquisa, destacamos a construção em pleno andamento de uma página de internet, cujo objetivo é a preservação e divulgação da LT, como patrimônio histórico imaterial da cidade de Bom Despacho, fornecendo bases para o estreitamento da relação entre língua, sociedade e cultura.

Referências

ALVES-MAZZOTTI, Alda J.; GEWANDSNAJDER, Fernando. **O método nas ciências naturais e sociais**: pesquisa quantitativa e qualitativa. 2. ed. São Paulo: Pioneira, 1999.

BLANCHE-BENVENISTE, Claire. *et al.* **Le Français parlé**: études grammaticales. Paris: Éditions du C.N.R.S., 1990.

CALLOU, Dinah; LOPES, Célia. Contribuições da Sociolinguística para o ensino e a pesquisa: a questão da variação e mudança linguística. **Revista do Gelne**, Fortaleza, ano 5, n. 1-2, p. 63-74, 2003.

COELHO, Izete Lehmkuhl *et al.* **Sociolinguística**. Florianópolis: LLV/CCE/UFSC, 2010.

KRIEGER, Márcia da Graça. Lexicografia: o léxico no dicionário. *In*: SEABRA, Maria Cândida Trindade (org.). **O léxico em estudo**. Belo Horizonte: Faculdade Letras da UFMG, 2006.

LABOV, William. **Padrões Sociolinguísticos**. Trad. Marcos Bagno. Maria Marta Pereira Scherre, Carolina Rodrigues Cardoso. São Paulo: Parábola Editorial, 2008.

RIBEIRO, Darcy. **O povo brasileiro**: a formação e o sentido do Brasil. Curitiba: Companhia das Letras, 1995.

SCOTT, Mike. **WordSmith Tools**, versão 06. Oxford: Oxford UP, 2012.

TARALLO, Fernando. **A pesquisa sociolinguística**. São Paulo: Ática, 1990.

Recebido em: 18.12.2022

Aprovado em: 10.04.2023