



Universidade Federal da Paraíba  
Centro de Ciências Humanas, Letras e Artes  
Departamento de Letras Estrangeiras Modernas  
Bacharelado em Tradução

Caio Cesar Martino

A elipse do sujeito pronominal “eu” em um corpus comparável de Monteiro Lobato

João Pessoa – PB  
Março de 2014



Universidade Federal da Paraíba  
Centro de Ciências Humanas, Letras e Artes  
Departamento de Letras Estrangeiras Modernas  
Bacharelado em Tradução

Autor: Caio Cesar Martino  
Orientador: Prof. Daniel Antônio de Sousa Alves

A elipse do sujeito pronominal “eu” em um corpus comparável de Monteiro Lobato

Trabalho realizado e apresentado como requisito para a conclusão do curso de Bacharelado em Tradução da Universidade Federal da Paraíba, sob orientação do Prof. MSc. Daniel Antônio de Sousa Alves, durante o período de 2013.2, para obtenção do título de Bacharel em Tradução.

João Pessoa – PB  
Março de 2014

A elipse do sujeito pronominal “eu” em um corpus comparável de Monteiro Lobato

Autor: Caio Cesar Martino

Banca Examinadora – TCC

---

Prof. MSc. Daniel Antônio de Sousa Alves  
ORIENTADOR

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Luciane Leipnitz  
EXAMINADORA

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Camila Nathália de Oliveira Braga  
EXAMINADORA

---

Prof. Dr. Roberto Carlos de Assis  
SUPLENTE

Trabalho de Conclusão do Curso de Bacharelado em Tradução da UFPB, período letivo 2013.2, defendido e aprovado no dia 19 de março de 2014, com nota final 9,5 (nove e meio).

João Pessoa – PB  
Março de 2014

## **AGRADECIMENTOS**

Ao professor Daniel Antônio de Sousa Alves, que aceitou orientar este trabalho e o fez com muita organização e responsabilidade;

A todos os professores do curso de Bacharelado em Tradução que me ensinaram, orientaram e contribuíram para meu crescimento como aluno, tradutor e cidadão.

Aos meus colegas, com quem e de quem aprendi muitas coisas, que me mostraram em muitas oportunidades que sempre há algo novo a se conhecer.

Em especial ao Prof. Roberto Carlos de Assis, que em agosto de 2009 assumiu um compromisso com a educação superior do Brasil, comprometendo-se ao mesmo tempo com um novo emprego, uma nova casa, um curso totalmente novo, e 50 alunos em quatro disciplinas.

## RESUMO

Esta pesquisa se insere na disciplina acadêmica dos Estudos da Tradução e na vertente teórico-metodológica da Linguística de Corpus e tem o objetivo de analisar quantitativamente a elipse do pronome “eu” – em casos os quais a gramática tradicional o denomina pronome pessoal do caso reto e o considera sujeito de um verbo –, em contraste ao seu uso explícito. A pesquisa utiliza um corpus comparável composto por dois textos em língua portuguesa: *Caçadas de Pedrinho* (1933), do escritor brasileiro Monteiro Lobato, e *Aventuras de Huck* (1934), que é tradução de Monteiro Lobato para o português da obra *The Adventures of Huckleberry Finn* (1884), do escritor estadunidense Mark Twain. Em virtude da opção por trabalhar com um corpus comparável, este trabalho restringe-se aos textos em língua portuguesa e não analisa o texto-fonte em língua inglesa, à luz da teoria de Baker (1998). A pesquisa é de natureza quantitativa e seu desenho metodológico é baseado em Maia (1998). Os dados em formato eletrônico foram processados com o concordanceador *AntConc*. Considerando as ocorrências de uma seleção de dois verbos marcados para 1ª pessoa do singular nas duas obras (total de quatro casos), em todos os quatro casos os resultados obtidos com este corpus apontam superioridade numérica da elipse do pronome pessoal “eu” na função de sujeito em comparação ao seu uso explícito. Entre as ocorrências de dois verbos não marcados nas duas obras (total de quatro casos), em somente um caso houve superioridade quantitativa da elipse.

Palavras-chave: Linguística de corpus. Corpus comparável. Elipse pronominal. Monteiro Lobato.

## ABSTRACT

This research is set in the academic discipline of Translation Studies and – concerning its theoretical-methodological trait – in Corpus Linguistics. This paper aims to conduct a quantitative analysis of the explicit use of the Portuguese pronoun *eu* – in cases traditional grammarians consider it the subject of a verb – in contrast with its ellipsis. A comparable corpus was used comprising two texts in Portuguese language: *Caçadas de Pedrinho* (1933), by the Brazilian writer Monteiro Lobato, and *Aventuras de Huck* (1934), Monteiro Lobato's translation into Portuguese of *The Adventures of Huckleberry Finn* (1884), by the American writer Mark Twain. Because the corpus used is comparable, by relying on the theory of Baker (1998), this work analyzes both texts in Portuguese and does not look up the source text in English. This research is quantitative and its methodology is based on the work of Maia (1998). Electronic data were processed through the freeware Concordancer *AntConc*. Considering the occurrences of a selection of two verbs marked for the 1<sup>st</sup> person singular in both texts (total of four cases), in all four cases the results obtained from this corpus show numerical superiority of the ellipsis of the personal pronoun *eu* when used as a subject when compared with its explicit use. Among the occurrences of two unmarked verbs in both texts (total of four cases), only in one case was the ellipsis quantitatively superior.

Key words: Corpus Linguistics, Comparable Corpus. Pronoun Ellipsis. Monteiro Lobato.

## LISTA DE ILUSTRAÇÕES

### FIGURAS

<i>Figura 1. Aba Word List do concordanceador AntConc: lista de palavras em T1.....</i>	<i>20</i>
<i>Figura 2. AntConc: linhas de concordância para o nóculo "como" em T1.....</i>	<i>22</i>
<i>Figura 3. Planilha do BrOffice Calc: levantamento dos dados do nóculo “quero” em T1. .....</i>	<i>28</i>
<i>Figura 4. Website Corpus do Português: resultado da pesquisa dos colocados “pernas” e “quero”.....</i>	<i>30</i>
<i>Figura 5. Website Corpus do Português: resultados da pesquisa dos colocados “não sei” e “quê”.....</i>	<i>31</i>
<i>Figura 6. Aba file view do AntConc: ocorrência 165 de “era” em T2.....</i>	<i>32</i>

### QUADROS

<i>Quadro 1. Ocorrências de verbos não marcados em T1 e T2.....</i>	<i>24</i>
<i>Quadro 2. Ocorrência de verbos marcados em T1 e T2.....</i>	<i>26</i>
<i>Quadro 3. Sujeito “eu” explícito e elipse pronominal — verbos marcados. ....</i>	<i>35</i>
<i>Quadro 4. Sujeito “eu” explícito e elipse pronominal — verbos não marcados.....</i>	<i>36</i>

### TABELAS

<i>Tabela 1. Levantamento das ocorrências de verbos com homônimos em T1.....</i>	<i>23</i>
<i>Tabela 2. Seleção final da amostra de verbos não marcados para análise.....</i>	<i>25</i>
<i>Tabela 3. Seleção final da amostra de verbos marcados para análise.....</i>	<i>27</i>

## SUMÁRIO

INTRODUÇÃO .....	8
1. FUNDAMENTAÇÃO TEÓRICA .....	13
2. DESCRIÇÃO DO CORPUS.....	17
3. METODOLOGIA .....	19
3.1. SELEÇÃO DA AMOSTRA.....	21
3.2. LEVANTAMENTO DOS DADOS.....	27
3.2.1. VERBOS MARCADOS .....	27
3.2.2. VERBOS NÃO MARCADOS.....	31
4. RESULTADOS E ANÁLISE DOS DADOS .....	34
4.1. VERBOS MARCADOS .....	34
4.2. VERBOS NÃO MARCADOS.....	36
5. CONSIDERAÇÕES FINAIS.....	39
REFERÊNCIAS BIBLIOGRÁFICAS.....	41



## INTRODUÇÃO

Este trabalho tem o objetivo de analisar o uso do pronome “eu” em um corpus comparável composto por duas obras literárias escritas por Monteiro Lobato. Esta análise se concentra nas orações em que o pronome “eu” – tradicionalmente chamado pronome pessoal do caso reto – desempenha a função de sujeito. Parte-se da identificação no corpus dos usos explícitos do pronome “eu” e das ocorrências de sua elipse. A relevância do estudo se encontra no fato de que as gramáticas da língua portuguesa mencionam (e algumas delas defendem) a possibilidade de elipse do pronome na função de sujeito, uma vez que este pode ser recuperado pela desinência pessoa-número presente no verbo conjugado. Portanto, este estudo visa conhecer dentro de um corpus comparável de um mesmo autor as ocorrências da elipse de um pronome em contraste ao seu uso explícito.

Esta pesquisa se insere na disciplina acadêmica dos Estudos da Tradução e na vertente teórico-metodológica da Linguística de Corpus. De maneira simplificada, Kenny<sup>1</sup> (2001, p. 23) citada por Bueno (2005, p. 19) define Linguística de Corpus como o “ramo da Linguística que estuda a linguagem a partir de corpora”. Sanchez<sup>2</sup> (1995), por sua vez, citado por Sardinha (2000, p. 338) apresenta uma definição para corpus – termo chave para o estabelecimento da matéria – que reúne muitos aspectos relevantes acerca do objeto e também do campo de estudo da Linguística de Corpus, a saber: a origem dos dados, o propósito, a composição, a formatação, a representatividade, e a extensão do corpus:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo

---

<sup>1</sup> KENNY, D. **Lexis and creativity in translation: a corpus-based study**. Manchester: St Jerome Publishing, 2001.

<sup>2</sup> SANCHEZ, A. Definición e historia de los corpus. In: \_\_\_\_\_ et al (org.). CUMBRE – Corpus Linguístico de Espanol Contemporaneo. Madrid: SGEL, 1995.

que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise<sup>3</sup>. (SARDINHA, 2000, p. 338)

O corpus investigado neste trabalho é do tipo comparável, composto por duas obras em língua portuguesa do mesmo autor, Monteiro Lobato. Uma das obras é diretamente produzida em língua portuguesa, *Caçadas de Pedrinho* (1933), e a outra é uma tradução para este mesmo idioma, *Aventuras de Huck* (1934), a partir de um texto-fonte em inglês — *The Adventures of Huckleberry Finn* (1884) — do escritor estadunidense Mark Twain.

Para o fim ao qual se prestam, as duas obras, objetos desta investigação, encontram-se em formato digital com extensão *.txt* e passíveis de processamento pelo concordanceador *AntConc*. Este programa permite – a partir de um corpus –, entre outras coisas, produzir uma lista de palavras em função de suas frequências, bem como listar as ocorrências de palavras ou frases de interesse nos contextos em que são utilizados. A escolha por esse *software* dá-se pelo fato de que suas funcionalidades cumprem o que está proposto em nosso objetivo de pesquisa, além de que a referida versão é disponibilizada gratuitamente para *download* e utilização<sup>4</sup>.

Nossa pesquisa tem o intuito de averiguar as ocorrências do uso explícito do pronome pessoal “eu”, usado como sujeito, e da sua elipse nas duas obras que compõem o corpus, e é baseada na metodologia de pesquisa apresentada em Maia (1998). No referido trabalho, a autora analisa a frequência e a natureza da estrutura do tipo SVO em inglês e em português, através dos dados coletados a partir de um corpus composto por uma obra produzida em língua portuguesa e outra em língua inglesa, e suas respectivas traduções

---

<sup>3</sup> Tradução para: “un conjunto de datos lingüísticos reales (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad, de manera que sean representativos de la totalidad del uso lingüístico o de alguno de sus ámbitos, y dispuestos de tal modo que puedan ser procesados por ordenador, con la finalidad de propiciar resultados diversos y útiles para la descripción y el análisis”.

<sup>4</sup> Disponível em <http://www.antlab.sci.waseda.ac.jp/software.html>. Acesso em 21 de agosto de 2013

para o outro idioma. Além de quantificar os casos de sujeito explícito e elipse nos textos em língua portuguesa (como texto-fonte e texto-alvo), o trabalho de Maia leva em conta o caráter pragmático, ao analisar as razões pelas quais o “eu” (sujeito) explícito ou elíptico é utilizado em cada caso.

Nosso trabalho está centrado no viés quantitativo e restringe-se a dois textos em língua portuguesa, sendo uma tradução e uma obra espontaneamente produzida. Em virtude da problemática envolvida no modo como nos referimos a um texto que se encontra no idioma em que foi primeiramente escrito, ou seja, que não é uma tradução, e tampouco um texto-fonte, adotamos o termo “espontaneamente produzido”<sup>5</sup>. Com isso, evita-se o uso do qualificador “original” em oposição a “traduzido”, uma vez que o primeiro poderia induzir à interpretação de que o texto traduzido não é original no sentido de faltar-lhe criatividade, inovação e/ou novidade.

A escolha de um corpus comparável para esta pesquisa se justifica pela necessidade, apontada por Baker (1995), citada por Olohan (2004), de se explorar como um texto produzido de modo relativamente livre de um roteiro em outra língua (ou seja, de um texto-fonte) difere de um texto produzido sob as condições normais que permeiam a tradução<sup>6</sup>. Baker (1998) defende ainda a avaliação de traduções através do uso de corpus comparável como uma alternativa ao cotejamento de textos-alvo com seus textos-fonte. Segundo Baker (1995, p. 2), a “metodologia empregada aqui não consiste em comparar os textos-fontes com suas traduções; ao contrário, consiste em comparar os textos originais e

---

<sup>5</sup> Nossa sugestão de tradução para *spontaneously produced*, de Zanettin (2000, p. 106), citado por Olohan (2004, p. 36). Zanettin, F. *Parallel Corpora in Translation Studies: Issues in Corpus Design*. In: Olohan, M. (ed.). *Intercultural Faultiness. Research Methods in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerone. 2000.

<sup>6</sup> Tradução nossa para “a shift in the focus of theoretical research in the discipline (...). We need to explore how text produced in relative freedom from an individual script in another language differs from text produced under the normal conditions which pertain in translation (...).”

traduções *dentro de um mesmo idioma e dentro de domínios próximos*”<sup>7</sup> (grifo no original). Acerca dos domínios próximos mencionados por Baker, as duas obras que compõem o corpus deste trabalho possuem traços de similaridade, como o fato de que foram publicadas em anos consecutivos (*Caçadas de Pedrinho* é de 1933 e *Aventuras de Huck* é de 1934) e de que os protagonistas, nos dois casos, são crianças do sexo masculino de idades aproximadas (Pedrinho com 10 anos e Huck com 13). Mais considerações acerca do corpus serão feitas na seção 2, Descrição do Corpus.

Face ao exposto, as perguntas que norteiam esta pesquisa são:

- Nos dois textos de Monteiro Lobato que compõem o corpus, prevalece o uso explícito do pronome pessoal “eu” na função de sujeito ou a sua elipse?
- Qual a diferença quantitativa na razão entre as ocorrências de uso explícito do pronome “eu” na função de sujeito e a sua elipse na comparação das duas obras?

Além deste capítulo introdutório que, principalmente, define Linguística de Corpus, justifica o uso de corpus comparável nesta pesquisa, introduz a metodologia de Maia e elenca as perguntas de pesquisa, este trabalho é composto de 5 seções, a saber:

1. Fundamentação teórica, que expõe as bases da gramática da língua portuguesa e inglesa para a elipse do pronome na função de sujeito, relata a visão de Baker acerca da normalização do texto traduzido, apresenta o conceito de tradução de Monteiro Lobato, e demonstra o desenho metodológico de Maia;
2. Descrição do corpus, que elenca características das publicações e enredos dos dois textos analisados;

---

<sup>7</sup> Tradução nossa para “La méthodologie employée ici ne consiste pas à comparer des textes sources à leurs traductions, mais plutôt à comparer des textes originaux et des traductions *dans une même langue* et dans des domaines apparentés”.

3. Metodologia, que descreve os fundamentos e procedimentos da coleta de dados a partir do corpus;
4. Resultados e Análise dos Dados, que apresenta numericamente as descobertas desta pesquisa e propõe uma discussão acerca dos índices encontrados;
5. Considerações Finais, que apresenta conclusões adicionais acerca dos resultados numéricos, responde às perguntas de pesquisa elencadas nesta Introdução, e indica possibilidades futuras de pesquisas com este tema.

## 1. FUNDAMENTAÇÃO TEÓRICA

Neste estudo utiliza-se um corpus comparável em língua portuguesa, composto de um texto espontaneamente produzido – conforme definido no capítulo introdutório deste trabalho – em português e uma tradução a partir de um texto-fonte em inglês. Conforme exposto na seção de Introdução, fundamentados em Baker (1998), não analisaremos dados referentes ao texto-fonte. No entanto, por uma questão comparativa, iniciaremos este capítulo com breves observações sobre o idioma inglês.

A sintaxe da língua inglesa é descrita como SVO. Isto significa dizer que a “ordem normal de uma oração afirmativa é a de sujeito + verbo + objeto”<sup>8</sup> (MAIA, 1998, p. 1), ou seja, que um verbo deve vir antecedido por um elemento que cumpra a função gramatical de sujeito. Um dos elementos que podem cumprir esta função é o pronome. Said Ali (1964, p. 61) define pronome como “a palavra que denota o ente ou a ele se refere, considerando-o (...) como pessoa do discurso”, entre as quais está a pessoa que fala, em oposição àquela de quem se fala e àquela com quem se fala. No padrão culto da língua inglesa, o imperativo apresenta-se como uma quebra da ordem SVO, em que o sujeito (*you*) pode ser elíptico sem impedimento de compreensão do comando imposto a uma segunda pessoa, tanto no singular como no plural (e.g.: *Shut the door*).

Assim como ocorre em língua inglesa, a estrutura do português é SVO. Diferentemente do inglês, no entanto, a língua portuguesa é flexível quanto à elipse do pronome e a norma prevê casos como o do sujeito indeterminado (e.g.: Comeram o bolo antes da festa) e dos verbos impessoais (e.g.: Choveu bastante esta tarde), além do imperativo (e.g.: Vá/Vão sem mim). O uso do sujeito é especialmente facultativo nos casos

---

<sup>8</sup> Tradução nossa para “the normal order of an affirmative sentence is that of subject + verb + object”.

mais tradicionalmente denominados de sujeito oculto (ou sujeito desinencial), o qual nos interessa investigar.

Perini (2006) classifica como redundante a informação acerca do agente de uma ação, como em “Eu comprei um carro novo”, uma vez que esta informação provém de duas fontes: o sujeito *eu* e a inflexão do verbo, em particular o sufixo de pessoa-número. Para tal autor, a opção pela elipse do sujeito mantém a carga semântica da oração e elimina a redundância.

Lima (2006, p. 323) reitera a defesa à elipse do sujeito ao dizer que “Por serem explícitas nossas desinências verbais, é comum a elipse do sujeito pronominal”. Neves (2000) aponta que podem ser excluídos os termos de valência nominal e acrescenta que a forma mais comum de elipse é a do elemento que corresponde ao sujeito. Tarallo et al. (2002, p. 27) afirmam que

A identificação e colocação da categoria subentendida apresenta menos problemas do que seria de se esperar uma vez que, quando o verbo está presente, os elementos ausentes podem ser identificados ou pela flexão do verbo, ou pela grande temática e estrutura de subcategorização, ou ainda por um antecedente na sentença ou fora dela, em posições estruturais que lhes permitam ser identificados como tais.

Na pesquisa de Maia (1998), o texto traduzido para o português apresenta maior frequência do sujeito explícito “eu” em relação ao texto espontaneamente produzido em português. Baker (1998) considera que esta influência do texto-fonte no texto traduzido é possível, uma vez que o segundo fica limitado ao primeiro (o que não acontece quando um texto é produzido espontaneamente). No entanto, a autora também concorda que os tradutores tendem a normalizar o texto de chegada, com exagero das características típicas da língua alvo:

Dois traços específicos sugerem que a tradução constitui provavelmente um comportamento linguístico e cultural único (...). O primeiro traço, é que um texto/enunciado traduzido é normalmente restringido por um texto (ou enunciado) inteiramente articulado em uma outra língua. (...) O segundo traço é

que os tradutores tendem, conscientemente ou não, a possuir a percepção do estatuto social do texto que produzem. (...). A tomada de consciência deste contexto especial de recepção permite melhor compreender porque os textos traduzidos tendem a se conformar às características típicas da língua alvo e mesmo a exagerá-las. (BAKER, 1998, p. 480)<sup>9</sup>

Após abordar a visão de Maia e Baker, vejamos alguns aspectos acerca do conceito de tradução de Monteiro Lobato, o escritor em questão.

De acordo com Mendes (2002), Lobato traduziu setenta e sete obras para o português, a maioria delas escritas em inglês, além de ter feito outras traduções de obras francesas, alemãs, italianas e dinamarquesas, algumas delas indiretamente a partir do inglês. Monteiro Lobato possuía uma opinião particular sobre as traduções. Para ele, estas deveriam ser “simples, fluentes, de modo a tornar o acesso do leitor à obra o mais fácil possível” (CAMPOS & OLIVEIRA, 2009, p. 4). Campos e Oliveira (2009, p. 4) defendem que essa “preferência pela fluência se faz mais presente nos textos infantis traduzidos para o português”, que é o caso de *Aventuras de Huck*, parte do corpus deste estudo, a despeito de toda a discussão em torno da crítica social contida na obra de Twain, a ser discutida na Seção 2 deste trabalho, Descrição do Corpus. Lobato cria que simplificar a linguagem dava ao tradutor a liberdade para melhorar o texto-fonte, onde este julgasse desejável (LOBATO, 1955). Algo que é pertinente ao nosso estudo é o conselho dado por Lobato para que o tradutor não se amarre “ao original em matéria de forma – só em matéria de fundo” (*ibid.*, p. 232); “o tradutor necessita compreender a fundo a obra do autor, e reescrevê-la em português, como quem ouve uma história e depois a conta com palavras

---

<sup>9</sup> Tradução nossa para “Deux traits spécifiques suggèrent que la traduction constitue vraisemblablement un comportement linguistique et culturel unique (...). Le premier trait, c'est qu'un texte/énoncé traduit est normalement contraint par un texte (ou énoncé) entièrement articulé dans une autre langue. (...) Le second trait est que les traducteurs ont tendance, consciemment ou non, à tenir compte de la perception du statut social du texte qu'ils produisent. (...) La prise de conscience de ce contexte spécial de réception permet de mieux comprendre pourquoi les textes traduits ont tendance à se conformer aux caractéristiques typiques de la langue cible et même à les exagérer”.



suas” (*ibid.*, p. 127). Estas considerações de Lobato dão indícios das razões pelas quais sua obra/tradução *Aventuras de Huck* tem tamanho reduzido em comparação ao texto-fonte de Twain: este possui mais de 113 mil *tokens*<sup>10</sup> enquanto a tradução possui pouco mais de 86 mil; uma redução de aproximadamente 24%. Do ponto de vista sintático, é possível, portanto, esperar que Lobato mantenha tanto no texto-alvo quanto no texto espontaneamente produzido a tendência ao sujeito desinencial, que é traço comum à língua portuguesa, conforme defendido pelos gramáticos citados no início deste capítulo.

---

<sup>10</sup> Em Linguística de Corpus, todas as palavras contidas em um texto/corpus, ou seja, cada uma das ocorrências destas palavras, são chamadas *tokens*. *Type* é como se convencionou chamar cada uma das palavras de um texto computada uma só vez, independente de quantas ocorrências/repetições haja neste texto. E.g.: em *to be or not to be*, contam-se 6 *tokens* e 4 *types*, uma vez que “*to*” e “*be*” são utilizadas duas vezes.

## 2. DESCRIÇÃO DO CORPUS

Esta seção discorre sobre dados e fatos acerca das publicações e enredos dos textos que compõem o corpus deste trabalho.

José Bento Renato Monteiro Lobato nasceu em 1882 e faleceu em 1948 e ficou conhecido por valorizar o folclore brasileiro e dar ênfase à vida no campo, na zona rural. Este é o cenário para *Caçadas de Pedrinho*, livro voltado ao público infantil, publicado em 1933. A história é contada em terceira pessoa, por um narrador onisciente, e é dividida em 12 capítulos com títulos. O enredo se desenvolve a partir de uma caçada a uma onça vista nos arredores do Sítio do Pica-Pau Amarelo, onde vivem os personagens, entre eles Pedrinho, de 10 anos de idade. Os outros animais da floresta, temendo serem as próximas vítimas, resolvem reagir à caça da onça e segue-se uma batalha. É então que um rinoceronte foge do circo e ao ser caçado por funcionários do governo, é ajudado por Emília, que o leva para viver no Sítio.

A obra foi alvo de acusação pelo Conselho Nacional de Educação (CNE) em 2011 por ser avaliada como racista, em decorrência dos comentários acerca de personagens negras, entre elas Tia Nastácia, cozinheira do Sítio. Exemplo disso é a passagem "Tia Nastácia, esquecida dos seus numerosos reumatismos, trepou, que nem uma macaca de carvão" (LOBATO, 1988. p. 45). Julgado inapropriado para crianças, seu uso foi proibido nas escolas públicas. Posteriormente, essa decisão foi revista e, em vez de censurar a obra, o colegiado do CNE indicou que as edições posteriores devem exibir uma nota técnica contextualizando o texto ao momento histórico em que foi produzido<sup>11</sup>.

---

<sup>11</sup> Parecer CNE/CEB nº 6/2011 disponível em [http://portal.mec.gov.br/index.php?option=com\\_content&view=article&id=16368:ceb-2011&catid=323:orgaos-vinculados](http://portal.mec.gov.br/index.php?option=com_content&view=article&id=16368:ceb-2011&catid=323:orgaos-vinculados). Acesso em 09/03/2014.

*Aventuras de Huck* é uma tradução de Monteiro Lobato publicada em 1934. O texto-fonte, *Adventures of Huckleberry Finn*, de 1884, foi publicado em inglês por Mark Twain, *nom de plume* de Samuel Longhorn Clemens, considerado um dos maiores escritores norte-americanos, que viveu entre 1835 e 1910.

A história é dividida em 43 capítulos apenas numerados. Huck, com 13 anos de idade, é o narrador-protagonista da história e tem como companheiro de aventuras Jim, um negro escravizado. Os dois fogem juntos, cada qual com suas razões: Jim quer ir para o norte, encontrar a liberdade onde não haja escravidão; Huck quer escapar da civilidade que lhe querem ensinar as irmãs Miss Watson e viúva Douglas, que o acolheram. O garoto também foge de seu pai alcoólatra. Huck e Jim separam-se e reencontram-se várias vezes durante a trama, envolvidos em várias aventuras.

Apesar de considerada por muitos, inclusive por Ernest Hemingway e William Faulkner, uma das principais obras na tradição literária norte-americana, *Adventures of Huckleberry Finn* também enfrenta críticas por utilizar palavras e mostrar comportamentos de personagens que podem ser considerados racistas, entre eles o uso da palavra *nigger*, que é ainda hoje considerada ofensiva e pejorativa quando direcionada a um negro.

### 3. METODOLOGIA

A fim de facilitar a referência, os textos que compõem o corpus deste trabalho, *Caçadas de Pedrinho* e *Aventuras de Huck*, passam a ser chamados T1 e T2, respectivamente. Na fase de levantamento dos dados, os referidos conteúdos encontravam-se previamente armazenados em arquivos em formato *.txt* e foram cedidos pelo orientador deste trabalho, por terem sido utilizados em outra pesquisa no campo da Linguística de Corpus. Procedeu-se com a retirada das etiquetas nos arquivos para maior acuidade na contagem de *tokens* e na ordenação dos *types* com base nas suas frequências de ocorrências.

Para o processamento do corpus, foi feito o *download* da versão 3.2.4w do concordanceador *AntConc* a partir do *website* do seu desenvolvedor. O *AntConc* foi a ferramenta utilizada nesta pesquisa pela sua facilidade de manuseio, a compatibilidade de suas funcionalidades com o objetivo deste trabalho, e o fato do *software* ser compartilhado gratuitamente para uso em pesquisas linguísticas como esta.

Procedemos com o processamento do corpus no *AntConc* objetivando localizar os verbos em primeira pessoa do singular, que se configuram como o ponto de partida para a identificação do sujeito “eu” explícito ou elíptico. Para isso foi gerada uma lista de palavras para T1 – e posteriormente para T2 – com suas frequências e ordenamento de base quantitativa, utilizando as seguintes configurações: na aba *word list* (lista de palavras), o campo *search term* (termo de busca) foi marcado em *words*, para limitar a busca a palavras; foi selecionado o campo *treat all data as lower case*, para que o programa lesse casos de caixa alta e baixa (maiúsculas e minúsculas) como “Fui”, “fui” e “FUI” sem diferenciá-los; o campo *sort by* (critério de seleção) foi ajustado em *sort by freq*, para que a frequência fosse o parâmetro da listagem; e a caixa *invert order* foi desmarcada, para que

as palavras da listagem fossem apresentadas da mais para a menos frequente no corpus. Outros campos e ajustes permanecem de acordo com a configuração *default*. A Figura 1 mostra a tela do programa na aba *word list* após a geração da lista a partir de T1.

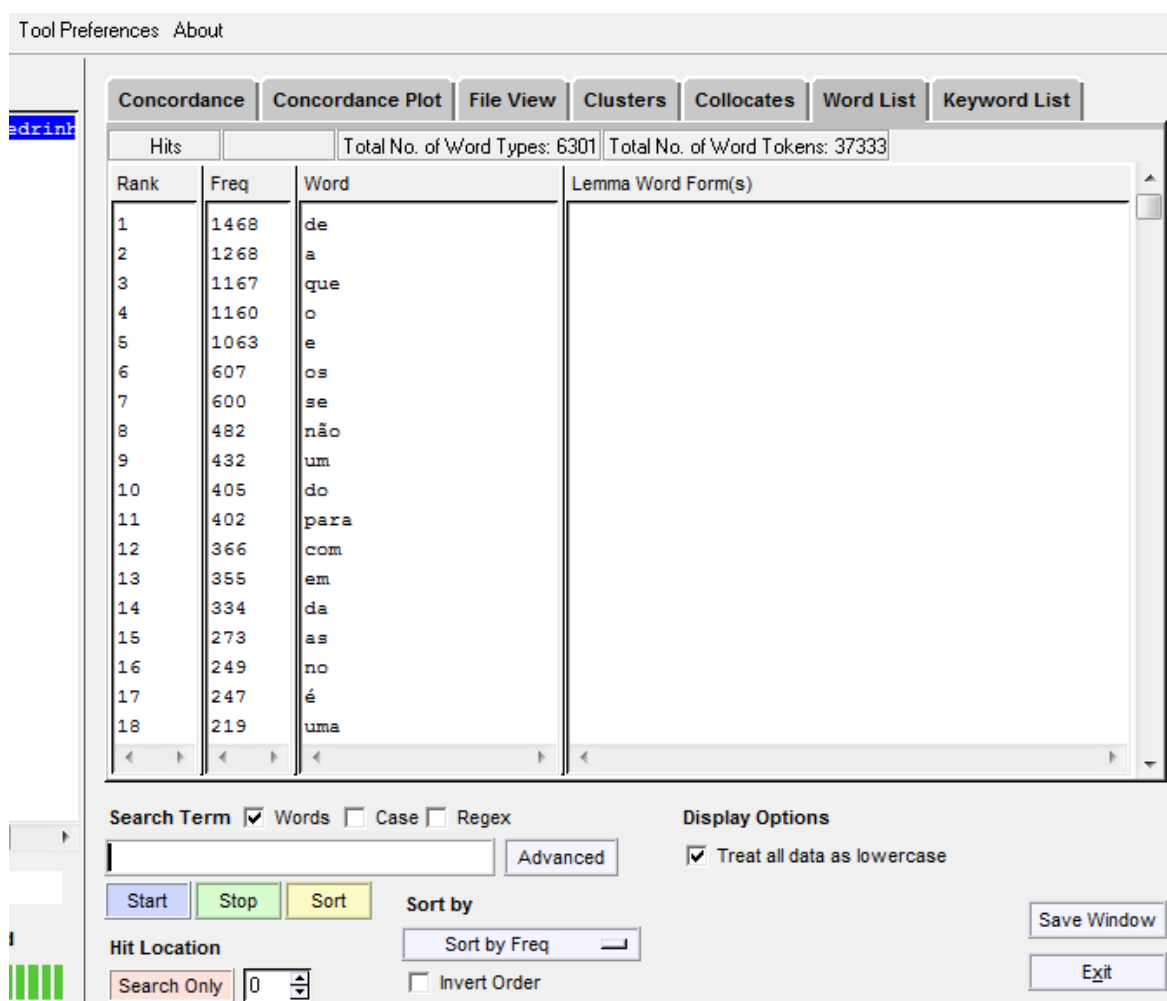


Figura 1. Aba Word List do concordanceador AntConc: lista de palavras em T1

Vê-se na Figura 1 a lista de palavras geradas a partir de T1 na coluna denominada *Word*.

À sua esquerda, na coluna *Freq*, pode-se conhecer o número de vezes que cada *type* aparece na obra. Mais à esquerda, a coluna *Rank* mostra a posição/classificação em ordem

crescente de cada *type* em relação ao total de *types*; quanto menor este número mais vezes o *type* aparece no corpus.

### 3.1. SELEÇÃO DA AMOSTRA

Buscando coerência no estudo, estipulou-se analisar os verbos comuns a T1 e T2 que se encontrassem mais bem posicionados no quesito *rank*. Para fins de delimitação, buscamos reduzir para quatro (4) a quantidade de verbos para análise, a ser subdivididos em dois grupos: dois verbos marcados para 1ª pessoa do singular e dois verbos não marcados. Para encontrar os verbos que satisfizessem a condição mencionada estipulou-se realizar o levantamento dos casos mais frequentes até a ducentésima (200ª) posição (da coluna *Rank* da aba *Word List* do *AntConc*).

O levantamento das palavras em T1 até a posição 200 resultou na seguinte lista de dezesseis possíveis verbos na 1ª pessoa do singular: “como”, “era”, “disse”, “estava”, “pelo”, “havia”, “sobre”, “fosse”, “caso”, “entre”, “fora”, “tinha”, “ia”, “porto”, “quero” e “sei”.

Como a seleção lexical feita pelo *AntConc* não tem base semântica, identificamos neste rol palavras que, sendo classificadas como verbos possuem homônimos com outra função gramatical. É o que ocorre com “como” (conjunção); “era” e “porto” (substantivo); “pelo” (substantivo ou preposição + artigo); “sobre” e “entre” (preposição); “caso” (substantivo); e “fora” (preposição ou substantivo). Em relação a estes casos, fez-se uma verificação manual das suas ocorrências em T1, através da aba “*Concordance*” do *AntConc*, com o objetivo de identificar e quantificar as ocorrências como verbo, que são nosso foco de interesse. A Figura 2 mostra a tela do programa no referido modo de visualização, durante a análise das linhas de concordância do nóculo “como” em T1.

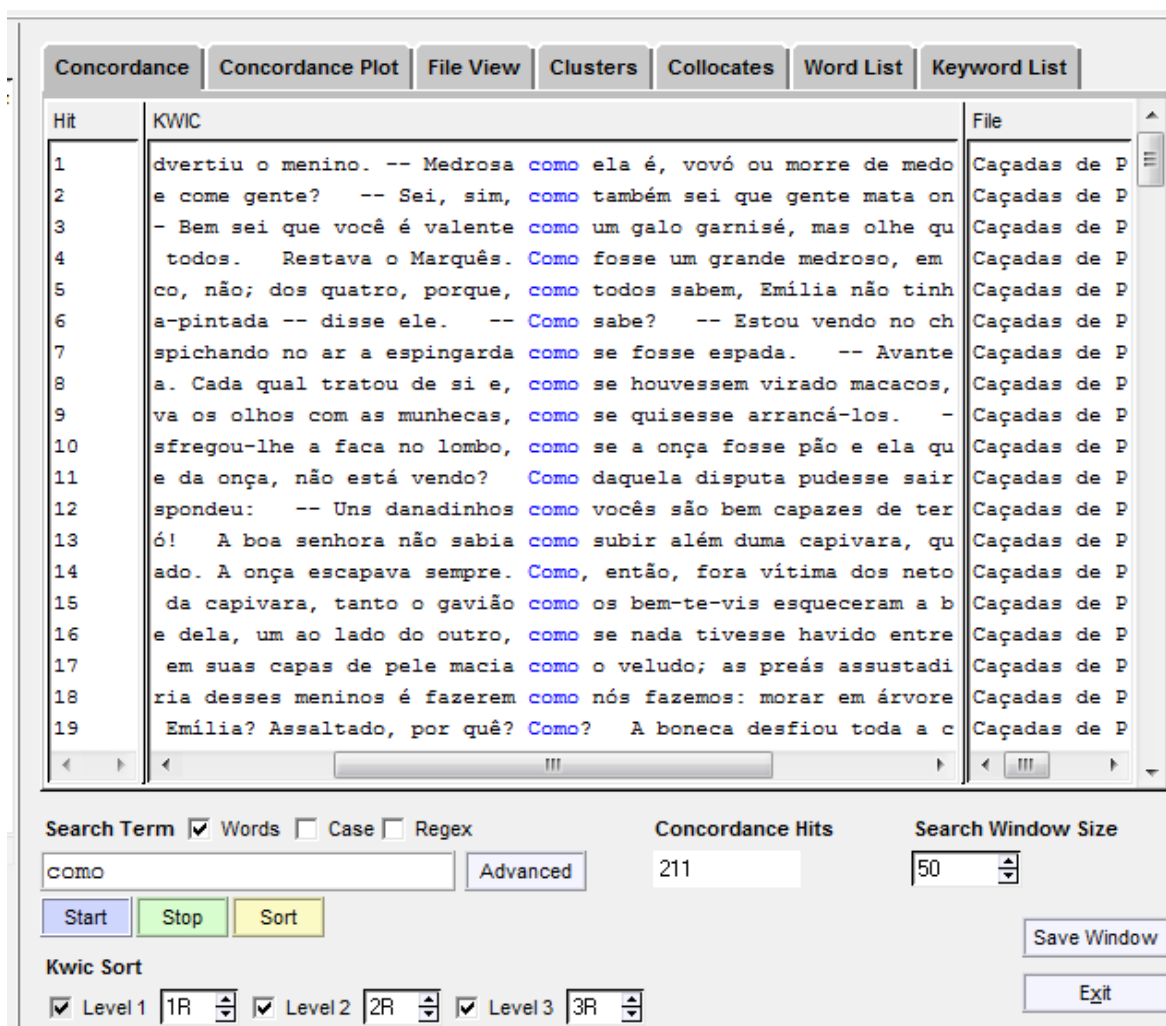


Figura 2. AntConc: linhas de concordância para o nóculo "como" em T1

A Figura 2 mostra a aba *Concordance* do *AntConc* com as linhas de concordância do nóculo selecionado “como” ao centro.

Para as palavras “como”, “pelo”, “sobre”, “caso”, “entre” e “porto”, de um total de 410 ocorrências em T1, nenhuma delas resultou ser um verbo. Deste modo, estes nóculos foram descartados.

Entre as 36 ocorrências da palavra “fora”, 19 são como verbo. No entanto, uma segunda análise mostra que todas elas são de 3ª pessoa. Como fica evidenciado que tais ocorrências não contribuem para nossa pesquisa, o nóculo “fora” também foi descartado da amostra.

Já a palavra “era”, que ocorre 168 vezes em T1 é utilizada em 100% dos casos como verbo, e por isso foi mantida na amostra.

Estes dados encontram-se sumarizados na Tabela 1.

Tabela 1. Ocorrências de verbos com homônimos em T1

T1: Caçadas de Pedrinho		
nóculo	Quantidade de ocorrências	Quantidade de ocorrências como verbo
como	211	0
era	168	168
pelo	59	0
sobre	42	0
caso	37	0
entre	36	0
fora	36	19
porto	25	0

Após os descartes, a lista de T1 até a 200ª posição resume-se a 9 verbos: “era”, “disse”, “estava”, “havia”, “fosse”, “tinha”, “ia”, “quero” e “sei”.

A partir de T2 o levantamento das palavras resultou na seguinte lista de dezenove possíveis verbos na 1ª pessoa do singular: “como”, “era”, “estava”, “disse”, “pelo”, “havia”, “sobre”, “rio”, “vi”, “fora”, “ia”, “fosse”, “tinha”, “caso”, “seria”, “sei”, “estou”, “entre” e “fui”.

Uma vez que fora estipulado analisar as ocorrências dos verbos coincidentes em T1 e T2, os nós que não resultaram ser verbos em T1, mas que se encontram em T2 até a 200ª posição (“como”, “pelo”, “sobre”, “caso” e “entre”) foram descartados da análise em T2. “Fora” que não resultou ser relevante para a análise em T1 também foi descartado em T2. O nó “porto” encontra-se na posição 1.610 de T2 (além da 200ª posição), com somente 6 ocorrências e também foi descartado.



Após os descartes, a lista resume-se a 12 verbos: “era”, “estava”, “disse”, “havia”, “vi”, “ia”, “fosse”, “tinha”, “seria”, “sei”, “estou” e “fui”, e um possível verbo, “rio”.

Após proceder com o levantamento ordenatório, a seleção e o descarte dos nódulos em T1 e em T2 até a 200ª posição, seguiu-se uma separação dos resultados finais entre formas marcadas para 1ª pessoa do singular e formas não marcadas. As formas marcadas são aquelas em cujo caso o uso explícito do pronome “eu” não é fator determinante para que se apreenda o sujeito do verbo – e.g.: “Só sei que ao voltarmos demos com a tia Sally de pé...” (T2, *sei*, ocorrência 56), “...vamos até o brejo que quero lhe mostrar um ninho de cobra d’água...” (T1, *quero*, ocorrência 16). As formas não marcadas, por sua vez, necessitam que o sujeito seja explicitado (ou a presença de um indicador semântico-pragmático) a fim de dirimir a ambiguidade de pessoa, – e.g.: “Não importa. Já disse que está nos livros e basta...” (T2, *disse*, ocorrência 3), “Não parece natural, mas garanto que era assim.” (T2, *era*, ocorrência 146).

No âmbito dos verbos não marcados, até o limite estipulado da 200ª posição, após os descartes, encontram-se 7 verbos em T1 e 8 em T2. Como buscamos 2 verbos não marcados para análise, uma visualização dos casos mais relevantes, com sua ordem de frequência, quantidade de ocorrências e posição no *ranking* de ocorrências, encontra-se no Quadro 1.

Quadro 1. Ocorrências de verbos não marcados em T1 e T2

verbo	T1			T2		
	ordem	Quantidade de ocorrências	rank	ordem	Quantidade de ocorrências	rank
era	1	168	25	1	273	33
disse	2	131	34	3	161	57
estava	3	72	49	2	164	56

Para cada uma das formas verbais não marcadas comuns a T1 e T2, vê-se no Quadro 1 a posição (ordem) dentre a amostra de verbos (de 1 a 3), a quantidade de ocorrências do *type* dentro de cada obra e a posição (*rank*) em relação à totalidade de *types* de cada obra.

O verbo mais frequente nas duas obras é coincidente: “era”. Para a escolha do segundo verbo a ser analisado, entre “disse” (2º mais frequente em T1 e 3º mais frequente em T2) e “estava” (3º mais frequente em T1 e 2º mais frequente em T2), levou-se em consideração que a proporção entre as quantidades de ocorrências de “disse” em T1 e T2 (131 e 161, respectivamente) é maior do que a de “estava” (72 e 164). Deste modo a amostra final dos verbos não marcados constituiu-se de “era” e “disse”. A Tabela 2 ilustra a escolha final da amostra dos verbos não marcados.

*Tabela 2. Seleção final da amostra de verbos não marcados para análise*

	T1	T2	
Verbo	Quantidade de ocorrências	Quantidade de ocorrências	Total
Era	168	273	441
Disse	131	161	292
Total	299	434	733

Da Tabela 2 pode-se constatar que T1 contribuiu com 299 ocorrências para os dois verbos somados, enquanto que T2 contribuiu com 434 ocorrências. O somatório das ocorrências de “era” e “disse” nas duas obras chega a 441 e 292, respectivamente. De qualquer viés que se considere, o número total de ocorrências de verbos não marcados analisadas em T1 e T2 foi de 733.

Após a delimitação da amostra de verbos não marcados, fez-se o mesmo para os verbos marcados para 1ª pessoa do singular. Até a 200ª posição, encontram-se 2 verbos em T1 e 3 em T2, sendo somente um deles (“sei”) coincidente nos dois textos. Uma

visualização destes casos, com sua ordem na lista de frequência, quantidade de ocorrências e posição no *ranking* de ocorrências, encontra-se no Quadro 2.

Quadro 2. Ocorrência de verbos marcados em T1 e T2.

verbo	T1			T2		
	ordem	Quantidade de ocorrências	<i>rank</i>	ordem	Quantidade de ocorrências	<i>rank</i>
quero	1	25	198	4	38	255
sei	2	25	199	2	63	159
estou	3	21	235	3	61	163
vi	x	6	876	1	115	90

Para cada uma das formas verbais marcadas comuns a T1 e T2, vê-se no Quadro 2 a ordem dentro da amostra de verbos (de 1 a 4; a letra x em T1 indica que “vi” não é o verbo mais frequente imediatamente após “estou”), a quantidade de ocorrências dentro de cada obra e a posição (*rank*) em relação à totalidade de *types* de cada obra.

Para a manutenção do critério pré-estipulado, foi necessário reduzir a amostra dos verbos marcados para uma quantidade de dois. Somente a forma verbal “sei” é coincidente nas duas listas até a posição 200 (ambos em 2º lugar, no quesito *ordem*). Para a escolha do segundo verbo desta amostra, nenhuma das combinações (“quero”, “estou” e “vi” em T1 e T2) permitiria respeitar o limite da 200ª posição. Portanto, optamos pelo verbo mais frequente em T1, por ser maior a proporção entre as quantidades de ocorrências de “quero” (que é o verbo marcado mais frequente em T1) em T1 e T2, 25 a 38, do que a mesma diferença para “vi” (que é o verbo marcado mais frequente em T2), 6 a 115. Deste modo a amostra final dos verbos marcados constituiu-se de “sei” e “quero”. Os números da amostra final para análise dos verbos marcados encontram-se na Tabela 3.

Tabela 3. Seleção final da amostra de verbos marcados para análise

	T1	T2	
verbo	Quantidade de ocorrências	Quantidade de ocorrências	Total
quero	25	38	63
sei	25	63	88
Total	50	101	151

Da Tabela 3 pode-se constatar que T1 contribuiu com 50 ocorrências para os dois verbos somados, enquanto que T2 contribuiu com 101 ocorrências. O somatório das ocorrências de “quero” e “sei” nas duas obras chega a 63 e 88, respectivamente. De qualquer viés que se considere, o número total de ocorrências de verbos marcados analisadas em T1 e T2 foi de 151.

À delimitação da amostra seguiu-se a coleta e análise das ocorrências dos 4 verbos selecionados nas duas obras, para a identificação dos casos de elipse ou uso explícito do pronome pessoal “eu” na função de sujeito.

## 3.2. LEVANTAMENTO DOS DADOS

### 3.2.1. VERBOS MARCADOS

Os verbos marcados trazem em sua desinência pessoa-número a marca da pessoa do discurso e não necessitam de um nome, pronome ou outro referente (ou um indicador semântico-pragmático) para que se compreenda o sujeito da oração. Neste estudo, os dois verbos marcados analisados no corpus comparável são “quero” e “sei”.

Para a análise, foram geradas quatro listas de linhas de concordância no *AntConc*, uma para cada verbo em cada uma das duas obras que compõem o corpus. Estas linhas foram analisadas manualmente uma a uma para identificar 1 de 3 casos:

- o uso explícito do pronome “eu” na função de sujeito do verbo analisado;

- elipse do pronome “eu” como sujeito do verbo analisado;
- outro uso do verbo que suscite o descarte da linha de concordância.

Para tal procedimento, primeiramente foi gerada uma lista de palavras no *AntConc*, na aba *Concordance*, usando o nóculo que se buscava analisar. As linhas de concordância foram então copiadas e coladas a uma planilha *BrOffice Calc* (similar ao *Microsoft Excel*) para que lá fossem feitos os devidos cálculos quantitativos.

Ao lado de cada linha do *BrOffice Calc* uma célula foi preenchida com o número 1, correspondente ao valor unitário de cada ocorrência em relação ao todo. A próxima célula à direita foi completada, após análise, com as marcações “expl” (nos casos de pronome sujeito “eu” explícito), “elip” (nos casos de elipse do pronome sujeito “eu”) ou “desc” (para indicar os casos descartados).

Alimentada a coluna que indica os tipos de ocorrência, o resultado quantitativo de cada caso foi obtido através da utilização de fórmulas inerentes ao *BrOffice Calc*, entre elas a soma condicional (*SOMASE*). A visualização desta etapa está disponível na Figura 3.

F11		fx		Σ	=	
A	B	C	D	E	F	
1						
2	T1: Caçadas de Pedrinho – verbos marcados – verbo 01 – quero					
3	1	O menino bateu no peito com arrogância. -- Pois QUERO ver isso! Vou organizar a caçada e juro que hei d	1	elip		
4	2	de perigo e temos de tomar providências. Por isso QUERO convocar uma reunião de todos os animais. Vocês.	1	elip		
5	3	os sobre essas pernas de pau, bem lá no alto – e QUERO ver!... -- E se as onças também subirem pelas p	1	elip		
6	4	orzinho de jardim, dos verdes -- de outra cor não QUERO. Terceiro que... -- Socorro! -- berrou, num tom	1	elip		
7	5	semelhante coisa. Dou tudo quanto me pedir. -- QUERO aquele pito de barro em que você pita -- responde	1	elip		
8	6	rinho e metade do dinheiro é meu. Serve? -- Não QUERO vendê-lo por dinheiro e sim trocá-lo pelo carrinh	1	elip		
9	7	dou desde já. Estou brincando. Não preciso, nem QUERO roda nenhuma. Ajudarei você a vender o rinoceront	1	elip		
10	8	xclamou Narzinho. -- Atravessou-se na porteira e QUERO ver agora quem entra ou sai. Estamos bloqueados.	1	elip		
11	9	ouvir aquilo Dona Benta protestou. -- Então não QUERO! -- disse ela. -- Se esses homens não têm boa pon	1	elip		
12	10	irem quebrar algum vidro das possas vidraças. Não QUERO! E voltando-se para a Cléu, que tinha muito bo	1	elip		
13	11	Dona Benta. -- Não entendo de tais serviços, nem QUERO entender. Aqui estamos nós para prestar aos serho	1	elip		
14	12	prestar aos senhores toda a ajuda possível. O que QUERO é que o quarto antes me livrem desse animalão. Ma	1	elip		
15	13	que não é da sua conta. O governo sabe o que faz, QUERO falar com a dona da casa. Cléu tapou com a mão	1	elip		
16	14	e quem falou. -- Minha senhora -- disse ele --, QUERO apresentar a Vossa Excelência o Senhor Fritz Müll	1	elip		
17	15	s por partes -- disse ela. -- Antes de mais nada, QUERO voar, conhecer o mundo, Adeus! -- Pois vai, meu	1	elip		
18	16	hem?... -- ... seja dono dum rinoceronte. Mas QUERO que prove que o rinoceronte dele é este, está ent	1	elip	ocorrências	
19	17	aqui nem na casa do diabo -- cortestou Emilia. -- QUERO prova de verdade. Alguma marca, algum sinal de na	1	elip	25	
20	18	berg. -- Adeus, meu pai! Não nasci para árvore, QUERO voar, conhecer o mundo, Adeus! -- Pois vai, meu	1	elip	descarte	
21	19	Francisco I, havia dito em certa ocasião: -- Eu QUERO que me mostrem o testamento de Adão que repartiu	1	expl	0	
22	20	... -- Pare um pouco, vovó -- pediu a menina. -- Eu QUERO dar um pulo lá dentro para trazer a Emilia. A coi	1	elip	total analisado	
23	21	aquele Darwin da Viagem ao redor do mundo, que eu QUERO que vocês leiam, minha ciência é igual a zero.	1	expl	25	
24	22	É minha, dizia um; o braço é meu, dizia outro; eu QUERO este pé, exclamava um terceiro. Em seguida ohi	1	expl	% explícito	

Figura 3. Planilha do *BrOffice Calc*: levantamento dos dados do nóculo “quero” em T1.

28			25			
29						

A figura 3 mostra as 25 linhas de concordância do verbo “quero” em T1; à direita de cada linha de concordância encontra-se uma célula preenchida com o número 1, que representa o valor de cada uma das ocorrências em relação ao total de ocorrências; a próxima coluna foi preenchida com “expl”, “elip” ou “desc” para o caso de a linha representar “eu” explícito, elíptico ou um descarte, respectivamente; a coluna seguinte já tem calculadas as quantidades totais de cada caso “expl”, “elip” ou “desc”; e a última coluna mostra a porcentagem de “eu” explícito e elíptico em relação ao total de ocorrências.

Com relação aos descartes feitos nesta etapa, cumpre mencionar os procedimentos utilizados.

Das 38 ocorrências de “quero” em T2, 2 foram descartadas em virtude de o verbo ser utilizado em meio a uma expressão (“... e 'pernas para que te quero!' Na gruta eu disse a Tom que não...” – ocorrência 3; “...a poeira dos calcanhares e pernas para que te quero, disparei correndo mais que um veado em direção...” – ocorrência 27). Para certificar-se do padrão de uso do verbo “quero” nestas expressões, utilizou-se o Corpus do Português (DAVIES, 2008)<sup>12</sup>. A pesquisa baseou-se nos colocados “pernas” e “quero” até um intervalo de 4 palavras à esquerda e à direita. Foram registrados 10 resultados dos quais todos se referem a “pernas para/prá/prá que te quero”, sem a presença de um sujeito pronominal explícito para “quero”. Uma visão da tela do Corpus do Português é disponibilizada na Figura 4.

---

<sup>12</sup> Disponível em <http://www.corpusdoportugues.org/>. Acesso em 20 de agosto de 2013.

The screenshot shows the interface of the 'CORPUS DO PORTUGUÊS' website. At the top, it states '45 MILLION WORDS, 1300s-1900s'. Below this, there are navigation options like 'DISPLAY', 'LIST', 'CHART', 'KWIC', 'COMPAR', 'SEARCH STRING', 'WORD', 'COLLO', 'POS LIST', and 'RANDOM'. The main content area displays search results for the keywords 'pernas' and 'quero'. The results are organized into a table with columns for document ID, source, and context. The context column shows the word 'pernas' or 'quero' highlighted in red within a sentence.

Doc ID	Source	Context
1	19:Fic:Br:Teixeira:Rua	A B C que veio de lá uma chuva de flechas? Eu, <b>pernas</b> para que vos <b>quero!</b> Quase fui parar em Andrad
2	19:Fic:Pt:Figueiredo:Oliveiras	A B C de borra, desde que decretado pescar da poda. * <b>pernas</b> para que vos <b>quero</b> , ao termo de mil an
3	19:Fic:Pt:Ribeiro:Planicie	A B C . Depois da República, o padre enrolou o estojo, <b>pernas</b> para que te <b>quero</b> , fechou-se a igreja, e t
4	18:Azevedo:Filha	A B C Rossio, quando ouvi gritar: " Pega! pega " <b>Pernas</b> pra que te <b>quero!</b> Olho um tilburi que saía! Brr..
6	18:França:Maldita	A B C de mim os seus agentes, e eu, - <b>pernas</b> para que te <b>quero!</b> Venho para a corte, meu pai soube do
7	18:Caminha:Normalista	A B C é para a gente namorar, encher de beijos e - <b>pernas</b> para que te <b>quero!</b> És muito calouro ainda ni
8	18:Menezes:Prosa	A B C tão ferozes que os marroquinos espíaram aquilo de longe e.. <b>pernas</b> prá que vos <b>quero!</b> Eu não di
9	18:Azevedo:Fritzmac	A B C Um soldado toma-me pelo desordeiro e vai prender-me; eu - <b>pernas</b> para que te <b>quero?</b> Embaraf
10	18:Azevedo:Coruja	A B C um diabo de um guarda desconfiou da marosca e eu - <b>Pernas</b> para que te <b>quero!</b> Foi o meu mal!

Figura 4. Website Corpus do Português: resultado da pesquisa dos colocados “pernas” e “quero”.

Das 63 ocorrências de “sei” em T2, 5 foram descartadas. Nestes 5 casos, “sei” é utilizado como no exemplo “... observei que todos traziam os bolsos cheios de *não sei que*, e muitos ocultavam embrulhos sob o paletó...” (T2, *sei*, ocorrência 24); “não sei (o) que/quanto” tomado como uma expressão. O Corpus do Português registra 357 resultados que confirmam que neste contexto não se utiliza sujeito pronominal explícito. Para complementar a verificação visual, procedemos com uma busca eletrônica do lema *eu*, utilizando o comando Localizar (*Ctrl+F* no *BrOffice Calc*), que visou evidenciar as ocorrências do pronome “eu” explícito agindo como sujeito de “sei”. A Figura 5 mostra a tela do Corpus do Português com resultados para a pesquisa com base nos colocados “não sei” e “que” até um intervalo de 4 palavras à esquerda e à direita, e também evidencia a localização do lema “eu”, que em nenhum dos casos surge como sujeito do verbo “sei”.

[?] 56987	SAVE LIST	CHOOSE LIST	-----	CREATE NEW LIST	[?]
C	Uns nos chamam de Geração 70. Outros, de Geração Intermediária, intermediária <b>não sei</b> de <b>quê</b> . O fato é que, depois da Geração de 45, a				
C	. Os críticos desrespeitavam-me muito - " Já vem mais uma mangaba " e <b>não sei</b> o <b>quê</b> . Quer dizer, não eram capazes de perceber que <b>eu</b> tinf				
C	aquela coisa do jogo - não era Flamengo e Vasco, era Palmeiras e <b>não sei</b> o <b>quê</b> -, as pessoas da ação tinham previsto a libertação do embaix;				
C	pra ela dar o conceito - porque na - me falaram aí sobre gramatica <b>não sei</b> o <b>quê</b> - aí <b>eu</b> nesse aspecto aí <b>eu</b> acho - <b>eu</b> sou contra a				
C	dizer até as letras da palavra - " o de - de ônibus e <b>não sei</b> <b>quê</b> " quer dizer a pessoa tinha que dizer a palavra - e sair dizendo				
C	você conhece bem daquela - terra da cebola - sei - lá - chinha <b>não sei</b> <b>quê</b> - sei quem é sei quem é - pronto - vai todo ano à				
C	não essa importância foi para o Rio de Janeiro - mas não voltou ainda <b>não sei</b> o <b>quê</b> você tem que aguardar " - <b>eu</b> tive que aguardar que essa				
C	de corrupções governamentais - de corrupções - da fazenda éh: Angra dos Reis <b>não sei</b> o <b>quê</b> e você fica - sem po / sem poder acreditar - ma				
C	burocracia não só do porto mas também - do governo porque não podia importar <b>não sei</b> o <b>quê</b> as peças ficaram estragadas lá com a maresia				
C	voleibol e fazer porque o senhor passou este castigo <b>eu</b> acho que está errado <b>não sei</b> o <b>quê</b> e coisa " - aí <b>eu</b> <b>eu</b> <b>eu</b> pedi a palavra não é				
C	disse " precisa nada disso " porque agora tem orientador né? um orientador <b>não sei</b> o <b>quê</b> - é um orientador - " é o orientador - devia estar aq				
C	não faça que papai do céu fica triste - desagrada mamãe do céu e <b>não sei</b> o <b>quê</b> " - hoje em dia - " manda pra lavanderia quando tá sujo				
C	de quinhentos cruzeiros - porque <b>eu</b> esqueci o dinheiro ou deixei o dinheiro ou <b>não sei</b> o <b>quê</b> " - aí o pai mandou mil né? - mandou logo mil				
C	colegas por amigos <b>meus</b> dizendo " áh você dá dinheiro demais a <b>seu</b> filho <b>não sei</b> o <b>quê</b> " <b>eu</b> digo " mas <b>eu</b> acho o seguinte que <b>eu</b> dando dinf				
C	vontade de casar mas não sabia então <b>eu</b> dei as melhores informações possíveis - <b>não sei</b> o <b>quê</b> - esse faz questão de se apresentar - <b>eu</b> tinha				
C	não cabe mais - aí começa a solicitar o vestido da mulher da filha <b>não sei</b> o <b>quê</b> porque a roupa não cabe - não é? - pesadona - porque				
C	: que não presta que <b>eu</b> não faço as coisas direito pra eles e <b>não sei</b> o <b>quê</b> cheia de defeito - vige Maria - a barra é pesada viu?				
C	muito - namorado só pra: () - " esses homens não presta <b>não sei</b> o <b>quê</b> : pra fazer barriga na gente " - ((ri)) - é				
C	três pontos - mas outra pessoa não voltou - ele se negou a responder <b>não sei</b> <b>quê</b> e disse que não sabia que não sabia - e tira tira zero -				
C	<b>eu</b> saia com aquela responsabilidade quer dizer pra - comprar - pacote - e <b>não sei</b> <b>quê</b> - e paga - e tira e vira e mexe e confere e procura				

Figura 5. Website Corpus do Português: resultados da pesquisa dos colocados “não sei” e “quê”

Em T1 não houve descartes entre as ocorrências de verbos marcados.

### 3.2.2. VERBOS NÃO MARCADOS

Os verbos não marcados não trazem em sua desinência pessoa-número a marca da pessoa do discurso e necessitam de um nome, pronome ou outro referente (ou um indicador semântico-pragmático) para que se compreenda o sujeito da oração. Neste estudo, os dois verbos não marcados analisados são “era” e “disse”.

Para a análise, foram geradas quatro listas de linhas de concordância no *AntConc*, uma para cada verbo em cada uma das obras que compõem o corpus. Estas linhas foram analisadas manualmente uma a uma para primeiramente separar os casos em que o verbo se refere à primeira pessoa do singular (eu) daqueles que se referem à terceira pessoa do singular (ele/ela).

Nesta primeira etapa fez-se necessário analisar um trecho mais amplo do que somente a linha de concordância, que – por mostrar um trecho reduzido do texto – nem sempre deixa claro a que pessoa se refere o verbo. Para proceder com a devida verificação,



recorreu-se à aba *file view* do *AntConc*, que dá acesso ao texto integral do arquivo apresentando em primeiro plano na tela o ponto da ocorrência selecionada na aba *concordance*. O aspecto da aba *file view* pode ser visualizado na Figura 6.

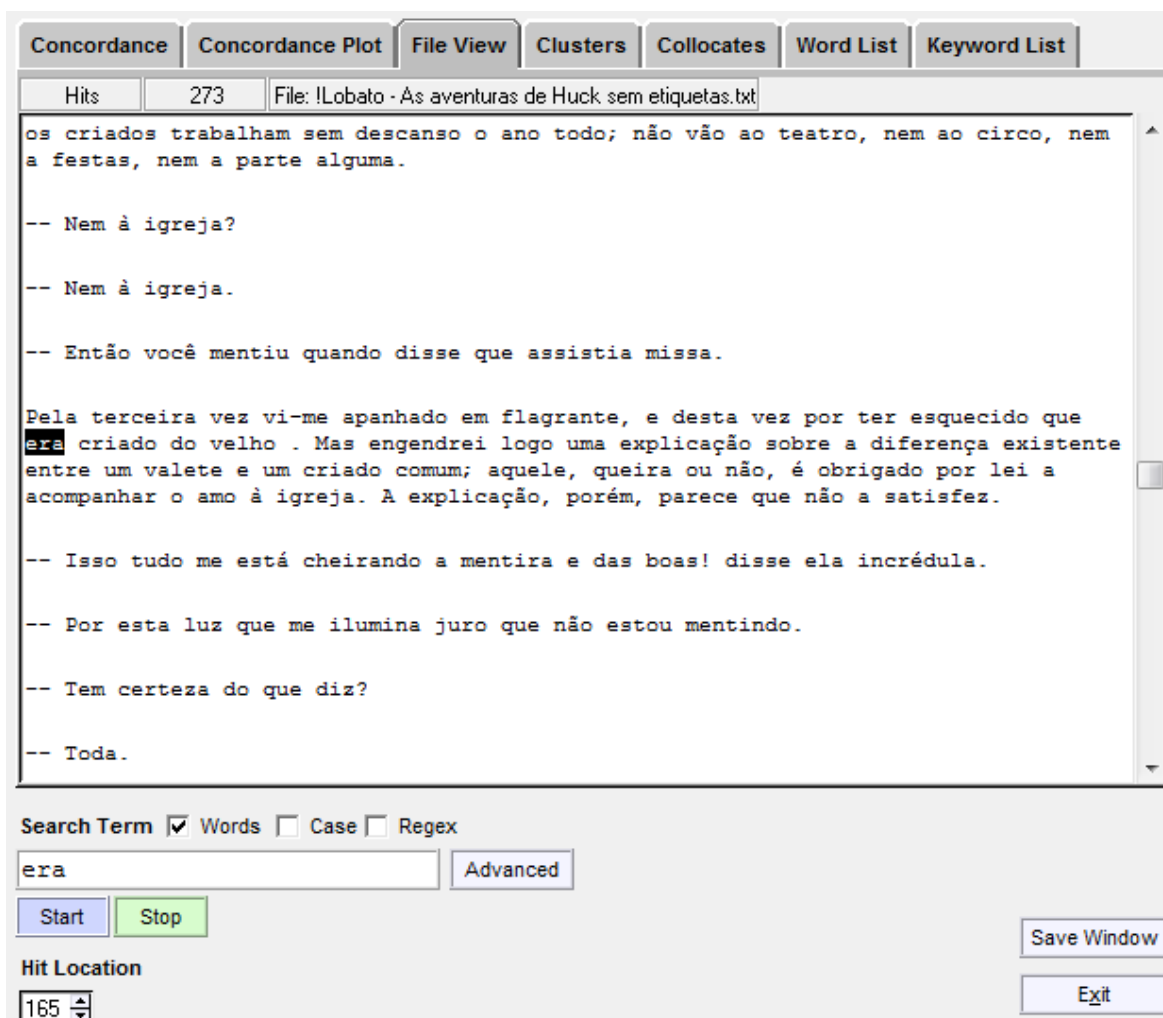


Figura 6. Aba *file view* do *AntConc*: ocorrência 165 de “era” em T2

A figura 6 mostra a ocorrência número 165 do nóduo/verbo “era” (observável no campo *Hit Location*), dentro do corpus utilizado. “Era” encontra-se na linha central evidenciado por um marca-texto escuro. Neste segmento, o contexto aponta para o verbo na 1ª pessoa do singular (“e desta vez por ter esquecido que [eu] era criado do velho”)

Após esta primeira aplicação, dentre os casos reconhecidos como de primeira pessoa do singular, a análise buscou identificar:

- o uso explícito do pronome “eu” na função de sujeito do verbo analisado;
- a elipse do pronome “eu” como sujeito do verbo analisado;
- outro uso do verbo que suscite o descarte da ocorrência.

Para tal procedimento, primeiramente foi gerada uma lista de palavras no *AntConc*, na aba *Concordance*, usando o nóculo que se buscava analisar. As linhas de concordância foram então copiadas e coladas a uma planilha *BrOffice Calc* (similar ao *Microsoft Excel*) para que lá fossem feitos os devidos cálculos quantitativos.

Ao lado de cada linha do *BrOffice Calc* uma célula foi preenchida com o número 1, equivalente ao valor unitário de cada ocorrência em relação ao todo. A próxima célula à direita foi completada, após análise, com as marcações “expl” (nos casos de pronome sujeito “eu” explícito), “elip” (nos casos de elipse do pronome sujeito “eu”) ou “desc” (para indicar os casos descartados).

Para complementar a verificação visual, procedemos com uma busca eletrônica do lema *eu*, utilizando o comando Localizar (*Ctrl+F* no *BrOffice Calc*), que visou evidenciar as ocorrências do pronome “eu” explícito agindo como sujeito.

Alimentada a coluna que indica os tipos de ocorrência, o resultado quantitativo de cada caso foi obtido através da utilização de fórmulas inerentes ao *BrOffice Calc*, entre elas a soma condicional (*SOMASE*).

## 4. RESULTADOS E ANÁLISE DOS DADOS

Conforme apontado na seção 3.1 deste trabalho, os dois textos que compõem o corpus, chamados aqui de T1 (texto produzido espontaneamente em língua portuguesa) e T2 (tradução para o português a partir de um texto-fonte em inglês) foram processados com o concordanceador *AntConc* para o levantamento dos verbos entre as duzentas palavras mais frequentes nos dois textos. Após cruzar os dados e descartar os casos não relevantes (por exemplo, palavras bem classificadas na lista de frequência, mas que se referiam a homônimos de formas verbais com uma função gramatical outra que não fosse um verbo), conforme programado inicialmente, chegou-se a 2 verbos marcados e 2 verbos não marcados para 1ª pessoa do singular, comuns às duas obras. Entre os quatro verbos da amostra final, procedemos com uma nova análise, descrita nas seções 3.2.1 e 3.2.2, desta vez com auxílio da ferramenta *BrOffice Calc* (planilha), de modo a identificar as ocorrências de “eu” explícito, elíptico ou casos passíveis de descarte. Após os devidos descartes, especialmente entre os verbos marcados, chegamos aos números finais deste trabalho, que são dispostos e discutidos na seção a seguir.

### 4.1. VERBOS MARCADOS

Neste estudo, os dois verbos marcados analisados no corpus comparável são “quero” e “sei”. “Quero” resultou 25 ocorrências em T1 e 38 em T2, totalizando 63. “Sei” resultou 25 ocorrências em T1 e 63 em T2, totalizando 88. Deste modo, T1 participou do estudo com um total de 50 ocorrências de verbos marcados e T2 com um total de 101.

Dentre as 25 ocorrências de “quero” em T1, 3 (12%) são de sujeito explícito e 22 (88%) são de sujeito elíptico.

Em T2, após o descarte de 2 casos, conforme descrito na seção 3.2.1, dentre as 36 ocorrências resultantes da amostra, obtém-se 2 casos (5,5%) de sujeito explícito frente a 34 casos (94,5%) de sujeito elíptico.

Quanto ao verbo “sei”, dentre as 25 ocorrências em T1, 4 (16%) são de pronome sujeito “eu” explícito e 21 casos (84%) são elipses.

Em T2, após o descarte de 5 casos, conforme descrito na seção 3.2.1, dentre as 58 ocorrências restantes da amostra, 7 (12%) são de sujeito explícito frente a 51 (88%) casos de sujeito elíptico.

O Quadro 3 resume a informação acerca do número de ocorrências e porcentagens dos dois verbos nas duas obras que compõem o corpus.

*Quadro 3. Sujeito “eu” explícito e elipse pronominal — verbos marcados.*

VERBOS MARCADOS					
		T1		T2	
		ocorrências	% da amostra	ocorrências	% da amostra
QUERO	Explícito	3	12%	2	5,5%
	Elíptico	22	88%	34	94,5%
SEI	Explícito	4	16%	7	12%
	Elíptico	21	84%	51	88%

Com base nestes números, pode-se concluir que:

1. Tanto em T1 como em T2, os dois verbos apresentam superioridade da elipse do pronome “eu” em comparação ao seu uso explícito como sujeito. A superioridade varia de 84% (“sei” em T1) a 94,5% (“quero” em T2).
2. Comparando os dois verbos nos dois textos, vê-se que a elipse do sujeito “eu” é superior em T2, ou seja, no texto traduzido (94,5% a 88% para “quero”; 88% a 84% para “sei”).

Portanto, com relação aos verbos marcados analisados, pode-se confirmar a tendência à elipse do sujeito “eu” em Lobato ao produzir *Caçadas de Pedrinho* (T1) espontaneamente em português. Esta tendência se vê ainda superior em sua práxis tradutória (T2). Estes dados e números corroboram a afirmação de Baker (1998), de que os tradutores tendem a normalizar a língua de chegada e a exagerar as características da língua de chegada no texto traduzido. Neste caso, a normalização ocorre através da elipse do sujeito, descrita, prescrita e mesmo defendida por gramáticos do português, conforme demonstrado na Seção 1.

#### 4.2. VERBOS NÃO MARCADOS

Neste estudo, os dois verbos não marcados analisados são “era” e “disse”. “Era” resultou 168 ocorrências em T1 e 273 em T2, totalizando 441. “Disse” resultou 131 em T1 e 161 ocorrências em T2, totalizando 292. Deste modo, T1 participou do estudo com um total de 299 ocorrências de verbos marcados e T2 com um total de 434.

Dentre as 168 ocorrências de “era” em T1, em 167 vezes ele é utilizado como verbo impessoal (e.g.: “Bem amarrada que foi a onça, era preciso conduzi-la até a casa” – T1, *era*, ocorrência 09) ou 3ª pessoa (e.g.: “Essa figueira parecia ter mil anos de idade. Era a maior árvore da zona.” – T1, *era*, ocorrência 15). Portanto, somente uma ocorrência resultou referir-se à primeira pessoa: “Eu, por exemplo, se fosse rico, era capaz de dar até trinta contos por um rinoceronte” (T1, *era*, ocorrência 47), caso este em que aparece com “eu” explícito. Deste modo, entre as ocorrências analisáveis de “era” em T1, 100% se referem ao uso de “eu” explícito.

Em T2, do total de 273 ocorrências de “era”, apenas 12 ocorrências foram identificadas como sendo de 1ª pessoa. Dentre estas, 4 ocorrências (33%) vêm antecedidas por “eu” explícito enquanto que em 8 ocorrências (67%) o caso é de elipse do pronome.

Quanto ao verbo “disse”, das 131 ocorrências em T1, 129 casos se relacionam à 3ª pessoa. Foram identificados somente dois casos em que é usado na 1ª pessoa, sendo 1 (50%) com sujeito explícito, “... aqui a vida inteira, se for preciso. – Eu não disse? – cochichou Dona Benta.” (T1, *disse*, ocorrência 43) e 1 (50%) com sujeito elíptico “...disseste a teu irmão? – perguntaram os índios. – disse-lhe – inventou Hans.” (T1, *disse*, ocorrência 116).

Em T2, “disse” resultou 161 ocorrências, das quais 38 são de 1ª pessoa. Estas estão divididas da seguinte maneira: 20 casos de pronome “eu” explícito (52,5%) e 18 casos de sujeito elíptico (47,5%).

O quadro 4 mostra as ocorrências analisáveis de “era” e “disse” (referentes à 1ª pessoa)

*Quadro 4. Sujeito “eu” explícito e elipse pronominal — verbos não marcados.*

VERBOS NÃO MARCADOS					
		T1		T2	
		ocorrências	% da amostra	ocorrências	% da amostra
ERA	Explícito	1	100%	4	33%
	Elíptico	0	0%	8	67%
DISSE	Explícito	1	50%	20	52,5%
	Elíptico	1	50%	18	47,5%

Com base nestes números pode-se concluir que:

1. Quantitativamente, T2 contribui com muitas mais ocorrências do que T1 no tocante aos 2 verbos. São 50 ocorrências no total em T2 contra 3 em T1. No caso de “era” a proporção é de 12 a 1; no caso de “disse” a proporção atinge 38 a 2.
2. Apesar do desbalanço quantitativo de ocorrências entre T1 e T2, há um equilíbrio percentual entre os casos de “eu” explícito e elíptico para “disse” em T1 (50% a 50%) e em T2 (52,5% a 47,5%, respectivamente).
3. Em T1, para os casos de “era”, a superioridade é do “eu” explícito. No entanto, para este caso, somente uma ocorrência foi computada.
4. Em T2, para os casos de “era”, a superioridade percentual é do “eu” elíptico (67% a 33%), o que corrobora os resultados observados para os verbos marcados.

## 5. CONSIDERAÇÕES FINAIS

Há diferenças entre os resultados obtidos para as ocorrências de verbos marcados e não marcados nas duas obras que compõem o corpus. Há um equilíbrio maior entre os verbos marcados nas duas obras, tanto em nível quantitativo quanto na relação entre as ocorrências de “eu” explícito e elíptico. Enquanto isso, entre os verbos não marcados, além do desequilíbrio quantitativo de ocorrências entre T1 e T2, as proporções também não são uniformes.

Deste modo, os números que respondem à primeira pergunta desta pesquisa, sobre a prevalência do uso explícito do pronome pessoal “eu” na função de sujeito ou da sua elipse no corpus estudado, indicam que prevalece a elipse entre os dois verbos marcados nas duas obras. Já entre os dois verbos não marcados, somente um dos verbos (“era”) em uma das obras (T2) demonstra superioridade da elipse e corrobora os resultados obtidos com os verbos marcados, mesmo que percentualmente o número seja inferior.

Com relação à segunda pergunta, sobre a diferença na razão entre as ocorrências de uso explícito do pronome “eu” na função de sujeito e a sua elipse na comparação das duas obras, os números indicam que para os verbos marcados, a diferença é de fato notável, sendo que a superioridade da elipse em relação ao “eu” explícito fica entre 84% e 94,5%. Já entre os verbos não marcados, a superioridade foi de 67% a 33% para a elipse somente no caso de “era” em T2.

Para “disse” tanto em T1 como em T2, há um equilíbrio percentual a despeito da amostra em T1 ter apresentado somente 2 ocorrências. Nota-se também que as ocorrências em T1 para os 2 verbos obedecem uma coerência no sentido de que se encontram em número reduzido.



Com base na comparação dos verbos não marcados entre T1 e T2, é possível que a diferença quantitativa geral de ocorrências se dê em razão de que os verbos sejam usados por agentes diferentes nas duas obras, uma vez que em *Caçadas de Pedrinho* (T1) o narrador é onisciente e em *Aventuras de Huck* (T2) o narrador é também protagonista.

Para refinar as descobertas desta pesquisa, algumas sugestões de futuros estudos são a ampliação do número de verbos marcados e não marcados analisados neste mesmo corpus e/ou a replicação da metodologia proposta neste trabalho com um corpus comparável ampliado de autoria de Lobato.

## REFERÊNCIAS BIBLIOGRÁFICAS

### Corpus de estudo

LOBATO, M. Caçadas de Pedrinho. São Paulo: Círculo do Livro. 1988.

TWAIN. M. Aventuras de Huck. Tradução de Monteiro Lobato. São Paulo: Editora Brasiliense, 1973. Primeira publicação 1934. Original em inglês.

### Software

ANTHONY, Laurence. AntConc3.2.4w. Tóquio, Japão: Waseda University, 2011. Disponível em <http://www.antlab.sci.waseda.ac.jp/software.html>. Acesso em 21 de agosto de 2013.

### Textos Teóricos

BAKER, Mona. *Corpus Linguistics and translations studies: An overview and some suggestions for future research*. Target – International Journal of Translation Studies. John Benjamins. 1995. Vol 7.2.

\_\_\_\_\_. *Meta: journal des traducteurs/Meta: Translators' Journal*, vol. 43, nº 4, 1998, p. 480-485. Disponível em: [www.erudit.org/revue/meta/1998/v43/n4/001951ar.pdf](http://www.erudit.org/revue/meta/1998/v43/n4/001951ar.pdf). Acesso em 18 de maio de 2013.

BECHARA, E., *Moderna Gramática Portuguesa*. São Paulo: Nova Fronteira, 2009. 37ª ed.

BUENO, L. T. *Transitividade, Coesão e Criatividade Lexical no Corpus Paralelo Macunaíma, De Andrade e Macunaíma, de Goodland*. Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, UFMG, Belo Horizonte. 2005.

CAMPOS, G. C.; OLIVEIRA, M. C. C. *O pensamento e a prática de Monteiro Lobato como tradutor*. In: *Ipotesi – Revista de Estudos Literários*, v. 13, nº 1, 2009. Disponível em: <http://www.editoraufjf.com.br/revista/index.php/ipotesi/article/view/415/389>. Acesso em 11 de agosto de 2013.

DAVIES, M. (2008) *Corpus do Português: One billion words, 4 countries*. Disponível em <http://www.corpusdoportugues.org>. Acesso em 20 de agosto de 2013.

LIMA, C. H. R. *Gramática Normativa da língua portuguesa*. Rio de Janeiro: José Olympio, 2006. 45ª ed.

LOBATO, M. A barca de Gleyre. São Paulo: Brasiliense. 1955. 2º tomo.

MAIA, B. *Word Order and the First Person Singular in Portuguese and English*. In: *Meta: journal des traducteurs / Meta: Translators' Journal*, vol. 43, nº 4, 1998. Disponível em: <http://www.erudit.org/revue/meta/1998/v43/n4/003539ar.pdf>. Acesso em 11/08/2013.

MENDES, D. R. Monteiro Lobato, o tradutor. Monografia (Bacharelado em Letras: Ênfase em Tradução – Inglês) Instituto de Ciências Humanas e de Letras, Universidade Federal de Juiz de Fora. Juiz de Fora, 2002. Disponível em [http://www.ufjf.br/bachareladotradingles/files/2011/02/denise\\_rezende\\_mendes.pdf](http://www.ufjf.br/bachareladotradingles/files/2011/02/denise_rezende_mendes.pdf). Acesso em 11/08/2013.

NEVES, M. H. M. *Gramática de Usos do Português*. São Paulo: UNESP, 2000.

OLOHAN, M. *Introducing Corpora in Translation Studies*. New York: Routledge, 2004.

PERINI, M. A. *Gramática descritiva do português*. São Paulo: Ática, 2006.

TARALLO et. al. *Rupturas na Ordem de adjacências canônicas no português falado*. In: *Gramática do Português Falado*. Vol. I: A ordem. Campinas: Editora da Unicamp, 2002. 4ª ed.

SAID ALI, M. *Gramática Secundária e Gramática Histórica da Língua Portuguesa*. São Paulo: Melhoramentos, 1964.

SARDINHA, T. B. *Linguística de Corpus: Histórico e Problemática*. D.E.L.T.A. São Paulo: EDUC, v. 16, n. 2, 2000.

## ANEXO I - TERMO DE COMPROMISSO DE ORIGINALIDADE

A presente declaração é termo integrante de todo trabalho de conclusão de curso (TCC) a ser submetido à avaliação da Coordenação do Curso de Tradução da UFPB como requisito necessário e obrigatório à obtenção do grau de bacharel em tradução.

Eu, **CAIO CESAR MARTINO**, portador do **RG 17.333.000-9 SSP/SP**, na qualidade de aluno da Graduação do Curso de Tradução da Universidade Federal da Paraíba, declaro, para os devidos fins, que:

- O Trabalho de Conclusão de Curso anexo, requisito necessário à obtenção do grau de bacharel em tradução pela Universidade Federal da Paraíba, encontra-se plenamente em conformidade com os critérios técnicos, acadêmicos e científicos de originalidade;
- O referido TCC foi elaborado com minhas próprias palavras, ideias, opiniões e juízos de valor, não consistindo, portanto, plágio, por não reproduzir, como se meus fossem, pensamentos, ideias e palavras de outra pessoa;
- As citações diretas de trabalhos de outras pessoas, publicados ou não, apresentadas em meu TCC, estão sempre claramente identificadas entre aspas e com a completa referência bibliográfica de sua fonte, de acordo com as normas vigentes da ABNT;
- Todas as séries de pequenas citações de diversas fontes diferentes foram identificadas como tais, bem como as longas citações de uma única fonte foram incorporadas suas respectivas referências bibliográficas, pois fui devidamente informado e orientada a respeito do fato de que, caso contrário, as mesmas constituiriam plágio;
- todos os resumos e/ou sumários de ideias e julgamentos de outras pessoas estão acompanhados da indicação de suas fontes em seu texto e as mesmas constam das referências bibliográficas do TCC, pois fui devidamente informado e orientado a respeito do fato de que a inobservância destas regras poderia acarretar alegação de fraude.

O Professor responsável pela orientação de meu trabalho de conclusão de curso (TCC) apresentou-me a presente declaração, requerendo o meu compromisso de não praticar quaisquer atos que pudessem ser entendidos como plágio na elaboração de meu TCC, razão pela qual declaro ter lido e entendido todo o seu conteúdo e submeto o documento em anexo para apreciação da Coordenação do Curso de Tradução da UFPB como fruto de meu exclusivo trabalho.

João Pessoa, 09 / 03 / 2014.

CAIO CESAR MARTINO