

**Atribuição de autoria: o enigma por trás da autoria do livro sátiras e outras
subversões**

**Attribution of authorship: the riddle behind the authorship of the book satire and
other subversions**

Joniel da Costa Barbosa¹

Universidade Federal do Piauí

Ivan dos Santos Oliveira²

Unisinós

Resumo: Esta pesquisa traz como tema a análise estilométrica do livro *Sátira e outras subversões* no qual investigamos a autoria de alguns dos textos contidos na obra. Tivemos como objetivo geral analisar a autoria dos textos reunidos por Felipe Botelho Corrêa, no livro, para comprovar ou não a autoria de Lima Barreto. Buscamos identificar se os métodos adotados são eficazes e podem ser utilizados para atribuir autoria a um texto e ainda investigar se é possível chegar a resultados satisfatórios, para que se possa comprovar a autoria do livro utilizando a estilometria, com o auxílio de programas de computador. Para a realização deste trabalho foram utilizados três programas de computador, o *PDFelement 6 Pro* responsável por transformar os textos digitalizados para o formato *Docx*, o programa *Word* que foi utilizado para transformar os textos para o formato *Txt*, e, finalmente o programa *Lexico 3* que é responsável por analisar o *corpus* e gerar dados estatísticos. Com base nos dados analisados, foi possível constatar que os 10 (dez) textos analisados e comparados ao texto extraído do livro *Recordações do Escrivão Isaias Caminha* realmente teriam sido escritos por Lima Barreto, fato este que corrobora a autoria atribuída por Felipe Botelho Corrêa. É importante ressaltar que, para se afirmar que todos os 164 textos contidos no livro teriam sido escritos por Lima Barreto, é necessário fazer uma pesquisa mais ampla, pois seria necessário analisar todos os textos.

Palavras-chave: Sátira e outras subversões. Estilometria. Lima Barreto.

Abstract: This research brings as a theme the stylometric analysis of the book *Sátiras e outras subversões* where we investigate the authorship of some of the texts contained in the work. Its general objective was to analyze the authorship of the texts gathered by Felipe Botelho Corrêa in the book, to prove or not the authorship of Lima Barreto. We try to identify if the adopted methods are effective and can be used to attribute authorship of a text and also to investigate if it is possible to arrive at satisfactory results, so that the authorship of the book can be verified using the stylometry, with the aid of computer programs. In order to carry out this work, three computer programs were used: *PDFelement 6 Pro* responsible for transforming the scanned texts into the "Docx" format, the "Word" program that was used to transform the texts into the *Txt* format, and finally the *Lexico 3* program which is responsible for analyzing the corpus and generating statistical data. Based on the analyzed data, it was possible to verify that the 10 (ten) texts analyzed and compared to the text extracted from the book *Recordações do Escrivão Isaias Caminha* really would have been written by Lima Barreto, a fact that corroborates the authorship attributed by Felipe Botelho Corrêa. It is important to emphasize that, in order to affirm that all 164 texts contained in the book would have been written by Lima Barreto, it is necessary to do a more extensive research, since it would be necessary to analyze all texts.

Key-words: Sátiras e outras subversões. Stylometry. Lima Barreto.

¹ Graduando do curso de Letras Português do PARFOR-UFPI 2017. E-mail: jonielcosta2010@hotmail.com

² Professor Orientador, Mestre em Educação pela UNISINOS-RS. E-mail: ivanoliveira@ifpi.edu.br

Recebido em 10 de setembro de 2018.
Aprovado em 22 de novembro de 2018.

Introdução

Buscando elucidar a real autoria dos textos publicados no livro “*Sátiras e outras subversões*”, até então atribuídas ao escritor brasileiro Lima Barreto, sentimos a necessidade de promover uma pesquisa que buscasse comprovar, ou não, a autoria desses textos valendo-nos de mecanismos tecnológicos para a obtenção dos resultados. Desse modo, a presente pesquisa surgiu para buscar examinar a autenticidade da autoria desses textos literários que recentemente foram atribuídos ao escritor Lima Barreto e publicados no livro *Sátiras e outras subversões*, organizado pelo professor e pesquisador Felipe Botelho Corrêa da *King’s College London* do Reino Unido.

Esta pesquisa tem como tema a análise estilística e busca a comprovação de autoria póstuma do livro *Sátiras e outras subversões*, na qual buscamos estabelecer um padrão de escrita e encontrar características lexicométricas que contribuíssem para que se possa comprovar ou não a autoria de alguns textos do livro estudado.

O presente trabalho buscou responder à seguinte problemática: o livro *Sátiras e outras subversões*, que reuniu 164 textos de autoria atribuída a Lima Barreto, teria sido realmente escrito por ele? É possível chegar a um resultado plausível sobre a autoria do livro utilizando a metodologia escolhida? Para responder a essas perguntas formulamos a seguinte pesquisa, em que enfatizamos que atribuição de autoria póstuma é sempre motivo de dúvida, por não ter certeza que essa autoria é válida. É necessário investigar a fundo esses textos para comprovar ou não essa autoria. Para isso, utilizamos o método de análise estilométrica comparativa computacional e análise estilística em que buscamos chegar a um resultado plausível utilizando programas sofisticados de computador como o *Word*, *PDFelement 6 Pro* e o *Lexico 3*.

Utilizamos como *corpus* para esta pesquisa os 10 (dez) primeiros textos do livro, sendo que o último texto foi cortado em parte para que chegássemos a um total exato de 5.000 (cinco mil) palavras, as quais foram comparadas com outras 5.000 (cinco mil) palavras do livro *Recordações do Escrivão Isaías Caminha*, cuja autoria é de Lima Barreto, de forma que foram escolhidas do segundo livro as 5.000 (cinco mil) primeiras

palavras que aparecem no romance, tendo como *corpus* 10.000 (dez mil) palavras analisadas na pesquisa.

Dessa forma, propomos como objetivo geral: analisar a autoria dos textos reunidos por Felipe Botelho Corrêa no livro *Sátiras e outras subversões* para comprovar ou não a autoria de Lima Barreto. E, como objetivos secundários, identificar se os métodos adotados são eficazes e podem ser utilizados para atribuir autoria a um texto e ainda investigar se é possível chegar a um resultado plausível sobre a autoria do livro, utilizando a metodologia estilométrica com o auxílio de programas de computador.

Afonso Henrique de Lima Barreto foi um dos mais influentes escritores brasileiros de sua época, autor de clássicos como *Recordações do Escrivão Isaías Caminha* e *Triste fim de Policarpo Quaresma*, entre outros. Assim como alguns outros autores de sua época, grande parte de sua obra foi redescoberta e publicada após sua morte, onde muitos dos textos atribuídos a Lima Barreto ainda são de autoria duvidosa fato esse que gerou o interesse que foi o norteador desta pesquisa.

A investigação sobre autoria de textos literários não é recente, diversos estudiosos da literatura já se embrenharam em pesquisas e conseguiram chegar a resultados que desmistificam e elucidam a autoria de obras que até então eram desconhecidas. Não é por acaso que o movimento pela busca autoral é de relevante importância, pois possibilita que textos que na sua origem foram assinados por pseudônimos ou até mesmo não foram assinados, possam receber seus respectivos autores, possibilitando assim, o enriquecimento do rol de obras produzidas pelo autor, fazendo com que a mesma adquira uma identidade.

Dessa forma, as análises foram feitas baseadas nos estudos de Corrêa (2016) que atribuiu a autoria do livro à Lima Barreto e nas pesquisas de Paiva (2013), Machado (2014) que estudam a estilometria de textos e Brandão (2006) que tem um trabalho que investiga a autoria de textos com autoria suspeita com a utilização de programas de computador, e que, em um estudo semelhante a este, investiga a autoria de *Cartas chilenas*, de autoria atribuída à Tomás Antônio Gonzaga.

Para essa análise, foram utilizados alguns programas de computador capazes de aferir algumas informações sobre a maneira de escrever do(s) autor(es) pesquisado(s), buscando-se resultados satisfatórios sobre as características presentes em cada obra analisada, utilizando para isso programas de computador. Com isso, esperamos que ao fim desse processo fosse possível determinar se a autoria da obra em questão é ou não de Lima Barreto.

Essa pesquisa é importante para revelar a verdadeira autoria do livro em questão e caso seja confirmada, esses textos venham contribuir para enriquecer o rol de textos literários de um dos maiores expoentes da literatura nacional, Lima Barreto.

A pesquisa está subdividida em seções para a melhor exposição e compreensão do texto, na qual começamos pela introdução em que trazemos as primeiras informações sobre o tema pesquisado; o referencial teórico em que discutimos as principais teorias que norteiam o trabalho, um pouco da história do uso da tecnologia para estudos de atribuição ou comprovação de autoria e ainda uma subseção em que tratamos da tecnologia e seu auxílio na busca pela veracidade da autoria de obras literárias; em seguida trazemos a metodologia que foi aplicada na pesquisa. Adiante, fizemos a sistematização dos dados obtidos durante a pesquisa, a análise e discussão dos dados, e a conclusão a que chegamos com o desenvolvimento do trabalho. Por fim, as referências que balizaram o presente trabalho acadêmico.

1. Lima Barreto: breve biografia

O autor Afonso Henriques de Lima Barreto nasceu em 1881, na cidade do Rio de Janeiro – RJ e faleceu em novembro de 1922 aos 41 anos de idade. Coursou engenharia na Escola Politécnica, não terminando o curso por motivos familiares.

Ainda estudante começou a escrever para jornais cariocas para complementar sua renda, já que trabalhava como escrevente copista na Secretaria de Guerra. Escreveu romances, sátiras, crônicas, contos, chegando a publicar textos na revista *Fon-Fon*. Em 1909 publicou em livro o romance *Recordações do Escrivão Isaías Caminha* na revista *Floreal* a qual dirigia.

Lima Barreto, como ficou conhecido, possui uma vasta produção literária, sendo autor ainda de clássicos como: *Triste fim de Policarpo Quaresma*, *Numa e Ninfa* e *Clara dos Anjos*.

É um escritor da fase de transição entre o Realismo e o Modernismo, possuindo um estilo de escrever distinto dos escritores de sua época. Em sua obra, Lima Barreto fazia críticas ao regime da República Velha em que o cenário político era dominado por Minas Gerais e São Paulo e ainda, criticava as injustiças sociais da época.

Lima Barreto nunca conseguiu em vida a notoriedade que merecia, sendo reconhecido como um grande escritor brasileiro apenas depois de sua morte, visto que

grande parte de sua obra foi publicada postumamente, como é o caso do livro alvo deste estudo.

Tendo sido internado para tratamento de alcoolismo, Lima Barreto candidatou-se por duas vezes para a Academia Brasileira de Letras, porém sem êxito.

2. Uso da tecnologia para atribuir ou comprovar a autoria de obras literárias

Para que se possa dar início a esta verificação, existe a necessidade de se reportar um pouco da história sobre os meios utilizados ao longo da história, para que se identifique se livros, textos, trabalhos e etc., possuem uma autenticidade concreta em relação a quem escreve os textos.

Pesquisas que busquem atribuir ou mesmo comprovar a autoria de obras até então desconhecidas e/ou questionadas não são comuns, especialmente no Brasil. Embora haja diversas obras brasileiras que merecem uma análise minuciosa a respeito de suas autorias.

Para entender melhor como essa utilização da tecnologia facilitou o trabalho de estudiosos que possuem trabalhos voltados para a investigação de autoria de obras literárias, trataremos a seguir uma pequena retrospectiva histórica.

2.1 Breve histórico

Desde muito tempo, no universo literário, indagações sobre a autoria de obras literárias povoam o imaginário e provocam curiosidade em pesquisadores e estudiosos da literatura, “[...] a determinação de autoria é ocupação importante de filólogos e teólogos, a mesma que passou a preocupar, mais recentemente, também os teóricos e críticos literários” (BRANDÃO, 2006, p. 02).

Os primeiros estudos que se propuseram a investigar a autoria de obras foram voltados a investigar a autenticidade de textos religiosos, como afirma Brandão (2002, p. 01):

Determinar a autoria de textos anônimos não é atividade nova no meio acadêmico e na tradição ocidental. Um exemplo didático e fecundo da antiguidade dessa atividade está primordialmente nos estudos dirigidos no meio religioso, com o objetivo de determinar a autenticidade de textos da tradição cristã, para afirmar o valor da reprodução dos ensinamentos e feitos de Jesus Cristo enunciados pelos apóstolos. O trabalho da determinação da autoria e, portanto, da autoridade desses textos para a fundação da tradição cristã, mobilizou estudiosos que, desde então e a cada época, retomam o

debate do tema com instrumentos de trabalhos novos em função do objetivo de consolidar o conjunto de princípios que servem de base para o sistema religioso que pretende se estabelecer doutrinariamente.

Comum na França e nos Estados Unidos, esse tipo de trabalho busca averiguar a veracidade da autoria em obras literárias que sejam questionáveis.

A verificação de autoria é feita quando dado um documento de autoria questionada é verificado se este documento foi elaborado por um autor determinado ou não. Trata-se, neste caso, de uma classificação binária, onde o resultado é uma resposta positiva ou negativa de autoria. (OLIVEIRA, 2011, p. 02)

Os primeiros trabalhos relacionados à atribuição e verificação de autoria utilizavam métodos menos sofisticados dos que os empregados atualmente, visto que as análises eram feitas por meio de especialistas humanos.

Os estudos de atribuição de autoria não tradicionais tiveram início com a pesquisa de Mosteller e Wallace (1964) e a partir dos anos 90 esses trabalhos foram dominados por tentativas de definir um estilo de escrita dos autores pesquisados, esse padrão de escrita do autor é conhecido como “Estilometria”. Atualmente, esse tipo de pesquisa é realizada em sua maioria com o auxílio de recursos computacionais que possibilitam mais praticidade, que se torna essencial para que o pesquisador analise os dados de maneira mais precisa, possibilitando assim que se chegue a um resultado mais confiável.

[... provavelmente, os primeiros trabalhos mais significativos sobre atribuição de autoria foram as pesquisas de Mosteller e Wallace, em 1964 (Mosteller, F. E Wallace, D. L., 1964), sobre uma série de ensaios políticos conhecidos como The Federalist Papers (“Os papéis federalistas”), composto por 134 documentos de autoria conhecida e 12 documentos de autoria questionada entre dois prováveis autores. (OLIVEIRA Jr., 2011, P. 01)

Segundo Mendenhall (1987, apud Stamatatos, E. 2009, p. 540), as pesquisas pioneiras relacionadas à atribuição e verificação de autoria foram baseadas em simples medidas e comprimentos de palavras e sentenças que eram comumente utilizadas no texto pesquisado. É importante dizer que esses trabalhos eram feitos manualmente, sem o auxílio de qualquer recurso computacional, o que dificultava e prolongava o tempo de análise.

3. A tecnologia e seu auxílio na busca pela veracidade da autoria de uma obra

Na iniciação da era da Informática, sabemos que tais máquinas tinham sua utilização específica para cálculos de guerra; efetuando assim, somente operações matemáticas, com a capacidade de superar as maiores mentes de matemáticos na execução de processamento de cálculos. No entanto, com o passar do tempo, inicia-se a necessidade de armazenar os dados, especialmente porque devido ao alto poder de processamento das máquinas, passou a sobrecarregar a mente de quem operava as mesmas. Dessa forma, houve a necessidade de aprimorar essa tecnologia que, nos primórdios, não era suficiente para atender todas as necessidades de seus usuários. Sendo assim, o rápido desenvolvimento das tecnologias dá a permissão para o surgimento de algo que possa garantir a autenticidade de assinaturas que possuem características próprias, ou seja, existe a possibilidade de se dar o mesmo significado e eficácia jurídica da assinatura manual.

A evolução das máquinas deu maior favorecimento ao usuário a partir do momento em que se tornaram acessíveis a todos, fato este que possibilitou que estudos como este viessem a se tornar mais presentes no meio acadêmico a partir do surgimento dos computadores.

Com o surgimento da internet apareceram também núcleos de pesquisa voltados para estudos relacionados à literatura digitalizada, o que favoreceu o acesso a obras que até então só estavam disponíveis de forma impressa. Esses núcleos de pesquisa são importantes ainda no desenvolvimento de *softwares*, capazes de detalhar diversos aspectos importantes e gerar resultados confiáveis a respeito de determinada obra literária.

Esta pesquisa se enquadra na linha de pesquisa literatura e tecnologia que se utiliza de meios tecnológicos computacionais para investigar o estilo e a riqueza lexical de textos literários. Dessa forma, buscaremos verificar a autoria do livro em questão utilizando-nos de recursos digitais que possibilitem uma análise confiável que seja capaz de aferir a autoria dos textos reunidos no livro *Sátiras e outras subversões*.

Já existem autores que trabalham diretamente com estudos estilométricos e lexicométricos como: Paiva (2013), Machado (2014) e Brandão (2006), este último além de trabalhar com a lexicometria e a estilometria, possui um trabalho pioneiro no Brasil de investigação de autoria de textos literários, mais especificamente os textos das *Cartas Chilenas* que foram atribuídos ao escritor português Tomás Antônio Gonzaga.

Buscando verificar a autoria do livro *Sátiras e outras subversões*, recorreremos aos estudiosos acima citados que desenvolveram pesquisas semelhantes, que certamente

contribuirão para verificar a autoria da obra supracitada de autoria até então atribuída à Lima Barreto.

Para a tarefa proposta utilizaremos, inicialmente, três programas de computador denominados *PDFelement 6 Pro*, *Word* e *Lexico 3*, responsáveis por fazer uma avaliação lexicométrica e estilométrica em alguns textos do livro *Sátiras e outras subversões* em que serão analisadas aproximadamente 5.000 (cinco mil) palavras e comparadas a uma parte do livro *Recordações do Escrivão Isaias Caminha* com a mesma quantidade de palavras.

Machado (2014) julga que as pesquisas estilométricas e lexicométricas são estudos estilísticos e estatísticos e que, portanto, baseia-se na coleta de dados que quando interpretados servem para dar embasamento para entender o estilo de cada autor.

Ainda nesta linha de estudo, Paiva (2013. p. 32 – 33) diz que:

Os estudos de atribuição de autoria têm muito em comum com os estudos estilométricos. Ambos são potencializados pela informática, fazendo uso da estatística textual para alcançar seus fins. No entanto, os estudos de atribuição de autoria partem das possíveis semelhanças entre os textos que podem embasar o reconhecimento de autoria. A estilometria, pelo menos em princípio, busca o contrário: aqueles aspectos que distinguem um texto do outro, que fazem daquele um texto singular, investiga o estilo próprio de uma obra, autor ou época.

Nesta perspectiva, procuramos estabelecer características próprias da literatura de Lima Barreto, para isso utilizaremos estes estudos estilométricos que servirão como base para o desenvolvimento deste trabalho que ajudarão a compreender o estilo dos textos estudados para ajudar a corroborar ou refutar a autoria dos textos contidos no livro *Sátiras e outras subversões*.

Vale lembrar que esta pesquisa se deu com o auxílio de programas de computador, e como qualquer máquina, está suscetível a falhas, sendo assim, carece de uma atenção muito grande principalmente no momento da conversão dos textos para o formato desejado, visto que é perceptível a aglutinação de caracteres que se não observados podem comprometer a idoneidade da pesquisa.

Sabendo que foram utilizados dados estatísticos para a obtenção e análise dos resultados, estas informações devem ser checadas antes que se dê um parecer, não se pode esperar que o programa *Lexico 3* sozinho chegue a uma interpretação consistente.

Paiva (2013, p.13) afirma que “como a estilometria se baseia em dados quantitativos, carece de parâmetros para que os números possam ser corretamente interpretados”.

Assim como Paiva (2013), Bernard, (1999, *Apud* Cúrcio, 2006, p. 04) ressalta que:

[...] para realizar uma melhor leitura dos resultados de um tratamento lexicométrico, é bom lembrar que os computadores têm apenas acesso a códigos numéricos e jamais ao sentido das palavras, quer dizer, a máquina é (ainda) incapaz de compreender informações semânticas ou simbólicas.

Deste modo, é importante ressaltar que o computador sozinho não é capaz de determinar a autoria de um texto, embora auxilie bastante esse processo de descoberta, é preciso que o pesquisador siga parâmetros para entender e transformar os dados em resultados que possam ser eficientes e possam determinar a autoria dos textos em questão.

4. Metodologia

Esta pesquisa consiste em uma análise estilométrica do livro *Sátira e outras subversões* em que foi feita uma comparação do estilo de escrita do possível autor desta obra e o autor do livro *Recordações do Escrivão Isaias Caminha*, Lima Barreto.

Para a realização dessa pesquisa investigativa, utilizamos alguns programas de computador capazes de fazer uma varredura textual, enumerando diversos aspectos da maneira de escrever do autor Lima Barreto e foram gerados alguns dados que foram imprescindíveis neste trabalho.

Essa pesquisa foi desenvolvida em etapas, nessas etapas foram feitos processos de conversão dos textos para diversos formatos, sendo que ao final utilizamos o formato *Txt*. Inicialmente fizemos necessário a utilização de um *Scanner*, já que os textos contidos no livro estudado não se encontram disponíveis em formato digital. Como os textos digitalizados foram salvos em formato *PDF*, houve a necessidade de convertê-los para o formato a ser utilizado. Para essa conversão utilizamos o programa *PDFelement 6 Pro* que foi responsável por transformar as imagens geradas em *PDF* para o formato *Docx* (documentos do Word) e posteriormente convertidos em *Txt* (texto sem formatação).

Finalizado o processo de conversão dos textos para o formato “Txt”, fez-se necessário a utilização do programa de computador “*Lexico 3*”, que possibilita que se possa aferir propriedades estatísticas presentes nos textos e ainda características do modo de escrever do (s) autor (es) e com isso, esperou-se que fossem obtidos resultados suficientes para comprovar ou negar a autoria em questão.

É importante mencionar, que trabalhamos com textos de tamanhos semelhantes, para que os resultados não fossem discrepantes, para isso fizemos recortes nos textos.

Vale lembrar, que todos os programas acima relacionados são interfaces gratuitas, de recursos limitados, e que sozinhos não são capazes de definir a autoria de um determinado *corpus*, sendo assim, fez-se necessário uma análise detalhada dos dados para que fosse possível chegar a resultados satisfatórios e com isso corroborar ou refutar a autoria do livro supracitado cuja autoria foi atribuída à Lima Barreto.

5. Análise e discussão

5.1 Caracterização do corpus

As características léxicas são as que consideram as características do conjunto de vocábulos de um idioma. As abordagens léxicas podem ser:

- baseadas em tokens;
- baseadas em riqueza vocabular;
- baseadas em frequência de palavras;
- baseadas em n-gramas de palavras;
- baseadas em erros ortográficos.

(Oliveira Jr., 2011. p.9)

Neste artigo utilizamos as abordagens baseadas em frequências de palavras, ou seja, analisamos a quantidade de vezes que uma palavra ou grupo de palavras aparece em um determinado *corpus*.

Estabelecer um parâmetro de escrita de um autor não é tarefa fácil, mas é possível reconhecer traços na sua escrita. Instintivamente cada autor escreve com um rol de palavras que faz parte de seu vocabulário, com repetições de palavras que são feitas sem que ele se dê conta. Ao analisar essas palavras é possível determinar quais palavras são mais usuais na escrita de cada autor, qual a função que elas exercem e com isso identificar as características presentes na forma de escrever de cada autor.

Nesta pesquisa, propôs-se fazer uma comparação entre duas obras, o livro *Sátiras e outras subversões*, cuja autoria está sendo estudada, e o livro *Recordações do*

Escrivão Isaias Caminha de autoria de Lima Barreto, com a finalidade de chegar à comprovação ou não da autoria da primeira obra. Para realizar esta comparação, foram levados em consideração diversos aspectos, que vão desde a quantidade que determinada palavra aparece nos textos até o percentual de palavras funcionais, aquelas que estabelecem relação entre os termos do discurso e palavras lexicais.

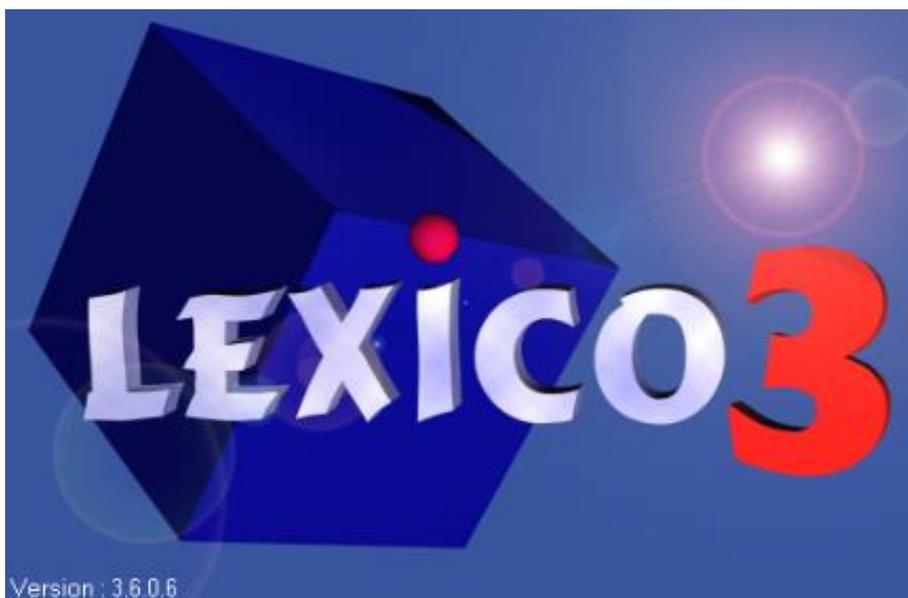
Para servir de *corpus* para nossa pesquisa utilizamos como dados 5.000 (cinco mil) palavras que foram escolhidas seguindo a ordem dos textos dispostos no livro *Sátiras e outras subversões*, para se chegar a esse quantitativo de palavras foram utilizados os 10 (dez) primeiros textos do livro estudado, sendo que o último não foi inteiramente utilizado, pois houve a necessidade de fazer um corte para que as palavras não ultrapassassem a quantidade a serem estudadas.

Assim como no livro *Sátiras e outras subversões*, no livro *Recordações do Escrivão Isaias Caminha* houve a necessidade de se fazer um corte no texto, visto que se trata de um trabalho comparativo. É importante que os textos apresentem tamanhos semelhantes para que não haja discrepâncias no momento da interpretação dos dados, escolhendo-se o mesmo quantitativo de palavras, seguindo os mesmos critérios, sendo assim, foram analisadas ao todo 10.000 palavras. Uma vez definidas as palavras que serviram de base para nossa pesquisa, coletou-se os dados transformando em informações valiosas para que se possa determinar a veracidade da autoria do livro.

5.2 O lexico 3

Esta pesquisa surgiu com o intuito de corroborar com a atribuição da autoria do livro *Sátiras e outras subversões* de Lima Barreto. Como essa autoria foi recentemente atribuída, veio a questão: esses 164 textos teriam mesmo sido escrito por Lima Barreto? Para se obter esta resposta este trabalho utilizou-se de uma metodologia diferente da utilizada pelo pesquisador responsável por essa atribuição, sendo assim, utilizamos prioritariamente o programa de computador *Lexico 3*.

O *Lexico 3*, lançado em 2001, foi desenvolvido pela equipe do Centro de Lexicometria e de Análise Automática de Textos – CLA2T, vinculado à Equipe SYLED (SYstèmes Linguistiques, Enonciation et Discursivité) da Université Sorbonne-Nouvelle, na França. É a terceira versão do software *Lexico*, cuja primeira data de 1990. Tem versões para sistema operacional MAC, Linux e Windows, todas no idioma francês. Adotamos sua atualização mais recente, a 3.5.2.4, disponibilizada em 2007, para a plataforma Windows. (PAIVA, 2013, p.45)



Fonte: Lexico 3

O programa *Lexico 3* é um programa de computador de interface gratuita, que como sugere seu nome, é uma ferramenta muito boa na obtenção de dados para pesquisas de cunho lexicométrico. Esse programa, entre outras atribuições, faz a contagem das palavras presentes em uma determinada obra, possibilitando assim que o pesquisador utilize suas informações em suas pesquisas.

É importante ressaltar que as análises dos dados oriundos do programa são de responsabilidade do pesquisador, visto que o programa disponibiliza a frequência em que uma determinada palavra aparece no texto, mas a utilização destas informações, bem como suas interpretações ficam a cargo do pesquisador.

5.3 Transformação para o formato txt (Documento sem formatação)

Em virtude de nossa pesquisa ser desenvolvida com o auxílio de recursos tecnológicos e o livro alvo da pesquisa não estar disponível em formato digital, houve a necessidade de utilização de outras ferramentas como Scanner, responsável por digitalizar a imagem do livro, e o programa de computador *PDFelement 6 Pro*, disponível gratuitamente na internet, que foi o responsável em transformar as imagens *scaneadas* no formato *PDF* para o formato *Docx* (documento do Word), tornando-as manipuláveis. Vale ressaltar que o programa *PDFelement 6 Pro* é muito eficiente, mas

como a maioria dos programas de computador apresentam algumas falhas em relação a erros na conversão, pois com muita frequência há a subtração ou troca de caracteres, sendo assim necessário uma revisão minuciosa de todo o texto convertido, utilizando para isso a versão impressa do livro pesquisado. É importante dizer que para a manipulação dos textos no *Lexico 3* é fundamental que os mesmos estejam formatados em caixa baixa (minúsculas), visto que o programa considera letra maiúscula como outra sentença.

Uma vez digitalizado, convertido e revisado o texto ainda não está pronto para ser utilizado, pois como descrito na metodologia, o trabalho foi desenvolvido com o auxílio do programa *Lexico 3*, interface gratuita disponível na web, e o referido programa trabalha prioritariamente com o texto em formato *Txt*. Para se chegar a esse formato é preciso que o documento em formato *Docx* seja salvo no programa *Word* no formato *Txt*. Ultrapassadas essas etapas de digitalização e conversão do texto, o mesmo está pronto para ser utilizado no programa *Lexico 3*.

5.4 Utilização do programa léxico 3

Com o texto pronto para ser trabalhado, começamos nossos trabalhos fazendo uma análise das formas que mais aparecem no *corpus* (*Sátiras e outras subversões e Recordações do Escrivão Isaias Caminha*).

A partir deste momento trataremos o *corpus* extraído do livro *Sátiras e outras subversões* como “TEXT01” e o *corpus* extraído do livro *Recordações do Escrivão Isaias Caminha* como “TEXT02”.

Eis duas tabelas com o resultado da análise dos dois textos:

TEXTO 01

Fichier Traitement Fenêtre

Lexico 3

Navigation Rapport Dictionnaire

Recherche :

Formes (ordre lexicométrique)	Frequence
de	238
o	180
e	169
a	137
que	124
do	89
—	89
não	68
da	63
em	60
os	54
um	54
com	42
as	41
dos	38
uma	38
para	34
é	31
se	31
ao	25
das	23
no	21
seu	20
à	19
foi	19
por	19
na	18
mais	17
como	16
rei	16
ministro	15
há	14
mas	14
em	14
2355 formes	

uma revista da cidade de salvador da bahia, há alguns meses, pôs em concurso a nacionalização de certos sobrenomes de origem estrangeira. queria ela que se pusessem em língua nacional os apelidos: tourmillon, laport, hasselmann, spinela, martinelli, silvany etc., e o leitor ou a leitora que melhor o fizesse receberia um prêmio.

não sei que fim teve a iniciativa da revista; mas não há como discordar que a ideia era genial. esses nomes de origem francesa, inglesa, italiana etc. sempre que qualquer um de nós, descendente do homem da lagoa santa,² topa com um deles, por exemplo, no diário oficial, causam-nos arrepios de indignação.

instintivamente, vemos neles a invasão do estrangeiro em coisas essencialmente nossas.

no despacho presidencial que precedeu a data de escrever estas reflexões, há a nomeação, para médico do exército, de um w. eisenlohr. esse nome é sueco ou dinamarquês, e o seu autor devia, previamente, ser intimado a nacionalizá-lo convenientemente.

há nomes estrangeiros em nacionais ilustres que só podem dar aos nossos inimigos de fora a noção de que somos governados por estranhos, tanto são eles respeitados nos jornais em virtude dos altos cargos da administração e da política que os seus portadores exercem.

vejam só: frontin, o dr. paulo; miller — o dr. general lauro; van erven — o das "águas"; schmidt, o senador; rondon, o general; ellis, o da defesa do café e da louça; ripper, o dos óculos; e tantos outros que sempre ocupam as colunas dos jornais oficiais, e não com os seus sobrenomes evidentemente estrangeiros que convém nacionalizar inteiramente, em obediência aos altos interesses da brasilidade.

há muitos outros em cargos menos importantes, mas não em menor destaque social, que também devem sofrer essa salvadora operação da nacionalidade; por exemplo: morize, bousquet, henninger, behring, lentes da escola politécnica; hynch e outros, na marinha de guerra; klinger, nicoll, no exército; mirineli, loretti, na diplomacia; além de muitos mais em várias e diferentes funções públicas que o brasil tem o direito de exigir que se deem a conhecer por nomes verdadeiramente nacionais, a menos que...

há ainda outros nomes de origem estrangeira que nós, os verdadeiros brasileiros, sentimos que sejam o de valorosos patricios. não é o caso do barão de teffé, que se chama hoonholtz? não é o caso do competente engenheiro militar ximeno villeroy? não é o caso do notável químico dr. nicolan ciancio? não é o caso do jovem e já célebre poeta paulista menotti del picchia? e quantos outros? cá em casa, até temos um, o nosso margiocco...

não é só na nossa onomástica que tal se dá; as denominações de nossos acidentes geográficos, de localidades, praças, ruas etc. estão enfeitados de nomes de outras línguas que não é a nossa. aqui, bem perto da rua do ourvidor, não temos nós; praia do russel, rua taylor, estação do méier, leblon, villegagnon etc. etc.?

as agremiações nacionalistas devem tomar uma providência a respeito, já e imediatamente, quanto a essas de proveniência inglesa, alemã, francesa, espanhola etc., deixando para depois as portuguesas, que não são, portanto, brasileiras geminamente, como: rio de janeiro, recife, porto alegre, são luis, bahia, campos, são paulo, campinas, e tantas outras.

em seguida, elas, as agremiações nacionalistas, devem encaminhar as suas vistas para os nomes originários de idiomas africanos: quilombo, munguengue³ etc.

depois dos sabidos estabelecerem que são autóctones nos nossos caboclos, poderemos conservar as denominações indígenas; mas, se o resultado das suas pesquisas for o contrário, devemos varrer da nossa nomenclatura topográfica: guanabara, niterói, manaus, itambé, cuiabá, goiás, ipanema e centenas de outras, por não serem perfeitamente nacionais.

a nossa nomenclatura contém me seja neminamente nacional em todos os aspectos e trata-se do que tratar. estamos na obrigação de radicalizá-la ao nosso solo. de

Prêt C:\Users\ioniel\OneDrive\TCC - LETRAS PORTUGUÊS\LIVRO EM WORD E TXT\TXT\Textos reunidos.par

Fonte: Lexico 3

TEXTO 02

Lexico3 - [TextPloreur]

Fichier Traitement Fenêtre

Navigation | Rapport | Dictionnaire | Segments repetes

Selectionnez une couleur

Recherche:

Formes (ordre lexicométrique)	Frequence
de	196
e	175
a	168
o	160
que	115
—	87
um	63
uma	62
não	57
da	49
do	49
para	46
com	45
me	43
minha	43
os	42
em	40
eu	39
as	38
meu	35
se	34
por	32
ao	28
era	26
no	26
dos	24
quando	24
como	23
na	23
mais	22
à	19
sem	17
das	15
isto	14

2271 formes

a tristeza, a compreensão e a desigualdade de nível mental do meu meio familiar, agiram sobre mim de modo curioso: deram-me anseios de inteligência. meu pai, que era fortemente inteligente e ilustrado, em começo, na minha primeira infância, estimulou-me pela obscuridade de suas exortações. eu não tinha ainda entrado para o colégio, quando uma vez me disse: você sabe que nasceu quando napoleão ganhou a batalha de marengo? arregalei os olhos e perguntei: quem era napoleão? um grande homem, um grande general... e não disse mais nada. encostou-se à cadeira e continuou a ler o livro. afastei-me sem entrar na significação de suas palavras; contudo, a entonação de voz, o gesto e o olhar ficaram-me eternamente. um grande homem!...

o espetáculo do saber de meu pai, realçado pela ignorância de minha mãe e de outros parentes dela, surgiu aos meus olhos de criança, como um deslumbramento. pareceu-me então que aquela sua faculdade de explicar tudo, aquele seu desembaraço de linguagem, a sua capacidade de ler línguas diversas e compreendê-las constituíam, não só uma razão de ser de felicidade, de abundância e riqueza, mas também um título para o superior respeito dos homens e para a superior consideração de toda a gente.

sabendo, ficávamos de alguma maneira sagrados, dedicados... se minha mãe me parecia triste e humilde — pensava eu naquele tempo — era porque não sabia, como meu pai, dizer os nomes das estrelas do céu e explicar a natureza da chuva... foi com estes sentimentos que entrei para o curso primário. dediquei-me afoadamente ao estudo. brilhei, e com o tempo foram-se desdobrando as minhas primitivas noções sobre o saber.

acentuaram-se-me tendências; pus-me a colimar glórias extraordinárias, sem lhes avaliar ao certo a significação e a utilidade. houve na minha alma um tumultuar de desejos, de aspirações indefinidas. para mim era como se o mundo me estivesse esperando para continuar a evoluir...

ouvira uma tentadora sibila falar-me, a toda a hora e a todo o instante, na minha glória futura. agia desordenadamente e sentia a incoerência dos meus atos, mas esperava que o preenchimento final do meu destino me explicasse cabalmente. veio-me a pose, a necessidade de ser diferente. relaxei-me no vestuário e era preciso que minha mãe me repreendesse para que eu fosse mais zeloso. fugia aos brinquedos, evitava os grandes grupos, punha-me só com um ou dois, à parte, no recreio do colégio; lá vinha um dia, porém, que brincava doidamente, apaixonadamente.

causava com isso espanto aos camaradas: oh! o isaías brincando! vai chover...

a minha energia no estudo não diminuiu com os anos, como era de esperar; cresceu sempre progressivamente. a professora admirou-me e começou a simpatizar comigo. de si para si (suspeito eu hoje), ela imaginou que lhe passava pelas mãos um gênio. correspondi-lhe à afeição com tanta força d'alma, que tive cúmes dela, dos seus olhos azuis e dos seus cabelos castanhos, quando se casou. tinha eu então dois anos de escola e doze de idade. daí a um ano, sai do colégio, dando-me ela, como recordação, um exemplar do poder da vontade, luxuosamente encadernado, com uma dedicatória afetuosa e lisonjeira. foi o meu livro de cabeceira. li-o sempre com mão diurna e noturna, durante o meu curso secundário, de cujos professores, poucas recordações importantes conservo hoje. eram banais! nenhum deles tinha os olhos azuis de dona ester, tão meigos e transcendentes que pareciam ler o meu destino, beijando as páginas em que estava escrito!...

quando acabei o curso do liceu, tinha uma boa reputação de estudante, quatro aprovações plenas, uma distinção e muitas sabatinas ótimas. demorei-me na minha cidade natal ainda dois anos, dois anos que passei fora de mim, excitado pelas notas ótimas e pelos prognósticos da minha professora, a quem sempre visitava e ouvia. todas as manhãs, ao acordar-me, ainda com o espírito acariciado pelos nevoentos sonhos de bom agouro, a sibila me dizia ao ouvido: vai, isaías! vai!...

isto ami não te hasta... vai para o rio!

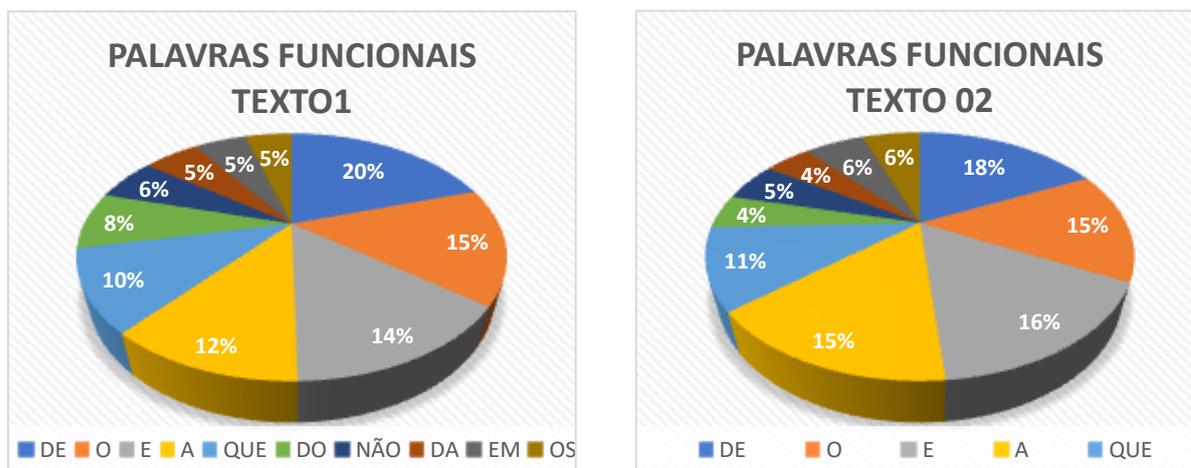
(C:\Users\ioniel\OneDrive\TCC - LETRAS PORTUGUÊS\LIVRO EM WORD E TXT\TXT\5000 palavras - Recordações do escrivão Isaías Caminha.par

Fonte: Lexico 3

Notamos que, embora se tratando de textos de gêneros literários diferentes, as principais palavras que se repetem nos textos são palavras funcionais, chamadas assim pelo fato delas por si só não carregarem valor semântico. Embora essa característica esteja presente na maioria dos textos de qualquer autor, dado que tais elementos textuais, em sua maioria, servem de conectivos e são responsáveis por dar coerência e coesão aos textos.

Por serem utilizadas inconscientemente, as palavras funcionais são muito importantes para analisar o perfil de escrita de cada autor, pois são palavras que ele utiliza sem ao menos se dá conta, pelo hábito e pelo fato de tal palavra está inserida no seu rol de palavras. Chama atenção o fato da palavra funcional mais utilizada em ambos os textos serem a mesma: “DE”, e apresentarem-se com um percentual de ocorrências muito semelhantes.

Veremos nos gráficos abaixo que quando se toma como base as 10 (dez) primeiras palavras funcionais de forma decrescente que aparecem nos textos é possível observar que os gráficos são muito semelhantes. Ao analisarmos a palavra “DE” constatamos que os textos apresentam praticamente a mesma porcentagem de repetições, que giram em torno de 20% (vinte por cento) das palavras analisadas. É importante ressaltar ainda que das dez primeiras palavras funcionais usadas pelo autor do TEXTO 01, oito delas estão presentes também no TEXTO 02, as exceções são as palavras “UM” e “UMA”.



Fonte: Lexico 3

Quando levamos em consideração as 100 (cem) primeiras palavras mais frequentes, nota-se que as palavras funcionais praticamente são as mesmas, com algumas alterações. O que diverge quando observamos essas palavras mais utilizadas são as palavras lexicais, o que é explicado pelo fato dos textos apresentarem formas divergentes e se tratarem de temáticas diferentes, visto que o livro *Sátiras e outras subversões* é composto por 164 pequenos contos, enquanto o livro *Recordações do Escrivão Isaias Caminha* trata-se de um romance.

Chama a atenção nestes dados a frequência em que os “*Hapaxes*”³ aparecem nos textos, em que mais uma vez há uma semelhança significativa nos dois textos, em que a diferença de ocorrência não ultrapassa 2% (dois por cento), um percentual que chama atenção dada a diferença nos gêneros literários das obras. Essa informação ajuda a entender o estilo dos escritores nos dois textos.

TEXTO 01 HAPAX		TEXTO 02 HAPAX	
OCORRÊNCIAS	PORCENTAGEM	OCORRÊNCIAS	PORCENTAGEM
2349	46,98%	2266	45,32%

Fonte: Lexico
3

Das 5.000 (cinco mil) palavras analisadas em cada texto, foi possível observar que cerca de 45% (quarenta e cinco por cento) das palavras usadas são “hapaxes”. Isso demonstra que o(s) auto (es) têm uma maneira de escrever muito semelhante.

Considerações finais

Neste trabalho abordamos o tema análise estilística e estilométrica computadorizada, no qual foi feita uma análise com base em dados gerados com o auxílio de alguns programas de computador, em especial o *Lexico 3*, que foi muito importante na geração de dados.

Concluimos, com base na análise dos dados, que os dez primeiros textos analisados possivelmente tenham mesmo sido escritos por Lima Barreto, o que vem ao encontro com a autoria atribuída por Felipe Botelho Corrêa, no ano de 2016, aos textos que foram por ele organizados no livro *Sátiras e outras subversões*, visto que ao se comparar as obras que compuseram o *corpus*, os dados que foram obtidos são muito semelhantes, quando analisamos estatisticamente.

Com base nos objetivos podemos afirmar que este trabalho cumpriu com o principal objetivo traçado que foi comprovar a autoria dos textos trabalhados, uma vez que os dados coletados foram suficientes para fazê-lo. E ainda provar que o método

³ Termo que aparece uma única vez em um determinado *corpus*.

aplicado tem o respaldo de estudiosos respeitados, portanto, são legítimos e confiáveis, bem como os mecanismos tecnológicos aplicados são eficientes e seguros.

O universo literário está cheio de escritos cuja autoria seja questionável, sendo assim, muito se tem a ser investigado. Cabe aos futuros estudiosos embrenharem-se por esses caminhos, buscando desvendar os segredos guardados por trás de algumas obras.

Nesse entendimento, essa pesquisa é de fundamental importância, pois se configura como um norte para aqueles estudantes que quiserem enveredar por esse universo fascinante da investigação de autoria de obras literárias.

Referências

BARRETO, L. *Sátiras e outras subversões: textos inéditos* / Lima Barreto; organização, introdução, pesquisa e notas Felipe Botelho Corrêa. 1ª ed. São Paulo: Penguin Classics Companhia das Letras, 2016.

BRANDÃO, S. C. de S. *Atribuição de autoria: um problema antigo, novas ferramentas*. Texto Digital, Florianópolis, ano 2, n. 1, Julho 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/textodigital/article/view/1325/1022>>. Acesso em 14 de set. 2016.

CÚRCIO, V. R. *Estudos estatísticos de textos literários*. Texto Digital, Florianópolis, ano 2, n. 2, Dezembro 2006. Disponível em: <https://www.google.com.br/search?q=C%C3%A9RCIO,+V.+R.+Estudos+estat%C3%ADsticos+de+textos+liter%C3%A1rios.+Texto+Digital,+Florian%C3%B3polis,+ano+2,+n.+2,+Dezembro+2006.&gws_rd=cr&ei=kxWbWMjMI8yvvgTEjZ7wCw>. Acesso em 08 de fev. 2017

OLIVEIRA Jr., W. R. *Atribuição de autoria de documentos em língua portuguesa utilizando a distância normalizada de compressão*. Curitiba, 2011. 151 p. Disponível em:

<https://www.ppgia.pucpr.br/pt/arquivos/mestrado/dissertacoes/2011/walter_ribeiro_ver_saofinal.pdf>. Acesso em 30 de ago. 2017

PAIVA, D. M. *Um poeta particular: estudo estilométrico da poesia de H. Dobal*. 2013. 96f. Dissertação (Mestrado em Letras), Universidade Federal do Piauí, Teresina.

MACHADO, S. L. S. *Estudo lexicométrico e estilométrico dos poemas de Fernando Pessoa e seus heterônimos* [manuscrito]. 2014. 86f. Dissertação (Mestrado) – Programa de Pós-Graduação em Letras, Universidade Federal do Piauí, Teresina.

STAMATATOS, E. A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology* (60), p. 538-556, 2009.

Disponível em

:<<https://pdfs.semanticscholar.org/d25c/27c7a3e9f41f150e8eadbad34c1c05d67510.pdf>>

. Acesso em 30 de ago. 2017.