



# O real-naturalismo na literatura pelo viés da Leitura Distante: um estudo do corpus OBras

*Real-naturalism in literature from the perspective of Distant Reading: a study of the corpus OBras*

Antonia Eduarda Trindade da Silva<sup>(a)</sup>; Emanuel Pires de Assis<sup>(b)</sup>

<sup>a</sup> Universidade Federal do Piauí, Piauí, Brasil – dhuda09@hotmail.com

<sup>b</sup> Universidade Estadual do Maranhão, Maranhão, Brasil – emanoel.uema@gmail.com

**Resumo:** Este artigo apresenta uma metodologia com recursos computacionais para a investigação de textos literários. Somado a esse fator, analisamos aspectos da construção de sentido de palavras ditas sintomáticas, conforme pontua Domício Proença Filho (1995), em obras realistas e naturalistas da literatura brasileira. Tais produções literárias estão disponíveis no corpus OBras, que é composto por um vasto conjunto de narrativas em domínio público (Linguatca, 2008). Nesse sentido, observamos o que o crítico literário supracitado afirma ao analisar o romance *O mulato* (1881). Na concepção do estudioso, os substantivos, os verbos e os adjetivos, nessa obra, são palavras sintomáticas, uma vez que criam, no leitor, uma ambiência sugestiva. Tal constatação serve para que o crítico faça um levantamento acerca do Real-Naturalismo brasileiro, em que, segundo ele, os autores transmitiam os aspectos negativos da sociedade em suas obras (Proença Filho, 1995). Para tanto, neste estudo, buscou-se uma base teórica pautada nas Humanidades Digitais, com enfoque no que se tem chamado de Leitura distante (Moretti, 2008). As Humanidades Digitais dizem respeito à transversalidade das áreas do conhecimento e das representações sociais comunitárias dentro do espaço eletrônico digital (Pereira, 2015) e a leitura distante diz respeito a um conceito que propõe a junção entre a computação e dados literários. O termo se refere, ainda, a métodos computacionais diversificados para pesquisas de dados quantitativos em obras literárias. Além da crítica de Moretti (2008), o nosso trabalho fundamenta-se nas acepções de Underwood (2017), Wolfreys (2009), Bastos (2019), Almeida (2019), Alves (2019), Rocha (2019), Ogiboski (2012), Higuchi (2021), Silva (2021), Santos (2019) e Hockey (2018).

**Palavras-chave:** Leitura distante. Real- naturalismo. Literatura brasileira.

**Abstract:** This article presents a computational resources based methodology for the investigation of natural language processing. Added to this factor, we analyzed aspects of the construction of

meaning of words said to be symptomatic, as Domício Proença Filho (1995) points out, in realistic and naturalist works of Brazilian literature. Such literary productions are available in the OBRas corpus, which is composed of a vast set of narratives in the public domain (Linguatca, 2008). In this sense, we observe what the before-mentioned literary critic affirms when analyzing the novel *O mulato* (1881). In the scholar's conception, the nouns, verbs and adjectives in this work are symptomatic words, since they create, in the reader, a suggestive ambience. This observation serves for the critic to make a survey about Brazilian real-naturalism, in which, according to Proença Filho, the authors conveyed the negative aspects of society in their works (Proença Filho, 1995). Therefore, in this study, a theoretical basis based on the Digital Humanities was sought, with a focus on what has been called Distant Reading (Moretti, 2008). Digital Humanities concern the transversality of areas of knowledge and of social communitarian representations within the digital electronic space (Pereira, 2015) and distant reading concerns an attributed concept that proposed the computation of literary data. The term also refers to diversified computational methods for researching quantitative data in literary works. In addition to Moretti's (2008) criticism, our work is based on the meanings of Underwood (2017), Wolfreys (2009), Bastos (2019), Almeida (2019), Alves (2019), Rocha (2019), Ogiboski (2012), Higuchi (2021), Silva (2021), Santos (2019) and Hockey (2018).

**Keywords:** Distant Reading. Real naturalism. Brazilian literature.

## Introdução

Este artigo intenta fazer uma análise tomando como ponto de partida o conceito e a abordagem metodológica da leitura distante. Conhecida pela expressão *distant reading*, tal concepção diz respeito a uma área dentro das Humanidades Digitais que propõe uma leitura não só dos elementos internos das obras literárias, mas, também, dos elementos quantitativos e dos metadados sobre os textos (Moretti, 2008), o que é possível por meio de técnicas computacionais.

Além disso, o interesse por essa abordagem tem origem em três pontos basilares. O primeiro deles parte das observações feitas por Proença Filho (1995), no estudo intitulado *Estilos de época na Literatura*. Nele, o crítico afirma, ao analisar o livro *O mulato* (1881), que os substantivos, os verbos e os adjetivos são palavras sintomáticas, uma vez que criam, no leitor, uma ambiência sugestiva, isto é, um estado mais materializado das coisas e uma

maior concretização da realidade. As afirmações feitas por Filho (1995) consideram não só a obra de Aluísio Azevedo como também o momento social que definia o Realismo e o Naturalismo no Brasil.

O segundo é o interesse pelo tema das Humanidades Digitais, área que nos permite estudar a aproximação entre a Literatura e a técnica computacional. Por isso, nossas análises são pautadas na leitura distante e consideram as concepções de Moretti (2008). Assim sendo, a *distant reading* se refere, também, a uma série de métodos computacionais diversificados para pesquisas de dados quantitativos em obras literárias.

O terceiro ponto de interesse surge dessa aproximação entre a Literatura e a Computação, levando-se em consideração os estudos de Filho (1995) que comparam o Realismo e o Naturalismo brasileiros em poucas obras e toma apenas uma como paradigma das suas observações. Nesse sentido, as afirmativas do crítico são algo passível de análise e, aqui, serão postas à prova por meio da técnica computacional e da busca em um corpus de obras literárias brasileiras.

Para o prosseguimento das análises, em relação ao Realismo e ao Naturalismo, inicialmente, adota-se uma dimensão conceitual e histórica, bem como uma visão de proximidade pautada no diálogo existente entre essas duas escolas literárias, já que ambas aconteceram no mesmo momento de realidade social e cultural no mundo e, principalmente, no Brasil.

Paralelo a isso, faz-se uma análise das duas escolas literárias brasileiras em questão objetivando estudar as afirmações feitas por Filho (1992), inclusive, contemplando a argumentação do crítico sobre o que seria a arte literária naquele momento. Para ele, essa arte é um espelho incumbido de ser o reflexo social de uma época. Assim, o autor, que é também o artista, reflete o cenário social em suas produções. Como se sabe, no Realismo e no

Naturalismo há uma predileção para que as narrativas sejam objetivas, fiéis ao que poderia ser, de fato, a realidade. Dessa forma, para Filho (1995), os adjetivos, assim como os verbos e os substantivos, bem como os outros materiais semânticos da produção analisada por ele, evidenciavam os aspectos sociais negativos (Filho, 1995). Na perspectiva do estudioso, isso faz com que os autores, em suas obras, apresentem uma visão unilateral e pessimista da realidade (Filho, 1995).

No contexto atual, várias são as investigações que avançam e que contribuem para o entendimento do uso de técnicas computacionais aliadas às pesquisas linguístico-literárias. Dentre elas, destacamos os estudos da Linguateca, “Centro de recursos – distribuído – para o processamento computacional da língua portuguesa” (Linguateca, 2008), que, entre outras coisas, se ocupa de desenvolver pesquisas sobre a literatura a partir de um *corpus* vasto de obras brasileiras em língua portuguesa e em domínio público. Além disso, o grupo atua para a formulação de métodos computacionais que visam ampliar as análises literárias, fomentando uma discussão mais pertinente dessa interdisciplinaridade.

Para mais, observamos muitas análises sobre a atuação crescente das Humanidades Digitais baseadas na interdisciplinaridade entre a informática e os estudos literários, e que adotam um panorama dos percursos das áreas tecnológicas, bem como as técnicas de pesquisas da leitura distante. A exemplo desses estudos, citamos a tese de doutorado intitulada *Extração automática de informações: uma leitura distante do Dicionário Histórico-Biográfico Brasileiro (DHBB)*, da pesquisadora Suemi Higuchi (2021). A autora reflete bem essa realidade quando propõe o estudo desses conceitos, cria um *corpus* que viabiliza as leituras distantes dentro do DHBB e faz uma análise da política contemporânea brasileira.

## As humanidades e as ferramentas computacionais

Um dos primeiros conceitos da disciplina das Humanidades é oriundo do termo *humanitas*, ou humanidades, cujo surgimento é anterior à informática e à internet. Em conformidade com o que aponta a crítica, sua origem se deu por volta do século II d. C, quando surgiram indícios de um primeiro registro no livro *Noites Áticas*, de Aulo Gélíio (*apud* Martire, 2017). No que se refere à noção de humanidades, este autor argumenta que "a busca desse tipo de conhecimento, e a formação dada por ele, foram concedidas, dentre todos os animais, ao homem apenas, e por isso são denominadas *humanitas*, ou 'humanidade'" (*apud* Martire, 2017, p. 23).

Por outro lado, as Humanidades, em termos digitais, são datadas da segunda metade do século XX, e o seu surgimento está atrelado ao advento das tecnologias de comunicação e da informática. Especificamente, atribui-se o primeiro trabalho dessa nova área de estudos e pesquisa científica a Roberto Busa, que, na intenção de indexar os trabalhos de Tomás de Aquino, procurou a ajuda do fundador e proprietário da IBM, Thomas J. Watson. A intenção de Busa era fazer um trabalho catalogado, quantitativo, qualitativo e interpretativo dos textos de Aquino. Assim, surgiu o *Index Thomisticus*, que se tornou público e foi disponibilizado<sup>1</sup> (Martire, 2017), no ano de 2005, para qualquer usuário ou estudioso ter acesso.

Nesse mesmo contexto, o computador e a internet são objetos que mediam as relações entre um sujeito e o ambiente ao seu redor, proporcionando, ao primeiro, a possibilidade de exercer alguma ação eficiente sobre o segundo, explica Gomes (2008). Dessa forma, a tecnologia proporciona às humanidades uma vasta rede de atividades que facilitam o acesso a informações. Dentre elas, destaca-se a leitura distante, cujo método

---

<sup>1</sup> <http://www.corpusthomisticum.org/it/index.age>

proporciona a aplicação de técnicas computacionais a dados literários. Por isso, nesse meio, as pesquisas ganham dinamicidade e modernizam a busca do pesquisador e da coisa pesquisada.

Para mais, também observamos que os estudos literários não estão tão distantes das atividades promovidas pela computação; ao contrário, pois a estreita relação entre as duas áreas promove uma expansão de suas reflexões. Além disso, o advento computacional, mediante a modernização da tecnologia e da internet, bem como o aumento dos seus usos, possibilitou, à humanidade, nessa nova cultura, a ampliação e a dinamização da busca por conhecimento. Por isso, os estudos interdisciplinares em Humanidades Digitais se tornam, dentro dessas novas linhas de pesquisa, um recurso muito utilizado, embora esses conceitos tenham surgido fora do ambiente digital.

#### *A leitura distante - O conceito de Franco Moretti*

Dentro da área das Humanidades Digitais, o conceito de leitura distante, normalmente, segue sendo associado ao italiano e estudioso Franco Moretti (2000). Este, que propôs um estudo da literatura por meio do suporte computacional, liderou uma série de pesquisas cujas metodologias, à época, foram vistas como polêmicas e desafiadoras. No entanto, o crítico redefiniu novos conceitos dentro das ciências humanas e digitais quando associou a história da Literatura, por ele chamada de sistema-mundo, às ferramentas e aos dados quantitativos virtuais.

Para Moretti (2000), a teoria evolutiva e o sistema-mundo davam margem a uma compreensão mais alargada da literatura nacional e mundial, pois o mundo se comportava de maneiras diferentes, em épocas distantes, e isso gerava um acúmulo de características passíveis de análises. Dentre essas especificidades, elencam-se as similaridades dos autores, dos leitores, dos estilos de épocas, assim também como as formas de escrita, dentre outros

recursos que poderiam passar despercebidos caso as afirmações fossem feitas com base em um conjunto limitado de textos.

Diante desse cenário, a leitura distante equivale a uma abordagem de estudos que aplica métodos computacionais a dados literários intrinsecamente ligados à história da literatura, à crítica literária, ao campo da tecnologia, à cultura, à Geografia, à linguística, bem como ao processo de decisão do leitor nas suas escolhas – os romances a serem lidos –, e aos métodos quantitativos associados às humanidades como um todo. A partir das definições da leitura distante, suscitadas no Encontro de Leitura Distante em Português (ELD), realizado em 2019, na Universidade de Oslo, temos:

Por oposição à leitura próxima, ou seja, a leitura de aspectos e elementos internos de uma dada obra literária, em que a perspectiva qualitativa prima sobre a quantitativa, a leitura distante pretende dar primazia inicialmente à perspectiva quantitativa, nomeadamente ao eleger como objeto de leitura – como faz Franco Moretti no início desse campo de estudos – «todas» as obras de um país, período, continente, testando as caracterizações existentes da História da Literatura, do género literário, da periodização literária, da estilística, entre outras (Santos *et al.*, 2020, p. 281).

Em seu artigo, *Conjectures on World Literature*, Moretti (2000) afirma, ainda, que o método de leitura distante incluiria obras que se encontram fora das conjecturas dos grandes cânones literários. Essas produções literárias eram denominadas, por ele, como "*Matadouro da Literatura*", termo que, pouco tempo depois, o próprio crítico substituiu, adotando, então, a expressão "o grande não lido" para se referir às obras não canônicas. Por isso, essa abordagem propõe uma inovação na maneira de se adentrar nos estudos históricos e críticos da Literatura mediante o uso de dados estatísticos, de amostras, de paratextos, de metadados, de gráficos e de algoritmos. Dessa forma, a leitura distante segue aliada à leitura atenta das

obras, a fim de ampliar a gama de discussão sobre as análises dos textos literários.

Nesse íterim, conforme explicita Ted Underwood (2017), a leitura distante tem raízes bem diferentes, provenientes de antes da modernidade dos processamentos de dados, cujas décadas são anteriores ao advento da internet e do computador. Com isso, constata-se que essa grandiosa genealogia suscitada por Moretti (2000) não está especificamente preocupada com a ascensão dos computadores. Além disso, essa área de estudo não deve ser preconizada como algo absolutamente novo, mas como um modelo que, com a informática, obteve uma maior dimensão de suas abordagens.

Embora a leitura distante já fosse preconizada por outros críticos, para Moretti (2008) ela passa a significar algo menos limitado e mais abrangente no que toca o número de obras literárias analisadas. Assim, o crítico conceitua a *distant reading* como uma leitura de muitos textos, formando um determinado grupo composto por várias narrativas. Essa abrangência – e consequentemente as análises – só se faz possível com a aplicação de ferramentas computacionais. Nessa empreitada, torna-se necessário que os textos, nas análises, estejam com poucas ou nenhuma divisão, sem rupturas nas leituras. Além disso, seria preciso, também, realizar a separação dos termos pelo método computacional, a fim de que os resultados indicassem, nessas obras, uma relação com o problema pesquisado.

Para Bruno Mattos (2019), esse método envolve, de certa maneira, os aparatos tecnológicos e os bancos de dados textuais para análises de grandes volumes de obras literárias. Dessa forma, principalmente por não se prender apenas à leitura aproximada de poucas obras (Mattos, 2019), uma investigação mais abrangente sobre os aspectos a serem analisados



nos *corpora*, possível com o suporte da ferramenta tecnológica, é necessária. Por isso, é preciso ter conhecimento de que, nessas análises, os resultados serão as probabilidades e as frequências de quantidade menores e maiores das aparições dos fenômenos filtrados, que são indicados pelos termos utilizados e aliados à técnica computacional, a partir de diferentes pontos analíticos como: distinções de personagens, de autores, de palavras e de espaços.

### *A crítica de Domício Proença Filho sobre a Literatura realista-naturalista no Brasil*

Após essas breves contextualizações, é necessário esclarecer por que delimitamos a crítica de Proença Filho (1995) para este trabalho. Assim, vale ressaltar que escolhemos a crítica de Proença Filho (1995) para cerne dessa pesquisa pelo teor do estudo estilístico que ele faz das escolas literárias Realista e Naturalista brasileiras. Por isso, utilizamos a leitura distante para confrontar a ideia estilística de análise que ele realiza.

Dessa forma, analisamos que sendo a literatura a arte da palavra, estudar aspectos dessa magnitude como algo central nessas escolas literárias nos leva à seguinte reflexão: é possível classificar o Real-Naturalismo recorrendo à análise de um recorte semântico de uma única obra literária como faz Proença Filho (1995)? Obviamente, percebemos que nem todos os problemas que emergem das teorias literárias, referentes a essas estéticas, cabem na visão do crítico, visto que de uma resposta surgem novas lacunas ainda mais complexas.

Somado a isso, notamos que os realistas e os naturalistas, ao influenciarem seus leitores sobre a lógica racional e desagradável, como aponta Proença Filho (1995), submergem sob a ótica de um mundo conflituoso com a realidade. Assim, os aspectos da racionalidade mais aceitável por esses autores são os que se referem ao exercício do transcrever, fielmente, a

solidez dos fatos (Pellegrini, 2018). No entanto, considerando que os aspectos em análise também constroem uma crítica à complexidade de se definir o que seria o Realismo e o que seria o Naturalismo (Filho, 1995), o desafio a ser definido, de início, parte do fato de que “ser real” significa, essencialmente, “viver sem ilusões sem se tornar desiludido” (*apud* Soares, 2000, p. 38).

Além desses aspectos, é pertinente destacar que Proença Filho (1995) desconsidera, em sua crítica, o fato de que o engajamento social, em muitos momentos, poderia intermediar a divulgação dos livros das estéticas brasileiras realista e naturalista. Assim, os escândalos que tanto incomodavam os costumes morais, a ética, bem como os temas de cunho patológico e as denúncias das mazelas cotidianas, tão corriqueiras nas obras em questão, rendiam uma maior visibilidade para as produções literárias, para os autores. O crítico também não aponta que fatores como o instinto de sobrevivência e/ou a adaptação dos sujeitos ao meio em que estão inseridos, a zoomorfização e as correntes científicas e filosóficas da época faziam com que tanto o Realismo quanto o Naturalismo buscassem a vida humana como matéria de exploração para as suas narrativas.

Além disso, ao contrário de Proença Filho (1995), outros críticos – a exemplo de Afrânio Coutinho (2003) – não só apontavam que o Naturalismo também era Realismo como acreditavam que, no Brasil, a dificuldade em separar e identificar os estilos literários surgia dos vários movimentos literários que se misturavam e se confundiam. Nesse cenário, havia o florescimento de outros movimentos literários como o Parnasianismo e o Simbolismo, na segunda metade do século XIX. Atrelado a esses fatores, enquanto a maior parte da crítica se preocupava em definir conceitos e distinguir as escolas literárias, críticos como Coutinho (2003) argumentavam que o estudo da Literatura era ciência e, assim sendo, necessitaria de métodos científicos em suas investigações.

## **A literatura sob o viés da tecnologia**

Partindo do que já foi exposto, usaremos um conjunto de características sintático-semânticas para a classificação dos substantivos, dos verbos e dos adjetivos a fim de categorizar as distinções e as semelhanças entre o Realismo e o Naturalismo brasileiros a partir de um conjunto de obras, colocando em análise as afirmações do crítico Proença Filho (1995). Nesse sentido, a nossa pesquisa aproxima a literatura das ferramentas digitais, pois utilizaremos técnicas de classificação da linguagem, com o intuito de verificarmos em que medida essas características e as complexidades do Real-Naturalismo condizem ou não com as reflexões e as análises de Proença Filho (1995), bem como se é possível obter dados que podem representar essas escolas literárias.

Também é de interesse deste trabalho verificar se os resultados que obtivemos com o uso da ferramenta digital e técnicas de processamento da linguagem (PLN) são pertinentes aos estudos literários e se podem ser usados para responder a questões no âmbito desses estudos. Mediante esses resultados, analisaremos formas que concordem com a autoridade literária, como também estudaremos os métodos de leitura distante que contribuem para a investigação dos problemas.

Nessa conjectura, a leitura distante nos oferece vantagens no estudo dessas duas estéticas literárias, pois, como estamos analisando o Real-Naturalismo brasileiro na visão de uma crítica, em específico, é oportuno começarmos mediante um número expressivo de obras que estão dentro e fora do cânone literário.

## **Os aspectos metodológicos da investigação**

Na área da Literatura e da estilometria textual<sup>2</sup>, pesquisar as interfaces da Leitura Distante é muito significativo para o desenvolvimento de mais pontos de vistas críticos sobre as hipóteses aqui levantadas. Almeja-se expandir a gama de estudos nessa ambiência, e, por isso, neste trabalho, as relações sociais e as incidências de casos que se referem ao sema “saude:doença” comporão a nossa metodologia. Isso será feito a partir de análises de textos do *corpus* OBras<sup>3</sup>. A nossa intenção, nessa conjectura, é estudar a relação de sentido das obras para que possamos distinguir os aspectos que definem e diferenciam o real-naturalismo, no Brasil, sob o viés do crítico Proença Filho (1995).

Assim, o *corpus* de análise escolhido trata-se de um *corpus* dinâmico, que recebe, constantemente, a inserção de novos textos em sua base e que possui um número expressivo de obras brasileiras em domínio público, 323 até o momento da escrita deste artigo. Dentre esses textos, há uma significativa quantidade de produções literárias realistas e naturalistas.

Partindo disso, como ferramenta de pesquisa, escolhemos o AC/DC, sistema que permite buscas por concordâncias de vocábulos, por categorias gramaticais, por escolas literárias, dentre outras. Para mais, esse sistema já possui o *corpus* OBras na sua interface. Além do *corpus* de análise e da ferramenta utilizada, nesta seção apresentam-se os procedimentos de coleta e as categorias de análises selecionadas dentro do AC/DC para executar a investigação de obras do Real-naturalismo brasileiro presentes no *corpus*. Dessa forma, para alcançar os objetivos propostos, foi feita uma pesquisa descritiva, de natureza qualitativa, cujo

---

<sup>2</sup> A estilometria textual é uma área dos estudos linguísticos e literários que avalia o estilo de um autor, ou de uma determinada escola literária, por meio da aplicação de análise estatística a um *corpus* de obras. Isso é feito com o auxílio da ferramenta computacional que faz a medida da escrita textual. Para considerações sobre essa técnica que também pode ser chamada de leitura distante, ver, por exemplo: Assis (2013) e Assis e Lopes (2019).

<sup>3</sup> Disponível em: <https://www.linguateca.pt/acesso/corpus.php?corpus=OBRAS>.

objetivo é selecionar um material semântico dentro do universo temático do sema “saude:doenca”.

Logo, tendo por base o nosso foco investigativo, os pressupostos teóricos da leitura distante e os grupos lexicais explicitados por Proença Filho (1995) – que aponta a relevância de vocábulos como pontos de influência na identificação de características semelhantes e diferentes entre o Realismo e o Naturalismo na literatura brasileira –, o percurso prático foi percorrido de acordo com o modelo de análise da linguagem do AC/DC. Dessa forma, as categorias dessa ferramenta selecionadas foram as seguintes: busca por distribuição da categoria gramatical (PoS), distribuição por autores (autor) e distribuição pela corrente literária (escola). Nesse sentido, elege-se um cotejo entre as afirmações de Proença Filho (1995), o *corpus* OBras e a busca das categorias de palavras com a utilização da ferramenta AC/DC.

#### *As funções de busca na ferramenta AC/DC*

As funções de buscas semânticas e a utilização do sistema AC/DC permitem que sejam consultados os fenômenos pretendidos anotados no *corpus*, bem como as suas frequências, que podem ser de um fenômeno ou mais (Linguatca, 2022). Dessa forma, para Alberto Simões (2014, p. 02), o AC/DC é “um serviço que disponibiliza o acesso a todos os corpos cujos donos nos deem autorização e o AC/DC é o local privilegiado para a comparação de corpos (ou melhor de pesquisas em corpos)”.

Logo, o AC/DC torna-se uma ferramenta acessível, à disposição para pesquisadores. E estes não precisam fazer qualquer tipo de cadastro para utilizá-la. Tal ferramenta permite a realização de buscas no *corpus* OBras, assim como em outros *corpora* alimentados pela Linguatca, por: concordância, distribuição das formas (word), distribuição dos lemas (lema), distribuição da categoria gramatical (PoS), distribuição do tempo

verbal e/ou do caso pronominal ([temcagr](#)), distribuição de pessoa e/ou número ([pessnum](#)), distribuição do gênero morfológico ([gen](#)), distribuição da função sintática ([func](#)), distribuição pelas obras (obra), distribuição por autores (autor), distribuição por gênero de texto (classe), distribuição pela corrente literária (escola), distribuição pelo sexo do entrevistado, do biografado ou do autor (sexo), distribuição por texto original ou traduzido (oritrad), distribuição por campo semântico (sema), distribuição por grupo (de cor, roupa, etc.) (grupo) e distribuição das dependências (dependências).

Além das supracitadas, na interface dessa ferramenta o pesquisador é informado de que pode obter outros resultados optando por buscas nas seguintes opções de categorias: resultados por ordem alfabética (só distribuições), resultados em formato separado por ponto e vírgula, ignorar maiúsculas/minúsculas (não admite parâmetros) e fazer uma nuvem com limite de espaçamentos utilizando a categoria de buscas: “de\_\_\_\_com amostra aleatória de\_\_\_\_ linhas”.

### *O corpus do projeto OBras*

Segundo Flávia Silva (2021, p. 12), “um *corpus* é um grande conjunto de enunciados ou textos que representa uma fonte de material ocorrido naturalmente”. O propósito do OBras é a criação de um *corpus* da literatura brasileira em língua portuguesa. O OBras, segundo seus idealizadores, tem o compromisso de atualizar e inserir em sua base as obras da literatura brasileira que vão surgindo (Linguatca, 2022). Em razão disso, o projeto está em constante atualização. Para mais, a Linguatca esclarece que se responsabiliza, também, por entrar em contato com as pessoas que forneceram listas de obras para esse processo, e por informar às outras plataformas que venham a fazer a digitalização da atualização de novas

obras. Por conseguinte, torna público o nome da equipe e dos responsáveis por cada obra coletada (Linguateca, 2022).

Atualmente, a equipe do OBras é formada, segundo as disposições da página, pelos professores e pesquisadores Alberto Simões, Cláudia Freitas, Diana Santos, Emanuel Pires, João Marques Lopes e Marcia Langfeldt. A pesquisadora Flávia Silva (2021) aponta que o “*corpus* OBras também é parte integrante da Gramateca e da Literateca, projetos da Linguateca para os estudos gramaticais e literários, respectivamente, possíveis em virtude da estrutura do AC/D”.

Diante da existência de vários recursos da Linguateca, dos outros *corpora* e de algumas listas distintas de referências bibliográficas em todos os projetos literários da Linguateca, o grupo intitula os trabalhos que se embasam em textos literários de [Literateca](#) (Linguateca, 2008).

Como o projeto é de domínio público, ele é aberto para as colaborações de pesquisadores que se interessam em contribuir, de alguma maneira, com os estudos das Humanidades Digitais. Alguns dos principais propósitos do projeto centram-se em expandir as análises da Linguística, da Literatura e da crítica cultural (LINGUATECA, 2022).

### *O sema de análise*

Conforme informado, a unidade de significação desta pesquisa é composta por metadados que nos remetem ao sema “saude:doenca”, no corpus OBras. Isso porque este sema comporta material semântico que evidencia as palavras sintomáticas, conceito usado por Domício Proença Filho (1995) para se referir aos substantivos, aos verbos e aos adjetivos em suas análises estilísticas. Logo, as contribuições desse sema servirão, posteriormente, para o reconhecimento de características que permitirão

identificar o Realismo e o Naturalismo brasileiros no *corpus* em questão, tendo em vista a crítica de Proença Filho (1995).

Tendo por base esse cenário, fizemos um levantamento das doenças que surgiram face a esse contexto, de modo a evidenciar como essas enfermidades eram retratadas nas obras literárias brasileiras pertencentes ao Realismo e ao Naturalismo. A historiografia aponta que a Segunda Revolução Industrial foi marcada por uma significativa desorganização, com as ruas das grandes cidades desproporcionais ao número de tráfego, o que fez com que as metrópoles se tornassem propensas a um acúmulo de sujeira, conseqüentemente, ocasionando o aumento do número de pessoas doentes.

Por outro lado, em que medida essas mazelas, postas como características, são capazes de refletir distinções relevantes entre as escolas literárias realista e naturalista? Tal problematização explica a escolha do sema “saude:doença”, que se justifica por observamos que, nesse contexto, existia um aglomerado de bairros e cortiços habitacionais, de indústrias e de estradas que iam sendo construídas. Paralelo a isso, veículos com animais eram utilizados para o transporte de cargas e de pessoas. Então, tendo por base a crítica suscitada por Proença Filho (1995), observamos que à medida que a sociedade brasileira se industrializava, a proliferação de novas doenças, espalhadas de maneira rápida, refletia o lado mais cruel e negativo dessa modernização.

Assim, como fazemos uso do AC/DC para investigar essas características, a nossa pretensão não é provar que a informática solucionará os problemas quanto à identificação e/ou diferenciação das escolas literárias em questão, mas, sim, explorar e analisar a informação que a ferramenta fornece. Por isso, adentraremos o funcionamento do programa e inicialmente, se almejávamos identificar a distribuição do campo



semântico nos textos do OBras, foi necessário pensar como seria possível identificar o sema “saude:doenca”, no *corpus*, e se esse campo semântico poderia ser gramaticalmente classificado em substantivos, verbos e adjetivos. Além disso, a nossa intenção é, também, apontar características relevantes que possam contrastar com a crítica de Proença Filho (1995). Para tanto, foi preciso observar como identificar o campo semântico no *corpus* e se seria possível classificar a classe gramatical dentro do contexto em que esse sema aparece, valendo-nos da seguinte categoria de buscas do AC/DC: distribuição da categoria gramatical (PoS) (pos). Após longo estudo, propomos a possibilidade de utilização do indicador fornecido na ferramenta digital:

a) Campo saúde indicativo de doença:

[sema="\*.saude:doenca.\*"]

Nessa empreitada, observamos os seguintes pontos para as análises dos textos disponíveis no OBras, utilizando o sema “saude:doenca”:

1. Como, tendo em vista a crítica de Domício Proença Filho, os substantivos os adjetivos e os verbos contribuem para o reconhecimento de características que permitem identificar o Realismo e o Naturalismo no corpus OBras, utilizando, para tanto, a leitura distante?
2. Quais pressupostos teóricos da leitura distante auxiliam na compreensão das características do Realismo e do Naturalismo levando-se em consideração a crítica de Proença Filho (1995)?
3. Quais grupos lexicais extraídos do *corpus* OBras concretizam os aspectos que definem as escolas literárias realista e naturalista com base na crítica de Proença Filho (1995)?

4. Como o campo semântico influencia na identificação de características semelhantes e diferentes entre o Realismo e o Naturalismo a partir da crítica de Domício Proença Filho (1995)?

Após a marcação dos nossos objetivos, seguimos, então, para as buscas e, consequentemente, para os resultados tendo como foco somente as obras brasileiras realistas e naturalistas inseridas no *corpus* de estudo.<sup>4</sup>

### **As análises: panorama na categoria gramatical (pos): substantivos**

Para a obtenção dos resultados, prosseguimos com os seguintes fatores que caracterizavam os adjetivos, os verbos e os substantivos no sema “saude:doenca”:

- 1- **Obtenção de frequências:** ocorrências do sema “saude:doenca” nas classes dos adjetivos, dos verbos e dos substantivos.
- 2- **Obtenção dos dados:** possibilidades de identificação das características semelhantes e diferentes do Realismo e do Naturalismo na literatura brasileira nos textos do *corpus* OBras sob o viés do crítico Domício Proença Filho (1995).
- 3- **Resultados obtidos:** listas das frequências para a concretização das análises nas categorias pesquisadas.

Dessa forma, para que a próxima etapa de análise se concretizasse, foi fundamental organizar as ocorrências do nosso eixo semântico, ou seja, as frequências do sema “saude:doenca” nas categorias pesquisadas. Para isso, obtivemos os seguintes dados como resultados:

---

<sup>4</sup> Para uma compreensão sobre a atribuição das escolas literárias, consultar Santos *et al.* (2020b).

a) Distribuição da categoria gramatical (PoS) ([pos](#))

Figura 01 - Imagem do corpus OBras com os indicativos do sema saude:doenca para as classes gramaticais: substantivos, verbos e adjetivos

Procura: [sema="saude:doenca.*"] Distribuição de pos Corpo: OBras v. 13.0 7636 casos.	
<b>Distribuição</b> Houve 12 valores diferentes de pos.	
N	3832
ADJ	3325
V	410

Fonte: Linguateca

Esses marcadores indicam as seguintes classes gramaticais e as possíveis indicações de subclasses:

Tabela 01 - Indicadores de classe e possíveis subclasses

Nome	Classe gramatical	Indicações de subclasse
N	Substantivo	Prop
ADJ	Adjetivo	NUMord, prop, KOMP, n, mente
V	Verbo	n, prop, fmc, quant

Fonte: Linguateca

b) Distribuição por autores (autor)

A imagem a seguir apresenta os resultados encontrados na categoria de busca por autores para o sema “saude:doenca”. Nessa busca, delimitamos os autores das escolas literárias brasileiras realista e naturalista presentes no *corpus* OBras.

**Tabela 02** - Indicativos do sema “saude:doenca” distribuído por autores do Realismo e do Naturalismo brasileiro disponíveis no corpus Obras<sup>5</sup>

Procura: [sema="\*.saude:doenca.\*"]  
 Distribuição de autor  
 Corpo: OBras v. 13.6  
 7636 casos.  
 Houve 61 valores diferentes de autor.

Marcador do autor	Resultados do sema “saude:doenca”	Número total de vocábulos
MacAss	1587	2208939
AluAze	627	708340
RodThe	188	95904
DomOli	168	150361
RauPom	146	206812
JulLAlm	107	129362
CosVel	90	69598
IngSou	76	166073
AdoCam	60	94863
JulRib	55	65657
EmiBdM	51	44061
ArtAze	15	32770

**Fonte:** Linguateca

Esses marcadores indicam os seguintes autores:

<sup>5</sup> Esses valores correspondem à quantidade de vezes que cada autor apresentou para o sema de busca da nossa pesquisa. Dessa forma, o valor maior é o total de todos os vocábulos nas obras de cada autor e o valor menor é a quantidade apresentada apenas para as palavras específicas do sema “saude:doenca”.

Tabela 03 - Indicadores para buscas de autores

Marcador	Autor
AdoCam	Adolfo Caminha
AluAze	Aluísio Azevedo
CosVel	Cosme Velho
DomOli	Domingos Olímpio
EmiBdM	Emília Bandeira de Melo
IngSou	Inglês de Sousa
JulLAlm	Júlia Lopes de Almeida
JulRib	Júlio Ribeiro
MacAss	Machado de Assis
RauPom	Raul Pompéia
RodThe	Rodolpho Theophilo
ArtAze	Artur Azevedo

Fonte: Linguateca

c) Distribuição pela corrente literária (escola)

Figura 02 - Imagem do corpus OBras com os indicativos do sema “saude:doenca” distribuído por escolas literárias

Procura: [sema="\*,\*saude:doenca.\*"] (1913 casos aleatórios)  
 Distribuição de escola  
 Corpo: OBras v. 13.0  
 7636 casos.

#### Distribuição

Houve 24 valores diferentes de escola.

desc	2698
realismo	1633
romantismo	868
naturalismo	690
realismo_naturalismo	483
modernismo	178
realismo_regionalismo_romantismo	140
naturalismo_realismo	132
histórico	131
romantismo_regionalismo	127
romantismo_realismo_naturalismo	105
neoparnasianismo	73
realismo_romantismo	62
naturalismo_realismo_romantismo	61
arcadismo	53
impressionismo_naturalismo_realismo_simbolismo	46
parnasianismo	31
naturalismo_regionalismo	30

Fonte: Linguateca

A discussão dos resultados, das características e das diferenças das escolas literárias brasileiras realista e naturalista no *corpus* OBras, sob o viés da crítica estilística de Proença Filho (1995), comporá a base das análises nas listas de frequências e dos gráficos nas categorias pesquisadas.

## As análises

### *Panorama na categoria gramatical (PoS): substantivos, verbos e adjetivos*

As ocorrências que serão analisadas, aqui, estão disponíveis na página da Linguateca (<https://www.linguateca.pt/>). Embora o campo semântico analisado esteja dentro do sema “saude:doenca”, isso não significa dizer que, obrigatoriamente, remeta somente à área da saúde e das doenças, porque pode, também, fazer parte de outros semas de pesquisa. Assim, os nossos estudos investigam esses resultados quanto às classificações referentes às classes gramaticais considerando o que pontua a crítica de Proença Filho (1995). Traremos o campo semântico do sema, que está distribuído nas seguintes categorias: <sup>6</sup>

<sup>6</sup> Disponível em: <https://www.linguateca.pt/Gramateca/Saude.html>

a) saude:saude

salubre, saudável, saúde, são

b) saude:acessorio

bisturi, estetoscópio, sanguessugas, termómetro

c) saude:medicina

anestesia, anestestista, botica, boticária, boticário, cardiologia, cardiologista, cirurgia, cirurgião, cirurgia, cirúrgico, consulta, consultório, curandeira, curandeiro, dentista, enfermagem, enfermaria, enfermeira, enfermeiro, facultativo, farmacêutico, farmácia, fármaco, ginecologia, ginecologista, hospital, hospitalizar, medicação, medicamentar, medicina, medicinal, médica, médico, operação, operar, obstetra, obstetrícia, odontologia, oftalmologia, oftalmologista, ortodontista, ortopedia, ortopedista, parteira, pediatra, pediatria, profilático, profilático, profilaxia, sanatório, terapêutica, terapia, terma, tratamento, tratar, vacina

d) saude:psico

depressão, deprimido, nervosismo, nervoso, spleen

e) saude:progressao

contagiar, contágio, convalescença, convalescente, cura, curar, curativo, epidemia, melhorar, melhora, melhoria, pandemia, piorar, pioria, recaída, restabelecer, sanar, sequele, surto

f) saude:doenca

acamar, acamado, achaque, alergia, alérgico, adoecer, adoentado, bexigas, bexigoso, câncer, canceroso, cancro, cólera, desmaiar, desmaio, diabetes, diabético, doença, doente, enferma, enfermiça, enfermidade, enfermo, febre, febril, ferida, ferimento, hepatite, histeria, histérico, infecção, infectar, infetar, insalubre, mal, lepra, leproso, malária, maleita, mazela, moléstia, moribunda, moribundo, paludismo, peste, pestilento, pneumonia, pus, pústula, pustulento, raiva, ranho, rubéola, ruséola, sarampo, sarna,

sarnoso, sintoma, symptoma, tifo, tifóide, tinoso, tísico, tosse, tuberculose, tuberculoso, tumor, varíola

g) saude:causa

bactéria, germe, gérmem, micróbio, viral, vírus

h) saude:remedio

antibiótico, aspirina, atadura, bálsamo, bandaid, compressa, drageia, emplastro, injeção, injeção, ligadura, medicamento, mezinha, penso, pílula, pomada, preparado, quinino, receita, remédio, supositório, tala, unguento, xarope

Colocamos essas categorias nos grupos de classes gramaticais que estamos analisando. Em seguida, buscamos investigar, nos textos do *corpus*, o sema “saude:doenca”, a fim de entendermos qual a incidência de suas ocorrências, como isso influenciaria o elo entre texto e sociedade, e se isso leva à área das impressões sintomáticas, conforme sugere a crítica de Proença Filho (1995).

### Substantivos

Para a análise dos substantivos, pesquisamos, no AC/DC, os casos por “Distribuição na categoria gramatical (PoS) ([pos](#))”. Ao fazê-lo, obtivemos os seguintes resultados em forma de lista de ocorrências:

Termo	Ocorrências	Termo	Ocorrências
Bisturi	15	Botica	124
Estetoscópio	1	Boticária	152
Sanguessugas	10	Boticário	152
Termómetro	1	Cardiologia	1
Anestesia	5	Cardiologista	1
Anestestista	1	Cirurgia	13



Termo	Ocorrências
Cirurgião	52
Cirurgiã	1
Cirúrgico	6
Consulta	119
Consultório	36
Curandeira	14
Curandeiro	39
Dentista	66
Enfermagem	1
Enfermaria	72
Enfermeira	77
Enfermeiro	58
Facultativo	33
Farmacêutico	32
Farmácia	85
Fármaco	1
Ginecologia	1
Ginecologista	6
Hospital	177
Hospitalizar	1
Medicação	11
Medicamentar	1
Medicina	182

Termo	Ocorrências
Medicinal	4
Médica	20
Médico	1633
Operação	176
Obstetra	1
Obstetrícia	2
Odontologia	1
Oftalmologia	1
Oftalmologista	1
Ortodontista	1
Ortopedia	1
Ortopedista	1
Parteira	26
Pediatra	1
Pediatrica	2
Profiláctico	1
Profilático	1
Profilaxia	5
Sanatório	1
Terapêutica	18
Terapia	1
Terma	1
Tratamento	193

Termo	Ocorrências
Vacina	5
Depressão	52
Nervosismo	24
Contágio	78
Curativo	26
Epidemia	79
Pandemia	1
Recaída	16
Sequela	1
Alergia	1
Alérgico	1
Bexigas	60
Bexigoso	7
Câncer	2
Cancro	19
Cólera	826
Desmaio	59
Diabetes	5
Doença	382
Doente	1261
Enfermiça	4
Enfermo	422
Febre	839

Termo	Ocorrências
Febril	183
Ferida	273
Ferimento	44
Hepatite	5
Histeria	24
Infeção	2
Infecção	16
Infectar	1
Insalubre	6
Mal	5832
Lepra	50
Malária	3
Maleita	7
Mazela	1
Moléstia	630
Paludismo	21
Peste	247
Pneumonia	14
Pus	101
Pústula	8
Raiva	643
Ranho	2
Antibiótico	1

Termo	Ocorrências
Aspirina	3
Atadura	3
Bálsamo	84
Bandaíd	1
Compressa	3
Drageia	1
Emplastro	6
Injeção	15
Injecção	1
Ligadura	9
Mezinha	20
Pílula	13
Pomada	16
Quinino	27
Receita	110
Remédio	854
Supositório	1
Tala	3
Unguento	2

Termo	Ocorrências
Xarope	40
Rubéola	1
Ruséola	1
Sarampo	7
Sarna	7
Sintoma	65
Symptoma	2
Tifo	9
Tifóide	3
Tísico	60
Tosse	1
Tumor	21
Varíola	3
Bactéria	1
Germe	52
Gérmen	53
Micróbio	10
Viral	1
Vírus	24
Pioria	1

Dispensamos o termo “mal”, pois, em muitos casos analisados, ele se referia tanto às classes gramaticais aqui investigadas quanto a outras não contempladas nesta pesquisa. Para melhor explicitar, o termo também

exerce função de advérbio e conjunção. Nesse sentido, o maior caso de ocorrências no grupo substantivos, em nossas buscas, foi do termo “médico”, com 1633 casos na expressão de busca por [word="médico" & pos="N.\*"]<sup>7</sup>, seguido do termo “moléstia”, com 616 casos e na expressão de busca por [word="moléstia" & pos="N.\*"]. Já outras ocorrências de impressões mais materiais como “bisturi” tiveram como resultado apenas 15 casos na expressão de busca por [word="bisturi" & pos="N.\*"], e “anestesia”, por sua vez, 5 casos na expressão de busca por [word="anestesia" & pos="N.\*"].

Dessa forma, como os substantivos, enquanto classe gramatical, designam e nomeiam os seres e os objetos, sejam eles concretos ou abstratos, é possível perceber que, nessa categoria, tanto as coisas materializadas quanto as não materializadas são passíveis de receberem nomes concretos, e que há uma probabilidade significativa de usos dos termos das áreas não materializadas. Quanto a isso, no conjunto de palavras dessa categoria, percebem-se ocorrências significativas dos seguintes termos: “pus”, 101 casos na expressão de busca por [word="pus" & pos="N.\*"]<sup>8</sup>, “consulta”, 119 casos na expressão de busca por [word="consulta" & pos="N.\*"], “peste”, 247 casos na expressão de busca por [word="peste" & pos="N.\*"], e “doença”, 382 casos na expressão de busca por [word="doença" & pos="N.\*"]. Com exceção de “pus”, todos estes se remetem a substantivos de ordem abstrata, referindo-se, ainda, às áreas ruins ou problemáticas da sociedade, o que intensifica uma tendência dos autores a refletirem as classes menos abastardas que, normalmente, eram as mais acometidas por essas condições.

---

<sup>7</sup> Essa expressão que se repetirá ao longo das análises, diferentes para cada palavra, corresponde ao comando de busca dos vocábulos no programa AC/DC e no *corpus* OBras. Assim, o nome de cada critério dessas buscas corresponde ao termo Lema.

<sup>8</sup> O delimitarmos pela categoria N (substantivo), excluimos as vezes em que a palavra faria referência ao pretérito perfeito do verbo pôr na primeira pessoa do singular.

Posto isso, em relação a essas características, Proença Filho (1995) elege como vocábulos indicativos dessas mazelas sociais, ainda, os substantivos “pedras”, “lampiões”, “vidraças”, “paredes”, “folhas”, “carroças de água”, “aguadeiros” e “pretos”. Este último tanto pode designar um substantivo quanto um adjetivo. E todos, segundo o crítico, traduzem uma preocupação com a realidade observada. Por isso, são elementos de ordem material, situados nas áreas de impressões sensoriais, mostrando a decadência social da cidade de São Luís (Proença Filho, 1995).

Além desses substantivos, Proença Filho (1995) ainda analisa outros, a saber: “‘fígados’, ‘rins’, ‘coração’, o termo ‘Ventrezinhos’ com tom irônico e ‘barriga’” (Proença Filho, 1995, p. 246). Logo, quando comparamos estes últimos termos com vocábulos que encontramos no sema da pesquisa, a exemplo de “bexiga”, “bactéria” e “ferida”, percebemos que há, no Realismo e no Naturalismo brasileiros, áreas que também estão situadas nas escritas de cunho cientificista, cujos autores focalizam as partes do corpo humano, evidenciando a ciência, característica das duas escolas literárias em questão, e a animalização do ser humano, isto é, o indivíduo como um animal pertencente ao plano natural em que todos os bichos convivem.

Somado a esses fatores, quando as afirmativas de Proença Filho (1995) são postas em análise e comparamos ambos os materiais semânticos, os do crítico e os do sema desta pesquisa, observamos, também, que os substantivos se referem tanto à materialização das coisas concretas, aqui já referenciado, – que, segundo a língua portuguesa, são os termos que englobam os objetos com formas definidas – quanto à materialização das coisas abstratas, termos que designam objetos sem forma definida.

Isso evidencia que as afirmativas de Proença Filho (1995) também estão situadas a esses fatos e às lacunas da crítica canônica em relação a essas escolas literárias. A prova disso é que, em termos de probabilidade, além

de o material semântico estudado mostrar uma característica muito marcante dessas escolas – que, no caso, seria a tentativa de ser o mais real e objetivado possível –, os autores não conseguem, por definitivo, separar objetividade e subjetividade dos fatos. Isso porque, quando se observa o cotidiano social, a maior parte dos literatos reflete, em suas obras, personagens psicologicamente confusos aqui evidenciados pelo uso de substantivos abstratos.

### Verbos

Prosseguiremos, agora, com a análise dos verbos. As pesquisas para essa classe gramatical também foram feitas no AC/DC, para obtermos o número de casos nessa categoria, na seção de buscas por “Distribuição na categoria gramatical (PoS) ([pos](#))”, obtivemos os seguintes resultados:

**Tabela 05** - Verbos no sema “saude:doenca” do *corpus* OBras

Verbo	Ocorrências
Curar	115
Melhorar	88
Adoecer	33
Acamar	7
Tratar	401
Operar	25
Restabelecer	37
Sanar	3
Infetar	1

Contagiar 2

Fonte: Linguaterra

Em nossas buscas, o maior número de ocorrências neste grupo foi do termo “tratar”, com 401 casos na expressão de busca por [word="trata" & pos="V.\*"], e “curar”, com 115 casos na expressão de busca por [word="cura" & pos="V.\*"]. Enquanto isso, outros vocábulos que indicam ações de impressões mais sintomáticas, característica do Realismo e do Naturalismo fomentada por Proença Filho (1995), têm menor ocorrência. Como exemplo, destacamos “restabelecer”, “tratar”, “cura” e “sanar” e seus contextos:

a)id="Luzia-Homem Prosa:Romance ou Novela DomOli 1903 naturalismo masc ":

Mas... era forçoso submeter-se à ordem do administrador, tão bom e compassivo, que lhe dera muitos dias de licença para **tratar** a pobre mãe enferma .

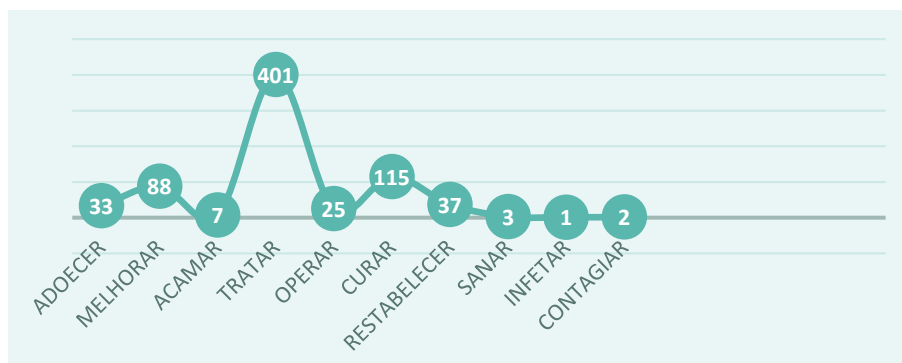
b)id="Luzia-Homem Prosa:Romance ou Novela DomOli 1903 naturalismo masc ":

Uma noite passava ele com o Belota e tivera o atrevimento de fazer-lhe uma serenata cantando à viola, quase no terreiro da casa, modinhas e canções eróticas, que terminavam nesta saudosa endecha: «Vou me embora, vou me embora, Como fez a saracura; Bateu asas, foi cantando: Mal de amores não se **cura!** ... " .

c)id= O\_Cabeleira Prosa:romance FT 1876realismo\_regionalismo

Confiava no poder da autoridade, e tinha por certo que havia de restaurar a tranqüilidade e a segurança privadas, e **restabelecer** o domínio das leis.

**Gráfico 1** - Trajetória dos casos do sema “saude:doenca” no grupo dos verbos no corpus Obras



**Fonte:** Linguateca.

As formas infinitivas verbais são uma das categorias nominais dos verbos da língua portuguesa. É possível verificar que, nesse eixo, os verbos perdem algumas de suas características: deixam de apresentar flexão de tempo e de modo, por exemplo, e assumem o papel, muitas vezes, de um substantivo, sendo este, em muitos casos, concreto ou abstrato.

Fizemos buscas com as formas infinitivas, buscas por lemas, e com as formas conjugadas dos verbos, buscas por palavras do sema “saude:doenca”. Logo, como nas nossas buscas houve um maior número de ocorrências dos verbos no infinitivo, infere-se que, possivelmente e em termos estatísticos, há uma preferência dos autores por uma escrita mais voltada para esta forma nominal dos verbos, o que significa que a escrita privilegia o lado da causa e do efeito, dos nomes e materialização das coisas, bem como dos sentimentos e explicação dos fatos. Tais aspectos são mais recorrentes que o enfoque de formas declinadas dos verbos, por exemplo. Isso evidencia as características dessas escolas literárias, tal como o fato de as personagens estarem entregues aos próprios destinos.

Com relação a essas especificidades, Proença Filho (1995) ainda destaca como vocábulos indicativos de causa e efeito – além dos que já citamos aqui



– os verbos “passavam”, junto ao vocábulo “reverberações”, e seguido da palavra “ruidosamente”, “abalando”, junto ao vocábulo “os prédios”, e, por último, o verbo “invadiam”. Segundo o crítico, essas características revelam uma má vontade dos moradores da cidade de São Luís, além de uma denotação-conotação que levam às áreas de impressões desagradáveis da região urbana da cidade (Proença Filho, 1995). Além desses vocábulos, ainda cita os verbos: “gemer”, “guinchavam” e “fermentava”.

Com isso, quando colocamos as afirmativas de Proença Filho (1995) à prova, observamos os verbos referidos pelo crítico. Dessa forma, quando os comparamos com um *corpus* maior de análise, percebemos que, no Realismo e no Naturalismo brasileiros, as questões de causa e efeito podem, também, estar associadas às atitudes das personagens que são justificadas por ações ocorridas anteriormente.

### Adjetivos

Para as análises dos adjetivos, fizemos o mesmo procedimento realizado anteriormente. Em seguida, pesquisamos a classe gramatical adjetivos no AC/DC para obtenção do número de casos nessa categoria, na mesma seção de buscas das classes anteriores – verbos e substantivos – e por “Distribuição na categoria gramatical (PoS) ([pos](#))”. Ao fazê-lo, obtivemos os seguintes resultados:

**Tabela 06** - Adjetivos no sema “saude:doenca” do corpus OBras

Adjetivo	Ocorrências
Convalescente	62
Acamado	6
Canceroso	1

Enferma	240
Histórico	15
Leproso	34
Moribunda	113
Moribundo	197
Pestilento	8
Tinhoso	21
Tuberculoso	24
Saudável	36
São	6418

**Fonte:** Linguateca

Dispensamos a referência ao vocábulo “são”, pois ele apresentou mais de uma classe gramatical nos textos. Assim, focamos em analisar apenas os casos referentes aos adjetivos. As ocorrências em nossas buscas no que se refere a essa classe gramatical mostram menções a termos como “moribundo”, que apresenta 197 casos pela expressão de busca por [word="moribundo" & pos="ADJ.\*"], e “leproso”, que contém 34 casos, expressão de busca por [word="leproso" & pos="ADJ.\*"]. Por outro lado, de impressões mais sintomáticas, característica do Realismo e do Naturalismo fomentada por Proença Filho (1995), a palavra “tuberculoso” teve como resultado apenas 25 ocorrências na expressão de busca por [word="tuberculoso" & pos="ADJ.\*"] com a variação de “tuberculosa” na expressão de busca por [word="tuberculosa" & pos="ADJ.\*"], e canceroso, 1 caso, na expressão de busca por [word="canceroso" & pos="ADJ.\*"]. Os

exemplos de obras que apresentaram muitas ocorrências para o sema desta pesquisa, a seguir, demonstramos alguns casos contextualizados:

a)id="Casa\_de\_Pensão Prosa:romance AA 1884 realismo masc ":

Cochichava-se pelos cantos, em magotes, discreateando-se projetos em voz de mistério, como se tratasse de um **moribundo**.

b)id="Dom\_Casmurro Prosa:romance MdA 1899 realismo masc ":

Morreu afinal, como os Estados morrem; no nosso caso particular, a questão é saber não se a Turquia morrerá, porque a morte não poupa a ninguém, mas se os russos entrarão algum dia em Constantinopla; essa era a questão para o meu vizinho **leproso**, debaixo da triste rota e infecta colcha de retalhos...

c)id="Casa\_de\_Pensão Prosa:romance AA 1884 realismo masc ":

Dr. Freitas deixou-se levar, sempre muito enfasiado; mas, antes de ir, bateu no ombro de Amâncio e segredou-lhe com a sua voz de **tuberculoso**:...

d)id="A\_afilhada Prosa:novela MdOP 1899 naturalismo masc ":

Pelas paredes sombrias, quadros do Juízo Final, da Morte do Justo, da horrenda Morte do Pecador; e ao lado da porta principal, fechada, que dava para a rua, escancarado, o armário dos milagres, ostentando per-rias de elefantíases, dependuradas, mãos inchadas e em pústulas, braços cortados, cabeças de crânio roído, ou de boca torta, ou de nariz **canceroso**, muletas, representações de naufrágios, em desenho rijo, ventres dilacerados, retratos desfigurados, peitos lancetados, pescoços escrofulosos, lombos mirrados pela tísica: a carne nua e podre na mais horrível confissão da miséria humana e do milagre divino.

Na obra *Casa de pensão*, de Aluísio Azevedo, também foi possível identificar um número significativo de adjetivos para o nosso sema de análise. Por

isso, também delimitamos dois trechos dessa narrativa para exemplificar os resultados na classe gramatical em questão. Esses excertos correspondem ao trecho A e ao trecho C, conforme disposto acima. É possível observar, a partir deles, a preferência pela utilização de adjetivos com a terminação -oso. Isso acontece não só nas obras de Aluísio Azevedo, pois também é recorrente nas produções literárias de Machado de Assis que analisamos.

Nesse sentido, no trecho A, da obra *Casa de pensão*, observamos o adjetivo “moribundo”. Esse termo tem origem no latim, *moribundus*. Enquanto classe gramatical, ele caracteriza algo ou alguém que está prestes a morrer. É oportuno observar que esse vocábulo é composto lexicalmente pela junção do termo “mor” ou “mortis”, que significa a morte, e do sufixo “bundus”, que também tem origem no latim e designa algo muito forte. Em termos de medicina, o moribundo seria um paciente que está no estágio terminal, ou seja, à beira da morte.

Entretanto, no excerto A, observamos que a palavra não está empregada em seu sentido literal, mas, sim, em um sentido figurado, dando a entender que, aparentemente, alguém se apresenta de forma descuidada e, por isso, parece um moribundo. Além desse sentido, para o vocábulo “moribundo” foram encontrados, ainda, trechos em que a palavra está empregada em sua forma material e indicando doença. Dentre eles, destaca-se o da obra *O coruja*, também de Aluísio Azevedo: “Assentou-se ao lado da cama e ajudou o **moribundo** a segurar uma vela de cera, que lhe haviam posto entre as mãos extensas e descarnadas” (Azevedo, 1889, p 45). Por conseguinte, essa obra apresentou como resultado para o sema desta pesquisa 81 casos sob 129655 vocábulos do material semântico da obra.

Para Proença Filho (1995), os adjetivos, nas produções desses autores, estão incumbidos de mostrar uma ambiência sugestiva. O crítico chega a

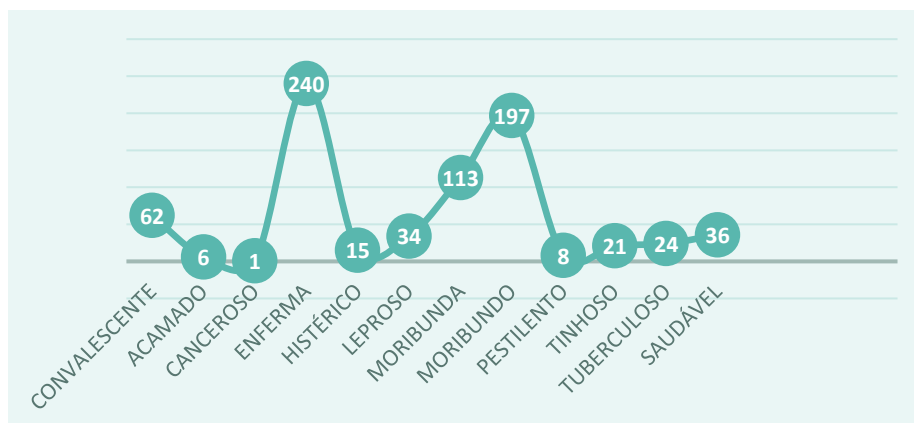
essa conclusão quando analisa alguns dos adjetivos presentes na obra *O mulato*, dando exemplos de vocábulos como “adormecido”, “fúnebre”, “sujo”, “seboso”, dentre outros (Proença Filho, 1995, p. 248). No entanto, se observamos que os adjetivos são a classe gramatical que atribui referência aos substantivos, como é possível que o material semântico das obras realistas e naturalistas indique somente sugestões aviltantes?

Nessa empreitada, parece-nos que o lado negativo e decadente da sociedade que fora evidenciado está ligado ao surgimento de novas classes sociais: médias e baixas. Esta, formada por sujeitos de tipos comuns, é representada por grupos que seriam facilmente encontrados nos cortiços, nas pensões e nas vielas. E eles surgem em uma representação geográfica das paisagens, isto é, em trechos que falam sobre os problemas urbanos das cidades transfiguradas. Vale acrescentar que esses grupos sempre existiram, com outros nomes, mas eram invisibilizados.

Assim, observamos que – para demonstrar essas novas tendências de escritas, que refletiam também as novas classes sociais, em sua grande maioria as menos abastadas – os autores do Realismo e do Naturalismo brasileiros optavam por fazer uso de adjetivos que indicassem atributos de rejeição e/ou negação.

Na esteira disso, o gráfico abaixo demonstra os adjetivos referentes ao sema “saude:doença”, no *corpus* OBras. Como resultado, há, nas narrativas, uma probabilidade maior do uso de adjetivos que caracterizam e qualificam as personagens em seus aspectos desfavoráveis.

**Gráfico 2** - trajetória dos casos do sema “saude:doença” no grupo dos adjetivos no corpus Obras



Fonte: Linguateca.

Os adjetivos, enquanto classe gramatical, são vocábulos que caracterizam os objetos e os seres, mas também podem se juntar aos substantivos, modificando os seus significados e os seus contextos. Dessa forma, os adjetivos podem ser constituídos e acrescidos das terminações -ado, -oso, -a e/ou -o, dentre outros. Em nossas buscas, além da preferência de escrita com adjetivos terminados em -oso, observamos, também, muitas ocorrências dessa classe gramatical com as terminações em -o e -a.

Proença Filho (1995) observa que os adjetivos conferem um sentido de ambientações sugestivas, sendo que, nos textos desses autores, eles remetem a recortes de paisagens. Como exemplo, ele situa alguns adjetivos, a saber: “miserável”, que caracteriza o objeto “casebre”, “tísica”, que se refere à “voz”, “gentil” e “bela”, que qualificam “Carolina”, personagem de *O Mulato*, “tortas”, que se refere ao substantivo “pernas”, e “maternas”, que qualifica “ilhas” (Proença Filho, 1995).

Além desses adjetivos, Proença Filho (1995) ainda situa os vocábulos “amarelento”, indicando a aparência desagradável de uma personagem, os

termos “vermelho” e “afogueado”, que se referem às pessoas brancas, bem como o termo “preto”, que designa pessoas negras (Proença Filho, 1995). Ainda situando o espaço da obra, o crítico apresenta como exemplos passagens que nos remetem à decadência do local: “‘arroz de Veneza’, ‘os cães uivam com uivos que pareciam gemidos humanos’ e ‘um cheiro acre de sabão da terra e aguardente’” (Proença Filho, 1995, p. 247).

Dessa forma, mais uma vez, ao focalizarmos as afirmativas de Proença Filho (1995), observamos, agora, os adjetivos colocados nas análises do crítico. Assim, quando os comparamos com o material semântico do *corpus* OBRAS, percebemos que, no Realismo e no Naturalismo brasileiros, além dos juízos de valor, como aponta Proença Filho (1995), nota-se, também, que os autores transmitiam as características das novas realidades em termo de novas tendências dos grupos sociais, por meio de construções das imperfeições que qualificam as coletividades.

### **O sema “saude:doenca” na distribuição pelos autores do Realismo e do Naturalismo brasileiro no *corpus* de análise**

Tendo em vista a utilização da leitura distante nas obras realistas e naturalistas brasileiras, esse método mostrou-se eficaz, uma vez que possibilitou a delimitação dos dados necessários para esta pesquisa de maneira constante e necessária. Isso ocorreu ampliando as formas de análises das obras literárias e tornando acessível o conjunto de recursos disponíveis na interface no AC/DC, junto à Liguatca. Logo, é perceptível que a leitura distante intermediou a análise quantitativa e qualitativa, análise estilística, dessas obras quando nos permitiu comprovar que a leitura aproximada dos textos, aliada a dados computacionais, aumentou as nossas análises e nos permitiu distinguir características mais exatas das frações nos resultados que obtivemos quando fizemos as buscas na ferramenta.

Dessa forma, de um ponto de vista estilístico, segundo Diana Santos (2019), a fração semântica por frase é um elemento que costuma ser utilizado na diferenciação de autores e de obras, ou por ser tendenciosa a apontar outros assuntos e não o que estava em alta, ou por demonstrarem aquilo que se é esperado. Como exemplo disso, as obras do escritor Joaquim Nabuco apontam um estilo de cunho abolicionista, mas ele não era colocado pela crítica como um escritor do Realismo ou do Naturalismo. De modo diferente, temos as obras de Julia Lopes de Almeida, que era naturalista e não divergia dessa corrente literária. Esses dados foram obtidos com a leitura distante.

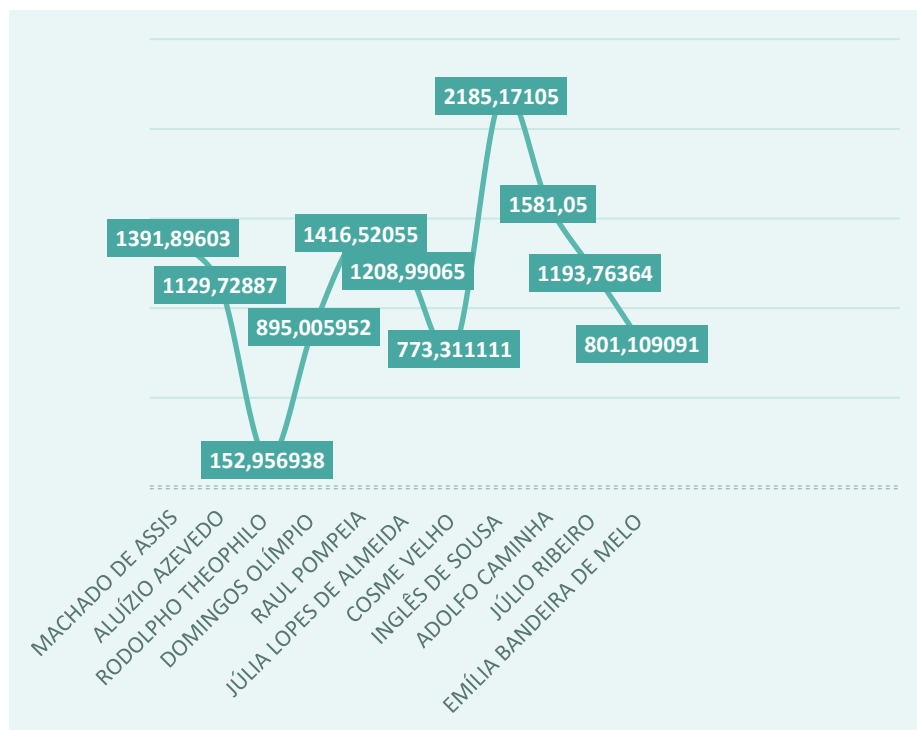
Com isso, como eram muitos os autores que divergiam das características de escrita do Realismo e do Naturalismo, fizemos, aqui, uma diferenciação por autores e buscamos as frequências do sema “saude:doenca” no *corpus* OBras, sempre observando o conjunto semântico nos resultados dos substantivos, dos verbos e dos adjetivos, e até que ponto essas características confirmam ou não a crítica de Proença Filho (1995). Para isso, delimitamos apenas os autores do Realismo e do Naturalismo brasileiros presentes no OBras e utilizamos como ponto teórico dessa análise a teoria evolutiva<sup>9</sup> proposta por Franco Moretti (2000). A imagem a seguir demonstra os resultados dessas frequências:

---

<sup>9</sup> A teoria evolutiva proposta por Franco Moretti (2000), uma das bases de obtenção de dados da leitura distante, analisa que as literaturas e os autores literários sofrem modificações ao longo do tempo considerando não só a mudança social de cada época, mas também, a experiência do autor o qual teve que conviver para sobreviver. Para tanto, nesse contexto, Moretti (2000) utiliza o modelo de árvores, a exemplo de Darwin que com essa metodologia explica a evolução das espécies, para mostrar as rupturas dos romances.



**Gráfico 3** - Frequência do sema “saude:doenca” por autores do Realismo e do Naturalismo no corpus Obras<sup>10</sup>



**Fonte:** Linguateca.

A imagem, que é o resultado dos cálculos das frequências nos casos pesquisados, reflete a busca pelo sema “saude:doenca” para o autor Machado de Assis, do Realismo, e para os autores do Naturalismo brasileiro que estão presentes no *corpus* de estudo. Como usamos a leitura distante, além dos recursos de processamento da linguagem, temos como vantagem o acesso ao texto em formato digital, o que permite distinguir, a partir dessas frequências, os dois grupos de autores analisados: um realista

<sup>10</sup>Essas frequências foram calculadas mediante os resultados da tabela 3, que contém os resultados de buscas por autor na ferramenta AC/DC. Assim, foi utilizado os valores maiores de cada autor, que são as características em todas as suas obras no corpus, dividido pelo valor menor, que representa o que cada autor aqui analisado mostrou na busca apenas do sema “saude:doenca”. Ou seja, o valor total de ocorrências de cada escritor foi dividido pelo valor dos casos do sema.

e outro naturalista, como também a propensão dos autores que refletem as duas escolas literárias.

Porque estamos analisando a literatura realista e naturalista brasileira, tendo por base o *corpus* OBras, investigamos, aqui, os estilos dos autores e não os gêneros de escrita produzidos por eles. Para essas análises, calculamos o conjunto semântico do sema dos textos para cada autor, gráfico 3, o que demonstra caracteres de interesse para uma possível descrição do estilo literário realista e naturalista que passamos a elencar.

No grupo de maior tendência de cunho realista, temos Machado de Assis. No naturalista, por sua vez, os seguintes autores: Aluísio Azevedo, Adolfo Caminha, Inglês de Sousa, Cosme Velho, Domingos Olimpo e Emília Bandeira de Melo. Porém, destacamos que algumas obras podem apresentar mais de uma escola literária. Assim, além dessas distinções, vale mencionar que os autores Júlio Ribeiro, Raul Pompeia e Júlia Lopes de Almeida apresentam traços tanto do Realismo quanto do Naturalismo, e Rodolpho Theophilo possui traços do regionalismo e do Naturalismo.

Por esse prisma, observa-se que autores consagrados pela crítica canônica – a exemplo de Machado de Assis – apresentam um bom índice de propensões para o sema “saude:doenca”. Entretanto, se considerarmos que Raul Pompeia demonstra tanto traços do Realismo quanto do Naturalismo, observaremos que, em termos estilométricos, ele ultrapassa Machado de Assis nos resultados do sema desta pesquisa. Isso porque, na sua frequência de menções para o “saude:doenca”, Raul Pompeia evidencia maiores índices, resultados, da escola literária referida. Diferente disso, é o caso da autora Júlia Lopes de Almeida. Embora ela também apresente propensões das duas escolas literárias, a frequência para o nosso sema de busca nessa escritora é bem menor com relação a Machado de Assis e Raul Pompeia.

Nas análises realizadas, observamos, ainda, que autores lidos em menor frequência, como Inglês de Sousa, apresentam uma probabilidade maior de propensões do sema de pesquisa. Esse material evidencia traços mais expressivos do Naturalismo, enquanto autores como Aluísio Azevedo, que é descrito pela crítica canônica como o precursor da escola literária naturalista, no Brasil, apresentou propensões menores do sema de pesquisa.

É importante observamos que estas características reforçam que a *distant reading* também pode ser utilizada nas análises por autores. Uma vez que esse método fornece dados nessa categoria. Um exemplo de estudo nessa modalidade, seria as investigações dos escritores tendo por base suas obras e as disseminações dos seus textos nas épocas de suas publicações. Assim, tendo por base essa metodologia de Moretti (2000), observamos que um fator muito importante para que os autores de ambas as escolas literárias Realistas e Naturalistas brasileiras fossem mais conhecidos seria a forma como esses escritores atraíam a atenção dos leitores para que estes lessem suas obras.

A exemplo dessas formas de propagação dos textos, os autores se colocavam o mais próximo possível do que de fato acontecia no cotidiano das pessoas. Isso é bastante observado nas obras de Aluísio Azevedo e Raul Pompeia que em sua grande maioria traziam narrativas que refletiam o cotidiano dos cortiços, dos pobres, das violências urbanas, dentre outros. Atrelado a esses fatores e com base na perspectiva de Moretti (2000), observamos que Proença Filho (1995), ao diferenciar o Realismo e o Naturalismo analisando trechos da obra de Aluísio Azevedo, aponta probabilidades de palavras dentro de uma ambiência sugestiva para o leitor, a partir de um material semântico que contempla verbos, substantivos e adjetivos e mostram o cotidiano no espaço geográfico da cidade de São Luís. Isso, conforme argumenta o crítico, mesmo tendo como

base um contexto de análise pequeno para essa conclusão, leva à área de impressões sintomáticas.

Em contraposição às afirmativas de Proença Filho (1995), observamos que o crítico, quando estuda trechos da obra *O Mulato*, analisa somente o material semântico de ambientação do espaço, e esses caracteres apontam, também, outras características da estética naturalista, a exemplo da forte influência do escritor Emile Zola sobre a escrita de Aluísio Azevedo, que abordava, em suas narrativas, os fatos sociais, a influência do meio e o desenvolvimento humano. Além disso, notamos que, se não há uma análise de *O mulato* como um todo, e tampouco um estudo de outras obras de Aluísio Azevedo, os traços mais recorrentes da escrita azevediana – a exemplo dos costumes da sociedade de São Luís, bem como da sociedade brasileira, as habitações, as análises das relações e o cotidiano de seus habitantes – são deixados de lado.

Dessa forma, analisar apenas trechos de uma única obra literária e de um único autor, como o faz Proença Filho (1995), e categorizar, de maneira conclusiva, o Realismo e o Naturalismo (Proença Filho, 1995) deixa alguns questionamentos a serem estudados. Isso demonstra uma escolha metodológica ainda imprecisa, embora o crítico não desacredite outros autores que são menos lidos e menos abordados pela crítica. E essa discordância já é apontada pelo próprio Proença Filho (1995), que reconhece que a sua visão é limitada, uma vez que gera uma divergência de hipóteses entre a parte analisada, cheia de lacunas de caracterização, e o que não foi analisado.

### **Considerações finais**

Nesse artigo, a nossa proposta foi colocar a crítica do estudioso Proença Filho (1995) em debate, e os nossos resultados se mostraram bem-sucedidos. Além disso, este trabalho contribuirá como um recurso a mais

para as pesquisas em estudos literários, ao passo que, também, colaborará com outras áreas do conhecimento como a das Humanidades Digitais. Concordamos com Mattos (2019) ao afirmar que:

A interdisciplinaridade inerente nas Humanidades Digitais, principalmente pela união da ciência da computação com as humanidades, trouxe o desenvolvimento de novos conceitos e metodologias para o campo e ajudou no avanço de algumas das principais linhas de pesquisas da área: conteúdo digital; marcação e codificação de texto; visualização e análise de texto; mineração de dados; e Organização do Conhecimento (Mattos, 2019, p.18).

Além das observações de Mattos (2019), percebemos que, junto às Humanidades Digitais, a ciência da computação também pode ser muito eficaz nos estudos críticos de obras literárias. A prova disso é que obtivemos sucesso na utilização de buscas semânticas na expressão do sema “saude:doenca”, a partir de padrões do português junto a outras expressões de buscas que utilizamos. Essas buscas nos indicaram os substantivos, os verbos e os adjetivos em um conjunto de obras da literatura realista e naturalista brasileira disponíveis no *corpus* OBras.

Notamos que o sema de pesquisa, na ferramenta AC/DC, foi capaz de extrair um número significativo de ações nos textos, nos possibilitando, assim, questionar se os substantivos, os verbos e os adjetivos nos levariam às áreas de impressões sintomáticas, e até que ponto as análises de Proença Filho (1995) condiziam ou não com a crítica canônica. Isso foi possível a partir dos resultados que obtivemos nas categorias de buscas e do *corpus* escolhido. Este foi essencial para os resultados, visto que é de acesso livre e já está inserido junto a outros *corpora* na ferramenta AC/DC.

As análises possibilitaram o entendimento de que é fundamental acontecer uma aproximação entre a leitura distante e a leitura aproximada de obras literárias nesses tipos de pesquisas, pois isso viabiliza o alcance dos resultados desejados. A prova disso é que analisamos um *corpus* que

contém 323 textos, sendo que 212 destas obras são do Realismo e do Naturalismo brasileiros, disponíveis em domínio público.

Observamos, ainda, que as classes gramaticais, nestes textos, evidenciam que o crítico Proença Filho (1995), ao concluir que o Realismo e o Naturalismo brasileiros funcionam como duas escolas literárias em que uma potencializa a outra, avulta uma visão que não responde e nem resolve os problemas de categorização das duas escolas literárias aqui estudadas. Além disso, em nossas análises, notamos uma preferência dos autores realistas e naturalistas por uma escrita mais ampliada do real, como uma espécie de denúncia social e não como uma via de produção focada em apenas mostrar o lado desagradável da sociedade, conforme aponta Proença Filho (1995). Isso porque, nos trechos das obras analisadas, a recorrência de autorreflexão das realidades é constante. Junto a esses fatores, observamos, ainda, que as probabilidades que obras literárias não canônicas têm de apresentar características estilísticas das escolas literárias brasileiras realista e naturalista são maiores do que as de narrativas mais conhecidas.

## REFERÊNCIAS

ASSIS, Emanuel Cesar Pires. Atribuição de autoria utilizando análises estatísticas: uma experiência com a relação abreviada. *Texto Digital*, v. 9, n. 1, p. 24-53, 2013.

ASSIS, Emanuel Cesar Pires; LOPES, Daniel. A estatística textual computadorizada e a literatura brasileira: uma análise do romance "Miragem", de Coelho Neto. *Studia Iberystyczne*, v. 18, p. 259-270, 2019.

AZEVEDO, Aluísio de. *O cortiço*. São Paulo: FTD, 1997

AZEVEDO, Aluizio de. *O mulato*. Teresina: Ática, 1994.

BARREIROS, Liliane Lemos Santana. *O VOCABULÁRIO DE EULÁLIO MOTTA*, 2017.

BOSI, Alfredo. *História Concisa da Literatura Brasileira*. São Paulo: Cultrix, 1994.

COUTINHO, Afrânio. *A literatura no Brasil: relações e perspectivas / conclusão*. 7 ed. São Paulo: Global Editora, 2003.

GOMES, Francisco Wellington Borges. *Linguagem e tecnologia: desafios para o ensino de línguas na sociedade da informação*. 2008.

HIGUCHI, Suemi. Do texto ao dado: debates sobre leitura distante nas humanidades. *Revista D2G: Humanidades Digitais*, Braga, v. 3, n. 2, 2021.

HOBSBAWM, E. J. A revolução industrial. In: HOBSBAWM, E. J. *A era das revoluções: 1789 - 1848*. São Paulo: Paz e Terra, 2015.

FRICK, Jean-Paul. *Auguste Comte, ou La République positive*. Nancy: Presses universitaires de Nancy, 1990.

LINGUATECA. Centro de recursos -distribuído- para o processamento computacional da língua portuguesa. Linateca. Disponível em: [www.linguateca.pt/Literateca/](http://www.linguateca.pt/Literateca/). Acesso em: 10 fev. 2023.

MARTIRE, Alex da Silva. *Ciberarqueologia em Vipasca: o uso de tecnologias para a reconstrução-simulação interativa arqueológica*. Museu de Arqueologia e Etnologia da USP, MAE, Brasil. 2017

MATTOS, Bruno Marques de. *A Organização do Conhecimento nas Humanidades Digitais e o Conceito de Leitura Distante de Franco Moretti*. Pontifícia Universidade Católica de São Paulo (PUC-SP), 2019.

MORETTI, Franco. *A literatura vista de longe*. Trad. Anselmo Pessoa Netto. Porto Alegre: Arquipélago Editorial, 2008

MORETTI, Franco. Conjecturas sobre a literatura mundial. In: SADER, E. (org.). *Contracorrente: o melhor da New Left Review em 2000*. Trad. Maria Alice Máximo et al. São Paulo: Record, 2000.

PELLEGRINI, Tânia. *Realismo e realidade na literatura: um modo de ver o Brasil*. São Paulo: Alameda Casa Editorial, 2018.

PELLEGRINI, Tânia. *Realismo: Postura e Método*. São Paulo: UFSCaR, 2007.

PROENÇA FILHO, Domício. *A Linguagem Literária*. São Paulo: Ática. 2001.

PROENÇA FILHO, Domício. *A linguagem literária*. São Paulo: Ática, 1992.

PROENÇA FILHO, Proença. *Estilos de Época na Literatura*. São Paulo: Ática, 1995.

RADWAY, JA. "Culturas Impressas Literárias e Eruditas em uma Era de Profissionalização e Diversificação". *Print In Motion: A Expansão da Publicação e Leitura nos Estados Unidos, 1880-1945* (Outono, 2008).

SANTOS, Diana et al. Leitura Distante em Português: resumo do primeiro encontro. *MatLit* (Centro de Literatura Portuguesa da Universidade de Coimbra), v. 8, n. 1, p. 279-298, 2020a.

SANTOS, Diana et al. Periodização automática: Estudos linguístico-estatísticos de literatura lusófona. *Linguamática*, v. 12, n. 1, p. 80-95, 2020b.

SILVA, Flávia Martins da Rosa Pereira da; FREITAS, Cláudia. Representações de gênero na caracterização de personagens: uma proposta metodológica e primeiros resultados. Rio de Janeiro, 2021. 166 p. Dissertação de Mestrado – Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro.

SOARES, Rosemary Dore. Gramsci, o Estado e a escola. *Revista Brasileira de Educação*, n. 19, 2000. Disponível em: [www.scielo.br/j/rbedu/a/C73fCRDCpbr9n54CbNrCyZr/](http://www.scielo.br/j/rbedu/a/C73fCRDCpbr9n54CbNrCyZr/). Acesso em: 10 set. 2022.



SOUSA, Maria Clara Paixão de. As Humanidades Digitais Globais. In: *Ciclo de Conferências: Congresso Humanidades Digitais em Portugal* (Universidade Nova de Lisboa, 8/10/2015, Conferência de abertura), CIDEHUS (Universidade de Évora, 6/10/2015), Programa Materialidades da Literatura (Universidade de Coimbra, 12/10/2015). 2015.

UNDERWOOD, Ted. A genealogy of distant eading. *Digital Humanities Quartely*. v.11, n. 2, 2017. Disponível em:  
<http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>.  
Acesso: 10 set. 2022.

WILLIAMS, Raymond, *Culture and Society*. João Pessoa: Paz & Terra, 1992.

#### NOTAS DE AUTORIA

**Antonia Eduarda Trindade da Silva** (dhuda09@hotmail.com): Mestranda do Programa de Pós-Graduação em Letras, Universidade Federal do Piauí (UFPI). Possui graduação em Letras-Língua Portuguesa, Literatura Brasileira e Portuguesa, pela Universidade Federal do Piauí (UFPI)-Campus Ministro Petrônio Portela, Centro de Ciências Humanas e Letras (CCHL) e Especialização em Literatura, estudos culturais e outras linguagens pela Universidade Estadual do Piauí (UESPI)- Campus Pirajá. Foi representante discente das bolsas no Colegiado de Letras do Programa de Pós-graduação em Letras- PPGEL(UFPI). Interessa-se pelos temas relacionados à literatura de Língua Portuguesa, teoria, história e crítica literária, relações entre as acepções teóricas literárias e a tecnologia, as Humanidades Digitais e a literatura fantástica. Atualmente, é membro dos grupos de pesquisa NUPLID e LAMID e coordena o grupo de discussão em Leitura Distante. Tem experiência na área de Letras e Literatura Fantástica, com ênfase em Literatura Tecnológica e Digitalizada, construção de simulacro digital de leitura de obras literárias brasileiras, bem como, pesquisas e monitorias em educação e também apresenta uma dinâmica de trabalho voltada para a escola, a didática e a cidadania.

**Emanoel Pires de Assis** (emanoel.uema@gmail.com): Possui graduação em Letras Licenciatura em Português/ Inglês e respectivas literaturas, pela Universidade Estadual do Maranhão - UEMA Campus Caxias (2008). Mestre em Letras - Estudos Literários - pela Universidade Federal do Piauí- UFPI. Doutorado em Literatura pela Universidade Federal de Santa Catarina - UFSC. Tem experiência na área de Letras, com ênfase em Teoria Literária e Literatura em Meio Digital. Atualmente, interessa-se pelo estudo de narrativas digitais, literatura brasileira contemporânea, ferramentas digitais para o ensino/aprendizagem de literatura e Teoria Literária. Líder do grupo de pesquisa CNPq: Literatura, Arte e Mídias - LAMID. Editor Gerente da Revista de Letras Juçara- UEMA. Bolsista de produtividade em pesquisa - UEMA 2018-2019 e 2021-2022. Professor

Adjunto II na Universidade Estadual do Maranhão - UEMA, está como vice-coordenador do Programa de Pós-graduação em Letras da UEMA e atua também como professor permanente no programa de pós-graduação em Letras da UFPI.

**Como citar este artigo de acordo com as normas da revista?**

SILVA, Antonia Eduarda Trindade da; ASSIS, Emanuel Pires de. O real-naturalismo na literatura pelo viés da Leitura Distante: um estudo do corpus OBras. *Texto Digital*, Florianópolis, v. 19, n. 1, p. 108-157, 2023.

**Contribuição de autoria**

Antonia Eduarda Trindade da Silva: concepção e elaboração do manuscrito; análise de dados; discussão dos resultados; revisão e aprovação.

Emanuel Pires de Assis: concepção e elaboração do manuscrito; análise de dados; discussão dos resultados; revisão e aprovação.

**Financiamento**

Não se aplica.

**Consentimento de uso de imagem**

Figura 01 - Imagem do corpus OBras com os indicativos do sema saúde:doença para as classes gramaticais: substantivos, verbos e adjetivos. Fonte: Languateca.

Figura 02 - Imagem do corpus OBras com os indicativos do sema “saúde:doença” distribuído por escolas literárias. Fonte: Languateca.

**Aprovação de comitê de ética em pesquisa**

Não se aplica.

**Licença de uso**

Este artigo está licenciado sob a Licença Creative Commons CC-BY. Com essa licença você pode compartilhar, adaptar, criar para qualquer fim, desde que atribua a autoria da obra.

**Histórico**

Recebido em: 04/07/2023.

Aprovado em: 10/07/2023.