RGe ISSN: 2177-3246

MÚLTIPLOS OLHARES SOBRE O CERRADO: UMA ANÁLISE COMPUTACIONAL POR MINERAÇÃO DE TEXTOS NA PRODUÇÃO ACADÊMICA HISTORIOGRÁFICA

MULTIPLE PERSPECTIVES ON THE CERRADO: A COMPUTATIONAL ANALYSIS THROUGH TEXT MINING IN ACADEMIC HISTORIOGRAPHICAL PRODUCTION

MÚLTIPLES MIRADAS SOBRE EL CERRADO: UN ANÁLISIS COMPUTACIONAL A TRAVÉS DE LA MINERÍA DE TEXTOS EN LA PRODUCCIÓN ACADÉMICA HISTORIOGRÁFICA

10.56238/revgeov17n6-105

Ronaldo Ferreira da Silva
Doutorando pelo Programa de Pós-graduação Territórios e Expressões Culturais no Cerrado (PPGTECCER)
Instituição: Universidade Estadual de Goiás (UEG)
E-mail: ronaldofsilva1@gmail.com

Edvilson Cerqueira Silva Júnior
Mestrando pelo Programa de Pós-graduação Territórios e Expressões Culturais no Cerrado (PPGTECCER)
Instituição: Universidade Estadual de Goiás (UEG)
E-mail: juniorlogi93@gmail.com

Poliene Soares dos Santos Bicalho
Docente pelo Programa de Pós-graduação Territórios e Expressões Culturais no Cerrado (PPGTECCER)
Instituição: Universidade Estadual de Goiás (UEG)
E-mail: poliene.bicalho@ueg.br

Fernando Lobo Lemes
Docente pelo Programa de Pós-graduação Territórios e Expressões Culturais no Cerrado (PPGTECCER)
Instituição: Universidade Estadual de Goiás (UEG)
E-mail: fernando.lemes@ueg.br

RESUMO
A busca por padrões e tendências na produção historiográfica sobre a ocupação e colonização do Cerrado nas narrativas acadêmicas, por meio da hermenêutica tradicional, constitui-se uma tarefa complexa e desafiadora, considerando a limitação cognitiva humana em analisar, com precisão, um extenso volume textual. Por outro lado, a técnica denominada Leitura Distante ou *Distant Reading* criada por Franco Moretti, propõe distanciar-se do objeto de leitura, não como um obstáculo, mas como uma forma de conhecimento, extraindo menos detalhes e mais relações, padrões e formas em uma obra ou entre obras. Esta técnica encontra suporte operacional nas Humanidades Digitais, área interdisciplinar que adota métodos e técnicas computacionais para apoiar o desenvolvimento de

estudos e análises nas áreas de ciências sociais e humanas. Esta pesquisa propôs analisar um conjunto de documentos históricos, organizados em um recorte específico, para encontrar padrões e tendências nas narrativas acadêmicas reproduzidas em um corpus textual, utilizando um grupo de ferramentas e técnicas computacionais por meio da mineração de textos. Os resultados auferidos no estudo, por meio de procedimentos sistematizados em uma Prova de Conceito, sintetizados quantitativamente, mostraram que a utilização de tecnologias digitais pode contribuir significativamente para análise, interpretação e compreensão de documentos históricos e apontar tendências nas narrativas acadêmicas. A mineração de texto confirmou as hipóteses levantadas sobre a mudança semântica de "Sertão" para "Cerrado", com a finalidade de reinventar discursivamente o território para categoria econômica e produção de *commodities* e o distanciamento vultoso das narrativas entre perspectivas desenvolvimentistas e a inserção das Comunidades Tradicionais neste processo.

Palavras-chave: Humanidades Digitais. Leitura Distante. Mineração de Texto. Povos Tradicionais.

ABSTRACT
The search for patterns and trends in the historiographical production on the occupation and colonization of the Cerrado in academic narratives, through traditional hermeneutics, constitutes a complex and challenging task, considering the cognitive limitations of humans in accurately analyzing a vast volume of text. On the other hand, the technique called Distant Reading, created by Franco Moretti, proposes distancing oneself from the object of reading, not as an obstacle, but as a form of knowledge, extracting fewer details and more relationships, patterns, and forms within a work or between works. This technique finds operational support in Digital Humanities, an interdisciplinary area that adopts computational methods and techniques to support the development of studies and analyses in the areas of social sciences and humanities. This research proposed to analyze a set of historical documents, organized within a specific framework, to find patterns and trends in the academic narratives reproduced in a textual corpus, using a group of computational tools and techniques through text mining. The results obtained in this study, through systematized procedures in a Proof of Concept, synthesized quantitatively, showed that the use of digital technologies can significantly contribute to the analysis, interpretation, and understanding of historical documents and point to trends in academic narratives. Text mining confirmed the hypotheses raised about the semantic shift from "Sertão" to "Cerrado," with the purpose of discursively reinventing the territory as an economic category and commodity production, and the substantial distancing of narratives between developmental perspectives and the inclusion of Traditional Communities in this process.

Keywords: Digital Humanities. Distance Reading. Text Mining. Traditional Peoples.

RESUMEN
La búsqueda de patrones y tendencias en la producción historiográfica sobre la ocupación y colonización del Cerrado en las narrativas académicas, a través de la hermenéutica tradicional, constituye una tarea compleja y desafiante, dadas las limitaciones cognitivas humanas para analizar con precisión un vasto volumen de texto. Por otro lado, la técnica denominada Lectura Distante, creada por Franco Moretti, propone distanciarse del objeto de lectura, no como un obstáculo, sino como una forma de conocimiento, extrayendo menos detalles y más relaciones, patrones y formas dentro de una obra o entre obras. Esta técnica encuentra apoyo operativo en las Humanidades Digitales, un área interdisciplinaria que adopta métodos y técnicas computacionales para respaldar el desarrollo de estudios y análisis en las áreas de las ciencias sociales y las humanidades. Esta investigación se propuso analizar un conjunto de documentos históricos, organizados dentro de un marco específico, para encontrar patrones y tendencias en las narrativas académicas reproducidas en un corpus textual, utilizando un conjunto de herramientas y técnicas computacionales a través de la minería de texto. Los resultados obtenidos en este estudio, mediante procedimientos sistematizados en una prueba de concepto y sintetizados cuantitativamente, demostraron que el uso de tecnologías digitales puede contribuir significativamente al análisis, la interpretación y la comprensión de documentos históricos, además de señalar tendencias en las narrativas académicas. La minería de textos confirmó las hipótesis

ISSN: 2177-3246

planteadas sobre el cambio semántico de "Sertão" a "Cerrado", con el propósito de reinventar discursivamente el territorio como categoría económica y productora de mercancías, y el distanciamiento sustancial de las narrativas entre las perspectivas de desarrollo y la inclusión de las comunidades tradicionales en este proceso.

Palabras clave: Humanidades Digitales. Lectura a Distancia. Minería de Textos. Pueblos Tradicionales.

1 INTRODUÇÃO

A história do Cerrado não é apenas uma narrativa ecológica ou econômica; é, antes de tudo, uma profunda lição sobre o poder da linguagem (Lemes, 2026). A trajetória do termo "cerrado", de um simples adjetivo para um substantivo carregado de significados, espelha a própria domesticação de um território vasto e complexo. Nomear, longe de ser um ato neutro de descrição, é um gesto fundador de apropriação. É a arte de dominar, transformando a realidade material e caótica em um conceito imaterial, delimitado, sujeito a controles, estudos, explorações e reinvenções. A invenção do Cerrado foi, em sua essência, um processo de colonização semântica.

Nesse processo, a escolha lexical é reveladora. Tal como Guimarães Rosa, em Grande Sertão: Veredas (1956), optou pela potência mítica e universalizante do "sertão" em detrimento de denominações mais técnicas, a elevação do "Cerrado" de adjetivo ("cerrado" como algo "obstruído, fechado, vedado") a substantivo não foi casual. "Sertão" carrega uma carga existencial, literária e humana. Já "Cerrado", embora também evoque uma paisagem – aberto, complexo, áspero –, possui uma ressonância mais técnica, descritiva e, portanto, mais facilmente apropriável pelos discursos da ciência e do desenvolvimento econômico (Vicentini, 2016).

Portanto, a invenção recente do Cerrado – política, científica e cultural – demonstra que o bioma, tal como o conhecemos hoje, é muito mais do que um conjunto de fauna e flora. É uma construção, um conceito forjado. Sua própria nomeação foi a arte inicial de sua conquista. Entender essa trajetória é essencial para desnaturalizar a relação que temos com o território. O Cerrado não estava simplesmente lá, esperando ser descoberto; ele foi sendo inventado, palavra por palavra, nas práticas discursivas – na legislação, nas instituições, nas ações governamentais, nos relatórios e mapas oficiais, até alcançar as falas cotidianas do senso comum.

Embora a transformação do Cerrado em área agrícola tenha se iniciado no final da década de 1950 e início da década de 1960, dentre outros programas governamentais, com o denominado Programa de Metas (Meta 17 - Mecanização da Agricultura), instituído no governo Juscelino Kubitschek (1956-1961), a intervenção estatal no território ocorre antes, na década de 1940, durante o primeiro governo de Getúlio Vargas (1930-1945), com o projeto de Colonização nos Cerrados e a criação das colônias agrícolas em Dourados, no Mato Grosso do Sul; e Ceres, em Goiás (Shiki et al., 1997).

A compreensão das dinâmicas territoriais ocorridas e as intervenções estatais no bioma, sob a ótica das narrativas acadêmicas, requer a organização, em séries temporais e discursivas dos documentos produzidos, pois a historiografia sobre a sociedade e o ambiente (Cerrado) é constituída por um elevado volume de estudos que reportam mais do que registros e relatos isolados (ou mesmo correlacionados) em distintas linhas historiográficas e ideológicas sobre o passado. Essa historiografia postula um campo de disputas no qual se chocam narrações acerca do vazio demográfico e da

sociobiodiversidade, considerando as condições dos Povos Tradicionais, bem como da política desenvolvimentista adotada pelo Estado brasileiro

Conforme aponta Vicentini (2016), a própria transição da terminologia "Sertão" para "Cerrado" carrega uma carga semântica que oscila entre o determinismo biológico (características naturais, clima, solo, vegetação) e a identidade cultural (história, simbolismo, cultura). Essa perspectiva acadêmica, que apresentou crescimento exponencial nas últimas décadas, aborda desde os efeitos da colonização e atividades antrópicas até a complexa relação do agronegócio com o bioma-território e os impactos nas mudanças climáticas.

Considerando essa grande massa textual inserida em documentos não estruturados, torna-se um grande desafio a realização de análises pertinentes, especialmente críticas para compreensão de tendências, vieses e silenciamentos dentro deste grande volume documental. Para Costa et al. (2021), a estruturação de dados ligados (*linked data*) para análise ou mesmo para compartilhamento é um dos maiores desafios das ciências humanas e sociais.

Na perspectiva da pesquisa bibliográfica tradicional, torna-se uma tarefa complexa realizar análises quantitativas para detectar padrões, entre eles frequência de ocorrências e possíveis ausências de menções aos Povos e Comunidades Tradicionais (PCTs) em narrativas mais antigas ou a mudança de foco geográfico nas novas tendências historiográficas. Além disso, é complexo analisar manualmente o avanço das discussões sobre a conversão do Cerrado (em lavoura e pecuária) contraposto à sua transformação em fronteira agrícola e os impactos ocasionados pelas pressões antrópicas sobre o bioma-território.

Portanto, a realização de análises, balizadas pela interdisciplinaridade, torna-se necessária neste contexto, não como simples justaposição, mas como a articulação crítica definida por Japiassu (1976) e Frigotto (2008). É neste cenário que as Humanidades Digitais (HD), campo interdisciplinar que adota métodos e técnicas computacionais para apoiar o desenvolvimento de estudos e análises nas áreas de ciências sociais e humanas, definidas por Burdick et al. (2020, p. 71) como novos modos de produção acadêmica e de unidades institucionais para a pesquisa, ensino e publicações colaborativa, oferecem o instrumental necessário para o estudo proposto.

As Tecnologias da Informação e Comunicação (TICs), compreendidas no contexto das HDs, emergem como suporte operacional fundamental para auxiliar na busca, organização e identificação de padrões, ênfases e lacunas na literatura, além de sistematização e visualização de informações.

O presente estudo tem por objetivo realizar uma análise da produção historiográfica sobre o Cerrado, em um recorte específico da sua ocupação, utilizando a mineração de texto e ferramentas computacionais para mapear as tendências e progressos das narrativas acadêmicas sobre o uso e a ocupação deste território. Este recorte considera um volume de estudos predefinido, contudo, as etapas

de análises construídas metodologicamente podem ser replicadas em outras massas textuais com variadas dimensões.

2 REFERENCIAL TEÓRICO

2.1 A CONSTRUÇÃO SEMÂNTICA DO BIOMA-TERRITÓRIO

A definição de Cerrado sofre variações semânticas ao longo da história, sendo um conceito em constante formação e evolução para adequações à luz das análises de componentes físicos, bióticos e abióticos que o compõem, além dos fatores econômicos, sociais e políticos presentes nos estudos realizados. Entre as variações semânticas constituídas, o Cerrado é definido como sistema biogeográfico (Barbosa, 1996; Ribeiro; Walter, 2008), domínio morfoclimático (Ab'sáber, 2003)e bioma-território (Chaveiro; Barreira, 2010).

Historicamente, este recorte espacial concentrou disputas discursivas intensas, oscilando entre representações de negatividade, o lugar do vazio, do inóspito; e representações de utilitarismo econômico, que o enxergam apenas como recurso a ser explorado. Para superar essa análise fragmentada, esta pesquisa adota o conceito de Bioma-território, definido na obra de Eguimar Felício Chaveiro. Segundo o autor, não é possível dissociar a dimensão natural do território da dimensão política. Ao analisar a apropriação do Cerrado, Chaveiro e Barreira (2010, p. 30) argumentam que a própria definição do espaço carrega intencionalidades:

> Por conta disso, sem se desfazer da importância dos estudos do bioma e dos ecossistemas, bem como dos que fazem a abordagem por meio da categoria região, propõe-se pensar o Cerrado pelo prisma de um Bioma-território. Esse prisma intenta envolver numa única perspectiva as dimensões físico-territoriais, as socioeconômicas e as culturais e simbólicas. (Chaveiro; Barreira, 2010, p. 30).

Essa perspectiva incompatibiliza-se com a ideia de neutralidade científica. A transição terminológica de "Sertão" para "Cerrado" e, posteriormente, para "Bioma", reflete uma mudança, não apenas terminológica, mas também conceitual na forma como o poder hegemônico lida com a região. Vicentini (2016) destaca como essa mudança semântica, ocorrida ao longo da segunda metade do século XX, serviu para cientificizar o território, ao remover-lhe a carga cultural e humana associada ao termo Sertão, facilitando assim sua conversão em mera mercadoria ou recurso produtivo. A conversão do Cerrado para, sobretudo, produção agrícola, articulada e financiada pelo poder público estatal, e a modernização agrícola, não operaram apenas sobre o solo, mas sobre o imaginário. Steinberger (1997) aponta que a ocupação do Centro-Oeste foi guiada por uma geopolítica estatal que via a região como um espaço a ser integrado e preenchido, ignorando as territorialidades pré-existentes. Essa visão de vazio demográfico apresentado pela autora é uma construção política que é corroborada e ampliada por Chaveiro e Barreira ao afirmar que:

> Embora o processo (de conversão do Cerrado em fronteira agrícola) tenha transformado o Cerrado num cinturão produtivo importantíssimo, principalmente para a balança comercial do país, ao gerar bens de exportação, houve uma concentração de terras, um aumento da desigualdade social e uma concentração espacial, especialmente fundada na urbanização desigual que espelha um território urbanizado e cheio de problemas. Segue, junto, um campo produtivo, mas vazio de gente e da cultura do local (Chaveiro; Barreira, 2010, p. 27).

Além disso, a crítica de Vandana Shiva (2003), sobre o que a autora definiu como "monoculturas da mente", dialoga diretamente com esse processo de mudança terminológica e produtivista do/no Cerrado brasileiro. A simplificação da sociobiodiversidade para a produção de *commodities* (monocultura agrícola) é precedida por uma simplificação do pensamento que reduz a complexidade da vida (sobretudo da existência dos povos originários e tradicionais) nesta faixa territorial a parâmetros estritamente econômicos e políticos. Nas palavras da autora,

> Quando o saber local [tradicional] aparece de fato no campo da visão globalizadora, fazem com que desapareça negando-lhe o *status* de um saber sistemático e atribuindo-lhe os adjetivos de “primitivo” e “anticientífico”. Analogamente, o sistema ocidental é considerado o único “científico” e universal. Entretanto, os prefixos “científico” para os sistemas modernos e “anticientífico” para os sistemas tradicionais de saber têm pouca relação com o saber e muita com o poder (Shiva, 2003, p. 23).

Portanto, a mineração de textos proposta neste trabalho busca identificar a materialidade desses discursos. Ao mapear a frequência de termos como "recurso", "agrícola", "fronteira", "monocultura" e "vazio" em oposição a "vida", "preservação", "povo", "cultura" e "território", pretendemos evidenciar o que Diniz (2006) define como geopolítica do grande Cerrado, em que a disputa semântica é a primeira etapa da disputa territorial.

2.2 A ETNOCONSERVAÇÃO NA HISTORIOGRAFIA: SILENCIAMENTOS E INVISIBILIDADE

A integração entre conhecimento acadêmico e saberes tradicionais amplia as perspectivas sobre a preservação e conservação dos bioma-territórios, além de lançar luz sobre a coevolução entre povos e ecossistemas ao longo do tempo. Contudo, a pressão crescente do avanço sistêmico da economia produz um cenário em que as preocupações ambientais são secundarizadas em favor de uma lógica que reduz as complexidades socioambientais, comprometendo o pleno aproveitamento dos saberes locais e a incorporação de práticas tradicionais de conservação. As externalidades integradas ao avanço capitalista sobre os territórios ensejam um desenvolvimento que revele pouco ou nenhum planejamento de critérios sociais e sustentáveis de crescimento. (Garcia, 1995)

Sob as lógicas capitalistas de ocupação e produção, os biomas passam a ocupar um espaço reducionista, sendo concebidos exclusivamente como unidades funcionais de produção, circulação e valoração econômica. Este valor é mensurado estritamente por sua relevância na produção de *commodities* e geração de riquezas quantificáveis, estabelecendo um paradoxo do desenvolvimento marcado por uma contradição estrutural: os princípios de exploração territorial objetificam os biomas,

reduzindo suas potencialidades e especificidades a meros vetores de expansão de sistemas produtivos. Essa prática corrobora com o que diz Diniz (2006) quando defende que uma cultura voltada para a exportação implica na menor atenção aos círculos internos de economia, onde estão inseridos os povos tradicionais. Tal redução submerge a discussão sobre limites ecológicos, ciclos regenerativos e os equilíbrios inerentes à interação dinâmica entre os diferentes agentes que compõem os ecossistemas.

Assim, os bioma-territórios carregam elementos capazes de prover a subsistência de povos diversos, desde que estes atuem como agentes integrados à própria natureza, respeitando limites que não suportam nem acompanham o ritmo desordenado da pressão econômica. A produção de conhecimento científico também sofre pressões indiretas, uma vez que a racionalidade do crescimento tende a incorporar apenas conhecimentos que não questionem fundamentalmente o sistema, deslegitimando formas alternativas de saber. Estas acabam esquecidas pela expansão contínua e pela subordinação da natureza a critérios puramente economicistas, que desprezam a pluralidade epistemológica na construção do conhecimento sobre conservação.

Essas vicissitudes geram tensões previsíveis no quadro social em correlação com o ambiente, os agentes produtivos e os povos tradicionais locais (Chaveiro, 2019). Ignoram-se, assim, os saberes tradicionais e impede-se a plena aplicação de princípios de etnoconservação que, historicamente, compuseram a base de manejo territorial antes do mito do desenvolvimento capitalista hegemônico.

2.3 HUMANIDADES DIGITAIS E ESTUDOS SOCIOESPACIAIS: O PARADIGMA DA LEITURA DISTANTE

A Leitura Distante, proposta pelo teórico Franco Moretti (2000), surge como uma resposta às limitações existentes nos métodos tradicionais de estudos literários que predominaram nos estudos científicos, sobretudo, ao longo do século XX. Segundo o autor, a potencialidade do método está na superação das balizas que existem quando os estudos são dirigidos a partir de poucos textos canônicos, uma vez que isso implica, necessariamente, na incapacidade de abranger todos os aspectos históricos, sociais e culturais que podem ser observados quando apoiados em um sistema de estudo que se ancora em textos mais amplos.

A inovação metodológica consiste na pesquisa de padrões, recorrências, variações e ausências num volume consideravelmente maior de textos analisados. Esse afastamento parte do princípio de que o método de análise se desloca para um novo tipo de interpretação, que foca no apreço de dados quantitativos e de modelos analíticos pontualmente aplicados. Como afirma Moretti (2000), trata-se de uma leitura menos intensa, mas mais abrangente, capaz de revelar estruturas invisíveis à leitura tradicional. Tal como uma necessidade da nossa época, as abordagens interdisciplinares carecem de modelos de estudo que supram as necessidades de investigação com otimização de meios que

percorrem longos *corpus* de análise, e que seriam impossíveis de serem apreciados puramente pela capacidade humana de análise e interpretação.

O método não está livre de tensões e carrega na sua própria emergência carências que lhe são características, como uma interpretação reducionista de obras consagradas, bem como diminuição da complexidade estética e semântica de obras que, quando analisadas em contextos mais amplos, perdem relevância e escopo no desenvolvimento de seus objetivos originais. Cabe ressaltar aqui que o método de *Distant Reading*[1] pressupõe, necessariamente, a apreciação reducionista das obras, mas estabelece uma parte posterior ao *Close Reading*[2] no trabalho de desenvolvimento e construção científica (Moretti, 2000). Assim, precisamos aceitar como natural a essa aplicação metodológica a abstração científica em alguns momentos, já que esta acaba dando lugar a uma interpretação de determinados textos como unidade de dados, de um corpo maior, que revela padrões e não necessariamente conclusões específicas e isoladas de determinado objeto investigado.

Segundo Japiassu (1976), é a interdisciplinaridade uma exigência do nosso tempo, pois surge como resposta à crescente complexidade dos fenômenos contemporâneos, que não podem mais ser compreendidos adequadamente a partir de abordagens disciplinares isoladas. Ora, não estamos então falando apenas de um novo método, mas da emergência de um paradigma nos padrões investigativos acadêmicos que surge agora munido de uma ferramenta capaz de percorrer longos corpos de análise que, por sua vez, colocam o pesquisador como piloto de uma busca que visa perquirir diferentes textos e gêneros, além de distintas formas de acumulação do saber teórico. O objeto de análise deixa de ser a obra isolada e passa a ser o conjunto de relações estruturadas pelo pesquisador em larga escala.

2.4 MINERAÇÃO DE TEXTOS COMO MÉTODO HISTORIOGRÁFICO

Reis (2011, p. 6) relata que "[...] a organização cronológica, a sucessão rigorosa dos momentos que constituem um evento (histórico) e dos eventos entre si, deve ser visível em uma documentação objetiva". Neste sentido, as pesquisas em história necessitam de métodos capazes de garantir que os registros documentados sejam fiéis ao evento estudado. Para Rüsen (2015), o método historiográfico deve ser organizado em unidade composta das seguintes etapas: heurística, crítica e interpretação. Portanto, a crítica e interpretação do pesquisador também estão inseridas nesses documentos históricos.

Considerando a estrutura organizacional nos documentos, a extração de informação dessa massa textual, considerando as unidades definidas pelo autor, traz desafios significativos. Portanto, com o avanço das tecnologias digitais, sobretudo da inteligência artificial, torna-se importante o uso de recursos, entre eles a mineração de textos, parte integrante do método para análise e interpretação de textos históricos.

[1] Leitura distante – metodologia de análise literária criada por Franco Moretti no ano de 2000.
[2] Leitura atenta – terno surgido no século XX.

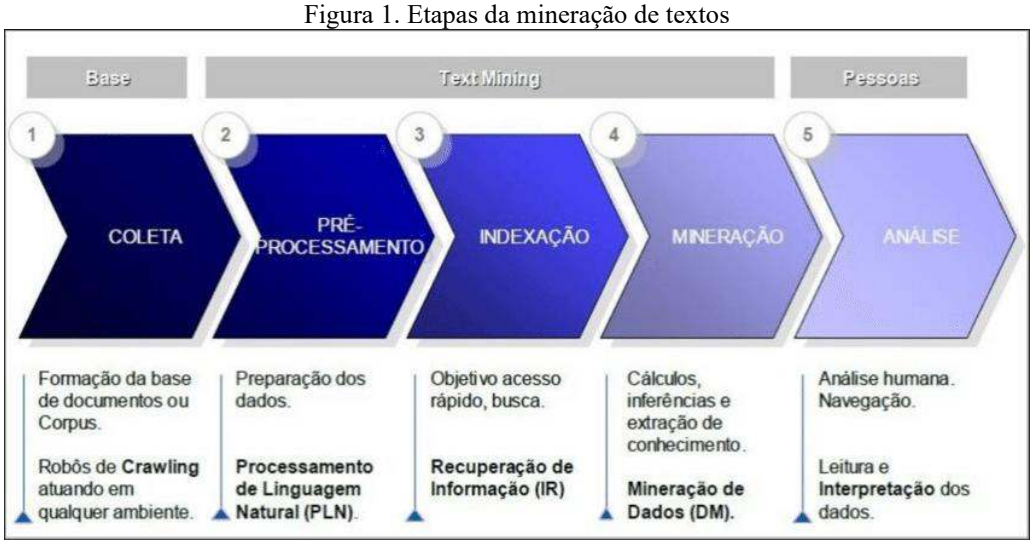
Mineração de textos (*Text mining*), conforme descreve Pezzini (2016, p. 58), consiste em "[...] uma extensão da mineração de dados, e pode ser definida como um processo de extração de informações desconhecidas e úteis de documentos textuais escritos em linguagem natural". A mineração de texto também é conhecida por outros nomes, como KDT (*Knowledge Discovery in Text* - Descoberta de Conhecimento em Textos) e TDM (*Text Data Mining* - Mineração de Dados em Textos), de acordo com Sakthi (2021). Sakthi (2021) corrobora com Pezzini (2016) e complementa:

> A mineração de textos pode ser compreendida como uma extensão da mineração de dados. Quando comparada à mineração de dados, a mineração de textos apresenta maior potencial comercial. As principais tarefas da mineração de textos incluem agrupamento de textos, produção textual, geração morfológica, categorização de textos, extração de conceitos e entidades, Reconhecimento de Entidades Nomeadas (NER), extração de termos, produção de taxonomias granulares, criação de dicionários, análise de sentimentos e sumarização de documentos, entre outras (Sakthi, 2021, p. 879).

Desta forma, a mineração de textos tem por objetivo transformar um grande volume textual disponível na forma de documentos não estruturados em unidades estruturadas, nas quais é possível aplicar análises quantitativas e extrair dados úteis mensuráveis e interpretáveis algoritmicamente. Esse processo se tornou fundamental devido ao aumento exponencial do volume de dados e informações produzidas em todas as áreas do conhecimento, incluindo a área acadêmica, na qual predominam documentos não estruturados, de tal modo que a utilização de técnicas apoiadas por tecnologias digitais torna-se fulcral.

De acordo com Aranha e Passos (2007), as etapas da mineração de textos são assimiladas entre *(i)* coleta; *(ii)* pré-processamento; *(iii)* indexação; *(iv)* mineração e *(v)* análise.

Figura 1. Etapas da mineração de textos



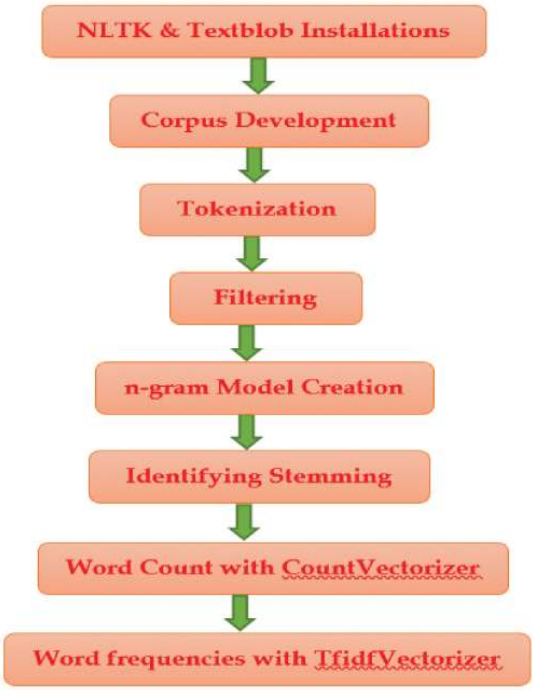
Fonte: Aranha e Passos (2007)

As etapas de mineração empregam técnicas desenvolvidas em diferentes áreas, como Extração de Informação (EI), Recuperação de Informação (RI), Processamento de Linguagem Natural (PLN),

Linguística Computacional, Estatística, Mineração de Dados (campo mais amplo da mineração que tem o objetivo de encontrar conhecimento em dados estruturados, semiestruturados e não estruturados) e Aprendizado de Máquina.

Uma das etapas mais importantes do processo é o pré-processamento, que se divide em diferentes subetapas e níveis, como instalações, desenvolvimento do *corpus*, tokenização, filtragem, criação de modelos de n-gramas, identificação de *stemming*, criação de vetorizador de contagem de palavras e vetorizador TF-IDF Sakthi (2021).

Figura 2. Etapas do pré-processamento em mineração de textos



Fonte: Sakthi (2021)

Neste processo, o autor destaca uma sequência de ações, iniciando pela organização de bibliotecas da linguagem de programação Python, comumente usada nas ferramentas computacionais para mineração de textos e de Inteligência Artificial, passando por tokenização e filtragens, até aplicação de métodos estatísticos como contagem de frequência de palavras no *corpus*.

O referencial teórico em um estudo compreende uma análise crítica e organizada da literatura pertinente ao tema, fornecendo uma contextualização teórica e definindo os conceitos-chave. Deve conter de maneira abrangente as teorias, modelos e pesquisas anteriores, identificando lacunas, contradições e consensos na literatura que são importantes para o foco do trabalho que está sendo desenvolvido.

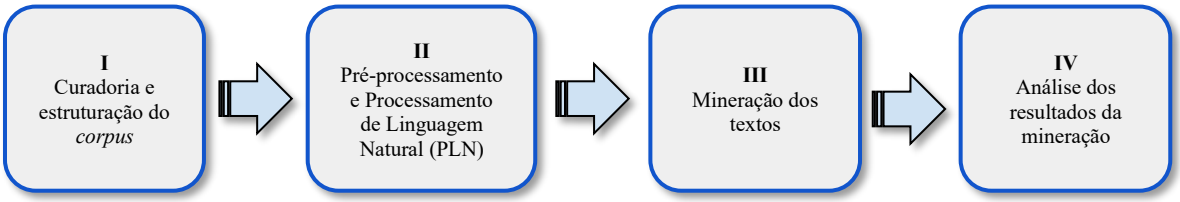
3 PROCEDIMENTOS METODOLÓGICOS

A pesquisa foi desenvolvida utilizando abordagem mista: qualitativa para análise documental e constituição do arcabouço teórico; e quantitativa, por meio de ferramentas e métodos computacionais, para análises estatísticas aplicadas ao *corpus*. O estudo fundamenta-se em conceitos embasados nas Humanidades Digitais, para instrumentalizar o conceito definido por Moretti (2013) como *Distant Reading* ou Leitura Distante. Diferente das interpretações tradicionais, que buscam a compreensão original de textos, a Leitura Distante utiliza algoritmos implementados em ferramentas computacionais para detectar padrões estruturais, tendências temporais e lacunas temáticas em conjuntos documentais extensos, impossíveis de serem alcançados pela leitura humana linear (Jockers, 2013).

A metodologia foi organizada para superar a capacidade cognitiva humana limitada de ler e processar a grande quantidade de documentos relativos à historiografia do Cerrado, permitindo assim, uma análise interpretativa realizada com tecnologias digitais, que cruza a evolução semântica dos conceitos (Vicentini, 2016) com a geopolítica da ocupação (Steinberger, 1997).

Os procedimentos metodológicos estão sintetizados em quatro etapas sequencialmente correlacionadas, envolvendo as técnicas da mineração de textos, sendo: (i) Curadoria e estruturação do *corpus*; (ii) Pré-processamento e Processamento de Linguagem Natural (PLN); e (iii) Mineração de textos e (iv) Análise dos resultados.

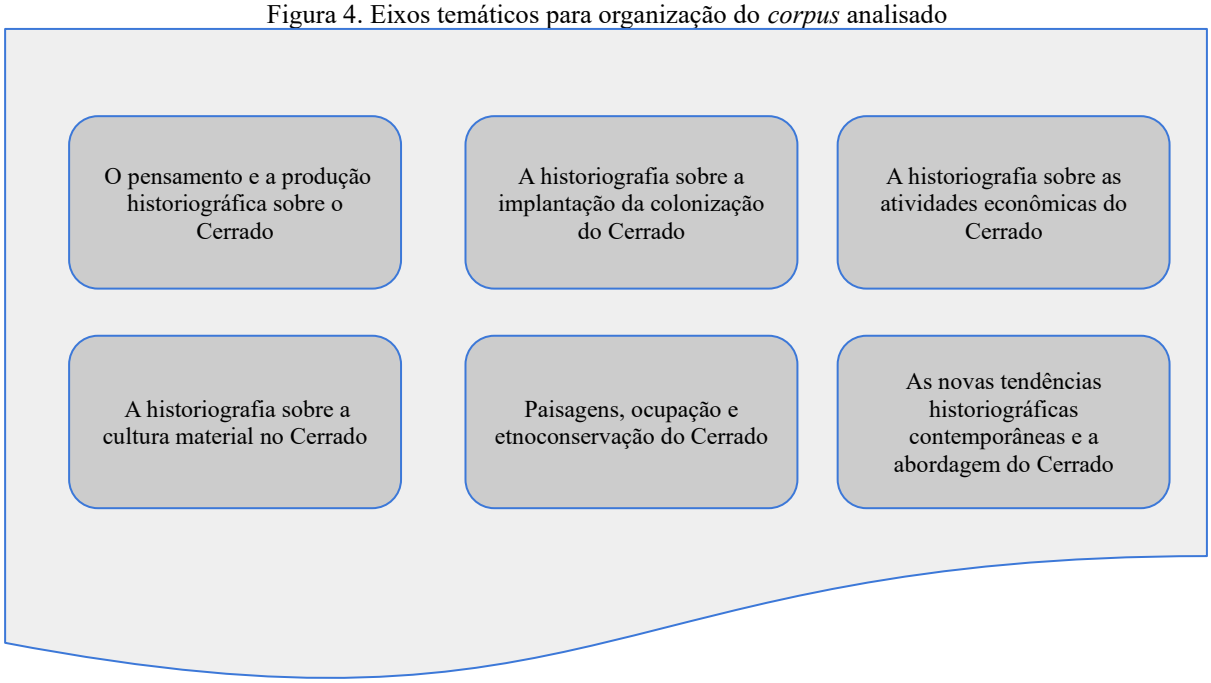
Figura 3. Etapas da análise documental



Fonte: Elaborada pelos próprios autores (2026)

A primeira etapa, que consiste na curadoria e estruturação, tem por finalidade preparar o corpo documental para análise, consistindo na seleção da produção bibliográfica acadêmica focada na historiografia, ocupação e aspectos socioambientais do Cerrado. Para o recorte desta pesquisa, a composição do *corpus* considerou primariamente as referências bibliográficas e artigos utilizados na disciplina Múltiplos Olhares sobre o Cerrado: pesquisas, estudos e abordagens, do Programa de Pós-graduação em Territórios e Expressões Culturais do Cerrado (PPGTECCER) da Universidade Estadual de Goiás (UEG), ofertada no semestre 2025/2, organizados em seis eixos temáticos, conforme ilustrado na Figura 4.

Figura 4. Eixos temáticos para organização do *corpus* analisado



Fonte: Elaborada pelos próprios autores (2026)

A seleção dos materiais complementares seguiu os critérios de relevância semântica e densidade teórica, incorporando obras, tais como Tal sertão, qual cerrado? (Vicentini, 2016) e Região Centro-Oeste: uma visão geopolítica (Steinberger, 1997), além de artigos sobre etnoconservação (Diegues, 2019) e modernização conservadora da agricultura (Shiki *et at*., 1997). Os documentos analisados não refletem toda a produção historiográfica acadêmica sobre o tema, porém, constitui-se uma Prova de Conceito (PoC - *Proof of Concept*) (Cruz et al., 2022), capaz de validar a metodologia adotada e os métodos e ferramentas computacionais utilizados no estudo.

Nas etapas I e II ocorre a mineração dos textos e foram executadas integralmente por meio de ferramentas e métodos computacionais (Tabela 1). Além das bibliotecas da linguagem de programação Python, implementadas utilizando o *Google Colab*, foram utilizadas ferramentas complementares como o software OCRmyPDF, que adiciona uma camada de texto pesquisável em arquivos PDFs digitalizados ou baseados em imagem, usando o motor *tesseract* (biblioteca OCR - *Optical Character Recognition*, ou Reconhecimento Óptico de Caracteres), necessário em alguns documentos disponibilizados em formato de imagens.

Nestas etapas ocorre o Pré-processamento e Processamento de Linguagem Natural (PLN), transformando os textos em dados computáveis e mensuráveis, reduzindo a complexidade, a dimensionalidade e o ruído linguístico. As ações foram realizadas seguindo a sequência de tratamento: *(i)* limpeza de ruído; *(ii)* tokenização e remoção de stopwords; e *(iii)* lematização, com emprego de três métodos computacionais analíticos distintos sobre o *corpus* para responder às questões historiográficas levantadas na fundamentação teórica.

A última etapa, a interpretação dos dados, parte dos resultados obtidos e sistematizados na mineração para explicar as evidências descobertas e conclusões, sendo realizada pelos pesquisadores e descritas na seção resultados deste trabalho.

Tabela 1. Descrição dos procedimentos técnicos realizados nas etapas metodológicas

Etapa	Procedimento técnico	Finalidade	Ferramenta(s) utilizada(s)
I	a) Ingestão de dados	Conversão de arquivos para texto plano (*text plain*, codificação UTF).	Software OCRmyPDF; Biblioteca Tesseract-ocr-por; Bibliotecas Python PyPDF2 / pdfminer.six.
	b) Segmentação	Estruturação dos textos em uma estrutura tabular denominada DataFrame.	Biblioteca Python Pandas.
II	a) Limpeza de ruído	Remoção automatizada, via Expressões Regulares (RegEx), de elementos não semânticos.	Biblioteca Python re.
	b) Tokenização e remoção de *Stopwords*	Segmentação do texto em unidades mínimas (*tokens*) e exclusão de palavras funcionais (artigos, preposições, conectivos).	Biblioteca Python NLTK (Natural Language Toolkit) / spaCy (modelo pt_core_news_sm).
	c) Lematização	Redução algorítmica das palavras à sua forma canônica ou raiz (ex: "colonização", "colonizar", "colonizada" → coloniz).	Biblioteca Python spaCy (modelo pt_core_news_sm).
III	a) Análise lexical e de relevância	Identificação da assinatura discursiva de cada período ou autor, aplicação técnica estatística TF-IDF (*Term Frequency-Inverse Document Frequency*).	Biblioteca Python Scikit-learn com o módulo TfidfVectorizer.
	b) Modelagem de Tópicos (*Topic Modeling* – LDA)	Descoberta de temas latentes e de silenciamentos (o que não foi descrito nas narrativas).	Algoritmo não supervisionado LDA (Latent Dirichlet Allocation) via biblioteca Gensim e pyLDAvis (para visualização).
	c) Cartografia digital e NER	Espacialização das narrativas com a identificação das localizações, permitindo visualizar quais áreas do Cerrado foram super-representadas e quais permanecem como "zonas de silêncio geográfico" na literatura acadêmica.	Biblioteca Python spaCy.
IV	a) Interpretação dos dados	Análise dos dados produzidos nas etapas anteriores, tópicos, agrupamentos sugeridos pelos algoritmos, frequência de palavras, localização das ocorrências das narrativas.	Interpretação realizada pelos pesquisadores.

Fonte: Elaborada pelos próprios autores (2026)

O *corpus* analisado foi construído com 23 arquivos historiográficos incluindo documentos estatais, livros, teses e artigos publicados entre os anos de 1958 a 2021, no idioma português, totalizando aproximadamente 2.400 páginas de dados não estruturados, conforme os critérios utilizados para o recorte conceitual da pesquisa. A massa textual foi organizada em dez tópicos e agrupados conforme a similaridade, cada documento, de acordo com o seu conteúdo, recebeu um peso variando entre 0 e 1, sendo o valor 1 o tópico onde o conteúdo do documento obteve maior similaridade, denominado "tópico predominante".

4 RESULTADOS E DISCUSSÕES

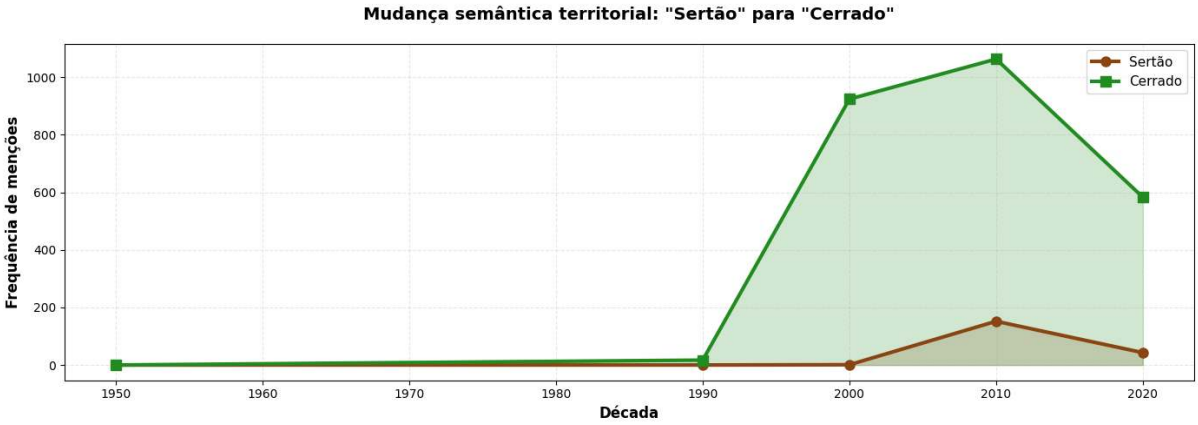
Assim como a mineração, genericamente, busca encontrar elementos de valor, a mineração de textos desenvolvida nesta pesquisa serviu para identificar padrões que se repetiam, além de ausências significativas nos trabalhos historiográficos examinados. Esse método permitiu observar aspectos do discurso que não aparecem de forma óbvia em um primeiro momento, mas que fundamentam como se relata a história do território e de sua ocupação. As análises foram permeadas pela correlação entre mudanças terminológicas "Sertão" e "Cerrado", estudo de tópicos, menções aos Povos Tradicionais e análise das narrativas dominantes sobre agricultura e suas correlações com os demais termos, conforme mostra a Figura 5.

Quanto aos contextos e ocorrências dos termos "Sertão", "Cerrado" e "Povos Tradicionais", centrais na mineração, foram encontradas, respectivamente, 196, 2636 e 1560 menções no *corpus*. A análise mostrou que 20 dos 23 documentos analisados mencionam Povos Tradicionais, correspondendo a 87% do *corpus*, e 13% omitem, totalizando 3 documentos. Os documentos que não mencionam Povos Tradicionais incluem o "território" em suas narrativas, mas ignoram quem vive nele. Este dado refuta, documentalmente, a hipótese dos silenciamentos (ou falta de menções) sobre os "Povos Tradicionais" nas narrativas acadêmicas dentro dos limites deste estudo.

Contudo, a análise de correlação mostrou que não existe relação dialógica entre "visão desenvolvimentista" e "preservação socioambiental", as menções ocorrem em documentos e épocas distintas. Assim, pode-se afirmar que a visibilidade nas narrativas acadêmicas não implica necessariamente em reconhecimento da autonomia dos Povos Tradicionais, em proteção territorial ou mesmo na produção de conhecimento comprometido com sustentabilidade social e cultural desses povos, mas, ao contrário, podem tomar direção oposta ao gerar uma apropriação simbólica que esvazia a dimensão sociológica das lutas e resistências em suas diferentes frentes de embate, reduzindo-os a elementos quase folclóricos ou a meros recursos de referência informacional para agendas externas ao Cerrado e seus habitantes originários.

A análise temporal confirmou a mudança semântica, o termo Cerrado passou a predominar nas narrativas ao mesmo tempo que "agricultura" também cresceu (Figuras 5 e 6).

Figura 5. Mudança semântica territorial

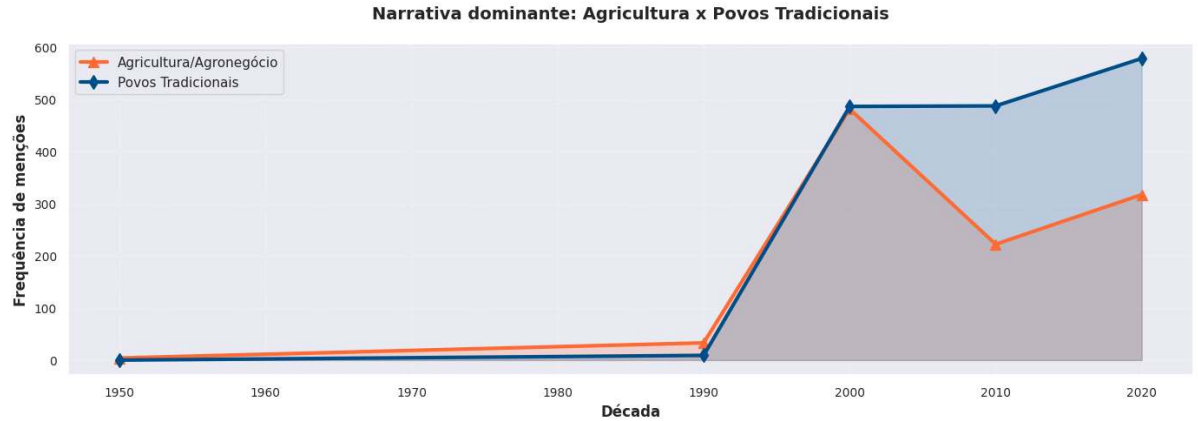


Fonte: Elaborado pelos próprios autores (2026)

A ausência de dados no período de 1960 a 1990 que ocorreu devido à falta de documentos inseridos no *corpus*, publicados neste espaço temporal, não significa que não houve discussões sobre os temas analisados, apenas não foram inseridos no recorte desta pesquisa.

Observa-se o aumento de densidade do termo "agrícola" na década de 1990, que, embora com queda na década seguinte, mantém-se em alta. Isso coincide com a consolidação do Cerrado como fronteira agrícola e expansão das monoculturas no bioma, impulsionada pela produção de *commodities*, iniciada na década de 1970 (Santos *et al.*, 2010), com vultuosos estímulos estatais (Steinberger, 1997). Essa expansão é marcada pela alta tecnologia, incentivos ao crédito e, posteriormente, a chamada Revolução Verde.

Figura 6. Narrativa dominante: Agricultura x Povos Tradicionais

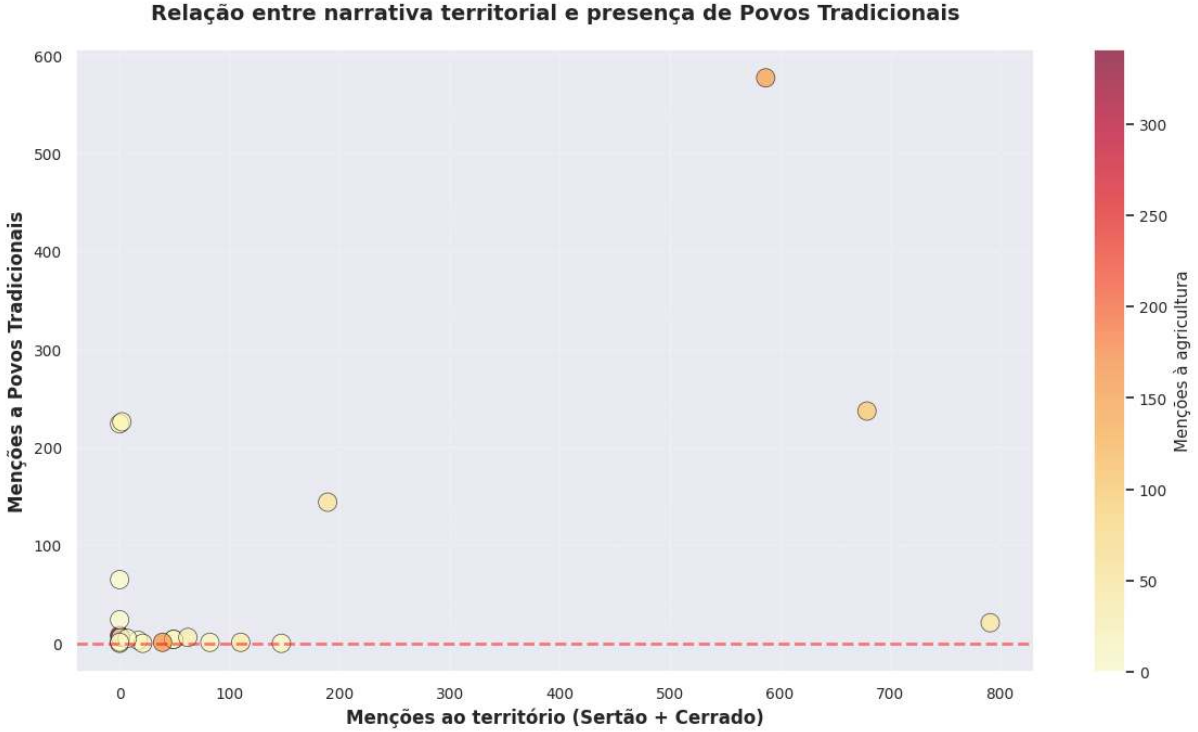


Fonte: Elaborado pelos próprios autores (2026)

O gráfico ilustrado na Figura 7 constata uma contradição na forma como o território é discutido em documentos oficiais e acadêmicos: existe uma invisibilização dos Povos Tradicionais que, contudo, ocorre nas discussões sobre o desenvolvimento do território. Embora o eixo horizontal mostre que existe um volume considerável de menções ao território (Sertão e Cerrado), e o eixo vertical registre a presença desses povos, a maioria dos dados está condensada na base do gráfico. Isso indica que, mesmo

quando se fala muito da terra e da agricultura (representada pelas cores mais escuras), a menção às comunidades que nela vivem, como indígenas e quilombolas, é frequentemente mínima ou inexistente. Isso evidencia uma narrativa que prioriza o espaço geográfico e a exploração econômica em detrimento da existência de suas populações originárias.

Figura 7. Relação entre narrativa territorial e presença de Povos Tradicionais



Fonte: Elaborado pelos próprios autores (2026)

A distribuição espacial dos pontos reforça também a hipótese de ausência de relação dialógica entre agricultura/território e Povos Tradicionais nos documentos, acentuando uma polarização das discussões em dois mundos distintos e acentuando as disputas territoriais: ou *(i)* o território é tratado como um vazio demográfico esperando o desenvolvimento, fala-se da geografia, do solo, da agricultura, mas as pessoas que sempre viveram ali não surgem nas narrativas, são aproximadamente um terço dos documentos analisados; ou *(ii)* os Povos Tradicionais são reconhecidos, seguindo um equilíbrio relativo ao falar do Cerrado sem apagar quem vive nele, mas sem menções à agricultura.

Os dados confirmam a hipótese de que o Cerrado foi "inventado" discursivamente para viabilizar a modernização. Em 1958, no documento relativo ao Programa de Metas de JK, o termo "Cerrado" não é citado nenhuma vez. O foco está no desenvolvimento nacional, mas a região ainda não é tratada como uma unidade biogeográfica específica de interesse econômico direto na linguagem moderna. Já em 2010, o termo "Cerrado" tem um salto para 661 menções apenas no livro Cerrados: perspectivas e olhares, totalizando 2636 em todo *corpus*. Isso valida a tese de Vicentini (2016) e Chaveiro (2010): o território deixou de ser o "Sertão" (cultural, misterioso) para ser cientificamente catalogado como "Cerrado" (bioma, recurso), facilitando sua inserção no mercado global.

A análise materializa também a invisibilidade do "vazio demográfico". A correlação entre os termos "agricultura" e "povos tradicionais" (Figura 7) expõe a exclusão discursiva denunciada por Diegues (2019) e Santos (2016). Dois documentos destacam essa materialidade: *(i)* documento técnico/estatal do ano 2000, intitulado "O papel do Estado no processo de ocupação das áreas de Cerrado entre as décadas de 60 e 80", menciona "Cerrado" 110 vezes, "agricultura" 39 e "povos tradicionais" apenas uma vez. A ocupação e agricultura são intensamente discutidas, mas a presença humana tradicional é estatisticamente apagada; *(ii)* o livro "Cerrados: perspectivas e olhares", publicado em 2010, sendo um documento crítico/sociológico, traz 661 menções para "Cerrado", 102 para "agricultura" e 237 para "povos tradicionais", isso evidencia a reação acadêmica e social às narrativas economicistas, trazendo os povos para o centro dos debates.

Podemos agrupar os dez tópicos criados na mineração dos textos em dois eixos: *(i)* visão desenvolvimentista e *(ii)* visão socioambiental. A partir disso, conclui-se, por meio da análise quantitativa, que, enquanto a discussão sobre "Agricultura" e "Cerrado" cresce exponencialmente a partir dos processos de modernização, a presença dos "Povos Tradicionais" permanece estatisticamente insignificante nos documentos técnicos de planejamento, aparecendo apenas nas obras de crítica sociológica. A Tabela 2 representa a análise de correlação agrupada por década. A coluna "Proporção Cerrado" mostra a correlação média entre "Cerrado" e "Agricultura" em cada década.

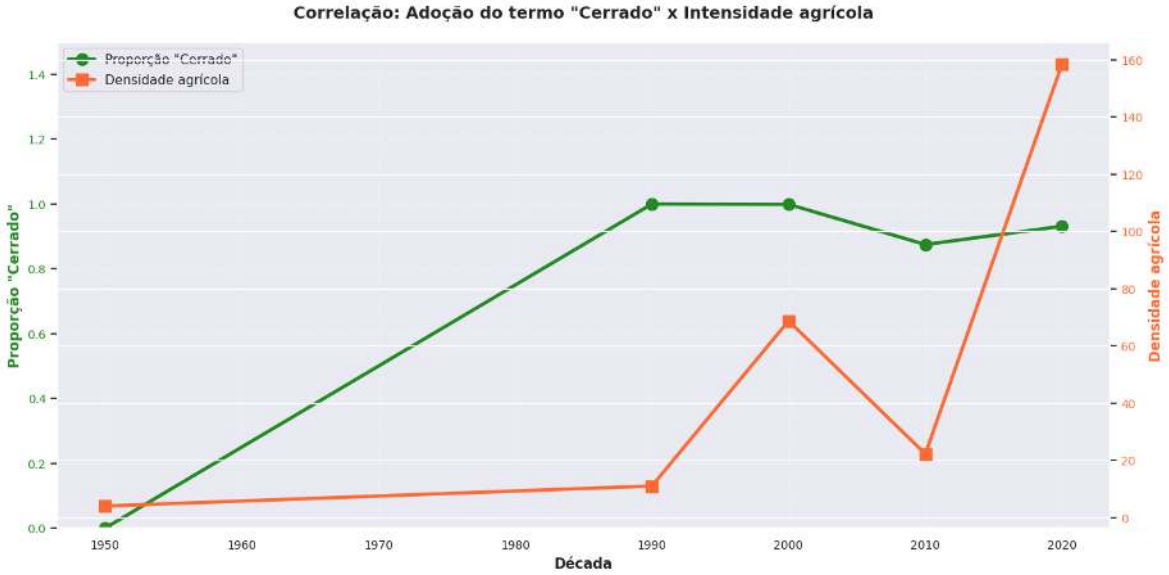
Tabela 2. Termo "Cerrado" correlacionado com densidade agrícola

Década	Sertão	Cerrado	Agricultura	Proporção do Cerrado	Densidade Agro
1950	0	0	4	0,000	4,00
1990	0	17	2	0,999	1,00
2000	1	924	481	0,999	68,71
2010	152	1.063	222	0,875	22,20
2020	43	583	317	0,931	158,50

Fonte: Elaborada pelos próprios autores (2026)

Os dados sintetizados mostram que a correlação entre a adoção de "Cerrado" e "agricultura" é de 0.413 (em uma escala que varia entre +1 a -1). Isso evidencia uma correção moderada, ou seja, existe uma tendência de aumento conjunto entre os dois termos nas narrativas. Os dados estatísticos pormenorizam um ponto fulcral: a década de 2010 indica ter sido o ponto de virada nas narrativas, o momento em que essas duas visões (Agricultura/Agronegócio e Povos Tradicionais) começaram a se cristalizar e se opor frontalmente.

Figura 8. Análise de correlação "Cerrado" x "Agro"



Fonte: Elaborado pelos próprios autores (2026)

Antes disso, nos anos 2000, a invisibilização era quase consensual, vários documentos falavam do território, mas simplesmente ignoravam sua população, perpetuando aquela ideia colonial de "vazio demográfico". Contudo, a partir de 2010, surgiu uma tensão visível no *corpus*: alguns textos intensificaram a narrativa desenvolvimentista enquanto outros começaram a questioná-la explicitamente. A organização dos tópicos permitiu visualizar essa corrente crítica, os documentos agrupados no eixo 2 (visão socioambiental) são os que mais mencionam Povos Tradicionais, sugerindo que existe um campo de pesquisa consolidado trabalhando com perspectivas de etnoconservação e reconhecimento territorial.

5 CONSIDERAÇÕES FINAIS

A análise apresentada confirma que o Cerrado foi discursivamente reinventado como categoria científica e econômica, suplantando a noção cultural de "Sertão" para viabilizar sua inserção na lógica capitalista global. A transição do termo não é apenas uma mudança semântica, mas a materialização de um projeto de poder que reconfigurou o território em unidade funcional, com vistas à produção de *commodities* agrícolas, fustigando deliberadamente a dimensão humana e cultural do bioma-território.

A menção dos Povos Tradicionais em 87% dos documentos analisados refuta a hipótese de silenciamento absoluto, porém, desvela uma forma mais sofisticada de invisibilização: a menção passiva. Os povos são citados, mas permanecem como objetos passivos das narrativas desenvolvimentistas ou como vítimas nas narrativas críticas, raramente como sujeitos portadores de conhecimentos válidos para pensar alternativas de resistência e conservação ambiental frente ao modelo econômico hegemônico.

Mais preocupante ainda é a constatação de que, mesmo nos documentos críticos, predomina uma lógica reativa, focada no diagnóstico dos danos, sem avançar na sistematização dos saberes

tradicionais como alternativa epistemológica legítima. É dado um lugar ao conhecimento dos povos originários, muitas vezes folclorizado, tratado como patrimônio cultural a ser "preservado", mas não como racionalidades válidas para orientar políticas públicas de uso territorial.

Fica sugerido que a modernização agrícola do Cerrado operou não apenas por meio da transformação material do território, mas fundamentalmente mediante uma pressão exercida sobre os significados semânticos que capturam sentidos e dão polidez a um território descaracterizado do que lhe é singular, integrado por valores economicistas que reduzem as perspectivas sociais complexas do Cerrado enquanto bioma-território.

REFERÊNCIAS

ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. RESI – Revista Eletrônica de Sistemas de Informação, n. 2, 2006.

AB'SÁBER, A. N. Os domínios de natureza no Brasil: potencialidades paisagísticas. São Paulo: Ateliê Editorial, 2003.

BARBOSA, A. S. Sistema biogeográfico do cerrado: alguns elementos para sua caracterização. Goiânia: Editora UCG, 1996.

BURDICK, A.; DRUCKER, J.; LUNENFELD, P.; PRESNER, T.; SCHNAPP, J. Um breve guia para as Humanidades Digitais. Tradução de Isabel Jungk. TECCOGS – Revista Digital de Tecnologias Cognitivas, São Paulo, n. 21, p. 69-98, jan./jun. 2020.

CADAVID GARCIA, E. A. Desenvolvimento econômico sustentável do cerrado. Pesquisa Agropecuária Brasileira, Brasília, DF, v. 30, n. 6, p. 759-774, jun. 1995.

CHAVEIRO, E. F.; BARREIRA, C. C. M. A. Cartografia de um pensamento de cerrado. In: PELÁ, Márcia; CASTILHO, Denis (Org.). Cerrados: perspectivas e olhares. Goiânia: Editora Vieira, 2010. p. 15-34.

COSTA, R.; ALMEIDA, B.; RAMOS, M.; CAMPOS, M. I. B. O papel da linguística na era das humanidades digitais. Linha D'Água, São Paulo, v. 34, n. 2, p. 1–8, maio/ago. 2021. DOI: 10.11606/issn.2236-4242.v34i2p1-8.

CRUZ, R. F.; COLAÇO JÚNIOR, M.; GOIS, V. M. Quão experimentais e estratégicas são as aplicações de *Business Intelligence* (BI) e data mining? *Iberoamerican Journal of Strategic Management (IJSM)*, São Paulo, v. 21, p. 1–36, 2022. DOI: https://doi.org/10.5585/riae.v21i1.17689.

DINIZ, B. P. C. O Grande Cerrado do Brasil Central: geopolítica e economia. 2006. Tese (Doutorado em Geografia Humana) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2006.

FRIGOTTO, G. A interdisciplinaridade como necessidade e como problema nas ciências sociais. Ideação, v. 10, n. 1, p. 41-62, 2008.

JAPIASSU, H. Interdisciplinaridade e patologia do saber. Rio de Janeiro: Imago, 1976.

JOCKERS, M. L. Macroanalysis: Digital Methods and Literary History. University of Illinois Press, 2013.

LEMES, Fernando Lobo. A invenção do cerrado: esquadrinhando discursos, reconstruindo histórias – Goiás (1890-1960). Relatório de Pesquisa. Programa de Pós-graduação Territórios e Expressões Culturais no Cerrado (PPGTECCER). 20p. Anápolis, Universidade Estadual de Goiás (UEG), 2026.

MORETTI, F. Conjectures on World Literature. New Left Review, nº 1, p. 54–68, Jan./Feb. 2000.

MORETTI, F. Distant Reading. London: Verso, 2013.

PEZZINI, A. Mineração de textos: conceito, processo e aplicações. Revista Eletrônica do Alto Vale do Itajaí – REAVI, v. 5, n. 8, p. 1–13, dez. 2016. DOI: 10.5965/2316419005082016058.

REIS, J. C. O lugar da teoria-metodologia na cultura histórica. Revista de Teoria da História, Goiânia, v. 6, n. 2, p. 4–26, 2011.

RIBEIRO, J. F.; WALTER, B. M. T. O conceito de savana e de seu componente Cerrado. In: SANO, S. M.; ALMEIDA, S. P. de; RIBEIRO, J. F. (Org.). Cerrado: ecologia e flora. Brasília, DF: Embrapa Informação Tecnológica; Planaltina, DF: Embrapa Cerrados, 2008. p. 21-45.

RIBEIRO, R. F. O Eldorado do Brasil central: história ambiental e convivência sustentável com o Cerrado. In: ALIMONDA, Héctor (Comp.). *Ecología política: naturaleza, sociedad y utopía*. Buenos Aires: CLACSO, 2002. p. 249-274.

RÜSEN, J. Teoria da história: uma teoria da história como ciência. Tradução de Estevão C. de Rezende Martins. Curitiba: Editora UFPR, 2015.

SHIVA, V. Monoculturas da mente: perspectivas da biodiversidade e da biotecnologia. São Paulo: Gaia, 2003.

STEINBERGER, M. Região Centro-Oeste: uma visão geopolítica. In: ENCONTRO NACIONAL DA ANPUR, 7., 1997, Recife. Anais... Recife: MDU/UFPE, 1997. p. 1902-1910.

SAKTHI VEL, S. Pre-processing techniques of text mining using computational linguistics and Python libraries. In: *INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND SMART SYSTEMS (ICAIS)*, 2021. Proceedings. [S. l.]: IEEE, 2021. ISBN 978-1-7281-9537-7.

SANTOS, M. A. dos; BARBIERI, A. F.; CARVALHO, J. A. M. de; MACHADO, C. J. O Cerrado brasileiro: notas para estudo. Texto para Discussão, n. 387, jun. 2010.

SHIKI, S; SILVA, J. G. da; ORTEGA, A. C. (Org.). Agricultura, meio ambiente e sustentabilidade do cerrado brasileiro. Uberlândia: EDUFU, 1997.

VICENTINI, A. Tal sertão, qual cerrado? Goiânia: Editora UFG, 2016.