



**XXVIII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA DA UEFS  
SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA - 2024**

**O USO DE CORPORA ELETRÔNICOS ANOTADOS PARA MAPEAR A  
SINTAXE DOS PRONOMES PESSOAIS DO PORTUGUÊS BRASILEIRO**

**Gilmara Oliveira dos Santos<sup>1</sup>; Antonio José Maria Codina Bobia<sup>2</sup>**

1. Voluntária PEVIC, Graduanda em Letras-Língua Portuguesa, Universidade Estadual de Feira de Santana, e-mail: [gilmaraoliveirasantos@outlook.com.br](mailto:gilmaraoliveirasantos@outlook.com.br)
2. Orientador, Departamento de Letras Artes, Universidade Estadual de Feira de Santana, e-mail: [ajmcbobia@uefs.br](mailto:ajmcbobia@uefs.br)

**PALAVRAS-CHAVE:** Pronomes pessoais; Português brasileiro; Edição eletrônica de textos.

## **INTRODUÇÃO**

No Português Brasileiro (PB), o uso do pronome tônico é mais prevalente do que o uso do clítico acusativo, como exemplificado em "João beijou ela" em vez de "João a beijou". Além disso, observa-se a coexistência dos pronomes "tu" e "você" em posição pré-verbal, assim como a coexistência de "te" e "você" com caso acusativo, exemplificados em frases como "Tu vai" e "Eu te vi". No entanto, em alguns dialetos do PB, há o licenciamento do pronome "tu" em posição pós-verbal, em casos acusativos ou dativos, substituindo as formas mais comuns "você" e "te", como em "Eu vi tu no cinema ontem". Este fenômeno do "tu" pós-verbal ainda é pouco estudado. Obras de referência como Castilho (2010) sequer mencionam o fenômeno, destacando a necessidade de pesquisas mais profundadas.

Dessa necessidade, surge o presente projeto, que visa investigar a ocorrência do "tu" pós-verbal e a sintaxe dos pronomes em geral no Corpus de cartas do Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS), sediado na UEFS e coordenado pelas professoras Mariana Fagundes e Zenaide Carneiro. Um corpus eletrônico é uma coleção de textos digitais utilizada para extrair informações sobre o conteúdo e as estruturas linguísticas presentes. Ferramentas computacionais como listadores de palavras, etiquetadores e analisadores sintáticos têm revolucionado a pesquisa em sintaxe diacrônica, proporcionando precisão e rapidez na análise de grandes quantidades de dados. Um exemplo proeminente é o Corpus Tycho Brahe, que contém textos em português de autores nascidos entre 1380 e o fim do século XIX, possibilitando buscas sintáticas automáticas em dados extensivos.

Nos últimos sete anos, a parceria entre o Corpus Tycho Brahe e o CE-DOHS resultou no Projeto CorPorA II, visando expandir a parte brasileira do Corpus Anotado do Português Histórico Tycho Brahe e criar a Plataforma Tycho Brahe para integrar diferentes corpora

online, incluindo o CE-DOHS. Neste contexto, propomos a edição eletrônica das 131 cartas do conjunto "Cartas de mãos inábeis" do CE-DOHS, que consiste em edições fac-similares e semidiplomáticas de cartas escritas por sertanejos das regiões rurais da Bahia. As cartas, escritas por indivíduos com pouca familiaridade com a língua escrita, oferecem uma amostra representativa da variedade popular do português brasileiro. O objetivo geral do projeto é realizar a edição eletrônica e a anotação morfossintática das cartas para estudar os aspectos sintáticos idiosincráticos do uso dos pronomes pessoais e verificar a prevalência dos pronomes de caso reto em posição pós-verbal. Especificamente, pretende-se anotar morfossintaticamente os textos, analisar morfossintaticamente o fenômeno dos pronomes de caso reto em posição pós-verbal e apresentar os resultados da pesquisa por meio de trabalhos científicos e relatórios detalhados das atividades desenvolvidas. Este trabalho contribuirá para a compreensão da variação linguística e das mudanças sintáticas no PB, além de enriquecer a base de dados disponível para futuras pesquisas.

### **MATERIAL E MÉTODOS OU METODOLOGIA (ou equivalente)**

A metodologia de preparação dos documentos seguiu os seguintes passos: Edição eletrônica dos textos com a ferramenta e-Dictor (cf. <https://edictor.net/>); 3. Etiquetagem de palavras (morfológica) no e-Dictor; 4. Anotação sintática automática e correção dos textos anotados; 5. Análise do fenômeno sob estudo

### **RESULTADOS E/OU DISCUSSÃO (ou Análise e discussão dos resultados)**

Devido à interrupção do trabalho, não foi possível chegar a resultados.