

Edição Modernizada dos *Dois Livros Manuscritos do Brejo do Campo Seco, Ba (XVIII-XIX): Desafios da Nova Filologia*

Modernized Edition of the Two Manuscript Books of *Brejo do Campo Seco, Ba (XVIII-XIX): Challenges of the New Philology*

Mariana Fagundes de Oliveira Lacerda*
Zenaide de Oliveira Novais Carneiro**
Bruno Lopes de Almeida***
Taine do Rosário****

RESUMO

Livro do Gado e Livro de Razão (XVIII-XIX), do Brejo do Campo Seco, Bahia, fazem parte do banco de textos do Núcleo de Estudos de Língua Portuguesa (NELP) da Universidade Estadual de Feira de Santana (UEFS), desde 2012. Encontram-se, atualmente, em edição semidiplomática, segundo critérios de transcrição do Projeto para a História do Português Brasileiro (PHPB) (SANTOS, 2019; SILVA, 2021), e

Recebido em 21 de fevereiro de 2022.

Aceito em 8 de agosto de 2022.

DOI: <https://doi.org/10.18364/rc.2023n64.1075>

* Universidade Estadual de Feira de Santana, E-mail marianafagundes@uefs.br, ORCID: <https://orcid.org/0000-0003-4335-3458>

** Universidade Estadual de Feira de Santana, E-mail zoncarneiro@uefs.br, ORCID: <https://orcid.org/0000-0001-5990-4854>

*** Universidade Estadual de Feira de Santana, E-mail: tainerosario99@gmail.com, ORCID: <https://orcid.org/0000-0003-4551-5246>

**** Universidade Estadual de Feira de Santana, E-mail: brunouefs@outlook.com, ORCID: <https://orcid.org/0000-0002-8860-7653>

em edição modernizada – com uso da ferramenta eDictor (PAIXÃO DE SOUSA; KEPLER; FARIA, 2009) –, disponíveis no Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS) (<http://www.uefs.br/cedohs/>). Nesta oportunidade, será apresentado o trabalho de edição modernizada dos livros manuscritos e os desafios que as especificidades desse material colocaram à ferramenta computacional utilizada, levando a seu aprimoramento. Trata-se de contribuições que o CE-DOHS oferece à Nova Filologia (CRANE *et al*, 2008), haja vista o objetivo do projeto de constituir *corpora* diacrônicos, sintaticamente anotados, para estudo da história do português brasileiro, com a possibilidade de busca automática de dados.

PALAVRAS-CHAVE: Português brasileiro. Corpus eletrônico. Edição modernizada. eDictor.

ABSTRACT

Livro do Gado and *Livro de Razão* (XVIII-XIX), from Brejo do Campo Seco, Bahia, are part of the bank Documentos of the Nucleus of Portuguese Language Studies (NELP) of the State University of Feira de Santana (UEFS), since 2012. They are currently in a semi-diplomatic edition, according to transcription criteria of the Project for the History of Brazilian Portuguese (PHPB) (SANTOS, 2019; SILVA, 2021), and in a modernized edition – with the use of the eDictor tool (PAIXÃO DE SOUSA; KEPLER; FARIA, 2009) –, available in the Electronic Corpus of Historical Documents of Sertão (CE-DOHS) (<http://www.uefs.br/cedohs/>). In this opportunity, the work of modernized edition of the manuscript books will be presented, and the challenges that the specificities of this material posed to the computational tool used, leading to its improvement. These are contributions that CE-DOHS offers to the New Philology (CRANE *et al*, 2008), given the project's goal of constituting diachronic corpora, syntactically annotated, for the study of the history of Brazilian Portuguese, with the possibility of automatic data search.

KEYWORDS: Brazilian Portuguese. Electronic corpus. Modernized edition. eDictor.

Introdução

Neste trabalho, apresentam-se contribuições que o projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS) (FAPESB 5566/2010) – do Núcleo de Estudos de Língua Portuguesa (NELP) da Universidade Estadual de Feira de Santana (UEFS) – oferece à Nova Filologia (CRANE *et al*, 2008), por meio da edição modernizada dos dois livros

manuscritos do Brejo do Campo Seco, Bahia, do final do período colonial, a qual demandou o aprimoramento da ferramenta computacional eDICTOR (PAIXÃO DE SOUSA; KEPLER; FARIA, 2009), adotada pelo CE-DOHS.

O CE-DOHS, coordenado pelas professoras Zenaide Carneiro e Mariana Lacerda, disponibiliza, na rede mundial de computadores (<http://www.uefs.br/cedohs/>), com acesso livre e gratuito, extensa base documental para estudo do português do Brasil, organizada em dois conjuntos: conjunto 1 – composto por textos escritos entre 1823 e 2000, por indivíduos nascidos no Brasil, a partir de 1724, e por amostras de fala de brasileiros, gravadas na década de 90 do século XX, na Bahia; conjunto 2 – composto por manuscritos produzidos entre 1640 e 1822 por diferentes populações nascidas no Brasil, a partir de 1590. São, atualmente, mais de um milhão de palavras, em textos editados em linguagem XML, modernizados com uso do eDICTOR, a partir do texto-fonte, em edição semidiplomática (LACERDA; CARNEIRO; SANTIAGO, 2016; CARNEIRO; LACERDA (2019)).

Documentação de foro privado, o *Livro do Gado* e o *Livro de Razão* – os dois livros manuscritos do Brejo do Campo Seco (BA), séculos XVIII-XIX – constituem, segundo Santos Filho (2012, p. 110), uma valiosa porção do arquivo do sobrado do Brejo. Esses documentos raros, escritos por três gerações, fazem parte do segundo conjunto de textos do CE-DOHS, encontrando-se concluídas, neste momento, as edições semidiplomáticas (SANTOS, 2019; SILVA, 2021; LACERDA, SANTOS, 2022; CARNEIRO, SILVA, SOUZA, 2022) e parcialmente realizadas as edições modernizadas, estas aos cuidados de Bruno de Almeida e Taine do Rosário, no âmbito do Programa de Iniciação Científica da UEFS, conforme consta nos metadados do processamento dos materiais, no site do projeto.

Está organizado o artigo da seguinte forma: na primeira seção, apresenta-se, em síntese, o trabalho de constituição de *corpus* para estudo da história do português brasileiro, no âmbito do NELP-UEFS, a começar pelo banco Documentos Históricos do Sertão (DOHS) criado em 1998, que ganhou a versão eletrônica CE-DOHS, em 2012. Na seção 2, estão

apresentados o *Livro do Gado* e o *Livro de Razão*. A terceira seção traz o processo de edição modernizada dos livros, os desafios que se colocaram e as soluções encontradas. Uma síntese das contribuições que o CE-DOHS dá à Nova Filologia, por meio dessa edição, está na seção 4. A ela seguem as considerações finais e as referências bibliográficas consultadas.

1. Constituição de *corpus* de PB no NELP/UEFS

O Núcleo de Estudos de Língua Portuguesa (NELP) – do Departamento de Letras e Artes (DLA) da UEFS –, coordenado, desde 2017, pela professora Mariana Lacerda, foi criado pelas professoras Norma Almeida e Zenaide Carneiro, em 1998, a partir das pesquisas desenvolvidas no âmbito do projeto A Língua Portuguesa no Semiárido Baiano, por elas coordenado e fruto de estudos iniciados, na instituição, pela professora Ilza Ribeiro e pelo professor Dante Lucchesi, sobre a constituição de banco de dados para investigação da história do português brasileiro. Hoje, fazem parte do NELP diversos projetos de pesquisa, em parceria com outras universidades brasileiras e universidades estrangeiras (conferir site do NELP: <https://nelpuefs.wordpress.com/>).



Figura 1: Página inicial do site NELP/UEFS. Fonte: <<https://nelpuefs.wordpress.com/>>

Entre os projetos do NELP, está o Projeto Vozes do Sertão em Dados: história, povos e formação do português brasileiro (CNPq 02/2009), coordenado, na fase 1, pela professora Zenaide Carneiro, e, nas fases 2, 3, 4 e 5, também pela professora Mariana Lacerda. Com o projeto Vozes, o banco de dados do NELP, intitulado Documentos Históricos do Sertão (DOHS), foi ampliado a partir da prospecção documental e da edição semidiplomática de textos, realizadas tanto no âmbito da Pós-Graduação como no Programa de Iniciação Científica da UEFS, ao longo de mais de 10 anos. As edições semidiplomáticas são o texto-fonte para as edições modernizadas, desenvolvidas como produto do projeto CE-DOHS – já referido no início desde trabalho –, que tem por objetivo o desenvolvimento da versão eletrônica do banco DOHS, disponibilizando para pesquisadores interessados a versão semidiplomática e a versão modernizada de *corpora* diacrônicos do português brasileiro, os quais vêm sendo, tendo em vista a busca automática de dados, anotados sintaticamente, em parceria com o projeto Corpus Histórico do Português Tycho Brahe(UNICAMP), coordenado pela professora Charlotte Galves.



Figura 2: Página inicial do site CE-DOHS. Fonte: <<http://www.uefs.br/cedohs/>>

O CE-DOHS, criado em 2012, pioneiro no Nordeste brasileiro, possui um conjunto documental bastante significativo, com controle sócio-histórico, em colaboração com o Projeto Nacional para a História do Português Brasileiro (PHPB); destacam-se, no *corpus*, os documentos de natureza epistolar:

A maior parte dos documentos do DOHS, datados e localizados – que hoje se encontram também em versão digital no CE-DOHS – são cartas manuscritas, dos séculos XIX e XX (1084 cartas, 422 remetentes), editadas sobretudo por Carneiro (2005), que investiu na busca e na organização de acervos documentais que pudessem contribuir para o processo de reconstrução sócio-histórica do PB, em um trabalho de investigação grandioso, percorrendo diversos arquivos, e publicadas em 2011, na obra, com três volumes, organizada por Carneiro, Lacerda, Almeida, Santiago e Oliveira, intitulada *Cartas brasileiras: coletânea de fontes para o estudo do português*(LACERDA, CARNEIRO, SANTIAGO, 2016, p. 131).

O NELP, como se vê, dispõe de um banco de *corpora* escritos (e também orais) – os *corpora* orais são produto, em sua grande parte, do projeto A Língua Portuguesa no Semiárido Baiano (1993-2016), coordenado pelas professoras Norma Almeida e Zenaide Carneiro – que possibilita estudos diversos (considerando informações do autor/informante e do documento/ amostra de fala) sobre a história do português brasileiro, particularmente do português no semiárido baiano:

No que diz respeito à constituição de *corpora* diacrônicos, os projetos A Língua Portuguesa no Semiárido Baiano, Vozes do Sertão em dados e CE-DOHS, da UEFS, destacam-se no PHPB-Bahia, apresentando material criteriosamente transcrito e editado, somando mais de um milhão de palavras. É uma grande contribuição aos estudos do português brasileiro, em sua realidade plural e polarizada entre normas vernáculas e normas cultas (LUCCHESI, 2015); de forma mais específica, ao estudo do processo de formação da língua portuguesa no espaço do semiárido baiano (CARNEIRO, LACERDA, 2019, p. 216).

O PHPB-Bahia, sem dúvida, no atendimento às três agendas do projeto nacional – o campo histórico-filológico, o campo gramatical e o campo da história social linguística (LOBO, 2009, p. 306), conta com o trabalho profícuo da equipe de pesquisadores do NELP-UEFS:

Com mais de 20 anos de existência, o NELP-UEFS, em seus diferentes projetos, já produziu uma variedade de trabalhos e tem contribuído, com rigor científico e seriedade, com os estudos sobre a língua portuguesa, no âmbito da Linguística Histórica, tanto com *corpora* escritos como com *corpora* orais (SANTIAGO; LACERDA; CARNEIRO, 2019, p. 366)

Como observam Lacerda, Carneiro e Santiago (2016, p. 135),

A aproximação entre o campo filológico e o campo computacional – observada desde a década de 1990 – encontra-se atualmente em plena expansão. O trabalho em ambiente digital no campo da Filologia e da Linguística Histórica tem sido cada vez mais significativo, fazendo surgir, segundo Crane *et al.* (2008), uma nova Filologia, a *e-philology*, ou determinando, de acordo com Schreibman *et al.* (2004), o nascimento das Humanidades Digitais.

Na era das Humanidades Digitais, o projeto CE-DOHS desenvolve a Linguística de Corpus e a Linguística Computacional (SARDINHA, 2000), montando um banco, cujas etapas são: localização e seleção de documentos (com controle sócio-histórico); transcrição segundo normas filológicas conservadoras; edição digital (modernizada), em linguagem XML; preenchimento dos metadados do documento e de seu processamento.

A edição em linguagem XML – etapa importante na formação de bancos eletrônicos – é feita com uso do eDictor (PAIXÃO DE SOUSA; KEPLER; FARIA, 2009), que combina um editor de XML e um etiquetador morfossintático e permite a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em HTML), e de versões com anotação morfossintática (em texto simples e XML). Para Paixão de Souza, Kepler e Faria (2009), a edição em linguagem XML busca ser o mais neutra possível, em relação ao conteúdo textual codificado, atendendo a

necessidades linguísticas e filológicas. Todas as intervenções feitas no texto, com uso do eDictor, ficam visíveis ao leitor, o que possibilita seu controle e mapeamento, garantindo a recuperabilidade das formas originais.

Atualmente, o eDictor, para o processo de modernização dos textos (que diz respeito apenas a intervenções de natureza gráfica), apresenta onze tipos de etiquetas: junção, segmentação, sobrescrito, ilegível, rasurado, subscrito, tachado, pontuação, expansão, padronização e modernizado.

Na seção 3, está apresentado – tomando, como exemplo, o *Livro do Gado* e o *Livro de Razão* do Brejo do Campo Seco (BA) – o passo-a-passo da edição eletrônica, em linguagem XML, como é realizada no projeto CE-DOHS.

Antes, contudo, apresenta-se uma breve descrição dos livros manuscritos em questão (seção 2).

2. A porção mais valiosa do arquivo do Sobrado do Brejo: o *Livro do Gado* e o *Livro de Razão*

De acordo com Santos Filho (2012, p. 110),

São dois os livros manuscritos que se conservaram no arquivo do Sobrado do Brejo. Velhos memoriais de 200 anos, manuseados e enebados quotidianamente pelas mãos de três gerações, guardados depois pela quarta e pelas seguintes, metamorfosearam-se hoje em preciosas relíquias, preciosas testemunhas de remoto passado.

Essa documentação de foro privado – o *Livro do Gado* (57 folhas), de 1755-1832¹, e o *Livro de Razão* (195 folhas), de 1795-1838² –, de que Santos

1 Escreveram no *Livro do Gado* o português Miguel Lourenço de Almeida (primeiro senhor do Brejo), de 1755 a 1785, e os brasileiros Antônio Pinheiro Pinto (segundo senhor), de 1794-1822, e Inocêncio José Pinheiro ou Inocêncio Pinheiro Canguçu (terceiro senhor), de 1822 a 1832.

2 Escreveu no *Livro de Razão* principalmente Antônio Pinheiro Pinto, de 1795 a 1821; a segunda mão a escrever foi de Inocêncio Pinheiro Canguçu, que escreveu

Filho (1956) dá notícia ao leitor, apresentando *Uma comunidade rural do Brasil antigo*, encontra-se atualmente sob a guarda de Lycurgode Castro Santos Neto, na cidade de Campinas, São Paulo.

O projeto CE-DOHS teve acesso aos documentos em 2012 e a autorização para editá-los e disponibilizar as edições na rede mundial de computadores. Foram os livros fotografados por Jorge Viana, segundo o método Lapelinc (SANTOS E BRITO, 2014). Lacerda e Carneiro (2016, p. 152) apontam essa agenda de trabalho:

São apresentados, aqui, os critérios adotados de edição semidiplomática dos referidos documentos, bem como a metodologia de edição digital, usada no âmbito do projeto CE-DOHS – *Corpus Eletrônico de Documentos Históricos do Sertão*, do qual os livros fazem parte. Essas edições servirão não apenas a pesquisadores interessados em aspectos linguísticos, da difusão da escrita, da leitura, das transmissões textuais, mas em aspectos históricos, políticos, econômico-sociais, entre outros. (...) Atualmente, já se encontram feitas a descrição extrínseca e descrição intrínseca do *Livro do Gado* e do *Livro de Razão*; ainda por concluir, a caracterização sócio-histórica da fazenda do Brejo do Campo Seco e a caracterização sociocultural dos escreventes, respondendo, de forma sistemática, às questões-problemas de Petrucci (2003). Em fase de execução, acha-se a edição semidiplomática dos livros. A etapa seguinte consiste na edição modernizada dos documentos, usando o eDictor.

Nos últimos 6 anos, foram feitas as edições semidiplomática (de acordo com normas de transcrição do PHPB, adaptadas) e modernizada dos livros manuscritos. As edições semidiplomáticas do *Livro do Gado* e do *Livro de Razão* foram feitas, respectivamente, por Santos (2019) e por Silva (2021), e as edições modernizadas, por Taine do Rosário – a edição do livro com

pouco, durante o período de 1822-1838. Segundo Santos Filhos (2021, p. 115), o último registro feito no *Livro de Razão*, quase duas décadas depois de Inocêncio Pinheiro Canguçu, é de Miguel Joaquim de Castro Mirante – genro de Canguçu –, “um indivíduo sorumbático e cismador”. Na terceira capa, escreveu, depois de revistar o livro: “Vistos e revistos q’ do ao fougo.”

anotações pastoris – e por Bruno Lopes, a edição do livro de contabilidade (esta se acha incompleta neste momento, pelas razões descritas em 3.2).

Documentação rara, esses livros escritos por três gerações, na fazenda de criação do Brejo do Campo Seco, da família Pinheiro Canguçu – povoado de Bom Jesus dos Meiras, Comarca de Rio de Contas (hoje Brumado), sertão baiano – têm inegável valor histórico, econômico, social e linguístico. Para Santos Filho (2012, p. 20), segundo o qual “como que armado de uma grossa lente de aumento, procurei focalizar, em ponto grande, um pequeno centro nordestino do Brasil antigo” (p. 19), o *Livro do Gado* e o *Livro de Razão* “são a porção mais valiosa do opulento arquivo do sobrado do Brejo”.

3. Edições modernizadas para o CE-DOHS: desafios e soluções

Nas seguintes subseções, serão descritos os desafios e soluções que se colocaram durante o processo de edição modernizada do *Livro do Gado* e do *Livro de Razão* do Brejo do Campo Seco. Também os desafios ainda sem solução efetiva, no que diz respeito à edição do *Livro de Razão*.

O *Livro do Gado*, como dito anteriormente, teve a sua edição filológica realizada por Santos (2019), como parte de sua dissertação de mestrado, defendida no âmbito do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) da Universidade Estadual de Feira de Santana (UEFS). As normas de transcrição adotadas pela autora seguiram os critérios do PHPB, tendo sido necessárias adaptações, em razão de especificidades do livro, especificidades essas que, até então, não eram encontradas em outros documentos editados a partir das normas do projeto. Essas adaptações – que buscavam preservar as características do manuscrito, em uma edição voltada, especialmente, aos estudos linguísticos – colocaram desafios no processo de edição modernizada do *Livro do Gado* da família Pinheiro Canguçu.

Tendo em mãos os fólios em edição semidiplomática (texto-fonte), em formato docx, como é possível observar na figura 3, fez-se a mesclagem das

células da tabela e das notas de rodapé, passo necessário, considerando dois motivos principais: a incompatibilidade das tabelas com o eDicator (notou-se que, ao transferir o documento para o *software*, sem fazer essa retirada, ocorria um erro que acarretava a perda da estrutura do texto); a disponibilização dos documentos na plataforma CE-DOHS (as informações presentes nas notas de rodapé devem fazer parte dos comentários sobre o fôlio).

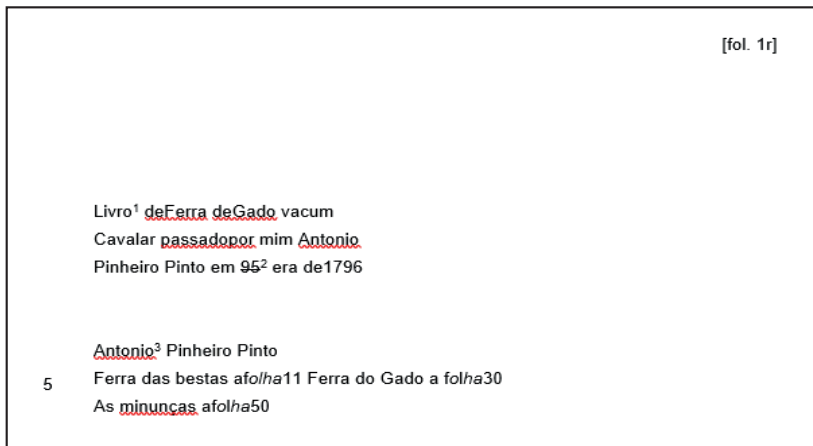


Figura 3: Edição semidiplomática do fôlio 1. Fonte: SANTOS (2019).

As modificações supracitadas, no âmbito filológico, não alteram o valor linguístico. Passou-se, sem dificuldades, para a próxima etapa do processo.

Transferiu-se o texto para a aba de transcrição do eDicator – conforme demonstrado na figura 4 –, onde é possível fazer a quebra de linhas, sentenças e parágrafos dos documentos, além de transformar as notas de rodapé em comentários, que figuram como uma espécie de cabeçalho descritivo do fôlio; a partir daí, já é possível gerar o XML do arquivo e começar a execução das intervenções necessárias, na modernização do documento.

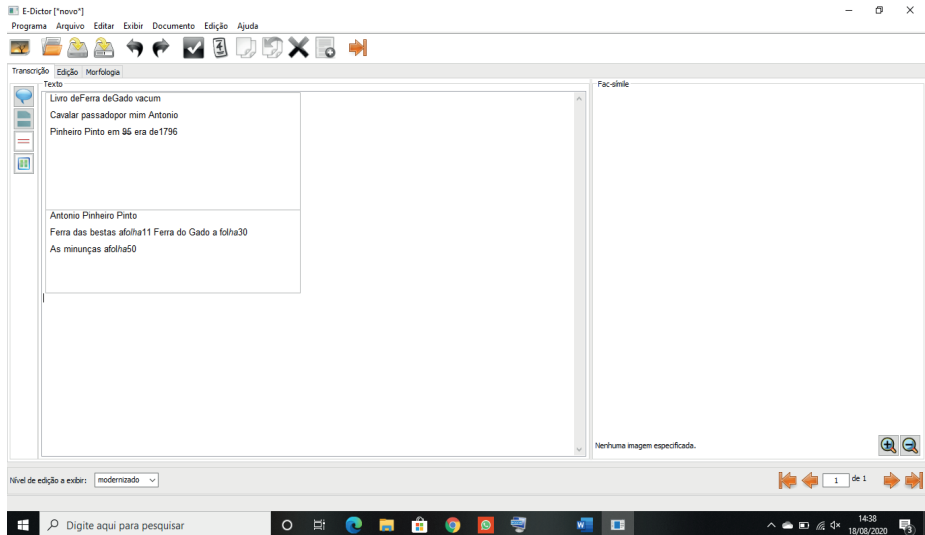


Figura 4: Aba de transcrição do eDICTOR. Fonte: Elaboração própria.

O eDICTOR apresenta a possibilidade de alteração das palavras, em vários níveis, ficando todas as intervenções disponíveis ao consulente, que pode acessar a versão (original ou modernizada) que melhor atender o seu interesse de pesquisa. Na figura 5, é possível perceber que algumas palavras estão em vermelho, indicação dada, pelo *software*, de que elas sofreram algum tipo de intervenção, em um ou mais níveis de edição, o que pode ser verificado no canto esquerdo superior da tela do programa, por meio da lista de edições.

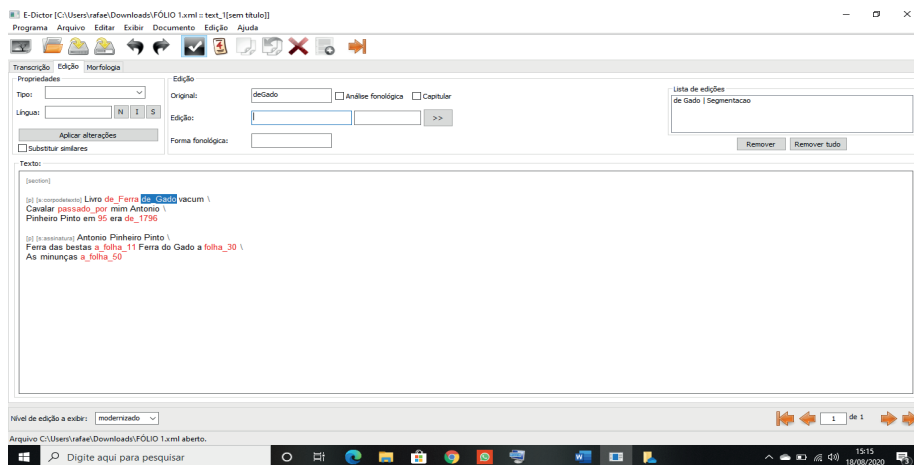


Figura 5 – Edição em XML do fôlio 1 do *Livro do Gado*. Fonte: Elaboração própria

Ainda observando a figura 5, e tomando como exemplo as palavras “do” e “gado”, pode-se conferir a lista de edições e verificar que o nível de edição presente é uma segmentação, o que implica dizer que as palavras foram grafadas juntas pelo escrevente e separadas na edição modernizada. Além dessas, há outras palavras com destaque em vermelho no fôlio em questão, que sofreram intervenções; para saber quais níveis foram aplicados, o consutente deve clicar na palavra, e sua respectiva lista de edição será disponibilizada.

O processo de etiquetagem, usando o eDictor, é de extrema importância, haja vista que se trata da preparação final do documento para a anotação morfossintática, que virá em seguida; as etiquetas, por essa razão, são pensadas para atender todas as especificidades dos documentos. Não havia, entretanto, no primeiro momento, etiquetas que contemplassem todas as especificidades do *Livro do Gado*.

edição modernizada, tendo sido necessário solicitar a criação de um novo arquivo de configurações do *software* que apresentasse essa e outras adições.

Para além das dificuldades decorrentes do uso de símbolos, existiram também percalços que foram notados ainda no processo de edição semidiplomática. Na figura 7, a seguir, está apresentada a edição modernizada do fôlio 3r; nela, é possível observar algumas marcações, como [inint], que, de acordo com os critérios do PHPB, indica um trecho do documento que não apresenta rasuras, mas que não pode ser decifrado, em virtude de algum outro fator.

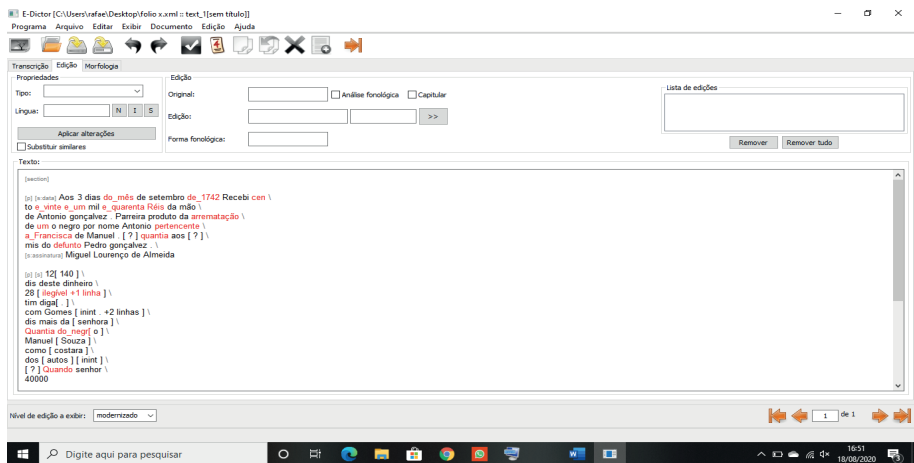


Figura 7: Edição em XML do fôlio 3r do *Livro do Gado*.

A edição modernizada do *Livro do Gado* foi um caminho de grandes desafios, resolvidos a partir do aperfeiçoamento da ferramenta eDactor.

Como já se disse, o texto-fonte de uma edição modernizada é a edição semidiplomática do documento. Com o *Livro de Razão* do Brejo do Campo Seco não seria diferente. A edição semidiplomática do *Livro de Razão*, que serviu ao trabalho de edição modernizada do material, aqui apresentada, foi realizada por Silva (2021), como parte de sua tese de

doutorado, defendida no Programa de Pós Graduação em Língua e Cultura (PPGLINC) da Universidade Federal da Bahia (UFBA). A figura 8 traz a edição semidiplomática do fôlio 17r:

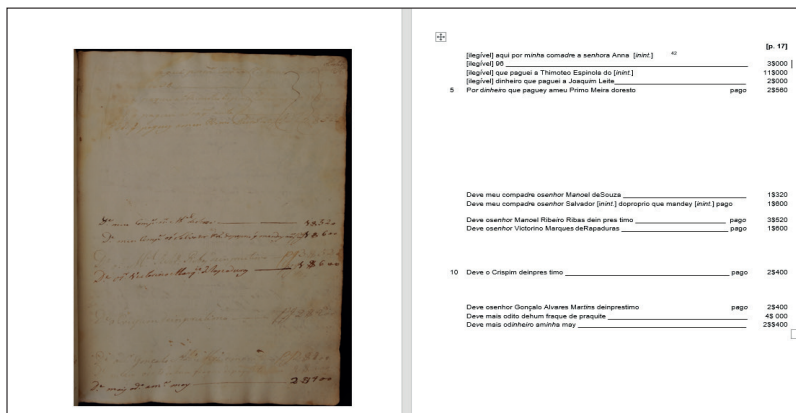


Figura 8: Edição semidiplomática do fôlio 17r. Fonte: SILVA (2021)

Transcreveu-se o texto para a aba de transcrição do eDictor e apagaram-seas informações que dizem respeito apenas à edição semidiplomática, como os números referentes à contagem de linhas, as tabelas e outras particularidades do labor filológico tradicional.

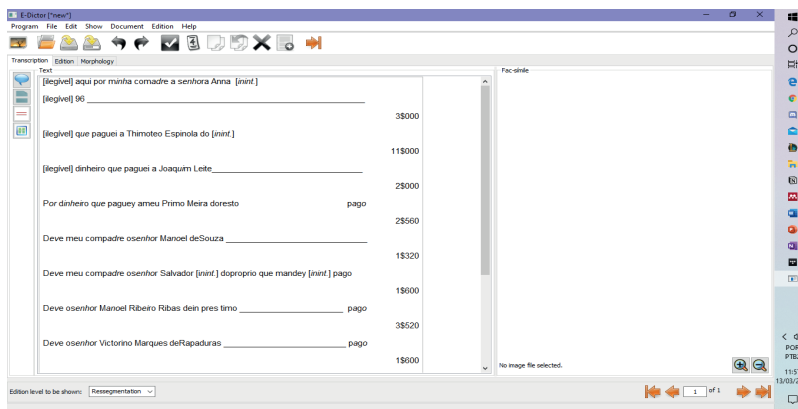


Figura 9: Aba de transcrição do eDictor.

Prosseguindo – com a contagem de linhas suprimida, as células mescladas e o texto devidamente transcrito –, gera-se a versão em XML, na qual a edição modernizada será efetivada. Ainda nesse processo inicial, já é possível observar peculiaridades do *Livro de Razão*. Ao contrário de procurações, atas e cartas (gêneros para os quais o eDictor foi, preliminarmente, pensado), o livro mais espesso da família Pinheiro Canguçu apresenta valores, extensas linhas, e suas sentenças, frequentemente, não obedecem à uma continuidade lógica ao leitor; como observado por Foisl (2009, p. 334), livros de razão não apresentam uma narrativa dentro de uma sucessão temporal continuada.

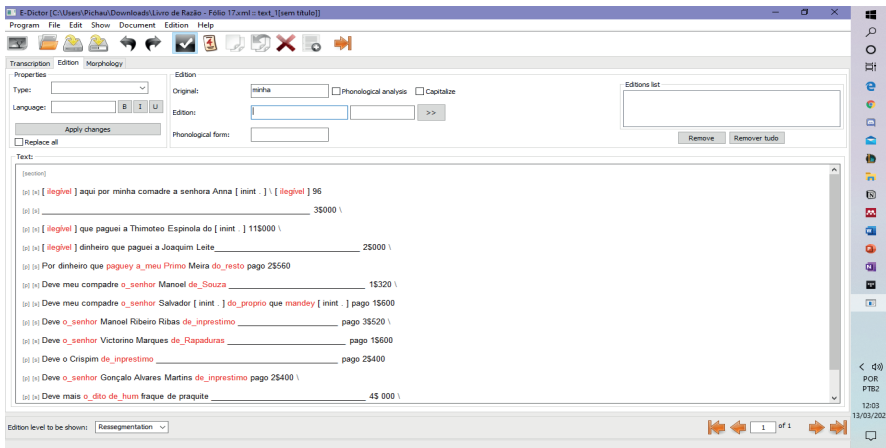


Figura 10: Edição em XML do fôlio 17r.

A cor vermelha, como já se disse, evidencia as palavras que sofreram modificações no processo de etiquetagem.

Com as etiquetas disponíveis no eDictor, foi possível realizar apenas a edição de 51 dos 195 fôlios do *Livro de Razão*. Como o *Livro do Gado*, de anotações pastoris, o livro de contabilidade do Brejo também apresenta muitos símbolos, laçadas, arabescos. Veja-se o fôlio 194r:

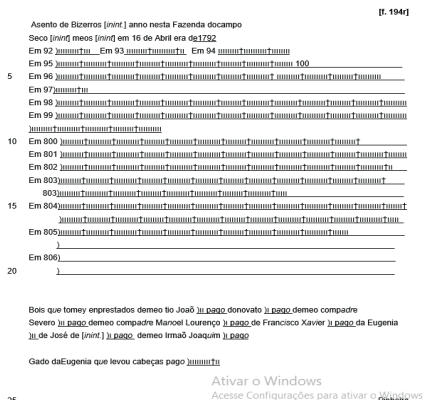
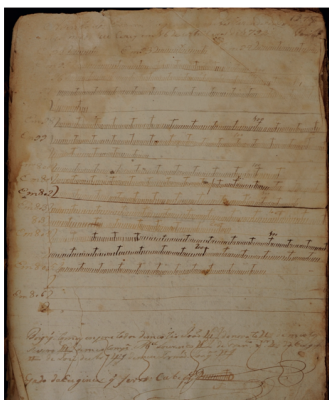


Figura 11 – Edição semidiplomática do fólio 194r. Fonte: SILVA (2021)

No referido fólio, o escrevente realiza o que parece ser a contagem escrita do “Asento de Bizerros” na fazenda do Campo Seco, naquela ocasião. Aqui, as mesmas dificuldades encontradas quando da edição modernizada do *Livro do Gado*: ausência de etiqueta para símbolos de contagem e para [inint]. Foi recorrente, em quase todo o documento, a aparição de símbolos e expressões não previstas pelo eDictor.

Do verso da capa ao fólio 42, o documento obedece à uma estrutura indispensável ao funcionamento do eDictor: a presença de apenas uma sentença por linha, em formato de coluna única. A partir do fólio 43r, no entanto, o livro passa a ser dividido, aleatoriamente, em duas ou até três colunas, fato que impediu a edição modernizada completa do *Livro de Razão*.

		[p.43]	
Deve [inint.] Lopes deSouza o seguinte		Deve o senhor Manoel Mathias	
	Por 10 varas depano deAlgodão --	2\$000	[inint.] que vem detras 38840
	[ilegível] couro diviado -----	\$160	do credito de Joaõ
	dofeito das calças deAlgodão ---	\$320	[inint.] + linha
5	Por dinheiro deinpres timo -----	\$640	deResto ---- 13400
	Por vara emeya depano deAlgodão	300	dehum credilo
	Por mey acoarta depolvora ----	320	demeu compadre
	Por meya vara depano ----	120	Manoel Lourenco ---- 30000
	[.] 1 couro deviado mateiro ----	480	deoutro credito
10	Por meya vara depano deAlgodão	400	[inint.] meu compadre 8000
	Por 1 xapeo de braga ----	2000	Soma 90240
	Por dinheiro que dey ao Lan din	800	69080
	Por dinheiro que aoMestre Jozé	800	21160
	Por dinheiro que dey aoPedro dos[inint]	1600	Eu devo odinheiro [?]
15	Soma ⁷⁹	9480	Manoel Mathias que vendeoutra Lau da ⁸⁰ restame odinheiro que pasou credito

Figura 12: Edição semidiplomática do fôlio 43r. Fonte: Elaboração própria

Observe-se que, na linha destacada na figura 12, as orações “Deve [inint.] Lopes de Souza o seguinte” e “Deve o senhor Manoel Mathias” não têm relação sintático-semântica; a continuidade sintático-semântica de cada uma aparece na linha subsequente, em suas respectivas colunas.

```
[p] [s] De ve [ inint . [s] ] Lopes deSouza o seguinte De ve o senhor Manoel Mathias \
Por 10 varas depano deAlgodão -- 2$000 [ inint . [s] ] que vem detras 38840 \
[ ilegível ] couro diviado ----- $160 do credito de Joaõ \
dofeito das calças deAlgodão --- $320 [ inint . [s] ] + linha \
Por dinheiro deinpres timo ----- $640 deResto ---- 13400 \
Por vara emeya depano deAlgodão 300 dehum credilo \
Por mey acoarta depolvora ---- 320 demeu compadre \
Por meya vara depano ---- 120 Manoel Lourenco ---- 30000 \
[ . [s] ] 1 couro deviado mateiro ---- 480 deoutro credito \
Por meya vara depano deAlgodão 400 [ inint . [s] ] meu compadre 8000 \
Por 1 xapeo de braga ---- 2000 Soma 90240 \
Por dinheiro que dey ao Lan din 800 69080 \
Por dinheiro que aoMestre Jozé 800 21160 \
Por dinheiro que dey aoPedro dos[ inint ] 1600 Eu devo odinheiro [ ? [s] ] \
Soma 9480 Manoel Mathias \
que vendeoutra \
Lau da \
restame odinheiro \
que pasou credito
```

Figura 13: Edição em XML do fôlio 43r. Fonte: Elaboração própria

Observando a figura 13, percebe-se que o eDictor não é capaz de reconhecer a divisão do texto em colunas, construindo uma coluna única eviolando um dos princípios da edição: o não comprometimento dos aspectossintáticos e semânticos. As frases que anteriormente pertenciam a colunas separadas e continuavam na linha seguinte agora estão fundidas na mesma sentença “Deve [inint.] Lopes deSouzaoseguinte Deve o senhor Manoel Mathias”. Confira-se destaque em azul, na figura 13. Apresentaram-se algumas soluções para essa incompatibilidade, mas, até o presente momento, nenhuma se mostrou realmente efetiva. A equipe de desenvolvimento do eDictor segue trabalhando, tendo em vista o aperfeiçoamento da ferramenta e a conclusão da edição modernizada do *Livro de Razão* do Brejo do Campo Seco.

4. Contribuições à Nova Filologia: uma síntese

Apontam-se aqui, objetivamente, as contribuições que o processo de edição modernizada do *Livro do Gado* e do *Livro de Razão*, no âmbito do projeto CE-DOHS, deram ao aperfeiçoamento do eDictor e à Nova Filologia:

- a) A elaboração de um novo arquivo de configurações, contendo as etiquetas necessárias para o desenvolvimento das edições citadas e que, no futuro, poderá ser utilizado para realizar trabalhos com outras documentações que apresentem características similares.
- b) A indicação da necessidade de uma reestruturação na ferramenta eDictor, para que seja possível a interpretação de documentos que se organizem de maneira diferente aos comumente editados (a exemplo das cartas, testamentos e atas).
- c) A colaboração com a Nova Filologia, por meio da ampliação do banco CE-DOHS.
- d) A inserção dos documentos do Acervo do Brejo do Campo Seco no campo das Humanidades Digitais, possibilitando o conhecimento, a difusão e a conservação das informações acerca do cotidiano do Sobrado do Brejo do Campo Seco, no final do período colonial.

Considerações finais

Lobo e Oliveira (2012, p. 3) afirmam que “reconstruir a história linguística do Brasil passa, também, por reconstruir a história da penetração da língua escrita no Brasil”. O projeto CE-DOHS tem, entre seus objetivos, colaborar com os estudos sobre a penetração e difusão da escrita no interior da Bahia, e os esforços dedicados à edição do *Livro do Gado* e do *Livro de Razão* do Brejo do Campo Seco trouxeram, até aqui, resultados importantes, conforme demonstrado, nesta síntese.

No domínio das Humanidades Digitais, a edição modernizada dos livros – focalizada neste artigo – trata-se de um trabalho muito importante na constituição do *corpus* eletrônico do NELP/UEFS. Essa edição, com uso do eDicator, apresentou desafios à Nova Filologia – dadas as especificidades da documentação –, para os quais se buscaram respostas, por meio do aperfeiçoamento da ferramenta.

O CE-DOHS – numa abordagem inter-relacionada entre língua, histórica e tecnologia – oferece um banco de dados sociolinguísticos, representativo do período colonial e do período imperial no Brasil; do período colonial são o *Livro do Gado* e o *Livro de Razão* aqui considerados, os primeiros documentos coloniais a fazer parte do banco do NELP. Uma rara e valiosa documentação do Brejo do Campo Seco, apresentada em diferentes versões de edição pela equipe da UEFS, que tem a grata satisfação de oportunizar à comunidade científica chegar mais perto de uma fazenda de criação dos sertões da Bahia e de sua riqueza arquivística, em 150 anos de história, dentro dos séculos XVIII e XIX.

Referências

ALVES, D. **As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico:** dos exemplos internacionais ao caso português. Lisboa: Ler História, 2016.

CARNEIRO, Z. O. N.; LACERDA, M. F. O. (Org). **CE-DOHS - Corpus Eletrônico de Documentos Históricos do Sertão (2012-2025)**. URL: <http://www.uefs.br/cedohs>. Acesso em 19 abr 2021.

CARNEIRO, Z. de O. N.; LACERDA, M. F. de O. **Corpus Eletrônico de Documentos Históricos do Sertão: etapa 1 (1750-2000)**. *Revista Binacional Brasil-Argentina:Diálogo entre às Ciências*, v. 8, p. 205-221, 2019.

CARNEIRO, Z. O. N., SILVA, A. J.; SOUZA, E. H. P. (Org.). **Livro de razão (XVIII-XIX):** edições fac-similar e semidiplomática: volume 2. Feira de Santana : UEFS Editora, 2022. 380 p. Coleção: O Sertão por Escrito: edição filológica dos livros do arquivo do Sobrado do Brejo (Bahia). Coordenação geral: Zenaide de Oliveira Novais Carneiro, Mariana Fagundes de Oliveira Lacerda.

CRANE, G. (et al.). **ePhilology: when the boooks talk to their readers**. BlackwellCampanionto Digital LiteraryStudies. Oxford: Blackwell, 2008.

FARIA, P.; GALVES, C. Criando “Bancos de Árvores”: O Sistema de Anotação e o Processo Automático. **Cadernos de Estudos Linguísticos**. Campinas: v. 58, n. 2 p. 299-315, maio/ago./2016. Disponível em 21 Feira de Santana, v. 19, n. 2, p. 8-2, 2018 <http://revistas.iel.unicamp.br/index.php/cel/article/view/5133>. Acesso em 25 mar. 2018.

FOISIL, M. A escritura do foro privado. In: ARIÈS, P.; CHARTIER, R. (Org.). **História da vida privada: Da Renascença ao Século das Luzes**. São Paulo: Companhia das Letras, 2009. v. 3. p. 331-369.

GONÇALVES, M. F.; BANZA, A. P. (org.). **Patrimônio textual e humanidades digitais:da antiga à nova Filologia**. Évora: CIDEHUS, 2013.

LACERDA, M. F. O.; CARNEIRO, Z. O. N. Edição filológica e digital do Livro do Gado e do Livro de Razão do Arquivo do Sobrado do Brejo (Bahia setecentista e oitocentista). In: **Labor Histórico**, Rio de Janeiro, 2, 2016. p. 151-163.

LACERDA, M. F. O.; CARNEIRO, Z. O. N.; SANTIAGO. **Núcleo de Estudos de Língua Portuguesa**. URL: <https://nelpuefs.wordpress.com/>. Acesso em 19 abr 2021.

LACERDA, M. F. O.; SANTOS, E. B. (Org.). *Livro do gado (XVIII-XIX):* edições fac-similar e semidiplomática: volume 1. Feira de Santana: UEFS Editora, 2022. 118 p. Coleção: O Sertão por Escrito: edição filológica dos livros do arquivo do Sobrado do Brejo (Bahia). Coordenação geral: Zenaide de Oliveira Novais Carneiro, Mariana Fagundes de Oliveira Lacerda.

LEAL, I; REZENDE, G. R. P; CARNEIRO, Z. O. N.; FARIA, P.; LACERDA, M. F. O. **A ferramenta de busca E-CORP aplicada ao Corpus Eletrônico de Documentos Históricos do Sertão**. A COR DAS LETRAS (UEFS), v. 19, p. 8-21, 2019.

LOBO, T. C. F. Arquivos, acervos e a reconstrução histórica do português brasileiro. In: OLIVEIRA, K.; CUNHA E SOUZA, H. F.; SOLEDADE, J. (Org.). **Do português arcaico ao português brasileiro: outras histórias**. Salvador: EDUFBA, 2009. p. 305-327.

MATTOS E LOBO, T. C. F.; OLIVEIRA, K. **História da cultura escrita no Brasil: um programa de investigação/HISCULTE**, 2012. No prelo.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. E-dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: **Anais do VIII Encontro de Linguística de Corpus**, realizado na UERJ, 13 a 14 de novembro de 2009. Rio de Janeiro, 2009. p. 69-105.

SANTOS, E. B. **O Livro do Gado do brejo do campo seco (Bahia):** edição semidiplomática e descrição de índices grafo-fonéticos. Dissertação (Mestrado em Estudos Linguísticos), UEFS, 2019.

SANTOS, J. V.; BRITO, G. S. Fotografia técnica de documentos para formação de corpora digitais eletrônicos: o método desenvolvido no Lapelinc. **Letras & Letras**, v. 30, n. 2, p. 421-430, 18 dez. 2014.

SANTIAGO, H. S.; LACERDA, M. F. O.; CARNEIRO, Z. O. N. A Filologia e a História das Línguas: contribuições do Núcleo de Estudos de Língua Portuguesa da UEFS. **Macabéa - revista eletrônica**, v. 8, p. 352-371, 2019.

SANTOS FILHO, L. **Uma comunidade rural do Brasil antigo**: aspectos da vida patriarcal no Sertão da Bahia, nos séculos XVIII e XIX. Salvador: UEFS: Fundação Pedro Calmon, 2012.

SARDINHA, Tony Berber. **Linguística de corpus**: histórico e problemática. *D.E.L.T.A.*, São Paulo, 2000. v. 16. n. 2. p. 323-367. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005>. Acesso em: 11 mar 2020.

SILVA, A. J. **O Sertão por Escrito no Livro de Razão**: um microcosmo sócio-histórico e linguístico da Bahia rural oitocentista. Tese (Doutorado em Língua e Cultura), UFBA, 2021.