

## A elaboração de uma edição digital d’*Os Lusíadas*

**Bruno Maroneze**  
UFGD / Dourados, MS  
brunomaroneze@ufgd.edu.br

**Vanessa Martins do Monte**  
USP / São Paulo, SP  
vmmonte@usp.br

**André Bertacchi**  
UFJF / Juiz de Fora, MG  
andre.bertacchi@ufjf.br

**Artur Costrino**  
UFOP / Ouro Preto, MG  
artur.costrino@ufop.edu.br

**Alexandre Agnolon**  
UFOP / Ouro Preto, MG  
alexandre.agnolon@ufop.edu.br

**Mário Eduardo Viaro**  
USP / São Paulo, SP  
maeviaro@usp.br

### Resumo

This article presents the *Lusíadas Digital* project, which proposes the development of a virtual philological edition of *Os Lusíadas* by Luís de Camões, integrating principles of textual criticism, Digital Humanities, and Natural Language Processing (NLP). The project aims to develop a digital platform bringing together facsimiles of the 1572 editions, diplomatic and modernized transcriptions, a dynamic critical apparatus, a lexical glossary with etymological information, historical and literary commentary, and translations aligned with the original text. The methodology combines traditional philological practices with XML-TEI text encoding, OCR techniques, automatic lemmatization, version alignment, and lexical mining. Initially focused on Canto I, the project seeks to establish a scalable and replicable model for the remaining cantos of the work. By proposing an open, interoperable, and data-oriented digital infrastructure, the initiative contributes to the advancement of e-Philology in Brazil and to the development of technologies applied to the digital critical editing of manuscripts and early printed editions.

### 1 Introdução

A digitalização de obras clássicas deixou de ser mero processo de reprodução fac-similar e estimulou a constituição de um campo de investigação interdisciplinar, as chamadas Humanidades Digitais. A edição filológica virtual (Monte e Paixão de Sousa, 2017) representa um dos produtos mais sofisticados de integração entre saber humanístico e tecnologia computacional.

Apesar da centralidade d’*Os Lusíadas* para a tradição literária da língua portuguesa, ainda não existe uma edição filológica virtual que integre, de forma sistemática e interativa:

- imagens fac-similares das primeiras edições;
- transcrições diplomáticas e modernizadas;
- aparato crítico comparativo;
- glossário lexical estruturado;
- traduções alinhadas;
- comentários histórico-literários incorporados à navegação.

O projeto *Lusíadas Digital* busca suprir essa lacuna, propondo uma edição filológica virtual que combine rigor crítico, modelagem formal de dados e ferramentas de Processamento de Linguagem Natural (PLN).

### 2 Problema filológico: as edições de 1572 e a tradição textual

O problema editorial d’*Os Lusíadas* é conhecido: há duas edições datadas de 1572, tradicionalmente identificadas pela orientação do pelicano no frontispício (sinistrógira e dextrógira). As diferenças entre ambas incluem:

- variantes ortográficas;
- divergências de pontuação;
- substituições lexicais;
- lapsos tipográficos;
- diferenças na disposição gráfica.

A dificuldade em determinar a precedência entre as edições gera implicações para a crítica textual e para a história da língua portuguesa. A edição crítica de Augusto Epiphanyo da Silva Dias (*Camões*,

1916) constitui referência incontornável, mas foi concebida para circulação impressa e para público especializado.

No ambiente digital, torna-se possível:

- visualizar variantes em paralelo;
- automatizar a detecção de divergências;
- representar formalmente a tradição textual;
- disponibilizar múltiplos níveis de leitura.

A edição virtual, portanto, não substitui a crítica tradicional, mas a potencializa.

### 3 Fundamentos teóricos: Filologia, e-Philology e modelagem textual

A Filologia tradicional sempre buscou a opção textual mais verossímil de acordo com evidências manuscritas e históricas, por meio da análise da tradição manuscrita ou impressa. Contudo, como propõem Crane et al. (Crane et al., 2008), as práticas filológicas precisam transformar-se qualitativamente diante das tecnologias digitais. A chamada e-Filologia envolve:

1. digitalização imagética de alta resolução;
2. transcrição estruturada e formalmente anotada;
3. representação explícita das variantes;
4. múltiplas formas de visualização;
5. interoperabilidade e reutilização dos dados.

A edição filológica virtual proposta ancora-se na definição de Monte e Paixão de Sousa (Monte and Paixão de Sousa, 2017), que usam o termo para se "referir ao objeto criado a partir de um trabalho que inclui a produção da réplica imagética digital do documento físico, a edição filológica digital, e as múltiplas possibilidades de representação final ou publicação digital. Todos esses processos são construídos com ferramentas e tecnologias computacionais, e são, portanto, 'digitais'; o conjunto dos processos forma o objeto que chamamos de virtual – pois que simula, representa, re-cria artificialmente os documentos originais.”.

A modelagem textual em TEI-XML (TEI Consortium, 2025) constitui padrão consolidado para representar desde estruturas textuais internas (como cantos, estrofes e versos) até variantes textuais e comentários críticos.

No caso d' *Os Lusíadas*, a marcação TEI permitirá representar, entre outras informações:

- diferenças entre as versões de 1572;
- distinção entre transcrição diplomática e modernizada;
- estrutura métrica;
- alinhamento com traduções.

## 4 Arquitetura da edição digital

A plataforma proposta está sendo organizada em módulos integrados:

### 4.1 Módulo fac-similar

Este módulo possibilitará:

- Visualização de imagens (tanto da edição sinistrógira quanto da dextrógira).
- Navegação por fólio e estrofe.
- Sincronização entre imagem e transcrição.

### 4.2 Módulo textual

O módulo textual disponibilizará:

- A transcrição diplomática das versões de 1572;
- A transcrição modernizada (seguindo os padrões ortográficos tanto do português europeu quanto do brasileiro);
- A alternância dinâmica entre versões (dextrógira, sinistrógira, modernizada).

### 4.3 Aparato crítico dinâmico

Em vez de um aparato crítico "tradicional", pretende-se usar os recursos digitais para exibir:

- Visualização paralela das duas edições de 1572, bem como da versão modernizada;
- Destaque automático de variantes;
- Filtro por tipo de divergência (ortográfica, lexical, tipográfica).

## 4.4 Glossário interativo

O formato digital permite que o glossário seja acessado por *hyperlinks* diretamente no texto.

Inicialmente, cada palavra do texto será associada ao verbete correspondente do dicionário de Peixoto e Pinto (Peixoto and Pinto, 1924), obra de referência para o estudo do texto camoniano. Futuramente, também se pretende elaborar um glossário próprio, contendo outras informações, tais como:

- definição contextual;
- etimologia;
- frequência na obra;
- ocorrências clicáveis.

Além disso, esse formato de glossário interativo permite que se acrescentem comentários de natureza histórica, literária e analítico-interpretativa, muito necessários para a plena compreensão do texto.

## 4.5 Traduções alinhadas

A obra de Camões foi traduzida inúmeras vezes. Pode-se mencionar, entre outras, as traduções para o espanhol (Garces, 1591), para o latim (Faria, 1622), para o inglês (Mickle, 1776) e para o italiano (Paggi, 1658). Assim, pretende-se disponibilizar as traduções com os seguintes recursos:

- Alinhamento por estrofe ou verso, quando possível;
- Visualização lado a lado com o texto português;
- Possibilidade de comparação lexical.

Abaixo pode-se ver a imagem de uma primeira versão da tela de visualização do texto.



Figura 1: Primeira versão da tela de visualização do texto.

## 5 Metodologia

A metodologia articula etapas fundamentais do trabalho da crítica textual ao processamento computacional, consolidada em quatro frentes principais.

### 5.1 Levantamento e curadoria

Nesta frente, procede-se à identificação das edições fac-similares disponíveis, bem como das traduções existentes em domínio público e das obras de comentaristas e críticos (também em domínio público) que podem trazer informações relevantes a serem incorporadas na edição digital.

### 5.2 Digitalização e OCR

Os materiais são transcritos automaticamente por meio do OCR (*Optical Character Recognition*) na ferramenta Transkribus. Após a transcrição, é feita uma revisão manual. A transcrição do Canto I já foi inteiramente finalizada.

Como "subproduto" deste projeto, será criado um modelo de OCR para o Transkribus, treinado especificamente para a tipografia portuguesa do século XVI.

### 5.3 Estruturação em TEI-XML

O texto transcrito é convertido para o formato estruturado TEI-XML, com a marcação de elementos XML tais como:

- versos (<l>);
- estrofes (<lg>);
- variantes (<app>, <rdg>);
- notas (<note>);
- nomes próprios (<name>).

Na figura 2 a seguir (captura de tela do *software* Oxygen XML), podem-se observar as diferenças entre as versões do texto, marcadas com o elemento <rdg>.

### 5.4 Processamento de Linguagem Natural

É necessário avaliar de que forma técnicas de PLN podem ser empregadas no texto camoniano. É possível pensar em três frentes principais:

#### a) Lematização e etiquetagem morfosintática

A lematização é necessária para que sejam feitos cálculos de frequência lexical, bem como para a associação de cada palavra ao verbete

```

<1>
<app>
<rdg wit="#V1">Por mares, nunca de antes na-uegados,</rdg>
<rdg wit="#V2">Por mares nunca de antes na-uegados,</rdg>
</app>
</1>
<1>Passaram, ainda alem da Taprobana,</1>
<1>Em perigos, &amp; guerras esforçados,</1>
<1>
<app>
<rdg wit="#V1">Mais do que prometia a força humana:</rdg>
<rdg wit="#V2">Mais do que prometia a força humana.</rdg>
</app>
</1>
<1>
<app>
<rdg wit="#V1">Entre gente remota edificáram;</rdg>
<rdg wit="#V2">E entre gente remota edificarão</rdg>
</app>
</1>
<1>
<app>
<rdg wit="#V1">Nouo Reino, que tanto sublimáram.</rdg>
<rdg wit="#V2">Nouo Reino, que tanto sublimarão.</rdg>
</app>
</1>
</1p>
<1g n="2">
<1>E tambem as memorias gloriosas</1>
<1>
<app>
<rdg wit="#V1">Daquelles Reis, que foram dilatando</rdg>
<rdg wit="#V2">Daquelles Reis, que forão dilatando</rdg>
</app>
</1>
<1>A Fee, o Imperio, &amp; as terras viciosas</1>

```

Figura 2: Trecho do texto com as marcações TEI-XML.

correspondente. Já a etiquetagem morfossintática possibilita buscas por construções sintáticas, o que é relevante para os estudos da linguagem literária.

Ferramentas de PLN usadas normalmente para lematização e etiquetagem (como o pacote spaCy) são treinadas com base em textos contemporâneos. Para o caso de textos do século XVI, é preciso adaptá-las para a multiplicidade de grafias da época (por exemplo, *Africa* e *Affrica*, *capitam* e *capitão*), ou fazer uma revisão "manual". Além disso, o texto poético, por ter sintaxe distinta do texto em prosa, traz mais dificuldade na etiquetagem morfossintática.

## b) Detecção automática de variantes

O mapeamento das diferenças entre as edições pode ser feito por meio de algoritmos de alinhamento textual. Uma possibilidade simples (mas limitada) é o alinhamento verso a verso. A comparação automática verso a verso das duas versões do Canto I (sinistrógira e dextrógira), efetuada por meio de um algoritmo simples de comparação escrito em Python (conforme a figura 3, a seguir) detectou 238 versos divergentes (de um total de 848 versos).

No entanto, como pode haver múltiplas divergências em cada verso, o alinhamento precisa ser feito palavra por palavra, considerando também os sinais de pontuação. Uma complicação adicional surge nos casos em que

```

1 Estrofe e Verso;Versão 1;Versão 2
2 Estrofe 1, verso 2;Que da occidental praya lusitana,;Que da occidental praya lusi-tana,
3 Estrofe 1, verso 3;Por mares, nunca de antes na-uegados,;Por mares nunca de antes na-uegados,
4 Estrofe 1, verso 6;Mais do que prometia a força humana;Mais do que prometia a força humana.
5 Estrofe 1, verso 7;Entre gente remota edificáram;E entre gente remota edificarão
6 Estrofe 1, verso 8;Nouo Reino, que tanto sublimáram,;Nouo Reino, que tanto sublimarão.
7 Estrofe 2, verso 2;Daquelles Reis, que foram dilatando;Daquelles Reis, que forão dilatando
8 Estrofe 2, verso 4;De Africa, & de Asia, andaram deoustado;De Affrica, & de Asia, andarão devastando,
9 Estrofe 3, verso 2;As nauegações grandes que fizeram;As nauegações grandes que fizero:
10 Estrofe 3, verso 3;Callese de Alexandro, & de Trajano;Callese de Alexandro, & de Trajano,
11 Estrofe 3, verso 4;A fama das victorias que tiueram,;A fama das victorias que tiuerão,
12 Estrofe 3, verso 6;A quem Neptuno, & Marte obedecéram;A quem Neptuno, & Marte obedecerão:
13 Estrofe 4, verso 3;Se sempre em verso humilde celebrado,;se sempre em verso humilde, celebrado
14 Estrofe 4, verso 6;Hum estillo grandiloco, & corrente;Hum estillo grandiloco, & corrente,
15 Estrofe 4, verso 8;Que nam tenham enueja ás de Hypocrene,;Que não tenham enueja aas de Hypocrene.
16 Estrofe 5, verso 2;E nam de agreste a vena, ou frauta ruda;E não de agreste a vena, ou frauta ruda:
17 Estrofe 6, verso 5;Vos ò nouo temor da Maura lança,;Vos o nouo temor da Maura lança,
18 Estrofe 6, verso 7;Dada ao mundo por Deos que todo o mande,;Dada ao mundo por Deos q todo o mande,
19 Estrofe 7, verso 4;Cesaria, ou Christianissima chamada;Cesarea, ou Christianissima chamada:
20 Estrofe 8, verso 3;Veo tambem no meyo do Hemispherio,;Veo tambem no meyo do Hemispherio,
21 Estrofe 9, verso 2;Que nesse tenro gesto vos contemplo,;Que nesse tenro gesto vos contemplo,
22 Estrofe 10, verso 1;Vereis amor da patria, nam mouido;Vereis amor da patria, não mouido
23 Estrofe 10, verso 2;De premio vil: mas alto, & quasi eterno,;De premio vil: mas alto, & quasi eterno
24 Estrofe 10, verso 3;Que nam he premio vil ser conhecido,;Que nam he premio vil, ser conhecido
25 Estrofe 10, verso 4;Por hum pregão do ninho meu paterno,;Por hum pregão do ninho meu paterno.
26 Estrofe 10, verso 6;Daquelles de quem sois senhor superno,;Daquelles de quem sois senhor superno.

```

Figura 3: Listagem de diferenças entre as versões, gerada automaticamente.

palavras inteiras encontram-se faltantes numa das versões, como o artigo "o" no exemplo a seguir:

*Eis nos bateis fogo se leuanta,* (versão dextrógira)

*Eis nos bateis o fogo se leuanta,* (versão sinistrógira)

(Canto I, estrofe 89, verso 1)

Assim, faz-se necessário o emprego de algoritmos mais complexos, que estão sendo avaliados.

## c) Alinhamento com traduções

O alinhamento do texto português com as traduções exige também o emprego de técnicas avançadas, visto que os textos traduzidos não correspondem ao texto original verso a verso. Assim, é possível que cada tradução precise ser alinhada de forma específica.

## 6 Desafios computacionais e inovações tecnológicas

A aplicação de PLN a textos da Primeira Modernidade apresenta alguns desafios, em especial a ausência de padronização ortográfica da época; tendo em vista que as ferramentas disponíveis são treinadas e programadas para o português atual (séculos XX e XXI), faz-se necessário recorrer a diversas adaptações.

Dessa forma, espera-se que o projeto contribua também para a criação de recursos computacionais que possam ser aplicados em outros textos históricos, tais como modelos de lematização e dicionários de formas ortográficas.

O projeto inova no sentido de que propõe:

1. Integração orgânica entre edição crítica e PLN;

2. Representação estruturada da tradição textual;
3. Interface interativa voltada tanto para especialistas quanto para estudantes;
4. Produção de dados reutilizáveis em pesquisa linguística.

Diferentemente de repositórios como o Project Gutenberg, que oferecem texto linearizado, o projeto propõe uma arquitetura baseada em dados estruturados e interoperáveis.

## 7 Resultados esperados e impactos

O principal resultado esperado com o projeto é a própria edição crítica digital do Canto I, com os vários recursos de consulta pretendidos. Parte integrante dessa edição crítica é o próprio texto anotado em TEI-XML, que poderá ser reutilizado em diversos outros projetos. Além disso, os algoritmos desenvolvidos poderão ser empregados em projetos similares, bem como o modelo de transcrição gerado pelo Transkribus poderá ser reaproveitado para outros textos da época. Por fim, também serão publicados artigos científicos sobre a metodologia empregada.

A médio prazo, o modelo será expandido aos demais cantos e a outras obras do período.

O projeto visa trazer impactos para:

### 7.1 Pesquisa filológica

Apesar de o texto camoniano já ter sido largamente estudado, a abordagem computacional pode trazer um olhar novo à obra (ainda que não necessariamente superior). Além disso, os métodos desenvolvidos podem ser aplicados a outras obras no futuro.

### 7.2 Linguística histórica

O texto d’*Os Lusíadas* anotado e estruturado computacionalmente servirá para apoiar estudos diacrônicos da língua portuguesa.

### 7.3 Ensino

Por ser *Os Lusíadas* um dos textos literários mais importantes da língua portuguesa, é uma obra amplamente estudada tanto na educação básica quanto no ensino superior. Assim, o projeto tem potencial para se tornar um recurso pedagógico interativo de grande relevância.

## 7.4 Desenvolvimento tecnológico

O principal impacto pretendido quanto ao desenvolvimento tecnológico é a reutilização dos algoritmos utilizados para outros projetos futuros, estabelecendo, assim, uma proposta de modelo de edição crítica digital.

## 8 Conclusão

O projeto Lusíadas Digital representa uma convergência entre tradição filológica e inovação tecnológica. Ao integrar modelagem textual, crítica editorial e processamento de linguagem natural, propõe-se um novo paradigma para a edição digital de textos clássicos em língua portuguesa. A edição do Canto I funcionará como projeto-piloto para uma iniciativa de maior alcance, capaz de posicionar a pesquisa brasileira em Humanidades Digitais em diálogo direto com experiências internacionais consolidadas.

## References

- Luís Vaz de Camões. 1916. *Os Lusíadas*. Companhia Portuguesa Editora, Porto. Comentados por Augusto Epiphany da Silva Dias. 2. ed. melhorada. Tomo I.
- Gregory Crane, David Bamman, and Alison Jones. 2008. [epihology: When the books talk to their readers](#). In Ray Siemens and Susan Schreibman, editors, *A Companion to Digital Literary Studies*. Blackwell, Oxford.
- Tomás de Faria. 1622. *Lusiadum Libri Decem*. Ex Officina Gerardi de Vinca, Lisboa.
- Henrique Garces. 1591. *Los Lusíadas de Luys de Camoes*. Impreso con licencia en casa de Guillermo Dreuy, Madri. Translator.
- William Julius Mickle. 1776. *The Lusiad; or, the Discovery of India. An Epic Poem*. Printed by Jackson and Lister, Oxford. Translator.
- Vanessa Martins do Monte and Maria Clara Paixão de Sousa. 2017. [Por uma filologia virtual: o caso das atas da câmara de são paulo \(1562–1596\)](#). *Revista da Abralín*, 16:239–264.
- Carlo Antonio Paggi. 1658. *Lusíada Italiana. Poema Heroico del grande Luigi de Camões*. por Henrique Valente de Oliveira, Lisboa.
- Afrânio Peixoto and Pedro A. Pinto. 1924. *Dicionário d’Os Lusíadas de Luis de Camões*. Francisco Alves, Rio de Janeiro.
- TEI Consortium. 2025. [TEI: Guidelines for Electronic Text Encoding and Interchange](#). Accessed: 18 Feb. 2026.