



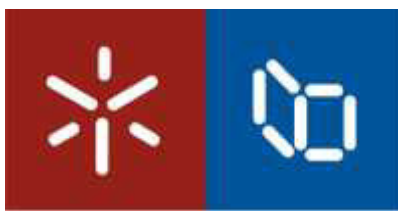
Universidade do Minho

Instituto de Letras e Ciências Humanas

Cátia Sofia Gomes da Silva

**TRANSCRIÇÃO DA CORRESPONDÊNCIA DIRIGIDA
A CAMILO CASTELO BRANCO: REFLEXÕES SOBRE
PRINCÍPIOS E NORMAS DE CODIFICAÇÃO ELETRÓNICA.**

Outubro 2015



Universidade do Minho

Instituto de Letras e Ciências Humanas

Cátia Sofia Gomes da Silva

**TRANSCRIÇÃO DA CORRESPONDÊNCIA DIRIGIDA
A CAMILO CASTELO BRANCO: REFLEXÕES SOBRE
PRINCÍPIOS E NORMAS DE CODIFICAÇÃO ELETRÓNICA.**

Relatório de Estágio
Mestrado em Mediação Cultural e Literária

Trabalho efetuado sob a orientação da:
Doutora Idalete Maria da Silva Dias

Outubro 2015

DECLARAÇÃO

Nome: Cátia Sofia Gomes da Silva

Endereço eletrónico: catiasgs20@hotmail.com

Número de Cartão de Cidadão: 14091335

Título do Relatório:

Transcrição da correspondência dirigida a Camilo Castelo Branco: reflexões sobre princípios e normas de codificação eletrónica.

Orientadora:

Doutora Idalete Maria da Silva Dias

Instituto de Letras e Ciências Humanas da Universidade do Minho

Ano de Conclusão: 2015

Designação do Mestrado:

Mestrado em Mediação Cultural e Literária

É AUTORIZADA A REPRODUÇÃO PARCIAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE;

Universidade do Minho, ____/____/____

Assinatura: _____

AGRADECIMENTOS

Aos meus pais, por todo o apoio e dedicação prestados desde sempre, e especialmente ao longo deste ano.

Aos meus avós por terem sido também meus pais, e pela inspiração que são todos os dias.

À minha orientadora de Mestrado, Professora Doutora Idalete Dias por todo o apoio e ajuda prestados, pela confiança e encorajamento, e pela disponibilidade e boa disposição, ao longo do decurso do projeto.

Ao Doutor José Manuel e à Doutora Carla Costa, pela colaboração e ajuda prestadas, e a toda a equipa do Centro de Estudos Camilianos, pela disponibilidade demonstrada desde o primeiro dia.

Um agradecimento especial à minha colega Cristiana Leal, pelo companheirismo e ajuda ao longo deste ano de trabalho.

TRANSCRIÇÃO DA CORRESPONDÊNCIA DIRIGIDA A CAMILO CASTELO BRANCO: REFLEXÕES SOBRE PRINCÍPIOS E NORMAS DE CODIFICAÇÃO ELETRÓNICA

RESUMO

Este projeto tem como principais objetivos a transcrição digital e a revisão de documentos epistolográficos pertencentes ao Arquivo da Correspondência dirigida a Camilo Castelo Branco, e a anotação estrutural dos mesmos documentos para integração posterior no Arquivo Digital da Correspondência para o escritor português do século XIX, a disponibilizar *online* para fins investigativos.

Os processos de transcrição e anotação digital das cartas basear-se-ão na metalinguagem de anotação eletrónica *Extensible Markup Language (XML)*, seguindo as diretivas de anotação da *Text Encoding Initiative (TEI)* e as propostas do projeto *Digital Archive of Letters in Flanders (DALF)*, referência incontornável no tratamento digital de documentos epistolográficos.

Palavras-chave:

Camilo Castelo Branco; edição eletrónica; correspondência; Text Encoding Initiative (TEI); Extensible Markup Language (XML)

TRANSCRIPTION OF LETTERS ADDRESSED TO CAMILO CASTELO BRANCO: A REFLECTION ON THE PRINCIPLES AND GUIDELINES OF ELECTRONIC TEXTUAL EDITING

ABSTRACT

The present project has as its main goals the transcription of correspondence material belonging to the Archive of the Letters addressed to Camilo Castelo Branco into electronic form and the respective annotation of structural and text-specific features. These electronic versions are to be included in the future Digital Archive of the Letters of the 19th century Portuguese writer. This archive will be made available online for research purposes.

The transcription and annotation processes will be based on the meta-markup language XML and the annotation guidelines provided by the *Text Encoding Initiative* (TEI) and the project *Digital Archive of Letters in Flanders*, considered an important source when it comes to describing and encoding correspondence material.

Keywords:

Camilo Castelo Branco; electronic edition; correspondence; Text Encoding Initiative (TEI); Extensible Markup Language (XML)

ÍNDICE GERAL

DECLARAÇÃO	ii	
AGRADECIMENTOS	iii	
RESUMO	v	
ABSTRACT	vii	
ÍNDICE GERAL	ix	
ÍNDICE DE FIGURAS	xi	
CAPÍTULO 1: INTRODUÇÃO AO PROJETO, OBJETIVOS E		
METODOLOGIA	13	
1.1. Introdução ao projeto	15	
1.2. Contextualização do estágio	16	
1.3. Objetivos do projeto	19	
1.4. Metodologia	21	
CAPÍTULO 2: TRANSCRIÇÃO E ANOTAÇÃO DE TEXTOS		
EPISTOLOGRÁFICOS	23	
2.1. Introdução	25	
2.2. Transcrição e anotação da correspondência de Camilo Castelo Branco	27	
2.2.1. <i>Text Encoding Initiative</i>	28	
CAPÍTULO 3: PRINCÍPIOS E NORMAS DE CODIFICAÇÃO ELETRÓNICA		31
3.1. Introdução	33	
3.2. Anotação da correspondência de Camilo Castelo Branco	34	
CAPÍTULO 4: METADADOS E TEIHEADER		51
4.1. Metadados	53	
4.2. O cabeçalho da TEI - TeiHeader	55	
CAPÍTULO 5: CONCEITO E PROCESSO DE VALIDAÇÃO DE DTD		71
5.1. Conceito de validação	73	
5.2. Processo de validação	73	
CAPÍTULO 6: CONSIDERAÇÕES FINAIS		77
REFERÊNCIAS BIBLIOGRÁFICAS E CIBERGRÁFICAS	81	

ANEXOS	89
Anexo 1 - Cabeçalho <teiHeader> bem formado da edição eletrónica da Carta 1 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	91
Anexo 2 - Estrutura da secção <text> da Carta 80 do signatário José de Azevedo e Menezes a Camilo Castelo Branco	97
Anexo 3 - Estrutura do documento DTD	103
Anexo 4 - Certificado da comunicação intitulada: "Arquivo Digital da Correspondência dirigida a Camilo Castelo Branco: problemas, soluções e desafios"	107

ÍNDICE DE FIGURAS

Figura 1 - Esquema das fases do projeto	21
Figura 2 - Esquema da estrutura de um documento TEI	33
Figura 3 - Excerto da Carta 57 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	34
Figura 4 - Excerto da Carta 89 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	35
Figura 5 - Excerto da Carta 16 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	36
Figura 6 - Excerto da Carta 57 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	37
Figura 7 - Excerto da Carta 57 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	39
Figura 8 - Excerto da Carta 25 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	40
Figura 9 - Excerto da Carta 16 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	41
Figura 10 - Excerto da Carta 57 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	42
Figura 11 - Excerto da Carta 69 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	43
Figura 12 - Excerto da Carta 46 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	44
Figura 13 - Excerto da Carta 46 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	44
Figura 14 - Excerto da Carta 46 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	45
Figura 15 - Excerto da Carta 16 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	46
Figura 16 - Excerto da Carta 46 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	47
Figura 17 - Excerto da Carta 44 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	48

Figura 18 - Excerto da Carta 69 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	48
Figura 19 - Excerto da Carta 44 do <i>Acervo da Correspondência de Camilo Castelo Branco</i>	49
Figura 20 - Exemplo de documento DTD mal formado	75
Figura 21 - Exemplo de documento DTD válido	76

CAPÍTULO 1:
INTRODUÇÃO AO PROJETO, OBJETIVOS E METODOLOGIA

1.1. INTRODUÇÃO AO PROJETO

O avanço tecnológico tem contribuído imenso para o desenvolvimento das humanidades digitais em diversos aspetos, principalmente no que diz respeito à preservação de todo o tipo de informação existente em formato impresso e em formato nato-digital¹. A criação de edições/arquivos digitais veio tornar todo o processo de investigação e consulta dos mesmos muito mais acessível. Assim sendo, a preservação digital através da criação de edições/arquivos digitais desempenha uma função importantíssima em todas as áreas de investigação e ensino, permitindo o fácil e rápido acesso à documentação pretendida, sem que seja necessária a consulta física da mesma. Como afirma Vanhoutte (2009, p. 78), "the goal of an electronic edition is to provide a version of the text which is encoded so as to permit electronic inspection, computer-assisted analysis, and retrieval, to which the raw image is inherently resistant".

O principal objetivo da criação de uma edição eletrónica consiste em providenciar uma versão de um documento/texto que permita a localização, a extração e a manipulação da informação constante no documento. Uma edição eletrónica pressupõe a codificação do documento/texto original, de tal forma que o mesmo possa ser pesquisado, processo que não é possível realizar com base no documento em formato de imagem. A transcrição do documento em formato eletrónico constitui um primeiro nível de uma edição eletrónica, que é criada com o intuito de corresponder aos objetivos de quem a cria. Assim, as edições são dotadas de informações adequadas e capazes de satisfazer os interesses de determinada audiência à qual o criador das mesmas pretende chegar. As edições eletrónicas são articuladas e extensíveis, pois, para além da informação acerca do documento original, podem conter informações relativas à sua própria criação. Podem, ainda, ser dotadas de mais informação ao longo dos tempos.

As informações presentes nas edições eletrónicas permitem uma análise do documento original por parte do utilizador das edições, através de computador, não necessitando da consulta do documento físico. Esta ação vai garantir a preservação do

¹ Erway define documentos "nato-digitais" da seguinte forma: "Born-digital resources are items created and managed in digital form" (2010: 1). Erway distingue os seguintes tipos de documentos "nato-digitais": fotografias digitais; documentos digitais (documentos PDF, etc.); conteúdo oriundo da *Web*; manuscritos digitais; registos eletrónicos (arquivos institucionais e empresariais); dados de pesquisa; dados dinâmicos; arte digital; publicações de média digital (publicações comerciais, tais como jogos de vídeo, filmes em DVD, músicas em CD). (2010: 1-3)

estado físico do documento original, uma vez que disponibiliza a informação do mesmo num formato digital. Uma edição eletrónica de um documento permite fornecer o documento original através do fac-símile e do texto transcrito e anotado eletronicamente, contribuindo para a preservação e recuperação do documento manuscrito. Uma edição eletrónica não deve apresentar barreiras informáticas, uma vez que deve ser processada com recurso a programas e aplicações interoperáveis e independentes, baseadas na dinâmica "código aberto". Assim, é possível afirmar que uma edição eletrónica deve conjugar três aspetos fundamentais: longevidade; a integridade intelectual; e a fácil acessibilidade sem a existência de aspetos técnicos que dificultem ou impossibilitem o seu uso.

Um exemplo de edição eletrónica similar à que se pretende criar neste projeto é o *The Van Gogh Letters Project*², que se encontra disponível *online* e contém a edição eletrónica da correspondência de e para o famoso pintor Vincent van Gogh.

1.2. CONTEXTUALIZAÇÃO DO ESTÁGIO

Este projeto, realizado no âmbito do Mestrado em Mediação Cultural e Literária da Universidade do Minho, teve como local de estágio o Centro de Estudos Camilianos, situado em S. Miguel de Seide, Famalicão, local onde se encontra o *Acervo da Correspondência de Camilo Castelo Branco*. O Centro de Estudos dispõe ainda de uma biblioteca que é

o espaço privilegiado de pesquisa e consulta dos acervos bibliográficos, documentais e iconográficos da instituição merecendo destaque do conjunto, o núcleo de cartas de e para Camilo, a sua biblioteca particular, as camilianas de Arzila da Fonseca, de Assis Chateaubriand, de Pinheiro Torres, de Alberto Pimentel, de Manuel Simões, de Alexandre Cabral, de Viale Moutinho e de Aníbal Pinto de Castro, mais o fundo que a Casa de Camilo adquiriu. Estes fundos produzem pela sua extensão e variedade, um inestimável repositório para o estudo e compreensão de Camilo e do seu tempo. (*Site oficial da Casa de Camilo*³).

² O *The Van Gogh Letters Project* envolve a criação de um arquivo digital da correspondência de Vincent van Gogh que contém as transcrições das cartas em língua original e respetiva tradução em língua inglesa, os fac-símiles das cartas e metadados contextuais. Pode ser consultado em: <http://www.vangoghletters.org>.

³ Cf. <http://www.camilocastelobranco.org/>

Para além destes aspetos, importa referir o apoio precioso prestado pelo Diretor do Centro de Estudos Camilianos, Doutor José Manuel Oliveira, e pela Doutora Carla Costa e restante equipa na concretização do plano de trabalhos do projeto: no acesso e na consulta das espécies do *Acervo da Correspondência de Camilo Castelo Branco*; na revisão das cartas transcritas para formato eletrónico; no fornecimento de informações relevantes sobre Camilo e seus correspondentes.

A correspondência tratada no âmbito deste projeto foi reunida pela *Comissão de Homenagem a Camillo Castello Branco*, constituída em abril de 1915 por José de Azevedo e Menezes, Nuno Simões e Francisco Correia de Mesquita Guimarães, com a incumbência de, por um lado, adquirir o espólio do escritor, incluindo as ruínas da Casa de Seide e o quintal, os livros, autógrafos, o mobiliário e os objetos de Camilo, que se encontrava sob a guarda dos netos pela quantia de 2000 escudos, e, por outro, reedificar a Casa de Seide e aí instalar o Museu de Camilo, hoje Casa Museu de Camilo. A obra *"Camillo Homenageado: O Escriptor da Graça e da Belleza"*, publicada pela Comissão de Homenagem (Menezes, José de Azevedo *et al.*, 1920), trata com minúcia e rigor contabilístico as diligências efetuadas, os apoios e donativos recebidos, registando os nomes dos doadores e os procedimentos levados a cabo nas várias etapas da missão acima descrita. Para além de relatórios, esta publicação inclui também várias fotografias e plantas da Casa de Camilo, antes e depois do restauro. De valor inestimável para o presente projeto é a secção desta obra dedicada aos documentos epistolográficos que a Comissão de Homenagem conseguiu reunir entre 1915 e 1920. Nesta secção encontram-se os resumos das 950 cartas reunidas, quase todas, exceto 15 espécies, dirigidas a Camilo Castelo Branco, no período compreendido entre 1864 e 1890⁴, num total de 278 correspondentes, incluindo muitas personalidades eminentes da cultura, da literatura e da política portuguesa da época. A título de exemplo, refira-se: Rei D. Luís; Bulhão Pato; D. Pedro II (do Brasil); Eça de Queirós; Eugénio de Castro; Fialho de Almeida; Guerra Junqueiro; Martins Sarmiento; Visconde de Ouguela (133 espécies); Fanny Owen; Dr. Ricardo Jorge; Tomás Ribeiro. Na obra *Camillo Homenageado*, os resumos das cartas encontram-se organizados por ordem alfabética do nome dos signatários. No presente projeto, foi trabalhado o conjunto de autores correspondente à letra A do alfabeto, um

⁴ 187 cartas do espólio não se encontram datadas.

total de 107 cartas, 32 signatários. Segue-se uma listagem dos 32 signatários cujas cartas foram transcritas no âmbito deste projeto: Camilo Castelo Branco, Abel Acácio Botelho, Augusto Cesário de Abreu, Eduardo de Abreu, Gaspar Ribeiro Gomes de Abreu, João Gomes de Abreu, Filippe d'Andrade Albuquerque, Conde de Aljezur, Eduardo Augusto Allen, António Maria de Almeida, António Nicolau de Almeida, Carlos de Almeida Braga, J. de Almeida, M. Duarte de Almeida, Joaquim Alves Matheus, António Alves Mendes da Silva Ribeiro, António Cândido Ribeiro da Costa, A. Barros, Alberto Braga, Alberto Pimentel, Alexandre Braga, Gaspar Nunez de Arce, José Maria Asenzio, Manuel de Assumpção, Luiz de Mello Athayde, José Joaquim Pereira de Azambuja Rodrigues, Alexandre Thomaz de Azevedo, J. A. de Azevedo Castro, Egydio de Azevedo, João Lúcio de Azevedo, José de Azevedo e Menezes e António José de Azevedo.

A correspondência de Camilo Castelo Branco possui informações inéditas acerca do romancista português, tanto a nível pessoal como social e profissional. O tratamento eletrónico e a disponibilização da correspondência de Camilo Castelo Branco constituirá uma peça fundamental para o conhecimento do papel e da importância de Camilo na sociedade portuguesa do século XIX. As cartas dão a conhecer, em primeira mão, as relações que Camilo mantinha com familiares, editores, amigos, políticos e notáveis da época, revelando pormenores da vida socio-cultural da época. Tomemos como exemplo as primeiras 15 cartas transcritas, nas quais Camilo é o remetente e o destinatário é José de Azevedo e Menezes⁵. Camilo troca informações com o destinatário, principalmente sobre genealogia, pretendendo saber mais sobre certas famílias e pedindo informações sobre este domínio. A forma como José de Azevedo e Menezes trata Camilo denota uma elevada admiração pelo escritor, tratando-o por "mestre": "Peço-lhe que me obrigue constantemente a manusear livros e cartapacios, para ir aprendendo alguma coisita com o meu querido e respeitavel mestre." (Carta 79, *Acervo da Correspondência de Camilo Castelo Branco*). A correspondência revela que Camilo gozava de um estatuto social ímpar na sociedade portuguesa. O pedido de favores a Camilo é um assunto recorrente ao longo da maioria das cartas tratadas, o que revela que o escritor era uma pessoa muito

⁵ José de Azevedo e Menezes Cardoso Barreto foi um erudito que colaborou em variadíssimos jornais: *Novidades*, *O Primeiro de Janeiro*, *Nova Alvorada*, *Correio do Minho*, *Progresso Católico* e *A Palavra*, de que foi um dos fundadores. Exerceu cargos importantes na sua terra natal, entre eles: Presidente da Câmara Municipal de Famalicão (1896-1898), e Presidente da Comissão de Homenagem a Camilo Castelo Branco. (Cabral, 1988).

influyente. Nas cartas do signatário Padre António José de Azevedo⁶, este pede a Camilo o favor de colocar o sobrinho num emprego público:

"(...) o vosso Am.º [Amigo] o Bispo de Viseu se acha colocado no governo; (...) he (segundo me parece) chegada a ocasião de levarmos a efeito a pertença de nosso sobr.º [sobrinho] Resende, cuja pertença he (como estareis lembrado) hũ logar, q [que] se acha vago no governo civil (...)" (Carta 98, *Acervo da Correspondência de Camilo Castelo Branco*).

A importância de Camilo a nível político e social também ultrapassava fronteiras. Vejamos a carta do poeta e político espanhol Gaspar Nuñez de Arce⁷ a Camilo, solicitando colaboração num periódico com o intuito de ajudar as vítimas dos terremotos de Andaluzia: "(...) nos atrevemos á suplicarle colabore en dicho periódico enviándonos una poesia, emcuento ó un articulo de reducidas dimensiones, un pensamiento ó una frase que por venir de V. será mas que limosna valioso donativo." (Carta 62, *Acervo da Correspondência de Camilo Castelo Branco*).

1.3. OBJETIVOS DO PROJETO

Este projeto de estágio enquadra-se num projeto mais amplo que visa a criação do Arquivo Digital da Correspondência dirigida ao romancista português Camilo Castelo Branco. O projeto geral tem como principais objetivos:

- (i) A preservação e o tratamento eletrónico das 950 cartas do espólio da correspondência reunida pela Comissão de Homenagem a Camilo;
- (ii) A disponibilização das transcrições dos documentos, segundo a ortografia da época dos documentos originais;
- (iii) A disponibilização dos respetivos fac-símiles;

⁶ O Padre António José de Azevedo foi mestre de Camilo Castelo Branco e viveu com o escritor. Camilo dedicou-lhe o romance *O Bem e o Mal* (1863). (Cabral, 1988).

⁷ Gaspar Nuñez de Arce foi um poeta e político espanhol, ocupando o cargo de Governador de Barcelona e o de Ministro do Ultramar. (*Biografías y Vidas. La Enciclopedia Biográfica en Línea*).

- (iv) A anotação estrutural e semântica da correspondência, trabalho que permitirá cruzar e desvendar elementos e relações desconhecidas através da representação visual da correspondência do acervo;
- (v) A criação do arquivo digital que permitirá a pesquisa das cartas, fac-símiles e dados contextuais.

O presente projeto, intitulado "*Transcrição da correspondência dirigida a Camilo Castelo Branco: reflexões sobre princípios e normas de codificação eletrónica*", tem como objetivos específicos:

- (i) A transcrição das 107 cartas que compõem a letra A da secção do *Camillo Homenageado* dedicada aos documentos epistolográficos;
- (ii) A anotação da estrutura e das propriedades textuais das cartas com base na metalinguagem de anotação eletrónica *eXtensible Markup Language* (XML), as diretrizes da norma de codificação textual *Text Encoding Initiative* (TEI) e as propostas do projeto *Digital Archive of Letters in Flanders* (DALF);
- (iii) A criação de um esquema de metadados que servirá de base para a elaboração das fichas bibliográficas das 107 cartas, designadas de cabeçalho TEI (*teiHeader*);
- (iv) A criação de um esquema de validação *Document Type Definition* (DTD) que permita a validação de todos os documentos XML.

1.4. METODOLOGIA

A elaboração deste projeto foi organizada nas seguintes quatro fases:

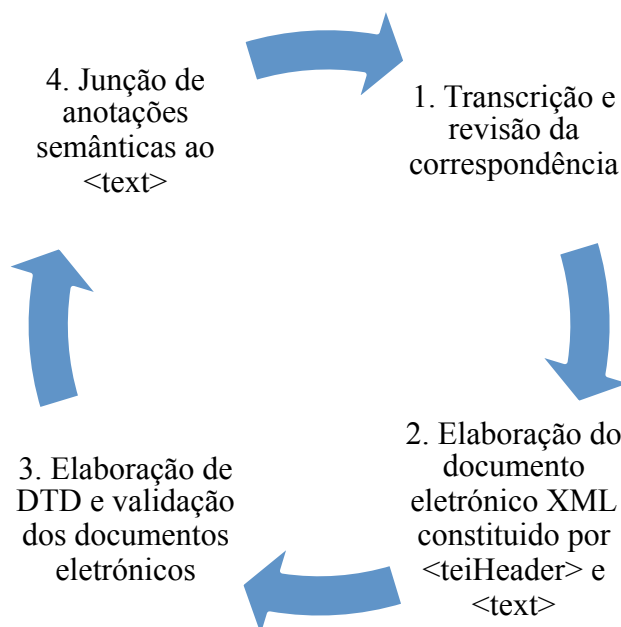


Figura 1 - Esquema das fases do projeto

Numa primeira fase, as 107 cartas foram transcritas respeitando as propriedades estruturais e textuais dos documentos originais. Para as versões transcritas tomou-se como base a metalinguagem de anotação XML. Nesta fase, as transcrições foram submetidas a um processo rigoroso de revisão.

Numa segunda fase, procedeu-se à criação de um documento eletrónico em formato XML para cada uma das cartas, documento esse constituído por duas secções distintas: (i) a primeira corresponde ao cabeçalho TEI (`<teiHeader>`), que documenta o processo de criação digital a partir do objeto original; (ii) a segunda parte corresponde ao conteúdo da carta em si (`<text>`). Em ambas as secções foram utilizadas as normas propostas pela TEI e pelo projeto DALF.

A terceira fase deste projeto consistiu na criação de um documento DTD e na respetiva validação dos 107 documentos XML criados.

A quarta e última fase consistiu na inserção de anotações semânticas no conteúdo das primeiras 25 cartas. Esta tarefa constituiu uma colaboração com o projeto intitulado *"Edição digital anotada da correspondência dirigida a Camilo Castelo Branco: problemas e desafios"*, elaborado no âmbito do Mestrado em Mediação Cultural e Literária da Universidade do Minho, da responsabilidade de Cristiana Leal.

CAPÍTULO 2:
TRANSCRIÇÃO E ANOTAÇÃO DE TEXTOS EPISTOLOGRÁFICOS

2.1. INTRODUÇÃO

Uma das principais características de uma edição eletrónica é o facto de esta poder conter os fac-símiles dos documentos originais que se pretendem representar, juntamente com as transcrições dos mesmos. A edição eletrónica, quando anotada conforme normas internacionais, que garantem o intercâmbio e a interoperabilidade dos dados, e complementada com ferramentas e aplicações tecnológicas, que permitem a análise e extração de informação, constitui um recurso valioso para investigadores. Uma edição eletrónica fornece metainformações sobre propriedades físicas do documento original (estado físico do documento, largura, altura, parafernália, etc.), dados relevantes para determinados estudos.

Uma versão transcrita de um documento que segue as normas da linguagem XML é composta por dois níveis de informação: o nível do texto e o nível das anotações. No que concerne ao nível das anotações, um projeto de edição que tenha como objetivo a transcrição fiel ao documento original deverá etiquetar descritivamente a estrutura lógica do mesmo: parágrafos, quebras de linha e página. Segundo Grésillon (1994, p. 129), toda a transcrição é simultaneamente mais rica e mais pobre do que o manuscrito original:

Plus riche, car elle [transcription] reflète un certain travail d'analyse : pour reproduire il ne suffit pas de copier, il faut d'abord *comprendre* les signes graphiques et les traduire en opérations d'écriture. Plus pauvre, car toute transcription dactylographiée perd irrémédiablement ce que l'écriture manuscrite véhicule comme charge affective (hâte, blocage, angoisse, jubilation).

Como refere Grésillon, por um lado, o ato de transcrição é obrigatoriamente um ato "inglório", uma vez que não é de todo possível passar para o meio eletrónico a carga emocional inerente ao processo de escrita em papel. Por outro, o próprio ato de transcrição implica uma análise documental, que permite ao transcritor conhecer de forma aprofundada o conteúdo do documento. Esta análise constituirá a base para a anotação descritiva da estrutura lógica e propriedades específicas do conteúdo do texto. Desta perspetiva, a versão transcrita e anotada pode ser considerada mais rica que o documento original.

Um projeto de transcrição envolve necessariamente a abordagem das seguintes questões:

- (i) O objetivo da edição eletrónica;
- (ii) O público ao qual a edição eletrónica se destina;
- (iii) As possíveis formas de interação do público com a edição eletrónica;
- (iv) As normas a utilizar nos processos de transcrição e anotação estrutural e textual dos documentos eletrónicos.

O processo de transcrição pressupõe delinear a finalidade para a qual a edição eletrónica é destinada. Neste sentido, o transcritor interpretará o documento à luz deste objetivo, transcrevendo e anotando o conteúdo de forma a representar as propriedades de interesse para o próprio transcritor e/ou anotador e público específico e geral. Uma vez que é difícil prever todas as formas de interação do público, geral e específico, com a edição eletrónica, é importante pensar no conceito de extensibilidade aquando do planeamento da edição eletrónica. Segundo Pichler (1995, p. 691), "the sign of a good transcription system is that it has extensibility. This means that it should be possible to revise and adjust the system to serve new interests as and when they arise". Neste sentido, importa salientar a relevância que o fator extensibilidade, isto é, a potencialidade da transcrição servir novos interesses de investigação, desempenha no ato de transcrição. Portanto, cabe à equipa responsável pela transcrição decidir qual o nível de detalhe da mesma. Refira-se, a título de exemplo, as decisões que terão de ser tomadas relativamente a aspetos como: palavras rasuradas, sublinhadas, acrescentadas, expansão de abreviaturas, normalização de ortografia, quebras de linha e página. A etiquetagem destes aspetos pode servir determinados fins investigativos, por exemplo no âmbito de estudos de crítica genética. Embora a crítica genética remeta para segundo plano a análise de textos epistolográficos, importa referir que este género textual constitui uma excelente fonte de informação nesta área, uma vez que é nestes textos que os escritores estabelecem um intercâmbio de pensamentos e ideias com os seus correspondentes. A correspondência é, assim, um meio privilegiado para a comunicação pessoal de sentimentos e estados de espírito.

As edições eletrônicas devem ser preservadas de modo a terem utilidade de longa duração, ou seja, não se devem tornar obsoletas, estando assim disponíveis para que se acrescente ao documento electrónico sempre algo mais, quando se considerar relevante. Estas edições devem ser concebidas com a finalidade de chegar a um público abrangente, seja ele composto por investigadores, estudantes, ou apenas utilizadores comuns.

2.2. TRANSCRIÇÃO E ANOTAÇÃO DA CORRESPONDÊNCIA DE CAMILO CASTELO BRANCO

Um projeto de edição eletrónica que tenha como objetivo o tratamento e processamento de textos epistolográficos, deverá, em primeiro lugar, definir o que é uma carta.

A letter is not a type of text as are prose, poetry, and drama, but the concept of a letter hints at a form which can contain all types of texts, images, or objects. The first step in building a textbase of correspondence material, therefore, is to define precisely what the project considers a letter. (Vanhoutte e Van den Branden, 2009, p. 83)

Tal como referem Vanhoutte e Van den Branden, documentos epistolográficos são complexos, dado que podem conter no seu conteúdo elementos de outros géneros textuais, tais como excertos de prosa, poesia, desenhos, fórmulas, pautas musicais, árvores genealógicas. Esta propriedade do material de correspondência constitui um desafio para os processos de transcrição e anotação. Nem sempre é fácil encontrar soluções para representar e etiquetar estes fenómenos. Veja-se, por exemplo, a Carta 11 do *Acervo da Correspondência de Camilo Castelo Branco*, de Camilo para José de Azevedo e Menezes, na qual aparece uma árvore genealógica. Neste caso, não encontramos propostas de transcrição e anotação na Iniciativa de Codificação Textual TEI e em outros projetos de edição de correspondência. Para anotar este caso específico, foi criado o elemento <genealogia>, de modo a identificar que o texto delimitado por este elemento contém informação genealógica.

No âmbito do presente projeto, adotando a definição de "carta" proposta pelo projeto *Digital Archive of Letters in Flanders*, uma carta é considerada, em primeira instância, uma estrutura composta por três secções:

- (i) Uma secção inicial, que pode conter os seguintes elementos: local de redação; data; saudação;
- (ii) A secção do conteúdo da carta em si, composta por uma sequência de parágrafos;
- (iii) Uma secção final, que pode conter os seguintes elementos: saudação; assinatura; local de redação; data; *postscript*.

O esquema de anotação proposto para o projeto baseia-se na estrutura de carta acima descrita.

2.2.1. TEXT ENCODING INITIATIVE

Após a transcrição dos manuscritos, e tendo em posse toda a informação necessária para a realização da edição eletrónica que se pretende, foi necessário proceder à investigação de normas e propostas de anotação existentes para a correspondência. Entre as normas existentes que propõem esquemas de anotação de metadados, a norma escolhida foi a *Text Encoding Initiative* (TEI)⁸, pois esta é a norma mais utilizada na edição e criação de arquivos digitais na área das Ciências Humanas, tendo como principal objetivo a elaboração de diretrizes para a preservação digital de documentos nas Humanidades. Para além deste aspeto, a TEI fornece propostas de anotação para vários tipos textuais, e facilita o trabalho de preservação desde a criação dos metadados à anotação do conteúdo do objeto a preservar. Facilita também o intercâmbio de informações em suporte digital através de uma anotação descritiva do texto:

- (i) Classificando os componentes do texto de acordo com a sua função na estrutura lógica do documento;
- (ii) Possibilitando o processamento do mesmo documento eletrónico, sem nenhuma alteração, por vários processadores diferentes.

⁸ As diretrizes da *Text Encoding Initiative* estão disponíveis em: <http://www.tei-c.org/index.xml>.

Na base da *Text Encoding Initiative* está a metalinguagem de anotação XML (*eXtensible Markup Language*), que apresenta as seguintes características:

XML, (...), has become an international success in a very short period of time. It is used in many fields, academic and commercial, for documents, data files, configuration information, temporary and long-term storage, for transmitting information locally or remotely, by new start-ups and multinational conglomerates. It is used as a storage format for almost everything, including word processing documents, installation scripts, news articles, product information, and web pages. More recently XML has been increasingly popular as a temporary storage format for web-based user interfaces. Its all-pervasive applicability has meant not only that there are numerous tools which read, write, transform, or otherwise manipulate XML as an application, operating system and hardware-independent format, but also that there is as much training and support available. (Cummings, 2008)

A metalinguagem XML, desenvolvida pelo Consórcio W3C (*World Wide Web Consortium*)⁹, é uma linguagem utilizada para descrever outras linguagens de anotação, caracterizando-se pela sua transversalidade em todos os domínios de aplicação informáticos. É uma linguagem extensível, baseada em regras de boa-formação e no conceito de validação formal de documentos. A linguagem XML representa dados estruturados através de uma anotação descritiva, ou seja, possibilita a anotação de partes constituintes de um texto de acordo com a função semântica que as mesmas desempenham na estrutura e no contexto textual.

A linguagem XML concebe os documentos como sendo instâncias de determinado tipo de documento (*document type*). Assim, o tipo de documento ao qual determinado texto pertence, depende dos componentes que o integram e da relação hierárquica entre os elementos/componentes. Se os documentos pertencem a vários tipos, é possível definir a estrutura formal dos mesmos através da criação de uma Definição do Tipo de Documento ou *Document Type Definition* (DTD). A definição da estrutura formal do DTD permite validar o documento XML, verificando, assim, se este obedece à estrutura formal definida.

A linguagem de anotação XML assegura, por um lado, a legibilidade dos ficheiros XML em diversos tipos de *softwares*, sem perda de informação, facto que facilita a conversão do ficheiro para formatos diferentes do original; por outro, a

⁹ Cf. a página oficial do Consórcio W3C em: <http://www.w3.org/>.

interoperabilidade, tornando assim os textos independentes de outras plataformas e aplicações.

No âmbito do presente projeto, tomaram-se como projetos modelo os arquivos digitais *The Van Gogh Letters Project*¹⁰ e o *Digital Archive of Letters in Flanders (DALF)*¹¹. Ambos estes projetos de edição seguem as diretrizes de anotação da TEI. Embora a TEI proponha diretrizes para um vasto número de géneros textuais, tais como textos em prosa, poesia, textos performativos, textos orais, textos lexicográficos, não apresenta diretrizes específicas para anotação de material epistolográfico. Confrontado com esta lacuna, o projeto DALF procedeu à extensão das diretrizes da TEI para anotação de elementos específicos de correspondência não previstos na TEI.

No que diz respeito à transcrição e anotação das cartas do acervo aqui em análise, a edição digital conjuga as propostas da TEI e as propostas específicas do projeto DALF.

¹⁰ Para uma descrição pormenorizada do projeto, consulte-se o seguinte site: <http://www.vangoghletters.org>.

¹¹ O projeto DALF está a ser desenvolvido pelo *Centre for Scholarly Editing and Document Studies*. Para mais informações sobre os princípios editoriais do projeto, veja-se: <http://www.ctb.kantl.be/>.

CAPÍTULO 3:
PRINCÍPIOS E NORMAS DE CODIFICAÇÃO ELETRÔNICA

3.1. INTRODUÇÃO

Todo o documento estruturado conforme as diretrizes da TEI apresenta duas secções: (i) um cabeçalho TEI (*teiHeader*) que fornece dados bibliográficos sobre o documento original e a criação do documento eletrónico a partir do documento original; (ii) o texto, a carta, em si. Segue abaixo a representação da estrutura de um documento TEI.

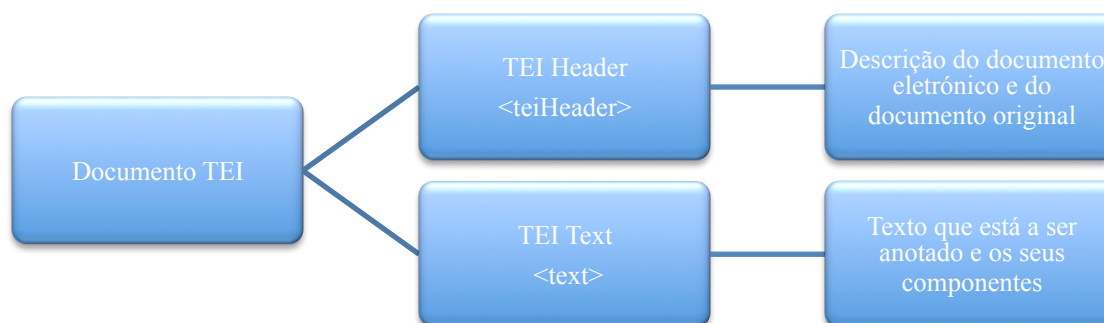


Figura 2: Esquema da estrutura de um documento TEI

Após o processo de transcrição das 107 cartas do *Acervo da Correspondência de Camilo Castelo Branco*, procedeu-se à anotação da estrutura e de determinados elementos textuais de cada carta. Isto significa que a primeira secção do documento TEI a ser trabalhada foi a secção `<text>`. A secção `<teiHeader>` foi elaborada posteriormente e será objeto de análise no Capítulo 4 deste relatório.

A secção `<text>` da versão eletrónica das cartas é composta por uma subsecção `<body>`, que, por sua vez, é constituída por três partes conforme a definição de "carta" descrita anteriormente: `<opener>`; sequência de parágrafos - `<p>`; `<closer>`. As anotações `<opener>` e `<closer>` são propostas do projeto DALF.

3.2. ANOTAÇÃO DA CORRESPONDÊNCIA DE CAMILO CASTELO BRANCO

Segue-se abaixo uma descrição detalhada de todo o processo de anotação das 107 cartas transcritas, incluindo as dificuldades e as soluções encontradas.

O elemento <opener> pode agrupar um conjunto de elementos que indicam o início da carta, tais como: a saudação inicial, a data e o local. A possibilidade de existência destes elementos numa carta é opcional, uma vez que nem todos os signatários seguem a mesma ordem de ideias quando iniciam uma carta.

Como podemos verificar na Figura 3, o signatário não inicia a carta com indicação da data e local de redação. Outros casos existem em que o signatário coloca esta informação da data e local de redação na secção <opener>. O elemento <opener> da carta 57, representado na Figura 3, é composto pelo texto: Ill.^{mo} Ex.^{mo} Snr. [Illustríssimo Excellentíssimo Senhor].

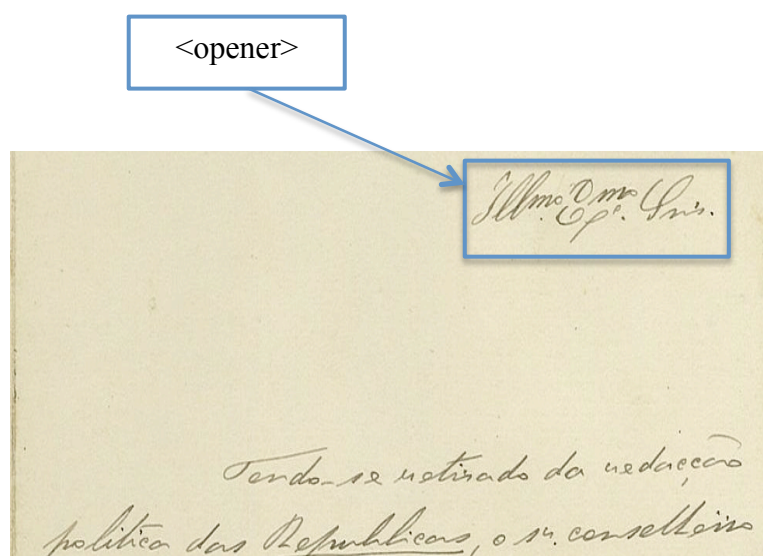


Figura 3 - Excerto da Carta 57 do Acervo da Correspondência de Camilo Castelo Branco¹²

¹² Carta de António Barros dirigida a Camilo Castelo Branco.

O elemento <closer> pode conter uma série de elementos que indicam o final da carta: a saudação final, a assinatura, a data, entre outros dados que possam ser acrescentados pelo signatário.

A Figura 4 apresenta a secção <closer> da carta 89 do Acervo, secção que agrupa as seguintes informações: local de redação (Villa Real); data (9 de Fevrº [Fevereiro] d'1868); assinatura do signatário (A. J. d'Azd.º [António José de Azevedo])

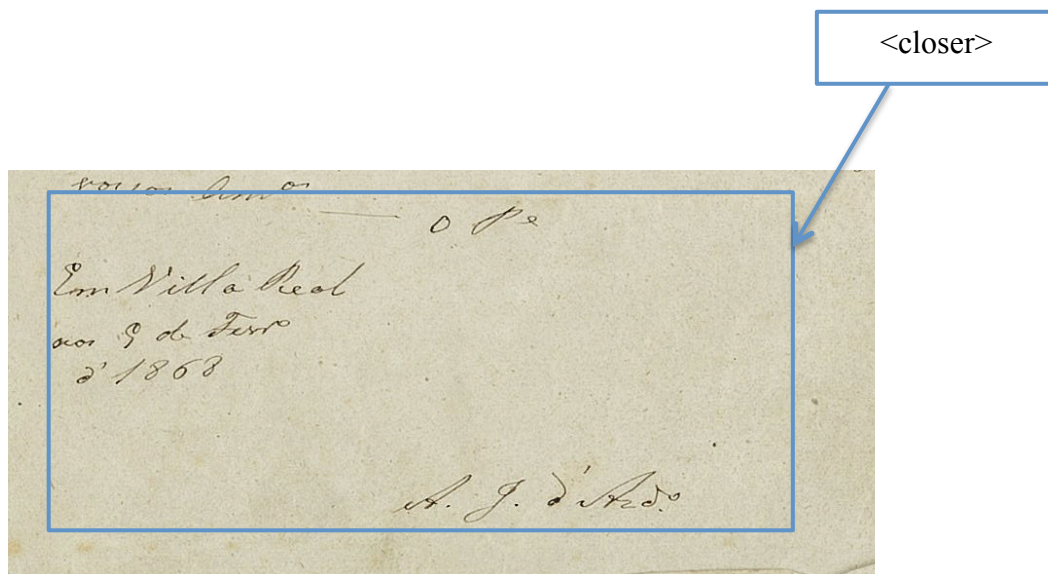


Figura 4 - Excerto da Carta 89 do *Acervo da Correspondência de Camilo Castelo Branco*¹³

Os elementos utilizados neste projeto para anotar o conteúdo das cartas em termos estruturais e textuais foram os seguintes: <p> (*paragraph*), <lb/> (*line break*) e <pb/> (*page break*). Todos estes elementos fazem parte das propostas de anotação da TEI.

O elemento <p> marca os parágrafos, podendo ter atributos associados a si.

O elemento <lb/> indica uma quebra de linha ou, se preferirmos, o início de uma nova linha. Este é um elemento vazio.

O elemento <pb/> é, tal como o <lb/>, um elemento vazio que indica uma quebra de página, ou seja, que no documento original se inicia uma nova página.

Os elementos podem admitir atributos que os qualificam ou servem para fornecer informação adicional. Vejamos de seguida o funcionamento dos atributos

¹³ Carta de António José de Azevedo dirigida a Camilo Castelo Branco

associados ao elemento <pb/>: o elemento <pb/> é complementado com o atributo *n* (*number*): <pb n="2"/>. Neste caso, o atributo dá-nos o conhecimento do número (n) de página que corresponde ao documento original, e o valor entre aspas corresponde à ordem numérica da página.

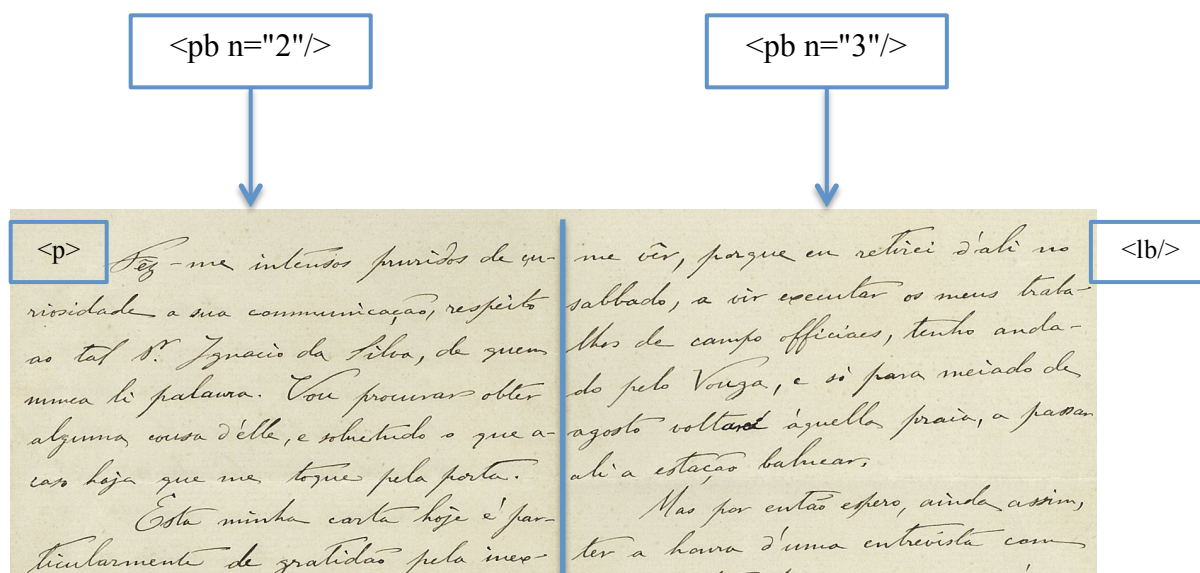


Figura 5 - Excerto da Carta 16 do Acervo da Correspondência de Camilo Castelo Branco¹⁴

A Figura 5 apresenta um excerto das páginas número 2 e 3 da Carta 16. A anotação do primeiro parágrafo da página 2 tem a seguinte estrutura:

```
<pb n="2"/>
<p>Fêz-me intensos pruridos de cu-
riosidade a sua comunicação, respeito
ao tal S.r [Senhor] Ignacio da Silva, de quem
nunca li palavra. Vou procurar obter
alguma cousa d'elle, e sobretudo o que a
caso haja que me toque pela porta.
```

No final da página 2, início da página 3, insere-se a anotação <pb n="3"/>, e assim sucessivamente.

Para fazer a distinção de parágrafos <p> que desempenham funções diferentes nas cartas, recorreu-se ao atributo *class*.

¹⁴ Carta de Abel Acácio Botelho dirigida a Camilo Castelo Branco

Neste projeto foram utilizados três tipos diferentes de atributos *class* associados ao elemento `<p>`:

- (i) `<p class="salute">` Este código atribui ao elemento `<p>` o valor de *salute*, ou seja, indica que se trata de um parágrafo que contém uma saudação ou cumprimento por parte do signatário da carta.
- (ii) `<p class="signature">` Este código atribui ao elemento `<p>` o valor de *signature*, indicando, assim, que o parágrafo anotado contém uma assinatura.
- (iii) `<p class="dateline">` Este código é classificado como um parágrafo que contém informação sobre o local e data de redação da carta.

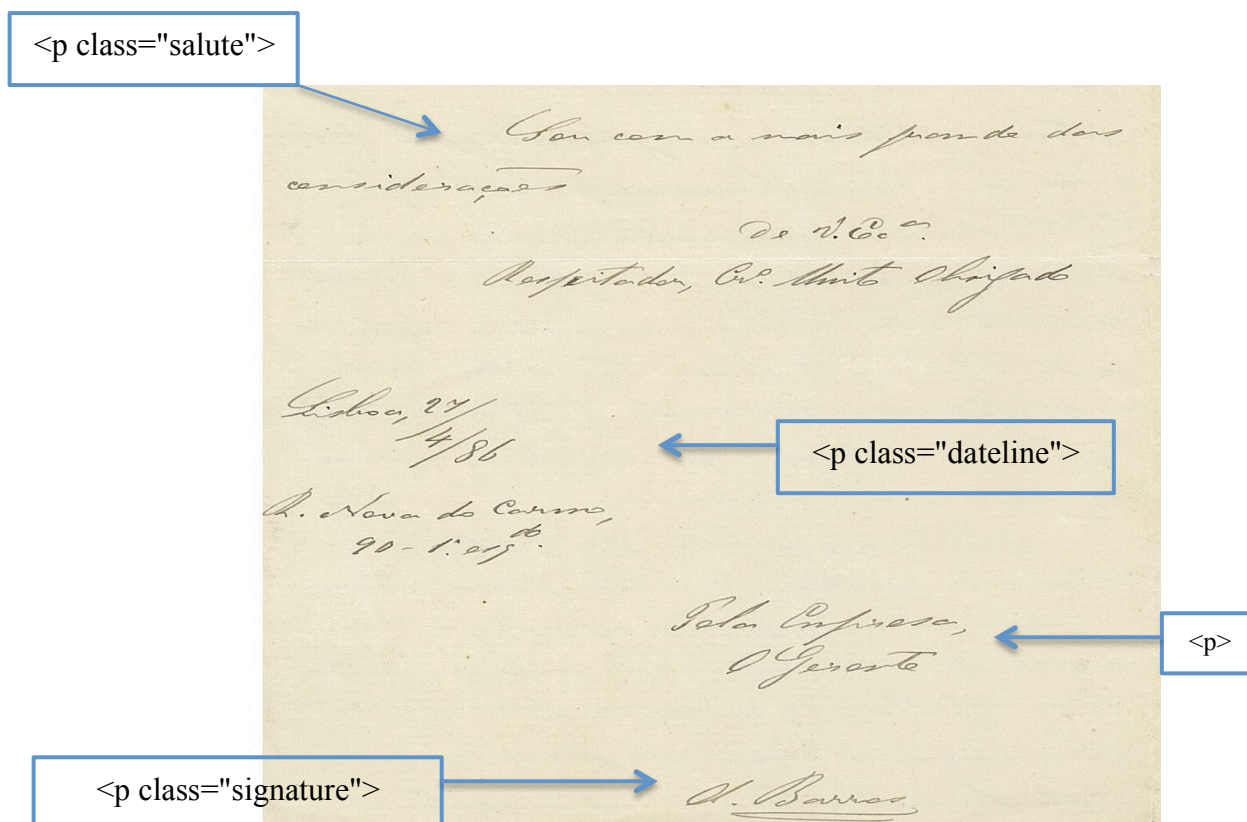


Figura 6 - Excerto da carta 57 do *Acervo da Correspondência de Camilo Castelo Branco*¹⁵

¹⁵ Carta de António Barros dirigida a Camilo Castelo Branco

A Figura 6 representa a secção <closer> da carta 57, cuja anotação segue abaixo:

```
<p class="salute">Sou com a mais grande das<b/>
considerações<b/>
De V.Ex.a [Vossa Excellencia]<b/>
Respeitador, Cr.o [Criado], Muito Obrigado</p>

<p class="dateline">Lisboa, 27/4/86<b/>
R. Nova do Carmo,<b/>
90 - 1.o esq.do [esquerdo]</p>

<p class="signature">A. Barros</p>
```

Como se pode verificar pela Carta 57, os documentos epistolográficos apresentam, não raras vezes, propriedades e informações que não são de fácil classificação e anotação. Neste caso, estamos perante um signatário que escreve e assina em nome da empresa que representa.

No caso específico da anotação <p class="dateline">, optou-se pela inclusão dos seguintes elementos:

<place>

Elemento que contém a informação do local da redação da carta.

<address>

Elemento que abarca toda a informação do endereço postal.

<addrLine>

Elemento que contém uma linha do endereço postal.

<date>

Elemento que contém a data em qualquer formato. Como seria de prever, as datas aparecem nas cartas nos mais diversos formatos. Foi, assim, necessário criar uma anotação, seguindo as normas da TEI, que fosse uniforme no que toca à representação das datas que aparecem nos documentos. Para tal, foi necessário

associar o atributo *when* ao elemento <date>. O valor atribuído por *when* a <date> é representado pelo seguinte formato *standard* de data: "dd.mm.aaaa". Não raras vezes, as datas aparecem incompletas. Nestes casos, optou-se pela utilização do hífen (-) para indicar falta de informação. Assim sendo, a data que aparece na carta 57 abaixo foi anotada da seguinte forma: <date when="27.04.1886">

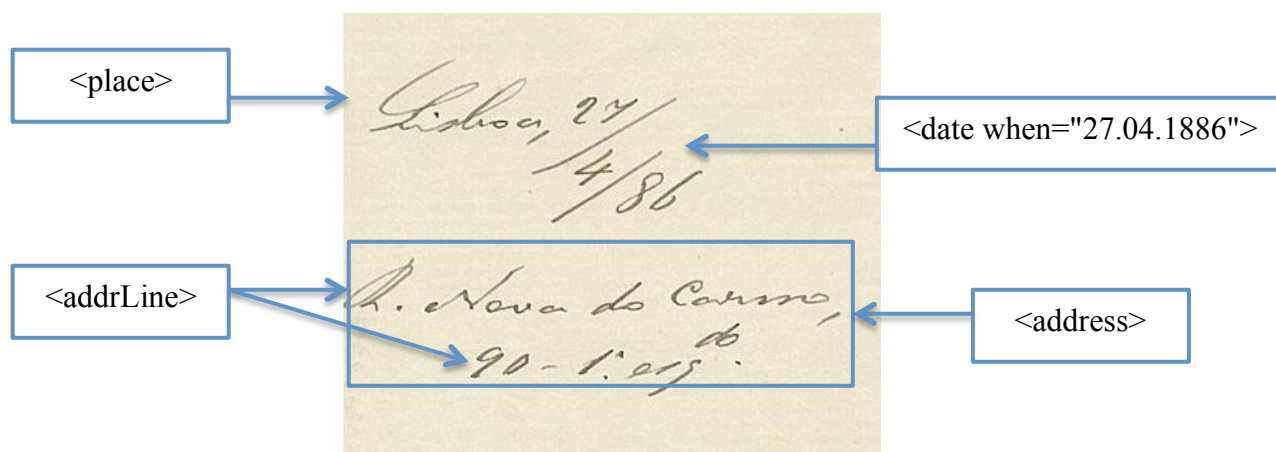


Figura 7 - Excerto da carta 57 do *Acervo da Correspondência de Camilo Castelo Branco*

Com base na Carta 57, a anotação <p class="dateline"> apresenta a seguinte estrutura:

```
<p class="dateline">
  <place>Lisboa</place>,
  <date when="27.04.1886">27/4/86</date>
  <address>
    <addrLine>R. Nova do Carmo,</addrLine>
    <addrLine>90 - 1.º esq.do [esquerdo].</addrLine>
  </address>
</p>
```

Vejamos de seguida as propostas para a anotação de texto "pré-impresso" nas cartas: O projeto DALF propõe para este género de texto/informação o elemento <print>, inserido antes do início do <opener>.

O elemento <print> tem a função de anotar material impresso no documento original, e que estava presente no suporte da carta, antes do ato de escrita. Este elemento visa preservar a anotação deste tipo de blocos de informação presentes na correspondência (texto ou material "pré-impresso" na carta).

As cartas podem ser escritas ou impressas, em papel ou noutro tipo de material, que contém texto previamente impresso ou inscrito. Estes blocos de informação ou informações devem ser vistos como parte integrante de uma carta dado que podem desvendar informação relevante sobre o signatário, por exemplo.

A anotação do elemento <print> requer a adição de um atributo, uma vez que, como já referido anteriormente, existem vários tipos de blocos de informação que a carta pode conter. O atributo *type* utilizado pelo projeto DALF fornece informações relativamente à posição da informação na carta.

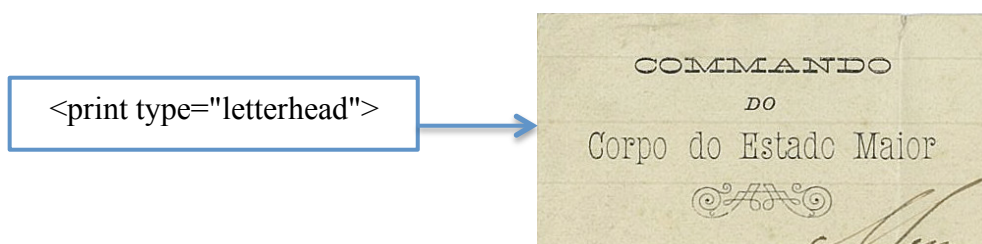


Figura 8 - Excerto da carta 25 do *Acervo da Correspondência de Camilo Castelo Branco*¹⁶

Tomando como exemplo a Carta 25, representada na Figura 8, a anotação da informação timbrada no cabeçalho da carta tem a seguinte forma:

```
<print type="letterhead">COMMANDO<b/>  
DO<b/>  
Corpo do Estado Maior</print>
```

¹⁶ Carta de Abel Acácio Botelho dirigida a Camilo Castelo Branco

Uma outra característica comum a textos epistolográficos é a ocorrência de bloco(s) de informação após o fecho da carta, designado(s) de *postscript(s)*.

No âmbito deste projeto, o elemento <postscript> é utilizado para anotar informação que ocorra após a despedida ou saudação e assinatura do signatário. Por norma, os *postscripts* são assinalados nas cartas através da abreviatura P.S. Nas cartas do *Acervo da Correspondência de Camilo Castelo Branco*, ocorrem *postscripts* introduzidos pelas abreviaturas P.S. e N.B. (*Nota Bene*). Estas abreviaturas são etiquetadas com o elemento <label>. O bloco de texto correspondente ao *postscript* em si é inserido num elemento <p>.

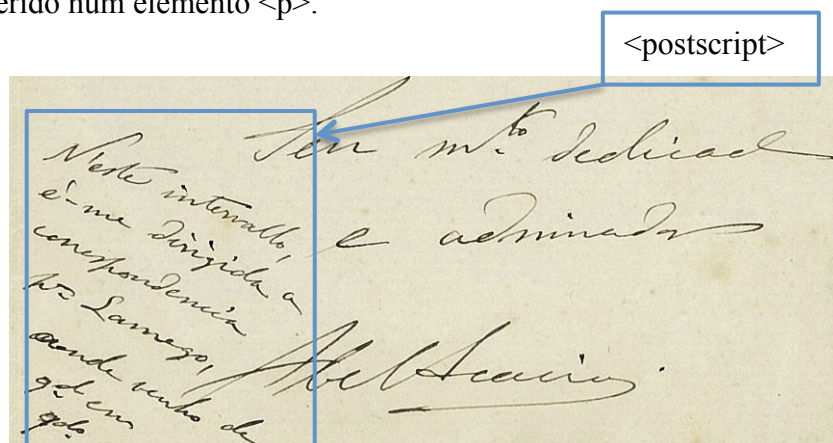


Figura 9 - Excerto da carta 16 do *Acervo da Correspondência de Camilo Castelo Branco*

A Figura 9 apresenta a secção final da carta de Abel Acácio Botelho¹⁷ a Camilo. Como é possível verificar, o autor da carta escreve um *postscript* na margem esquerda da carta, informação que foi assinalada da seguinte forma:

<postscript>

<p>N´este intervallo,<lb/>

é-me dirigida a<lb/>

correspondência<lb/>

p.^a [para] Lamego,<lb/>

aonde venho de<lb/>

q.^{do} [quando] em<lb/>

q.^{do} [quando]</p>

</postscript>

¹⁷ De nome completo Abel Acácio Almeida Botelho, foi coronel e chefe do Estado Maior do Exército. Apaixonado pelas letras, desenvolveu intensa atividade jornalística, conquistando uma posição de relevo na literatura nacional, sobretudo como romancista. (Cabral, 1988)

As anotações anteriormente apresentadas referem-se principalmente à anotação estrutural do documento epistolográfico, estrutura essa que o define e identifica como tal. Para além de este nível de anotação, existe um segundo nível de anotação, nomeadamente a etiquetação de propriedades textuais, tal como: abreviaturas; conteúdo ilegível, sublinhado, apagado, rasurado ou até mesmo adicionado; a marcação de espaços de alguma dimensão existentes no documento original da carta; a anotação de citações presentes no texto, entre outros. Para efetuar estas anotações foram seguidas as diretrizes propostas pela TEI.

Uma grande parte da correspondência trabalhada no âmbito deste projeto contém abreviaturas. É comum encontrar palavras abreviadas, sendo que este fenómeno é mais comum em certos signatários do que em outros. Dado o número significativo de abreviaturas em cada uma das cartas, foi decidido expandir as abreviaturas de forma a ser possível facultar uma versão de leitura ao utilizador do Arquivo Digital. As abreviaturas foram anotadas com o elemento <abbr>. A expansão da abreviatura é colocada no atributo *expan*. Vejamos o seguinte exemplo:

`<abbr expan="Dom">D.</abbr>`

Uma particularidade das abreviaturas em português é o facto das mesmas conterem caracteres superiores à linha. Por norma, o signatário ao abreviar uma palavra, escreve a(s) última(s) letra(s) da palavra superiores à linha. Estes caracteres foram anotados com o elemento <sup> (*superscript*).

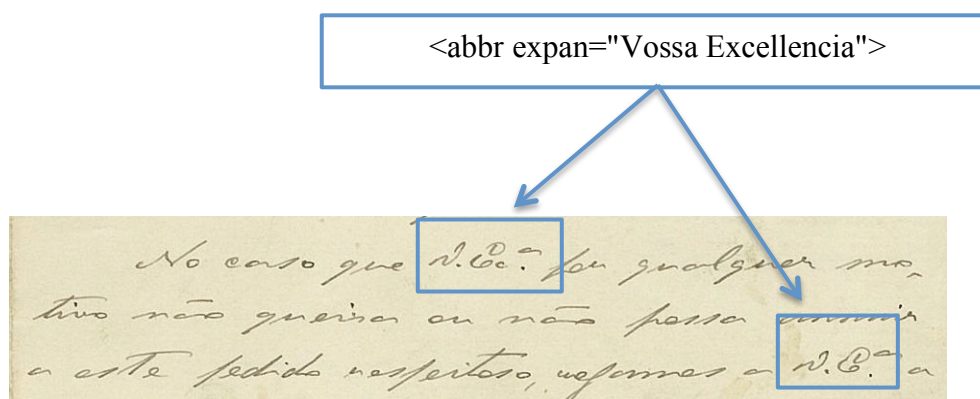


Figura 10 - Excerto da carta 57 do Acervo da Correspondência de Camilo Castelo Branco

A Figura 10 apresenta uma das abreviaturas mais comuns ao longo de toda a correspondência: V. Ex^a (Vossa Excelência). Esta abreviatura foi anotada da seguinte forma:

```
<abbr expand="Vossa Excellencia">V.Ex<sup>a</sup></abbr>
```

Embora no exemplo acima estejamos perante duas palavras abreviadas, estas são tidas como uma unidade. Daí o valor do atributo *expand* conter a expansão da unidade abreviada.

Ainda em relação às abreviaturas, quando confrontados com uma abreviatura desconhecida, o valor associado ao atributo *expand* é *unclear*.

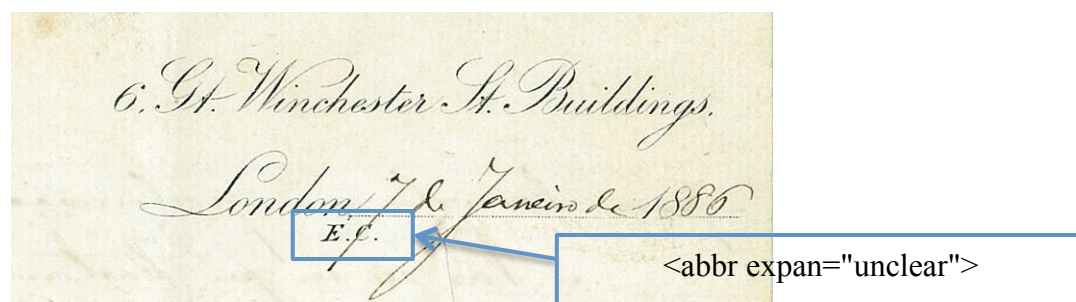


Figura 11 - Excerto da carta 69 do *Acervo da Correspondência de Camilo Castelo Branco*¹⁸

A Figura 11, representa o cabeçalho da Carta 69, que contém uma morada e a data de redação da carta. Para além desta informação, aparece a abreviatura desconhecida "E.C.". Ao anotar o facto da abreviatura ser desconhecida através do atributo *unclear*, ver código abaixo, é possível localizar esta informação mais tarde, caso a dúvida se venha a esclarecer.

```
<abbr expand="unclear">E.C.</abbr>
```

O processo de transcrição das 107 cartas foi dificultado pelo facto de terem sido transcritos documentos de 32 correspondentes diferentes, incluindo Camilo, cada um com a sua caligrafia. Segue-se uma descrição das dificuldades encontradas no ato de transcrição e as soluções de anotação encontradas.

¹⁸ Carta de J. A. de Azevedo Castro dirigida a Camilo Castelo Branco

O elemento `<unclear>` foi utilizado para marcar texto que apresentava dificuldades em termos de decodificação, mas para o qual era possível apresentar uma proposta de transcrição.

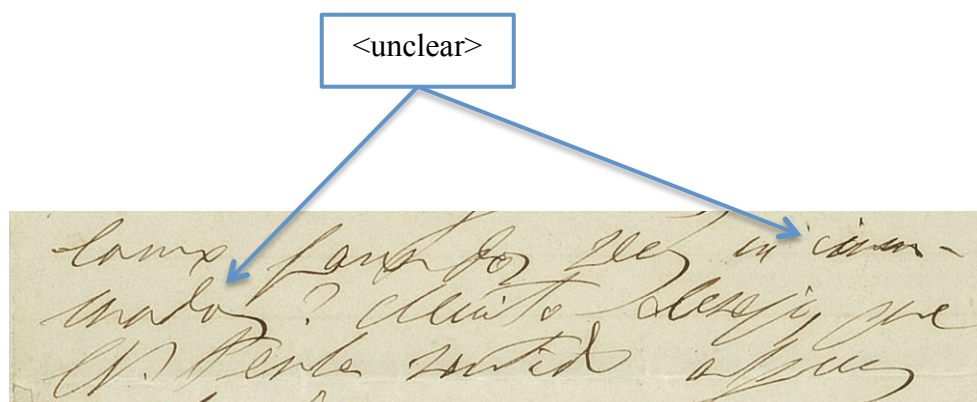


Figura 12 - Excerto da carta 46 do Acervo da Correspondência de Camilo Castelo Branco¹⁹

O elemento `<unclear>` pode ser utilizado juntamente com vários atributos propostos pelas diretrizes da TEI. No âmbito deste projeto, este elemento apenas se encontra associado ao atributo *reason*.

Tal como o próprio atributo *reason* indica, o valor do mesmo explicita o motivo que levou à dificuldade de transcrição.

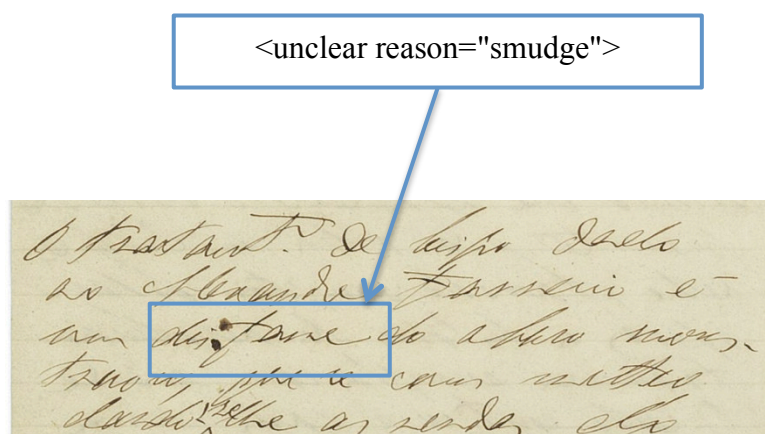


Figura 13 - Excerto da carta 46 do Acervo da Correspondência de Camilo Castelo Branco

A Carta 46, anteriormente representada, contém uma palavra difícil de decifrar devido a manchas de tinta. De qualquer forma, foi possível fazer uma proposta de anotação, esta marcada com o código `<unclear>`.

¹⁹ Carta de Joaquim Alves Mateus dirigida a Camilo Castelo Branco

`<unclear reason="smudge">disfarse</unclear>`

No caso da impossibilidade de decifrar uma palavra, frase ou blocos de texto, foi utilizado o elemento `<gap/>`, indicando que naquele determinado local do documento original existe texto omitido, invisível ou ilegível. Como já havíamos referido, a parte do texto que é anotada como ilegível pode ser de menor ou maior extensão. Tal informação é registada através do atributo *extent* que especifica a extensão do texto impossível de decifrar. No caso de encontrarmos uma palavra ilegível em determinada carta, a estrutura final desta anotação seria: `<gap extent="1 word"/>`.

Também no caso do elemento `<gap/>`, é possível especificar o motivo que esteve na origem da não transcrição do texto através do atributo *reason*. No exemplo abaixo, o código `<gap/>` diz respeito à não transcrição de texto com a extensão de uma palavra por motivo de ilegibilidade.

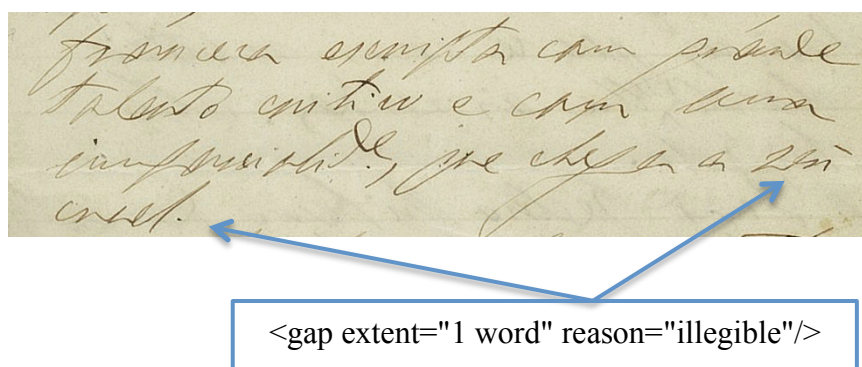


Figura 14 - Excerto da carta 46 do *Acervo da Correspondência de Camilo Castelo Branco*

Um dos objetivos principais da edição eletrónica aqui em análise é a criação de representações eletrónicas muito próximas dos documentos originais. Isto significa que é necessário prestar atenção a pormenores, como: texto sublinhado; rasurado; riscado; texto adicionado; espaçamentos que causam estranheza, etc. A anotação destas particularidades levanta, não raras vezes, questões de interpretação. De que forma deve ser interpretado/anotado texto sublinhado? Como texto enfatizado pelo autor (neste caso, código TEI `<emph>` (*emphasized*) ou como texto destacado tipograficamente (neste caso, código TEI `<hi>` (*highlighted*). Uma vez que é difícil

saber-se com qual intenção o autor de uma carta sublinhou determinado texto, neste projeto recorreu-se ao elemento `<hi>` para anotar texto sublinhado. Tal como é possível verificar no exemplo abaixo, o elemento `<hi>` pode ser complementado pelo atributo *rend* (*rendition*) para especificar a forma como o texto se encontra destacado.

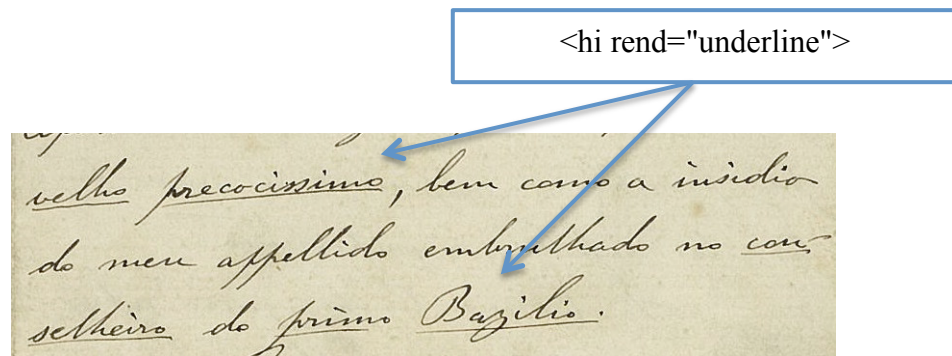


Figura 15 - Excerto da carta 16 do *Acervo da Correspondência de Camilo Castelo Branco*

`<hi rend="underline">velho precocissimo</hi>`, bem como a
insidia`<lb/>`
do meu appellido embrulhado no `<hi rend="underline">con-`
`<lb/>`
selheiro`</hi>` do `<hi rend="underline">primo Bazilio</hi>`.

No processo de transcrição foram algumas as cartas que apresentavam texto riscado pelo autor. Nestes casos optou-se pela utilização do elemento `` (*delete*), complementado pelo atributo *rend*, que especifica a forma como o texto foi riscado. Vejamos o seguinte excerto da Carta 46 do Acervo:

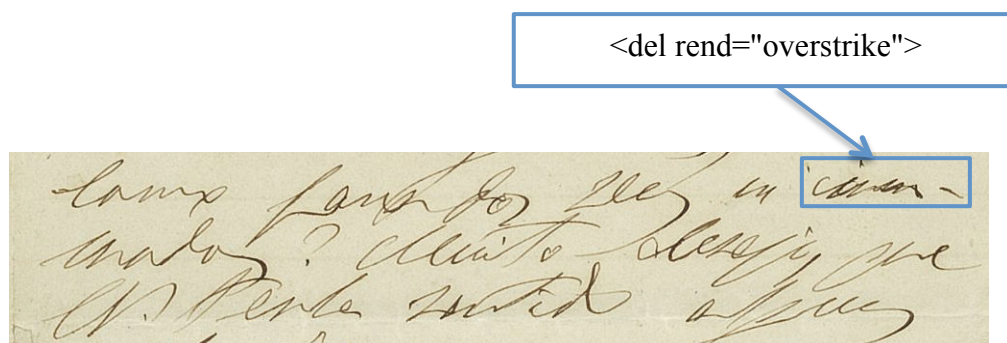


Figura 16 - Excerto da carta 46 do Acervo da Correspondência de Camilo Castelo Branco

Embora este projeto incida em primeira instância na anotação descritiva da correspondência, por vezes é necessário recorrer a anotações de natureza procedimental ou tipográfica para conseguirmos criar uma edição eletrónica mais próxima do documento original.

Uma carta pode, como já referido no Capítulo 2 do relatório, conter outros géneros textuais ou texto proveniente de outras fontes, como, por exemplo, citações. Neste caso particular do *Acervo da Correspondência de Camilo Castelo Branco*, encontrámos em certas cartas versos e/ou citações. Estes aspetos foram identificados na edição eletrónica com o elemento `<quote>`, que nos remete para uma frase/passagem citada pelo signatário. O atributo *rend* foi utilizado para complementar o elemento `<quote>` no que respeita a aspetos de disposição do texto citado. No exemplo que se segue, o texto citado (colocado no quadrado em azul) não se encontra tipograficamente destacado do resto do texto.²⁰

²⁰ A citação da Carta 44 foi anotada da seguinte forma:

```
<quote rend="margin"><p>Na onda turva, inquieta, glabra,<b/>
Passa a veloz dansa macabra<b/>
De imperadores, papas, reis...<b/>
Jograes de um dia ou d'uma hora...<b/>
Baixel, que vê partir a <unclear reason="smudge">aurora</unclear><b/>
E a noite some entre <unclear reason="smudge">pasceis</unclear>.</p></quote>
```

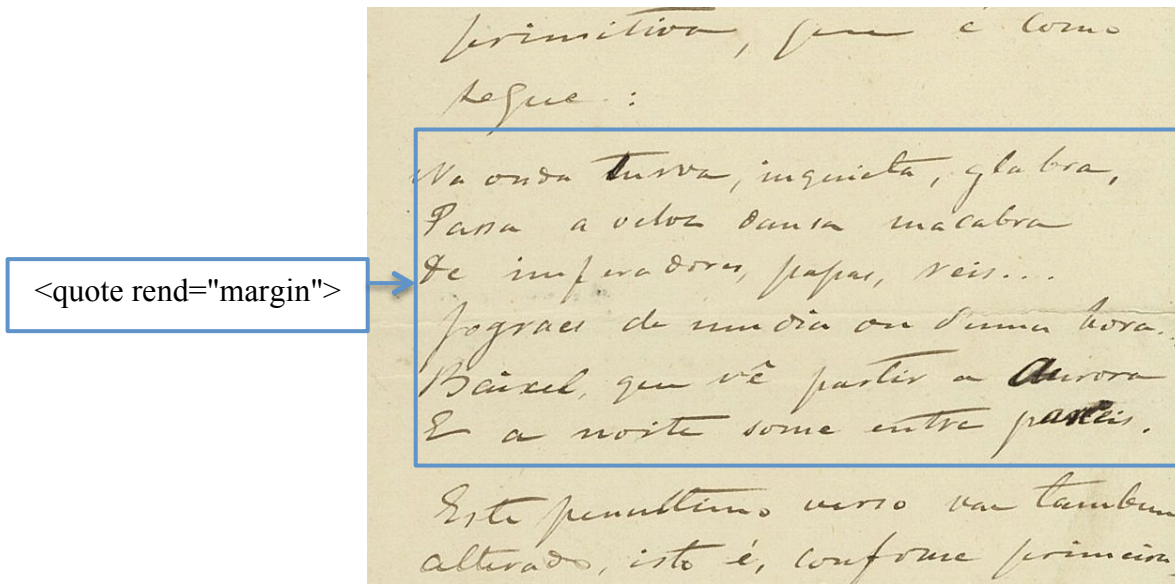


Figura 17 - Excerto da carta 44 do Acervo da Correspondência de Camilo Castelo Branco²¹

Compare-se este exemplo com a citação que consta na Figura 18, no qual o texto citado se encontra visivelmente afastado da margem.²²

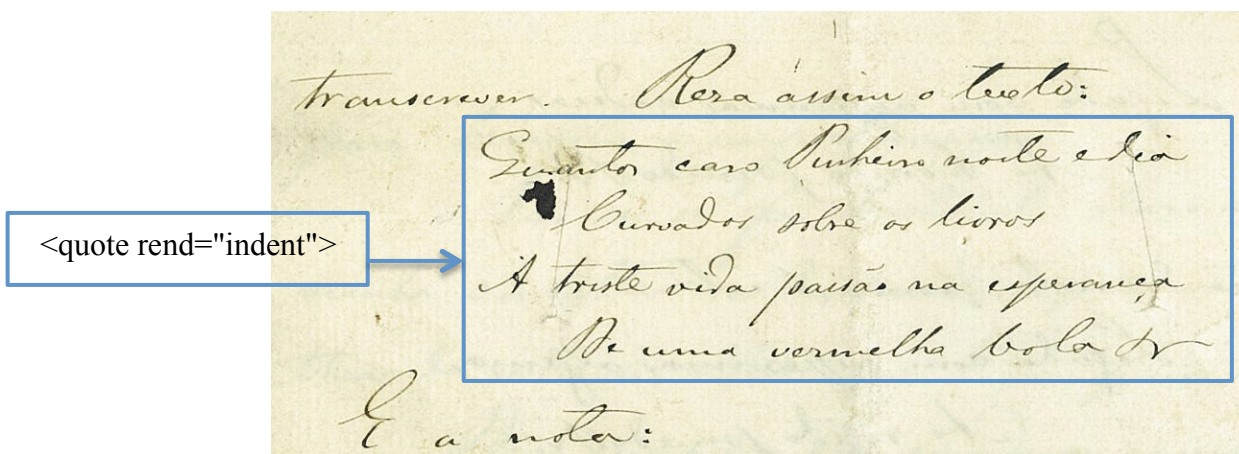


Figura 18 - Excerto da carta 69 do Acervo da Correspondência de Camilo Castelo Branco

O atributo *rend* foi utilizado no elemento `<quote>` para especificar a disposição do texto citado: *margin* (texto encostado à margem); *indent* (texto afastado da margem)

²¹ Carta de Manuel Duarte de Almeida dirigida a Camilo Castelo Branco

²² A citação da Carta 69 foi anotada da seguinte forma:

`<quote rend="indent"><p>Quantos caro Pinheiro noite e dia<lb/>`

`Curvados sobre os livros<lb/>`

`A triste vida paixão na esperança<lb/>`

`De uma vermelha bola <gap extent="1 letter" reason="illegible"/><p/></quote>.`

Por vezes o remetente tem necessidade de acrescentar texto ao que já escreveu, por exemplo, por esquecimento no momento de escrita, ou simplesmente para acrescentar mais informação à ideia que pretende transmitir.

Para anotar tais situações, foi utilizado o elemento <add> (*addition*), mencionando assim, uma adição de texto ao já existente. De modo a identificar, na edição eletrónica, o local onde foi adicionado o texto, foi inserido o elemento <add> com o atributo *place*, e este último com o valor *above* - <add place="above"> - quando o texto adicionado aparece acima do texto que vem complementar.

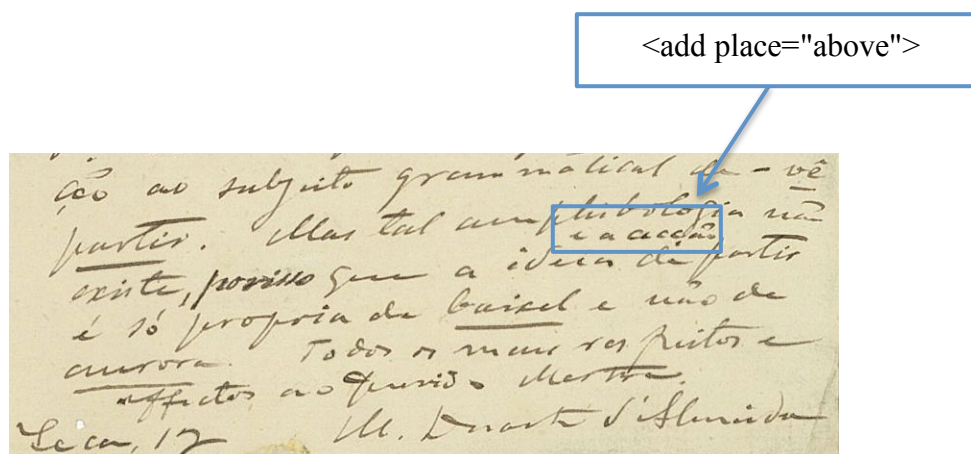


Figura 19 - Excerto da carta 44 do *Acervo da Correspondência de Camilo Castelo Branco*

A anotação de situações como a que se encontra representada na Figura 19, tem a seguinte forma:

<add place="above">e a acção</add>

Como foi possível verificar pela exposição acima dos processos de transcrição e anotação das primeiras 107 cartas do *Acervo da Correspondência de Camilo Castelo Branco* reunido pela Comissão de Homenagem no início do século XX, a criação de uma edição eletrónica envolve a definição de todo um conjunto de princípios de edição a aplicar na transcrição e anotação dos documentos a preservar. O processo de definição dos princípios de edição deverá ter como elementos orientadores, por um lado, o público interessado no material do acervo e, por outro, a ideia do produto final que se pretende conceber e disponibilizar, por exemplo, a

integração do material tratado e processado num arquivo digital com funcionalidades de pesquisa específicas. No caso do tratamento de documentos epistolográficos, não é possível, pela experiência adquirida no âmbito deste projeto, definir com precisão todos os princípios e normas de edição a adotar no início do projeto. À medida que se vai conhecendo cada um dos documentos que compõe o acervo, vai-se repensando e acrescentando às normas de edição inicialmente propostas.

CAPÍTULO 4:

METADADOS E TEIHEADER

4.1. METADADOS

Os metadados desempenham um papel importante no que respeita a catalogação e classificação de informação, processos imprescindíveis para o intercâmbio de informação entre instituições e projetos, nomeadamente projetos de edição eletrónica. Um projeto de edição eletrónica depara-se inevitavelmente com a questão do esquema de metadados a adotar para os objetivos definidos. Esta tarefa não é fácil, uma vez que existem vários esquemas ou normas de metadados que foram sendo desenvolvidas. A secção que se segue é o resultado do levantamento de questões relevantes que se colocam aquando da elaboração de um projeto de edição eletrónica (Schmidt, 2014; Smith, 2004.).

Os metadados têm várias funções e vários propósitos:

- Representam a informação estruturada que descreve o recurso de informação em termos de conteúdo, contexto e estrutura;
- Ajudam a localizar e identificar o recurso da informação do documento;
- Ajudam a extrair, recuperar, utilizar e gerir o recurso da informação;
- Esclarecem questões de autenticidade, autoridade, acessibilidade, interoperabilidade e usabilidade;
- Identificam relações estruturais internas e entre objetos de informação.

Quanto mais pormenorizada e consistente for a descrição do objeto de informação, mais facilmente essa estrutura pode ser pesquisada, manipulada e interrelacionada com outros objetos de informação.

Existem vários tipos de metadados: administrativos, descritivos, de preservação, técnicos e de uso.

Os metadados administrativos identificam o objeto digital e o respetivo original no repositório ao qual pertence, através da indicação dos dados do repositório, coleção, o número de identificação, a entidade responsável pela publicação e distribuição, os direitos de autor, utilização e a sua reprodução. Este tipo

de metadados descreve a menção de responsabilidade de quem criou o objeto original e quando o mesmo foi criado, e também quem criou e é responsável pela preservação e manutenção do objeto original. Os metadados administrativos têm ainda a função de descrever o historial do objeto a preservar, contendo informação acerca da história do objeto antes da sua aquisição, mencionando a sua origem e proveniência, e também o processo e termos de aquisição do objeto. Contêm também informação acerca do registo de outras versões e/ou reproduções do objeto.

Os metadados descritivos identificam e descrevem os objetos e/ou coleções através da descrição do objeto original, como por exemplo, de um manuscrito, identificando o tipo textual do mesmo, a localização geográfica, a descrição do seu suporte físico, dimensões e a condição ou o estado do objeto; os elementos não textuais relacionados com o ato de escrita (como por exemplo, esquemas, gráficos, desenhos...); a parafernália; no caso concreto da correspondência, o sobrescrito; e o resumo do conteúdo.

Os metadados de preservação: (i) registam todo o processo de criação de um objeto digital a partir do objeto original; (ii) descrevem os processos de preservação digital de longo prazo, garantindo que os conteúdos digitais possam ser consultados e interpretados no futuro; (iii) documentam decisões técnicas de preservação e edição adotadas, ou seja, os métodos e estratégias tomadas para a preservação.

Os metadados técnicos descrevem quais as tecnologias utilizadas para criar e/ou aceder ao objeto digital, documentando assim, qual o *hardware* e *software* utilizados; a migração de conteúdo; e os processos de digitalização: qual o formato, a resolução, os esquemas de compressão.

Os metadados de uso contêm dados acerca da gestão de direitos autorais e propriedade intelectual e dados acerca da autenticidade do recurso digital.

Apesar da existência de uma grande variedade de metadados, existe o desafio da decisão da utilização dos mesmos. O grande problema diz respeito à não existência de um padrão de metadados capaz de atender a todas as necessidades de preservação

dos diferentes tipos de recursos e contextos digitais. Existem, sim, vários esquemas e padrões de metadados desenvolvidos por comunidades diferentes, entre elas, a TEI.

De modo a fazer a anotação dos metadados, é necessário antecipar a informação que será necessária para atingir determinados objetivos de preservação e identificar o(s) esquema(s) que melhor poderá/poderão corresponder às necessidades do gestor de informação, às propriedades do objeto de informação, às propriedades do repositório/acervo onde o objeto está inserido, e às necessidades dos utilizadores.

4.2. O CABEÇALHO DA TEI - TEIHEADER

O presente projeto prevê a criação de uma ficha bibliográfica detalhada conforme as diretrizes da *Text Encoding Initiative*, denominada de *TEI Header* (Cabeçalho TEI) para cada carta transcrita e anotada.

Metadata has come to mean many things, including any form of data that describes digital resources on the Web. But in the context of TEI, metadata means data describing the digital document as a whole, and in TEI it is embodied in the Header, which is an obligatory part of every TEI document. (Schmidt, 2014)

Segundo Schmidt, uma das vantagens da TEI é o facto de as suas diretrizes permitirem a descrição do documento digital como um todo. Este aspeto ficará claro com a descrição do esquema do cabeçalho TEI elaborado para as edições eletrónicas das cartas do *Acervo da Correspondência de Camilo Castelo Branco*.

A ficha bibliográfica *teiHeader* de cada documento eletrónico criado documenta o processo de criação do documento digital a partir do objeto original. A anotação que identifica o cabeçalho TEI tem a seguinte forma:

```
<teiHeader>descrição do documento eletrónico e do documento original</teiHeader>
```

A secção <teiHeader> é composta por quatro partes, uma delas obrigatória, enquanto as restantes são opcionais.

A primeira parte diz respeito ao <fileDesc> (*File Description*), que contém a descrição do processo de criação do documento digital a partir do documento original. O <fileDesc> é o elemento obrigatório do <teiHeader>.

A segunda parte <encodingDesc> (*Encoding Description*) apresenta a descrição das práticas de anotação e edição implementadas.

A terceira parte intitulada <profileDesc> (*Profile Description*) descreve aspetos do texto eletrónico úteis para fins investigativos, tais como: língua(s), variedade linguística, registo(s), contexto de criação, intervenientes, nomes, locais, instituições mencionados nas cartas.

Por fim, a quarta parte <revisionDesc> (*Revision Description*) contém metadados acerca do historial das alterações e correções editoriais efetuadas ao texto eletrónico, por exemplo: o responsável, a data e o motivo que originou as alterações.

No âmbito deste projeto em concreto, foi apenas utilizado o elemento <fileDesc> na elaboração do <teiHeader>, mas na edição eletrónica final do *Arquivo Digital da Correspondência de Camilo Castelo Branco*, a ficha bibliográfica <teiHeader> irá conter duas partes constituintes: o <fileDesc> e também o <profileDesc>, este último trabalhado no âmbito do projeto intitulado "*Edição digital anotada da correspondência dirigida a Camilo Castelo Branco: problemas e desafios*".

O elemento <fileDesc> contém a descrição completa da ficha bibliográfica do documento eletrónico e da fonte do documento original. As informações do <fileDesc> podem ser utilizadas para catalogar, identificar e recuperar o documento; podem ainda ser utilizadas por investigadores que pretendam citar o documento eletrónico.

O <fileDesc> é composto por três elementos obrigatórios:

(i) <titleStmt> (*Title Statement*), que contém dados bibliográficos do documento eletrónico e a menção de responsabilidade dos investigadores envolvidos nas tarefas associadas à criação do documento eletrónico;

(ii) <publicationStmt> (*Publication Statement*), que contém dados relacionados com a publicação e distribuição do documento eletrónico;

(iii) <sourceDesc> (*Source Description*), que fornece uma descrição bibliográfica da fonte principal/original.

As descrições e declarações presentes no cabeçalho TEI podem ser:

- em formato prosa - parágrafos delimitados por <p> e </p>;
- em formato de elementos propostos pela TEI ou outros adequados a determinado propósito/projeto.

Assim, um cabeçalho mínimo TEI tem a seguinte estrutura:

```
<teiHeader>
  <fileDesc>
    <titleStmt> . . . </titleStmt>
    <publicationStmt> . . . </publicationStmt>
    <sourceDesc> . . . </sourceDesc>
  </fileDesc>
</teiHeader>
```

Os exemplos que se seguem são parte integrante do ficheiro <teiHeader> da Carta 1 do *Arquivo Digital da Correspondência de Camilo Castelo Branco*.

A TEI propõe vários elementos para a composição do <titleStmt>, mas no âmbito deste projeto foram utilizados os seguintes: <title>; <author>; <sponsor> e <respStmt> (*Statement of Responsibility*).

O elemento <title> contém o título do documento eletrónico e tem como objetivo distinguir o texto digital do texto original.

<title>Carta de Camilo Castelo Branco a José de Azevedo e Menezes a 14 de outubro de 1878: uma versão eletrónica**</title>**

O título atribuído ao documento eletrónico contém a informação da identidade do remetente, do destinatário e da data da carta. Dado que é necessário distinguir o

documento eletrônico do documento original, o título do documento eletrônico deverá explicitar que se trata da versão eletrônica do original.

Nem todos os títulos atribuídos a cada cabeçalho do documento eletrônico possuem todas estas informações, uma vez que em algumas cartas não existe informação da data, ou parte dela. Assim, de modo a resolver esta questão e a anotar esta informação no documento eletrônico, o título dos documentos que não apresentem quaisquer dados relativamente à data contém a seguinte informação:

<title>Carta de António Alves Mendes da Silva Ribeiro a Camilo Castelo Branco - sem data: uma versão eletrônica**</title>** (Exemplo da Carta 53 do *Acervo da Correspondência de Camilo Castelo Branco*.)

As cartas que possuem apenas dados relativos ao dia e mês de redação têm o seguinte título no seu documento eletrônico:

<title>Carta de Alberto Pimentel a Camilo Castelo Branco - 28 - sem mês . sem ano: uma versão eletrônica**</title>** (Exemplo da Carta 60 do *Acervo da Correspondência de Camilo Castelo Branco*, na qual existe apenas a indicação do dia.)

Nos casos onde existe informação do dia e mês da redação da carta, mas não há informação acerca do ano, o título da edição eletrônica possui a seguinte informação:

<title>Carta de Conde de Aljezur a Camilo Castelo Branco a 20 de dezembro - sem ano: uma versão eletrônica**</title>** (Exemplo da Carta 36 do *Acervo da Correspondência de Camilo Castelo Branco*.)

O elemento **<author>** indica o nome do autor da carta, ou seja, do signatário. Para especificar o nome em questão, foi acrescentada outra anotação com o elemento **<name>** associado ao atributo *type*, especificando assim, o tipo de nome que está a ser anotado, neste caso enunciando que o mesmo se trata de uma pessoa. Esta anotação tem a seguinte estrutura final:

```
<author>
  <name type="person">Camilo Castelo Branco</name>
</author>
```

Para identificar a entidade patrocinadora da elaboração das edições eletrônicas, foi utilizado o elemento `<sponsor>`. De modo a especificar, neste caso, as entidades em questão, foram acrescentados dados acerca das mesmas, anotando o elemento `<name>` juntamente com o atributo *type*, de modo a identificar o tipo de entidade de que se trata e o endereço postal das mesmas.

A anotação das entidades patrocinadoras deste projeto, no cabeçalho das edições eletrônicas, tem a seguinte estrutura:

```
<sponsor>
  <name type="municipality">Câmara Municipal de Vila
    Nova de Famalicão</name>
  <address>
    <addrLine>Praça Álvaro Marques</addrLine>
    <addrLine>4764-502 Vila Nova de
      Famalicão</addrLine>
  </address>
</sponsor>

<sponsor>
  <name type="institute">Universidade do Minho</name>
  <name type="institute">Instituto de Letras e
    Ciências Humanas</name>
  <address>
    <addrLine>Campus de Gualtar</addrLine>
    <addrLine>4710-057</addrLine>
  </address>
</sponsor>
```

A última anotação utilizada, pertencente ao elemento obrigatório <titleStmt> do <fileDesc>, foi o elemento <respStmt>. Este elemento anota a menção de responsabilidade, ou seja, indica a pessoa, ou as pessoas responsáveis pela criação e anotação do documento eletrónico a partir do original.

Responsável de uma publicação entende-se a pessoa ou pessoas singulares e/ou colectivas, que tenham contribuído para a realização de uma obra, quer intelectual quer artisticamente, na qualidade de autor do texto da obra, de ilustrador, de comentador, anotador, revisor, editor literário, etc. (Nobre de Gusmão, *et al.*, 1997).

Para além do elemento <respStmt>, foi incluído também o elemento <resp> que indica qual a responsabilidade desempenhada por determinada pessoa. Devido ao facto de ter sido feita a transcrição e anotação estrutural da carta pela mesma pessoa, existe a necessidade de criar dois blocos <respStmt>:

```
<respStmt>
  <resp>Transcrição da carta efetuada por</resp>
  <name type="person">Cátia Silva</name>
</respStmt>
```

```
<respStmt>
  <resp>Anotação estrutural da carta efetuada
  por</resp>
  <name type="person">Cátia Silva</name>
</respStmt>
```

O segundo elemento obrigatório do <fileDesc> é o <publicationStmt>. No âmbito deste projeto, foram utilizados os seguintes elementos pertencentes ao <publicationStmt>: <authority>, <idno> (*identifier*) e <availability>.

O elemento <authority> designa a pessoa ou entidade responsável pela distribuição eletrónica do documento. Neste caso do *Arquivo Digital da Correspondência de Camilo Castelo Branco*, são quatro as entidades responsáveis. Para a anotação diferenciada de todas elas, foram anotados os nomes das entidades, com o atributo *type* associado, e ainda foram anotados os endereços postais das

mesmas. Os blocos de elementos <authority> têm a seguinte estrutura no <teiHeader> das edições eletrônicas do arquivo digital da correspondência:

```
<authority>
```

```
  <name type="municipality">Câmara Municipal de Vila  
  Nova de Famalicão</name>
```

```
  <address>
```

```
    <addrLine>Praça Álvaro Marques</addrLine>
```

```
    <addrLine>4764-502</addrLine>
```

```
  </address>
```

```
</authority>
```

```
<authority>
```

```
  <name type="institute">Casa de Camilo - Centro de  
  Estudos Camilianos</name>
```

```
  <address>
```

```
    <addrLine>Avenida de São Miguel, 758</addrLine>
```

```
    <addrLine>4770-631 São Miguel de  
    Seide</addrLine>
```

```
  </address>
```

```
</authority>
```

```
<authority>
```

```
  <name type="institute">Universidade do Minho</name>
```

```
  <name type="institute">Instituto de Letras e  
  Ciências Humanas</name>
```

```
  <address>
```

```
    <addrLine>Campus de Gualtar</addrLine>
```

```
    <addrLine>4710-057</addrLine>
```

```
  </address>
```

```
</authority>
```

O elemento <idno> indica a referência ou o número que identifica o texto digital no *Arquivo Digital da Correspondência de Camilo Castelo Branco*. Neste caso, a

ordem numérica atribuída às cartas coincide com a ordem das mesmas na obra “*Camillo Homenageado: O Escriitor da Graça e da Belleza*” (Menezes, José de Azevedo *et al.*, 1920).

Assim sendo, esta anotação tem a seguinte forma:

```
<idno type="Arquivo Digital da Correspondência de  
Camilo Castelo Branco">Carta 1</idno>
```

O elemento <availability> diz respeito aos direitos de utilização associados ao texto eletrónico. Este elemento pode ser associado ao atributo *status*, e este pode ser dotado dos seguintes valores: *restricted*, *unknown* e *free*. No âmbito deste projeto foi aplicado o valor *restricted*, mencionando assim que o utilizador das edições eletrónicas, deverá respeitar as disposições legais em vigor relativas à propriedade literária, artística ou científica, mais propriamente as presentes no Código do Direito de Autor e dos Direitos Conexos. Uma vez que as edições eletrónicas criadas têm o propósito de servir interesses de investigação, sempre que os investigadores pretendam fazer uso dos documentos, deverão obter junto dos responsáveis legais a respetiva autorização, estes que estão enunciados no elemento <availability> na anotação delimitada por <p>.

A anotação deste elemento tem a seguinte forma:

```
<availability status="restricted">  
  <p>Copyright - Casa de Camilo - Centro de Estudos  
  Camilianos</p>  
</availability>
```

O terceiro elemento obrigatório do <fileDesc> trata-se do elemento <sourceDesc> (*Source Description*). Este elemento proposto pela TEI visa descrever as características bibliográficas do documento original. Como já foi referido, a TEI não propõe características específicas para a anotação de correspondência. Assim sendo, no que respeita à secção <sourceDesc> foram aplicadas as propostas do projeto DALF.

O DALF segue as propostas da TEI para a criação dos cabeçalhos dos seus documentos. No entanto, dada a falta de propostas para o tratamento de

correspondência por parte da TEI, o projeto DALF elaborou um esquema pormenorizado para anotar as propriedades do documento original.

Assim, no elemento da TEI <sourceDesc>, o DALF introduziu o elemento <letDesc> (*Letter Description*). Este elemento é obrigatório e agrupa todos os metadados específicos acerca de correspondência, como, por exemplo, a localização geográfica das cartas, os dados de catalogação, o resumo da carta, a descrição física, e a existência de envelope. Opcionalmente, pode ser anotada informação relativa ao conteúdo das cartas, a sua história, ou partes específicas de cada carta.

O elemento <letDesc> pode conter vários elementos obrigatórios e facultativos. No âmbito deste projeto foram utilizados os seguintes elementos: <letIdentifier> (*Letter Identifier*), <letHeading> (*Letter Heading*), <physDesc> (*Physical Description*), <envOcc/> (*Envelope Occurrence*), <letContents> (*Letter Contents*), <history> e <additional>.

O elemento <letIdentifier> é o primeiro elemento obrigatório do <letDesc>: contém informações sobre a localização de cada carta: país, instituição, repositório, número de identificação do documento e do acervo ao qual o mesmo pertence.

Neste projeto foram anotados os seguintes elementos pertencentes ao <letIdentifier>:

<country>

Elemento TEI utilizado para indicar o país onde o documento está localizado;

<region>

Elemento TEI utilizado para indicar a região onde o documento está localizado;

<settlement>

Elemento pertencente à TEI utilizado para mencionar a localidade onde o documento se encontra localizado;

<institution>

Contém o nome da instituição na qual está localizado o repositório do documento original;

<repository>

Este elemento é utilizado para indicar o nome do repositório ao qual o documento pertence;

<collection>

Contém a designação da coleção da qual a carta é parte integrante;

<idno>

Elemento TEI que fornece o número de identificação utilizado para identificar o documento na instituição, repositório e/ou coleção.

A estrutura final que compõe o elemento <letIdentifier> tem a seguinte forma:

<letIdentifier>

<country>Portugal**</country>**

<region>Norte**</region>**

<settlement>Vila Nova de Famalicão - São Miguel de Seide**</settlement>**

<institution>Casa de Camilo - Centro de Estudos Camilianos**</institution>**

<repository>Arquivo da Correspondência de Camilo Castelo Branco**</repository>**

<collection>Arquivo da Correspondência a Camilo Castelo Branco: "Camillo Homenageado"**</collection>**

<idno type="Arquivo da Correspondência a Camilo Castelo Branco-Camillo Homenageado">Carta 1**</idno>**

</letIdentifier>

O segundo elemento obrigatório do <letDesc> é o elemento <letHeading> que contém informações essenciais acerca da identificação bibliográfica de uma carta. Este elemento identifica os participantes da carta: o signatário; o destinatário; o local e a data de redação.

Neste projeto, o `<letHeading>` é composto pelos seguintes elementos:

`<author>`

Identifica o autor da carta - signatário.

`<addressee>`

Identifica a pessoa a quem a carta é endereçada - destinatário.

`<placeLet>`

Identifica o local onde a carta foi escrita.

`<dateLet>`

Identifica a data de redação da carta.

A cada um destes elementos foi associado o atributo *attested* que fornece informação sobre o grau de certeza relativamente aos dados relacionados com cada elemento acima descrito. Este atributo permite especificar se a informação relativamente ao signatário, destinatário, data e local de redação está explícita na carta (valor *yes*) ou se a carta não fornece essa informação (valor *no*).

O `<letHeading>` apresenta a seguinte estrutura:

`<letHeading>`

`<author attested="yes">`Camilo Castelo

Branco**`</author>`**

`<addressee attested="yes">`José de Azevedo e

Menezes**`</addressee>`**

`<placeLet attested="no">`desconhecido**`</placeLet>`**

`<dateLet attested="yes">`14.10.1878**`</dateLet>`**

`</letHeading>`

Neste projeto, as cartas trabalhadas não ofereceram dificuldades no que respeita aos signatários e destinatários. Tal situação não acontece com os restantes dois elementos. Por vezes, não existe informação acerca do local de redação `<placeLet>`. Neste caso, optou-se pela atribuição do valor *no*, ao atributo *attested*, indicando o local como desconhecido. Esta anotação permitirá ao anotador recuperar facilmente as informações que necessitam ser confirmadas ou certificadas. O elemento `<dateLet>`

apresenta o mesmo problema, pois muitas vezes não existe informação acerca da data de redação da carta.

O elemento `<physDesc>` (*Physical Description*) é o terceiro elemento obrigatório da secção `<letDesc>`. Esta secção permite ao anotador descrever propriedades físicas da carta: tipo de correspondência (carta ou bilhete), disposição da informação na carta, elementos decorativos, parafernália, etc.

Neste projeto, foram anotados os seguintes elementos pertencentes ao `<physDesc>`:

`<type>`

Identifica o documento epistolográfico como sendo do tipo "carta", "bilhete", "telegrama", etc.

`<extent>`

Elemento pertencente à TEI que especifica o tamanho da carta.

O elemento `<extent>` é composto pelo elemento `<dimensions>`, que, por sua vez, é composto pelos seguintes elementos:

`<height units="mm">`

Especifica a altura da carta, utilizando milímetros como unidade de medida.

`<width units="mm">`

Especifica a largura da carta, utilizando milímetros como unidade de medida.

No âmbito deste projeto não foi possível, devido a restrições no que respeita ao manuseamento dos manuscritos, a recolha da informação relativa às dimensões de cada uma das cartas transcritas e anotadas.

Assim, o <physDesc> tem a seguinte estrutura:

```
<physDesc>
  <type>Carta</type>
  <extent>
    <dimensions>
      <height units="mm">yy</height>
      <width units="mm">xx</width>
    </dimension>
  </extent>
</physDesc>
```

O quarto elemento obrigatório pertencente à secção <letDesc>, o elemento <envOcc/>, que associado ao atributo "occ" especifica a presença (valor *yes*) ou a ausência (valor *no*) de um envelope. Este elemento <envOcc/> é um elemento vazio.

As cartas trabalhadas neste projeto não dispõem de envelopes, portanto, a anotação da estrutura final deste elemento é a seguinte:

```
<envOcc occ="no"/>
```

O elemento <letContents> é o primeiro elemento opcional que pode aparecer associado à secção <letDesc>. Este elemento contém o resumo do conteúdo da carta, e pode ser extremamente útil para quem consulta a edição eletrónica, uma vez que possibilita um rápido acesso ao(s) assunto(s) abordado(s) em cada um dos documentos. O resumo das cartas é anotado através da utilização do elemento <p>, delimitando assim o mesmo por parágrafos. Os resumos presentes nas edições eletrónicas foram retirados da obra utilizada como base de apoio ao projeto: *"Camillo Homenageado: O Escriptor da Graça e da Belleza"* (Menezes, José de Azevedo *et al.*, 1920). A identificação da fonte a partir da qual o resumo foi transcrito é realizada através do elemento <note>.

O elemento <letContents> possui a seguinte estrutura:

<letContents>

<p>Carta de agradecimento ao destinatário, José de Azevedo e Menezes, por pedir notícias ao fecundo romancista do seu estado de saúde, após o descarrilamento do comboio, perto de Ermezinde, e no qual seguia para o Porto. *Queixa-se mais dos nervos que das feridas.***</p>**

<note>Fonte: Menezes, José de Azevedo *et al.* (1920)
"Camillo Homenageado: O Escriptor da Graça e da Belleza.", Famalicão: Minerva**</note>**

</letContents>

O segundo elemento opcional da estrutura hierárquica do elemento <letDesc> é o elemento <history>, que identifica aspetos específicos da história de uma carta. A descrição histórica de uma carta abrange detalhes de origem da mesma, da sua história antes da aquisição por parte de determinada instituição, e do processo de aquisição.

Neste projeto, o elemento <history> é composto por uma primeira parte relativa ao historial da carta, parte essa delimitada por parágrafos. A segunda parte desta secção, <note> contém referências bibliográficas relativas à informação fornecida na primeira parte da secção.

O elemento <history> apenas foi inserido nos documentos eletrónicos das primeiras 15 cartas, que correspondem às únicas cartas dos 950 documentos do Acervo em que Camilo aparece como signatário. Todas as 15 cartas têm como destinatário José de Azevedo e Menezes. Foi possível incluir o historial destes 15 documentos porque esta informação encontra-se descrita na obra "*Camillo Homenageado: O Escriptor da Graça e da Belleza*" (Menezes, José de Azevedo *et al.*, 1920). Não foi possível obter informação histórica relativamente às restantes cartas transcritas.

Assim, o elemento <history> apresenta a seguinte forma:

<history>

<p>Carta oferecida pelo destinatário José de Azevedo e Menezes ao Museu - Camilo**</p>**

<note>Fonte: Menezes, José de Azevedo *et al.* (1920)
"Camillo Homenageado: O Escriitor da Graça e da Belleza" Famalicão: Minerva**</note>**

</history>

O último elemento pertencente ao elemento <letDesc>, é o elemento opcional <additional>. Esta secção fornece ao utilizador da edição eletrónica informações sobre aspetos, tais como: outras representações formais do documento, detalhes bibliográficos sobre a publicação do mesmo.

Neste projeto, o elemento <additional> é composto pelos seguintes elementos:

<surrogates>

Este elemento contém informação acerca de outras representações formais existentes da carta;

<listBibl>

Este elemento pertence à TEI; contém a descrição de listas bibliográficas relativas à publicação de documentos.

O elemento <surrogates> identifica as várias representações formais da carta que possam existir. Neste projeto, as transcrições foram efetuadas com base nas cópias das cartas.

O elemento <listBibl> contém referências bibliográficas relativamente a publicações relacionadas com a carta. Este elemento é composto pelos seguintes subelementos: <bibl>; <author>; <title>; <pubPlace>; <publisher>; <date> e <idno>.

<bibl>

Elemento que abarca os restantes elementos que compõem o <listBibl>.

<author>

Indica o nome do autor da publicação.

<title>

Identifica o título da publicação.

<pubPlace>

Identifica o local da publicação.

<publisher>

Indica a editora da publicação.

<date>

Identifica a data da publicação.

<idno>

Indica o número de identificação da publicação.

O elemento <additional> tem a seguinte estrutura:

<additional>

<surrogates>

<p>Cópia da carta</p>

</surrogates>

<listBibl>

<bibl>Esta carta encontra-se publicada em

<author>Camilo Castelo Branco**</author>**

<title>Camilo - Obras Completas**</title>**

<pubPlace>Porto**</pubPlace>**

<publisher>Lello & Irmão - Editores**</publisher>**

<date>2002**</date>**

<idno type="unknown">sem informação**</idno>**

</bibl>

</listBibl>

</additional>

O esquema completo do <teiHeader> da Carta 1, documento a integrar no *Arquivo Digital da Correspondência de Camilo Castelo Branco*, encontra-se disponível para consulta no Anexo 1.

CAPÍTULO 5:
CONCEITO E PROCESSO DE VALIDAÇÃO DE DTD

5.1. CONCEITO DE VALIDAÇÃO

O conceito de validação assume um papel fundamental no âmbito da criação do *Arquivo Digital da Correspondência de Camilo Castelo Branco* pelas seguintes razões:

- (i) Estamos perante um número elevado de documentos eletrónicos.
- (ii) O projeto conta com a colaboração de vários investigadores, sendo necessário definir um esquema de anotação para a estrutura e conteúdo das cartas, esquema a seguir por todos os colaboradores.

A validação dos documentos XML encontra-se interligada com a definição da estrutura hierárquica do documento a anotar, estrutura essa denominada por DTD (*Document Type Definition*). Assim, com o documento DTD é possível verificar se a estrutura do documento anotado (documento XML) está conforme o esquema definido no DTD.

Um documento DTD (i) define e permite validar a estrutura de uma classe ou tipo de documentos; (ii) descreve as anotações que serão utilizadas para anotar cada um dos elementos que formam os documentos de determinado tipo; (iii) identifica quais os atributos ligados a cada elemento; caracteriza os valores de cada atributo, indicando se o mesmo é facultativo ou obrigatório; (iv) define a estrutura do conteúdo de cada elemento.

Com a criação de um documento DTD é possível validar os documentos XML anteriormente anotados. Este processo é efetuado através de um validador que liga e compara as estruturas do documento XML e DTD, ou seja, verifica se a estrutura do documento XML constituído pelo <teiHeader> e pelo <text>, obedece à estrutura das regras formais definidas no documento DTD.

5.2. PROCESSO DE VALIDAÇÃO

O processo de validação é fundamental no âmbito da criação do *Arquivo Digital da Correspondência de Camilo Castelo Branco*, uma vez que se trata de um número elevado de ficheiros e os documentos constituintes do arquivo pertencem ao mesmo género textual, ou seja, com a criação de um único documento DTD é possível

validar todas as cartas do arquivo. A utilização e validação do mesmo esquema de anotação, em todos os documentos, facilita a conversão dos mesmos para outros formatos.

Existem três tipos de documentos DTD: (i) DTD interno - documento que se encontra dentro do documento XML; (ii) DTD externo - documento que se encontra externo ao XML; (iii) DTD público - documento que se encontra disponível para o público em geral, *on-line*.

No âmbito deste projeto, foi criado e utilizado o documento DTD externo, uma vez que todas as cartas foram validados pelo mesmo documento.

A validação de documentos implica a associação do documento DTD ao XML através da declaração DOCTYPE (*DOCTYPE declaration*), que se encontra no interior do documento XML, com a seguinte estrutura:

```
<!DOCTYPE TEI.2 SYSTEM "DTD.dtd">
```

A declaração inicia-se com a palavra-chave DOCTYPE, segue-se o elemento raiz do documento XML (TEI.2), que é o primeiro elemento hierarquicamente presente no documento XML. A palavra que se segue é SYSTEM, indicando que o documento DTD se encontra no sistema eletrónico do utilizador. Por fim, a informação "DTD.dtd" refere-se ao nome atribuído ao ficheiro DTD.

O processo de validação das 107 cartas trabalhadas no âmbito deste projeto foi efetuado com o programa *Oxygen XML Editor*²³. O processo de validação encontra-se exemplificado nas figuras 20 e 21 que se seguem.

Na figura 20, temos uma captura de ecrã do programa *Oxygen XML Editor* aquando do momento da validação do documento DTD criado para as cartas deste projeto. Como podemos verificar, através da marca (i), o programa possibilita a abertura, em simultâneo, do documento XML (Carta 1.xml) e do documento DTD (DTD.dtd). Para consultar cada uma das páginas dos documentos é apenas necessário selecionar o separador correspondente a cada uma.

A marca (ii) demonstra que a validação do DTD falhou, tendo ocorrido um erro. A marca (iii) explicita qual o erro que foi encontrado e que não permite a

²³ Página oficial do *Oxygen XML Editor*: <http://www.oxygenxml.com>.

validação do DTD. A marca (iv) fornece informação relativamente ao local onde ocorreu o erro. Na marca (v), obtemos a informação de que o erro ocorre no DTD.

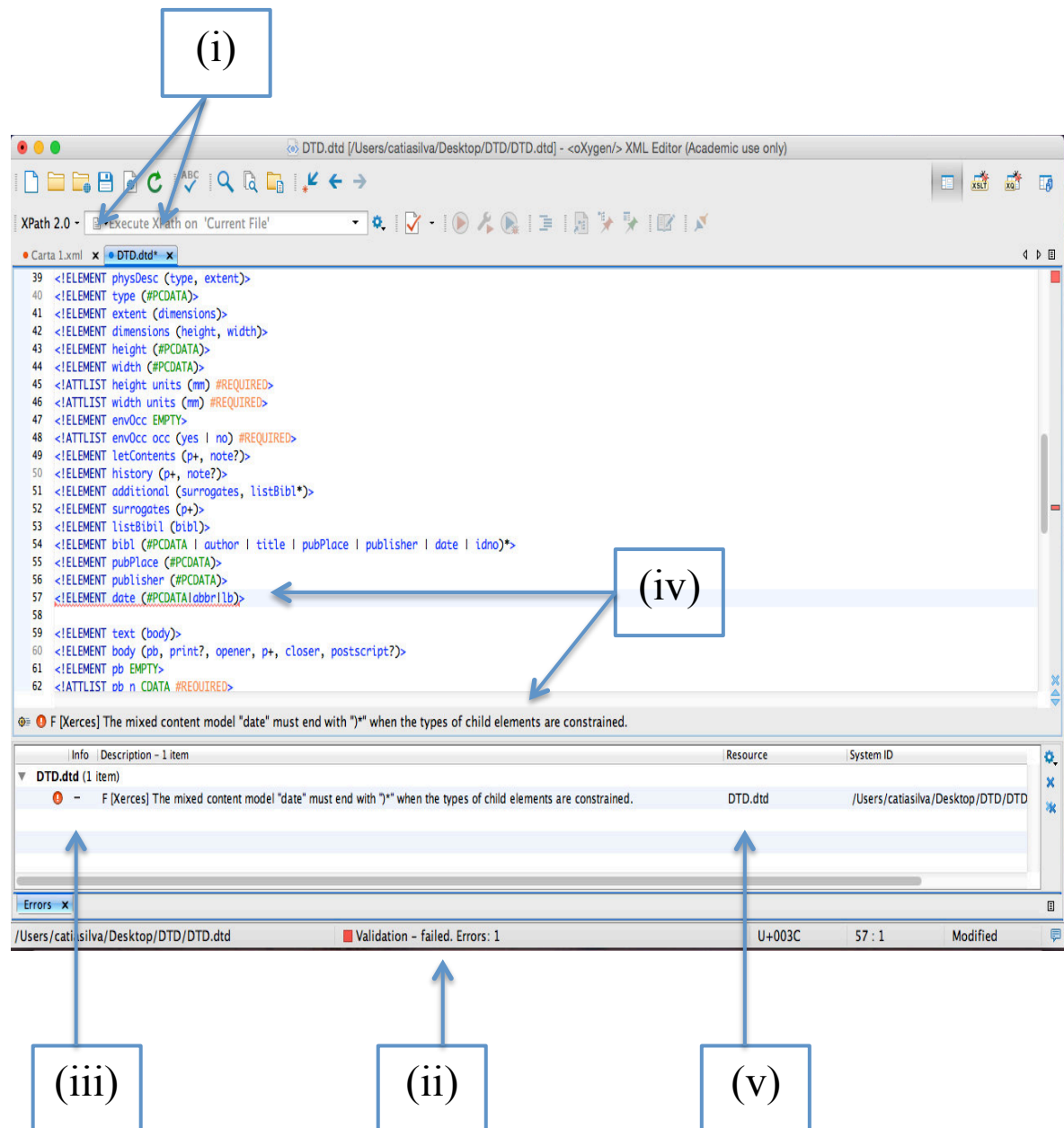


Figura 20 - Exemplo de documento DTD mal formado

Na Figura 21, estamos perante uma captura de ecrã do programa *Oxygen XML Editor* que apresenta um documento DTD válido.

A marca (i) indica, tal como na Figura 20, os ficheiros XML e DTD que estão a ser utilizados no programa. A marca (ii) mostra qual a mensagem que o programa apresenta quando o documento que está a ser validado é válido.

A marca (iii) da Figura 21, remete-nos para o erro identificado pelo validador na Figura 20. Dado que o conteúdo do elemento <date> é de tipo "misto", ou seja, o elemento <date> contém texto intercalado com elementos, a declaração DTD deverá apresentar um asterisco (*) no final. Este operador significa que os elementos intercalados com o texto, neste caso os elementos <abbr> (*abbreviation*) e <lb/> (*line break*), podem não aparecer no elemento <date>, como poderão aparecer uma ou mais vezes em qualquer posição do elemento.

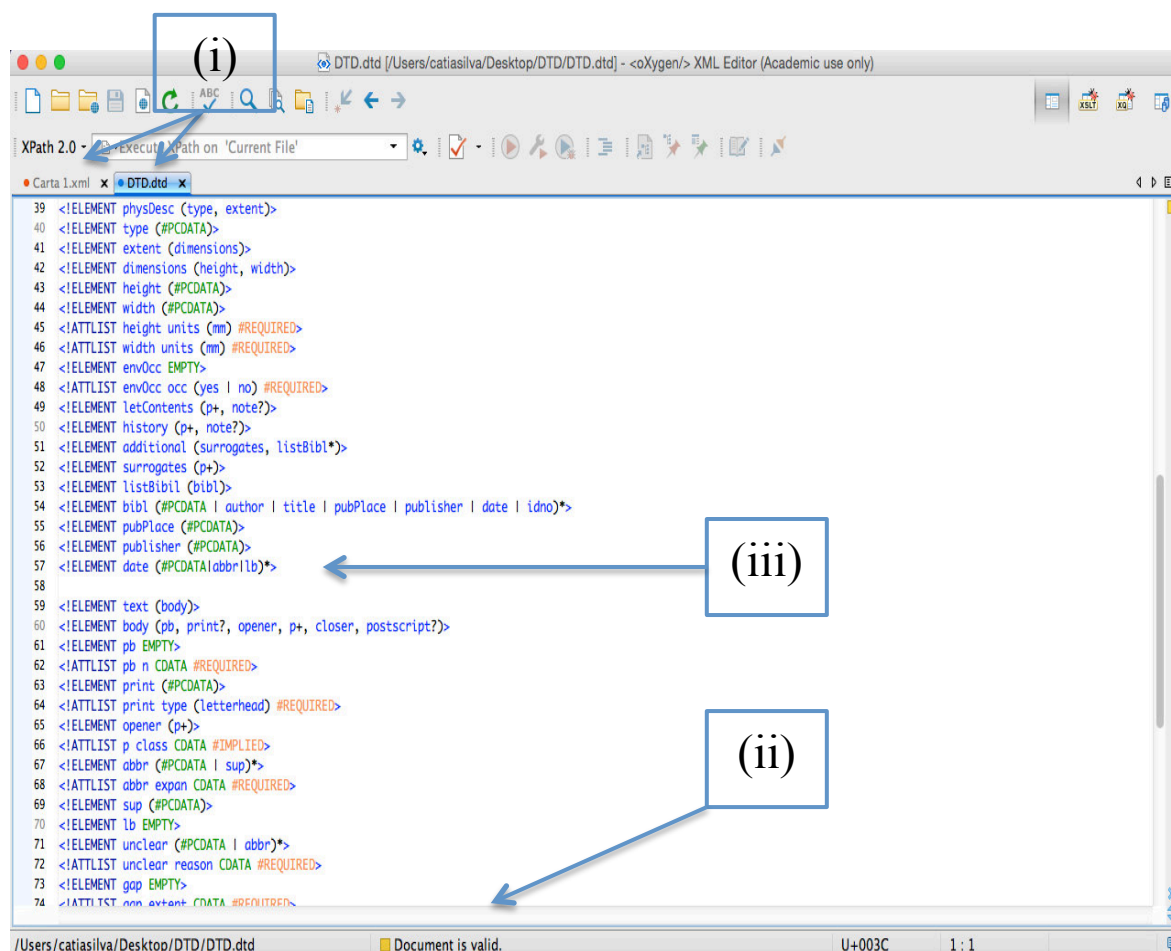


Figura 21 Exemplo de documento DTD válido

Todos os 107 documentos XML foram validados com o documento DTD que se encontra no Anexo 3 deste relatório.

CAPÍTULO 6:
CONSIDERAÇÕES FINAIS

CONSIDERAÇÕES FINAIS

Neste projeto, foi efetuada a transcrição da correspondência dirigida a Camilo Castelo Branco entre 1864 e 1890, com o intuito de criar uma edição eletrónica da mesma. Foram transcritas 107 cartas de um universo de 950 cartas e criada uma edição eletrónica anotada para cada uma delas, utilizando a metalinguagem XML (*eXtensible Markup Language*), seguindo as diretrizes de anotação da TEI (*Text Encoding Initiative*) e do projeto DALF (*Digital Archive of Letters in Flanders*). Os princípios de edição definidos e adotados no âmbito deste projeto servirão de orientação para a transcrição e anotação das restantes cartas do *Acervo da Correspondência de Camilo Castelo Branco*, acervo reunido pela Comissão de Homenagem ao escritor entre 1915 e 1920. As edições eletrónicas das cartas criadas neste projeto, juntamente com o trabalho desenvolvido no projeto "*Edição digital anotada da correspondência dirigida a Camilo Castelo Branco: problemas e desafios*" pela colega Cristiana Leal, serão posteriormente integradas no *Arquivo Digital da Correspondência de Camilo Castelo Branco*.

Este projeto deu-me a oportunidade de conhecer o trabalho de bastidores envolvido na criação de uma edição eletrónica. Muitas foram as dificuldades encontradas: a dificuldade de leitura dos manuscritos; dúvidas relativamente à anotação de dificuldades derivadas do ato de transcrição; dúvidas em termos de anotação da estrutura das cartas e de propriedades específicas dos textos; entre outros. As dificuldades transformaram-se em desafios: conhecer e problematizar as diretrizes da TEI ao pormenor e aplicar as anotações propostas ao projeto; conhecer e analisar as propostas do projeto DALF para a anotação de documentos epistolográficos e aplicar as anotações propostas ao projeto; criar um esquema de metadados (*TEI Header*) que servisse de orientação para todos os documentos eletrónicos; criar um documento DTD que validasse as 107 cartas transcritas e anotadas; entre outros.

Foi uma honra e um prazer ter desenvolvido este projeto, uma vez que o mesmo contribui para a preservação do espólio de um dos maiores nomes da literatura portuguesa.

REFERÊNCIAS BIBLIOGRÁFICAS E CIBERGRÁFICAS

REFERÊNCIAS BIBLIOGRÁFICAS

Dicionários

Cabral, Alexandre. (1988). *“Dicionário de Camilo Castelo Branco”* (2ªed.). Lisboa: Caminho.

Manuscritos

Acervo da Correspondência de Camilo Castelo Branco: cartas 1 a 107. In Casa Museu de Camilo - Centro de Estudos Camilianos, Seide, Vila Nova de Famalicão.

Bibliografia secundária

Burnard, Lou. (2005). "Encoding Standards for the Electronic Edition". In Matija Ogrin (ed.), *Scholarly Editions and the Digital Medium*. Ljubljana: Studia Litteraria ZRC ZAZU, pp. 12-67. Disponível em: <http://nl.ijs.si/e-zrc/bib/eziss-Burnard.pdf>

Burnard, Lou, Katherine O'Brien O'Keeffe & John Unsworth. (eds.) (2006). *Electronic Textual Editing*. New York: Modern Language Association of America.

Buzzetti, Dino & Jerome McGann. (2006). "Critical Editing in a Digital Horizon". In Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth (eds.) *Electronic Textual Editing*. New York: Modern Language Association of America. Disponível em: http://www.tei-c.org/About/Archive_new/ETE/Preview/mcgann.xml

Committee on Scholarly Editions. (2006). "Guidelines for Editors of Scholarly Editions". In Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth (eds.), *Electronic Textual Editing*. New York: Modern Language Association of America. Disponível em: http://www.tei-c.org/About/Archive_new/ETE/Preview/guidelines.xml

Committee on Scholarly Editions. (2006). "Guiding Questions for Vettors of Print and Electronic Editions". In Lou Burnard, Katherine O'Brien O'Keefe & John Unsworth (eds.), *Electronic Textual Editing*. New York: Modern Language Association of America.

Disponível em: http://www.tei-c.org/About/Archive_new/ETE/Preview/hirst.xml

Cummings, James. (2008). "The Text Encoding Initiative and the Study of Literature". In Susan Schreibman & Ray Siemens (eds.), *A Companion to Digital Literary Studies*. Oxford: Blackwell, pp. 451-476.

Disponível em: <http://www.digitalhumanities.org/companionDLS/>

Driscoll, M. J. (2006). "Levels of transcription". In Lou Burnard, Katherine O'Brien O'Keefe & John Unsworth (eds.), *Electronic Textual Editing*. New York: Modern Language Association of America, pp.254-261. Disponível em: http://www.tei-c.org/About/Archive_new/ETE/Preview/driscoll.xml

Erway, Ricky. (2010). *Defining "Born Digital": An Essay*. Disponível em: <http://www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf>.

Grésillon, A. (1994). *Éléments de critique génétique: lire les manuscrits modernes*. Paris: Presses Universitaires de France.

Gusmão, A. N., Fernanda Maria Guedes de Campos & José Carlos Garcia Sottomayor (eds.). (1997). *Regras Portuguesas de Catalogação*. Lisboa: Biblioteca Nacional.

Hockey, Susan. (2000). *Electronic Texts in the Humanities: Principles and Practice*. Oxford: Oxford University Press.

McGann, Jerome. (2004). "Marking Texts of Many Dimensions". In Susan Schreibman, Ray Siemens & John Unsworth, (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 198-217.

Disponível em: <http://www.digitalhumanities.org/companion/>

Menezes, José de Azevedo *et al.* (1920). "*Camilo Homenageado: O Escriptor da Graça e da Belleza*". Famalicão: Minerva.

Pereira, António Maria. (1973). "*Cartas de Camilo aos Editores*". Lisboa: Parceria A. M. Pereira.

Pichler, A. (1995). "Transcriptions, Texts, and Interpretations". In K. Johannessen & T. Nordenstam (eds.), *Culture and Value: Philosophy and the Cultural Sciences*. Vienna: Austrian Ludwig Wittgenstein Society, pp.690-695.

Portela, Manuel (2013). "Nenhum Problema Tem Solução: Um Arquivo Digital do "Livro do Desassossego"". *MatLit*, 1 (1), pp. 9-33.
Disponível em: <http://iduc.uc.pt/index.php/matlit/article/view/1618/1012>

Price, Kenneth M. (2008). "Electronic Scholarly Editions". In Susan Schreibman & Ray Siemens, (eds.), *A Companion to Digital Literary Studies*. Oxford: Blackwell, pp. 434-450.
Disponível em: <http://www.digitalhumanities.org/companionDLS/>

Renear, Allen H. (2004). "Text Encoding". In Susan Schreibman, Ray Siemens & John Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 218-239.
Disponível em: <http://www.digitalhumanities.org/companion/>

Schmidt, Desmond (2014). "Towards an Interoperable Digital Scholarly Edition". *Journal of the Text Encoding Initiative*, Issue 7.
Disponível em: <http://jtei.revues.org/979>

Schreibman, Susan & Ray Siemens (eds.). (2008). *A Companion to Digital Literary Studies*. Oxford: Blackwell.

Schreibman, Susan, Ray Siemens & John Unsworth (eds.). (2004). *A Companion to Digital Humanities*. Oxford: Blackwell.

Silva, António de Moraes. (1949). *Grande Dicionário da Língua Portuguesa* (10ª ed.). Lisboa: Confluência.

Smith, Abby. (2004). "Preservation". In Susan Schreibman, Ray Siemens & John Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. Disponível em: <http://www.digitalhumanities.org/companion/>

Smith, Martha N. (2004). "Electronic Scholarly Editing". In Susan Schreibman, Ray Siemens & John Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell, pp. 306-322. Disponível em: <http://www.digitalhumanities.org/companion/>

Sperberg-McQueen, C. M. & Burnard L. (1995). "The Design of the TEI Encoding Scheme". *Computers and the Humanities*, 29 (1), pp. 17–39.

Vanhoutte, Edward (2004). "Presentational and Representational Issues in Correspondence Reconstruction and Sorting". *Literary and Linguistic Computing*, Vol. 19 (1), pp. 45-54.

Vanhoutte, Edward. (2006). "Prose Fiction and Modern Manuscripts: Limitations and Possibilities of Text-Encoding for Electronic Editions". In Lou Burnard, Katherine O'Brien O'Keeffe & John Unsworth (eds.) *Electronic Textual Editing*. New York: Modern Language Association of America. Disponível em: http://www.tei-c.org/About/Archive_new/ETE/Preview/vanhoutte.xml

Vanhoutte, Edward & Ron Van den Branden (eds.) (2003). *DALF Guidelines for the Description and Encoding of Modern Manuscript Material, Version 1.0.* Gent: Centrum voor Teksteditie en Bronnenstudie. Disponível em: <http://ctb.kantl.be/project/dalf/dalfdoc/index.html>.

Vanhoutte, Edward, & Ron Van den Branden (2009). "Describing, Transcribing, Encoding, and Editing Modern Correspondence Material: A Textbase Approach". *Literary and Linguistic Computing*, Vol. 24 (1), pp.77-98.

REFERÊNCIAS CIBERGRÁFICAS

Biografias y Vidas

http://www.biografiasyvidas.com/biografia/n/nunez_gaspar.html

Centro de Estudos Camilianos

<http://www.camilocastelobranco.org/>

Digital Archive of Letters in Flanders

<http://ctb.kantl.be/project/dalf/dalfdoc/index.html>

Text Encoding Initiative:

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

Vincent van Gogh: *The Letters*

<http://www.vangoghletters.org/vg/>

World Wide Web Consortium (W3C)

<http://www.w3.org/>

ANEXOS

ANEXO 1

Cabeçalho <teiHeader> bem formado da edição eletrónica da Carta 1 do *Acervo da Correspondência de Camilo Castelo Branco*.

```
<teiHeader>
<fileDesc>
<titleStmt>
<title>Carta de Camilo Castelo Branco a José de Azevedo e
Menezes a 14 de outubro de 1878: uma versão
eletrónica</title>
<author><name type="person">Camilo Castelo
Branco</name></author>
```

```
<sponsor>
<name type="municipality">Câmara Municipal de Vila Nova
de Famalicão</name>
<address>
<addrLine>Praça Álvaro Marques</addrLine>
<addrLine>4764-502 Vila Nova de Famalicão</addrLine>
</address>
</sponsor>
<sponsor>
<name type="institute">Universidade do Minho</name>
<name type="institute">Instituto de Letras e Ciências
Humanas</name>
<address>
<addrLine>Campus de Gualtar</addrLine>
<addrLine>4710-057 Braga</addrLine>
</address>
</sponsor>
```

```
<respStmt>
<resp>Transcrição da carta efetuada por</resp>
```

```

<name type="person">Cátia Silva</name>
</respStmt>
<respStmt>
<resp>Anotação estrutural da carta efetuada por</resp>
<name type="person">Cátia Silva</name>
</respStmt>
</titleStmt>

<publicationStmt>
<authority>
<name type="municipality">Câmara Municipal de Vila Nova
de Famalicão</name>
<address>
<addrLine>Praça Álvaro Marques</addrLine>
<addrLine>4764-502 Vila Nova de Famalicão</addrLine>
</address>
</authority>
<authority>
<name type="institute">Casa de Camilo - Centro de Estudos
Camilianos</name>
<address>
<addrLine>Avenida de São Miguel, 758</addrLine>
<addrLine>4770-631 São Miguel de Seide</addrLine>
</address>
</authority>
<authority>
<name type="institute">Universidade do Minho</name>
<name type="institute">Instituto de Letras e Ciências
Humanas</name>
<address>
<addrLine>Campus de Gualtar</addrLine>
<addrLine>4710-057 Braga</addrLine>
</address>
</authority>

```

<idno type="Arquivo Digital da Correspondência de Camilo Castelo Branco">Carta 1</idno>

<availability status="restricted">

<p>Copyright - Casa de Camilo - Centro de Estudos Camilianos</p>

</availability>

</publicationStmt>

<sourceDesc>

<letDesc>

<letIdentifier>

<country>Portugal</country>

<region>Norte</region>

<settlement>Vila Nova de Famalicão - São Miguel de Seide</settlement>

<institution>Casa de Camilo - Centro de Estudos Camilianos</institution>

<repository>Arquivo da Correspondência de Camilo Castelo Branco</repository>

<collection>Arquivo da Correspondência a Camilo Castelo Branco: "Camillo Homenageado"</collection>

<idno type="Arquivo da Correspondência a Camilo Castelo Branco-"Camillo Homenageado"">Carta 1</idno>

</letIdentifier>

<letHeading>

<author attested="yes">Camilo Castelo Branco</author>

<addressee attested="yes">José de Azevedo e Menezes</addressee>

<placeLet attested="yes">São Miguel de Seide </placeLet>

<dateLet attested="yes">14.10.1878</dateLet>

</letHeading>

```

<physDesc>
<type>Carta</type>
<extent>
<dimensions>
<height units="mm">yy</height>
<width units="mm">xx</width>
</dimensions>
</extent>
</physDesc>

<envOcc occ="no"/>

<letContents>
<p>Carta de agradecimento ao destinatário, José de
Azevedo e Menezes, por pedir notícias ao fecundo
romancista do seu estado de saúde, após o descarrilamento
do comboio, perto de Ermezinde, e no qual seguia para o
Porto. Queixa-se mais dos nervos que das feridas.</p>
<note>Fonte: Menezes, José de Azevedo et al. (1920)
"Camillo Homenageado: O Escriptor da Graça e da Belleza".
Famalicão: Minerva</note>
</letContents>

<history>
<p>Carta oferecida pelo destinatário José de Azevedo e
Menezes ao Museu-Camilo</p>
<note>Fonte: Menezes, José de Azevedo et al.(1920)
"Camillo Homenageado: O Escriptor da Graça e da Belleza".
Famalicão: Minerva</note>
</history>

<additional>
<surrogates>
<p>Cópia da carta</p>

```



```
</surrogates>
<listBibl>
<bibl>Esta carta encontra-se publicada em
<author>Camilo Castelo Branco</author>
<title>Camilo - Obras Completas</title>
<pubPlace>Porto</pubPlace>
<publisher>Lello & Irmão - Editores</publisher>
<date>2002</date>
<idno type="unknown">xx</idno>
</bibl>
</listBibl>
</additional>
</letDesc>
</sourceDesc>
</fileDesc>
</teiHeader>
```


ANEXO 2

Estrutura da secção <text> da Carta 80 do signatário José de Azevedo e Menezes a Camilo Castelo Branco.

<text>

<body>

<pb n="1"/>

<opener><p class="salute"><abbr expan="Excellentissimo Senhor">Ex.^{mo} Snr.</abbr> e meu presado <abbr expan="Amigo">Am^o</abbr></p></opener>

<p>Pelo que vejo nos meus livros<lb/>
os Eças procedem de <abbr expan="Dom">D.</abbr> Fer-<lb/>
nando, filho do infante <abbr expan="Dom">D.</abbr><lb/>
João e de sua <abbr expan="primeira">1^a</abbr>
mulher, a in-<lb/>
feliz <abbr expan="Dona">D.</abbr> Maria Telles. O
infante<lb/>
foi senhor da <hi rend="underline">Villa de Eça,<lb/>
que é no extremo d´Aragão</hi>,<lb/>
<abbr expan="por">p.^r</abbr> doação do seu
parente, <abbr expan="Dom">D.</abbr> Fradi-<lb/>
que deCastro, Duque d´Arjona,<lb/>
e filho de <abbr expan="Dom">D.</abbr> Pedro
Henriques,<lb/>
Conde de Trastamasa, e de sua<lb/>
mulher, <abbr expan="Dona">D.</abbr> Izabel deCastro,
<abbr expan="sobrinha">sobr.^a</abbr><lb/>
da rainha <abbr expan="Dona">D.</abbr> Ignez, mae do di-<lb/>

to infante <abbr expan="Dom">D.</abbr> João, <hi rend="underline">que deu<lb/>
aquella Villa a seu <abbr expan="filho Dom">f.^o D.</abbr> Fernando.</hi><lb/>
Este <abbr expan="Dom">D.</abbr> Fernando <hi rend="underline">D´Eça</hi>, - <hi rend="underline">o
<abbr expan="primeiro">1^o</abbr> de<lb/>
nome e mourisco</hi>, - <hi rend="underline">foi de
tão<lb/>
larga consciencia que casou<lb/>
com seis mulheres, sendo todas<pb n="2"/>
vivas, e entre ligítimos e bastar<lb/>
dos teve 42 filhos!</hi></p>
<p>Depois o homem fez vida san-<lb/>
ta, vestindo o habito de <abbr expan="Sao">S.</abbr>
Fran-<lb/>
cisco; e d´ahi vem o cordão que<lb/>
trazem nas suas armas os d´es-<lb/>
ta familia.</p>

<p><abbr expan="Dona">D.</abbr> Catharina d´Eça,
abbadessa do Lor-<lb/>
vão, trasladou para a egreja do Con-<lb/>
vento do <abbr expan="Espirito">Esp.</abbr> Santo de
Gouveia, da<lb/>
ordem de <abbr expan="Sao">S.</abbr> Francisco, os ossos
de<lb/>
seu pae, e pôz na sepultura este e-<lb/>
pitaphio: <quote>"aqui jaz <abbr expan="Dom">D.</abbr>
Fernando<lb/>
d´Eça, filho do <abbr expan="infante Dom">inf.^e D.</abbr> João, e neto d´el-<lb/>
rei <abbr expan="Dom">D.</abbr> Pedro de Portugal e de
<abbr expan="Dona">D.</abbr> Ignez<lb/>

deCastro, sua mulher, e bisneto d'el -<lb/>
rei deCastella, <add place="above"> (<abbr
expan="Dom">D.</abbr> Sancho IV pai de <abbr
expan="Dom">D.</abbr> Prates <gap extent="1 word"
reason="illegible"/> d'Affonso IV)</add> o que venceu a
batalha<lb/>
do Salado. Este <abbr expan="Dom">D.</abbr> Fernando foi
padre<lb/>
de <abbr expan="Dona">D.</abbr> Catharina, abbadessa de
Lorvão,<pb n="3"/>
que aqui mandou trasladar na<lb/>
era do nascimento de <abbr expan="Nosso Senhor Jesus">N.
S. J.</abbr> Christo<lb/>
de 1479, aos 25 de Janeiro."</quote></p>

<p>Com elle tem aqui suas ossadas <abbr
expan="Dona">D.</abbr> Gra-<lb/>
cia e <abbr expan="Dom">D.</abbr> João, todos de
appellido d'Eça."</p>
<p><abbr expan="unclear">V. Hist. Seraf.</abbr>, Tomo
<abbr expan="segundo">2.^o</abbr>, <abbr
expan="pagina">pag.</abbr> 647. A<lb/>
<abbr expan="unclear">M.</abbr> Lusitana diz que estão os
ossos<lb/>
na <abbr expan="capela">cap.</abbr> mór de Lorvão.</p>

<p>Entre os numerosos filhos d'este pa-<lb/>
<hi rend="underline">triarcha</hi> ha um tambem de no-
<lb/>
me <abbr expan="Dom">D.</abbr> <hi
rend="underline">Fernando d'Eça</hi>, que foi creado<lb/>
do <abbr expan="primeiro">1^o</abbr> Duque de
Bragança e alcaide<lb/>

mor de Villa Viçosa. Casou com <abbr
expan="Dona">D.</abbr><lb/>
Joanna de Saldanha ou d´Aragão,<lb/>
<abbr expan="filha">f^a</abbr> de Fernão Lopes
de Saldanha,<lb/>
coretador mor deCastella, da qual<lb/>
teve 5 filhos.</p>

<p>Voltando ao infante <abbr expan="Dom">D.</abbr>
João,<lb/>
teve elle da <abbr expan="segunda">2^a</abbr>
mulher, <abbr expan="Dona">D.</abbr> Constan-<pb n="4"/>
ça, bastarda d´el-rei <abbr expan="Dom">D.</abbr>
Henrique<lb/>
<abbr expan="segundo">2.^o</abbr> deCastella,
duas filhas -<abbr expan="Dona">D.</abbr> Be-<lb/>
atriz ou Maria, e <abbr expan="Dona">D.</abbr> Marinha
<gap extent="1 word" reason="illegible"/><lb/>
ambas com geração, que não<lb/>
vem para o caso.</p>

<p>O tal infante manteve lá<lb/>
por Hespanha as tradições hon-<lb/>
radas da familia, e entre os<lb/>
seus filhos bastardos vejo um<lb/>
<hi rend="underline"><abbr expan="Dom">D.</abbr>
Fernando</hi>, que foi <abbr expan="Senhor">Sr.</abbr> de
Bragan-<lb/>
ça, e marido de <abbr expan="Dona">D.</abbr> Leonor Cou-
<lb/>
tinho, de <abbr expan="quem">q.^m</abbr> teve
dois filhos, <abbr expan="Dom">D.</abbr> <del
rend="overstrike">Antonio Affonso deCascaes, - e
<abbr expan="Dom">D.</abbr> Pedro<lb/>

da Guerra, ambos <abbr expan="com geração">c.
g.</abbr></p>

<p>Veja <abbr expan="Vossa Excellencia">V.
Ex.^a</abbr> se percebeu esta em-<lb/>
brulhada noticia, e <abbr
expan="muito">mt.^o</abbr> estimarei<lb/>
que em algum ponto o satisfaça.</p>

<p>É certo que estamos em divergen-<lb/>
cia; mas eu desejo que me obrigue a<lb/>
a estudar até achar a concordancia.<lb/>
Nada sei de tal ama com uma crean-<lb/>
ça, em que falla o Fernão Lopes. Ne-<lb/>
nhum dos meus linhagistas se refere a<lb/>
isto, e admiro. Será peta?</p>

<closer><p class="salute">Sou com <abbr
expan="muita">mt.^a</abbr> estima e<abbr
expan="consideração">consid.^{ção}</abbr><lb/>
De <abbr expan="Vossa Excellencia Amigo creado">V.
Ex.^a</abbr><lb/>
Am^o</abbr> cr.^{do}</abbr> Admirador<lb/>
<abbr expan="obrigadissimo">obg.^{mo}</abbr></p>
<p class="signature">José D´Azevedo e Menezes.</p>
<p class="dateline">
<place><abbr expan="Sua Casa">S. C.</abbr> do
Vinhall</place>
<date when="12.01.1887">-12-<abbr
expan="primeiro">1^o</abbr>-
1887</date></p></closer>

</body>
</text>

ANEXO 3

Estrutura do documento DTD

```
<!ELEMENT TEI.2 (teiHeader, text)>
<!ELEMENT teiHeader (fileDesc, encodingDesc?,
profileDesc?, revisionDesc?)>
<!ELEMENT fileDesc (titleStmt, publicationStmt,
sourceDesc)>
<!ELEMENT titleStmt (title, author, sponsor+, respStmt+)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT author (#PCDATA | name)*>
<!ELEMENT name (#PCDATA)>
<!ATTLIST name type CDATA #IMPLIED>
<!ELEMENT sponsor (name+, address)>
<!ELEMENT address (addrLine+)>
<!ELEMENT addrLine (#PCDATA | abbr | gap | sup | hi)*>
<!ELEMENT respStmt (resp,name+)>
<!ELEMENT resp (#PCDATA)>
<!ELEMENT publicationStmt (authority+, idno,
availability)>
<!ELEMENT authority (name+, address)>
<!ELEMENT idno (#PCDATA)>
<!ATTLIST idno type CDATA #REQUIRED>
<!ELEMENT availability (p+)>
<!ATTLIST availability status CDATA #REQUIRED>
<!ELEMENT p (#PCDATA | abbr | lb | pb | sup | unclear |
gap | hi | del | space | quote | add | place | date |
address | label | genealogia)*>
<!ELEMENT sourceDesc (letDesc)>
<!ELEMENT letDesc (letIdentifier, letHeading, physDesc,
envOcc, letContents, history?, additional?)>
<!ELEMENT letIdentifier (country, region, settlement,
institution, repository, collection,idno, note?)>
<!ELEMENT country (#PCDATA)>
```

```

<!ELEMENT region (#PCDATA)>
<!ELEMENT settlement (#PCDATA)>
<!ELEMENT institution (#PCDATA)>
<!ELEMENT repository (#PCDATA)>
<!ELEMENT collection (#PCDATA)>
<!ELEMENT note (#PCDATA)>
<!ELEMENT letHeading (author, addressee, placeLet,
dateLet)>
<!ATTLIST author attested (yes | added | no | unk)
#IMPLIED>
<!ELEMENT addressee (#PCDATA)>
<!ATTLIST addressee attested (yes | added | no | unk)
#IMPLIED>
<!ELEMENT placeLet (#PCDATA)>
<!ATTLIST placeLet attested (yes | added | no | unk)
#IMPLIED>
<!ELEMENT dateLet (#PCDATA)>
<!ATTLIST dateLet attested (yes | added | no | unk)
#IMPLIED>
<!ELEMENT physDesc (type, extent)>
<!ELEMENT type (#PCDATA)>
<!ELEMENT extent (dimensions)>
<!ELEMENT dimensions (height, width)>
<!ELEMENT height (#PCDATA)>
<!ELEMENT width (#PCDATA)>
<!ATTLIST height units (mm) #REQUIRED>
<!ATTLIST width units (mm) #REQUIRED>
<!ELEMENT envOcc EMPTY>
<!ATTLIST envOcc occ (yes | no) #REQUIRED>
<!ELEMENT letContents (p+, note?)>
<!ELEMENT history (p+, note?)>
<!ELEMENT additional (surrogates, listBibl*)>
<!ELEMENT surrogates (p+)>
<!ELEMENT listBibl (bibl)>

```

```

<!ELEMENT bibl (#PCDATA | author | title | pubPlace |
publisher | date | idno)*>
<!ELEMENT pubPlace (#PCDATA)>
<!ELEMENT publisher (#PCDATA)>
<!ELEMENT date (#PCDATA | abbr | lb | hi | unclear)*>

<!ELEMENT text (body)>
<!ELEMENT body (pb, print?, opener, p+, closer,
postscript?, p*)>
<!ELEMENT pb EMPTY>
<!ATTLIST pb n CDATA #REQUIRED>
<!ELEMENT print (#PCDATA | abbr | lb)*>
<!ATTLIST print type (letterhead) #REQUIRED>
<!ELEMENT opener (p+)*>
<!ATTLIST p class CDATA #IMPLIED>
<!ELEMENT abbr (#PCDATA | sup | lb | hi | pb | unclear)*>
<!ATTLIST abbr expan CDATA #REQUIRED>
<!ELEMENT sup (#PCDATA)>
<!ELEMENT lb EMPTY>
<!ELEMENT unclear (#PCDATA | abbr | lb | del | sup)*>
<!ATTLIST unclear reason CDATA #IMPLIED>
<!ELEMENT gap EMPTY>
<!ATTLIST gap extent CDATA #REQUIRED>
<!ATTLIST gap reason CDATA #REQUIRED>
<!ELEMENT hi (#PCDATA | abbr | sup | pb | lb)*>
<!ATTLIST hi rend CDATA #REQUIRED>
<!ELEMENT del (#PCDATA | lb)*>
<!ATTLIST del rend CDATA #REQUIRED>
<!ELEMENT space EMPTY>
<!ATTLIST space rend CDATA #REQUIRED>
<!ELEMENT quote (#PCDATA | lb | gap | hi | pb |
unclear)*>
<!ATTLIST quote rend CDATA #REQUIRED>
<!ELEMENT add (#PCDATA | abbr | gap)*>

```

```
<!--ATTLIST add place CDATA #REQUIRED-->
<!--ELEMENT place (#PCDATA| abbr | lb | hi)*-->
<!--ATTLIST date when CDATA #IMPLIED-->
<!--ELEMENT closer (p+ | address)*-->
<!--ELEMENT postscript (label*, p+, pb*)-->
<!--ELEMENT label (#PCDATA)-->
<!--ELEMENT genealogia (#PCDATA | unclear | lb | abbr |
del)*-->
<!--ENTITY ntilde "&#241;"-->
<!--ENTITY iquest "&#191;"-->
<!--ENTITY utilde "&#x169;"-->
<!--ENTITY qtilde "q&#x0303;"-->
<!--ENTITY etilde "e&#x0303;"-->
```

ANEXO 4

Certificado de comunicação proferida no âmbito do *Congresso de Humanidades Digitais em Portugal - Construir pontes e quebrar barreiras na era digital*, intitulada "Arquivo Digital da Correspondência dirigida a Camilo Castelo Branco: problemas, soluções e desafios", Faculdade de Ciências Sociais e Humanas - Universidade Nova de Lisboa, dia 9 de outubro de 2015.

