

COLEÇÃO ATAS

JORNADA DE HUMANIDADES DIGITAIS DO CIDEHUS

Renata Vieira e Ana Paula Banza

ATAS

Renata Vieira
Ana Paula Banza

Universidade de Évora

Ficha Técnica

Título: Jornada de Humanidades Digitais do CIDEHUS

Autores: Renata Vieira e Ana Paula Banza

© Os autores

Editor: Imprensa da Universidade de Évora

Coleção: Atas

Paginação: Divisão de Comunicação da Universidade de Évora

ISBN: 978-972-778-268-0

Évora 2022

ÍNDICE

Prefácio	4
<i>Renata Vieira e Ana Paula Banza</i>	
Abstracts	5
Conferências convidadas	
Simulação e Performatividade: O Arquivo LdoD em 8 Diagramas.....	9
Manuel Portela	
Ler de longe o quê? Humanidades Digitais e Anotação Linguística.....	18
Cláudia Freitas	
Comunicações	
A edição digital da <i>História do Futuro</i> , de António Vieira: arquivo e ferramentas.....	28
Ana Paula Banza, António Rito Silva e Irene Rodrigues	
As Humanidades Digitais aplicadas à Paleografia e Diplomática: contributos internacionais.....	39
Ana Pereira Ferreira	
The Monsoon project: digital perspectives of the Habsburg <i>Estado da Índia</i>	48
Ana Sofia Ribeiro	
Processamento de Linguagem Natural em Arqueologia: o megalitismo português como estudo de caso.....	58
Ivo Santos	
História Digital – da transcodificação à ressignificação.....	68
Renata Vieira	
<i>As Memórias Paroquiais</i> : do manuscrito ao digital.....	75
Fernanda Olival, Helena Cameron e Renata Vieira	
A investigação colaborativa em Humanidades Digitais: o caso português.....	93
Luísa Alvim e Teresa Costa	

PREFÁCIO

As Humanidades Digitais estão na interseção da computação e das Humanidades, de maneira transdisciplinar e colaborativa. O seu estudo incorpora ferramentas e métodos digitais e considera novas formas de produzir conhecimento, ao mesmo tempo que assume uma posição crítica sobre a forma como o digital tem impacto sobre o património cultural e a cultura. Esta área do conhecimento está a consolidar-se cada vez mais no meio científico, nacional e internacional.

O Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora (CIDEHUS) congrega investigadores das Ciências Sociais e Humanas e das Ciências da Informação e Documentação, que, nos últimos anos, em conjunto com investigadores das Ciências da Computação, têm vindo a unir-se em torno das Humanidades Digitais, partilhando as grandes questões transversais, sem prejuízo dos desafios próprios de cada disciplina.

No sentido de agrupar esses investigadores, propusemos a realização de uma Jornada de Humanidades Digitais do CIDEHUS, um evento *online* e gratuito, aberto à comunidade. Os membros do CIDEHUS com interesse em participar na Jornada manifestaram-se e propuseram um conjunto de comunicações. Dessa forma, no dia 29 de outubro de 2021, em reunião virtual, compartilhamos um conjunto de trabalhos em desenvolvimento no CIDEHUS, nas diferentes áreas que se congregam em torno das Humanidades Digitais, tendo como principais objetivos a divulgação de projetos, a discussão e partilha de experiências e o desenvolvimento de novas parcerias neste domínio.

Este volume apresenta os artigos referentes às comunicações apresentadas pelos membros do CIDEHUS nesse dia e conta, ainda, com a participação especial de dois artigos dos conferencistas convidados, os professores doutores Cláudia Freitas, da PUC-Rio, e Manuel Portela, da Universidade de Coimbra. Na formatação dos vários contributos, respeitámos a opção de cada autor, no que toca ao uso da norma portuguesa ou brasileira, ao uso do AO90, bem como a opção pelo inglês, num dos textos.

O evento obteve 179 inscrições, distribuídas por mais de 50 instituições. De acordo com o registo de inscritos, o maior número de participantes veio do Brasil (87) e de Portugal (72), além de Angola (14), Espanha (5) e Cabo Verde (1); a maior parte professores e investigadores (47%), mas também estudantes (41%), além de outros (12%). As áreas de maior interesse informadas pelos inscritos foram a linguística, a filologia e a história. O evento e suas atas foram financiados por fundos nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto UIDB/00057/2020.

Renata Vieira e Ana Paula Banza
Évora, 15/7/2022

ABSTRACTS

SIMULATION AND PERFORMATIVITY: THE LDO D ARCHIVE IN 8 DIAGRAMS

Manuel Portela

University of Coimbra, Centre for Portuguese Literature (CLP)

mportela@fl.uc.pt

This chapter describes the LdoD Archive as an evolutionary environment that simulates the social multidimensionality of texts. By recreating the literary experience through a network of interventions according to a role-playing structure (reading, editing, writing), experimentation with the medium as a playful space of performative simulation moves away from Cartesian epistemology and the formalist methods that dominate the Digital Humanities. A socialized model of distributed cognition and creation, the LdoD Archive invents a textual space that transcends the bibliographic imagination.

WHAT DISTANT READING? DIGITAL HUMANITIES AND LINGUISTIC ANNOTATION

Cláudia Freitas

Pontifical Catholic University of Rio de Janeiro

claudiafreitas@puc-rio.br

I defend, in this brief text, the linguistic annotation as a practice of digital humanists, illustrating the point with data from literary collections and taking as categories of analysis human and gender characterizations. The main objective of the explorations is to exemplify how "new" forms of reading, such as distant reading and text mining, can be articulated, leading to the insertion of reading layers in the texts and, consequently, enriching analyses.

THE DIGITAL EDITION OF THE HISTORY OF THE FUTURE, BY ANTÓNIO VIEIRA: ARCHIVE AND TOOLS

Ana Paula Banza¹, António Rito Silva² and Irene Rodrigues³

¹ CIDEHUS, University of Évora, ² INESC, ID,

³ Department of Informatics, University of Évora

anabanza@uevora.pt, Rito.Silva@tecnico.ulisboa.pt, ipr@uevora.pt

This text discusses how the new conceptions of texts and editing, resulting from the arrival of Digital in the world of Philology, were considered in the project of creating a digital archive of the work *História do Futuro*, by Father António Vieira - "Digital archive and assisted analysis of the *History of the Future*, by António

Vieira" - which, by its nature, incomplete and fragmentary, constitutes an unattainable philological challenge outside the new digital paradigm. The structure and objectives of the project are also briefly presented, focusing on the construction of the archive and the implementation of certain natural language processing tools, selected, and adapted according to the type of text, the challenges it raises and the readings it allows. Some of its possible scenarios of use are presented to illustrate the added value that the project will bring, both from an academic point of view and in the dissemination of the work of Father António Vieira.

DIGITAL HUMANITIES APPLIED TO PALEOGRAPHY AND DIPLOMATIC: INTERNATIONAL CONTRIBUTIONS

Ana Pereira Ferreira
CIDEHUS - University of Évora and CH-University of Lisbon
anapsf@uevora.pt

From 6 to 8 of May 2021, took place the first international congress under the theme of Paleography and Diplomatic in the Digital Humanities. With lecturers from all over the world, there were several contributions that have shown the application of the Digital Humanities in the disciplines of Paleography, Diplomatic, Sigillography and Archives management. The purpose of this paper is to make an analysis, among the presentations made at the congress, of the various options available for the study of these sciences, verifying their strengths and weaknesses, future potentialities and uses to the assistance of the researcher in History. From more complex methodologies related to photography, to computer programs for transcription of historical documents or to specialized databases, we synthesize the main technologies available to the historian of the 21st century.

THE MONSOON PROJECT: DIGITAL PERSPECTIVES OF THE HABSBURG ESTADO DA ÍNDIA

Ana Sofia Ribeiro
CIDEHUS – Universidade de Évora, Évora, Portugal
asvribeiro@uevora.pt

MONSOON aims to study the internal dynamics of the Portuguese macro-region of *Estado da Índia* in its various dimensions in a holistic and complex way during the period in which Portugal was part of the Hispanic Monarchy (1580-1640). The challenge is to combine, at the same time and in the same instruments of analysis, the formal and informal presence of the Portuguese in the Asian seas since literature usually analysed the two perspectives separately [1] [2].

MONSOON intends to test a different methodological approach to the study of regions that were part of European empires in the early modern period based upon massive narrative collection of sources. It expects to develop a flexible digital approach that could be applied to other European historical spaces and documentary sources. To achieve these goals the team will use a set of documents known as *Livros das Monções*. The information extracted from the sources will be analyzed according to three variables – agents, events, spaces – and using digital humanities tools: digital ontological semantics applying Natural Language Processing techniques, social network analysis and digital mapping.

NLP IN ARCHAEOLOGY: PORTUGUESE MEGALITISM AS A CASE STUDY

Ivo Santos
CIDEHUS - University of Évora
ifs@uevora.pt

By its own nature, part of the research in Archaeology is destructive, making the archaeological record (data collection) central. In this work, it is assumed that it is essential that the register includes the relationship between all parties and allows us to understand, by man and machine, the semantic relationship between the existing data. The first steps of a PhD project that seeks to apply NLP and Archaeological Information Extraction are presented from a case study of the Megalitism.

DIGITAL HISTORY: FROM TRANSCODING TO RESIGNIFICATION

Renata Vieira
CIDEHUS, University of Évora
renatav@uevora.pt

From orality to physical symbolic registration, from clay to paper, from volumen to códex, from manuscripts to press, from paper to digital: in all these transformations there are many technological revolutions involved. The current revolution in computer language processing technologies opens new doors to older language registers. When considering the development of areas such as artificial intelligence, machine learning, natural language processing, semantic web, virtual reality and so many others, the question arises: how will the future historian robots be? This work brings a look at the past and future of human records and technologies. We approach the issues involved in these transformation processes and relate them to the research developed within CIDEHUS. We invite for an expanded reflection on the theme Digital Humanities and the evolution of human thought.

THE PARISH MEMORIES: FROM MANUSCRIPT TO DIGITAL

Fernanda Olival¹, Helena Freire Cameron², Renata Vieira¹

¹ CIDEHUS, University of Évora, ² CIDEHUS, Polytechnic Institute of Portalegre
mfo@uevora.pt, helenac@ippportalegre.pt, renatav@uevora.pt

This text aims to trace the custodial history of the parish memories, from the dissemination of surveys in 1758, to current projects aimed at converting them into digital objects and data. It should be noted that it is a very relevant resource to know Portugal from the mid-eighteenth century, and that it interests not only the historians, but also many other scholars and actors in the local, regional, and country field. Within CIDEHUS, due also to its size, the Parish Memories have promoted interdisciplinary teamwork for more than a decade. Reflecting on this effort, and others made in Portugal, is also a way to evaluate and rethink work strategies on this collection.

COLLABORATIVE RESEARCH IN DIGITAL HUMANITIES: THE PORTUGUESE CASE

Luísa Alvim¹ and Teresa Costa²

¹ CIDEHUS - University of Évora

² CIDHEUS-University of Évora; Center for Classical Studies - University of Lisbon
luisaalvim@uevora.pt, teresa.costa.23@gmail.com

We present in this paper the Portuguese research around the area of digital humanities, through a study on the publications deposited in the Scientific Repositories of Open Access in Portugal, from 2012 to 2020. Being an emerging area in Portugal, several dynamics and projects have appeared. Many researchers, linked to research institutions, already constitute a community organized with common purposes and objectives, formalized in some universities, in lines of research. A bibliographical survey was carried out, where several types of documents from twenty-two national repositories were retrieved. It was found that most of the studies recovered were in the field of Information Science, followed by Literature and History, showing that researchers are dispersed throughout the Arts and Humanities, Social Sciences and Technologies. The lines of research are constituted by collaborations among researchers, creating bridges between the Humanities and other Sciences, proving its interdisciplinarity. The relevance of the collaborative work that the Digital Humanities require is verified, where experts in digital technologies and tools assist in the production of new knowledge in the Social and Human Sciences.

SIMULAÇÃO E PERFORMATIVIDADE: O ARQUIVO LDOD EM 8 DIAGRAMAS

Manuel Portela

Universidade de Coimbra, Centro de Literatura Portuguesa (CLP)

mportela@fl.uc.pt

Este capítulo descreve o Arquivo LdoD como um ambiente evolucionário que simula a multidimensionalidade social dos textos. Ao recriar a experiência literária através de uma rede de intervenções abertas segundo uma lógica de desempenho de papéis (ler, editar, escrever), a experimentação com o meio enquanto espaço lúdico de simulação performativa afasta-se da epistemologia cartesiana e dos métodos formalistas que dominam as Humanidades Digitais. Modelo socializado de cognição e criação distribuídas, o Arquivo LdoD inventa um espaço textual que transcende a imaginação bibliográfica.

1. UM AMBIENTE TEXTUAL EVOLUCIONÁRIO

Como usar o meio digital para simular a leitura, a edição e a escrita enquanto processos literários? Como conceber e desenhar o meio digital de um modo que transcenda a imaginação bibliográfica? O Arquivo LdoD é uma tentativa de responder a estas duas perguntas, juntando uma teoria performativa da ação literária e um ambiente computacional altamente dinâmico [1]. O *Livro do Desassossego*, de Fernando Pessoa (1888–1935), é a obra selecionada para esta simulação da multidimensionalidade social da experiência do texto. Os procedimentos algorítmicos foram imaginativamente explorados para mostrar de que modo a experiência literária é gerada por uma rede de intervenções abertas realizadas de acordo com uma lógica de desempenho de papéis.

Partindo da premissa de que a experiência literária é sustentada por ações de leitura, edição e escrita, os sujeitos são convidados a explorar múltiplos percursos de leitura, criar edições virtuais e escrever as suas próprias variações sobre textos do *Livro do Desassossego*. A simulação literária consiste em abstrair os atos de leitura, edição e escrita como papéis que se podem desempenhar dentro do ambiente textual [2]. As ações dos interatores passam assim a fazer parte do campo textual e todo o sistema evolui à medida que os utilizadores experimentam com esses processos literários. A magna obra em curso de Fernando Pessoa torna-se, portanto, um modelo de como um livro assume uma forma textual emergente que resulta das intervenções colaborativas de leitores, editores e escritores. Esta estrutura lúdica permite que os utilizadores mudem de perspetivas e de papéis, experienciando conscientemente os processos através dos quais o literário é dinamicamente constituído durante as interações de improvisação entre diferentes agentes.

Esta experiência literária e computacional é também uma tentativa de responder aos desafios apresentados pela modelação textual e desenho de interfaces digitais de acordo

com princípios humanísticos. A obra conhecida como *Livro do Desassossego* transformou-se assim no ponto de partida para uma tradução da teoria literária contemporânea no meio computacional. Enquanto experiência de humanidades digitais, o Arquivo LdoD: Arquivo Digital Colaborativo do *Livro do Desassossego* está informado pelas complexidades da crítica pós-estruturalista, teorias performativas da cultura, teoria dos sistemas, estudos de *software* e materialidades da literatura, entre outros. Em vez de ser apenas um artefacto funcional ou uma plataforma operacional – produzida segundo modelos representacionais de textualidade e princípios de transparência na engenharia de interação humano-computador –, apresenta-se antes como uma experiência concetual e técnica cujo resultado pode ser descrito como um ambiente textual evolucionário [5].

A noção de ambiente textual evolucionário pretende referir a evolução da sua estrutura e funcionalidades à medida que os investigadores continuam a experimentar o que é possível neste meio, por um lado, e às mudanças no seu conteúdo e usos conforme os interatores realizam as suas ações desempenhando papéis programados e não-programados, por outro. As funções dinâmicas e socializadas do Arquivo LdoD proporcionam uma interface interpretativa por meio da qual os sujeitos são instados a experimentar e a observar de que modo as suas ações se tornam constitutivas do próprio ambiente textual e de que modo elas são constrangidas pelas convenções que estão disponíveis para as interações, incluindo aquelas que eles próprios trazem para o jogo. Neste espaço de leitura-edição-escrita, as condições de produção e receção da própria *World Wide Web* foram integradas na ecologia da sua forma literária e computacional. O Arquivo LdoD muda o conteúdo e as relações entre os seus elementos à medida que as intervenções por esta comunidade de utilizadores modifica a configuração do seu espaço textual e computacional ao longo do tempo.

2. OITO DIAGRAMAS

Os interatores são convidados não apenas a ver a processualidade da leitura, da edição e da escrita tal como estas se instanciam a partir de múltiplas perspetivas através do conteúdo e das funções do Arquivo LdoD, mas podem fazer experiências com os seus próprios atos de leitura, edição e escrita, inscrevendo-os no sistema. O sistema evolui conforme essas inscrições e, por sua vez, as análises das interações são usadas iterativamente para alterar e adaptar as funcionalidades, recursos e restrições da plataforma. A modelação da performatividade literária opera assim de acordo com uma dupla lógica – representacional e simulatória –, gerando um ambiente textual cujo conteúdo e estrutura evoluem de modo a registar as intervenções dos seus interatores e a estimular a sua criatividade (Fig. 1). Este sistema textual aberto é o resultado imprevisto da tentativa de dar uma resposta teórica e computacional à pergunta: “o que é o *Livro do Desassossego*?”



Fig. 1

O Arquivo LdoD como representação e simulação

O projeto foi desencadeado pelo desejo de modelar a relação entre duas respostas a essa pergunta: em primeiro lugar, o *Livro do Desassossego* é uma obra inacabada, composta por um conjunto semideterminado de textos modulares em vários estágios de composição; em segundo lugar, ele é também um conjunto de edições específicas que utilizaram diversos critérios de transcrição, seleção e organização dos respetivos textos. Quando começámos a modelar a relação entre estas duas perspectivas, o nosso foco transferiu-se da intenção de representar a história dessas relações, tal como está documentada no arquivo editorial e autoral da obra, para o desejo de simular os processos através dos quais uma obra se torna obra. Estes processos passaram a ser referidos, no modelo reconcetualizado, como performatividade literária, conceito com que pretendo referir o conjunto de processos materiais e sociais que sustentam a produção e a reprodução de significados, formas e experiências literárias. Estes processos de natureza temporal foram, por seu turno, abstraídos como função-leitor, função-editor, função-autor e função-livro (Fig. 2).



Fig. 2

Um modelo da performatividade literária

Deste modo, o Arquivo LdoD transformou-se gradualmente numa complexa arquitetura de *software*, baseada num modelo de dados e numa série de funcionalidades que visam expressar a relação entre uma camada de representação e uma camada de simulação. A evolução da dinâmica entre essas camadas permite formular um conjunto adicional de perguntas fundamentais: como podemos representar e simular a processualidade do livro em construção (enquanto produção material e operador concetual)? Como podemos representar e simular a processualidade dos atos envolvidos na sua construção, tais como a leitura, a edição e a escrita? Como podemos abri-los a novos atos de leitura, edição e escrita através dos quais as nossas próprias interações interpretativas se inscrevem recursivamente no seu espaço bibliográfico e imaginário? Como usamos o conhecimento gerado por esta experimentação para alterar as suas condições de inscrição e análise ao nível do código e da interface?

Cada uma das dimensões da performatividade é modelada enquanto representação: os atos de escrita pela codificação do *Livro do Desassossego* como um conjunto de materiais autógrafos; os atos de edição pela codificação do *Livro do Desassossego* como um conjunto de edições de materiais autógrafos pelos peritos; os atos de leitura mostrando o *Livro do Desassossego* como um conjunto de percursos efetivos de leitura. Ao mesmo tempo, a simulação dessas ações enquanto gama de possibilidades requer que sejam experienciadas por meio de intervenções nos materiais textuais que retroalimentam a camada representacional. Modelado por meio desta camada simulatória, o *Livro do Desassossego* dentro do Arquivo LdoD transforma-se num conjunto aberto de interações através do desempenho de papéis – isto é, um conjunto aberto de leituras, edições e escritas consteladas pela força gravitacional de um livro imaginado – e, em última análise, num ambiente textual evolucionário (incluindo o seu próprio código-fonte).

Tendo em conta que o conteúdo e a estrutura do sistema respondem às ações em curso dos sujeitos, a processualidade e a performatividade daquelas ações literárias são experienciadas como constitutivas do sistema. Se levarmos em consideração que leitura, edição e escrita foram concebidas para explorar uma ampla gama de ações humanas e automáticas – apenas humanas, assistidas por máquina, assistidas por humanos, apenas máquinas –, o modelo também pode ser entendido como um ambiente digital experimental em que um grande conjunto de ferramentas e aplicações digitais podem ser adaptados, usados e transformados (Fig. 3). A modelação deste ambiente textual dinâmico torna-se, portanto, uma experimentação combinatória com a performatividade do próprio código por meio de um conjunto de técnicas, algoritmos e métodos.

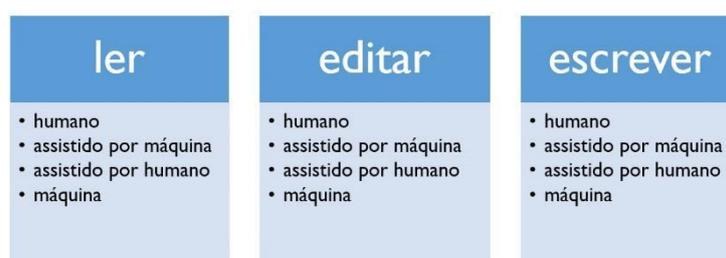


Fig. 3
Um ambiente digital experimental:
cognição distribuída numa ecologia algorítmica aberta e socializada

Os interatores experienciam as suas ações não apenas como uma série de possibilidades textuais que surgem de uma gama de valores e parâmetros predefinidos, mas também como uma exploração aberta da própria performatividade literária. O que é um ato de leitura? O que é um ato editorial? O que é um ato de escrita? De acordo com sua lógica de simulação, cada *output* é o resultado de uma ação colaborativa em que a produção algorítmica do sistema é modificada pelas intenções e procedimentos dos interatores. *Outputs* — seja sob a forma de percursos de leitura, sequências editadas, taxonomias de classificação, macrovisualizações ou novos textos — não estão inteiramente determinados pela lógica interna do sistema, uma vez que irão capturar o processamento humano do seu processamento programado. As improvisações constrangidas fomentadas pela lógica do Arquivo LdoD contribuem para a natureza emergente de leituras, edições e escritas como possíveis atualizações da experiência da obra (Figs. 4, 5 e 6).



Fig. 4

Da lógica da leitura humana à lógica da leitura virtual

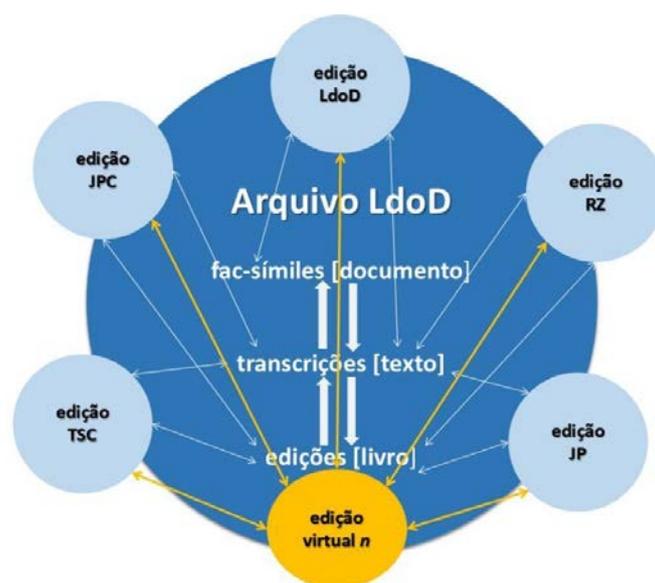


Fig. 5

Da lógica da edição crítica à lógica da edição virtual



Fig. 6

Da lógica da escrita autógrafa à lógica da escrita virtual

Esta abertura do Arquivo LdoD à ação interpretativa dos seus sujeitos ocorre em dois níveis entrelaçados de performatividade. A um nível, a ação interpretativa acontece de acordo com as perspetivas flexíveis e mutantes sobre as representações textuais. Noutro nível, a ação interpretativa funciona de acordo com práticas simulatórias de desempenho de papéis num jogo. À medida que os utilizadores inscrevem os seus atos de leitura, edição e escrita, o Arquivo LdoD documenta os resultados dessas interações e oferece-as a atos de leitura, edição e escrita adicionais, e a análises adicionais (Figs. 7 e 8). Tanto os conteúdos criados pelos utilizadores quanto as análises reflexivas fazem parte do sistema textual aberto. Tomado como um todo, o ambiente textual evolucionário poderia ser redescrito como uma prática literária socializada em desenvolvimento nos média programáveis em rede. A fusão entre modelo técnico e concetual, por um lado, e atos reais de leitura, edição e escrita, por outro, materializam a natureza ilimitada e iterativa da condição textual.



Fig. 7

O Arquivo LdoD como constelação das interações de leitores, editores, autores e livros



Fig. 8

O Arquivo LdoD como espaço textual evolucionário alimentado por análises textuais reflexivas

3. O JOGO LITERÁRIO COMO IMPROVISACÃO

Se pensarmos em termos éticos, poderíamos dizer que o Arquivo LdoD não é um sistema para automatizar a produção, recepção ou análise literárias [3]. Nesse aspecto, diferencia-se quer das abordagens dominantes na engenharia para a criatividade computacional por meio de sistemas de inteligência artificial, quer das abordagens dominantes nas humanidades digitais para processamento e análise textual. Pelo contrário, a sua ecologia de ação humana assistida por computador e ação computacional assistida por humano transforma os processos algorítmicos em procedimentos literários, com o objetivo de abrir espaços textuais para explorações críticas e criativas. O resultado é um ambiente alimentado pela imprevisibilidade e criatividade dos interatores humanos num processo social vivo, distribuído no tempo e colaborativo. Um processo em que as regras e os papéis do jogo literário podem ser desprogramados e reimaginados [4, 6].

O seu objetivo é instanciar as condições que permitem a reiteração da relação entre potencialidade e atualidade para cada ação individual de desempenho de um papel. Dado que este complexo ambiente computacional teve origem na tentativa de modelar a processualidade de dar existência a um livro (neste caso particular, o *Livro do Desassossego* de Fernando Pessoa), podemos dizer que a própria imaginação bibliográfica se reconfigura como força linguística e física na dinâmica que produz o literário como poética material e prática semiótica social. Torna-se um modelo vivo da identidade improvisada e emergente da obra de arte literária como uma construção cultural.

REFERÊNCIAS

1. Portela, M., Rito Silva, A., eds.: *Arquivo LdoD: Arquivo Digital Colaborativo do Livro do Desassossego*. Coimbra: Centro de Literatura Portuguesa da Universidade de Coimbra, (2017). <https://ldod.uc.pt>.
2. Portela, M.: The *Book of Disquiet* Archive as a Collaborative Textual Environment: From Digital Archive to Digital Simulator. In *The Writing Platform: Digital Knowledge for Writers*. Brisbane: Queensland University of Technology, (2017). <http://thewritingplatform.com/2017/07/book-disquiet-archive-collaborative-textual-environment-digital-archive-digital-simulator/>.
3. Portela, M.: The LdoD Archive as a creative textual environment and a model of literary performativity. In *Computational Creativity Meets Digital Literary Studies, Dagstuhl Reports*, Edited by Tarek Richard Besold, Pablo Gervás, Evelyn Gius, and Sarah Schulz, Vol. 9, Issue 4 (2019): 87–106, p. 98. | http://drops.dagstuhl.de/opus/volltexte/2019/11305/pdf/dagrep_v009_i004_p087_19172.pdf.
4. Portela, M.: An Evolutionary Textual Environment: The Unfinished Machine. *Textual Cultures*, 14.2 (2021): 220-238. <https://doi.org/10.14434/tc.v14i2.33659>.
5. Portela, M.: *Literary Simulation and the Digital Humanities: Reading, Editing, Writing*. New York: Bloomsbury, (2022).
6. Portela, M.: Literary Fields Forever: Playing with the Book of Disquiet. *Textual Cultures*, 15.1 (2022): 135-154. <https://doi.org/10.14434/tc.v15i1.34506>.

LER DE LONGE O QUÊ? HUMANIDADES DIGITAIS E ANOTAÇÃO LINGUÍSTICA

Cláudia Freitas
Pontifícia Universidade Católica do Rio de Janeiro
claudiafreitas@puc-rio.br

Resumo. Defendo, neste breve texto, a anotação linguística como uma prática também dos humanistas digitais, ilustrando o ponto com dados de acervos literários e tomando como categorias de análise caracterizações humanas e gênero. As explorações apresentadas têm como objetivo principal exemplificar de que maneira “novas” formas de leitura, como a leitura distante e a mineração de textos, podem ser articuladas, levando à inserção de camadas de leitura nos textos e, conseqüentemente, enriquecendo análises.

1. APRESENTAÇÃO

A incapacidade humana de lidar com o volume colossal de textos que temos à disposição, nas mais variadas áreas e associados a interesses igualmente variados, é uma das motivações para o desenvolvimento de métodos e abordagens para o processamento automáticos desses textos.

Mais recentemente, a parcela das Humanidades Digitais que tem nos textos seu objeto de estudo trouxe para cena o conceito de “leitura distante” (*distant reading*), uma forma de ler que combina, originalmente, Estudos Literários, Linguística Computacional e Computação na análise de grandes coleções de textos. Ainda que sua utilização venha contemplando cada vez mais material de fontes diversas, os primeiros trabalhos desenvolvidos com esta abordagem tiveram como alvos textos literários (por exemplo [15]).

A leitura distante surge como uma abordagem da história literária tendo em vista a impossibilidade de ler “de perto” (uma leitura convencional, o *close reading*) o acervo literário da literatura mundial ([14], [8]). Nos termos em que inicialmente é proposta, podemos pensar na leitura distante como uma estratégia para mitigar um problema de escala: a história literária depende de um cânone que seja, necessariamente, pequeno (do contrário, não poderá ser lido em um tempo de vida). A única maneira, portanto, de ir além do cânone é criar uma outra abordagem sobre o(s) texto(s), diferente da leitura convencional¹.

¹São dignas de nota, por outro lado, as seguintes observações de [26]: (i) *close reading* é um termo genérico que abriga diferentes perspectivas do que seja leitura – leituras estruturalistas, pós-estruturalistas, dos estudos culturais, dos estudos feministas, de ideologia crítica etc –, sinalizando para as várias abordagens do que seja *leitura* nos estudos literários – e (ii) a história literária – campo para o qual Moretti e Jockers recomendam abordagens “distantes” ou “macroanálises” – nunca

Podemos associar este outro tipo de leitura, distante, ao movimento de dar um passo para trás, de tomar distância, para ver melhor um objeto muito grande e complexo. Ver melhor, nesse caso, é ver a partir de um outro ângulo, um que leva à percepção de padrões, tendências e relações. Nas palavras de Franco Moretti, que cunhou o termo “um pouco por brincadeira e um pouco não”, a leitura distante é um modo de trabalhar em que a distância é uma aliada, uma forma específica de conhecimento, e não um obstáculo: é graças à distância que conseguimos ver melhor os padrões e as formas, tornando possíveis perspectivas sobre os dados que a leitura convencional não permite.

A leitura distante tem como insumo padrões que são, na imensa maioria das vezes, padrões baseados na forma: as palavras e suas frequências. Relacionamos essas contagens de palavras (e computadores são ótimos contadores) com outras variáveis, que podem estar fora do texto, como o tempo (séculos, décadas, anos), gêneros literários, etc. Com isso, vemos tendências, como a frequência da palavra escravidão ao longo do tempo [13], ou a distribuição da palavra amor entre a fala de personagens femininas ou masculinas [8]. Em consequência, os padrões que vemos refletem aquilo que está mais diretamente observável.

Por outro lado, e de maneira complementar, o Processamento de Linguagem Natural (PLN) vem, desde meados do século XX, se dedicando ao tratamento computacional das línguas e à exploração do conteúdo de grandes volumes de textos. Dentre os diversos subcampos do PLN, a Mineração de Textos (*Text Mining*) e, de maneira mais ampla, a Extração de Informação, tem aproximações nítidas com a leitura distante. Na Mineração de textos, o que está em jogo é a descoberta, por meio de técnicas computacionais, de informações consideradas “novas”, desconhecidas e/ou implícitas, tornadas visíveis — isto é, “mineradas” — por meio da detecção de padrões e extração de informações que estão nos textos [5]. A ideia geral é encontrar informações até então inacessíveis por meio de uma leitura convencional; explorar textos para encontrar coisas novas e, eventualmente, encontrar padrões e tendências. No entanto, técnicas e abordagens do PLN até pouco tempo eram apenas raramente aplicadas a objetos de interesse das Humanidades (e a exceção fica por conta da utilização de corpora eletrônicos nos estudos linguísticos, como mostra [17]).

E à medida em que as HDs vão avançando, avança também a integração de diferentes métodos e áreas; fica mais orgânica a combinação de abordagens quantitativas (mais frequentemente associadas à computação) e qualitativas (mais frequentemente associadas às Humanidades). Este ponto merece destaque porque sabemos ser comum, nas Humanidades, a tendência a associar métodos quantitativos a paradigmas objetivistas e métodos qualitativos a práticas interpretativas.

Nesse sentido, é ilustrativa a visão de Jokers [8], um dos proponentes e entusiastas da leitura distante, sobre esta “nova” abordagem para os textos: métodos computacionais e *big data* permitem que pesquisadores evitem ou mitiguem a subjetividade e o viés embutidos na observação humana, assumindo uma posição que evidencia, nas palavras de

tomou como fundamentação básica evidências coletadas por meio da leitura convencional (*close reading*), mas evidências vindas de categorias como autores, movimentos, gêneros e também da história social e institucional.

Smith [26], “sua visão antiquada de ciência”².

Mas sabemos também que essa associação é arbitrária: tanto é possível contar sem a ilusão de objetividade, como contar frequentemente pressupõe uma qualificação anterior ([4], [16] e [12]). Nesse sentido, um dos objetivos deste texto é mostrar como, na leitura distante auxiliada pela anotação linguística, é possível diluir esta dicotomia, sem abrir mão da dimensão interpretativa e sem cair na armadilha da neutralidade da tecnologia.

No PLN, tanto a extração de informação quanto a mineração de textos com frequência se beneficiam da anotação linguística. Anotação, no PLN, se refere à inclusão de informação linguística a porções de texto (palavras, sintagmas, frases, parágrafos...). O tipo de anotação mais popular talvez seja a anotação de classes de palavras, o que contribui para a ideia (equivocada) de que a anotação é uma atividade puramente técnica, ou objetiva³. Pelo contrário: a anotação é, em primeiro lugar, uma atividade interpretativa. De um ponto de vista prático, o que a anotação faz é classificar – ou organizar – os dados (as formas linguísticas), criando camadas de informação que, por sua vez, poderão enriquecer aquilo que, do ponto de vista das Humanidades, podemos chamar de leituras distantes.

Mas... ler (ou ver) de longe o quê? Leitura distante do quê? No PLN, as categorias de análise/categorias de anotação vêm das tarefas, majoritariamente vinculadas aos interesses da indústria. E, nas Humanidades, que categorias de análise podem levar a padrões relevantes? Até agora, muito do que se tem feito nas HDs aproveita ferramentas de anotação (e, conseqüentemente, classes de anotação) já disponíveis no PLN: anotadores de entidades permitem, por exemplo, um levantamento dos locais e pessoas (personagens) em obras ficcionais (Veja-se, respectivamente, [18] e [24] para a língua portuguesa, por exemplo); a anotação de sentimento permite um estudo sobre como personagens de duas obras de Kafka se caracterizam emocionalmente ([9]). O cruzamento de anotações do tipo local e sentimento, por sua vez, permite associar a presença de certas emoções e espaços geográficos em obras da literatura inglesa ([6]). Em [20] e [22] damos um passo na caracterização de textos e gêneros literários a partir dos diversos tipos de anotação disponíveis nos *corpora* do projeto AC/DC: além das anotações gramaticais, utilizamos anotação referente à identificação de discurso relatado, cores, roupas, emoções, relações de parentesco e corpo humano. Tais anotações, entretanto, embora até possam ser de interesse às HDs, em geral, e aos estudos literários, em particular, não foram motivadas por questões literárias ou sociais específicas. Já o trabalho de Higuchi [7] voltado para a História contemporânea brasileira, cria classes de anotação especialmente relevantes para este domínio.

²Jockers is not alone among contemporary literary scholars in his enthusiasm for, but rather old-fashioned view of, 'science'." (2016:67-68).

³Certamente contribui para esse entendimento a nossa falta de hábito de discutir as classes de palavras, em particular, e a própria gramática, em geral (no âmbito do PLN, algumas exceções são [21] e [3]). No entanto, não podemos esquecer que gramáticas – e classes de palavras, funções sintáticas, etc. — são construtos humanos, situados no tempo e motivados por interesses [2]. O que temos, nestes casos, é tão somente uma longa tradição disfarçada de neutralidade/objetividade.

A forma pela qual a leitura distante — e aquilo que chamarei de técnicas associadas do PLN, como a extração de informação e a mineração de textos — se apresenta pode fazer parecer que aquilo que deve ser lido — ou a informação a ser extraída/minerada — está adormecido (ou latente) — no texto, pronto para ser descoberto desde que tenhamos as técnicas adequadas. Partindo do pressuposto de que o que pode ser lido — e o que é informação — não está exatamente no texto, e que a própria noção do que é informação pode variar entre situações, tempo, espaço, pessoas, defendo, neste breve texto, a anotação como uma prática também dos humanistas digitais, ilustrando o ponto com dados de acervos literários e tomando como categorias de análise caracterizações humanas e gênero. As explorações apresentadas têm como objetivo principal exemplificar como essas “novas” formas de leitura (mineração de textos e leitura distante) podem se articular para enriquecer análises.

2. O ACERVO E A MINERAÇÃO DOS TEXTOS

Os resultados que relato a seguir vêm do *corpus Literateca*, que é a junção de todos os *corpora* literários disponíveis no serviço AC/DC [19]⁴. O acervo da *Literateca* é dinâmico, e alguns dos *corpora* ali contidos estão constantemente recebendo novas obras. Especificamente, a *Literateca* é a junção dos seguintes *corpora*: Vercial, OBRas, NOBRE, Tycho Brahe e Colonia⁵. Todo o material é anotado morfossintática e semanticamente, o que permite uma mineração de textos precisa. A versão com que trabalhei para as análises apresentadas a seguir tem as seguintes características: 33.6 milhões de palavras; 857 obras e 246 autores, dos quais apenas 20 são mulheres e, dessas, apenas 4 são brasileiras. O acervo utilizado, assim, não contém nenhum recorte específico relativo à época, gênero literário, nacionalidade ou autoria, ainda que estas informações estejam incluídas nas obras, e a maior preocupação aqui está na discussão relativa aos procedimentos de análise.

Para encontrar as caracterizações humanas, buscamos estruturas linguísticas que reconhecidamente sinalizam a presença de algum tipo de predicação, como apostos, adjuntos adnominais e predicativos (encontráveis devido à anotação sintática), associados a pessoas (encontráveis devido à anotação semântica de nomes próprios e à utilização de itens lexicais tipicamente associados a pessoas, como homens, mulheres, cavalheiros, donzelas, etc). Assim, graças à anotação, é fácil encontrar o seguinte tipo de construção:

- (1) id = “O Coruja Prosa: romance AA 1889 realismo masc”: é feio, mas enfim, sempre há homens sérios, cujo nome o público não ignora;
- (2) id = “O Cabeleira Prosa: romance FT 1876 realismo regionalismo romantismo masc”: Mas José César era ativo, enérgico, esforçado e de grandes espíritos.
- (3) id = “Uma lágrima de mulher prosa: romance AA 1880”: E Rosalina, meiga, encarava com chorosa ternura o olhar sombrio de Miguel.

⁴O AC/DC é um serviço de acesso de disponibilização de *corpora* anotados, criado e mantido pela Linguatca [23]. Pode ser acessado em <https://www.linguatca.pt/ACDC/>

⁵<https://www.linguatca.pt/acesso/corpus.php?corpus=LITERATECA>.

2.1. O QUE VEMOS? (PARTE 1)

A figura 1 mostra os resultados relativos às predicções associadas a personagens femininas e masculinas. Nas caracterizações femininas, destaca-se, em primeiro lugar, a palavra *bonita*, seguida de *casada*, *formosa*, *bela* e *condessa*; *amada*, *honest*a, *alta*. Nas masculinas, destacam-se, na mesma proporção, *alto* e *conde*, mas também *padre*, *rico*, *doutor*, *forte*, *livre* e *pároco*. Podemos esboçar uma análise desses dados, isto é, do que vemos: das caracterizações femininas, temos 3 associadas à aparência (*bonita*, *formosa*, *bela*), e duas a algum tipo de papel social (*casada* e *condessa*). Já nas masculinas, exceto por *alto*, as demais caracterizações frequentes estão na esfera social: *conde*, *padre/pároco*, *doutor*, *rico*. Será possível falar em tendências a partir desses resultados? Caracterizações femininas em sua grande maioria associadas à beleza, e masculinas a ocupações sociais? No entanto, o que não vemos nas figuras é a imensa quantidade de predicadores que ocorrem pouco, e formam uma espécie de sombra ao fundo da figura. No total, temos mais de 1.200 predicções femininas e mais de 1.800 predicções masculinas. No entanto, em ambos os casos, cerca de 60% das predicções aparece uma única vez, e, portanto, ficam invisíveis à distância. Como observar padrões em ocorrências que não se repetem? Que quadro veríamos, levando em conta todas as predicções?



Fig. 1

Predicadores femininos (esquerda)
e predicadores masculinos (direita)

Para conseguir atribuir sentido às ocorrências únicas, precisamos organizá-las em classes — que podemos chamar aqui de classes de anotação. Toda classificação é motivada por interesses, e nesse caso o interesse está na caracterização humana. A partir da leitura e análise dos dados brutos, foram criadas 4 classes: aparência, caráter, papel social e emoção (veja-se [4] para uma descrição inicial da classificação). Após a distribuição manual de todos os predicadores nos 4 eixos⁶, anotamos automaticamente todo o material com essa classificação, criando um cenário que possibilita a apreensão de padrões. Este trabalho deu origem a um dos projetos de anotação da *Literateca*, chamado Predicação Humana^{7,8}.

⁶Características emocionais de pessoas (a classe *emoção*) também podem ser consideradas traços de *caráter*, mas optamos por manter a distinção.

⁷Material disponível em <http://www.linguateca.pt/Gramateca/PredicaçãHumana.html>.

⁸Um estudo sobre essa classificação e seus resultados apenas no corpus OBRas estão descritos em [25].

2.2. O QUE VEMOS? (PARTE 2)

Se levamos em conta só os 20 lemas mais frequentes (os mais visíveis), temos o quadro da figura 2. Por outro lado, se consideramos todos os lemas, vemos a distribuição da figura 3. Em cada imagem vemos padrões diferentes (e podemos fazer leituras distantes diferentes), mas isto só é possível porque, por meio da distribuição das palavras em classes, incluímos informação de sentido (as classes de anotação) no material.



Fig. 2
Distribuição conforme 20 lemas mais frequentes

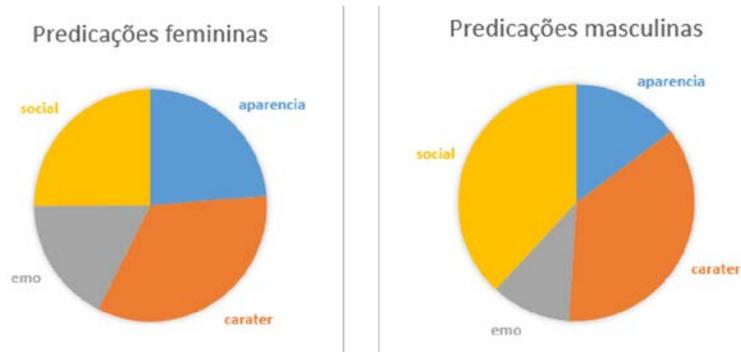


Fig. 3
Distribuição conforme todos os lemas

3. ALGUMAS PALAVRAS SOBRE ANOTAÇÃO

Frequentemente – e de maneira equivocada - a anotação é entendida, tanto no PLN como nas HDs, como uma atividade técnica, objetiva e independente de teoria (veja-se por exemplo a visão de [10]⁹). Mas, como também já indiquei, desde sua origem trata-se de uma tarefa (ou atividade) interpretativa¹⁰ e a impressão é que, com a sua popularização, esta dimensão interpretativa e humana vem, aos poucos, sendo apagada.

A anotação organiza e distribui os dados, para que possamos generalizar e falar melhor sobre eles – ou, nos termos das HDs, para que possamos ler de longe padrões. Mas padrões que estão lá porque foram incluídos por nós, conforme nossos interesses. Quando me refiro a “nós”, não remeto ao aspecto de execução da anotação (se manual, automática e semi-automática), mas ao aspecto humano de desenho das classes de anotação, de decisão a respeito de quais categorias criar, em função de quais interesses. A organização (ou distribuição) dos dados propiciada pela anotação, portanto, não tem nada de “natural” ou de “ausência de viés”, para retomar os termos de [8]; muito pelo contrário, é o resultado de uma classificação interessada.

Além da não-neutralidade, outro aspecto da anotação que gostaria de salientar aqui é que, para além do produto final que ela viabiliza (um texto anotado), a anotação interessa a nós, humanistas, enquanto processo.

Retomando as análises das figuras 2-5, poderíamos (e deveríamos) nos perguntar por que os números que indicam que mulheres são mais caracterizadas pela aparência do que os homens são confiáveis. Se a anotação é um processo interpretativo, por que devemos confiar na interpretação/classificação — e, conseqüentemente, confiar nos dados? Quem garante a validade dessa interpretação? Não estaria ela carregada de viés?

Esta confiança é garantida pela concordância interanotadores (CIA), procedimento que mede o consenso entre diferentes análises. Uma alta CIA indica que podemos ter confiança na interpretação subjacente à classificação dos dados.

Quando inserimos a nossa interpretação nos dados por meio da anotação, o que teremos, em grande escala, é essa mesma nossa análise (que por sua vez poderá ser reproduzida/aprendida pelas máquinas). Ainda, assumir a anotação como atividade interpretativa indica também que podemos discordar dela — da classificação — e que é possível propor classes e classificações diferentes, ainda que a motivação para a anotação seja a mesma. Afinal, se a anotação não é “natural”, ela é passível de discussão.

A anotação aparece, assim, como uma maneira de amenizar críticas de que a leitura

⁹CL methodology demands high inter-annotator agreement and *theory-independent categories*, while the categories in hermeneutic reasoning are often tied to a particular interpretive approach (viz. a theory of literary interpretation) and may bear a non-trivial relation to a reader’s pre-understanding.” (Kuhn, 2019:53[10]) (grifo meu).

¹⁰“But what is corpus annotation? It can be defined as the practice of adding *interpretative*, linguistic information to an electronic corpus of spoken and/or written language data.” (Leech, 1997:2[11]) (grifo meu).

distante é uma abordagem meramente quantitativa, como as feitas em [1] e [26], por exemplo. Tomando-se as classes e o processo de anotação como atividades interpretativas, trata-se de uma maneira de concatenar a dimensão qualitativa à quantitativa, fazendo uso do melhor dos dois mundos.

E o que mais pode interessar inserir em textos, para que seja possível este outro tipo de leitura? Que outras classificações interessam às humanidades ou o que seria uma anotação para as humanidades? Que conceitos, que operadores analíticos julgamos relevantes para analisar um dado fenômeno ou conjunto de dados? Utilizar a anotação como um método, que vai do desenho do esquema de anotação até a sua aplicação no texto, passando pela validação por meio de uma CIA, abre uma série de possibilidades para as pesquisas em HDs.

4. CONSIDERAÇÕES FINAIS

Talvez seja um pouco anacrônico tratar de anotação e de esquema de anotação, de uma perspectiva quase artesanal, em um mundo em que a IA guiada pelos dados explode e supera as expectativas quanto aos resultados. Vemos atualmente modelos de língua gigantescos apresentarem resultados muito satisfatórios em tarefas complexas como responder perguntas e escrever textos. No que se refere à anotação, tais modelos apresentam a possibilidade de “ajuste fino” (*fine tuning*) a partir de dados especializados. Isto é, se até há alguns anos uma das limitações para experimentar anotações era o custo humano envolvido na preparação dos dados — quanto mais complexa uma anotação, mais dados são necessários para que as máquinas aprendam, e esses dados serão produzidos por pessoas — com o ajuste fino é possível, ao menos teoricamente, anotar apenas o suficiente para que as máquinas possam aprender a reproduzir as nossas interpretações.

Além disso, também sabemos que muitos dados não é sinônimo de dados melhores: dados reproduzem vieses e preconceitos presentes na sociedade. Neste sentido, a anotação é uma garantia de que não ficamos à deriva; pelo contrário, ela assegura que nós estamos no comando.

REFERÊNCIAS

1. Araújo, N.: Vista de longe, a literatura é o que desaparece... (acerca de um fracasso programático em Franco Moretti). In *Variações sobre o romance*, pp. 259–272. Rio de Janeiro: Makunaima (2016).
2. Auroux, S.: *A revolução tecnológica da gramatização*. Editora da Unicamp, Campinas, Brasil (1992).
3. Freitas, C., Trugo, L.F., Chalub, F., Paulino-Passos, G., Rademaker, A.: Tagsets and datasets: Some experiments based on Portuguese language. In *International Conference on Computational Processing of the Portuguese Language*. pp. 459–469. Springer, Canela, RS, Brazil (2018), <https://rd.springer.com/book/10.1007/978-3-319-99722-3>.
4. Freitas, C., Martins, F., Biar, L.: Um ‘olhar discursivo’ sobre predicação e gênero: aproximações metodológicas entre corpus e discurso. *Texto Livre* 15, e 36213 (fev 2022). <https://doi.org/10.35699/1983-3652.2022.36213>, <https://periodicos.ufmg.br/index.php/textolivre/article/view/36213>.
5. Hearst, M.A.: Untangling text data mining. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*. p. 3–10. ACL ’99, Association for Computational Linguistics, USA (1999). <https://doi.org/10.3115/1034678.1034679>, <https://doi.org/10.3115/1034678.1034679>.
6. Heuser, Ryan and Moretti, Franco and Steiner, Erik: The emotions of London. In *Literary Lab Pamphlet* 13 (2016), <https://litlab.stanford.edu/LiteraryLabPamphlet13.pdf>.
7. Higuchi, S.: *Extração automática de informações: uma leitura distante do Dicionário Histórico-Biográfico Brasileiro (DHBB)*. Tese de Doutorado. Departamento de Letras, PUC-Rio (2021).
8. Jockers, M.L.: *Macroanalysis: Digital Methods and Literary History*. Chicago: University of Illinois Press (2013).
9. Klinger, R., Suliya, S.S., Reiter, N.: Automatic Emotion Detection for Quantitative Literary Studies – A case study based on Franz Kafka’s “Das Schloss” and “Amerika”. In *Digital Humanities 2016: Conference Abstracts*. pp. 826–828. Jagiellonian University and Pedagogical University, Kraków, Poland (July 2016). <http://dh2016.adho.org/abstracts/318>.
10. Kuhn, J.: Computational text analysis within the humanities: How to combine working practices from the contributing fields? *Language Resources & Evaluation* 5e, 565–602 (2019).
11. Leech, G.: Introducing corpus annotation. In *Corpus annotation: Linguistic information from computer text corpora*, pp. 1–18. Taylor & Francis (1997).
12. Leite, M., Freitas, C.: Pesquisa em educação: perspectivas (qualitativas?) na exploração de grandes corpora. *OSLa-Oslo Studies in Language* 7, 139–152 (2015). <https://doi.org/https://doi.org/10.5617/osla.1458>.
13. Michel, J.B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Pickett, J.P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M.A., Aiden, E.L.: Quantitative analysis of culture using millions of digitized books. *Science* 331(6014), 176–182 (2011). <https://doi.org/10.1126/science.1199644>, <https://www.science.org/doi/abs/10.1126/science.1199644>.
14. Moretti, F.: Conjectures on world literature. *New Left Review* 1, 54–68 (2000).

15. Moretti, F.: *A literatura vista de longe*. Porto Alegre: Arquipélago Editorial (2008).
16. Santos, D.: *Podemos contar com as contas?*, pp. 194–213. Cambridge Scholars Publ (maio 2014), <http://www.linguateca.pt/Diana/download/Santos2014contas.pdf>.
17. Santos, D.: *Literature studies in Literateca: between digital humanities and corpus linguistics*, pp. 89–109. Novus Forlag, Oslo (2019).
<http://www.linguateca.pt/Diana/download/DianaSantosFSCEO.pdf>.
18. Santos, D., Alves, D.: *Placing GIS and NLP in literary geography: experiments with literature in portuguese* (submitted).
<https://www.linguateca.pt/Diana/download/SantosAlvesBILLIGcasestudy.pdf>.
19. Santos, D., Bick, E.: Providing Internet access to Portuguese corpora: the AC/DC project. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*. European Language Resources Association (ELRA), Athens, Greece (May 2000), <http://www.lrec conf.org/proceedings/lrec2000/pdf/85.pdf>.
20. Santos, D., Freitas, C., Lopes, J.M.: Comparando a literatura lusófona com outras literaturas: recursos para leitura a distância em português. In *I Congresso Internacional em Humanidades Digitais no Rio de Janeiro (HdRio2018)*. CPDOC/FGV (24 de setembro 2018), <https://www.linguateca.pt/Diana/download/SantosFreitasLopes2018Anais.pdf>.
21. Santos, D., Gasperin, C.: Evaluation of parsed corpora: experiments in user-transparent and user-visible evaluation. In *Proceedings of LREC 2002, the Third International Conference on Language Resources and Evaluation*. pp. 597–604. European Language Resources Association (ELRA), Las Palmas de Gran Canaria, Spain (May 2002), www.linguateca.pt/Diana/download/SantosGasperinLREC2002.pdf.
22. Santos, D., Pires, E., Freitas, C., Fuão, R.S., Lopes, J.M.: Periodização automática: Estudos linguístico-estatísticos de literatura lusófona. *Linguamática* 12(1), 81–95 (Jun 2020).
<https://doi.org/10.21814/lm.12.1.314>,
<https://linguamatica.com/index.php/linguamatica/article/view/314>.
23. Santos, D., Simões, A., Frankenberg-Garcia, A., Pinto, A., Barreiro, A., Maia, B., Mota, C., Oliveira, D., Bick, E., Ranchhod, E., et al.: *Linguatca: um centro de recursos distribuído para o processamento computacional da língua portuguesa* (2004).
24. Santos, D., Willrich, R., Langfeldt, M., Moraes, R.G.d., Mota, C., Pires, E., Schumacher, R.: Identifying literary characters in portuguese: Challenges of an international shared task. In *International Conference on Computational Processing of the Portuguese Language, PROPOR 2022*. Springer (2022).
25. Silva, F.M.d.R.P.d.: *Diferenciações de gênero na caracterização de personagens: uma proposta metodológica e primeiros resultados*. Dissertação de Mestrado. Departamento de Letras, PUC-Rio (2021).
26. Smith, B.H.: What was close reading? a century of method in literary studies. *Minnesota Review* 87, 57–75 (2016).

A EDIÇÃO DIGITAL DA *HISTÓRIA DO FUTURO*, DE ANTÓNIO VIEIRA: ARQUIVO E FERRAMENTAS¹¹

Ana Paula Banza¹, António Rito Silva² e Irene Rodrigues³

¹ CIDEHUS, Universidade de Évora, ² INESC, ID,

³ Departamento de Informática, Universidade de Évora

anabanza@uevora.pt, Rito.Silva@tecnico.ulisboa.pt, ipr@uevora.pt

Resumo. Aborda-se, no presente texto, a forma como as novas concepções de edição e de texto, resultantes da chegada do Digital ao mundo da Filologia, se reflectiram no projecto de criação de um arquivo digital da obra *História do Futuro*, do Padre António Vieira - “Arquivo digital e análise assistida da *História do Futuro*, de António Vieira” - que, pela sua natureza, incompleta e fragmentária, constitui um desafio filológico inatingível fora do novo paradigma digital. Apresenta-se também brevemente a estrutura e objectivos do projecto, tomando como foco a construção do arquivo e a implementação de determinadas ferramentas de processamento de linguagem natural, seleccionadas e adaptadas em função do tipo de texto, dos desafios que este suscita e das leituras que permite, desenvolvendo-se, a título de exemplo, alguns dos seus possíveis cenários de utilização, para ilustrar as mais valias que o projecto trará, quer do ponto de vista académico, quer na divulgação da obra do Padre António Vieira.

1. O PROJECTO ARQUIVO DIGITAL E ANÁLISE ASSISTIDA DA *HISTÓRIA DO FUTURO*, DE ANTÓNIO VIEIRA

A *História do Futuro* (HdoF) é uma obra muito diferente daquelas pelas quais o autor é geralmente conhecido, que são os sermões e as cartas. Neste caso, estamos perante uma obra de natureza escatológica, sobre o final dos tempos e sobre o papel que o reino de Portugal teria no estabelecimento do Império de Cristo na terra, o Quinto Império, sendo este último aspecto, que é verdadeiramente original e nada ortodoxo, o que causaria ao autor os problemas com a Inquisição que viria a enfrentar.

Apesar de ter trabalhado nesta obra desde 1649 até à sua prisão, em 1665, Vieira deixou-a inacabada e muito longe do plano, previsto para sete volumes. Do que chegou até nós desta obra, pensada para ser monumental, o Livro Antepimeiro da *História do Futuro*, organizado por Vieira, antes de ser preso, como uma introdução à obra, e enviado para a Corte numa espécie de manobra publicitária, teve uma primeira edição em 1718 e várias depois dessa. O resto dos papéis já escritos, de forma fragmentária, para a *História do Futuro* foi apreendido pela Inquisição na altura da prisão de Vieira, em 1665, e ficou apenso ao

¹¹Trabalho desenvolvido no âmbito do projeto UIDB/00057/2020 - FCT - Portugal.

Processo. Lúcio de Azevedo publicou, em 1918, o conjunto mais organizado destes fragmentos (cf. Sérgio e Cidade, 1953: 1-160), permanecendo ainda apenas parcialmente editado um outro conjunto de fragmentos, mais dispersos, cuja identificação e organização levantam vários problemas (cf. Muhana, 1994).

A Representação dos motivos que tive para me parecerem prováveis as proposições de que se trata, escrita em cativo, entre 1665 e 1666, como defesa da obra perante as acusações de heresia da Inquisição, sintetiza e, mais importante, completa de forma resumida a matéria que não tinha chegado a ser escrita.

Assim, do ponto de vista conceptual, da obra *História do Futuro*, fazem parte três obras fisicamente distintas, além dos fragmentos referidos, que é nossa intenção integrar no Arquivo digital da Obra.

No paradigma impresso, seria impossível colocar todos estes textos em diálogo, mas, como bem escreveu Paixão de Sousa (2013: 129):

“a lógica da difusão digital tem [...] como característica central (e [...] mais interessante) a desvinculação entre o texto lógico e seu suporte material. Ela descorporifica o texto”.

É precisamente nesta descorporificação do texto, isto é, na separação entre o que é físico e o que é conceptual, que se baseia este projecto, cujas questões de partida são as seguintes: Pode o meio digital permitir uma leitura nova e mais completa dos textos de António Vieira? Se sim, como?

Acreditamos que sim, porque o arquivo digital integrará toda a matéria escrita para a *História do Futuro*, reconstruindo “arqueologicamente” a Obra, a partir do seu plano, o que vai permitir uma visão nova e mais completa.

Por outro lado, a introdução de ferramentas de análise assistida permitirá identificar automaticamente as relações conceptuais entre os textos, comprovando a relação entre os diferentes textos físicos, além de assessorar a análise e comentário (linguístico, literário, histórico-cultural).

Assim, o projecto organiza-se em quatro tarefas, que se relacionam entre si: o texto, estabelecido na primeira tarefa, é integrado e organizado no arquivo, cuja construção constitui a segunda tarefa; as ferramentas, a implementar na terceira tarefa, usam o conteúdo do arquivo e são também incorporadas nele. Finalmente, a anotação e comentário, previstos para a quarta tarefa, usam o texto e as ferramentas incorporadas no arquivo e são, eles próprios, também incorporados no arquivo.

Relação entre Tarefas

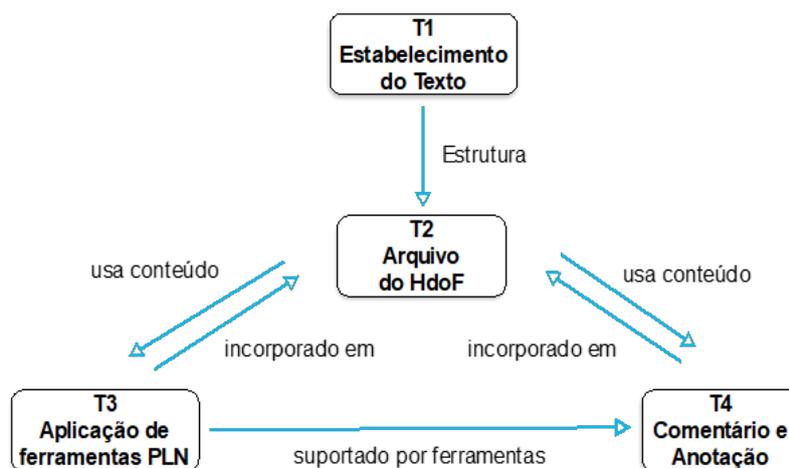


Fig. 1
Relação entre tarefas

2. O ARQUIVO

O objectivo do Arquivo da *História do Futuro* (HdoFArq.) é estruturar os fragmentos da obra do Padre António Vieira de modo a facilitar a sua interacção, promover a integração de ferramentas de Processamento de Linguagem Natural (PLN) e suportar a categorização e os comentários dos conteúdos da HdoF.

No que diz respeito ao primeiro objectivo, a interacção deve ser dirigida pela interpretação teórica associada à HdoF, a desenvolver na primeira fase do projecto, mas para a qual existe já uma base, estabelecida em trabalho anterior (Banza, 2008). Assim, é central fornecer mecanismos de ligação entre as diferentes partes de reescrita da HdoF, facilitando a sua identificação e comparação. Enquanto paradigma global de interacção com o arquivo HdoF, pretende-se adicionar mecanismos de navegação horizontal, entre as partes reescritas da HdoF, ao mecanismo usual de navegação vertical na obra. A interacção deve permitir, em qualquer momento, durante a leitura da HdoF, a possibilidade de ler e analisar todas as reescritas efetuadas pelo Padre António Vieira do fragmento em foco. Esta estrutura combinatória deve, contudo, preservar a ideia de uma obra, uma estrutura vertical, promovendo, não obstante, a combinação de reescritas. O maior desafio que se apresenta à definição da estrutura do arquivo HdoF será o estabelecimento da granularidade dos elementos reescritos, dado que o grão fragmento e sequência de fragmentos, como será inicialmente codificada a HdoF na sua dimensão vertical, pode não corresponder de forma directa às reescritas efetuadas pelo Padre António Vieira. Ou seja, a relação entre partes reescritas pode ser uma relação de muitos para muitos, quando considerado o grão fragmento. Ainda no que diz respeito a este primeiro objectivo, o

arquivo HdoF deverá disponibilizar as funcionalidades usuais de um arquivo associadas à pesquisa do seu conteúdo e ao suporte de meta-dados acerca dos seus fragmentos.

A integração com ferramentas de PLN deverá ter duas vertentes: análise e exploratória. Na vertente exploratória, o investigador, apoiado por um especialista de PLN, deve poder usar os conteúdos do arquivo HdoF para explorar as suas hipóteses de investigação através da aplicação de ferramentas de PLN. A vertente de análise permitirá aos investigadores, mas também a leigos, uma interacção sobre o arquivo HdoF através da utilização de ferramentas de PLN. Evidentemente, a vertente de análise resulta da estabilização dos resultados alcançados pelos investigadores na vertente exploratória, que, uma vez considerados de relevo, serão codificados no arquivo HdoF através da integração das ferramentas de PLN usadas na fase exploratória. Desta forma, poderão utilizadores leigos, não especialistas, usar as ferramentas na sua interacção com o arquivo HdoF. Com vista a suportar a vertente de exploração, o arquivo deverá fornecer um conjunto de interfaces de disponibilização de dados para serem consultados pelas ferramentas de PLN. Note-se que a bancada de experimentação deverá ser externa ao arquivo HdoF, uma vez que cada experimentação terá as suas características e tenderá a evoluir dinamicamente com o próprio processo de exploração. Já a vertente de análise exigirá uma adaptabilidade do arquivo HdoF por forma a permitir a fácil integração de novas ferramentas e dados. Para isso, será necessário que o arquivo HdoF possua uma boa modularidade, que limite o impacto das alterações a efectuar.

Finalmente, para se atingir o objectivo de suporte à categorização e à adição de comentários aos conteúdos da HdoF, será necessário que o arquivo permita, para além da meta-informação estática associada aos seus conteúdos, a adição dinâmica de meta-informação. Pretende-se suportar dois tipos de meta-informação: categorias e comentários. Esta meta-informação poderá ser gerada manualmente por um utilizador ou automaticamente por uma ferramenta de PLN. Tal como para a estruturação do arquivo HdoF, também aqui se levanta a questão de qual o grão do conteúdo sobre o qual se aplica a meta-informação. Sendo provável que a granularidade do fragmento não seja suficiente, serão provavelmente necessárias granularidades maiores, por exemplo, grupos de fragmentos, e menores, por exemplo, parte de um fragmento. Assim, o arquivo HdoF deverá possuir uma estrutura que suporte a introdução dinâmica de novos tipos de meta-informação.

Pretende-se desenvolver o arquivo HdoF estendendo o actual arquivo do *Livro do Desassossego* (LdoD), de Fernando Pessoa (Portela e Rito Silva, 2015; Rito Silva e Portela, 2015) com o objectivo de, através de processos de refactorização da sua base de código, construir um sistema que seja altamente flexível, de forma a poder ser facilmente adaptado para implementar diferentes repositórios de humanidades digitais. Assim, o trabalho terá por base o actual arquivo que se encontra em produção, assim como o trabalho já realizado na sua modularização e que se encontra descrito em Gonçalves et al. (2021).

Um dos resultados finais que se pretende alcançar é um sistema adaptável e reutilizável de módulos que possam ser usados na construção de diferentes arquivos. A Figura 2 apresenta este processo de definição de repositórios por integração de módulos específicos de um repositório com módulos comuns adaptados, exemplificado para os repositórios LdoD e

HdoF. Pretende-se que este processo permita desenvolver novos repositórios de forma robusta e reduzindo o esforço de desenvolvimento.

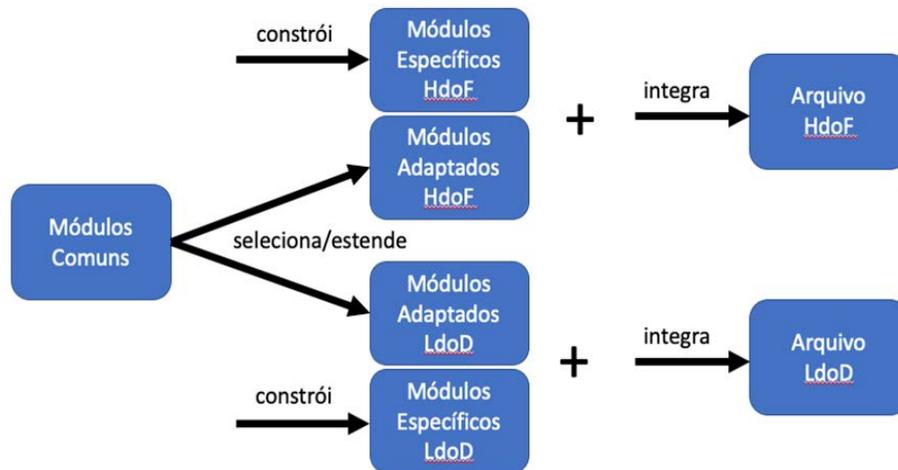


Fig. 2
Processo de Definição de Repositórios

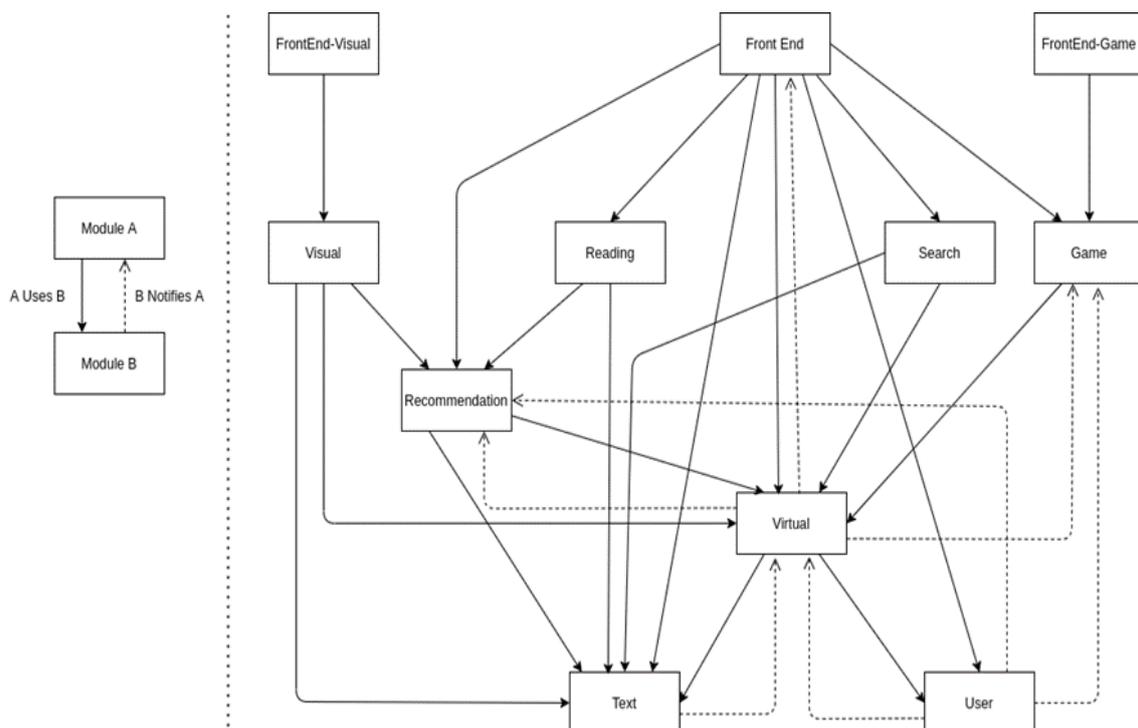


Fig. 3
Arquitetura de Módulos do Arquivo LdoD

Os módulos base que poderão ser adaptados do actual repositório de código são: *Text*, *User*, *Virtual*, *Search* e *Recommendation*. Estes módulos, ainda que requerendo alguma

adaptação e generalização, permitirão suportar alguns dos objectivos definidos para o arquivo HdoF.

O módulo de *Text* permitirá suportar a estrutura do arquivo HdoF, em particular as navegações verticais e horizontais. No arquivo LdoD, a navegação vertical está associada à estrutura de uma edição do *Livro do Desassossego*, enquanto a navegação horizontal está associada às diferentes transcrições de cada um dos fragmentos por parte dos seus editores. No arquivo da HdoF, haverá apenas uma navegação vertical, pelo que a adaptação do repositório existente será trivial. Contudo, a adaptação da navegação horizontal será mais complexa, pois terá de passar a aceitar mais graus de liberdade, devido às relações de muitos para muitos que poderão ocorrer entre diferentes fragmentos da HdoF, como anteriormente descrito.

O módulo de *User* será trivialmente adaptado para suportar a autenticação e autorização de utilizadores do arquivo HdoF, uma vez que estas funcionalidades não são centrais na semântica dos arquivos, sendo apenas logísticas.

Os módulos de *Search* e *Recommendation* podem ser adaptados para suportar as funcionalidades de pesquisa do arquivo; funcionalidades estas que estão associadas ao primeiro objectivo do arquivo.

O módulo *Virtual* implementa as funcionalidades de gestão de edições virtuais do arquivo LdoD. Estas funcionalidades são específicas deste arquivo, contudo incluem um conjunto de sub-funcionalidades que permitem a categorização da totalidade, ou partes, de um fragmento. Adicionalmente, associadas a essa categorização, também existem funcionalidades para comentar. Uma vez que um dos objectivos do arquivo HdoF é o suporte à categorização e comentário dos conteúdos da HdoF, a estratégia que se pretende seguir será a da extracção de um novo módulo de categorização e comentário do actual módulo *virtual*. Uma vez efectuada essa extracção, será possível usar este novo módulo para a categorização e comentário de fragmentos da HdoF.

Dos objectivos definidos para o arquivo HdoF, a integração de ferramentas de PLN não encontra nenhum módulo que possa ser estendido ou adaptado do actual repositório do LdoD. Assim, para se atingir este objectivo, deverá ser definido um novo módulo. Este módulo deverá ser integrado com os restantes módulos, por forma a manter a sua independência, como previamente descrito.

3. AS FERRAMENTAS

No que diz respeito às ferramentas a integrar no Arquivo HdoF, conforme referido anteriormente, elas servem essencialmente dois objectivos: a confirmação da identidade/complementaridade dos textos das diferentes obras físicas e a análise, em diferentes níveis.

Para o primeiro objectivo, serão essenciais as ferramentas de identificação de similaridades. No entanto, verifica-se que as ferramentas actualmente disponíveis não se adequam necessariamente a este tipo de texto e aos objectivos pretendidos, pelo que será necessário, em alguns casos, adaptar os modelos actuais ao caso da HdoF, conseguindo, essencialmente, mais eficiência, em termos de precisão e cobertura.

Nomeadamente, num caso como o dos excertos abaixo, em que as similaridades são evidentes, um dos objectivos será identificar automaticamente, não apenas correspondências exactas, na mesma linha ou em linhas diferentes, como as assinaladas a amarelo, mas também paráfrases, como as assinaladas a rosa, flexões da mesma palavra, como as assinaladas a verde, ou ainda outros tipos de semelhança semântica, como os assinalados a azul.

- **História do Futuro**

- Parecia-te que vias defronte de ti uma **estátua** grande, de estatura alta e sublime e de aspecto terrível e temeroso. **A cabeça** desta estátua era de ouro, **o peito** e os braços **de prata**, **o ventre** até os joelhos **de bronze**, **os joelhos** de ferro, **os pés** de ferro e de barro. **Estando assim suspenso** no que vias, **viste** mais que se **arrancava uma pedra de um monte**, cortada dele sem mãos, e que, dando **nos pés da estátua**, a derrubava. Então **se desfizeram** juntamente o barro, o ferro, o bronze, a prata, o ouro, e se converteram **em pó e cinza**, que foi levada dos ventos, e nem aqueles metais apareceram mais, nem o lugar onde tivessem estado; porém **a pedra** que tinha derrubado a estátua **cresceu**, e fazendo-se **um grande monte**, **ocupou e encheu toda a terra**.

- **Representação**

- Tinha a **estátua a cabeça de ouro**, **os peitos** de prata, **o ventre de bronze**, **ũa e outra perna de ferro** e **os pés de ferro e barro**. E enquanto Nabucodonosor **estava admirado** do que **via** (porque a estátua, como diz o texto, era grande e de aspecto terrível), **viu** mais que **de um montecão** ou se **arrancava, sem mãos, ũa pedra**, a qual deu um golpe **nos pés da estátua** com tão maravilhoso efeito que a estátua e seus metais **se desfizeram todos em pó e cinza**, e a **pedra** **crescendo** se converteu em **um monte** de tão imensa grandeza que **cobriu toda a terra**.

Fig. 4

O episódio da estátua dos quatro metais: tipos de similaridades

Para a análise, nos domínios linguístico, histórico-cultural e literário, serão também usadas diferentes ferramentas já disponíveis (como o reconhecimento de entidades nomeadas ou a análise de sentimentos), mas que poderão também ser adaptadas ao tipo específico de texto em causa. Tomando como exemplo as ferramentas de análise linguística, prevê-se a utilização do esquema de Dependências Universais, por exemplo, para a análise morfo-sintáctica.

Neste caso, o desempenho das ferramentas existentes é, no geral, bastante razoável. No entanto, tratando-se de um texto do séc. XVII e de um texto de Vieira, há, apesar de tudo, desafios importantes a resolver, como é o caso da interpretação de grafias antigas (e.g. ãa = uma) ou da interpretação da função sintáctica de constituintes deslocados (Teixeira et al., 2019).

Tomem-se como exemplo as frases assinaladas nos dois excertos já referidos:

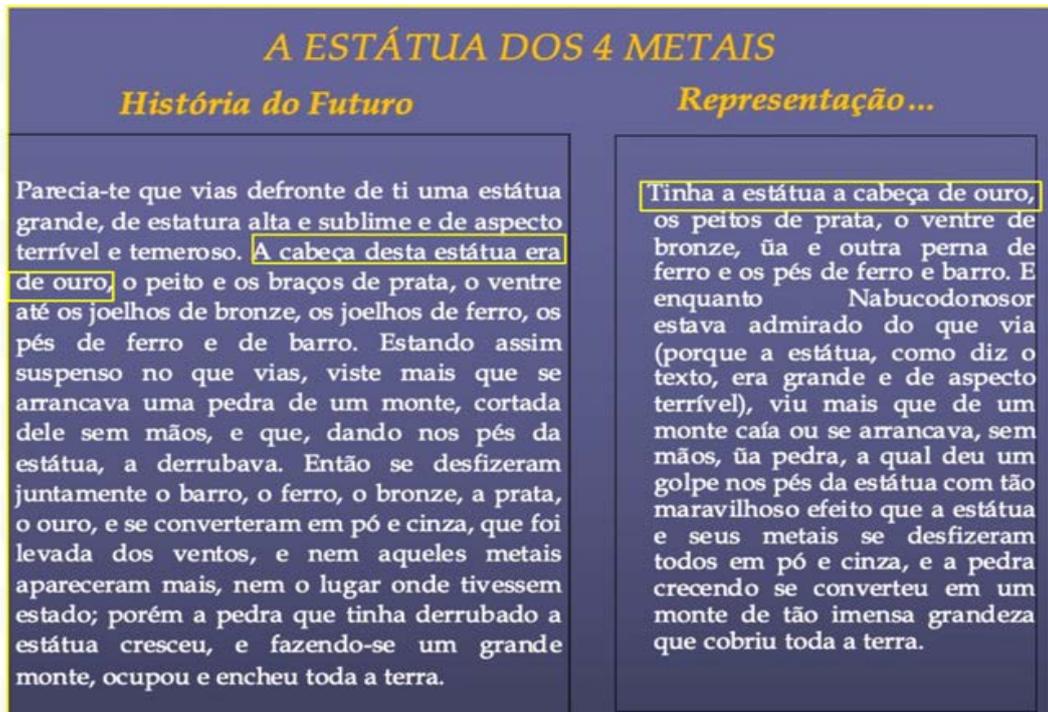


Fig. 5

O episódio da estátua dos quatro metais: ex. de similaridade morfo-sintáctica

Na frase 1, “a cabeça desta estátua era de ouro” (HdoF), a análise de um analisador sintáctico como o stanza.run é a seguinte:

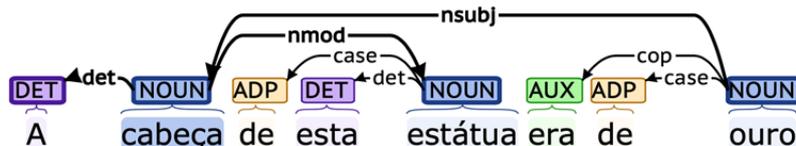


Fig. 6

Análise sintáctica da frase 1

O grafo semântico desta representação sintáctica (Silva et al., 2021) representa as seguintes relações entre os diferentes objectos, permitindo a extracção das seguintes anotações (“parte_de (cabeça,estátua) e parte_de (ouro,cabeça)”).

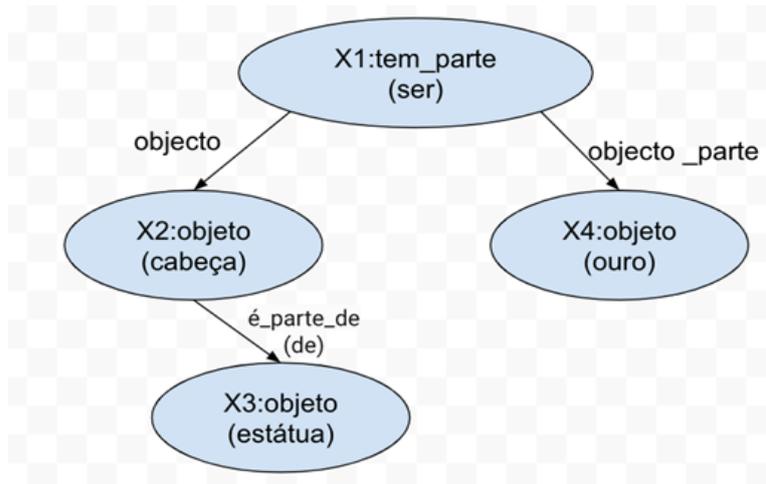


Fig. 7
Grafo semântico da frase 1

Na frase 2, “tinha a estátua a cabeça de ouro” (Representação...), o stanza.run identifica “a estátua” como objecto, por estar depois do verbo, e não como sujeito, como se apresenta na Figura 8:

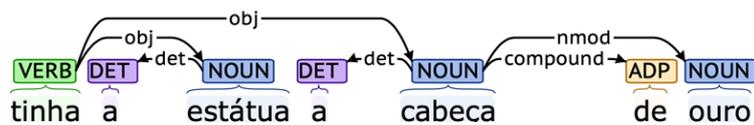


Fig. 8
Análise sintáctica da frase 2

No entanto, apesar de não identificar correctamente o sujeito da frase, esta representação sintáctica pode ser a entrada de uma ferramenta de análise semântica que, usando recursos como dicionários e taxonomias, permite chegar à representação de um termo num grafo semântico como o da Figura 9:

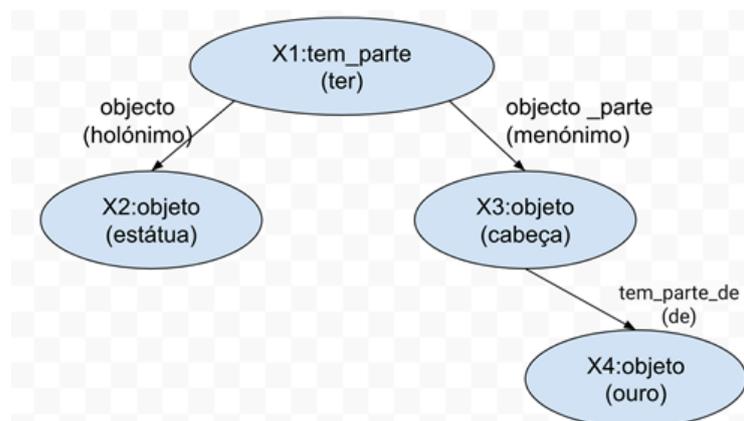


Fig. 9
Grafo semântico da frase 2

A partir deste grafo, podem gerar-se anotações ligadas ao texto na plataforma que auxiliam na detecção, manual ou automática, de similaridades entre textos.

Assim, a partir de ferramentas existentes, como o analisador sintáctico stanza.run, propomos o desenvolvimento de um analisador semântico que use os recursos melhorados no projecto (dicionários e taxonomias adequados ao vocabulário dos textos do Vieira), permitindo extrair e representar alguma informação dos textos e possibilitando a detecção automática de similaridade baseada no seu conteúdo semântico.

Em resumo, a ideia será partir de ferramentas existentes, melhorá-las, tornando-as mais versáteis, de forma a serem úteis neste tipo de texto, e desenvolver novas ferramentas que permitam representar parte do conteúdo semântico dos textos.

4. CONCLUSÕES

O Arquivo HdoF, contendo todos os textos conceptualmente pertencentes à obra, suporta leituras a duas dimensões: vertical (progressão do texto de acordo com o Plano) e horizontal (variantes).

Por outro lado, o arquivo suporta também ferramentas de PLN, para análise comparativa e textual da obra.

Finalmente, o Arquivo HdoF apoia a interacção perito-arquivo, na gestão dos resultados obtidos nas actividades de investigação.

REFERÊNCIAS

1. Banza, A.: P. *Representação perante o Tribunal do Santo Ofício, de Padre António Vieira. Edição crítica e estudo filológico*. Lisboa, Imprensa Nacional – Casa da Moeda (2008).
2. Gonçalves N., Faustino, D., Rito Silva, A., Portela, M.: Monolith Modularization Towards Microservices: Refactoring and Performance Trade-offs. *IEEE 18th International Conference on Software Architecture Companion (ICSA-C)*, pp. 1-8 (2021). Doi: 10.1109/ICSA-C52384.2021.00015.
3. Muhana, A.: *Apologia das coisas profetizadas*. Lisboa, Cotovia (1994).
4. Paixão de Sousa, M. C.: A Filologia Digital em Língua Portuguesa: alguns caminhos. In Gonçalves, M. F., Banza, A. P. (orgs.). *Património Textual e Humanidades Digitais: da Antiga à Nova Filologia*. Évora, CIDEHUS, pp. 113-138 (2013). Disponível em: <https://books.openedition.org/cidehus/1073>.
5. Portela, M., Rito Silva, A.: A model for a virtual LdoD, *Digital Scholarship in the Humanities*, Volume 30, Issue 3, September, 354–370 (2015). <https://doi.org/10.1093/llc/fqu004>.
6. Sérgio, A., Cidade, H.: História do Futuro. In *Obras escolhidas do Padre António Vieira*, IX, pp. 1-160 (1953 [1918]).
7. Rito Silva, A., Portela, M.: TEI4LdoD: Textual Encoding and Social Editing in Web 2.0 Environments. *Journal of the Text Encoding Initiative [Online]*, Issue 8 (2015). <https://doi.org/10.4000/jtei.1171>.
8. Silva, J.Q., Melo, D., Rodrigues, I.P., Seco, J.C., Ferreira, C., & Parreira, J.: An Ontology based Task Oriented Dialogue. *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management, IC3K 2021*, Volume 2: KEOD, Online Streaming, October 25-27, 2021, Book Series: KEOD, pp. 96-107 (2021).
9. Teixeira, C., Rodrigues, I.: Deciphering Latin sentences using traditional linguistic resources. *Digital Scholarship in the Humanities*, Volume 34, Issue 4, 791–805 (2019).
10. Vieira, A.: *Historia do Futuro. Livro Antepimeiro*. Lisboa Occidental, Oficina de Antonio Pedrozo Galram (1718).

AS HUMANIDADES DIGITAIS APLICADAS À PALEOGRAFIA E DIPLOMÁTICA: CONTRIBUTOS INTERNACIONAIS¹²

Ana Pereira Ferreira¹³

CIDEHUS – Universidade de Évora e CH-Universidade de Lisboa

anapsf@uevora.pt

Resumo. Nos dias 6 a 8 de maio de 2021, realizou-se o primeiro congresso internacional sob a temática da Paleografia e Diplomática nas Humanidades Digitais. Com conferencistas de todo o mundo, houve contributos variados que deram a conhecer a aplicação das Humanidades Digitais nas disciplinas de Paleografia e Diplomática, mas também na Sigilografia e na gestão de Arquivos. O propósito da presente comunicação é fazer uma análise, de entre as apresentações realizadas no supracitado congresso, das várias opções disponíveis para o estudo das referidas ciências, verificando as suas mais-valias e pontos fracos, potencialidades futuras e usos presentes no auxílio ao investigador em História. Desde metodologias mais complexas relacionadas com a fotografia, a programas informáticos de transcrição de documentos históricos ou a bases de dados especializadas, procuraremos sintetizar as principais tecnologias disponíveis para o historiador do século XXI.

1. INTRODUÇÃO

Em maio de 2020, teve início um congresso internacional baseado em ciências como a Paleografia e a Diplomática¹⁴. Seria o primeiro congresso internacional organizado pelo CIDEHUS sobre estas matérias e queríamos que fosse um congresso pioneiro. Optámos, por isso, por cingir as comunicações àquelas ciências e à sua integração nas humanidades digitais, cada vez mais essenciais ao investigador em ciências sociais e humanas e, aqui, ao historiador em particular.

¹² Este trabalho é financiado por fundos nacionais através da Fundação para a Ciência e a Tecnologia, sob o projecto UIDB/00057/2020, e fundos internacionais através do Fundo Social Europeu.

¹³Bolseira de doutoramento FCT (SFRH/BD/137506/2018). Doutoranda PIUDHist, investigadora integrada do CIDEHUS – Universidade de Évora e investigadora colaboradora do CH-Universidade de Lisboa.

¹⁴Paleografia e Diplomática na senda das Humanidades Digitais: caminhos e propostas, congresso realizado via Zoom entre os dias 5 e 7 de maio de 2021, organizado em parceria entre o Cidehus-UÉvora e o CH-ULisboa.

Desse colóquio saíram várias comunicações e o nosso contributo na *Jornada de Humanidades Digitais* consistiu precisamente numa análise do que foi apresentado nessa altura e de outras aplicações que as humanidades digitais têm nas ciências referidas.

No congresso *Paleografia e Diplomática na senda das Humanidades Digitais: caminhos e propostas*, foram várias as comunicações que procuraram demonstrar o uso e correlação entre as humanidades digitais e a paleografia, diplomática, sigilografia, agentes da escrita, arquivos e uso do digital na universidade. No geral, estas podem organizar-se em três grandes grupos:

1. Bases de dados de pesquisa/catálogos *online*;
2. Imagens digitais *online*.
3. Técnicas e *softwares* de trabalho.

Cada um destes grupos apresenta um variado número de opções, de onde destacaremos adiante algumas, organizando-as pelos grupos das sessões que marcaram o congresso, já mencionados *supra*.

Porém, importa ter em conta que cada um destes grupos de alternativas digitais apresenta umas quantas vantagens e pontos fortes, mas tem igualmente pontos fracos e desvantagens.

De entre os pontos fortes, nomeadamente para os grupos 1 e 2, podemos nomear aquele que será o mais relevante em tempos de pandemia e restrições de circulação: o acesso remoto fácil a um enorme manancial de informação e documentação. A acrescentar a este ponto e em estreita ligação com ele, uma maior possibilidade de comparação entre fontes, permitida pelo monitor de um computador ou *tablet*, que nem sempre é possível com o documento físico, dadas algumas restrições que arquivos e bibliotecas impõem, como a presença de mais do que um suporte na mesa do investigador.

Quanto aos pontos fracos, destacam-se, nomeadamente: a perda de acesso ao documento físico, de onde resulta, por vezes, a dificuldade ou mesmo impossibilidade de algumas análises mais específicas. A título de exemplo, por vezes no documento original, físico, conseguimos detectar o que se encontra debaixo de rasuras do suporte, nomeadamente quando se trata de tinta pintada por cima da escrita ou de fungos; contudo, tal não é, muitas vezes, perceptível através da digitalização simples do documento.

Outro aspecto que temos de considerar como um ponto fraco ou desvantagem, é alguma possível perda de noção global do documento, fundo, colecção, entre outra catalogação possível.

No grupo 3, sobre as técnicas e *software* de trabalho, podemos igualmente enumerar um conjunto de vantagens e desvantagens, nomeadamente:

Pontos fortes:

- Acesso gratuito, em qualquer local;
- Algumas técnicas de implementação pouco dispendiosas;
- Leitura de documentos por não-especialistas;
- Leitura de documentos em outras línguas e caracteres se o modelo estiver criado;

Pontos fracos:

- Necessidade de criação de modelos para cada escrita;
- Necessidade de correções posteriores à transcrição;
- Necessidade de adaptação às normas de transcrição;
- Algumas técnicas dispendiosas e necessidade de formação especializada.

Parece-nos importante fazer aqui uma breve referência a algumas destas bases de dados, técnicas, *software* que temos à disposição, analisando-as caso-a-caso.

Importa, ainda, considerar que existem muitas mais opções disponíveis que não foram abordadas no congresso, mas a que faremos breve menção.

De referir, também, a volatilidade destas aplicações, o que podemos constatar, por exemplo, numa breve análise a um manual publicado em 2010¹⁵, onde a maioria dos *sites* de projectos lá referidos já não se encontra *online*.-Tal deve-se ao facto de alguns serem simples projectos universitários, sem seguimento posterior no mercado.

2. DIPLOMÁTICA E CODICOLOGIA

Passando a uma análise com base nas sessões do congresso, iniciamos a nossa observação pela Diplomática.

Na verdade, na Diplomática cabem aquelas três vertentes que verificámos supra: a disponibilidade de bases de dados, bancos de imagens de documentos, mas também técnicas e *software* que permitem, de acordo com a linguagem de programação (XML, TEI, DTD, ...) trazer mais valias à Diplomática e à análise dos documentos, permitindo ao investigador um maior e mais rápido acesso, comparação entre documentos e difusão de informação e resultados.

São inúmeros os exemplos que poderíamos dar de recursos de digitalização e bases de dados ao serviço da Diplomática, sendo que enumeraremos apenas alguns, a título exemplificativo¹⁶: ANTT, BNP, *Catalogue of Digitized Medieval Manuscripts, Manuscripta Mediaevalia*, entre outros.

Mas temos igualmente ao dispor, e mais úteis do ponto de vista da análise e sintetização de informação, vários *softwares* ou técnicas de trabalho. Nas técnicas, que nos permitem aceder a informação por vezes impossível de outra forma, destacamos a Fotografia Digital/Imagens Multi-espectro, o recurso a luz ultravioleta ou ainda o recurso a RX. A este respeito, o exemplo do trabalho desenvolvido por Laura Esposito, em que se utiliza a

¹⁵ BRESCIANO, Juan Andrés (2010) - *Ciencias Auxiliares de la Historia. Guía crítica de recursos electrónicos*. [s.l.]: Maat Libros Editorial.

¹⁶ <https://digitarq.arquivos.pt/> - <https://bndigital.bnportugal.gov.pt/explorar-colecoes/> - <https://english.ucla.edu/manuscripts/> - <http://www.manuscripta-mediaevalia.de/#|4>.

fotografia digital e imagem multi-espectro para detectar escrita em palimpsestos¹⁷ que, de outra forma, seria impossível ou, pelo menos, mais difícil detectar.

Quanto a *software*, há a considerar alguns, muitas vezes criados à medida, baseados no recurso a bases de dados. A título de exemplo, o Proyecto Carabela¹⁸, um projecto espanhol que nos permite, num documento, detectar, através de um campo de pesquisa, palavras ou termos que queiramos encontrar na documentação disponível na base de dados, à qual podemos igualmente acrescentar novos documentos, o que facilita muitíssimo o trabalho de investigadores que procuram assuntos muito específicos.

Mas existem igualmente *softwares* que permitem compreender as partes externas de um documento. É o caso do *eCodicology*¹⁹, que, através de linguagem informática e de digitalização de documentação, consegue fazer um processamento de algoritmos e de imagens e extrair metadados a partir da linguagem de programação.

Por fim, outro exemplo-Em Portugal, temos bases de dados e registos de informação para os estudos relacionados com a Inquisição e na investigação de Tomás Lacerda e Israel Aquino, um projecto da Universidade Federal do Rio Grande do Sul criou uma base de dados, a NACAOB, onde fazem registos paroquiais através da leitura dos assentos disponíveis em arquivos e num-conjunto de informação e filtros, conseguindo comparar e analisar a informação de forma mais rápida e profícua.

3. PALEOGRAFIA

No caso da Paleografia, o conjunto de ferramentas que temos ao nosso dispor é mais alargado. No congresso, foram apresentadas duas grandes ferramentas: o *Transkribus* e o *eScriptorium*. Ambos são *softwares* que permitem transcrever documentos, sejam de que época forem. A estes, podíamos juntar outros, como o *Transcripte.exe*, o *GitHub*, o *Genscriber* ou o *D-Scribes*²⁰. A lógica do funcionamento de todos é muito similar e começaram como projectos universitários, tendo o *Transkribus* evoluído já para a *ReadCoop*. Com *layouts* ligeiramente diferentes, ambos necessitam de um *upload* do documento que se pretende transcrever, o qual pode ser carregado para o *software* (que é

¹⁷ Palimpsestos são pergaminhos que foram reutilizados posteriormente, isto é, foi raspado o seu conteúdo original e escrito novo conteúdo por cima.

¹⁸ <http://carabela.prhlt.upv.es/es>.

¹⁹ <http://www.ecodicology.org/index6249.html?id=1&L=2>.

²⁰ <https://readcoop.eu/transkribus/>.

<https://www.escriptorium.uk/>.

<http://www.jacobboerema.nl/en/Freeware.htm>.

<https://github.com/virtualvinodh/scriptanalyzer>.

<http://genscriber.com/genapps/genapps.php>.

<https://d-scribes.philhist.unibas.ch/en/>.

gratuito e pode ser descarregado *online*) e, após estar disponível, o utilizador deve escolher o modelo que pretende que faça a transcrição.

E é aqui que começam os problemas deste tipo de *software*: temos de criar um modelo, isto é, explicar à máquina como deve transcrever o documento. Os modelos podem ser públicos, em que qualquer pessoa pode usar um modelo já existente, elaborado por outras pessoas ou grupos, ou pode ser privado (a maioria) e, neste caso, apenas podem ser usados por quem os criou, o que constitui, logo à partida, um problema. Como tal, principalmente para grafias medievais, é grande, a probabilidade de termos de criar o nosso próprio modelo.

Para criar um modelo, são necessárias a imagem do documento e a transcrição das 10 a 15 primeiras linhas, para que o *software* aprenda a corresponder imagens a letras. A partir daí, o programa conseguirá transcrever textos com grafias iguais às do modelo criado. Tal facto traz-nos outro problema: imaginemos as grafias da chancelaria régia portuguesa entre D. Dinis e D. Manuel I. Nestes cerca de 200 anos - nada de extraordinário, em termos de extensão cronológica, para um medievalista - temos, pelo menos, só entre a documentação régia, quatro grafias diferentes, ainda que todas góticas; isto sem considerar as fases de transição de umas para as outras. Assim, um historiador que analise estes 200 anos de História iria necessitar de pelo menos quatro modelos diferentes e, voltamos a reforçar, estamos a falar somente do âmbito da documentação emanada dos escrivães régios, sem considerar a documentação redigida por tabeliães para instituições particulares, por escrivães concelhios ou por notários apostólicos. A este facto, acresce a posterior adaptação que o utilizador terá de efectuar às normas de transcrição que pretende usar. Para alguém ágil em paleografia, acaba por ser mais prático transcrever de origem os textos. Este *software* será útil, porém, para quem apenas queira saber o conteúdo do documento e não tenha habilidades paleográficas ou para alguém, mesmo com habilidades paleográficas, que tenha de lidar com um documento em paleografia grega, hebraica ou árabe, por exemplo. Ainda assim, tem de pré-existir o referido modelo.

A grande vantagem destes *softwares* está na sua utilização em grandes projectos que utilizem a mesma grafia e que tenham de transcrever centenas de documentos. Além disso, o facto de ter um acesso gratuito também se torna uma mais-valia. Mas, para além destes dois exemplos que mencionámos, existem outros projectos e *softwares*, com funções paleográficas ligeiramente diferentes. É o caso do *DigiPal*²¹ — base de dados e fonte digital de manuscritos — desenvolvido pelo *King's College* e que tem como propósito utilizar as humanidades digitais para estudar as várias tipologias de grafias: o *software* reconhece na imagem as várias letras e consegue catalogá-las de acordo com os tipos de grafias conhecidas (por exemplo, unciais, carolíngias, insulares, góticas...). O *software* permite ainda pesquisar, entre os outros documentos da sua base de dados, letras iguais.

Também diferente é o propósito do *Graphoskop*²². Neste caso, foi inicialmente desenvolvido como *plug-in* e permite uma maior comparação de informação, seja de documentação manuscrita, seja de *fac-similes*. O seu intento prende-se com a identificação do surgimento

²¹ <http://www.digipal.eu/>.

²² <http://www.palaeographia.org/graphoskop/index.htm>.

de um determinado fenómeno gráfico e desenho da sua curva de difusão, a fim de poder questionar mais claramente as causas e métodos dessa evolução.

Verificamos, desta forma, que as potencialidades das humanidades digitais aplicadas, principalmente, à paleografia, podem ser imensas e permitir, com pouco esforço para o investigador, comparar dezenas, centenas, milhares de documentos e traçar perfis e conclusões bastante mais assertivas que “a olho nu” pela observação e comparação pontual de documentos “físicos”.

4. SIGILOGRAFIA

No caso da sigilografia, voltamos a ter dois tipos de influência das humanidades digitais: as bases de dados e catálogos *online* e as técnicas de tratamento e preservação.

No caso das primeiras, são várias as bases de dados que existem em vários países. Normalmente têm uma imagem de qualidade do selo, a sua descrição externa e interna, isto é, as suas características em termos de formato, cor, tipo de material, fita e o seu conteúdo em termos de frases, imagens, simbologia. Frequentemente têm a indicação de outros documentos que surgem com a mesma imagem, o que permite uma análise completa.

Exemplos destas bases de dados são a portuguesa *Sigillvm*, mas também outras bases de dados europeias como a *DigiSig* ou a *Sigilla*²³, entre outras.

De forma complementar e menos frequente, temos as técnicas de tratamento e preservação. Neste caso, tivemos presente de igual modo, no congresso, uma apresentação de um projecto levado a cabo pela Universidade de Granada²⁴, que recorre à conjugação do uso de fotogrametria *Structure from Motion* (SfM) e técnicas de iluminação com Sistema de Informação Geográfica (SIG). Por meio do uso de técnicas de fotogrametria digital, procura-se obter modelos tridimensionais de alta precisão a partir de fotografias. Depois de gerados estes modelos tridimensionais, o *Reflectance Transformation Imaging* (RTI) simula virtualmente o selo, obtendo detalhes macroscópicos do relevo através da manipulação da luz. A combinação das técnicas de RTI e SfM permite alcançar melhores resultados, ultrapassando algumas limitações decorrentes do seu uso isolado.

Desta forma, consegue-se, não só uma perfeita visualização, mas igualmente uma forma fenomenal de conservação futura. No caso do projecto apresentado, trata-se de selos de cera sobre papel e não de selos pendentes, o que permite, muitas vezes, uma melhor compreensão dos elementos que constituem o selo. De referir ainda que esta técnica não implica custos desmesurados, na medida em que uma câmara fotográfica de um telemóvel

²³ <http://portugal-sigillvm.net/> <http://www.digisig.org/> <http://www.sigilla.org/>.

²⁴GARCIA, Marta Durbán (2021) – “Digital Documentation of Papered Seals By Means Of Photogrammetry And Digital Illumination Techniques”, apresentação oral no congresso “Paleografia e Diplomática na senda das Humanidades Digitais: caminhos e propostas”, 6 de maio de 2021.

permite o efeito que se pretende, desde que o utilizador vá movendo o ângulo em que tira a fotografia, de forma a ser depois processada pelo *software*.

5. AS HUMANIDADES DIGITAIS NO MEIO ACADÉMICO

Outra sessão do congresso dizia respeito ao uso das humanidades digitais nas universidades. Sobre este tópico, as possibilidades são quase infinitas, com tanto que pode ser explorado para tornar as aulas mais atraentes para os alunos e os métodos mais eficazes.

Ainda que não directamente relacionados com o ensino nas universidades, os portais que temos hoje à disposição são um enorme depósito científico, gratuito e acessível *online* a todos os estudantes e investigadores. É o caso de portais como o *Academia.edu*, a *Dialnet* ou o *JStor*, para citar somente alguns. Nesses portais, são colocados artigos, capítulos, recensões e demais materiais, dos quais, na maioria dos casos, se pode fazer *download* gratuitamente, o que facilita o acesso à informação, numa época em que a produção é diária e de grande qualidade, permitindo a jovens estudantes beberem do conhecimento através dos especialistas nas mais variadas áreas.

A par destes portais, onde estão guardados milhares de ficheiros acessíveis através de um *click*, temos portais onde são igualmente disponibilizados por docentes e regentes de disciplinas outros materiais, mais dirigidos e seleccionados previamente para o estudo e de acordo com um programa de leccionação. É o caso do que sucede na base de dados *Theleme*²⁵, um *site* onde os alunos podem encontrar *dossiers* e técnicas de estudo aplicados à História, com manuais, estudos, artigos, exercícios e bibliografia.

Por outro lado, a presença de computadores e projectores nas salas de aulas das universidades, hoje um pouco por todo o mundo, é mais uma fonte de possibilidades de aplicação das humanidades digitais. Actualmente, já não é mais necessário usar imagens fotocopiadas de álbuns de paleografia com 40 ou 50 anos, pode facilmente aceder-se ao portal de um arquivo e projectar uma imagem de um documento, com muito maior qualidade, a cores, e desenvolver, a partir daí, o exercício de leitura e transcrição do documento. Desta forma, vai-se ensinando ao aluno duplamente, a forma de pesquisar e aceder à documentação em arquivos e bases de dados *online*, mas também a leitura e transcrição paleográfica ou análise diplomática ou codicológica de um documento.

Por fim, falemos dos arquivos. Na realidade, aquilo que se possa referir a respeito das humanidades digitais nos arquivos e bibliotecas, já foi referido ao longo destas linhas. Ainda assim, importa salientar a importância de um investimento em boas máquinas de digitalização e a consequente disponibilização *online*, principalmente após o mundo experimentar um longo período de restrições devido à pandemia. Mas, além das imagens *online* e bases de dados, há técnicas que são importantes apostas por estas entidades, permitindo aceder a informação que, de outra forma, estaria dificultada ou mesmo

²⁵ <http://theleme.enc.sorbonne.fr/>.

interdita. É o que sucede, por exemplo, com as técnicas de *Rx*. Temos, a este respeito, o exemplo do projecto REXII, desenvolvido, em 2020, pelo Arquivo Nacional Francês²⁶, que permitiu ao investigador aceder a informação contida em missivas que, de outro modo, seriam de difícil escrutínio. O projecto trouxe à luz do dia a informação escondida nas cartas trocadas entre a rainha francesa Maria Antonieta e o seu alegado amante. Nas cartas redigidas pela rainha, tinham sido riscadas várias linhas, pelo alegado amante. Acedendo a tecnologias *Rx*, que permitem separar a composição química das duas tintas usadas (a da rainha e a do seu correspondente), conseguimos compreender o que havia sido escrito por Maria Antonieta e que, de outra forma, estaria irremediavelmente interdito aos olhos dos investigadores.

Outras técnicas, como a fotogrametria e multi-espectro, tornam-se também essenciais para compreender e preservar documentos e selos, principalmente se estivermos a falar de selos de cera, tão frágeis.

6. CONCLUSÕES

Terminamos estas páginas com algumas breves notas conclusivas. Por um lado, importa realçar a indubitável importância do digital (bases de dados, imagens, *software*), que, em tempos de pandemia, permitiu ao investigador continuar a desenvolver o seu trabalho remotamente, em qualquer parte do mundo; e, por outro, o facto de que, ao assegurar-se a digitalização da documentação, está-se igualmente a permitir a sua conservação e permanência para as gerações vindouras. De frisar igualmente a existência, actualmente, de uma grande diversidade de *softwares*, principalmente para uso na Paleografia, mas que requerem melhorias.

Pela negativa, importa assinalar os riscos que a proliferação destes sistemas digitais trazem para o trabalho dos paleógrafos. O facto de muitos *softwares* terem acesso gratuito aumenta estes riscos para os recursos humanos, ainda que tal seja uma vantagem no âmbito do utilizador. Se este acesso facilitado acontece nos *softwares*, não podemos dizer o mesmo para algumas técnicas mais dispendiosas e que implicam maior investimento, nomeadamente o acesso a *scanners* profissionais.

Ainda assim, foi possível observarmos que existem técnicas relativamente económicas, recorrendo somente a telefones com câmara fotográfica ou a máquinas semi-profissionais para o uso da fotografia multi-espectro.

As Humanidades Digitais têm de começar a ser vistas também por outro prisma. É certo que elas podem ser o futuro das Humanidades e Ciências Sociais, numa sociedade cada vez mais preocupada com o digital e com as ciências exactas e menos com as ciências humanas. Porém, esta adesão das ciências sociais às humanidades digitais traz consigo a necessidade de incorporar novas técnicas e de formar especialistas nas universidades, para que as gerações futuras possam lidar com os desafios que as esperam.

²⁶ <https://rex.sciencesconf.org/>.

REFERÊNCIAS

1. BRESCIANO, Juan Andrés (2010) - *Ciencias Auxiliares de la Historia. Guía crítica de recursos electrónicos*. [s.l.]: Maat Libros Editorial.
2. Sessões do Congresso “Paleografia e Diplomática na senda das Humanidades Digitais: caminhos e propostas”:
Abertura e sessão I: <https://www.youtube.com/watch?v=7vZN1kveT5k>
Sessão II e III: <https://www.youtube.com/watch?v=bpcWDa8yCNo>
Sessão IV: <https://www.youtube.com/watch?v=z7iXghA-5rU>
Sessão V: <https://www.youtube.com/watch?v=CnsRTsLKWr4>
Sessões VI, VII e sessão de encerramento:
<https://www.youtube.com/watch?v=xHnzVXN-q2Q>.
3. Projectos e *software*:
<https://digitalq.arquivos.pt/>
<https://bndigital.bnportugal.gov.pt/explorar-colecoes/>
<https://english.ucla.edu/manuscripts/>
<http://www.manuscripta-mediaevalia.de/#|4>
<http://carabela.prhlt.upv.es/es>
<http://www.ecodicology.org/index6249.html?id=1&L=2>
<http://theleme.enc.sorbonne.fr/>
<https://readcoop.eu/transkribus/>
<https://www.escriptorium.uk/>
<http://www.jacobboerema.nl/en/Freeware.htm>
<https://github.com/virtualvinodh/scriptanalyzer>
<http://genscriber.com/genapps/genapps.php>
<https://d-scribes.philhist.unibas.ch/en/>
<http://www.digipal.eu/>
<http://www.palaeographia.org/graphoskop/index.htm>
<http://portugal-sigillvm.net/>
<http://www.digisig.org/>
<http://www.sigilla.org/>
<https://rex.sciencesconf.org>.

THE MONSOON PROJECT: DIGITAL PERSPECTIVES OF THE HABSBURG *ESTADO DA ÍNDIA*²⁷

Ana Sofia Ribeiro

CIDEHUS – Universidade de Évora, Évora, Portugal

asvribeiro@uevora.pt

Abstract. MONSOON aims to study the internal dynamics of the Portuguese macro-region of *Estado da Índia* in its various dimensions in a holistic and complex way during the period in which Portugal was part of the Hispanic Monarchy (1580-1640). The challenge is to combine, at the same time and in the same instruments of analysis, the formal and informal presence of the Portuguese in the Asian seas since literature usually analysed the two perspectives separately [1] [2]. MONSOON intends to test a different methodological approach to the study of regions that were part of European empires in the early modern period based upon massive narrative collection of sources. It expects to develop a flexible digital approach that could be applied to other European historical spaces and documentary sources. To achieve these goals the team will use a set of documents known as *Livros das Monções*. The information extracted from the sources will be analyzed according to three variables – agents, events, spaces – and using digital humanities tools: digital ontological semantics applying Natural Language Processing techniques, social network analysis and digital mapping.

1. INTRODUCTION

MONSOON aims to study the internal dynamics of the Portuguese macro-region of *Estado da Índia* in its various dimensions in a fully integrated way during the period in which Portugal was part of the Hispanic Monarchy (1580-1640). Political affairs, socioeconomic realities, cultural and religious matters, relationships with autochthonous neighboring powers and communities, as well as the relations with competing European powers are key variables to be addressed.

The *Estado da Índia* constituted the most complex overseas Portuguese set of territories, encompassing a geography as wide as the Indian Ocean borders. Literature underlines that this political unit also should be perceived within the geographical scope of Portuguese informal presence in Asia, that is, the places where free-riding Portuguese individuals constituted significant communities (mostly mestizos), albeit there were nor Portuguese formal structures of any sort, nor under Portuguese jurisdictional power [3].

²⁷ This work is financed by national funds by the Fundação para a Ciência e a Tecnologia, within the project UIDB/00057/2020.

Combining the study of both realities and intertwining different locations in the same study are some of the great challenges that this project faces ahead. During the Habsburg rule of Portugal, and even if the administration of the Iberian overseas empires remained separated after 1581, historians traditionally underline a critical situation for Portuguese position in Asia, due to the extension of the Dutch-Spanish conflict to overseas territories and the English competition for the control of Asian commodities trade destined to Europe, together with several disruptive events and conflicts with local authorities [1] [4]. On the other hand, this juncture has hardly ever been fully taken into consideration encompassing the *Estado da Índia* as a whole region (except Murteira's work [5]). Instead, micro-regional studies for Macao, Japan, Ceylan, or Persia have been the focus of attention. An aprioristically athenatic approach is yet to be explored and MONSOON wants to fulfill it.

The historiographical discourse was coeval with the own pace of events evolution. The most relevant modern historiography mostly dates from the 1960s and goes until the 1990s, when solid syntheses of the Portuguese Eastern settlements were elaborated, encompassing political, administrative, economic, social, religious or cultural affairs [6] [1] [7] [8]. They were the first to fully understand the meaning and concept of the *Estado da Índia* as a set of littoral locations around the Indian Ocean, yet discontinuous in space (and in time) and comprising territories with diverse juridical status (conquests, factories, fortresses), or different types of relationships with local authorities (conflict, diplomatic alliances, vassalage, or pure segregation). Yet their focus was mostly on formal territories that were at a certain time administered by the Portuguese. Some of these works were part of an Eurocentric approach, still prevailing in the Dutch and British historiographies, claimed the precocity of the Portuguese in the process of European overseas expansion and empire-building. They were the first Europeans to arrive in the Indian Ocean, provoking a disruptive effect on an almost immutable world, due to its superior naval force and warship power. For authors like Boxer [6] or Panikkar [9], the history of Asia could be divided in two periods: before and after Vasco da Gama.

Since the 1980s, there was a reaction against this discourse and soon the relative ascendancy of the Portuguese in the sixteenth and seventeenth centuries has been questioned. In fact, Chaudhuri [10], Das Gupta [11], Frank [12] or Mukherjee [13] have proved how the arrival of the first Europeans represented a minor role in a world already well interconnected both economically or socially, complex and cosmopolitan where Muslims, Hindus, Christians, Persians, Chinese or Malays already intermingled and circulated. In fact, since the beginning of the twenty-first century, Portuguese historiography has revealed how the organization of Eastern Portuguese societies in conquered territories has become possible by the integration of autochthonous populations, since they were key elements for establishing economic and political (essential) relations with the bordering regions [14].

Following this perspective, MONSOON wants to approach the Portuguese presence into the functioning of this Indian Ocean world, considering the Portuguese agents as one more element in the existent local immense human diversity. In the last two decades, literature has realized that an immense part of the Portuguese presence in Asia during the Early Modern period took place outside the boundaries of the formal territories of the *Estado da Índia*, in two different vectors. The first relates with the establishment of more or less small communities of Portuguese or Portuguese mestizo descendants in spaces like Ayutthaya or

in some port cities around the Bengal Bay, e.g. [3] [15]. The second reinforces the role of private agency not only in establishing alternative economic connections, but also stressing the action of individuals as motors of empire-building. Individuals not only acted side by side and within the strategies of the Crown, but also simultaneously acted outside the legal frameworks or explored, at their own cost, partnerships outside the influence of the *Estado da Índia* [16]. But until now, formal and informal «empire(s)» have been considered separately by historiographical discourse. MONSOON wants to shorten this gap by combining the analysis of the two types of Portuguese presence in Eastern seas assuming their complementarity.

If the period of the «Habsburg» Portuguese empire in Asia has been neglected by historiography for a long time, since the last decade of the twentieth century it has been subject to a profound and innovative reflection regarding the place it had in the strategies of the overall Hispanic Monarchy. Young and senior researchers have highlighted relations with the new competing European powers in Asia or the loss of strategic positions for the control of spice trade not only in the Persian Gulf, but also in Ceylon, Melaka, or Japan [5] [17] [18]. Although some punctual considerations have been taken on complementarities between Portuguese and Spanish presence in Asian seas [19] [20], an integrative and athenatic vision of this *Estado da Índia* is still to be pursued. MONSOON aims to fulfill this hole in historiographical literature which we believe can only be made using computational processes of source analysis.

MONSOON team, disposing of the expertise in dealing with digital humanities, expects to produce an holistic perspective of the Portuguese presence in the East during the Iberian political union of crowns intermingling the internal functional of the State of India (considered here from Eastern coast of Africa until Macao, Japan or Timor), the strategies of central administration in metropolitan court and the informal presence of Portuguese influence, without forgetting the relationships with the formal European competitors and indigenous communities.

The team is composed of historians and computer scientists, mostly from Évora University. Trained early modern historians are Ana Sofia Ribeiro from Évora University (the principal investigator) and Liliana Oliveira from the University of Porto. The co-principal investigator, Renata Oliveira is an expert on digital humanities and shares a computer science background with Paulo Quaresma, Marlo Souza and Anderson Sacramento. These last two belong to the Universidade Federal da Bahia in Brazil. The team is concluded with Ivo Santos (Évora University), an archeologist, but an expert on digital humanities and holds postgraduate training in Computer Science.

The following pages seek to present the project to the public. More than a history project, Monsoon intends to assert itself as a digital humanities paradigm in the field of history, building up a digital framework for historical analysis. The following section enunciates the main scientific goals of the project and presents the main documental corpora used by the project to the analysis of the *Estado da Índia*. The third section concretizes the main features of the digital methodology pursued by MONSOON, underling how this choice necessarily derived from the scientific and historical goals of the project. The section also evidences the main challenges faced by the team in the field of the digital humanities.

2. HISTORICAL THEORETICAL FRAMEWORK

2.1. HYPOTHESES AND GOALS

The main hypothesis of MONSOON is that considering casuistic relationships between historical events of diverse typology is insufficient to fully interpret the complexity of relationships between the formal presence of the Portuguese Crown and the agency of individual action and interests [16], namely in a period of reordering and change as when the Portuguese East was under the Habsburg rule. On the other hand, understanding the complexity of the *Estado da Índia* without an in-depth analysis of its internal functioning, that is, without considering the Portuguese presence in the vast social, political and economic Asian equilibria and as just another player in the sixteenth and seventeenth centuries is not possible. Knowing that the Indian Ocean is the first sea where long distant exchanges of all sorts were established, the project intends to avoid Eurocentric assumptions and civilizational superiority ideas considering the dialectics between people from diverse backgrounds. To test the veracity of such hypothesis the project proposes to depart from a macro point of view by developing an insight which comprehends the Indian Ocean in all its regional subsystems and local particularities, overcoming solid casuistic syntheses and disposing of the most recent global approaches. Such perspective underlines that not only the local is “the reception ground where global phenome impact”, it is also where an individual scale of human intervention is possible [21]. Despite its advantages, this view is often fragmented, as Pearson underlines [22].

By pursuing this general problem, MONSOON intends to test a different methodological approach to the study of regions that were part of European empires in the early modern period based upon massive narrative sources. Although the project will implement this framework concerning the eastern part of the Portuguese empire, it expects to develop a flexible digital approach that could be applied to other European historical spaces and documental sources, such as the Dutch Dagregisters, for example.

MONSOON goals are to:

- i. Understand the historical junctures of social, economic and political systems of the *Estado da Índia* during the Iberian Union, in a complex and integrated way.
- ii. Study simultaneous formal and informal systems of Portuguese presence in Asian seas and how they were related.
- iii. Comprehend communication mechanisms between Portuguese local governments, the authorities of the viceroyalty and the Portuguese and Hispanic metropolitan authorities.
- iv. Build up a computational hierarchical analysis of agents, spaces and events.
- v. Reconstitute relations between individuals, events and spaces.
- vi. Comprehend relationships between the Portuguese State of India and other imperial Iberian spaces, to infer the integration of the viceroyalty into the Hispanic Monarchy empire.

- vii. Establish formal and informal relations with Asian communities and other competing European powers in the region.

2.2. THE MONSOON BOOKS

To achieve such goals the project will use a set of documents known as *Documentos Remetidos da Índia* or *Livros das Monções* (*Monsoon books*). These books collect letters exchanged between the monarchs and the Portuguese government councils and India viceroys where all types of affairs concerning from Eastern Africa to Japan were discussed. The use of this collection is paramount to understand the internal dynamics of the Portuguese *Estado da Índia* until the nineteenth century. In fact, they are considered the core of the documents produced by Portuguese authorities in Asia. The fact of being a type of documental corpora concerning all types of issues makes the *Monsoon Books* unique and a privileged lab for building a new analytic model and approach to understand internal dynamics of colonial empires macro-regions.

The collection is dispersed between the National Archive of Torre do Tombo in Portugal and the Historical Archive of Goa, in India. However, due to its vast volume MONSOON has chosen to work with a small set, in order to implement a stable in-depth automatic methodology which will enable a future analysis of the collection for the whole time period. To explore and build a solid computational structure of analysis the team will use 5 of the 10 published volumes of *Monsoon Books* transcriptions of the first 20 volumes remaining at Torre do Tombo (1605-1619) [23]. Although available in a digital format at the National Library of Portugal, these copies of the *Monsoon Books* were made using an image format. Therefore, they demand a costly and time-consuming process of conversion to a digital text-format towards the accomplishment of our goals. Although ordinarily used for all thematic studies concerning Asian parts of the Portuguese empire, these sources have never been analyzed as a whole body, and even less using computational technology.

3. A DIGITAL HUMANITIES METHODOLOGY

MONSOON expects to develop a holistic approach to internal historical dynamics of *Estado da Índia*. The integration of multiple variables consists in two distinct approaches: one quantitative, by evaluating the most frequent words, and therefore identifying themes and concepts present in the *Monsoon Books* in determined timeframes accomplishing with resorting methods of Natural Language Processing; another, by establishing relations within the three variables under study and between each other. The analytic variables are agents involved, events and geographical locations. They will be identified using text annotation techniques.

Digital humanities tools are the most suitable methods to such accomplishment. MONSOON will use different digital approaches. The technical challenge is to combine the four of them:

i.) Ontology based semantics: by applying a system of automatic classification of the *Monsoon Books* data into the three analytical variables to make clear the relevant topics, as well as pointing out their properties and relationships. For that we will use software for text annotation, using machine learning processes that will gradually and automatically annotate the text according to the chosen variables, based on manually annotated examples.

ii.) Automatic event extraction acknowledgment: an event detection system should be able to identify whether the sentences contain events of interest by means of the identification of event trigger terms (event identification) and classify them into specific event types (event classification). We will have the collaboration of Computer Science researchers from the Universidade Federal da Bahia who have developed well trained event extraction systems [24].

iii.) Digital mapping: both general and thematic GIS projections of spatial data extracted from *Monsoon Books* will help to understand specific functions and roles of each location.

iv.) Social network analysis (SNA): SNA will permit reconstituting a system of relationships between agents, events, and places, creating hierarchies. Applying SNA structural statistics enhances observation of the resilience of this overseas subsystem. This will be done according to the results provided by text annotation and event detection.

3.1. WORKFLOW

To comply with this approach the project's team must be able to attend to specific challenges that each step of the workflow comprises. First, it must create a system for automatically transforming the current PDFs of the *Monsoon books* from an image format to a text format. Since we are dealing with five volumes of around 300 printed pages it will be a time-consuming task, although mandatory to perform the following tasks. It can be done by simply using a PDF editor and migrating the text into a TXT file. Nevertheless, a process of text normalization should be pursued. It will be made in two distinct aspects: first, by editing the text abolishing possible, yet likely, extra blank spaces or indented text; second, by unfolding the abbreviations that the original paleographic transcription maintained. Although the team should not annotate abbreviated text, the Portuguese language of the sources will remain as truthful to the documental sources as possible. Apparently simple this task is mandatory, since these texts are the material resource which enables pursuing a digital methodology for the analysis of internal dynamics of early modern European colonial spaces.

The second step focuses on applying machine learning text annotation techniques to *Monsoon Books*, extracting the information and relationships regarding the three studied variables: agents, events, and locations. It is based on Natural Language Processing

techniques, an artificial intelligence-driven process of making human Natural Language decipherable to software which will then permit analysis and interpretation of a massive set of data.

This is a three-step process. One, the team will have to focus on a decision-making phase. It must experiment with an immense range of text annotation software and clearly and accurately decide what text to annotate, restricting it to the three variables under study and their relations. It will determine which text labels to be created, as well as the attributes considered in each one, designing the annotation guidelines for the project. The second step consists in humanly annotating the text. Since the process of machine learning used by softwares rely on training data (initially manually annotated text samples), the team must provide the information to the machine. The team will apply a technique called Named Entity Recognition (NER). NER consists in tagging our variables or entities within the text. The repetition enables its automatic recognition and their future extraction for further analysis, for instance how many times those entities occur within the dataset. The process of machine-learning requires continuous monitoring from the team members, both by accessing the quality of historical data extracted by the machine, both by accessing the feasibility of the Artificial Intelligence process to the challenges evoked by using historical texts. They often present the same word in several different ways, for instance. Therefore, it is necessary to ensure the accuracy of the learning process, and it could comprise the adaptation of the chosen tool to the reality of working with historical sources. The third moment will be the process of entity linking, that is, connecting the entities tagged in the text with each other in order to obtain relationships between entities that can be extracted.

The project will also develop an already existing system of event identification and classification, developed by Anderson Sacramento and Marlo Souza, by adapting it to narrative historical corpora. The goal of event identification and classification is to detect the event mentions of target event types in plain text. Given an input text, an event detection system should be able to identify whether the sentences contain events of interest by means of the identification of event trigger terms (event identification) and classify them into specific event types (event classification). The system TEF (text event frame extraction) was trained on an enriched TimeBankPT corpus with event types from the FrameNet project, using deep neural networks and contextualized word embeddings from a Portuguese BERT (Bidirectional Encoder Representations from Transformers) model.

First, we need to adapt it to the ancient Portuguese language. Secondly, the team will discuss which new word triggers and new event types emerge from sample data. Then, we work on the arguments, that is, the information that should automatically be aggregated to each event, especially agents, one of the projects' main variables of study. It is a less time-consuming form of text annotation. If events are automatically identified and extracted to different analytic software, text annotation can mostly focus on agents and places.

Other challenge the team will face within the field of digital humanities is to prepare and adapt textual data to different digital methods of analysis. This task consists in designing digital support to organize textual data in order to perform data analysis. Since each tool usually needs to prepare data in a different way, the challenge of the team is to potentiate a form of unique data organization to supply each one of the used tools: semantic analysis, network analysis and geo-referenced analysis. For such a purpose a digital common

platform should be implemented. This task is not easy to accomplish and will be developed by a trial-error approach. We will try to connect different existing software using the computational skills of the team, in such form that obtained results in each step of the process can migrate between each other and be integrated in a common digital platform. In such a way, MONSOON seeks to comprise with the FAIR principles of scientific data by making available to the public the preliminary results obtained by this Exploratory Project [25].

4. CONCLUSION

MONSOON intends to create a model of analysis for the European overseas empires before the “Colonialism” period (19th-20th c.), that enhances the understanding of the complexity of relations between different-origin people and helping to stress the relativity of European position in Asian seas. Computational instruments are the most indicated to deal with a massive volume of textual information and to extract regularities, exceptions, changes, and organized patterns of relationships. To be feasible, the project focuses on a small chronological interval to foreclose an in-depth analysis. It combines text annotation techniques, like NER and event extraction, and search for specific solutions to project extracted events, locations and individuals in network graphs and geographical maps.

The team is aware of the multiple technical challenges facing each task of the workflow. And MONSOON should be perceived as the first step to consolidate a method that is not being further explored by historians on the contrary of other humanities fields, such as linguistics. Once the method is consolidated and validated by peers, MONSOON will try to obtain a larger financing to apply it to the extended chronology of the Habsburg domain in Portugal and will try to establish partnerships with foreign universities where the field of Dutch or British Empires in Asia are core areas, so the team could work with similar narrative sources. Crossing the different documental corpora should enable a richer knowledge of the Asian seas during the sixteenth and seventeenth centuries. In the future, MONSOON intends to apply for financing to digitalize the *Monsoon Books* kept in Goa that are truly in danger due to problems in the documents’ conservation. In this sense, the project results will not only be disseminated across the scientific community but will also become available to the community since one of our main goals is to contribute to strengthen ties and maintain a common heritage between Portugal and Goa.

REFERENCES

1. Subrahmanyam, S.: *The Portuguese empire in Asia, 1500-1700: a political and economic history*. Longman, London/ New York (1993).
2. Polónia, A.: Indivíduos e redes auto-organizadas na construção do império ultramarino português. In: *Economia, instituições e império. Estudos em homenagem a Joaquim Romero Magalhães*, ed. Garrido, A., Costa, L. F., Duarte, L. M., pp. 349-371. Almedina, Coimbra (2012).
3. Hespanha, A. M.: *Filhos da terra. Identidades mestiças nos confins da expansão portuguesa*. Tinta da China, Lisboa (2019).
4. Disney, A. R.: *Twilight of the pepper empire: Portuguese trade in Southeast India in the early seventeenth century*. Harvard University Press, Cambridge, MA (1978).
5. Murteira, A.: *A navegação portuguesa na Ásia e na Rota do Cabo e o curso neerlandês, 1595-1625*. Lisboa, Ph diss. NOVA University of Lisbon (2016).
6. Boxer, C.: *The Portuguese seaborne empire*. Hutchinson and Co, London (1969).
7. Godinho, V. M.: *Os descobrimentos e a economia mundial*. 4 vols. Presença, Lisboa (1963-71).
8. Thomaz, L. F. R.: *De Ceuta a Timor*. Difel, Lisboa (1998).
9. Panikkar, K. M.: *Asia and Western dominance: a survey of the Vasco da Gama epoch of Asian History, 1498-1945*. George Allen & Unwin, London (1953).
10. Chaudhuri, K. N.: *Trade and civilization of the Indian Ocean from the rise of Islam to 1750*. Cambridge University Press, Cambridge (1985).
11. Das Gupta, A. *The world of the Indian Ocean merchants, 1500-1800: collected essays of Ashin Das Gupta*. Oxford University Press, New Delhi (2001).
12. Frank, A. G.: *ReOrient: global economy in the Asian Age*. University of California Press, Berkeley, CA (1998).
13. Mukherjee, R. (ed.): *Networks in the first global age, 1400-1800*. Primus Books, Hyderabad (2011).
14. Xavier, A. B.: *A invenção de Goa. Poder imperial e conversões culturais nos séculos XVI e XVII*. Imprensa de Ciências Sociais, Lisboa (2015).
15. Flores, J.: Relic or springboard? A note on the rebirth of Portuguese Hugli, ca. 1632-1820. *The Indian Economic and Social History Review* 39, 381-395 (2002).
16. Antunes, C., Polónia, A.: Introduction. In: *Beyond empires: global, self-organizing, cross-imperial networks, 1500-1800*, Antunes, C, Polónia, A. (eds.). Brill, Leiden, 1-11 (2016).
17. Biederman, Z.: *The Portuguese in Sri Lanka and South Indian. Studies in the history of diplomacy, empire and trade, 1500-1650*. Harrassowitz Verlag, Wiesbaden (2014).
18. Pinto, P.: *The Portuguese and the Straits of Melaka, 1575-1619: power, trade and diplomacy*. NUS Press, Singapore (2012).
19. Sousa, L.: *The Early European Presence in China, Japan, The Philippines and Southeast Asia, (1555-1590) – The Life of Bartolomeu Landeiro*. Fundação Macau, Macau (2010).
20. Ribeiro, A. S.: *Depending from the other. The role of Hindustan businessmen in the Treasure of the Portuguese India State and their interactions with Portuguese merchants (1580-1640)*. (in press).
21. Polónia, A.: Think globally, Act locally. Environmental history as global history in the first global age. *Asian Review of World Histories* 3/1, 59-80 (2015).

22. Pearson, M. N.: *The Portuguese in India*. Cambridge University Press, Cambridge (1987).
23. Patto, A. B. (dir.): *Documentos remetidos da Índia ou Livros das Monções*. Tomos I-V. Academia Real das Ciências de Lisboa, Lisboa (1880-1935). Rego, A. S. (dir.): *Documentos remetidos da Índia ou Livros das Monções*. Tomos VI-X. Academia Real das Ciências de Lisboa, Lisboa (1974-1982). [<http://purl.pt/26875/4/>].
24. Sacramento, A., Souza, M.: Joint event extraction with contextualized word embeddings for the Portuguese language. In: *Brazilian Conference on Intelligent Systems*, 496–510. Springer (2021).
25. Wilkinson, M., Dumontier, M.; Aalbersberg, I. et al.: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (2016).

PLN EM ARQUEOLOGIA: O MEGALITISMO PORTUGUÊS COMO ESTUDO DE CASO²⁸

Ivo Santos
CIDEHUS - Universidade de Évora
ifs@uevora.pt

Resumo. Por natureza, parte da investigação em Arqueologia é destrutiva, tornando o registo arqueológico (a recolha de dados) fulcral. Neste trabalho, pressupõe-se ser fundamental que o registo inclua a relação entre todas as partes e permita compreender, por homem e máquina, a relação semântica entre os dados existentes. Apresentam-se aqui os primeiros passos de um projeto de doutoramento que procura aplicar PLN e Extração de Informação Arqueológica a partir de um caso de estudo: o Megalitismo.

1. INTRODUÇÃO

Este trabalho faz parte de um projeto de doutoramento que procura aplicar métodos de Processamento de Linguagem Natural para Extração de Informação (EI) arqueológica em língua portuguesa. Apresentam-se aqui os primeiros passos para a Extração de Informação de um caso de estudo: o Megalitismo.

Antes de se iniciar um trabalho arqueológico é necessário obter determinadas informações para caracterização do objeto de estudo, por exemplo: uma listagem dos sítios arqueológicos de um período cronológico e/ou em determinada região; em que sítios foi exumado um artefacto de uma tipologia em concreto (simplesmente por ser o alvo do seu estudo, ou porque a tipologia indica um período cronológico, ou uma ocupação por um grupo com características culturais distintas); ou quais os sítios arqueológicos cuja arquitetura tem determinadas características (por exemplo, a forma da planta ou mesmo as dimensões), entre muitas outras informações.

Quando não existe uma ferramenta para tal, como no caso português, a compilação desta informação pode ser obtida através de bases de dados de gestão pública (de entre as quais se destaca o Portal do Arqueólogo [11]), ou pelo contacto com os pares, ou pela verificação de possíveis referências na bibliografia. Os estudos, em geral, são naturalmente condicionados à exaustão desta recolha e o esforço para realizar estudos supra-locais é progressivamente maior. Consequentemente, quando não há partilha de dados, a comunidade multiplica-se em esforços isolados com o mesmo objetivo. Em última instância, podemos pressupor a existência de dificuldades na reprodutibilidade das

²⁸ Este trabalho é financiado por fundos nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto UIDB/00057/2020.

conclusões, pela destruição do objeto de estudo, mas também pelas dificuldades inerentes ao acesso aos dados.

Por natureza, parte da investigação em Arqueologia é destrutiva, tornando o registo arqueológico (a recolha de dados) fulcral e muito valorizado pela generalidade dos investigadores. Idealmente, tudo o que for passível de registo deve ser registado. Não é nosso objetivo elencar tudo o que se pode registar em Arqueologia, mas vamos resumir com o contexto. Isto inclui desde o próprio contexto arqueológico ao contexto geográfico, seja o contexto político dos trabalhos arqueológicos, referências históricas, a relação entre camadas estratigráficas, a relação entre estas e os artefactos, entre muitos, muitos outros. O contexto no seu todo é, assim, essencial para a investigação arqueológica.

Concordamos com Cripps em que *without a data model capable of adequately describing not only the archaeological data contained within records but the archaeological processes used to generate records, any database is limited in terms of its capabilities and suitability for reporting, assessment and analysis and ultimately its suitability as an archive* [10]. Assim, do ponto de vista da informação, consideramos que se torna fundamental que o registo inclua a relação entre todas as partes e compreenda a relação semântica entre os dados existentes. Idealmente, uma base de dados para gestão arqueológica beneficiária da possibilidade de realizar pesquisas com booleanos (por exemplo, a presença / ausência de determinado artefato), mas também pela caracterização das relações entre os dados e, indubitavelmente, pela possibilidade de aplicar mecanismos de inferência automática a partir destes mesmos dados.

2. OBJETIVOS

O objetivo principal do trabalho aqui apresentado é testar o mapeamento de informação para um modelo de dados alinhado com o CIDOC-CRM [4], conforme anteriormente proposto [18]. Este modelo procura responder aos requisitos da comunidade para uma ferramenta de gestão em arqueologia e, assim, é fundamental cumprir com sucesso outros objetivos:

- constituir um *corpus* a partir da documentação existente;
- adaptar os dados arqueológicos recolhidos aos princípios FAIR (*Findable, Accessible, Interoperable and Reusable*);

De entre a informação a recolher destacam-se:

- elementos que permitam caracterizar a arquitetura do monumento e as suas partes constituintes;
- listagens dos artefactos recolhidos;
- bibliografia associada;
- referências à existência de outras análises (arqueociências) para posterior inclusão.

3. TRABALHOS RELACIONADOS - BREVE SÍNTESE

Relativamente ao Megalitismo, por exemplo, Leonor Rocha considera que o Megalitismo funerário deve ser visto de uma perspetiva alargada, independente de divisões administrativas [20] enquanto Ana Catarina Sousa critica os estudos de Megalitismo regional pela sua tendência para inserir dados da população como se fosse uma caracterização geográfica, não incorporando os dados na explicação da evolução das práticas funerárias e separando-os das linhas interpretativas globais [21].

No que concerne à recuperação de informação, essencial para os nossos objetivos, devemos salientar projetos como *Archeotools* [16,15] e STAR [1] que aplicaram EI a partir de relatórios de trabalhos arqueológicos. Por outro lado, entre as iniciativas internacionais para aplicar PLN, destacam-se trabalhos como: [3], os resultados do projeto ARIADNE [1], e noutras línguas, como o holandês [19,14] ou italiano [6].

Noor et al. defendem que abordagens recorrendo a ontologias em Arqueologia não são sempre utilizadas devido a requerimentos específicos e limitações dos próprios projetos arqueológicos [7]. Contudo, a nível europeu, têm surgido vários projetos estruturantes, como, por exemplo, o ARIADNE [2,22].

Num trabalho anterior, propusemos um modelo de dados alinhado com o CIDOC-CRM em conjunto com o *software* de acesso aberto ARCHES para construir um atlas digital para o Megalitismo [18]. Nesse trabalho, sintetizamos os vários conceitos relacionados com o Megalitismo da Península Ibérica e identificamos a utilização de conceitos importados e alguns sinónimos. No que concerne aos trabalhos académicos relacionados com o Megalitismo, verificou-se que, muitas vezes, não são disponibilizados os dados originais, para reutilização e confirmação de interpretações, limitando drasticamente o processo científico.

Desse modo, considerou-se que o estudo do Megalitismo deve ser baseado em dados normalizados, independentes dos conceitos existentes, e acreditamos que a utilização de dados normalizados em conjunto com modelos 3D enriquecidos com dados semânticos permitiria uma análise supranacional, cálculo automático de índices e impulsionaria a partilha de informação entre os *stakeholders*[18].

4. OS DADOS

Para constituir um *corpus* do Megalitismo através de métodos de PLN para Extração de Informação arqueológica em língua portuguesa, procedemos à caracterização das principais fontes de informação (relatórios de trabalhos arqueológicos, teses académicas e bibliografia especializada).

De entre estas fontes, o Portal do Arqueólogo (ferramenta estatal de gestão de dados

arqueológicos) destaca-se por alojar informação estruturada, assim como breves descrições e relatórios de trabalhos arqueológicos. Esta decisão deve-se ao facto de as características formais dos relatórios se encontrarem regulamentadas e, assim, possivelmente formarão um conjunto de dados mais apropriado aos nossos objetivos. Estes relatórios constituem, no entanto, um *data pit* de características desconhecidas.

4.1. PORTAL DO ARQUEÓLOGO

O Portal do Arqueólogo não tem um *Thesaurus* diretamente ligado à própria página, no entanto, surge referência, em página introdutória da Direção Geral do Património Cultural (DGPC), para um PDF com o título *Listas de termos para preenchimento dos campos da base de dados Endovélico*, que propõe 158 tipologias de sítios [8]. No que concerne ao Megalitismo, verificou-se a utilização de termos que se poderão considerar sinónimos (por exemplo, Dólmen e Anta), mas também a utilização de nomenclatura ambígua (como é o caso de Monumento Megalítico). Por exemplo, qualquer Anta e Menir podem ser classificados como Monumento Megalítico. No entanto, na própria base de dados, são utilizadas 9 tipologias que não estão presentes no *Thesaurus*, com destaque para a introdução de Anta/Dólmen.

Em termos formais, a informação do Portal do Arqueólogo encontra-se distribuída em três conjuntos principais: sítios, trabalhos e projetos. Um projeto pode ser constituído por vários trabalhos e um trabalho pode referir-se a um ou mais sítios.

No momento de redação deste trabalho, não existe uma forma pública de recolher a informação do Portal do Arqueólogo. Conforme a ficha técnica: Estes conteúdos não podem ser copiados ou reproduzidos, a não ser quando no âmbito do estritamente necessário à respectiva visualização *online* [12].

É possível visualizar em mapa a localização exata dos sítios, sem coordenadas, mas os utilizadores registados obrigam-se à confidencialidade das informações sobre a localização dos sítios arqueológicos [11] e é possível aceder a alguns relatórios digitalizados, no entanto, apenas é permitida a sua leitura página a página.

4.2. CARACTERIZAÇÃO

Em Junho de 2021, existem 36275 registos de sítios arqueológicos e 39947 trabalhos, dos quais 5855 têm uma ligação para um relatório em formato digital. Verificou-se, no entanto, que 1151 das ligações para os relatórios não funcionam, por erro.

Quanto ao Megalitismo, identificaram-se 4640 registos de sítios (cerca de 12.8%), 5494 trabalhos e 826 relatórios. Porém, 107 relatórios encontram-se inacessíveis. A tipologia

Anta/Dólmen é a mais comum nos registos, mas, por outro lado, a única ocorrência de Dolmen possivelmente corresponde a um erro.

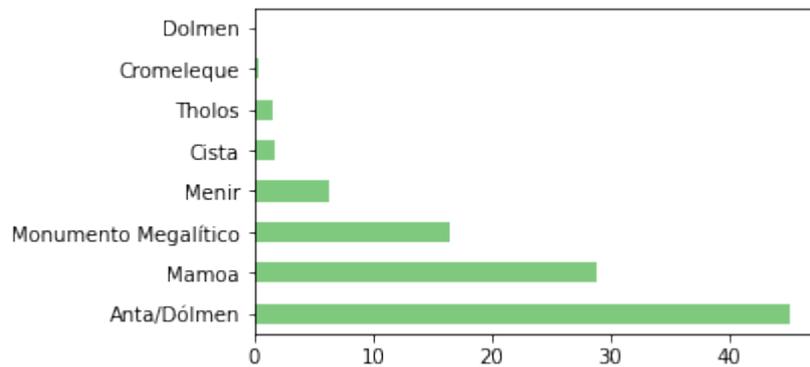


Fig. 1

Percentagem de ocorrências de cada tipologia correspondente ao Megalitismo

Podemos considerar que, em geral, a distribuição geográfica dos trabalhos presentes no Portal reflete a ocorrência dos trabalhos de salvaguarda a acompanhar as grandes obras públicas, com destaque para os distritos de Évora e Beja (Alqueva e Sistema Global de Rega associado).

Do mesmo modo, a distribuição geográfica dos registos de sítios atribuídos ao Megalitismo corrobora o defendido na bibliografia especializada, em geral. Como podemos verificar na Fig.2, a distribuição demonstra alguma predominância em determinados distritos, principalmente em Évora, Portalegre e Viseu.

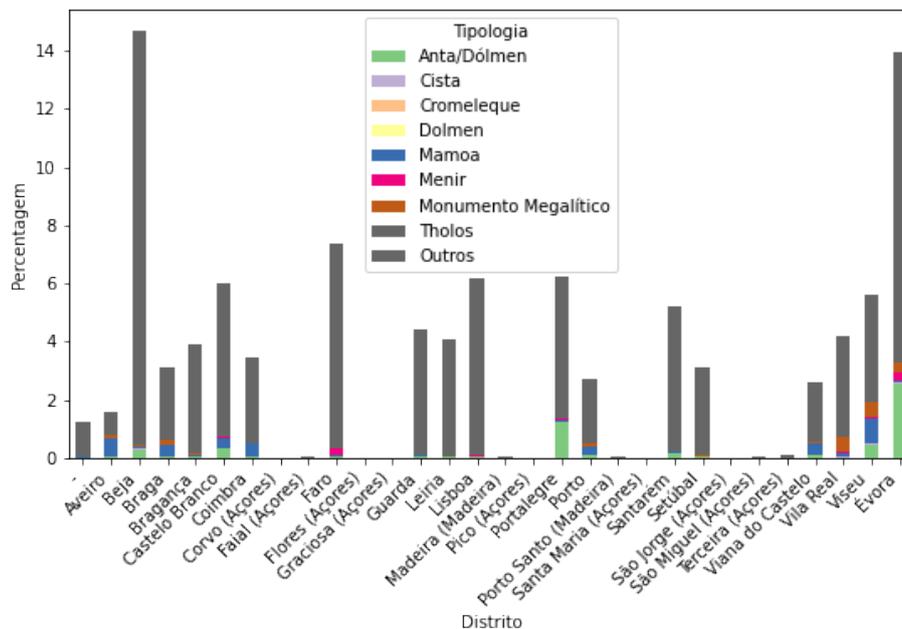


Fig. 2

Distribuição de registos por Distrito (ênfase para as tipologias do Megalitismo)

Quanto aos períodos cronológicos atribuídos aos sítios megalíticos, existe, tal como era

expectável, uma predominância do Neo-Calcolítico, Neolítico e Calcolítico, cronologias às quais se atribui normalmente a construção destes monumentos no território atualmente português. De salientar a representatividade do Indeterminado na atribuição da classificação de Mamoa e Monumento Megalítico.

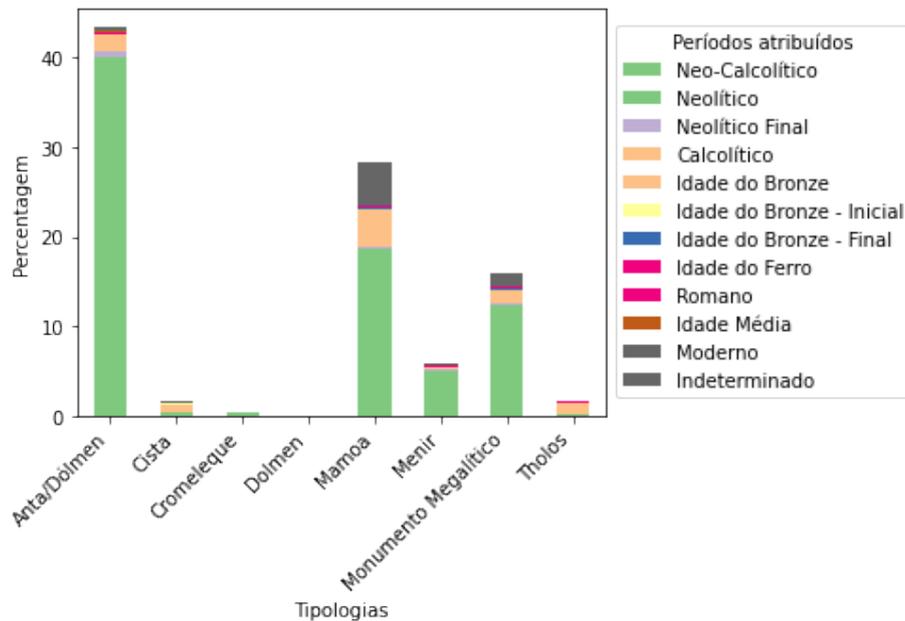


Fig. 3
Relação entre períodos e tipologia

Da análise cronológica dos metadados resultou também a confirmação de que existe uma tendência geral de acompanhamento da economia nacional. Em geral, observa-se um aumento do número de trabalhos arqueológicos, com destaque para a Prospeção e um declínio por volta de 2010, reflexo provável dos anos de crise económica.

No que concerne aos primeiros resultados do pré-processamento para aplicar PLN aos relatórios *online*, optou-se por apresentar aqui a forma não lematizada das palavras, por se ter verificado que estavam a ser atribuídas formas canónicas incorretas a diversos conceitos arqueológicos, assunto que pretendemos desenvolver mais detalhadamente noutra trabalho.

Da análise inicial, representada pelas nuvens de palavras das figuras 4 e 5, destaca-se a presença de unigramas como sítio, projeto e “m”, seguidos de conservação, trabalho, impacte, obra e localização, entre outras.

No caso dos bigramas, confirma-se a tendência para um vocabulário associado à chamada arqueologia de salvamento, pois, após o maior destaque de acompanhamento arqueológico, surgem referências como vistas na figura 5: *medidas minimização, impacte ambiental, parque eólico, estudo impacte ou mesmo alta velocidade e ligação ferroviária.*



Fig. 4
Nuvem de palavras - Unigrama



Fig. 5
Nuvem de palavras - Bigrama

5. DESAFIOS

Várias características dos próprios dados se perfilam como desafiadoras:

- Algumas categorias de trabalho arqueológico, pela sua natureza, possibilitam que os relatórios incluam descrições de diversas tipologias de sítio arqueológico. Por exemplo, um relatório de um Estudo de Impacto Ambiental poderá incluir referências a sítios romanos, como também a sítios pré-históricos.
- Sítios arqueológicos construídos em determinado período poderão incluir reutilização noutros períodos.
- Existe uma linguagem própria com conceitos e taxonomias específicas a de determinados períodos que se refletem naturalmente na redação dos relatórios de trabalhos arqueológicos.
- O grau de detalhe na descrição dos trabalhos arqueológicos irá limitar a possibilidade de recolher informação sobre a arquitetura do monumento (número de elementos, a sua dimensão, preservação, etc.), assim como o grau de preservação e registos gráficos.
- Maioritariamente os relatórios foram redigidos em português, mas também se identificaram documentos em inglês (o seu número total ainda nos é desconhecido);
- Resultou das nossas primeiras análises de preparação para o PLN a verificação de um grande número de termos relacionados apenas com trabalhos arqueológicos em contexto de obra. Este resultado, apesar de expectável, levanta-nos dúvidas sobre as reais proporções de dados científicos presentes nos relatórios arqueológicos.

6. PRÓXIMOS PASSOS

Para atingir os nossos objetivos, iremos continuar com o tratamento da informação e testar várias abordagens na Extração de Informação. Pretendemos utilizar os campos descritivos da base de dados do Portal do Arqueólogo para iniciar um processo iterativo de consolidação de um primeiro conjunto de dados de treino, a partir do qual pretendemos testar a Extração de Informação na generalidade dos relatórios e, em particular, nos dedicados ao Megalitismo.

Pretendemos também tentar identificar nos relatórios, em geral, os textos que incluem referências ao Megalitismo, mas não estão identificados como tal. Para esse efeito, por um lado, poderá bastar a identificação de palavras-chave, por outro, poderá ser útil a utilização de abordagens de sumarização dos dados nos relatórios em geral.

REFERÊNCIAS

1. Vlachidis, A., Tudhope, D.: A knowledge-based approach to Information Extraction for semantic interoperability in the archaeology domain. *Journal of the association for information science and technology*. 67, 1138 1152 (2016).
2. Migliorini, S., Grossi, P., Belussi, A.: An Interoperable spatio-temporal model for archaeological data based on ISO standard 19100. *Journal on Computing and Cultural Heritage (JOCCH)*. 11, 1 28 (2017).
3. Byrne, K., Klein, E.: Automatic extraction of archaeological events from text. In: *Proceedings of Computer Applications and Quantitative Methods in Archaeology*. Oxford: Archaeopress (2010).
4. CIDOC-CRM, <http://www.cidoc-crm.org>.
5. DGPC - Página introdutória ao Portal do Arqueólogo.
<http://www.patrimoniocultural.gov.pt/pt/patrimonio/patrimonio-imovel/patrimonio-arqueologico/gestao-da-atividade-arqueologica/portal-do-arqueologo/>.
6. Di Buono, M.P.: Information extraction for ontology population tasks. An application to the Italian archaeological domain. *International Journal of Computer Science: Theories and Applications*. 3, 40 50 (2015).
7. Noor, S., Jamil, S., Gohar, N., Shah, L.: Knowledge retrieval of historic concepts using semantic web. *Cluster Computing*. 22, 7321 7332 (2019).
8. *Listas de termos para preenchimento dos campos da base de dados Endovélico*.
http://www.patrimoniocultural.gov.pt/static/data/patrimonio_arqueologico/listadecampos.pdf.
9. Fiedukowicz, A., Glazewski, A., Kolodziej, A., Koszewski, K., Kowalski, P., Olszewski, R., Wlochynski, L.: Map portal as a tool to share information on cultural heritage illustrated by the national heritage board geoportal. In: *Advances in Digital Cultural Heritage*. pp. 48 64. Springer (2018).
10. Cripps, P.J.: Places, people, events and stuff; building blocks for archaeological information systems. In *Computer Applications and Quantitative Methods in Archaeology CAA 2012* (2012).
11. Portal do Arqueólogo, <https://arqueologia.patrimoniocultural.pt>.
12. Portal do Arqueólogo - Ficha Técnica.
<https://arqueologia.patrimoniocultural.pt/index.php?sid=about>.
13. Portal do Arqueólogo - Termos e Condições de Uso.
http://www.patrimoniocultural.gov.pt/static/data/patrimonio_arqueologico/portaldoarqueologo/termos_condicoes_relatorios_pa_versao_def.pdf.
14. Paijmans, J., Brandsen, A.: Searching in archaeological texts: Problems and solutions using an artificial intelligence approach. *PalArch's Journal of Archaeology of Egypt/Egyptology*. 7, 1 6 (2010).
15. Richards, J., Jeffrey, S., Waller, S., Ciravegna, F., Chapman, S., Zhang, Z.: The Archaeology Data Service and the Archaeotools project: faceted classification and natural language processing. *Archaeology*. 2, 31 56 (2011).
16. Jeffrey, S., Richards, J., Ciravegna, F., Waller, S., Chapman, S., Zhang, Z.: The Archaeotools project: faceted classification and natural language processing in an

- archaeological context. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 367, 2507-2519 (2009).
17. De Kleijn, M., van Manen, N., Kolen, J., Scholten, H.: Towards a user-centric SDI framework for historical and heritage European landscape research. *International Journal of Spatial Data Infrastructures Research*. 9, 1-35 (2014).
 18. Santos, I.: *Um atlas digital para o megalitismo: uma infraestrutura de dados espaciais (sudoeste da Península Ibérica)*, (2018).
 19. Paijmans, H., Brandsen, A.: What is in a name: Recognizing monument names from free-text monument descriptions. In: *Proceedings of the 18th Annual Belgian Dutch Conference on Machine Learning (Benelearn)*. pp. 2-6. (2009).
 20. Rocha, L.: *Espaços de necrópoles das primeiras sociedades camponesas no concelho de Arraiolos: um ponto da situação*. (2015).
 21. Sousa, A.C.: *O Penedo do Lexim e a sequência do Neolítico Final e Calcolítico da Península de Lisboa*. (2010).
 22. Tchienehom, P.: ModRef Project: from creation to exploitation of CIDOC-CRM triplestores. In *The Fifth International Conference on Building and Exploring Web Based Environments (WEB 2017)* (2017).

HISTÓRIA DIGITAL: DA TRANSCODIFICAÇÃO À RESSIGNIFICAÇÃO²⁹

Renata Vieira
CIDEHUS, Universidade de Évora
renatav@uevora.pt

Resumo. Da oralidade ao registo simbólico físico, da argila ao papel, do *volumen* ao códex, do manual à imprensa, do papel ao digital: em todas essas transformações são muitas as revoluções tecnológicas envolvidas. A revolução atual em tecnologias computacionais de tratamento de linguagem abre novas portas aos mais antigos registos de linguagem. Ao considerar o desenvolvimento de áreas como a inteligência artificial, o aprendizado de máquina, o processamento de linguagem natural, a *web* semântica, a realidade virtual e tantas outras, surge a pergunta: como serão os robôs historiadores do futuro? Este trabalho traz um olhar sobre o passado e o futuro dos registos e das tecnologias. Abordamos as questões envolvidas nesses processos de transformação, e as relacionamos com as pesquisas desenvolvidas no âmbito do CIDEHUS. Fazemos um convite de reflexão ampliada sobre o tema Humanidades Digitais e a evolução do pensamento humano.

1. INTRODUÇÃO

Vivemos a onda da transformação digital, ampliamos nossa capacidade de armazenar, processar e distribuir todo o tipo de informação. É um movimento que segue o curso da evolução humana e sua linguagem, vivenciado no caminho da oralidade, da escrita, da codificação, da criação das mais diversas tecnologias de suporte e transporte de informação. A evolução humana e sua inteligência pressupõe distintas estruturas cognitivas, relacionadas com as capacidades de memória e planeamento, e que estão ligadas a noções de tempo, passado e futuro. Assim, temos a capacidade de lembrar, observar e refletir sobre um possível passado e de prevermos e criarmos um provável futuro. A História então se estabelece como a área do conhecimento que reconstrói o passado, a partir de evidências. Muitas das evidências são documentais, registos codificados, impressos e mantidos em suportes que podem muitos anos depois serem observados, reinterpretados e transformados. Neste artigo, irei refletir sobre esses processos e discutir a transformação digital, as humanidades digitais e sua relação com a história. Nesse contexto, iremos ainda situar a investigação realizada no CIDEHUS em uma perspectiva inter e transdisciplinar que envolve, além da História, várias áreas, como

²⁹ Este trabalho é financiado por fundos nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto UIDB/00057/2020.

Informática, Processamento de Linguagem Natural, Linguística, Arqueologia e Turismo, entre outras.

2. A (R)EVOLUÇÃO DIGITAL COMO EVOLUÇÃO DO PENSAMENTO

No livro *How Homo became sapiens: On the evolution of thinking*³⁰, de Peter Gardenfors (2003), o autor discute amplamente como a inteligência humana difere da inteligência dos animais. A ideia central é a capacidade de simular a realidade, criar uma realidade imaginária, não necessariamente atrelada ao que se pode perceber no exato contexto e momento em que o observador se situa no presente, ou seja, os humanos, pela capacidade de simular e reter memória, trabalham muito bem com a noção de tempo, passado e futuro. O autor explica como essa capacidade é fortemente atrelada ao desenvolvimento da linguagem, de forma que se possa desenvolver uma memória compartilhada entre gerações. A oralidade e a transferência de cultura entre gerações deram origem a sociedades que se organizam com base em planejamento. Posteriormente, a escrita, o registo dessas memórias em um suporte perene, ampliou largamente o acesso ao que é distante no espaço e tempo, as culturas passam a ser construídas e transmitidas ao longo de várias gerações. A nossa geração e as que nos precederam mais recentemente tiveram o privilégio das bibliotecas, do conhecimento publicado em livros produzidos por estudiosos, chancelados por editores, organizados nas prateleiras pelos bibliotecários, mantidos pelos governos e organizações. As novas gerações vivem o privilégio do acesso digital. Dessa forma, podemos considerar transformação digital como um passo na externalização, evolução e transformação do pensamento.

3. HISTÓRIA E TRANSCODIFICAÇÃO

Pode-se dizer que a história se ocupa de recuperar um passado vivido, ao simular ou imaginar fatos que se sucederam, com base em evidências. Muitas das evidências são documentais, registos de linguagem realizados em alguma espécie de suporte. Geralmente é preciso decifrar um código antigo e, para interpretá-lo, se faz necessário transcodificá-lo, adaptá-lo a um novo código ou novo suporte. São processos de tradução, necessários para converter para um formato moderno ou com melhor suporte dados que estão registados ou armazenados de forma obsoleta. Nesse processo de evolução, são observados inúmeros suportes e meios de produção: podemos recordar os registos simbólicos, a escrita cuneiforme, o papiro, o pergaminho, o códex, o livro.

A forma de geração dos registos passa de manuscrita para impressa até chegar no digital. Cada uma dessas etapas tem um grande efeito no acesso, que se amplifica. O digital agiliza

³⁰ Traduzido em 2014 para *De Homo a Sapiens: sobre a evolução do pensamento*.

muito a distribuição, altera limitações espaciais e, além disso, flexibiliza as possibilidades de transformações no registo. Torna-se mais prática a complementação da informação, a adição de meta-dados ou de interpretações possíveis. Mas uma diferença fundamental é que passa a permitir o processamento do conteúdo de forma automática, computacional.

O movimento da digitalização ganhou grande força, permitindo gerar cópias de obras e documentos que pudessem ser mais facilmente acessados e distribuídos. No entanto, muitas dessas cópias foram digitalizadas como imagens, que apesar de atenderem às questões de acesso e leitura, ainda não permitem o processamento textual, importante para processos de busca, extração, ou seja, o tratamento computacional da informação. Para permitir o tratamento computacional do conteúdo com técnicas de recuperação e extração de informação e processamento de linguagem natural, a conversão da imagem para texto digitalizado é necessária. A conversão da imagem textual para o texto digital pode ser feita de maneira manual ou automática, através de tecnologia de reconhecimento de caracteres, o OCR. Porém, esse tipo de transcrição ainda gera muito ruído na tradução. Até mesmo registos em suportes muito antigos, como a escrita cuneiforme, podem ser, e são, transcodificados para o digital (Liu et al., 2016) e a partir daí podem passar por processos computacionais, como, por exemplo, o de tradução automática. O processamento de linguagem natural permite o tratamento (busca, seleção, comparação, relação) do conteúdo de diversas fontes. As áreas de processamento de linguagem natural e extração de informação podem auxiliar no processo de organização do conhecimento. Um exemplo de tarefa de extração de informação que tem sido aplicada a registos históricos é a anotação semântica ou o reconhecimento de entidades nomeadas (Álvarez-Mellado et al., 2021, Hubková et al., 2020; Allepuz et al., 2020). Este tipo de processamento está fortemente relacionado com áreas de estudo ligadas à História, como a Prosopografia.

4. HISTÓRIA DIGITAL

Romein (2020) apresenta um panorama geral do desenvolvimento da História Digital e aponta para o caminho dos dados abertos ligados, de forma que os esforços de diferentes grupos de pesquisa possam ser relacionados, permitindo que o conhecimento seja compartilhado, enriquecido e amplificado. Para isso, a comunidade apresenta e defende que se observem princípios para o compartilhamento de dados, como os princípios FAIR³¹, para que os dados sejam fáceis de serem encontrados, sejam acessíveis, interoperáveis e reusáveis.

No CIDEHUS, Centro Interdisciplinar de História, Culturas e Sociedades, da Universidade de Évora, diversos projetos têm se dedicado a adaptar suas fontes para permitir o tratamento computacional que possa auxiliar a organização do conhecimento ali representado. Um exemplo de grande sucesso é o caso do projeto das *Memórias Paroquiais*, que a partir do percurso inicial de digitalização já possibilitou a aplicação de técnicas de processamento de linguagem natural, baseadas em inteligência artificial, para a extração de informação

³¹ <https://www.go-fair.org/fair-principles/>.

(Olival et al., 2022). Nomeadamente, foram aplicadas técnicas de reconhecimento de entidades nomeadas que permitem identificar as menções feitas a entidades de determinadas categorias, como pessoas, locais e organizações. Uma base de dados foi criada e disponibilizada³². Esses elementos são relevantes para a análise da fonte histórica, uma vez que permitem gerar redes de relações pessoais e institucionais, associadas a uma distribuição geográfica.

Para além das *Memórias Paroquiais*, outras fontes textuais se preparam para o tratamento computacional, como é o caso da obra de Curvo Semedo (Finatto et al., 2018), do Padre Antônio Vieira, a *História do Futuro* (Banza et al., 2022); o *Livro das Monções* (Ribeiro, 2022); e muitos outros (Vieira et al., 2022).

O CIDEHUS dedica-se ao estudo do património e, como tal, engloba pesquisas interdisciplinares de várias naturezas. Ricas fontes de estudos históricos, abrigadas no entorno geográfico desse Centro de pesquisa, são os sítios arqueológicos. Irei destacar aqui o megalitismo. Nesse caso, temos as estruturas edificadas em rocha como fonte. Como tal, são fontes de acesso bastante limitado, por isso a importância dos relatórios descritivos e interpretativos em papel, produzidos pelos arqueólogos. No entanto, os relatórios em PDF, ainda assim, são fontes de acesso limitado. Como mencionado anteriormente, apresentam restrições para o tratamento computacional do conteúdo, e é preciso recorrer a tratamentos específicos para o formato PDF antes de poder trabalhar com o conteúdo propriamente dito (Santos & Vieira, 2022).

Outra maneira de trazer tais riquezas às pessoas, pelo digital, são as representações espaciais, ou 3D. Essas reproduções digitais permitem uma apresentação e descrição dessas fontes por imagens tridimensionais, que procuram reproduzir percepções de inserção e pertença ao ambiente.

5. TRANSCODIFICAR E RESSIGNIFICAR

Para além do textual, o património cultural ligado à preservação arqueológica promove uma ressignificação de espaços e ambientes. Esse tipo de reinterpretação histórica permite ligações com atividades de turismo, e não só. Muitas vezes está também ligado ao conceito de bem-estar e pode se configurar como ambientes, atividades físicas e de interiorização (Rodrigues e Vieira, 2020). Um exemplo bastante conhecido é o sítio arqueológico de Stonehenge, que abriga eventos e celebrações ligados à natureza e aos ciclos anuais e sua interrelação com os astros.

Considerando tais manifestações culturais tão longínquas, como os monumentos megalíticos, irei ainda mencionar as recentes descobertas de um novo local de pinturas rupestres em Castellote (Teruel).³³ Diz-se que a pintura apresenta uma cena de uma pessoa

³² <https://zenodo.org/record/4946479>.

³³ <https://archaeologynewsnetwork.blogspot.com/2021/07/well-preserved-cave-painting-of-honey.html> 14/07/2021.

subindo em uma árvore para pegar mel de uma colmeia e que foi produzida há cerca de 7.500 anos. Isso remete-nos à ideia de termos, ainda hoje, novos acessos a registros simbólicos que foram encravados nas pedras há 7500 anos. Essa notícia chega até nós pela tela do computador, através de algoritmos que avaliam nossos interesses e nos apresentam informações conforme esta avaliação. Podemos voltar a pensar na sequência de evolução da oralidade, escrita, imprensa, computação, internet e inteligência artificial. O que alcançamos atualmente nessa linha evolutiva vai além de uma nova modalidade de registro e sua distribuição, mas passa por uma manipulação do registro (através da IA), onde ferramentas ampliam nossas capacidades de coleta, síntese e interpretação. Somos uma nova espécie, híbrida, com ampliação da capacidade mental e da manipulação do natural, manipulação do pensamento externalizado.

Para além da história, outra maneira de ressignificar e refletir sobre o que significa ser humano se dá pela arte. Na mesma temática e na mesma época em que escrevia esse artigo, tive a oportunidade de visitar a Exposição Santuários, de Renée Gagnon, na Fundação Eugénio de Almeida, em Évora. No evento da sua abertura e com participação da artista e do arqueólogo Manuel Calado, eles explicaram o processo criativo das fotos dos monumentos megalíticos com manipulação digital e pintura e o significado desses monumentos do ponto de vista arqueológico. Através do processo de transcodificação e ressignificação artística dos registros do passado, estes deslocam-se das colinas do Alentejo para a sala do museu, por meio de representações digitais manipuladas.

A História vislumbra o passado e prepara o futuro, ao dar um sentido presente para um registro antigo. Em sua participação na transformação digital, a história depara-se com questões de suporte: armazenamento, distribuição e preservação; e questões de codificação: transcrição, processamento e extração. Podemos ainda apontar para um novo elemento: o raciocínio externalizado. A leitura atual é computacional, é interativa, o novo texto (Portela, 2022). E o novo livro requer novos leitores e escritores, acompanhar e compreender o que os algoritmos estão a realizar e aquilo que podem ou o que não podem é fundamental, é a nova alfabetização. No processo de evolução da externalização do pensamento, o “raciocínio” é agora executado/simulado por algoritmos (códigos) em dados (que podem estar nas nuvens) e, para além disso, as novas ferramentas se auto codificam, podem, por exemplo, sugerir códigos em linguagem de programação a partir de uma fórmula matemática. Tanto as humanidades precisam se apropriar desse processo, como os desenvolvedores de algoritmos precisam conhecer bem as questões humanas envolvidas com aquilo que desenvolvem. Se tornam muito importantes as humanidades nesse processo, para entender a relação entre o humano e não humano e colocar em cena essas e outras discussões, como o papel central da ética nesse desenvolvimento. A história envolve-se com a produção de registros para a permanência da história e vive o impacto do digital, computacional nesse âmbito. Deixo, na sequência, a pergunta: como serão os robôs historiadores do futuro?

REFERÊNCIAS

1. Allepuz, E. T., del Fresno Bernal, P., Martí, A. M., & Gordo, S. M.: The Semantics of Historical Knowledge. Labelling Strategies for Interdisciplinary and Digital Research in History. In *Hybrid Intelligence for Natural Language Processing HI4NLP@ ECAI* (2020).
2. Álvarez-Mellado, E., Díez-Platas, M. L., Ruiz-Fabo, P., Bermúdez, H., Ros, S., González-Blanco, E.: TEI-friendly annotation scheme for medieval named entities: a case on a Spanish medieval corpus, *Language Resources and Evaluation*, vol.55, no 2, 525–549. <https://doi.org/10.1007/s10579-020-09516-2> (2021).
3. Banza, A.P., Silva, A. R. Rodrigues, I.: A edição digital da História do Futuro, de António Vieira: arquivo e ferramentas. *Atas da Jornada de Humanidades Digitais do CIDEHUS* (2022).
4. Finatto, M. J., Quaresma, P., & Gonçalves, M. F. et al.: Portuguese Corpora of the 18th century: old medicine texts for teaching and research activities. *Proceedings of the Conference on Language Technologies & Digital Humanities* (2018).
5. Gardenfors, Peter: *How homo became sapiens: On the evolution of thinking*. Oxford University Press (2003).
6. Hubková, H., Kral, P., & Pettersson, E.: Czech historical named entity corpus v 1.0. *Proceedings of the 12th Language Resources and Evaluation Conference*, Marseille: France (2020).
7. Liu, Y., Hearne, J., & Conrad, B.: Recognizing proper names in ur iii texts through supervised learning. In *The Twenty-Ninth International Flairs Conference* (2016).
8. Olival, F., Cameron, H., Vieira, R.: *As Memórias Paroquiais: do manuscrito ao digital*. *Atas da Jornada de Humanidades Digitais do CIDEHUS* (2022).
9. Portela, M.: Simulação e Performatividade: O Arquivo LdoD em 8 Diagramas. *Atas da Jornada de Humanidades Digitais do CIDEHUS* (2022).
10. Ribeiro, A. S., Sacramento, A., Souza, M. & Vieira, R.: Event identification in the Monsoon Books (1616-1618). *Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing*, co-located with the International Conference on the Computational Processing of Portuguese (2022).
11. Rodrigues, A. & Vieira, R.: New Trends in the Sport Tourism industry, a case study of outdoor yoga at the lake Alqueva, Portugal. *Bet on Health, International Scientific Conference* (2020).
12. Romein, C. A., Kemman, M., Birkholz, J. M., Baker, J., DeGrujter, M., Meroño-Peñuela, A., Ries, T., Ros, R., & Scagliola, S.: State of the field: digital history. *History*, 105(365), 291–312. DOI: <https://doi.org/10.1111/1468-229X.12969> (2020).
13. Santos, I. & Vieira, R.: Semantic Information Extraction in Archaeology: Challenges in the Construction of a Portuguese Corpus of Megalithism. In: Garoufallou, E., Ovalle-Perandones, MA., Vlachidis, A. (eds) *Metadata and Semantic Research*. MTSR 2021. Communications in Computer and Information Science, vol 1537. Springer, Cham. https://doi.org/10.1007/978-3-030-98876-0_21 (2022).
14. Vieira, R., Olival, F., Cameron, H. F., Santos, J., Sequeira, O., & Santos, I.: Enriching the 1758 Portuguese Parish Memories (Alentejo) with Named Entities. *Journal of Open Humanities Data*, 7, 20. DOI: <http://doi.org/10.5334/johd.43> (2021).

15. Vieira R. et al.: Digital Humanities and Portuguese Processing: a research pathway. *Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing*, co-located with the International Conference on the Computational Processing of Portuguese (2022).

AS MEMÓRIAS PAROQUIAIS: DO MANUSCRITO AO DIGITAL³⁴

Fernanda Olival¹, Helena Freire Cameron², Renata Vieira¹

¹ CIDEHUS, Universidade de Évora

² CIDEHUS, Instituto Politécnico de Portalegre

mfo@uevora.pt, helenac@ipportalegre.pt, renatav@uevora.pt

Resumo. Este texto tem o intuito de traçar a história custodial das *Memórias Paroquiais*, desde a difusão dos inquéritos em 1758, até aos projetos atuais que visam convertê-las em objetos e dados digitais. Note-se que se trata de um recurso muito relevante para conhecer o Portugal de meados do século XVIII e que interessa não só ao historiador, como a muitos outros estudiosos e intervenientes no campo local, regional e do país. No interior do CIDEHUS, até pela dimensão, as *Memórias Paroquiais* têm suscitado, desde há mais de um decénio, trabalho de equipa interdisciplinar. Refletir sobre este e outros esforços em Portugal é também uma forma de avaliar e repensar estratégias de trabalho sobre esta coleção.

1. INTRODUÇÃO

“Memórias Paroquiais” é a designação pela qual hoje é conhecida a coleção das respostas a um inquérito mandado efetuar aos párcos de todas as freguesias de Portugal Continental, em 1758.

Com este texto, pretende-se dar a conhecer a trajetória das *Memórias Paroquiais*, desde a sua produção aos atuais projetos para a sua conversão em dados digitais. Conhecer este itinerário é importante para explicar como foi constituída esta coleção, as lacunas que apresenta, as instituições a que esteve associada e as apropriações e transformações que sofreu ao longo do tempo, também elas fruto dos diferentes contextos político-culturais e tecnológico-científicos. Do ponto de vista metodológico, usaremos enfoques diversificados, sem escamotear os de escala micro, sempre que qualquer indício o justificar, por mais ténue que seja. Nas histórias custodiais, muitas destas pequenas pistas podem ser, às vezes, esclarecedoras.

³⁴ Trabalho desenvolvido no âmbito dos projetos UIDB/00057/2020 e PTDC/ART-HIS/32327/2017 - FCT – Portugal.

2. OS MANUSCRITOS: A PRODUÇÃO

O objetivo do inquérito de 1758 era conhecer melhor o país, avaliar o impacto do sismo de 1755 e contribuir para um dicionário histórico-geográfico de Portugal. A iniciativa partiu da Secretaria de Estado dos Negócios do Reino, mas teria em vista - de certa forma - também dar continuidade ao trabalho do Padre oratoriano Luís Cardoso (Pernes, 1697- Lisboa, 1769). Este último era académico do número da Academia Real da História, desde 1737.

Em 1747 e em 1751, o Padre Luís Cardoso publicou o primeiro e o segundo volumes de um dicionário geográfico (letras A-C), para o qual reunira dados através de um inquérito, com 65 perguntas, efetuado aos párocos em 1732-1733, a que juntara outras fontes. Com o sismo de 1755, ter-se-ia perdido o material para os restantes volumes. Por isso, em 1758, o projeto tinha condições para ser retomado. Reenviaram-se interrogatórios para todas as freguesias do país, mesmo para aquelas cujo nome se iniciava por A, B, C, com textos já publicados, pois havia outras questões a apurar, certamente do interesse da Coroa, como os correios e as reconstruções, na sequência do Terramoto. Desta vez, o interrogatório continha 60 perguntas, organizadas em três conjuntos: terra, serra e rio. Foi impresso na época, num formato de 19,5 x 28,5 cm. A folha solta, com registos nos dois lados, devia ser dobrada a meio, formando um pequeno caderno de quatro páginas, facilitando o seu manuseio. Em dois exemplares que se conservam na Biblioteca Nacional de Portugal (BNP, Ms. 5, nº 3 e 4) é fácil observar com clareza essa dobragem (ver um deles em fig.s 1 e 2, em anexo). Não é possível confirmar se algum destes foi o exemplar transcrito e publicado no final do século passado por Gabriel Pereira, sem indicar qualquer cota, apenas que se encontrava naquela biblioteca (Pereira, 1895). É de realçar que, em qualquer deles, na pergunta 3, sobre a terra, “Quanto vizinho tem?”, foi acrescentado à mão, com letra da época, “e o Número das pessoas”.

O impresso foi distribuído através dos bispos (ou congéneres) e foram eles, ou os cabidos ou outras autoridades eclesíásticas, a fazê-lo chegar às mãos dos destinatários. Junto com as respostas, é provável que alguns párocos tenham devolvido o questionário, como aconteceu com o de S. Romão do Sado, no termo de Alcácer, que assim o refere:

“Por ordem de Vossa Magestade expedida pella Sacrataria de Estado dos negocios do Reyno ao Reverendíssimo e Excelentíssimo Senhor Arcebispo de Evora que me manda com suspensão do meu officio Parochial o seu arbitrio diga o que souber a respeito de huns interrogatorios impresso[s], que me remeteo em dezasete de Março desta prezente era de mil e setecentos e sincoenta, e oitavo, cujo impresso torno a remeter com êsta informação en que digo o que me consta sobre cada húm dos interrogatorios nelle expressados” (ANTT, *Memórias Paroquiais*, vol. 33, nº 13, p. 111).

Também o prior da Matriz de Alcácer do Sal terá acostado o papel impresso às suas respostas, sem que este hoje faça parte do acervo da Torre do Tombo (ANTT, vol. I, nº 71, pp. 501-560). Vários outros terão feito o mesmo. Quem compilou as respostas terá

eliminado esse material, considerando-o supérfluo, pois apenas teria interesse no que escreviam os párocos; o compilador conheceria bem o teor da lista de perguntas.

Provavelmente, um exemplar da Biblioteca Nacional de Portugal conservou-se por estar junto de apontamentos para preparar as respostas sobre Caminha, no Norte de Portugal (BNP, Ms. 5, nº 3). O outro porque se encontrava entre as respostas originais dos párocos ao Provisor do Priorado do Crato, mais em concreto, no meio do texto sobre a Flor da Rosa que, por sua vez, estava cosido com várias freguesias daquele Priorado. Sendo este último um território *nullius diocesis*, foi o provisor quem mandou distribuir o inquérito pelos párocos da sua jurisdição. Tudo indica que os clérigos terão respondido ao seu superior hierárquico ainda em 1758, cabendo a este remetê-los para a Secretaria de Estado. Alguns dirigiram-se mesmo a “Sua Senhoria” e não a “Sua Majestade”, quando responderam':

“Reverendíssimo Senhor

Vossa Senhoria me detremina lhe responda aoz interrogatorioz, que impressoz me forão remetidoz por ordem de Vossa Senhoria; e satisfazendo na parte, que metera á sua continência, informo a Vossa Senhoria, que esta aldea do Val do Pezo he situada na Provincia de Alentejo (...)”

(ANTT, *Memórias Paroquiais*, vol. 38, nº 52, p. 275)

Após ter sido recepcionada uma primeira resposta de cada paróquia, em 1759, deve ter sido solicitada uma cópia autêntica, por qualquer razão: extravio, mudança de provisor ou outra. Esta segunda hipótese afigura-se mais credível, a julgar pela resposta da Amieira que, em 1758, na sua réplica, aludia a uns interrogatórios que recebera do provisor Filipe Abranches de Castelo Branco e no ano seguinte indicava ter recebido, nesse ano, do Doutor João Nunes da Silveira, idêntico pedido (cf: ANTT, *Memórias Paroquiais*, vol. 3, nº 71, p. 541 e BNP, Ms. 5, nº 5). Em 11 terras alentejanas do Priorado do Crato examinadas (Amieira, Nossa Senhora da Conceição do Crato, Vale do Peso, Flor da Rosa, Gáfete, Mártires, Monte da Pedra, Tolosa, Monte Chamiço, Aldeia da Mata e Comenda), quando há data (em 3 freguesias não há), a lição da Biblioteca Nacional é de Março ou Abril de 1758 e um pouco mais minuciosa. A outra, que acabou inserida na coleção do Padre Luís Cardoso e que hoje se encontra na Torre do Tombo, quase sempre foi subscrita em Setembro ou Outubro de 1759 (falta a datação em 3 freguesias). Nem sempre os dados são rigorosamente coincidentes numa versão e na outra, nomeadamente nas informações demográficas. É difícil explicar como chegou a lição mais antiga à BNP, porque até ao momento não se descobriu qualquer pista segura, mas é muito provável que tenha entrado com outros fundos eclesiástico-conventuais, perto de meados do séc. XIX.

3. OS MANUSCRITOS NOS CONVENTOS ORATORIANOS

Na Secretaria de Estado, desconhece-se o tratamento dado às respostas, se algum tiveram. Dali teriam sido enviadas para o Convento das Necessidades, em Lisboa, sem que se saiba a data precisa. Destinavam-se ao Padre Luís Cardoso, que faleceu vários anos depois sem ter voltado a publicar qualquer volume do seu projetado dicionário. No entanto, Maria José Mexia Bigotte Chorão definitivamente comprovou que ele editou, sob o pseudónimo de Paulo Dias de Niza, ainda no final da sua vida, uma obra em 3 tomos, com informações que certamente resultaram destes inquéritos de 1758. Trata-se do *Portugal-sacro profano*, cujo volume 3 contava com dados sobre os correios, organizados pelo seu irmão mais novo, Pedro Nolasco dos Reis Cardoso, que era naquele período cónego da Colegiada de Ourém. Aliás, o pseudónimo já aparecia descodificado no catálogo da Biblioteca das Necessidades, de cerca de 1780 (Biblioteca da Ajuda (BA), 51-XIII-16, f. 183). Provavelmente, a tudo isto não seria indiferente a situação dos oratorianos que, a partir de 1760 e até ao final do reinado de D. José, deixaram de poder admitir noviços e conheceram várias limitações à sua atuação (Santos, 2010).

Os padres daquela agremiação ocupavam o convento das Necessidades, em Alcântara, desde 1750. Ou seja, quando acabou de ser construído, a mando de D. João V, que ali também erguera um palácio real e transformara uma ermida em capela real, sagrada em 1751. Com o Terramoto de 1755, a igreja saiu ileso e o convento exigiu obras. Ali, os padres oratorianos estavam obrigados a manter atividades de Ensino e tinham dotação régia para a conservação e aumento da biblioteca, que cresceria muito depois de 1755. Antes, as suas livrarias estavam no convento-sede, o do Espírito Santo da Pedreira, na baixa de Lisboa, onde hoje se situa o Centro Comercial “Armazéns do Chiado”. Todo o edifício, que lhes fora concedido em 1671, e a zona foram fortemente devastados pelo sismo e incêndio de 1755. Por isso, as espécies bibliográficas que conseguiram de alguma forma recuperar foram levadas para as Necessidades.

Depois da morte do Padre Luís Cardoso, a 3 de Julho de 1769, estariam as respostas ao inquérito de 1758 nesta Casa das Necessidades, onde aquele religioso faleceu e foi enterrado, na sepultura 8 (ANTT, *Congregação do Oratório*, L^o 9, p. 8).

Talvez já no século XIX, teria sido outro clérigo da mesma filiação, cujo nome se desconhece, que organizou os textos em forma de dicionário, com paginação numérica grafada. As freguesias foram numeradas, uma a uma, e ao cimo foi registado o nome da comarca a que cada uma pertencia. Por fim, cada volume foi encadernado, com o rótulo de "Diccionario Geographico de Portugal", na lombada e outro do número do tomo. Receberam a etiqueta tipográfica daquela biblioteca conventual, no verso da capa, como seria usual na instituição, neste período (Campos, 2015, p. 117). Ainda hoje consta de 80,5% dos 41 volumes de respostas ditas originais daquele Dicionário. Falta em 8 (vol.s 4, 5, 9, 11, 12, 16, 18, 36) e nos suplementos, mas talvez pela degradação da encadernação, tendo algumas delas sofrido alterações (*verbi gratia*, vol 18). É de notar que estes *ex-libris* apresentavam espaços para inserir a cota da obra (com um separador de letra e número), embora estes dados nem sempre fossem preenchidos, como aconteceu com os volumes do *Dicionário Geográfico*.

No que respeita às respostas ao inquérito de 1758, por se terem extraviado as descrições de cerca de 500 freguesias, no convento oratoriano foram acrescentados dois volumes de suplemento (vol.s 42-43). A estes adicionou-se um de índices do conjunto (vol. 44), com um prólogo. De acordo com uma nota que aparece na folha de rosto deste volume, estas intervenções datariam de 1832 ou ficaram concluídas nesse ano (Ver fig. 3).

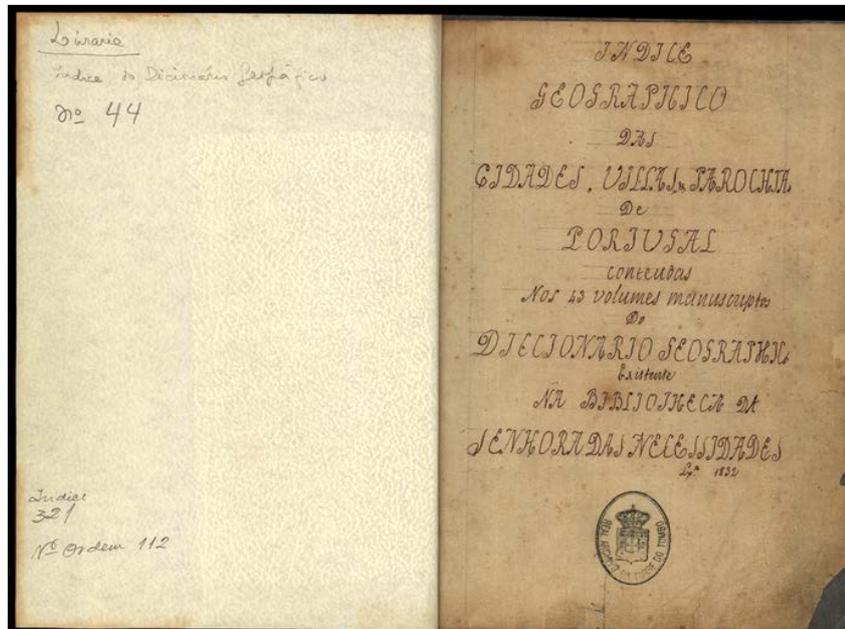


Fig. 3
Verso da capa e folha de rosto do vol. 44 das
Memórias Paroquiais (imagem cedida pelo ANTT)

Na altura da extinção das ordens religiosas masculinas, em 1834, a coleção encontrava-se de novo no Convento de Espírito Santo da Pedreira, cuja reconstrução ficara concluída em 1792 e para onde os oratorianos voltaram em Julho de 1833. Fizeram-no quando D. Pedro, como regente na menoridade da filha D. Maria II, desembarcou em Lisboa, exigindo que, em escassas horas, os oratorianos despejassem o Convento das Necessidades (Costa, 1959, p. 102). Foi uma mudança apressada, mas estes 44 volumes não ficaram para trás.

4. DO DEPÓSITO GERAL DOS LIVROS DOS CONVENTOS EXTINTOS À TORRE DO TOMBO

Em finais de Maio de 1834, com a causa liberal a vencer a guerra civil, foram decretados extintos todos os conventos, mosteiros e casas de religiosos. Os seus bens deviam ser incorporados “nos próprios da Fazenda Nacional”. Por isso, os recursos destas instituições, fossem de que natureza fossem, passaram a ser inventariados e apropriados pelo Estado.

Assim aconteceu com os do convento do Espírito Santo da Pedreira. Talvez nos finais do Verão de 1834, o escrivão que redigiu o levantamento dos bens móveis e imóveis, ao chegar à “livraria” e ao museu, limitou-se a dizer que todos estes pertencentes e os de igual perfil

ficavam entregues ao Doutor António Nunes de Carvalho (ANTT, *Ministério das Finanças, Casa do Espírito Santo de Lisboa*, Cx. 2233). Ou seja, seria responsabilidade dele arrecadá-los.

Quem era António Nunes de Carvalho (1786-1867)? Era um liberal, doutorado em Direito pela Universidade de Coimbra. A partir de meados de Outubro de 1834 e até 1836, ficaria oficialmente encarregado do *Depósito Geral das Livrarias dos Conventos Extintos* (DLEC), no edifício que pertencera ao convento de S. Francisco da cidade. Este organismo funcionou entre 1834 e 1841, quando se fundiu com a Biblioteca Nacional de Lisboa.

Se o escrivão não inventariou os livros, entre os papéis de Nunes de Carvalho foi possível localizar uma “Relação dos livros existentes no cubículo nº 4 da extinta Casa do Espírito Santo pertencentes a extinta Casa de Nossa Senhora das Necessidades recebidos no Hospício extinto da Terra Santa em 30 de Dezembro do anno de 1834”. Tinha letra do punho de José Mendes Dinis, que efetuaria outros inventários de conventos, neste âmbito. Ali, na relação invocada, a primeira obra descrita era precisamente o “Diccionario Geographico do Reino de Portugal”, em 43 volumes, em formato “fólio”, ao qual não atribuía autor, local de edição, nem ano. Em observações, esclarecia: “Em manuscripto com hum indice em folio” (BNP, Cód. 6963, f. 195). Nas quase 7 páginas do rol efetuado, entre 10 livros manuscritos e 102 títulos impressos, esta obra destacava-se pelo quantitativo de volumes (ver fig. 4, em anexo).

O Hospício da Terra Santa, para onde foi levado o Dicionário Geográfico, era igualmente dos franciscanos até à extinção das ordens religiosas. Situava-se no quarteirão do Convento de S. Francisco da Cidade, onde se viria a instalar o governo civil de Lisboa, nos anos de 1840. Terá servido de complemento ao DLEC, aparentemente sobretudo para armazenar pintura. Era tudo muito perto do convento oratoriano. Por uma relação de letra de José Mendes Dinis respeitante à despesa feita na condução de livros, quadros e outros objetos sob a alçada do Dr. Nunes de Carvalho é possível saber que o transporte dos materiais do Convento do Espírito Santo começou a 25 de Outubro e talvez tenha terminado a 31 de Dezembro de 1834, com os livros (BNP, Cód. 6963, f. 264-264v). Se estes ficaram para os 2 últimos dias, logo no início da operação foi adquirida uma escova, água e duas varas de pano, tudo para servir “na livraria” do convento do Espírito Santo. Provavelmente destinavam-se a limpezas, embora o pano pudesse ter outros usos. Logo a 29 de Outubro, foram transportadas as estantes, bancas e cadeiras.

Como na sua vigência o DLEC era uma entidade receptora e redistribuidora de acervos (Barata, 2002, p. 76), o *Dicionário* acabou sendo enviado para a Torre do Tombo, ainda na década de 1830. A partir de 1837, desenvolve-se todo um debate sobre a legitimidade de confiscar as bibliotecas contíguas a outras instituições, como paços régios, o que poderia implicar devoluções por apropriação indevida. A das Necessidades foi uma das abrangidas pela polémica, pois tinha ao lado o palácio real (Barata, 2003, pp. 127-132). Assim, em 4 de Março de 1843, o então bibliotecário da Ajuda, Alexandre Herculano (1810-1877), mais tarde conhecido como historiador e homem de letras, pretendia que fosse restituída à Biblioteca da Casa Real “a grande collecção d’apontamentos e memorias para o Diccionario Geographico de Portugal pelo Padre Luiz Cardoso”. Alegava que, por morte do padre oratoriano, os volumes tinham ficado a pertencer à “Bibliotheca do Real Paço das Necessidades (...) sendo tambem certo que a referida Bibliotheca pertence hoje ao apanagio da Corôa”. Como prova, citava o catálogo da Real Biblioteca onde figurariam os livros e rematava dizendo que lhe constara que a obra no seu conjunto fora levada pelo Doutor

Nunes de Carvalho para o DLEC e dali para a Torre do Tombo (Freitas, 1910, p. 11). É de realçar que Herculano, mais do que bibliotecário, era “uma espécie de superintendente, como que inspector, das bibliotecas dos Paços Reais da Ajuda e Necessidades e do Real Gabinete de Física, anexo à Biblioteca da Ajuda” (Santos, 1965, p. 103). Na sequência do seu pedido, a 21 de Março, uma portaria assinada por Costa Cabral e dirigida ao guarda-mor do Arquivo Nacional pretendia saber se aqueles volumes da coleção existiam na Torre do Tombo e em que estado se encontravam, depois de terem sido remetidos do Depósito acima apontado (ANTT, *Arquivo do Arquivo, Avisos e Ordens*, mç. 21, doc. 111). A Torre do Tombo confirmou a incorporação dos 44 volumes, mas não os deixou sair. Respondeu José Manuel Severo Aureliano Basto a 27 do mesmo mês, nos seguintes termos:

“não deve semelhante Collecção sahir deste archivo, por quanto não he propriedade particular, he uma diligencia, que o governo mandou fazer, incumbindo o referido Padre de coordenar as ditas Discripções Geograficas, para depois se poder fazer, e publicar o Diccionario em forma, o que não pôde ser feito pelo mesmo Padre talvez por falecer poucos annos dêspois; e se não fosse a curiosidade de hum outro Padre da mesma Congregação do Oratorio ficariã todas em montão confuso, e sem serventia alguma” (ANTT, *Ministério do Reino*, mç. 3723).

Note-se que este oficial-maior da Torre do Tombo conhecia bem o prólogo do volume 44 das *Memórias*, onde parte desta história ficou registada. Realce-se o papel do anónimo compilador oratoriano invocado. Na realidade, foi ele quem criou a coleção, ao organizá-la e encaderná-la. Porque a coleção ainda não existia, ela não aparece sequer entre as espécies inventariadas nos catálogos manuscritos das duas bibliotecas oratorianas das Necessidades, datados do séc. XVIII (ANTT, *Manuscritos da Livraria*, nº 627A, de cerca de 1758; BA, 51-XIII-16 a 19; 51-XIII-9; 51-XIII-11 e 12). O mesmo acontece nos catálogos da Real Biblioteca, que se encontram hoje identificados na Biblioteca da Ajuda (BA, 51-XI-54-58; 51-XIII-8; 51-XIII-14-15). As duas primeiras cotas do catálogo régio, que se acaba de mencionar, nem reportam espécies posteriores a 1831. É altamente provável que houvesse outros catálogos produzidos no séc. XIX. No entanto, não foram por ora localizados, até porque se passou a usar o sistema de “bilhetes” (entenda-se fichas) para a catalogação dos livros.

Alexandre Herculano, em 5 de Julho de 1843, voltou a insistir no direito de reaver “a colleção mss. dos apontamentos e noticias colligidas pelo Padre Cardoso para a feitura de um Diccionario Geographico de Portugal, collecção que por um motu proprio do Dr. Antonio Nunes de Carvalho foi retirada d’entre outros livros pertencentes á Biblioteca Real” (Freitas, 1910, pp. 11-12), mas não teve sucesso. A ser verdade o protagonismo de Nunes de Carvalho no envio da coleção para o Arquivo Nacional, então os 44 volumes terão entrado na Torre do Tombo entre 1835 e 1836, antes da saída dele do DLEC. A expressão *motu proprio* sugere que a iniciativa foi dele e sem o documento do governo a autorizar.

Nunes de Carvalho, depois da sua demissão do cargo de gestor do Depósito, foi nomeado Bibliotecário da Casa Real, em Junho de 1836 (ANTT, *Registo Geral de Mercês de D. Maria II*, L^o 5, fl.212v-213). Todavia, pouco tempo esteve na Biblioteca da Ajuda e nunca ali se albergou o *Dicionário Geográfico*, ao contrário do que já se escreveu (Portugal; Matos, 1974, p. 13). Em Setembro desse ano, transitou para Guarda-mor interino da Torre do Tombo (ANTT, *Registo*

Geral de Mercês de D. Maria II, L^o 8, f.13), onde tratou logo de constituir a Biblioteca do Arquivo, como se comprova por uma portaria de 5 de Outubro de 1836 a dar corpo à ideia (ANTT, *Ministério do Reino*, L^o 1716, f. 64v-65). Foi perto desta que, possivelmente, ficaram integrados os 44 volumes do *Dicionário Geográfico*, quando chegaram à Torre do Tombo. Outra possibilidade é ter ele solicitado a obra ao DLEC, quando exerceu o cargo de guardador (1836-1838). Até agora, não foi localizado qualquer documento comprovativo da data de entrada da coleção naquele arquivo. Talvez tenha sido na Torre do Tombo que cada volume recebeu uma cota alfanumérica e que ainda hoje é possível observar na maioria dos tomos (ver fig. 5). Está grafada a lápis, nas primeiras páginas em branco que sucedem à capa.

Vol.	Letra	Cota	Vol.	Letra	Cota
1	A1	21/C/15	23	M2	21/D/7
2	A2	21/C/16	24	M3	Sem cota
3	A3	21/C/17	25	M.N.	21/D/9
4	A4	21/C/18	26	O1	21/D/10
5	A5	21/C/19	27	P1	Sem cota
6	B1	21/-C/20	28	P2	21/D/12
7	B2	21/C/21	29	P3	21/D/13
8	C1	21/C/22	30	P9	21/D/14
9	C2	21/C/23	31	R1	21/D/15
10	C3	21/C/24	32	R2	21/D/16
11	C4	Sem cota	33	S1	21/D/17
12	C5	21/C/26	34	S 2	21/D/18
13	D-E1	21/C/27	35	S 3	21/D/19
14	E2	21/C/28	36	T 1	21/D/20
15	F1	21/C/29	37	T 2	21/D/21
16	G2 H J	21/D/2	38	V 1	21/D/22
17	G1	21/D/1	39	V 2	21/D/23
18	G2 H I	21/D/2	40	V 3	21/D/24
19	L1	21/D/3	41	V Z	21/D/25
20	L2	21/D/4	42	Suplemento	21/D/26
21	L3	21/D/5	43	Suplemento	21/D/27
22	M1	Sem cota	44		

Fig. 5

Primeiras cotas da coleção, na Torre do Tombo. Nos volumes onde se registou “sem cota” é provável que a folha se tenha soltado e perdido

Nesta altura, a cota muitas vezes coincidia com a localização no depósito. Hipoteticamente, estas cotas reportar-se-iam a uma estante 21, corpos C e D. Quando a Torre do Tombo, em 1862, mudou da ala da Calçada da Estrela para a ala da Rua de S. Bento, é possível que tenha havido ajustes no armazenamento. Em 1905, sabemos que os 44 volumes do *Dicionário Geográfico* estavam guardados na mesma sala da Biblioteca do Arquivo, sala essa designada “livraria” e ali também estava a coleção das *Genealogias Manuscritas* (Azevedo; Baião, 1905, p. 52). Em 1910, ocupavam “duas prateleiras da Estante 17, Segunda Divisão da Sala S

(Livraria)” (Freitas, 1910, p. 10). Os volumes do *Dicionário Geográfico* permaneceram na “Livraria”, enquanto a Torre do Tombo residiu em S. Bento.

5. DAS TRANSFERÊNCIAS DE SUPORTE AOS PROJETOS DE PUBLICAÇÃO DIGITAL

No século XIX, estando na Torre do Tombo, esta coleção continuou a despertar algumas atenções. Em Maio de 1868, o secretário geral da Academia das Ciências de Lisboa, Latino Coelho, endereçou um ofício ao Guarda-mor daquele arquivo, também ele sócio efetivo da Academia, para o empréstimo do “Diccionario Geographico de Cardoso (...) manuscripto que se torna necessario por algum tempo na bibliotheca d’esta Corporação scientifica para o desempenho d’uma commissão de serviço publico” (ANTT, *Arquivo do Arquivo, Avisos e Ordens*, mç. 30, nº 150).

Vários eruditos consultaram o Dicionário, como fonte, realçando a sua riqueza informativa (Azevedo, 1896, pp. 62-3). Uns publicaram pequenos trechos e outros não. Por exemplo, o advogado e jornalista Almeida Carvalho (1817-1897), que se interessava pelo passado de Setúbal, resumiu o prólogo do volume 44 e copiou pelo menos o inquérito. Hoje, este material faz parte do seu espólio (Arquivo Distrital de Setúbal, PT/ADSTB/PSS/APAC/C/0136). Muito contribuiu para divulgar a riqueza desta fonte Pedro A. de Azevedo. Publicou inúmeros extratos com interesse para a Arqueologia e para a História, entre 1896 e 1903, em números sucessivos d’ *O Archeologo Português*. É a partir desta iniciativa e para salientar o protagonismo dos párocos, que o rótulo de “Memórias Paroquiais” para designar a coleção se difunde, fazendo olvidar a de “dicionário geográfico”. Aliás, no volume de índices que ganhou em 1832, ainda era assim designada, tanto na folha de rosto com uma caligrafia cuidada, como numa nota a lápis, no verso da capa (ver fig. 3). O mesmo título de “Dicionário geográfico” aparece grafado a lápis na folha de rosto do vol. 3, com letra do mesmo punho do volume 44, e igualmente nas folhas de rosto dos volumes 16 e 18 ainda é visível: “Decionario Geographico de Portugal”.

A nova designação impôs-se com rapidez. Em 1905, num livro que Pedro de Azevedo publicou em co-autoria com António Baião sobre a Torre do Tombo, é como “Memórias Paroquiais” que se referem aos 44 volumes.

Também as cotas alfanuméricas acima citadas terão caído em desuso ou tornaram-se desajustadas, ainda no século XIX³⁵. A coleção passou a ser referenciada pelos tomos e fólhos (na realidade páginas), como fez Pedro A. Azevedo no artigo citado e nos que se seguiram. Outros autores nem mencionavam qualquer cota, confiando na simples identificação através do nome da freguesia. As *Memórias Paroquiais* ganhavam identidade por si mesmas. Com o passar do tempo, foram as indicações de volume e número dado à freguesia aquelas que prevaleceram, conforme o registo do volume 44, produzido ainda no Convento das

³⁵Hoje praticamente só as *Genealogias Manuscritas* continuam a manter cotas alfanuméricas, no ANTT.

Necessidades. Este instrumento de descrição, e depois uma fotocópia do mesmo, manteve-se longamente em uso, na sala de referência da Torre do Tombo: até 2014.

Não foi por erro que a coleção em apreço foi designada “Dicionário Geográfico”, nos séculos XIX e XX, como já se escreveu (Portugal; Matos, 1974, p. 13).

Em 1935-37, as *Memórias Paroquiais* já seriam, possivelmente, a coleção mais conhecida e consultada da Torre do Tombo, por historiadores e curiosos, no dizer de Rocha Madahil, que as considerava um tesouro merecedor de ser impresso e conhecido de todos (Madahil, 1937, p. 7, 49). Terá sido ele o primeiro a editar integralmente uma Memória (Ílhavo), neste caso uma freguesia inteira, a principal daquele município (perderam-se, aliás, as restantes).

No século XX, sobretudo a partir dos anos de 1980, com o poder democrático a dinamizar o país, fosse em formato de livro ou de artigo, houve um surto de publicações destes textos, quase sempre usando o recorte concelhio.

Cerca de 2002, por ser “leitura muito solicitada”, a coleção começou a ser microfilmada, na Torre do Tombo, para preservar os originais. Poucos anos depois, foi feita nova transferência de suporte. Ou seja, os microfilmes a preto e branco foram digitalizados e postos online, apesar da falta de qualidade de algumas imagens, o que por vezes dificulta as transcrições. O facto de ficarem consultáveis à distância, redobrou o interesse por esta fonte.

Assim, este importante conjunto documental chegou ao primeiro decénio do século XXI em diferentes suportes e com diferentes acessibilidades. Por esta altura, começaram também a surgir projetos de maior fôlego. São de referir cinco iniciativas de publicação, que arrancaram todas na década invocada.

A primeira foi desenvolvida no âmbito de uma colaboração entre a Torre do Tombo e uma equipa da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa, liderada por Luís Espinha da Silveira (1954-2021). Este projeto inseriu-se no âmbito do *Atlas histórico do SIGMA* - Sistema de Informação Geográfica e Modelação de Dados aplicado à História de Portugal. Num mapa online, as imagens das *Memórias Paroquiais* manuscritas do ANTT ficaram acessíveis a partir da carta de freguesias (Cf. <http://atlas.fcsh.unl.pt/cartoweb35/atlas.php?lang=pt>).

O segundo empreendimento foi gerado na Universidade do Minho. Este, nos últimos vinte anos, tem editado os textos em formato de dicionário, indexando-os por freguesias, dentro dos atuais concelhos e organizados por distritos. Entre 2003 e 2021, saíram dos prelos tipográficos onze grossos volumes, que correspondem a toda a zona a norte do Rio Tejo, ao distrito de Setúbal e Portalegre, bem como ao território de Olivença, que em 1758 fazia parte de Portugal (Capela, 2003,2005; Capela, et al., 2006-2021). José Viriato Capela tem sido o coordenador do projeto e, ao mesmo tempo, promoveu o estudo destes textos. A título individual ou coletivo, publicou importantes introduções sobre especificidades das zonas tratadas.

A terceira iniciativa de publicação foi promovida a partir de 2005 pelo Centro de História da Universidade de Lisboa. Editou topograficamente os volumes do ANTT, tendo publicado em papel, até ao momento, nove volumes, que correspondem às letras A-C na sequência alfabética (Cosme & Varandas (ed.), 2009 a 2016).

Quase sempre com supervisão de Saul Gomes, também de 2005 e até 2012, uma equipa da Universidade de Coimbra envolveu-se na edição de textos sobre a região centro no séc. XVIII, nos quais incluiu as *Memórias Paroquiais*. Organizados por concelhos, foram publicados 11 volumes: Viseu, Marinha Grande, Batalha, Almeida, Alvaiázere, Castelo Branco, Mangualde, Leiria, Lousã, Ansião e Pombal (*Notícias e memórias paroquiais setecentistas*, 2005-2012).

A quinta iniciativa de publicação foi levada a cabo pelo Centro Interdisciplinar de História, Culturas e Sociedades da Universidade de Évora (CIDEHUS). Este projeto desenvolveu-se em várias fases e visou, *ab initio*, uma disponibilização integralmente digital, em acesso aberto, num *site* intitulado Portugal 1758. Desde 2016, as Memórias de quase todas as paróquias do Alentejo foram reunidas num portal agregador de dados, o CIDEHUSDigital, onde também podem ser encontradas outras coleções (Cf. <http://www.cidehusdigital.uevora.pt/>). Dispõe de um motor de busca e de licenças *creative commons 4.0*.

Com efeito, em 2007, o CIDEHUS iniciou o processo de transcrição dos textos, quase sempre a partir das digitalizações da Torre do Tombo. Recorreu a um sistema colaborativo, quer para fazer as transcrições, quer para, e sempre que possível, fazer acordos com outros transcritores que já tivessem editado essas *Memórias*. Este processo, ainda que moroso, revelou-se eficaz, pelo que estão já disponíveis *online* 366 freguesias no CIDEHUSDigital. Ressalve-se, contudo, que dado que entrevistaram diversos paleógrafos e, inclusive, estudantes universitários, os textos agora disponíveis apresentam diferentes critérios de transcrição paleográfica.

A partir de 2020, o CIDEHUS-UÉvora alargou o âmbito do projeto em duas vertentes: transcrição e processamento. Relativamente à primeira, pretende-se transcrever as *Memórias* em falta, para que se possa disponibilizar na íntegra e em acesso aberto todas as freguesias do Sul de Portugal, respeitando os princípios FAIR³⁶. No que respeita ao processamento, neste momento, estão a ser desenvolvidos estudos em Humanidades Digitais, beneficiando da acessibilidade deste *corpus*. Os textos estão a ser enriquecidos através de um processo de anotação de Entidades Nomeadas. Nesta fase, estão a ser estudadas as categorias semânticas de Pessoas, Locais e Organizações (Santos et al., 2020; Vieira et al., 2021). A equipa já publicou um *dataset* no CLARIN - *Common Language Resources and Technology Infrastructure*, e no ZENODO - Repositório Científico de Dados no âmbito da estrutura de Ciência aberta *OpenAIRE* - EOSC (Vieira et al., 2021), além de vários artigos (Santos et al., 2020; Vieira et al., 2021).

Este importante conjunto textual propicia, também, estudos em linguística. Neste momento, estão a ser desenvolvidas abordagens no que se refere ao tratamento computacional da ortografia do século XVIII, cuja variação gráfica é bastante expressiva nas *Memórias* (Cameron et al., 2021).

O funcionamento da sociedade neste século XXI tem sido cada vez mais digital. Os serviços e estruturas acompanham esta crescente transição digital e este *modus operandi* reflete-se também num grande incentivo à digitalização do património (UE, 2021).

Fruto deste apelo para uma transição digital, atualmente já se encontram disponíveis muitos dados em suporte virtual, mas, frequentemente, estes não estão em acesso aberto ou não

³⁶ <https://www.go-fair.org/fair-principles/>

conseguem ser pesquisados. Assim, a mera transferência de suporte não representa um verdadeiro funcionamento digital.

Para estudos históricos e linguísticos sobre texto antigo, é essencial que seja feita uma disponibilização digital de dados textuais com qualidade, com rigor nos metadados, seguindo as diretrizes de Timothy Berners-Lee³⁷. A informação em ambiente digital deverá continuar a ser, quanto possível, fidedigna ao original, mas desenvolvida de modo a propiciar processamento computacional, trazendo a realidade passada para os adventos deste século da informação.

As Humanidades Digitais, enquanto abordagem de investigação, utilizam novos métodos e ferramentas para conseguir solucionar questões cuja resposta seria difícil ou mesmo impossível de conseguir sem ajuda de cada vez mais sofisticados instrumentos de processamento computacional (Edmond, 2021). Desenhados na sua génese a partir de uma perspetiva interdisciplinar, os estudos em Humanidades Digitais sobre texto antigo com valor patrimonial assumem, assim, novas oportunidades de pesquisa, nunca antes pensadas, e as *Memórias Paroquiais* enfrentam mais esse desafio.

6. EPÍLOGO

Como se demonstrou, o inquérito de 1758 teve um itinerário com muitas intervenções, mudanças e transformações. Deu origem a uma coleção, que quase seguramente só foi constituída por volta de 1832, ainda no Convento dos Oratorianos. Foi no final dessa centúria que ganhou uma nova identidade, ao começar a ser designada por “Memórias Paroquiais”, pelo oficial arquivista da Torre do Tombo, Pedro A. Azevedo (1869-1928).

Uma vez incorporada no principal arquivo do país, é de realçar a microfilmagem e a digitalização que sofreu, permitindo-lhe esta última maior difusão.

Apesar da sua relevância e das atenções que tem merecido, só no século XXI começaram a surgir projetos mais alargados, tendo em vista a publicação integral, até agora ainda não concretizada em versão transcrita. De notar que o século XX foi dominado pela publicação parcelar, sempre em suporte de papel. A transformação para o mundo digital é muito recente e tem sido marcada pelo digital passivo. Urge alterar a estratégia e transformar as *Memórias Paroquiais* em dados, aptos a serem ligados com outros, e assim contribuir para uma maior divulgação e fruição deste tão relevante conjunto documental, quer a nível nacional, quer internacional. O seu conteúdo pode ser importante para diversos investigadores e intervenientes, em várias áreas do conhecimento, muito para além das Humanidades.

³⁷ https://www.w3.org/2011/gld/wiki/5_Star_Linked_Data

REFERÊNCIAS

1. Azevedo, Pedro A. de. «Extractos archeologicos das “Memorias parochiaes de 1758”», *O Archeologo Português*, II -VIII (1896-1903).
2. Azevedo, Pedro de; Baião, António. *O Archivo da Torre do Tombo: sua historia, corpos que o compõem e organização*. Lisboa: s.n. (1905).
3. Barata, Paulo J. S. "O Depósito das Livrarias dos extintos conventos: uma instituição desconhecida do período oitocentista português". *Revista Portuguesa de História do Livro e da Edição*, n. 10, pp. 75–105 (2002).
4. Barata, Paulo J.S. *Os livros e o liberalismo: da livraria conventual à biblioteca pública uma alteração de paradigma*. Lisboa: Biblioteca Nacional (2003).
5. Cameron, H.F., Gonçalves, M.F., Olival, F.; Vieira, R. “Spelling variants in 18th-century Portuguese: constraints on natural language processing”, *Programming in Digital Humanities*, Springer; High Performance Computing Chair – Univ. Évora [approved for publ.] (2022).
6. Cameron, H. F.; Gonçalves, M. F.; Quaresma, P. “Linguistic and orthographical classic Portuguese variants. Challenges for NLP”. In: M.J.Finatto, R. Vieira, S.Pollak and S. Luz (ed.), *Proceedings of the Workshop on Digital Humanities and Natural Language Processing*, co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2020), vol. 2607. Évora: CEUR-WP org, 43-48. <http://ceur-ws.org/Vol-2607/short1.pdf> (2020).
7. Campos, Fernanda Maria Guedes de. *Para se achar facilmente o que se busca: bibliotecas, catálogos e leitores no ambiente religioso: séc. XVIII*. Casal de Cambra, Portugal: Caleidoscópio (2015).
8. Capela, José. *As freguesias do distrito de Braga nas «Memórias paroquiais» de 1758: a construção do imaginário minhoto setecentista*, Braga, Universidade do Minho (2003).
9. Capela, José V. *As freguesias do distrito de Viana do Castelo nas memórias paroquiais de 1758: Alto Minho memória, história e património*, Monção, Casa Museu Universidade do Minho (2005).
10. Capela, José Viriato; Matos, Henrique; Borralheiro, Rogério. *As freguesias do Distrito de Vila Real nas memórias paroquiais de 1758: memórias, história e património*. [Braga]: J.V.C. (2006).
11. Capela, José Viriato; Borralheiro, Rogério; Matos, Henrique; Oliveira, Carlos Prada de. *As freguesias do distrito de Bragança nas Memórias Paroquiais de 1758: memórias, história e património*, Braga, s.n. (2007).
12. Capela, José Viriato; Matos, Henrique; Borralheiro, Rogério. *As freguesias do distrito do Porto nas Memórias Paroquiais de 1758: memórias, história e património*, Braga, s.n. (2009).
13. Capela, José Viriato; Matos, Henrique. *As freguesias do distrito de Viseu nas memórias paroquiais de 1758: memórias, história e património*, Braga, s.n. (2010).
14. Capela, José Viriato; Matos, Henrique. *As freguesias dos distritos de Aveiro e Coimbra nas Memórias Paroquiais de 1758: memórias, história e património*, Braga, Ed. José Viriato Capela (2011).
15. Capela, José Viriato; Matos, Henrique. *As freguesias do distrito da Guarda. Memória, história e património*, Braga, Ed. José Viriato Capela (2013).

16. Capela, José Viriato, Matos, Henrique; Castro, Sandra. *As freguesias dos distritos de Lisboa e Setúbal nas «Memórias Paroquiais de 1758»: memórias, história e património*, Braga, Casa Museu de Monção/Universidade do Minho (2016).
17. Capela, José, Matos, Henrique; Castro, Sandra. *As freguesias dos distritos de Castelo Branco, Portalegre e Olivença nas «Memórias Paroquiais» de 1758: memórias, História e património*, Braga, Dos Autores (2019).
18. Capela, José Viriato; Matos, Henrique; Castro, Sandra. *As freguesias dos distritos de Leiria e Santarém nas Memórias Paroquiais de 1758: memórias, história e património*, Braga (2021).
19. Cardoso, Luís. *Diccionario geografico, ou noticia historica de todas as cidades, villas, lugares, e aldeas, rios, ribeiras, e serras dos Reynos de Portugal, e Algarve, com todas as cousas raras, que nelles se encontraõ, assim antigas, como modernas*. Vol. I. 2 vols. Lisboa: na Regia Officina Sylviana, e da Academia Real (1747).
20. Cardoso, Luís. *Portugal sacro-profano*. 3 Vols, Lisboa: na Officina de Miguel Manescal da Costa (1757-1768).
21. Cosme, J. & Varandas, J. *Memórias Paroquiais (1758)*, vol. 1 [Abação-Alcaria] - 9 [Canedo-Castedo], Lisboa; Casal de Cambra: Centro de História da Universidade de Lisboa; Caleidoscópio (2009-2020).
22. Costa, Mário. "O Palácio Barcelinhos e o seu antecessor o Convento do Espírito Santo da Pedreira". *Olisipo*, n. 86, pp. 91-103 (1959).
23. Chorão, Maria José Mexia Bigotte. "Inquéritos promovidos pela Coroa no século XVIII". *Revista de História Económica e Social*, 1a série, n. 21, pp. 93-130 (1987).
24. Edmond, Jennifer (ed). *Digital Technology and the Practices of Humanities Research*. UK: Open Book Publishers, <https://doi.org/10.11647/OBP.0192> (2020).
25. European Commission: *Recommendation of 10.11.2021 on a Common European data space for cultural heritage*, Brussels, 10.11.2021 - C(2021) 7953 final, <https://digital-strategy.ec.europa.eu/en/news/commission-proposes-common-european-data-space-cultural-heritage>.
26. Freitas, Jordão de. "Herculano, bibliothecario". *Boletim da Real Associação dos Archeologos Portuguezes*, Número commemorativo do Centenário de Herculano, pp.10-12 (1910).
27. Madaíl, António Gomes da Rocha. *Ílhavo no século XVIII: as informações paroquiais de 1721 e de 1758: integralmente publicadas pela 1a vez*. Figueira da Foz: Tipografia Popular (1937).
28. *Notícias e memórias paroquiais setecentistas*. 11 Vols,. Viseu: Palimage / Centro de História da Sociedade e da Cultura (2005-2012).
29. Pereira, Gabriel. «Interrogatorios para a organização do “Diccionario Geographico” do Pe. Luís Cardoso», *O Archeologo Português*, I, pp. 268-271 (1895).
30. Portugal, Fernando; Matos, Alfredo de. *Lisboa em 1758*. Lisboa: s.n. (1973).
31. Santos, Eugénio. “Oratorianos”, *Dicionário Histórico das Ordens e Instituições afins em Portugal*, Lisboa, Gradiva (2010).
32. Santos, I., Olival, F. & Sequeira, O. “Excavating the data pit: the Portuguese Parish Memories (1758) as a gold standard”. *Workshop on Digital Humanities and Natural Language Processing* (2020).
33. Santos, Mariana A. Machado. “Alexandre Herculano e a Biblioteca da Ajuda”, *O Instituto*, Vol. CXXVII, Coimbra, pp. 99-174 (1965).

34. Vieira, R., Olival, F., Cameron, H. F., Santos, J., Sequeira, O., & Santos, I. Enriching the 1758 Portuguese Parish Memories (Alentejo) with Named Entities. *Journal of Open Humanities Data*, 7, 20. DOI: <http://doi.org/10.5334/johd.43> (2021).
35. Vieira, R.; Olival, F.; Helena F. C.; Santos, J.; Santos, I.; Sequeira, O. "*Parish Memories with Named Entities*". Zenodo. <https://doi.org/10.5281/ZENODO.4946479> (2021).

ANEXO (AS MEMÓRIAS PAROQUIAIS: DO MANUSCRITO AO DIGITAL)

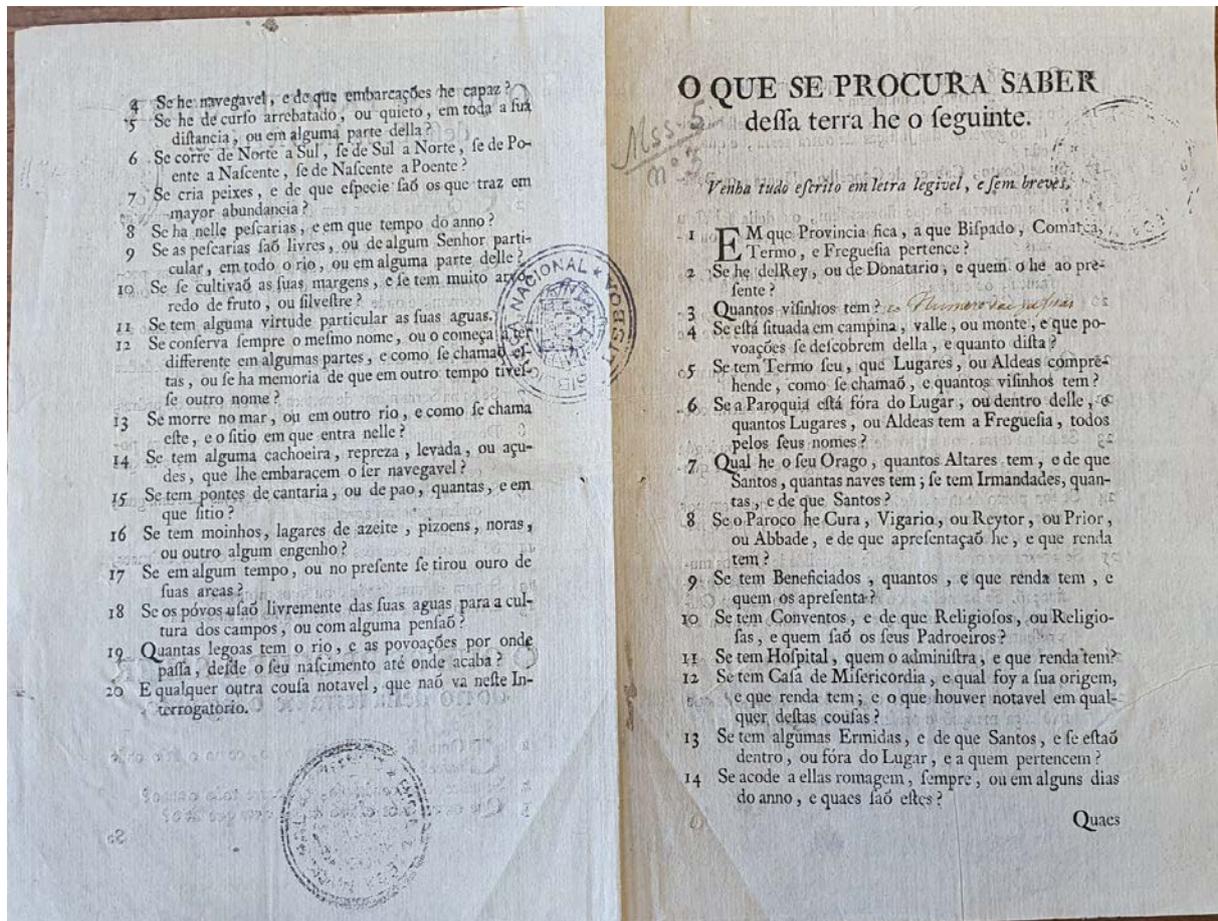
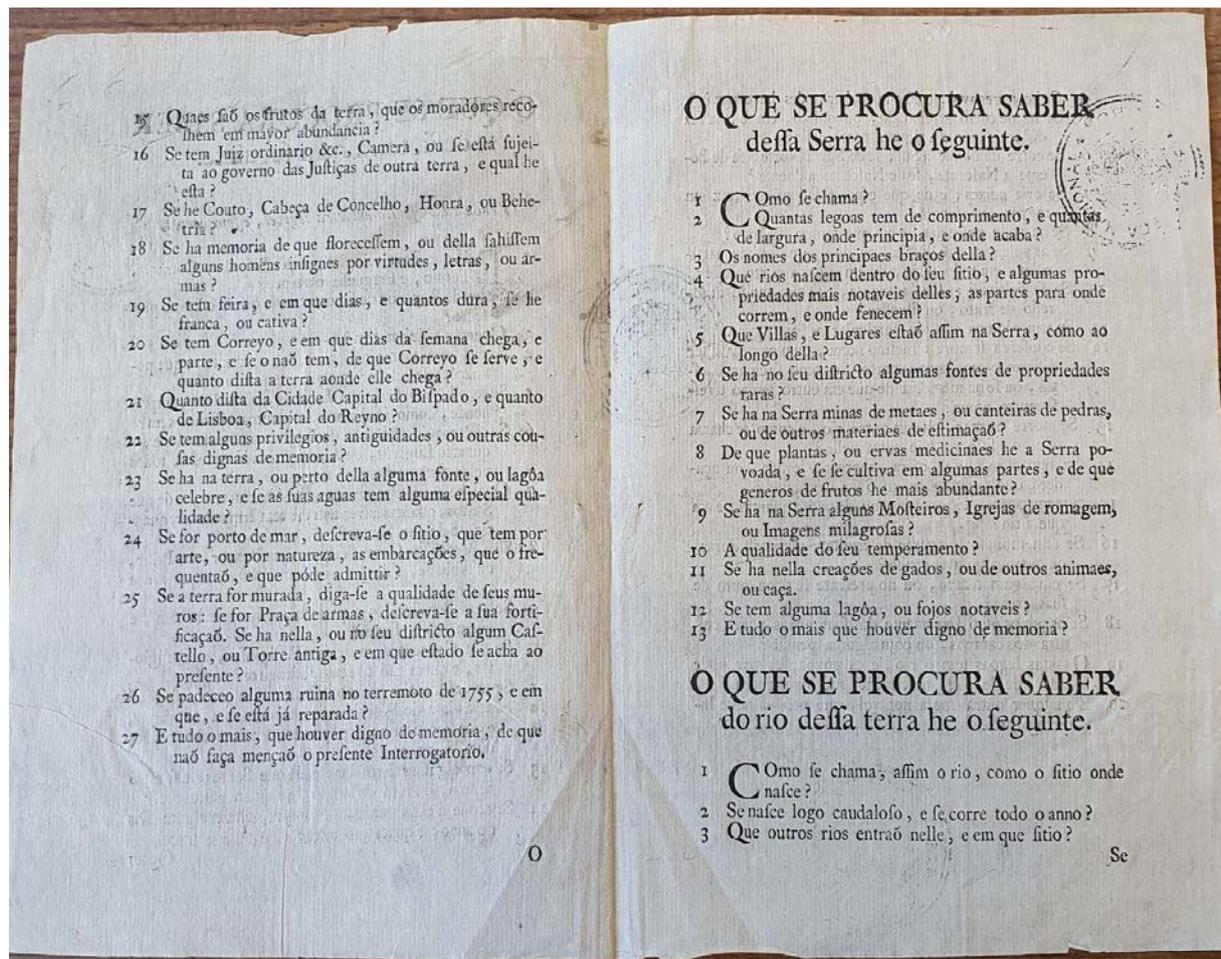


Fig. 1
Inquérito impresso distribuído em 1758 (BNP, Ms. 5, nº 3). Imagem cedida pela BNP



- 15 Quaes são os frutos da terra, que os moradores recohem em maior abundancia?
- 16 Se tem Juiz ordinario &c., Camera, ou se está sujeita ao governo das Justicas de outra terra, e qual he esta?
- 17 Se he Couto, Cabeça de Concelho, Honra, ou Behetría?
- 18 Se ha memoria de que florecessem, ou della sahísem alguns homêns insignes por virtudes, letras, ou armas?
- 19 Se tem feira, e em que dias, e quantos dura, se he franca, ou cativa?
- 20 Se tem Correyo, e em que dias da semana chega, e parte, e se o não tem, de que Correyo se serve, e quanto dista a terra aonde elle chega?
- 21 Quanto dista da Cidade Capital do Bispado, e quanto de Lisboa, Capital do Reyno?
- 22 Se tem alguns privilegios, antiguidades, ou outras cousas dignas de memoria?
- 23 Se ha na terra, ou perto della alguma fonte, ou lagôa celebre, e se as suas aguas tem alguma especial qualidade?
- 24 Se for porto de mar, descreva-se o sitio, que tem por arte, ou por natureza, as embarcações, que o frequentão, e que pôde admitir?
- 25 Se a terra for murada, diga-se a qualidade de seus muros: se for Praça de armas, descreva-se a sua fortificação. Se ha nella, ou no seu districto algum Castello, ou Torre antiga, e em que estado se achá ao presente?
- 26 Se padecco alguma ruina no terremoto de 1755, e em que, e se está já reparada?
- 27 E tudo o mais, que houver digno de memoria, de que não faça menção o presente Interrogatorio.

O QUE SE PROCURA SABER dessa Serra he o seguinte.

- 1 Como se chama?
- 2 Quantas legoas tem de comprimento, e quantas de largura, onde principia, e onde acaba?
- 3 Os nomes dos principaes braços della?
- 4 Que rios nascem dentro do seu sitio, e algumas propriedades mais notaveis delles; as partes para onde correm, e onde fenecem?
- 5 Que Villas, e Lugares estáo assim na Serra, como ao longo della?
- 6 Se ha no seu districto algumas fontes de propriedades raras?
- 7 Se ha na Serra minas de metaes, ou canteiras de pedras, ou de outros materiaes de estimaçã?
- 8 De que plantas, ou ervas medicinaes he a Serra povoada, e se se cultiva em algumas partes, e de que generos de frutos he mais abundante?
- 9 Se ha na Serra alguns Mosteiros, Igrejas de romagem, ou Imagens milagrosas?
- 10 A qualidade do seu temperamento?
- 11 Se ha nella creações de gados, ou de outros animaes, ou caça.
- 12 Se tem alguma lagôa, ou fojos notaveis?
- 13 E tudo o mais que houver digno de memoria?

O QUE SE PROCURA SABER do rio dessa terra he o seguinte.

- 1 Como se chama, assim o rio, como o sitio onde nasce?
- 2 Se nasce logo caudaloso, e se corre todo o anno?
- 3 Que outros rios entrao nelle, e em que sitio?

Se

Fig. 2 Verso da folha impressa do inquérito de 1758 (BNP, Ms. 5, nº 3). Imagem cedida pela BNP

195^{ma.}

Relação dos Livros existentes no Gabinete M.^o do extinto Convento do Espírito Santo fundidos a extinta Casa de S. Penha das Necessidades recolhidos no Hospício extinto da Terra Santa em 30 de Dezembro do anno de 1834

<i>Auctor.</i>	<i>Título da Obra.</i>	<i>Edição.</i>	<i>Anno.</i>	<i>Pel.</i>	<i>Folios.</i>	<i>Observ.^{ões}</i>
"	<i>Dictionnaire Geographique de l'Asie et de Portugal.</i>	"	"	15.	Folios.	<i>Em Manuscripto com duas Anexas em Folios.</i>
<i>Calvet.</i>	<i>Histoire Universelle Sacree et profane depuis le commencement du Monde jusqu'à nos jours.</i>	<i>Strasbourg.</i>	<i>1755.</i>	8.	<i>Quarto.</i>	"
<i>M. P. L. Henry.</i>	<i>Histoire Generale des Auteurs.</i>	<i>Paris.</i>	<i>1729.</i>	9.	<i>Quarto.</i>	"

Fig. 3

Parte superior da página de início do inventário dos livros do extinto Convento do Espírito Santo e que deram entrada no Hospício da Terra Santa, a 30 de Dezembro de 1834 (BNP, Cód. 6963, f. 195). Imagem cedida pela BNP

A INVESTIGAÇÃO COLABORATIVA EM HUMANIDADES DIGITAIS: O CASO PORTUGUÊS³⁸

Luísa Alvim¹ e Teresa Costa²

¹CIDEHUS - Universidade de Évora

²CIDHEUS-Universidade de Évora; Centro de Estudos Clássico - Universidade de Lisboa
luisaalvim@uevora.pt, teresa.costa.23@gmail.com

Resumo. Pretende-se dar a conhecer o estado da questão da investigação portuguesa na área das Humanidades Digitais, através de um estudo sobre as publicações depositadas nos Repositórios Científicos de Acesso Aberto em Portugal, dos anos de 2012 a 2020. Tratando-se de uma área emergente, em Portugal, têm-se afirmado várias dinâmicas e projetos. Muitos investigadores, ligados às instituições de investigação, constituem já uma comunidade organizada com propósitos e objetivos comuns, formalizada nalgumas universidades em linhas de investigação. Foi realizada uma pesquisa bibliográfica, tendo sido recuperados vários tipos de documentos, provenientes de vinte e dois repositórios nacionais. Verificou-se que a maioria dos trabalhos recuperados era no âmbito da Ciência da Informação, seguida da Literatura e da História, concluindo-se que os investigadores estão dispersos pelas Artes e Humanidades, Ciências Sociais e Tecnologias. As linhas de investigação são constituídas por colaborações de investigadores, criando pontes entre as Humanidades e outras Ciências, comprovando-se a interdisciplinaridade. Verifica-se a pertinência do trabalho colaborativo que as Humanidades Digitais exigem, onde os especialistas nas tecnologias e ferramentas digitais, e noutras, auxiliam na produção de novos conhecimentos nas Ciências Sociais e Humanas.

1. INTRODUÇÃO

Neste trabalho, pretende-se dar a conhecer a investigação portuguesa na área das Humanidades Digitais, através de um estudo sobre as publicações depositadas no RCAAP (Repositórios Científicos de Acesso Aberto em Portugal). Muito se tem escrito e teorizado acerca da definição das Humanidades Digitais (Hockey, 2004; Unsworth, Schreibman, & Siemens, 2004). Estas encerram uma nova prática, um novo paradigma de investigar nas Humanidades. Tratando-se de uma área emergente, existe uma ampla discussão internacional sobre a definição de Humanidades Digitais, sendo difícil de formular e delimitar as suas fronteiras (Priani Saisó et al., 2014).

³⁸ Este trabalho é financiado por fundos nacionais através da Fundação para a Ciência e a Tecnologia, no âmbito do projeto UIDB/00057/2020.

As Humanidades Digitais visam ajudar a interpretar o impacto cultural e social da realidade digital, em que se torna necessário ter técnicas de análise de conteúdos, assim como de recolha e produção da informação; respondem e disponibilizam informação, favorecendo a sua divulgação e o seu estudo e englobam um conjunto de pesquisas e de tarefas que visam facilitar a utilização de recursos digitais nas Ciências Sociais e Humanas (Guerreiro, 2017).

A nível internacional, é visível a consolidação deste novo campo de saber através de indicadores como o surgimento de associações, departamentos de universidades que trabalham nesta área, revistas especializadas, congressos sobre o tema e programas de ensino na academia, etc.

Em Portugal, têm-se afirmado várias dinâmicas e projetos (Guerreiro, 2017, Guerreiro & Borbinha, 2014; Rollo, 2020). Muitos investigadores, ligados às instituições de investigação, constituem já uma comunidade organizada com propósitos e objetivos comuns, formalizada, em algumas universidades, em linhas de investigação, como a linha Humanidades Digitais e Investigação Histórica; ou em estruturas, cursos e grupos como o Laboratório de Humanidades Digitais (Lab_HDig) do Instituto de História Contemporânea (FCSH/UNL) (Alves, 2016); o Núcleo Património e Humanidades Digitais (Universidade de Coimbra); o curso de Mestrado em Humanidades Digitais (Universidade do Minho) e o Grupo de Investigação em Humanidades Digitais do CEHUM (U.Minho); a fundação da Associação das Humanidades Digitais (<http://ahdig.org/>). E temos ainda visto surgir novos grupos de interesse, como na Universidade de Évora.

2. HUMANIDADES DIGITAIS

Kim & Stommel (2018) advogam que a definição de uma disciplina se torna mais um exercício de exclusão do que de inclusão e, relativamente às Humanidades Digitais, afirmam que estas não são passíveis de uma definição clara, porque ao fazê-lo exclui-se o trabalho que é diverso e que constitui este campo de saber. É necessário tocar as bordas das Humanidades Digitais, abrindo-se a outros horizontes, para as entender e não as fechar numa definição. O seu rigor e a sua heterogeneidade são encontrados na discrepância do cânone das Humanidades Digitais (Kim & Stommel, 2018).

O espaço comunitário das Humanidades Digitais requer compromissos e diálogos que originam trabalhos académicos, e outros, menos convencionais, gerando novos paradigmas. O trabalho de investigação nesta área rompe com a hegemonia e o isolamento das disciplinas tradicionais das Humanidades.

Na introdução à obra sobre as Humanidades Digitais e o mundo lusófono (Pimenta & Alves, 2021), a investigadora G. del Rio Grande apresentou esta disciplina através de uma trajetória geoespacial, partindo do projeto *Index Thomisticus* de Robert Busa (1950), cruzando outras comunidades académicas como a Linguística computacional e a Ciência da Informação, etc., advogando a necessidade de se fazer história da investigação ao nível global nas Humanidades Digitais, para se obter um campo científico consolidado. Referiu que

inicialmente a língua inglesa salientava-se, nos primeiros projetos, mas foi possível encontrar outras árvores genealógicas para as Humanidades Digitais.

Na atualidade, o global é que define o campo geográfico das Humanidades Digitais, sendo a tecnologia o que une a investigação em todas as latitudes. A investigação aplicada contempla zonas mais teóricas, como Ciências da Comunicação, Sociologia, estudos sobre a tecnologia e o cânone tradicional: as edições digitais, as bibliotecas, os arquivos, a Linguística Computacional, a extração de textos e a *Web* semântica.

O Manifesto das *Digital Humanities* (THATCAMP & Dacos, 2010) considera as Humanidades Digitais uma “transdisciplina, portadora dos métodos, dos dispositivos e das perspectivas heurísticas ligadas ao digital no domínio das Ciências Humanas e Sociais”.

Constata-se que no campo das Humanidades Digitais existem inúmeras comunidades científicas específicas oriundas de diversas práticas, instrumentos ou objetos transversais: como arquivos orais, artes e literatura digital, lexicografia, sistemas de informação geográfica, digitalização de património cultural, científico e técnico, etc. (THATCAMP & Dacos, 2010). As Humanidades Digitais comportam uma comunidade alargada de investigadores de diferentes áreas afins ou complementares das Humanidades, que conciliam os conhecimentos e os métodos usados nas Ciências Sociais e Humanas com a realidade digital (Guerreiro, 2017).

Relativamente ao mundo lusófono, Pimenta e Alves (2021) traçaram um cenário das Humanidades Digitais em Portugal e no Brasil, referindo qual a produção do conhecimento nas diversas disciplinas que compõem as Humanidades, mediadas pelos recursos digitais e pelo uso da computação. O resultado do levantamento foi muito enriquecedor e diverso nas abordagens, métodos e conceitos, assim como nos objetivos e questionamentos.

Decorreram dois congressos pioneiros no campo das Humanidades Digitais em língua portuguesa, o primeiro na cidade de Lisboa, em 2015, organizado pela Universidade de Lisboa, Universidade Nova de Lisboa e Universidade de Évora, e em 2018, no Rio de Janeiro, organizado pela Fundação Getúlio Vargas, Universidade Federal do Estado do Rio de Janeiro, Fundação Casa Rui Barbosa, Universidade Federal Fluminense, Universidade Federal Rural do Rio de Janeiro. Estes congressos afirmaram e potenciaram as Humanidades Digitais no meio académico português e brasileiro. Constatou-se, a partir dos artigos destes congressos (Pimenta & Alves, 2021), uma diversidade de disciplinas e uma ampla gama de metodologias exploradas, sendo as áreas da Linguística e da Computação as mais representadas, situação esperada, já que esta foi a primeira relação nas Humanidades Digitais. Outras áreas se destacam, como a História, a Filologia, os Estudos Literários e Artísticos, a Comunicação, a Ciência da Informação e a Tradução. Estão também presentes a Literatura, as redes sociais, os arquivos e os museus. As aplicações dos métodos digitais aos temas de cada disciplina, usam os sistemas de informação geográfica, as técnicas de linguística computacional, as aplicações e linguagens *Web* e as ferramentas de desenvolvimento 3D.

Alves (2016) salientou que as Humanidades Digitais são mais uma comunidade de prática do que uma área ou campo de estudo, justificando que o termo Humanidades Digitais, em Portugal, começou a firmar-se no ano 2010, a ser utilizado em conferências, textos, blogues, etc. Mas, anteriormente, já existiam práticas nas investigações e na interpretação dos

dados, nas áreas da História, Filosofia, Literatura e Linguística, sem que se utilizasse formalmente o termo. Portanto, reafirmou a existência, em Portugal, de uma comunidade de prática em Humanidades Digitais, um grupo de investigadores e alunos que, dentro das Humanidades, incorporam tecnologias no método de trabalho. Os investigadores de áreas científicas e de cursos/universidades diferentes trabalham em conjunto, em parceria, logo em comunidade, sendo que a sua relação não é formal.

Desenvolveu-se, no meio académico português, um trabalho nas Humanidades Digitais de muita conexão e interdisciplinaridade entre investigadores da Geografia, Literatura, Linguística e Filosofia. O investigador Daniel Alves (IHC/NOVA FCSH) relatou que na universidade tentam a formalização e institucionalização desta área, tendo criado um Laboratório de Humanidades Digitais (Lab_HDig), onde se reúnem, discutem e elaboram projetos comuns e onde tentam formalizar um curso de pós-graduação na área de Humanidades Digitais na Universidade Nova de Lisboa. Salientou, também, que se pode correr um risco ao vincular o seu estudo à área que tiver desenvolvido mais projetos dentro da universidade, perdendo-se a característica da interdisciplinaridade. As Humanidades Digitais precisam de duas componentes, o digital e o cruzamento com outras disciplinas das Humanidades, com métodos e experiências diferentes (Aquino, 2020).

Em Portugal, várias equipas, nos centros de investigação em algumas universidades, têm dado passos na direção da interdisciplinaridade, por exemplo com a incorporação dos SIGs (Sistemas de Informação Geográfica), a Análise de Redes Sociais (adotada por investigadores da História Social), a Linguística Computacional (Processamento de Linguagem Natural, Mineração de Textos, entre outros), Arquivos Digitais, incorporação de metodologias 3D, etc. (Alves, 2016). Apesar da disparidade metodológica e conceptual, as Humanidades Digitais são forçadas a comprometer-se e a equacionar o seu âmbito epistémico, desvalorizando-se as questões relacionadas com a multiplicidade de teorias e de práticas (Gonçalves & Banza, 2013).

3. METODOLOGIA

Para a revisão da literatura, fez-se uma pesquisa sobre o conceito de Humanidades Digitais e explorou-se sobretudo a literatura relacionada com a realidade portuguesa.

Foi realizada, no dia 1 de junho de 2021, uma pesquisa bibliográfica no RCAAP (Repositórios Científicos de Acesso Aberto em Portugal), tendo sido utilizada a expressão de pesquisa “humanidades digitais”. A pesquisa visou obter resultados exclusivamente de Portugal pelo que se limitou a mesma apenas ao RCAAP, excluindo à partida os dados do Brasil, dos anos 2012 a 2020. Os resultados foram colocados num ficheiro Excel e identificados e excluídos os duplicados. Após a recolha de toda a informação, os dados foram trabalhados com recurso a tabelas *pivot* para a elaboração das várias tabelas e gráficos de suporte ao artigo e que ajudam, não só na análise dos dados obtidos, mas também na sua visualização e interpretação.

Analisou-se a produção científica portuguesa referente a todas as publicações recuperadas, com a expressão de pesquisa já mencionada. A abordagem a cada recurso depositado centrou-se nas seguintes variáveis: 1) distribuição dos recursos por anos de publicação; 2) tipologia de recursos; 3) proveniência dos recursos; 4) tipologia dos documentos; 5) assuntos; 6) autores.

Para a análise da variável assuntos dos recursos obtidos, utilizou-se o método de investigação documental para recolha e análise das informações de cada recurso, nomeadamente sobre as palavras-chave e sobre o resumo de cada artigo, por vezes fazendo uma leitura do texto integral dos mesmos.

Em termos metodológicos, recorreu-se à Bibliometria, que permitiu quantificar a informação recolhida, descrevê-la e, por fim, interpretá-la. Desta forma, foi possível analisar a produção científica portuguesa depositada no RCAAP sobre Humanidades Digitais até 2020.

4. RESULTADOS

No total, recuperaram-se 98 documentos, porém, e após acesso e leitura dos mesmos, foi possível reduzir a lista para 83 trabalhos únicos (excluíram-se duplicados que, apesar de estarem depositados em repositórios distintos, devido às afiliações dos autores, eram o mesmo documento).

4.1. POR ANO

Foi possível verificar que o interesse pela temática das Humanidades Digitais, em Portugal, tem vindo a aumentar nos últimos anos, como é possível observar na tabela 1, com 14 trabalhos em 2018 e 2020 e 16 em 2019.

Tabela 1. Número de trabalhos em Humanidades Digitais publicados por ano

Ano	N.º
2012	2
2013	4
2014	9
2015	7
2016	8
2017	9
2018	14
2019	16
2020	14
	83

Como já referido, os congressos de Humanidades Digitais, em língua portuguesa, decorreram em 2015 e 2018 e provocaram a criação de projetos de investigação e trabalhos colaborativos.

4.2. TIPOLOGIA

Obtiveram-se 83 documentos de 20 recursos distintos, dos quais 14 são repositórios institucionais, quatro publicações periódicas, um centro de recursos e umas atas. A maioria dos trabalhos (62) provém de repositórios, como se pode observar na tabela 2.

Tabela 2. Número de trabalhos em Humanidades Digitais por tipo de recurso

Recursos	N.º de Recursos	N.º de trabalhos
Repositorios	14	62
Publicações periódicas	4	8
Centro de recursos	1	10
Actas	1	3
Total	20	83

4.3. PROVENIÊNCIA

Relativamente à proveniência (tabela 3), destaca-se o Repositório da Universidade de Évora, com 20 trabalhos, o Estudo Geral da Universidade de Coimbra, com 11, o RUN, da Universidade de Lisboa, com 10, e, com igual número, a Linguateca (10).

Tabela 3. Número de trabalhos em Humanidades Digitais e sua proveniência

Recursos	N.º
Repositório Científico da Universidade de Évora	20
Estudo Geral - Universidade de Coimbra	11
Linguateca	10
RUN - Repositório Institucional da UNL	10
Repositório da Universidade de Lisboa	8
Cadernos BAD	4
Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas	3
Gaudium Sciendi Revista electrónica - Universidade Católica Portuguesa	2
Repositório Aberto da Universidade Aberta	2
Repositório Científico do Instituto Politécnico de Santarém	2
Repositório Institucional da Universidade Fernando Pessoa	2
Camões - Repositório Institucional da Universidade Autónoma de Lisboa	1
Journal of Digital Media & Interaction	1
Repositório Aberto da Universidade do Porto	1
Repositório Científico do Instituto Politécnico de Lisboa	1
Repositório Científico do Instituto Politécnico do Porto	1
Repositório do ISCTE-IUL	1
Revista Lusófona de Estudos Culturais	1
Sapientia - Universidade do Algarve	1
UBibliorum - Repositório Institucional da Universidade da Beira Interior	1

83

A Linguateca é um centro de recursos para o processamento computacional da língua portuguesa que mantém vários serviços, entre eles um catálogo de publicações. Outro tipo de recurso são as publicações periódicas científicas, das quais se pode destacar os Cadernos BAD, com quatro artigos, e as Actas do Congresso Nacional BAD, com três. Destaca-se, em número de trabalhos científicos, o Repositório Científico da Universidade de Évora, com 20 trabalhos, dos quais sete (da área da Ciência da Informação) são da autoria de Dália Guerreiro (CIDEHUS-UE), quatro dos quais em coautoria com outros investigadores. Salienta-se também Ana Paula Banza (CIDEHUS-UE), com quatro trabalhos (no âmbito da Filologia), dos quais dois em coautoria com Maria Filomena Gonçalves (CIDEHUS-UE).

4.4. TIPOLOGIA DOCUMENTAL

Dos trabalhos recuperados, a maioria são artigos científicos (39), seguidos de objetos de conferências (17), que vão desde comunicações, posters, apresentações em *PPT/slides* ou vídeos, capítulos de livros (12), entre outras tipologias documentais (tabela 4). Destaca-se, nesse caso, o número de objetos de conferências, bem representativo da realização de eventos e iniciativas que têm evidenciado as Humanidades Digitais em Portugal.

Tabela 4 Número de trabalhos em Humanidades Digitais por tipologia documental

Tipologia documental	N.º
Artigo científico	39
Objeto de conferência	17
Capítulo	12
Dissertação de Mestrado em Ciência da Informação	4
Ebook	3
Livro	2
Tese de Doutorado	2
Documento de trabalho	1
Índice Livro	1
Outro (editorial)	1
Recensão	1
	83

4.5. ASSUNTO

Relativamente aos assuntos, e sendo as Humanidades Digitais transversais a várias áreas do conhecimento, verificou-se que a maioria dos trabalhos recuperados era no âmbito da Ciência da Informação (16), seguida da História (13) e da Literatura (13) (tabela 5).

Tabela 5 Número de trabalhos em Humanidades Digitais por assunto

Assunto	N.º
Ciência da Informação	16
História	13
Literatura	13
Linguística	8
Humanidades digitais	6
Arquivos Digitais	4
Filologia	4
Turismo	4
Museologia	3
Educação	2
Património	2
TIC	2
Bibliotecas digitais	1
Cibernética	1
Codicologia	1
Cultura Digital	1
Filosofia	1
Preservação digital	1
	83

4.6. AUTORES

Relativamente à autoria, destaca-se Dália Guerreiro (CIDEHUS-UE), com doze trabalhos, cinco como única autora e sete em coautoria. Em termos de assunto, estes trabalhos são, na sua maioria, como já anteriormente referido, no âmbito da Ciência da Informação. Segue-se Diana Santos (Linguatca & ILOS; Universidade de Oslo), com oito trabalhos na área da Literatura e da Linguística, dos quais sete são trabalhos em coautoria. Por fim, refere-se Daniel Alves (IHC/NOVA FCSH), com cinco trabalhos, tendo a História como tema central. Tal evidencia bem a interdisciplinaridade das Humanidades Digitais, com trabalhos aqui em três áreas científicas distintas.

5. CONCLUSÕES

Nas publicações científicas recuperadas nesta pesquisa, constata-se um interesse crescente pelas Humanidades Digitais e que os investigadores estão dispersos pelas Artes e Humanidades, Ciências Sociais e Tecnologias.

Neste trabalho, recuperaram-se e analisaram-se documentos depositados no RCAAP, portanto todos os outros trabalhos científicos, de autores portugueses, por exemplo *ebooks*, livros, etc., quer pelas suas características físicas, quer outras (não estarem em acesso aberto, não estarem em repositórios institucionais I&D), não foram incorporados e estudados.

As linhas de investigação encontradas são constituídas por colaborações de investigadores, constituindo-se redes e criando pontes entre as Humanidades e outras Ciências, comprovando-se a interdisciplinaridade. Verifica-se a pertinência do trabalho colaborativo que as Humanidades Digitais exigem, onde os especialistas nas tecnologias e ferramentas digitais, e noutras, auxiliam na sua aplicação aos métodos de investigação tradicionais e à produção de novos conhecimentos nas Ciências Sociais e Humanas. As Humanidades Digitais, enquanto uma comunidade de prática, têm a característica de romper fronteiras e quebrar barreiras nas áreas científicas, proporcionando um diálogo que vem enriquecer o conhecimento científico.

REFERÊNCIAS

1. Alves, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. *Ler História*, (69), 91–103. Obtido de <https://journals.openedition.org/lerhistoria/2496> (2016).
2. Aquino, I. Digital Humanities como uma comunidade de práticas: entrevista com o professor Daniel Alves (IHC / NOVA FCSH). *Revista Aedos*, 12(26), 740–761 (2020).
3. Guerreiro, D. *Bibliotecas digitais para as Humanidades*. Universidade de Évora. Tese de doutoramento. Obtido de <http://dspace.uevora.pt/rdpc/handle/10174/23282> (2017).
4. Gonçalves, M.F., & Banza, A.P. (coord.). *Património Textual e Humanidades Digitais: da antiga à nova Filologia*. Évora: CIDEHUS (2013).
5. Guerreiro, D., & Borbinha, J. Humanidades Digitais: Novos desafios e oportunidades. *Cadernos BAD*, (1), 63–78 (2014).
Obtido de <http://www.bad.pt/publicacoes/index.php/cadernos/article/view/1060>
6. Hockey, S. The History of Humanities Computing. In *A Companion to Digital Humanities*. Oxford: Blackwell (2004).
Obtido de <http://www.digitalhumanities.org/companion/>.
7. Kim, D., & Stommel, J. (ed.). *Disrupting the Digital Humanities* (Vol. 148). punctum books. Obtido de <https://punctumbooks.com/titles/disrupting-the-digital-humanities/> (2018).
8. Pimenta, R., & Alves, D. (org.). *Humanidades Digitais e o mundo Lusófono* (Vol. 148). Rio de Janeiro: FVG Editora (2021).
9. Priani Saisó, E., Spence, P., Galina Russell, I., González-Blanco, E., Alves, D., Barrón Tovar, J., ... Sousa, M. C. Las humanidades digitales en español y portugués. Un estudio de caso: DíaHD/DiaHD. *Anuario Americanista Europeo*, 12, 5–18 (2014).
10. Rollo, M. F. Desafios e responsabilidades das Humanidades Digitais: preservar a memória, valorizar o Patrimônio, promover e disseminar o Conhecimento. O Programa Memória Para Todos. *Estudos Históricos* (Rio de Janeiro), 33(69), 19–44. <https://doi.org/10.1590/s2178-149420200001000003> (2020).
11. THATCAMP, & Dacos, M. *Manifesto das humanidades digitais*. Paris: ThatCamp. Obtido de <http://tcp.hypotheses.org/497>(trad. 2010).
12. Unsworth, J., Schreibman, S., & Siemens, R. (ed). *A Companion to Digital Humanities*. Oxford: Blackwell (2004).
Obtido de <http://www.digitalhumanities.org/companion/>.