



Ricardo M. Pimenta
Daniel Alves
ORGANIZADORES

HUMANIDADES DIGITAIS



IE O MUNDO LUSÓFONO



FGV EDITORA

COMITÊ CIENTÍFICO

Rita Marquilhas (Universidade de Lisboa)

Manuel Portela (Universidade de Coimbra)

Dália Guerreiro (Universidade de Lisboa)

Gimena del Rio Riande (CAICYT/CONICET; Universidade de Buenos Aires)

Sarita Albagli (IBICT)

Ivana Bentes (Universidade Federal do Rio de Janeiro)

**Ricardo M. Pimenta
Daniel Alves**
ORGANIZADORES

HUMANIDADES DIGITAIS



E O MUNDO LUSÓFONO



Copyright © 2021 Ricardo M. Pimenta e Daniel Alves

Direitos desta edição reservados a
FGV EDITORA
Rua Jornalista Orlando Dantas, 9
22231-010 | Rio de Janeiro, RJ | Brasil
Tels.: 0800-021-7777 | 21-3799-4427
Fax: 21-3799-4430
editora@fgv.br | pedidoseditora@fgv.br
www.fgv.br/editora

Todos os direitos reservados. A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação do copyright (Lei nº 9.610/98).

Os conceitos emitidos neste livro são de inteira responsabilidade dos autores.

1ª edição – 2021

Revisão de texto: Janaynne Carvalho do Amaral
Capa: Estúdio 513

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pela Biblioteca Mario Henrique Simonsen/FGV

Humanidades digitais e o mundo lusófono / Organização Ricardo M. Pimenta,
Daniel Alves. – Rio de Janeiro : Editora FGV, 2021.
375 p.

Inclui bibliografia.
ISBN: 978-65-5652-070-4

1. Humanidades digitais. 2. Ciência da informação. I. Pimenta, Ricardo
Medeiros. II. Alves, Daniel R. III. Fundação Getulio Vargas.

CDD – 306

Elaborada por Rafaela Ramos de Moraes –CRB 7/6625

Sumário

Prefácio

GIMENA DEL RIO RIANDE ----- 8

Apresentação

RICARDO M. PIMENTA
DANIEL ALVES----- 12

TRACEdb: Construcción de un archivo digital para investigadores de Humanidades Digitales

ALEJANDRO BIA
JESÚS JAVIER RODRÍGUEZ-SALA----- 15

WebSinC: buscas *on-line* em corpora sintaticamente anotados

ALINE SILVA COSTA
CRISTIANE NAMIUTI----- 37

Lisboa operária na última década do século XIX: um Sistema de Informação Geográfico aplicado à investigação histórica

ANA ALCÂNTARA----- 48

Novos desafios para antigas fontes: a experiência DOVIC na nova linguística histórica

CRISTIANE NAMIUTI
JORGE VIANA SANTOS----- 69

Narrativas transmídia como performances de leitura

FERNANDA BONACHO----- 90

A fotografia digital na transposição de documentos manuscritos históricos

JORGE VIANA SANTOS
GIOVANE SANTOS BRITO-----103

Métodos não supervisionados de *clustering* para a análise de textos literários: a seleção de materiais sobre grandes volumes de *corpus*

MARÍA LUISA FERNÁNDEZ RODRÍGUEZ
ROBERTO SAMARTIM-----116

Comunicação digital em museus e instituições de património cultural do Alentejo: estado da arte

NICOLA SCHIAVOTTIELLO
JOÃO BRIGOLA -----133

A obra potencial – para uma edição digital dos projetos editoriais de Fernando Pessoa

PEDRO SEPÚLVEDA
ULRIKE HENNY-KRAHMER -----161

O uso de mapas interativos como ferramenta de aprendizagem colaborativa e multimodal

SÍLVIA ARAÚJO
DIOGO CUNHA
ADRIANA SILVÉRIO
FRANCISCO ABRUNHOSA -----175

A Digital Platform for Literary Translation: collaborative translation and teaching

SUSANA VALDEZ
ISABEL OLIVEIRA MARTINS -----195

Projeto Corpus Eletrónico de Documentos Históricos do Sertão: a documentação epistolar

ZENAIDE DE OLIVEIRA NOVAIS CARNEIRO
MARIANA FAGUNDES DE OLIVEIRA LACERDA -----215

Humanidades Digitais e as instituições memoriais brasileiras: evolução em descompasso

AQUILES ALENCAR BRAYNER -----234

Sistemas de Informação para a Ciberdemocracia

RENATA MENDES DE ARAUJO -----249

Uso de Mineração de Dados no contexto eleitoral brasileiro: um caso de pesquisa à luz das Humanidades Digitais

JOSIR C. GOMES
MARCO SCHNEIDER
RICARDO M. PIMENTA -----267

Análise e mapeamento da organização técnica dos acervos digitais das instituições federais vinculadas ao Ministério da Cultura: fontes de informação para o desenvolvimento das Humanidades Digitais

DALTON LOPES MARTINS

JOYCE SIQUEIRA -----283

O processo de representação de ativos de informação: Web Semântica e Web de Dados no contexto das Humanidades Digitais

CLÁUDIO JOSÉ SILVA RIBEIRO -----310

Humanidades Digitais, Memória e Filatelia: uma aplicação prática para a Web

DIEGO A. SALCEDO -----330

A arte espectro neural no limiar das ciências e tecnologias: uma luta pela qualidade da vida

LUIZA HELENA GUIMARÃES -----349

Sobre os autores-----369

Humanidades Digitais e o mundo lusófono. Prefacio para las “otras” Humanidades Digitales

Gimena del Rio Riande (CONICET)

Por más de que siga resultando sorprendente la emergencia de las Humanidades Digitales en pleno siglo XXI, los abordajes y las reflexiones computacionales aplicados a las Humanidades cuentan con una larga trayectoria cuyo posible primer ya mítico mojón es el proyecto del Index Thomisticus del jesuita italiano Roberto Busa y la empresa IBM, comenzado hacia 1950. Este mito, compartido por las Ciencias de la Información, la Lingüística Computacional y las Humanidades Digitales, siempre ha resultado lo suficientemente expresivo y global para la mayor parte de estas comunidades académicas y suele mencionarse en manuales y trabajos científicos. No obstante, recuperar la memoria de los que nos antecedieron es parte ineludible del cualquier proceso de emancipación intelectual y, por ello, es aún urgente explorar los inicios de las Humanidades Digitales en los diferentes países, si es que realmente buscamos construir un campo científico consolidado a nivel local y global.

Contamos con varios posibles árboles genealógicos para las Humanidades Digitales, aunque, como es esperable, la mayor parte de ellos figura en los archivos más anglófonos. Si bien es interesante destacar que, desde el inicio encontramos proyectos en los que participaron investigadores de diferentes países, como es el caso del Dictionary of the Old Spanish Language (DOSL) en el Hispanic Seminary of Medieval Studies de la Universidad de Wisconsin, donde fuentes de financiación y profesores estadounidenses trabajaron codo a codo con especialistas españoles y fuentes medievales castellanas, lo que indica que las “otras” Humanidades Digitales (todas las que no hablan inglés) existían, aún con una menor capacidad de producción autónoma. Siempre es útil recordar que el conocimiento es un producto social, geopolítico y situado, y que los desbalances y las inequidades que encontramos en las macroeconomías se trasladan a los intercambios académicos. Así, no llama la atención que, si el inglés es hoy día “la” lengua científica, este aparato cultural de dominación instaurado desde corporaciones de software para investigación,



bibliotecas, publicación científica, que además dominan la construcción de los más sesgados rankings académicos, las Humanidades Digitales más visibles sean las Digital Humanities, aunque las “otras” Humanidades Digitales estén vivas y gocen de buena salud.

El problema es que ya hace tiempo que las Humanidades Digitales vienen esforzándose en definirse como “globales”, un término peligroso que, como bien nos enseñó Boaventura de Sousa Santos, a medida que se expande, excluye. En las Humanidades Digitales, congresos, mapas, programas y publicaciones vuelven una y otra vez sobre esta posible dimensión global, entendiendo a la tecnología –casi siempre mimetizada con la tecnología web- como el elemento capaz de aglutinar a investigadores e investigaciones de diferentes latitudes. Así, la apertura y la participación en el campo parecen desprenderse de una lectura algo simplificada del espíritu de la primera World Wide Web, sin una reflexión sopesada del mismo campo en las batallas globales del poder y la representación. Evidentemente, que una comunidad comparta algunos estándares o herramientas multilingües para realizar su trabajo (pensemos en la Text Encoding Initiative ó TEI para los primeros, o Voyant Tools para las segundas), y también para “encontrarse” (Twitter y las más variadas mailing lists), es un muy buen comienzo para construir comunidad, pero ello no implica equidad en el acceso a la investigación o a la tecnología, en el sentido de infraestructuras o alfabetización digital. Además, el elemento situado, que caracteriza a cualquier tipo de investigación (desde el uso de una lengua para comunicar ciencia, materiales, o las distintas herencias disciplinarias), tiñe indefectiblemente a la globalidad con distintos colores locales.

¿Qué pasa entonces cuando las voces de las “otras” Humanidades Digitales buscan insertarse en el “canon global”? ¿Cómo construimos un campo propio y, a la vez, cómo colaboramos con el mundo? Es verdad que en los últimos años podemos hablar de Humanidades Digitales hispanohablantes, lusófonas, o Humanidades Digitales en India, Australia, en África, o en el Sur Global, y que estas miradas complejizan aún más el problema de la globalización como proceso y resultado en el universo académico, al buscar los cortes en la lengua, los países o las regiones, a sabiendas de que, la mayoría de las veces no hay un único camino para comprender cómo y con quiénes estamos investigando y que, asimismo, la globalización ha fomentado, como terapia a los males que produce, la diáspora de académicos a países ricos.

Este libro da buena cuenta de la existencia de unas Humanidades Digitais lusófonas, de un mundo hipercomplejo e hiperconectado, que, tal y como lo señalan gran cantidad de los trabajos que lo componen, está construido por investigadores de Brasil, de Portugal, España, Alemania, algunos trabajando en sus países de origen, otros como parte de las diásporas por las Academias del Norte Global, otros como investigadores extranjeros especialistas en Lusofonía o cuestiones tecnológicas. De igual manera, si el lugar de enunciación es, desde su título, el del mundo lusófono, los textos que encontramos hablan tres lenguas, portugués, español e inglés. También, más allá de la geopolítica y el multilingüismo, la cuestión de la diversidad juega aquí un papel importante. Como ya hemos afirmado en otros lugares, aunque en el ámbito anglófono la etiqueta *Digital Humanities* encuentran una práctica disciplinar acotada por las convocatorias de las mismas agencias de financiación – la National Endowment for the Humanities, las agencias europeas – y alianzas como la Alliance for the Digital Humanities Organization (ADHO), en otras latitudes donde estas infraestructuras son más débiles o inexistentes, el campo se vuelve más difuso en lo que hace a las disciplinas que lo componen. A pesar de que no existe una sola y única explicación, una de las tantas ancla en los históricos fuertes lazos de colaboración y financiación de proyectos de investigación y publicaciones científicas Norte-Norte, es decir, América del Norte y países anglófonos europeos. A ello se suma un contexto tecnológico homogéneo, favorecido asimismo por corporaciones de producción de código, software, hardware, etc. Como es esperable, en las latitudes Sur del planeta este horizonte es muy diferente, lo que ha hecho que, para el tema que aquí nos toca, la investigación digital aplicada a las Humanidades haya discurrido por zonas más teóricas, a veces más cercanas a los Estudios sobre Medios, de Ciencias de la Comunicación, Sociología, o de los Estudios sobre Ciencia, Tecnología y Sociedad. Y esto también se lee en este compilado que da lugar a trabajos sobre el canon de las Digital Humanities – las ediciones digitales, bibliotecas y archivos, la minería de textos, la Lingüística Computacional y las innovaciones de la web semántica- y a otros sobre Transmedia y Neurociencias. Es decir, las tensiones de lo local y lo global, lo situado y lo globalizado que más arriba exponíamos con respecto a la geopolítica del conocimiento, quedan de manifiesto y sostienen, a su vez, al volumen.

Más allá de estas tensiones, una posible lectura que podemos realizar sobre *Humanidades Digitais e o mundo lusófono*, y que contribuye con una

mejor comprensión del campo sin entenderlo sesgadamente desde lo global o lo local, es la de la construcción de una cultura epistémica. El término, acuñado por la socióloga austríaca Karin Knorr-Cetina, hace referencia a los mecanismos históricos y coyunturales que, en un campo dado, determinan cómo conocemos lo que conocemos. Las culturas epistémicas son las culturas que crean y garantizan el conocimiento. Y esto creo es lo que surge de la lectura de las páginas que siguen. Todo punto de vista es la vista de un punto: si se quiere, este libro es una mirada lusófono sobre las Humanidades Digitales Globales. No hay en él una repetición colonizada de métodos y categorías de las ciencias desarrolladas en el Norte; todo lo contrario, en la amalgama de colaboraciones, proyectos y textos que ofrece parecen resonar con fuerza las palabras del gran José Martí: “Injértese en nuestras Repúblicas el mundo, pero el tronco ha de ser nuestras Repúblicas”.

Apresentação

O que se tem feito em Humanidades Digitais em língua portuguesa? A partir desta coletânea esperamos contribuir decisivamente não somente para responder perguntas como essa, mas igualmente para o esclarecimento sobre o que vem a ser esse cenário das Humanidades Digitais no Brasil e em Portugal. Apresentamos aqui uma produção lusófona do conhecimento nas diversas disciplinas que compõem as Humanidades mediadas pelo uso da computação, dos recursos digitais, seja como ferramenta, como metodologia, seja como efetivo constituinte do objeto de pesquisa.

Não pretendemos fazer um estado da arte exaustivo, mas fixar uma amostragem tão diversificada quanto possível do que tem sido as Humanidades nos últimos anos. Nas próximas páginas, veremos temas e abordagens distintas. Tão ricas para uma pretensa verdade, existência e identidade sobre as Humanidades Digitais lusófonas quanto as demais personalidades de Fernando Pessoa para a literatura portuguesa, sabemos que é tarefa quase inviável conseguirmos representar todas as Humanidades Digitais em uma única coletânea. Há riqueza demais de abordagens, métodos e conceitos, assim como objetivos e questionamentos egressos de campos disciplinares distintos. Uma coletânea em Humanidades Digitais é como uma breve fotografia tirada no calor do momento/movimento cuja beleza e curiosidade que suscita nos olhos que a fita não fazem jus aquilo que ela é, ou que já deixou de ser haja vista a exponencial aceleração tecnológica e informacional a qual todos nós estamos submetidos. Então porque fazê-la? Por que empreender em sua construção?

Os textos agora reunidos resultam essencialmente de uma pequena selecção de comunicações feitas a dois congressos pioneiros no campo das Humanidades Digitais em língua portuguesa. O primeiro realizado em 2015, em Lisboa, organização conjunta da Universidade Nova de Lisboa, Universidade de Lisboa, Universidade de Coimbra e Universidade de Évora (<https://congressohdpt.wordpress.com/>). O segundo organizado no Rio de Janeiro, em 2018, pela Fundação Getúlio Vargas, Universidade Federal do Estado do Rio de Janeiro, Fundação Casa de Rui Barbosa, Universidade Federal Fluminense, Universidade Federal Rural do Rio de Janeiro e Universidade do Estado do Rio de Janeiro (<https://eventos.fgv.br/hdrio2018>).



Embora existissem redes e contactos prévios no campo das Humanidades Digitais dos dois lados do Atlântico, o certo é que esses dois eventos potenciam este mesmo campo e de forma decisiva contribuíram para a sua afirmação no meio académico e no espaço público brasileiro e português.

Quer em 2015, quer em 2018, os organizadores desta coletânea lançaram o desafio aos colegas que apresentaram os seus trabalhos para produzirem os capítulos aqui publicados. Por vicissitudes várias, só agora é possível dar à estampa a esses contributos. Temos a noção que isso traz um risco a esta publicação, o de alguns dos temas, projetos e perspectivas estarem já a sofrer da pressão do tempo, que no caso das Humanidades Digitais é bem mais rápido a contribuir para a desatualização inerente à produção científica. A coletânea de textos tem uma estrutura cronológica, seguindo as datas de realização dos dois congressos já referidos.

A diversidade de disciplinas representadas pelos autores que aceitaram o desafio, a riqueza de temáticas estudadas e a ampla gama de metodologias exploradas levam-nos a considerar que ainda faz sentido esta publicação. No caso dos autores, vemos que as áreas da Linguística e da Computação são as mais representadas, algo que não se apresenta como estranho face ao que foi tradicionalmente a ligação precoce entre essas duas disciplinas, ainda antes daquilo que hoje chamamos de Humanidades Digitais. Mas a diversidade é significativa e passa pela História, pela Filologia, pelos Estudos Literários, Artísticos e Culturais, pela Comunicação, pela Ciência da Informação ou pela Tradução.

Quanto às temáticas dos textos, temos um panorama bastante amplo e representativo do que são hoje as Humanidades Digitais, área em que os objectos de estudo podem ir do texto simples ao multimédia complexo. Textos de reflexão sobre as Humanidades Digitais também estão presentes, mas em menor número. O centro de gravidade dos materiais desta coletânea são as aplicações concretas de métodos digitais a temas, de certo modo, clássicos dentro de cada disciplina, seja através do uso de sistemas de informação geográfica, de técnicas da linguística computacional, de aplicações e linguagens web ou de ferramentas de desenvolvimento 3D. Estuda-se a história, mas também se reflete sobre e analisa a democracia; a arte e fotografia têm destaque, assim como a literatura e as redes sociais; o património e a memória, os museus e os arquivos estão também presentes.

Os organizadores desta coletânea querem agradecer aos colegas que aceitaram este desafio e tiveram a paciência de esperar pela sua publicação em tempos tão difíceis

como os que vivemos em meio à pandemia da COVID-19. Agradecemos também os organizadores e demais participantes que individualmente e coletivamente contribuíram para o sucesso dos dois eventos científicos realizados em Portugal e no Brasil, sendo a razão inicial de pensamento e escrita dos textos que agora se publicam. Por fim, queremos agradecer ao comitê científico pelo apoio, à professora doutora Gimena del Rio Riande pelo generoso prefácio, ao Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) juntamente com seu programa de pós-graduação em ciência da informação (PPGCI/Ibict-UFRJ) — que completa 50 anos — e a Universidade Nova de Lisboa, que sempre nos deram a liberdade e o incentivo para perseverar em nossos intuitos, respectivamente representados pelo Laboratório em Rede de Humanidades Digitais – Larhud (<http://www.larhud.ibict.br>) e pelo Laboratório de Humanidades Digitais (<https://dhlab.fcsh.unl.pt/>). Acreditamos estar a dar um passo importante e definitivo para o desenvolvimento das Humanidades Digitais de língua portuguesa com esta bela coletânea.

Boa leitura!

Ricardo M. Pimenta
Daniel Alves

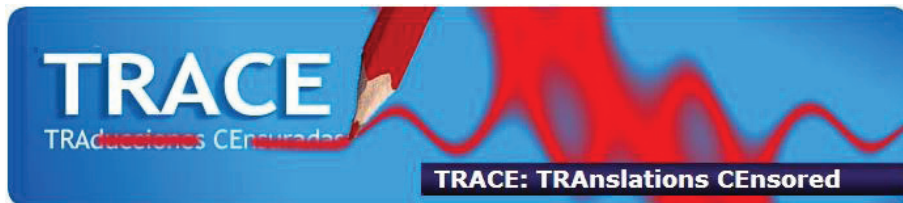
TRACEdb: Construcción de un archivo digital para investigadores de Humanidades Digitales

Alejandro Bia
Jesús Javier Rodríguez-Sala

INTRODUCCIÓN

El Proyecto TRACE comenzó alrededor del año 1997¹, cuando un grupo de investigadores del Departamento de Lenguas Modernas de la Universidad de León (ULE) y de la Universidad del País Vasco (UPV / EHU) decidió abordar un estudio coordinado de la historia de la traducción en España en el siglo XX.

FIGURA 1 - LOGO DEL PROYECTO TRACE



Fuente: Creación propia de los autores (2020).

Una de las fuentes más fiables para entender el paisaje cultural de la época es el fondo Archivo General de la Administración española (AGA) en el que se conservan los archivos que mantienen un registro de los procedimientos de los censores durante el régimen de Franco. Este archivo contiene información, no sólo acerca de los productos culturales que llegaron a la audiencia, sino también de los que no lo hicieron. Estos datos, que se han comparado y complementadas por otras fuentes (índices bibliográficos de otras bibliotecas o archivos, bases de datos, etc.), han resultado ser esencialmente relevantes para evaluar qué material textual

¹ Disponible en: <<https://trace.unileon.es/en/introduction/>>. Acceso en: 22 jun. 2020.

fue importado y traducido, en lugar de establecer el alcance del estudio utilizando una selección de criterios a priori basado en la cultura de origen (TRACE/ULE, History of TRACE).

El proyecto TRACEssoftTools (TST) es uno de los cuatro subproyectos que forman parte del proyecto TRACE, éste implica el estudio de la traducción y la censura en España de diferentes géneros textuales en diferentes combinaciones de idiomas entre 1939 y 1985. TRACE realiza un intensivo uso de marcado XML, TEI, y traduce memorias (TMX) para construir y aprovechar un corpus alineado paralelo multilingüe. Para este propósito se han utilizado herramientas como Giza++ (LIANG; WONG; CHAO, 2011), MDR (MONOSTORI; ZASLAVSKY; BIA, 2001), Bitext2tmx (SAVELLI, 2008; SCHREIBMAN *et al.*, 2007).

FIGURA 2 - ALGUNAS HERRAMIENTAS UTILIZADAS PARA ALINEAR TEXTOS



Fuente: Creación propia de los autores (2020).

El trabajo del grupo TST implica el desarrollo de ciertas herramientas y técnicas informáticas para facilitar el análisis contrastivo de textos en una doble vertiente:

- En primer lugar, da soporte al resto de subproyectos TRACE, desarrollando herramientas para el alineamiento paralelo de textos y marcado automático (BIA, 2015), y desarrollando la base de datos y la aplicación informática para el registro digital descrito en el presente trabajo.

- Por otra parte, se alimenta del corpus paralelo TRACE y de metadatos para llevar a cabo su propia investigación sobre temas como Data Mining, Text Mining y Estilometría.

EL ARCHIVO DIGITAL TRACE

Los corpus textuales digitalizados, por su parte, permiten la aplicación de herramientas avanzadas de análisis textual sistemático. El grupo TRACE-ULE ha desarrollado en proyectos previos algunos programas básicos para limpiar textos escaneados y digitalizados y los programas de alineamiento y conversión de los corpus a bases de datos SQL. La finalidad ahora es dar un salto cualitativo e innovador con la integración de todos los subcorpus desarrollados por TRACE con programas de análisis-textual más sofisticados.

Entre los objetivos del proyecto están: la creación, publicación online, ampliación y mantenimiento de la base de datos TRACE-ULE (1939-1985) y del corpus paralelo inglés-español, el desarrollo de las herramientas y el software ya existentes y la creación de nuevas herramientas de análisis textual, así como la creación de un archivo digital de recursos online, que es el tema del presente artículo. El propósito de éste archivo digital es desarrollar una plataforma Web que soporte tanto las bases de metadatos de TRACE, como el corpus bilingüe para su consulta online.

Preferimos llamar a esta aplicación un “archivo digital” en lugar de una “biblioteca digital”, ya que los archivos son bastante distintas de las bibliotecas con respecto a sus funciones y organización, aunque en este caso la arquitectura de la aplicación de software es equivalente a la de una biblioteca digital (uno altamente especializado), que comprende un catálogo de búsqueda y los documentos correspondientes. El propósito y funciones son de alguna manera diferentes a las de una típica biblioteca digital. El archivo digital de TRACE, descrito aquí, es una aplicación orientada a la investigación que tiene que incluir no sólo las obras originales, sino que también sus múltiples traducciones a diferentes idiomas. El catálogo es en realidad una base de datos especializada de metadatos de investigación que procura apoyar consultas avanzadas.

Debido a la variada naturaleza de las obras (teatro, cine, libros) y el gran número de investigadores, la situación de partida era en cierto modo caótica, ya que cada investigador poseía una base de datos de carácter personal que, a pesar de ser similar a las bases de datos del resto de los investigadores del grupo, presentaba algunas diferencias derivadas de la naturaleza particular de su propio trabajo. Para dificultar más las cosas, el grupo utilizaba dos tecnologías de bases de datos diferentes: Microsoft Access y FileMaker. Esta situación hizo que la integración de las diferentes bases de datos individuales en una única base de datos integrada para la aplicación (MySQL), fuera una tarea muy difícil.

La aplicación permite cuatro tipos de usuarios o roles para trabajar con los datos de la aplicación, de los cuales, uno es un tipo de usuario no registrado y el resto serán usuarios registrados con diferentes niveles de privilegios, en todos los casos, cualquier usuario registrado deberá estar adscrito a alguno de los subcatálogos o grupos de trabajo de la aplicación. Estos roles son:

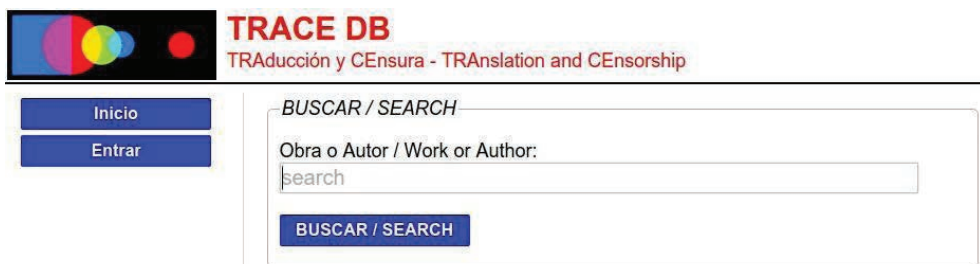
- Rol “Guest” (usuario no registrado): podrá hacer búsquedas sencillas y visualizar los metadatos y ficheros públicos de cualquier obra registrada en la base de datos.
- Rol “Visit” (usuario registrado): Podrá hacer búsquedas avanzadas y visualizar los metadatos y ficheros públicos y privados de las obras pertenecientes al subcatálogo al cual está adscrito. Adicionalmente dispondrá de un historial en el que se irán guardando todas las búsquedas realizadas durante una sesión de trabajo.
- Rol “User” (usuarios registrados): Además de las funciones del rol “Visit” podrá crear nuevas obras y editarlas.
- Rol “Admin” (usuario registrado): Es el usuario que administra un grupo o subcatálogo (solamente podrá haber un administrador por grupo), el cual, además de poder visualizar y editar todas las obras de dicho subcatálogo, también podrá gestionar los ítems de las listas desplegables del formulario de edición de obras, así como activar/desactivar y cambiar el rol (“Visit” \Leftrightarrow “User”) de los usuarios adscritos al grupo que administra.

Adicionalmente, y sólo para tareas administrativas, existe un quinto rol, que hemos denominado “Jefe”, encargado de gestionar los usuarios y los grupos

(subcatálogos) de la aplicación, es decir, darlos de alta, de baja o modificarlos (p.e.: cambiar el nombre o el administrador de un grupo, asignar usuarios a un grupo, etc.).

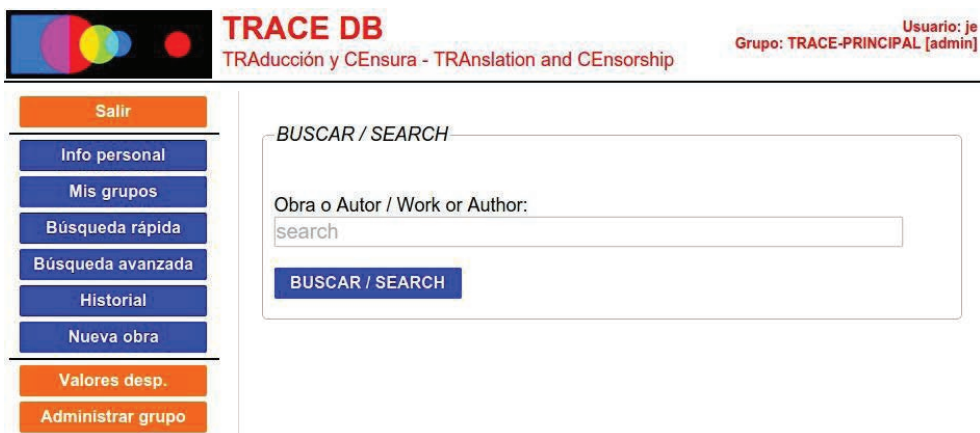
Las Figuras 3 y 4 muestran el formulario de búsqueda sencilla, cuando la aplicación es usada por un usuario “Guest” o un usuario “Admin”.

FIGURA 3 - EL USUARIO “GUEST” SOLO PUEDE HACER BÚSQUEDAS SENCILLAS



Fuente: Captura de pantalla del programa TRACE (2020).

FIGURA 4 - UN USUARIO REGISTRADO (P.E.: “ADMIN”) PUEDE REALIZAR MÁS OPERACIONES

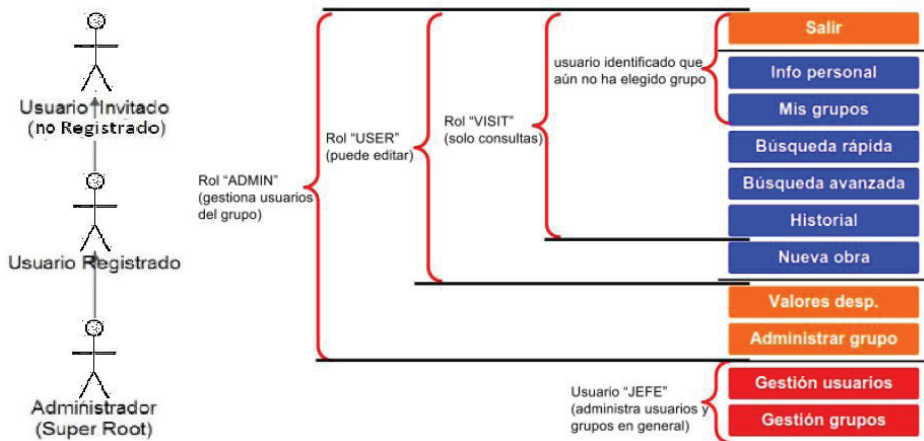


Fuente: Captura de pantalla del programa TRACE (2020).

Como se observa en la Figura 4, el usuario registrado puede realizar algunas opciones especiales como, en el caso de “Admin”, administrar los usuarios del grupo o editar la lista de opciones de los cuadros desplegables que utilizados para seleccionar valores para muchos de los campos de entrada de datos.

La Figura 5 muestra el menú de opciones completo para un usuario “Admin”, que además desempeña el rol de “Jefe”, es decir, también realiza las tareas administrativas de gestionar los usuarios y los grupos de la aplicación.

FIGURA 5 - MENÚ DE OPCIONES COMPLETO DE LA APLICACIÓN



Fuente: Creación propia de los autores (2020).

Cuando un usuario se identifica en el sistema, antes de poder trabajar, debe indicar con qué grupo quiere hacerlo, por lo que la aplicación sólo muestra los dos primeros botones azules, “Info personal” y “Mis grupos”, hasta que el usuario a seleccionado dicho grupo (ver Figura 6). Un mismo usuario puede pertenecer a varios grupos y tener un rol diferente en cada uno de ellos.

FIGURA 6 - GRUPOS DISPONIBLES (Y ROLES) PARA UN USUARIO IDENTIFICADO



Fuente: Captura de pantalla del programa TRACE con dibujos superpuestos de los autores (2020).

La aplicación permite realizar diversas clases de consulta, desde consultas sobre las obras originales, a consultas sobre las obras derivadas (por ejemplo, traducciones). La Figura 7 muestra el formulario de búsqueda avanzada (sólo disponible para usuarios registrados). El usuario podrá construir cualquier consulta añadiendo diferentes criterios de filtrado, además de los más comunes que se muestran por defecto. Los resultados se podrán ordenar por diferentes campos.

FIGURA 7 - FORMULARIO DE BÚSQUEDA AVANZADA

Fuente: Captura de pantalla del programa TRACE (2020).

Sobre los datos que se almacenan sobre una obra, en las Figuras 8 a 13 se muestran los campos disponibles agrupados en diferentes bloques o pestañas, todos ellos conforman el formulario de entrada de datos de obras (botón “Nueva obra”).

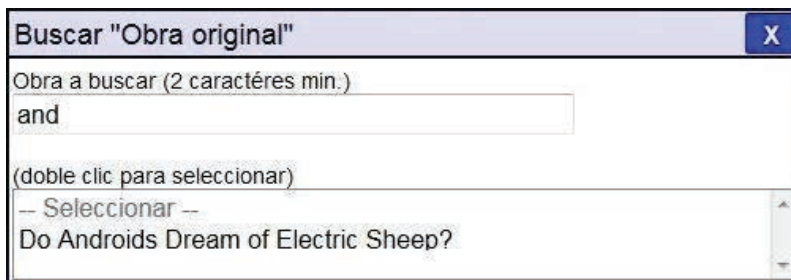
FIGURA 8 - FORMULARIO DE ENTRADA DE DATOS: INFORMACIÓN GENERAL

Fuente: Captura de pantalla del programa TRACE (2020).

Se pueden ingresar tres títulos diferentes para una misma obra (incluyendo títulos traducidos): el título real, tal como aparece en el documento de origen, el título normalizado (decidido por los catalogadores para unificar todas las versiones o derivados de una obra determinada), y el título de la obra fuente, cuando el trabajo actual se deriva de otro. La Figura 8 muestra un ejemplo, donde un guión de una película bajo el título español de *El cazador implacable*, se corresponde con el título normalizado *Blade Runner* (la famosa película de Ridley Scott de 1982), que a su vez es una modificación de la adaptación al cine de la novela *Do Androids Dream of Electric Sheep?* de Philip K. Dick (campo “obra fuente”). Este es un ejemplo interesante para el caso, ya que, para empeorar las cosas, existe otra versión de la película titulada *Blade Runner (Director’s Cut)*, especialmente editada en 1992, en el décimo aniversario de la primera versión. Se ha utilizado este ejemplo para crear diferentes registros para la novela original, las dos versiones de la película, y las diferentes traducciones de cada una de ellas.

Para simplificar la entrada de datos y ayudar a enlazar correctamente un título concreto con su título normalizado o con el título de su obra fuente, se proporciona una búsqueda de subcadena, escribiendo algunas letras normalmente será suficiente para encontrar el título correcto para el campo dado (Figura 9).

FIGURA 9 - BÚSQUEDA DE SUBCADENA PARA SELECCIÓN RÁPIDA



Fuente: Captura de pantalla del programa TRACE (2020).

Hemos reducido la información redundante en el diseño de la base de datos de la aplicación, mediante la creación de un tipo de registro de usos múltiples que pudiera contener información de los diferentes tipos de obras (libros, guiones escénicos de diferentes clases, etc.). Sin embargo, usando un enfoque adaptativo,

las interfaces son diferentes para cada tipo de obra, lo cual permite presentar visualizaciones diferenciadas para cada una de ellas.

La Figura 7 muestra la pantalla de entrada de datos para el guión desarrollado, la cual, es diferente de la pantalla para otros tipos de trabajos.

FIGURA 10 - FORMULARIO DE ENTRADA DE DATOS: INFORMACIÓN DETALLADA

The screenshot shows a web-based data entry form for the TRACE database. At the top, there are navigation tabs: 'Información General', '+ Info' (selected), 'TRACE', 'Responsabilidades', 'Archivos', and 'Logs'. The form is organized into several sections with input fields:

- Top Section:** 'Cine, teatro o canal' (text input), 'Compañía' (text input).
- Second Section:** 'Hora' (text input), 'Duración' (text input), 'Nº espectadores' (text input), and 'Rendimiento' (dropdown menu with '-- Seleccionar --').
- Third Section:** 'Lugar' (text input), 'Evento' (text input).
- Fourth Section:** 'Colección' (text input), 'Serie / Ciclo' (text input), and 'Número serie / Episodio' (text input).
- Fifth Section:** 'Fuentes de la obra' (text input) and 'Ubicación de la obra' (text input).
- Sixth Section:** 'Editorial' (text input), 'Edición' (text input), and 'Páginas' (text input).
- Seventh Section:** 'ISBN' (text input), 'Tirada' (text input), and 'Precio/Recaudación' (text input).

At the bottom of the form, there are two blue buttons: 'Guardar' and 'Guardar y siguiente'.

Fuente: Captura de pantalla del programa TRACE (2020).

Los metadatos de investigación se obtienen de varias fuentes, siendo el Archivo General de la Administración (AGA) la más frecuente.

El AGA mantiene los detalles administrativos e históricos sobre las traducciones, como horas y fechas, traductor, clasificaciones oficiales y morales, localización geográfica, nombre del solicitante, correcciones de la censura, y referencias a otros ficheros relacionados (véase la Figura 11).

FIGURA 11 - FORMULARIO DE ENTRADA DE DATOS: INFORMACIÓN DE INVESTIGACIÓN OBTENIDA DE LOS FONDOS AGA

The screenshot shows a web-based data entry form for the TRACE program. At the top, there are navigation tabs: 'Información General', '+ Info', 'TRACE' (which is highlighted), 'Responsabilidades', 'Archivos', and 'Logs'. The form is divided into several sections:

- PARES / Fondos AGA:** A dropdown menu with "-- Seleccionar --".
- Expediente:** A text input field.
- Signatura AGA:** A text input field.
- Año expediente:** A text input field.
- Fecha formación expediente:** A text input field.
- Fecha entrada:** Three separate input fields for Day, Mes, and Año.
- Fecha resolución:** Three separate input fields for Day, Mes, and Año.
- Fecha depósito:** Three separate input fields for Day, Mes, and Año.
- Signatura TRACE:** A text input field.
- Caja TRACE:** A text input field.
- Calificación oficial:** A dropdown menu with "-- Seleccionar --".
- Calificación moral:** A dropdown menu with "-- Seleccionar --".
- Provincia:** A dropdown menu with "-- Seleccionar --".
- Peticionario:** A text input field.
- Corrección censura:** A text input field.
- Expedientes relacionados:** A text input field.

At the bottom of the form, there are two buttons: "Guardar" and "Guardar y siguiente".

Fuente: Captura de pantalla del programa TRACE (2020).

Una característica de este proyecto, es la necesidad de tratar con diferentes roles y responsabilidades relacionadas con los diferentes tipos de obra y géneros literarios. Dependiendo del tipo de obra, será necesario asignar diferentes responsabilidades (roles): autores, editores, directores, intérpretes, dobladores, productores, letristas, compositores, etc.

El uso de roles personalizados definibles en lugar de un conjunto fijo de posibles roles predefinidos permite una mayor flexibilidad. De esta manera, los nuevos roles pueden ser añadidos por el propio usuario que introduce los datos del registro, siempre que surja la necesidad (véase la Figura 12).

FIGURA 12 - FORMULARIO DE ENTRADA DE DATOS: RESPONSABILIDADES PARA CADA TIPO DE TRABAJO



Fuente: Captura de pantalla del programa TRACE (2020).

Adicionalmente, en el apartado de “Responsabilidades” se dispone de opciones de búsqueda de subcadena para selección rápida, tanto para los nombres de responsables, como para el rol o tipo de responsabilidad desempeñada (análogos a los mostrados en la Figura 9).

Al igual que en una biblioteca digital, se pueden asociar uno o más documentos de cualquier formato a cualquier registro del catálogo. Dichos documentos pueden ser de muy distinta naturaleza, desde la obra actual digitalizada, hasta cualquier otro documento remotamente relacionado con dicha obra: escritos, resúmenes, ediciones alineadas paralelas, ficheros del archivo AGA, notas de investigación, facsímiles, biografías, e incluso recursos audiovisuales o enlaces (URLs) a diversas webs (véase la Figura 13).

FIGURA 13 - FORMULARIO DE ENTRADA DE DATOS: DOCUMENTOS ASOCIADOS

Editar obra

Información General + Info TRACE Responsabilidades Archivos Logs

Dirección web Subir fichero

W Link a libro en PDF (25/04/2016 21:22:05)

I Blade_Runner_poster.jpg (File - 16/6/2015 11:17:30)

digital resources

text multimedia

Fuente: Captura de pantalla del programa TRACE, con dibujos agregados por los autores (2020).

Adicionalmente, el usuario que introduce los archivos, puede indicar si dicho recurso será de acceso público o privado (véase las casillas “check” en la Figura 13) para quien esté realizando una búsqueda en el catálogo.

ADMINISTRACIÓN DE GRUPOS

Una de las tareas específicas de los administradores de grupo (rol “Admin”) es la de gestionar el contenido de los cuadros desplegables del formulario de entrada de obras, en varios apartados dicho formulario (ver Figuras 8, 10 y 11), también en el formulario de búsqueda avanzada (ver Figura 7), se dispone de cuadros de selección cerrada en los que el usuario puede seleccionar uno de los ítems disponibles, dichos ítems se gestionan mediante esta funcionalidad.

FIGURA 14 - GESTIÓN DE VALORES EN CAMPOS DESPLEGABLES

Gestión de valores en campos desplegables

Seleccionar categoría
Área textual

Lista de ítems

Audiovisual
Narrativa
Poesía
Teatral

Actualización de la lista: 'Área textual'

Nuevo ítem
Nuevo ítem

Guardar

Ítem: 'Audiovisual'
Audiovisual

Modificar Eliminar

Fuente: Captura de pantalla del programa TRACE (2020).

En la Figura 14 se muestra como para la categoría “Área textual” hay disponibles cuatro ítems, con las opciones a la derecha del formulario es posible añadir nuevas opciones a esta categoría, modificar algunas de ellas o eliminarla.

La otra tarea específica del rol “Admin” es la de administrar a los usuarios de su grupo.

FIGURA 15 - ADMINISTRACIÓN DE GRUPO

The screenshot shows the TRACE DB interface for user management. At the top left, it says "TRACE DB" and "TRAducción y CEnsura - TRAnslation and CEnsorship". At the top right, it shows the user "jefe" and the group "TRACE-PRINCIPAL [admin]". The main content is titled "Aministración del grupo 'TRACE-PRINCIPAL'". On the left, there is a list of users: "camino (user)", "cristina (user) [*]", "elena (visit)", and "sergio (user)". On the right, there is a form for the user "elena". The form has a checkbox for "Activo:" which is checked. Below that, there is a dropdown menu for "Rol de usuario:" with "Consulta" selected.

Fuente: Captura de pantalla del programa TRACE (2020).

Con esta opción el administrador podrá activar o desactivar los usuarios de su grupo, un usuario desactivado no podrá entrar en la aplicación (aparecerá en cursiva y con un asterisco en la lista de usuarios). Además, el administrador podrá modificar el rol de los usuarios de su grupo, cambiándolos de “Visit” (consulta) a “User” (edición) y viceversa.

Explícitamente, se indicó como prerequisite de la aplicación que no fuese tarea del administrador de un grupo el añadir o eliminar usuarios del grupo que gestiona, esa debía ser función del rol “Jefe”.

ADMINISTRACIÓN DE USUARIOS Y GRUPOS (SUBCATÁLOGOS)

Como se ha comentado anteriormente, existe un rol “Jefe” encargado de realizar las tareas administrativas de gestionar los usuarios y los grupos de la aplicación, entendiendo por grupo al conjunto de usuarios que tienen acceso a un subcatálogo en concreto, es decir, aquellos usuarios que pueden agregar, editar y/o consultar las obras de un determinado subcatálogo dentro del catálogo general de la aplicación TRACE. Naturalmente, si por alguna razón, se estima necesario

que un usuario pueda tener acceso a varios subcatálogos (p.e.: investigadores o personal de biblioteca), la aplicación admite que un mismo usuario pueda tener acceso a dichos subcatálogos, pudiendo éste indicar con cuál de ellos va a trabajar en cada momento (ver Figura 6, opción “Mis grupos”).

FIGURA 16 - GESTIÓN DE USUARIOS 1. AÑADIR USUARIOS

Gestión de usuarios

Usuarios registrados en la aplicación TRACE

<p>Lista de usuarios (nº grupos):</p> <ul style="list-style-type: none"> camino (9) cristina (2) elena (2) jefe (3) leticia (11) luisa (11) sergio (2) 	<p>Nombre usuario:</p> <input type="text" value="Nombre usuario"/>
	<p>Contraseña:</p> <input type="text" value="Contraseña"/>
	<input type="button" value="Añadir"/>

Fuente: Captura de pantalla del programa TRACE (2020).

La Figura 16 muestra un listado de todos los usuarios registrados en la aplicación, indicando entre paréntesis el número de grupos a los que pertenece. Para dar de alta un nuevo usuario bastará con indicar un nombre de usuario y una contraseña que el usuario podrá cambiar al entrar a la aplicación.

FIGURA 17 - GESTIÓN DE USUARIOS 2. MODIFICAR/ELIMINAR USUARIOS

Gestión de usuarios

Usuarios registrados en la aplicación TRACE

<p>Lista de usuarios (nº grupos):</p> <ul style="list-style-type: none"> camino (9) cristina (2) <li style="background-color: #f4a460;">elena (2) jefe (3) leticia (11) luisa (11) sergio (2) 	<p>Nombre usuario:</p> <input type="text" value="elena"/>
	<p>Contraseña:</p> <input type="text" value="Contraseña"/>
	<input type="button" value="Modificar"/> <input type="button" value="Cancelar"/> <input type="button" value="Eliminar"/>

Fuente: Captura de pantalla del programa TRACE (2020).

Al elegir un usuario de la lista la opciones del formulario cambiarán para que se pueda modificar su contraseña (medida que se tomaría si el usuario notifica haber olvidado su contraseña) o eliminarlo de la aplicación, teniendo en cuenta que no se puede eliminar a un usuario que esté adscrito a al menos un grupo, es decir, antes de eliminar un usuario definitivamente del sistema es preciso haberlo eliminado de cada uno de los grupos a los que pertenezca (entre paréntesis aparecería el valor “0”).

Una vez se han creado uno o varios usuarios, se pueden crear grupos, ya que una condición necesaria para crear un grupo es la de asignarle un administrador (no puede haber un grupo sin administrador). Una vez creado un grupo, éste pasa a definir un nuevo subcatálogo en la aplicación TRACE (véase campo “Subcatálogo” en Figura 8). La Figura 18 muestra el formulario de creación de nuevos grupos junto con el listado de grupos existentes y la información del grupo seleccionado.

FIGURA 18 - GESTIÓN DE GRUPOS I: DETALLE INFORMACIÓN DE GRUPO

Fuente: Captura de pantalla del programa TRACE (2020).

Seleccionando un grupo y pulsando el botón rojo “Ver Usuarios” (Figura 18) se tiene acceso a la función para gestionar los usuarios del grupo, se dispone de un listado de los usuarios que ya tiene el grupo y la opción de añadir nuevos, indicando su nombre y el rol (“Visit” o “User”) que dicho usuario desempeñará en este grupo.

FIGURA 19 - GESTIÓN DE GRUPOS 2: AÑADIR USUARIOS A UN GRUPO

The screenshot shows the 'Gestión de grupos' interface. At the top, there is a 'Nuevo Grupo/Catálogo' section with input fields for 'Nombre grupo:' (containing 'Nombre grupo') and 'Administrador:' (containing 'ID Administrador'), and an 'Añadir Grupo' button. Below this is a 'Listado grupos/catálogo' list containing various group names like 'TRACE-PRINCIPAL', 'TRACEci (1951-1975)', etc. To the right, the 'Información de grupo/catálogo 'TRACE-PRINCIPAL'' panel is visible. It includes a red 'Info. Grupo' button, a 'Lista de usuarios:' list with 'camino (user)', 'cristina (user)', 'elena (visit)', and 'sergio (user)'. The 'Usuario:' field contains 'Nombre usuario', and the 'Rol:' dropdown is set to '-- Seleccionar --'. An 'Añadir' button is at the bottom.

Fuente: Captura de pantalla del programa TRACE (2020).

Si se selecciona un usuario de la lista (Figura 19), cambian las opciones del formulario para permitir, bien cambiar su rol (de “Visit” a “User” o viceversa), o bien eliminarlo del grupo. Pulsando el botón rojo “Info. Grupo” se volvería al formulario tal y como se muestra en la Figura 18.

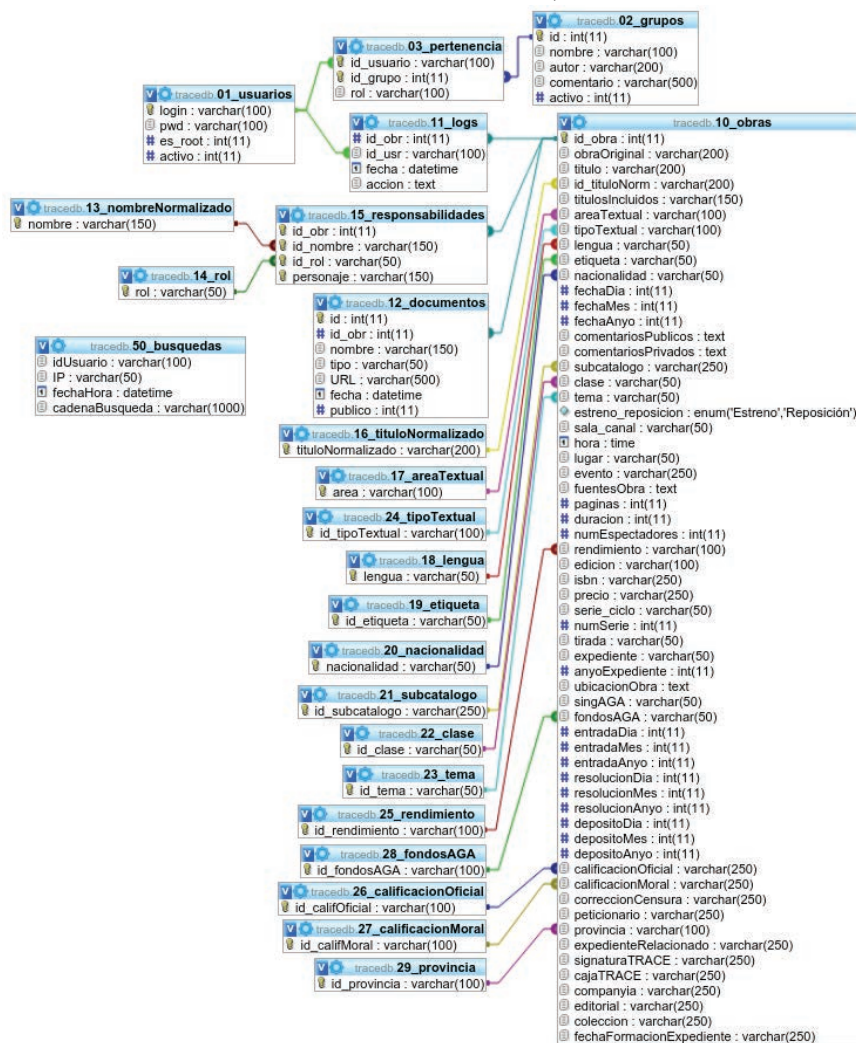
FIGURA 20 - GESTIÓN DE GRUPOS 3. MODIFICAR/ELIMINAR USUARIOS DE UN GRUPO

This screenshot shows the same interface as Figure 19, but with 'cristina (user)' selected in the 'Lista de usuarios:' list. The 'Usuario:' field now contains 'cristina', and the 'Rol:' dropdown is set to 'user'. The 'Info. Grupo' button is red. At the bottom of the panel, there are three buttons: 'Modificar' (blue), 'Cancelar' (blue), and 'Eliminar' (orange).

Fuente: Captura de pantalla del programa TRACE (2020).

diseño en forma de modelo Entidad-Relación es el que se muestra la Figura 21, en la que la tabla de mayor tamaño, “10_obras”, es en la que se guarda registro de las diferentes obras del catálogo TRACE completo (incluidos todos los subcatálogos), todos esos campos se corresponden con los observados en el formulario de entrada de datos para nuevas obras, en las Figuras 8, 10 y 11.

FIGURA 21 - BASE DE DATOS DE LA APLICACIÓN TRACE (MODELO ENTIDAD-RELACIÓN)



Fuente: Creación propia de los autores (2020).

Adicionalmente, las tablas entre la “16_títuloNormalizado” y la “29_provincia”, son las que contienen, cada una de ellas, la lista de ítems de los cuadros desplegables del formulario de entrada de registros, la gestión del contenido de dichas tablas se realiza a través de la opción “Valores Desp.,” disponible para todos los administradores de grupos (ver Figura 14).

Las tablas “13_nombreNormalizado” y “14_rol” contienen nombres de personas que son (o pueden ser) autores, directores, intérpretes, etc., la primera, y las denominaciones de posibles papeles que se pueden desempeñar con respecto a una obra (autor, director, escritor, editor, productor, etc.), la segunda. La razón por la que se ha separado en dos tablas esta información es porque, aunque no es frecuente, sí puede ocurrir que una misma persona puede desempeñar varios roles diferentes, incluso con respecto a la misma obra (por ejemplo: productores de cine que también dirigen sus películas y, a veces, también son los guionistas, o escritores que son también los editores de sus propias obras o los traductores de las mismas a otro idioma). La tabla “15_responsabilidades” permite almacenar, para una obra determinada, una lista de pares {nombre, rol}, para identificar los diferentes responsables y roles desempeñados en el desarrollo de una obra concreta, incluso, en el caso de intérpretes, es posible registrar el nombre del personaje interpretado en la obra (o personajes si fueran más de uno). Toda esta información se registra a través de la pestaña “Responsabilidades” del formulario de edición de obras (véase la Figura 12).

La tabla “12_documentos”, es en la que se almacenan los archivos relacionados con la obra, dichos archivos pueden ser documentos que se suben al servidor y que posteriormente, en una operación de consulta, podrían descargarse para su estudio, o también pueden ser direcciones de Internet (URLs) a páginas o recursos en línea relacionados con la obra. La Figura 13 muestra la pestaña “Archivos”, donde se edita esta parte de la información relacionada con la obra.

Las tablas “01_usuarios”, “02_grupos” y “03_pertenencia” son las que contienen los datos de los usuarios registrados en la aplicación, los grupos o subcatálogos existentes, y a que grupo pertenece cada usuario y que rol desempeña en dicho grupo. Estas tres tablas permiten definir la forma de trabajo de los equipos encargados de cada subcatálogo TRACE, su contenido es editado a través de los formularios de las Figuras 15 a 20.

Adicionalmente, la tabla “01_usuarios” está relacionada con la tabla “10_obras” a través de la tabla “11_logs”, en esta última se registran todas las operaciones de actualización que se hacen sobre una obra y quién las ha hecho, es una forma de poder auditar quién ha trabajado en que obra y cuáles son los cambios que ha incorporado, esta información se guarda automáticamente con cada nueva actualización y está disponible, sólo en modo de consulta, en la pestaña “Logs” del formulario de edición de obras.

Por último, los responsables del grupo de investigación TRACE, a quien va destinada esta aplicación, expresaron su deseo de que en un futuro se pudiera ampliar para poder estudiar qué tipo de consultas se hacen sobre el catálogo y cuáles eran las más habituales. Actualmente no se ha desarrollado ninguna funcionalidad que permita mostrar este tipo de información a través del interfaz gráfico de la aplicación, pero en previsión de que en un futuro se pudiera desarrollar un módulo de tales características, la aplicación actual, de forma automática, ya almacena en la tabla “50_búsquedas” todas las búsquedas, tanto simples como avanzadas, que se realizan sobre el catálogo. Estos datos sobre consultas estarán disponibles si en un futuro se decidiera realizar la citada ampliación.

CONCLUSIONES

Digna de destacar es la experiencia adquirida en reducir la brecha entre el desarrollo de una biblioteca/archivo digital convencional, y una biblioteca con requisitos especializados de investigación en humanidades.

Uno de los aspectos más interesantes de este desarrollo fue la realimentación cruzada entre los humanistas y los ingenieros de software. En muchos casos hemos tenido que elegir entre lo que los diseñadores de software consideran un diseño de base de datos óptimo, y lo que los humanistas necesitan para poder representar su base de conocimientos.

La participación de estudiantes del Grado en Ingeniería Informática fue también una buena idea, aunque el proyecto se retrasó un poco por ello. Fue una interesante experiencia de fusión de la investigación y la docencia, donde varios grupos de prácticas de gestión de proyectos compitieron en un entorno creativo

para construir los mejores prototipos que han permitido poner a prueba diferentes aspectos del problema.

También cabe mencionar, como un valor añadido de la implementación desarrollada, las posibilidades tan potentes de trabajo colaborativo que proporciona el hecho de disponer de una gestión de usuarios, grupos y roles tan versátil como la implementada para esta aplicación. Esta funcionalidad, sumado al hecho de que se trata de una aplicación web, es decir, ésta va a estar disponible a través de Internet, cualquier persona autorizada, podría colaborar en el proyecto desde diversas universidades o centros de investigación ampliando el catálogo y/o revisando las entradas actuales; también sería posible que investigadores, internos y externos, puedan consultar el catálogo para sus propias investigaciones. En ciertas asignaturas relacionadas con el tema del catálogo TRACE, se podría dar acceso a alumnos para hacer algunas prácticas, o incluso abre la opción de poder realizar Trabajos Fin de Grado o Fin de Master dando a un alumno acceso a la aplicación con el rol que resulte más conveniente según el tipo de tareas que deba realizar. Como puede verse las posibilidades son muchas, en todo caso, todo dependerá del tipo de uso que quieran darle los responsables del proyecto TRACE a esta aplicación.

REFERENCIAS

BIA, A. Down to TEI: Use of Extended Markdown to Speed-up the Creation of TEI Documents. In: TEI CONFERENCE AND MEMBER'S MEETING, 15., 2015, Lyon, France, 2015. France: TEI, 2015.

KRASNER, G. E.; POPE, S.T. A Cookbook for Using the Model-View-Controller User Interface Paradigm in Smalltalk-80. *Journal of Object-Oriented Programming*, v. 1, n. 3, p 26-49,1988.

LIANG, T.; WONG, F.; CHAO, S. Word Alignment Using GIZA++ on Windows. *Proceedings of Machine Translation Summit XIII*, p. 369-372, 2011.

MONOSTORI, K.; ZASLAVSKY, A.; BIA, A. Using the MatchDetectReveal System for Comparative Analysis of Texts. *Proceedings of the Sixth Australasian Document Computing Symposium*, Pacific Bay Resort, Coffs Harbour, p. 51-58, 2001.

ROUSE, M. *What is LAMP?* 2008. Disponível em: <<http://searchenterpriselinix.techtargert.com/definition/LAMP>>. Acesso em: 22 jun. 2020.

SAVELLI, R. Bitext2tmx: free text aligner and TMX exporter. 2008. Disponível em: <<http://blog.albatrossolutions.com/alignment/bitext2tmx-free-text-aligner-and-tmx-exporter>>. Acesso em: 22 jun. 2020.

SCHREIBMAN, S.; HANLON, A.; DAUGHERTY, S.; ROSS, T. The Versioning Machine v3.1: A Tool for Displaying and Comparing Different Versions of Literary Texts. Digital Humanities (2007). 2007. Disponível em: <<https://trace.unileon.es/en/introduction/goals/>>. Acesso em: 22 jun. 2020.



WebSinC: buscas *on-line* em *corpora* sintaticamente anotados

Aline Silva Costa
Cristiane Namiuti

INTRODUÇÃO

Buscas automáticas por categorias sintáticas ou morfossintáticas em textos de *corpora* anotados têm sido utilizadas como metodologia de obtenção de dados para várias pesquisas na área da Linguística. A utilização de softwares que realizem tais buscas é fundamental, uma vez que permitem a análise de grandes *corpora* com grande volume de dados textuais. Apresentamos neste trabalho¹ o software WebSinC², um aplicativo web que fornece o recurso de buscas automáticas com interface gráfica. A ferramenta também provê o recurso de gerenciamento de *corpora* anotados, disponibilizando os textos e as imagens dos manuscritos na Internet, para fins de pesquisa.

O software WebSinC foi criado para gerenciar e disponibilizar os textos do *corpus* digital Documentos Oitocentistas de Vitória da Conquista e região

¹ Este trabalho apresenta resultados de pesquisa relativos aos projetos: Fapesp 2012/06078-9, Fapesb PET0034/2010, CNPq 471753/2014-9, CNPq 485098/2013-0. Vincula-se também aos projetos Fapesb APP0007/2016, Fapesb APP0014/2016, CNPq 436209/2018-7.

² Aplicativo Web idealizado pelos professores coordenadores do Laboratório de Pesquisa em Linguística de *Corpus* (LAPELINC/UESB) – Prof.ª. Dr.ª. Cristiane Namiuti e Prof. Dr. Jorge Viana Santos –, e programado por Aline Silva Costa como produto de sua dissertação de mestrado: COSTA, A. S. *WebSinC: uma ferramenta web para buscas sintáticas e morfossintáticas em corpora anotados; Estudo de Caso do Corpus DOViC-Bahia*. 2015. 187 f. Dissertação (Mestrado) - Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2015). Disponível em: <<<http://memoriaconquistense.uesb.br/websinc/>>. Acesso em: 10 mar. 2020.

(DOViC)³ que, formado por documentos manuscritos do século XIX guardados nos arquivos do Fórum de Vitória da Conquista – Bahia, forneceu requisitos para implementação e testes com o software. O WebSinC foi programado para buscas por categorias morfossintáticas e sintáticas em *corpora* anotados nos moldes do *Corpus Histórico do Português Tycho Brahe* (CTB)⁴ com todo o esquema de anotação e buscas no padrão XML (*Extensible Markup Language*)⁵. O CTB é um *corpus* digital composto atualmente de textos em português de autores nascidos entre 1380 e 1845, desenvolvido na Universidade Estadual de Campinas (Unicamp).

Tanto o CTB quanto o *Corpus* DOViC são *corpora* de textos antigos. Faz parte da metodologia de trabalho destes *corpora*, a anotação de edições realizadas nos textos para tratamento computacional, com a preservação de características originais para estudos linguísticos e filológicos. A motivação para o desenvolvimento do software WebSinC consiste no fato de que os recursos desenvolvidos e aplicados para a compilação, anotação e busca de dados no CTB não seguem um padrão único de linguagem. A linguagem XML é utilizada para a anotação de edição e para a anotação morfossintática. Já a anotação sintática segue o formato *Penn TreeBank*⁶, um outro formato que implica duplicação do texto, perda de informação e uso de outra linguagem para as buscas morfossintáticas. Com o WebSinC, as buscas morfossintáticas podem explorar as potencialidades das anotações de edições no formato XML, possibilitando implementação do retorno de resultados na versão

³ *Corpus* compilado no âmbito do projeto *Memória conquistense: implementação de um corpus digital* (NAMIUTI; SANTOS, 2013), que dá continuidade ao trabalho iniciado no projeto *Memória Conquistense: recuperação de documentos oitocentistas na implementação de um corpus digital* (SANTOS; NAMIUTI, 2009) e que atualmente sua contínua construção está sob a vigência do projeto *Corpora digitais de documentos históricos da imperial Vila da Victoria, atual Vitória da Conquista - Bahia: resgate e preservação do patrimônio linguístico e da memória da escravidão na Bahia* (SANTOS, NAMIUTI, 2016).

⁴ GALVES, C.; ANDRADE, A.; FARIA, P. *Corpus Histórico Anotado do Português Tycho Brahe Parsed Corpus of Historical Portuguese*. Campinas: UNICAMP, 2017. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>>. Acesso em: 10 mar. 2020.

⁵ Segundo Silva Filho (2004), XML é uma linguagem de editoração que oferece um formato universal para estruturação de documentos e dados na Web. Ela foi proposta pelo World Wide Web Consortium (W3C) como uma nova alternativa à linguagem HTML, linguagem dominante na Web; a XML combina extensibilidade, poder e flexibilidade com a simplicidade exigida pela Web (SILVA FILHO, 2004).

⁶ De acordo com Santorini (2010), o *Penn TreeBank Format* (Formato *Penn TreeBank*) é um esquema de anotação sintática de *corpora* desenvolvido na Universidade da Pensilvânia. O esquema utiliza uma representação arbórea delimitada por parênteses etiquetados.

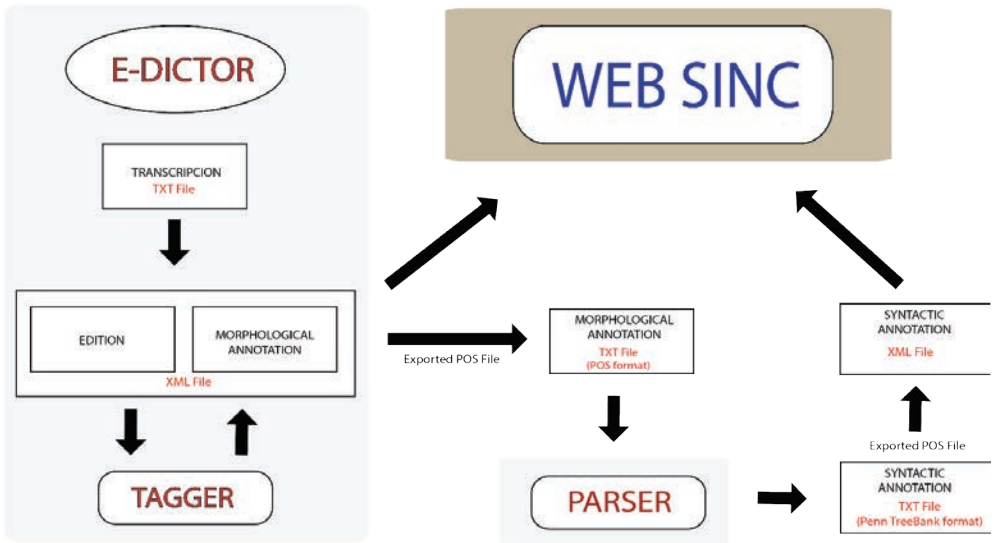
Atualmente, não há uma ferramenta com a funcionalidade de recuperação das informações de edições anotadas no arquivo XML de textos do CTB. As buscas automáticas baseadas em categorias morfossintáticas são realizadas em um arquivo de texto com anotação *Part-Of-Speech* (POS) realizada pelo etiquetador POS. Assim, as anotações de edições com intervenções dos editores e as versões originais dos textos contidas no arquivo XML gerado pelo eDictor, não podem ser exploradas nas buscas. Os resultados com as ferramentas utilizadas pelo CTB trazem apenas sentenças em versões editadas. A ferramenta WebSinC foi programada para realizar as buscas morfossintáticas no arquivo XML, podendo explorar as potencialidades das anotações com retorno de sentenças de resultado em versão original ou editada, minimizando a perda de informação.

A versão atual do programa eDictor (versão 1.0 beta 10) não realiza anotação da estrutura sintática. Tal informação é gerada separadamente utilizando um *parser* que recebe como entrada um arquivo anotado no texto (.txt) no formato POS, com as etiquetas morfossintáticas, e gera como saída um outro arquivo texto no formato *Penn TreeBank*. Os textos recebem anotação morfossintática (categoria POS) e sintática dentro dos moldes propostos pelo *Penn-Helsinki Parsed Corpus of Middle English* (PPCME), cuja proposta sugere que a etiquetagem POS deve preceder o processo de anotação sintática (GALVES; BRITO, 2008).

O processo de compilação de corpora realizado para a construção do CTB apresenta, portanto, heterogeneidade de linguagem. Entretanto, na busca por obter homogeneidade de linguagem para reuso de tecnologia, o WebSinC converte a anotação *Penn TreeBank* para o formato XML, reutilizando assim a mesma tecnologia para todos os tipos de buscas. Além do ganho com reuso de tecnologia, é possível obter mais flexibilidade para as buscas e exibição dos resultados, bem como obter independência tecnológica, uma vez que ferramentas de busca para o formato *Penn TreeBank* tornam-se dispensáveis (NAMIUTI; COSTA, 2014).

A Figura 1 mostra o resumo dos processos realizados com os textos do *Corpus DOViC*. Toda a parte à esquerda é idêntica aos processos do CTB. Os retângulos brancos representam os processos realizados e o texto em vermelho no interior dos retângulos representa o formato da saída gerada. As ferramentas computacionais utilizadas nos processos estão indicadas (i. eDictor; ii. Tagger incorporado ao eDictor; iii. Parser; iv. WebSinC). As setas indicam a interação e direção entre os processos.

FIGURA 1 - RESUMO DOS PROCESSOS PREVISTOS PARA A COMPILAÇÃO DO *CORPUS* DOVIC



Fonte: Santos, Namiuti e Costa (2015a).

O SOFTWARE WEBSINC

De acordo com Namiuti e Santos (2015), o software WebSinc provê uma interface gráfica de fácil utilização pelo usuário, que pode ser um pesquisador interessado no *corpus*, ou um administrador do sistema que o gereencie, fazendo *upload* e cadastro de documentos digitais imagem (DDI) e documentos digitais texto (DDT), tornando o *corpus* disponível na Internet. Ao pesquisador da Internet estarão disponíveis os documentos do *corpus* para visualização do documento original em imagem digital (fotos do original físico) e também das versões do texto original (transcrita, editada e morfossintaticamente anotada). A ferramenta provê o recurso de buscas baseadas em categorias sintáticas ou morfossintáticas para auxiliar pesquisas na área da Linguística, sem que o usuário aprenda qualquer linguagem de consulta, pois as buscas poderão ser feitas graficamente através de componentes Graphic User Interface (GUI) como links, botões e caixas de seleção.

A Figura 2 mostra a tela de apresentação do software na qual são exibidas informações gerais e os campos para acesso com *login* e senha. O usuário deve ser registrado para ter acesso ao *corpus*.

FIGURA 2 - TELA DE LOGIN DO WEBSINC

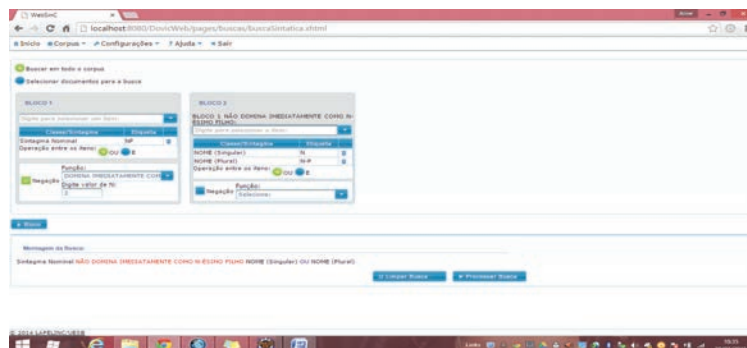


Fonte: Namiuti, Santos e Costa (2015b).

BUSCAS SINTÁTICAS NO WEBSINC

A Figura 3 mostra a tela do WebSinC para o recurso de buscas sintáticas. As buscas podem ser feitas pela manipulação dos componentes gráficos na interface.

FIGURA 3 - TELA DE BUSCAS SINTÁTICAS DO WEBSINC



Fonte: Namiuti, Santos e Costa (2015b).



Não é necessário que o usuário conheça o conjunto de etiquetas da anotação do *corpus* pesquisado. A busca é construída através da montagem de um conjunto ou *container* de itens, denominado na ferramenta como um “bloco” com categorias de itens lexicais e/ou sintagmas. Os itens do bloco podem ser relacionados pela operação lógica “OU” ou “E”. A opção “OU” vem pré-selecionada por *default*. O usuário pode selecionar um item lexical/sintagma usando o recurso de pesquisa do componente (recurso do tipo *autocomplete*), que funciona de maneira que, a cada caracter digitado, um filtro é aplicado retornando apenas itens que se iniciem com os caracteres digitados. Após a seleção do item, ele é inserido automaticamente no bloco.

Feita a montagem do bloco, o usuário deve selecionar uma função de busca. Há funções que requerem apenas um argumento, como é o caso da função “Existência”. As outras funções requerem dois ou mais argumentos e, ao selecioná-las, os componentes da interface são exibidos dinamicamente para seleção. As funções de busca sintática implementadas no WebSinC são: 1) Existência; 2) Precedência; 3) Precedência imediata; 4) Dominância; 5) Dominância imediata; 6) Irmandade; 7) C-Comando; 8) Dominância de N palavras; 9) Dominância de mais de N palavras; 10) Dominância de menos de N palavras; 11) Dominância imediata como primeiro filho; 12) Dominância imediata como último filho; 13) Dominância imediata como N-ésimo filho; 14) Dominância imediata como único filho; 15) Dominância imediata de N filhos; 16) Dominância imediata de menos de N filhos; 17) Dominância imediata de mais de N filhos.

Para realização das buscas sintáticas no WebSinC, é necessária a transformação do arquivo com anotação sintática gerado pelo *parser* em um arquivo XML com representação da estrutura sintagmática. Tal conversão é realizada pelo WebSinC. Sendo assim, o arquivo XML para buscas sintáticas se trata de um outro arquivo diferente e não relacionado ao arquivo com anotações morfossintáticas e de edições do texto gerado pelo eDictor. Isto porque o *parser* utilizado atualmente não pode ser aplicado a um arquivo XML, mas unicamente ao formato POS.

Ao processar a consulta, o resultado da busca é exibido numa tabela, com cada sentença de resultado numa linha. Na tabela também é exibido um atalho para visualização de sentença no formato gráfico de árvore, além de informações sobre a busca: texto utilizado para a busca, a consulta realizada e o total de sentenças encontradas. A Figura 4 mostra como exemplo a tela de resultado para uma busca

XML. A aplicação usou um Sistema Gerenciador de Banco de dados (SGBD) livre com suporte a armazenamento XML, o PostgreSQL⁹.

CONCLUSÃO

O desenvolvimento e utilização do software WebSinC com o *Corpus* DOViC trouxeram como resultados o recurso de gerenciamento e a disponibilização do *corpus* na Internet, permitindo também a realização de buscas automáticas para fins de pesquisa. Ressaltamos que o uso de XML para anotação sintática tem a vantagem de reutilizar a mesma tecnologia já utilizada para anotações morfossintática e de edições no *Corpus* DOViC. Como XML é um padrão, usá-lo para todas as representações nos textos do *corpus* favorece a criação de recursos padronizados, permitindo reuso de tecnologia, oferecendo mais flexibilidade para as buscas e exibição dos resultados, e independência tecnológica para grupos de pesquisa interessados em estudos a partir desse *corpus*.

REFERÊNCIAS

GALVES, C.; BRITO, H. *A construção do corpus anotado do Português Histórico Tycho Brahe: o sistema de anotação morfológica*. 2008. Disponível em: <http://www.tycho.iel.unicamp.br/~tycho/pesquisa/artigos/GALVES_Cetal-Fase1a.pdf>. Acesso em: 05 ago. 2014.

GALVES, C.; ANDRADE, A. L. de; FARIA, P. Tycho Brahe Parsed Corpus of Historical Portuguese. 2017. Disponível em: <<http://www.tycho.iel.unicamp.br/~tycho/corpus>>. Acesso em: 13 mar. 2020.

NAMIUTI, C.; COSTA, A. S. Reflexão sobre anotação sintática e ferramentas de busca: uso da linguagem XML para anotação sintática no *corpus* digital DOViC. *Letras & Letras*, v. 30, n. 2, p. 82-103, 2014. Acesso em: 15 set. 2015.

NAMIUTI, C.; SANTOS, J. V. New Challenges for Ancient Sources: DOViC experience in the New Historical Linguistics. CONGRESSO DE HUMANIDADES DIGITAIS EM PORTUGAL: CONSTRUINDO PONTES E QUEBRANDO BARREIRAS NA ERA DIGITAL, 2015, Lisboa, Portugal. Anais... Lisboa: Universidade Nova de Lisboa.

⁹ O PostgreSQL é um sistema robusto, confiável e largamente utilizado como SGBD de sistemas empresariais e baseados na Web por todo o mundo. (PEREIRA NETO, 2003).

NAMIUTI, C.; SANTOS, J. V.; COSTA, A. S. New challenges for ancient sources: an important dialogue between Computer Science and new Historical Linguistics. Trabalho apresentado no Workshop: The New Historical Linguistics and the World of Annotated Corpora, Campinas, Brasil, Março 9-13, 2015a.

NAMIUTI, C.; SANTOS, J. V.; COSTA, A. S. *WebSinC*. versão 1.0. Vitória da Conquista: UESB, 25 abr. 2015b. Disponível em: <<https://memoriaconquistense.uesb.br/websinc>>. Acesso em: 15 mar. 2020.

PAIXÃO DE SOUSA, M. C. Memórias do Texto. *Revista Texto Digital*, v. 2, n. 1, 2006. Disponível em: <<http://www.textodigital.ufsc.br/num02/paixao.htm>>. Acesso em: 05 ago. 2014.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. FARIA. P. eDictor: novas perspectivas na codificação e edição de corpora de textos históricos. In: SHEPHERD, T.; SARDINHA, T. B.; PINTO, M. V. *Caminhos da Linguística de Corpus*. Campinas: Mercado de Letras, 2010.

PEREIRA NETO, A. *PostgreSQL: técnicas avançadas; versões open source; soluções para desenvolvedores e administradores de banco de dados*. São Paulo: Editora Érica, 2003.

SANTORINI, B. Annotation Manual for the Penn Historical Corpora and the PCEEC. 2010. Disponível em: <<http://www.ling.upenn.edu/hist-corpora/annotation/index.html>>. Acesso em: 8 out. 2013.

SILVA FILHO, A. M. *Programando com XML*. Rio de Janeiro: Elsevier, 2004.

Esta forma de ver a espacialidade dos fenómenos e processos, tanto em termos históricos como sociais, foi grandemente fomentada pela publicação de *La production de l'espace*³ de Henri Lefebvre. Este autor introduziu a ideia de que o espaço não é simplesmente uma geografia natural nem um contentor vazio que foi sendo cheio pela História, sendo antes algo que a humanidade foi produzindo ao longo do tempo, já que as relações no e com o espaço se alteram. O espaço é por si só histórico e reflete as mudanças e evoluções das comunidades humanas. Esta atenção dedicada à questão espacial, nomeadamente à forma como as modificações dos espaços e das relações entre eles explicam e implicam alterações no tempo, não era então um movimento completamente novo. Já com a escola dos Annales, em autores como Marc Bloch e, fundamentalmente, Fernand Braudel – que concebe o tempo histórico como decorrendo em três níveis, o da “longa duração”, o da “conjuntura” e o dos “eventos” – surge como inconcebível uma “história total”, se desligada do espaço onde decorrem os processos.

A ideia do território como uma construção social complexa que afeta as práticas e perceções humanas e a sua importância e influência na reprodução social conduziram, em grande medida, ao interesse de alguns historiadores em utilizar tecnologias e metodologias desenvolvidas para outras áreas do conhecimento científico e aplicá-las nas questões da investigação histórica⁴. Nota-se, portanto, um crescente interesse em “localizar” o passado, em apor a evolução temporal à variação espacial – movimento conhecido por *spatial turn* – revelando a intenção em entender o papel do espaço no desenrolar da História humana (BODERNHAMER; CORRIGAN; HARRIS, 2010). Procuram-se, assim, novas perspetivas no examinar/reexaminar de relações, padrões e contextos que emergem quando a História das comunidades humanas é deslindada através de uma lente “espacial”. O âmbito da análise espacial dos fenómenos históricos situa-se, no fundo, no exame de atributos e relações entre os dados, levando em conta a localização de um acontecimento ou de um processo em estudo. Esta pode ser feita de forma simples, através da observação do fenómeno e de sua distribuição

³ A primeira edição desta obra é datada de 1974.

⁴ Nomeadamente os Sistemas de Informação Geográficos e as metodologias relacionadas com estatística espacial, de que são exemplo os seguintes estudos: Bodenhamer, Corrigan e Harris (2010); Frank (2007); Gregory e Ell (2007); Healey e Stamp (2000); Knowles (2000, 2002); Schwartz (2011); Silveira et al. (2011).

no espaço, ou através de uma análise mais elaborada, que considere a interação de vários fenómenos para explicar uma determinada situação no espaço geográfico.

A espacialidade dos processos e fenómenos históricos é elemento indispensável ao entendimento das relações sociais e humanas e toda a ação humana acontece num determinado espaço-tempo, que se materializa num “lugar” da História (LEFEBVRE, 2007). Estas conceções impõem, então, que a variável espacial assuma um papel fundamental na análise de como era e se organizava a concentração fabril lisboeta na última década do século XIX.

Nesta época operaram-se grandes mudanças económicas, sociais e urbanísticas na cidade de Lisboa, impulsionadas parcialmente pelo desenvolvimento industrial. No entanto, estas não se processaram de forma instantânea nem tão pouco homogénea no território urbano da capital portuguesa. Neste texto procuro analisar a implantação industrial da capital portuguesa na última década de Oitocentos, a partir da construção de uma cartografia digital pormenorizada. Sendo, portanto, identificada a localização dos estabelecimentos industriais e oficinais, as diferentes formas de habitação operária e as sedes das respetivas associações, granjeando as diferenças na espacialidade da vida operária em distintos contextos desta Lisboa em mutação.

Como se caracterizava e distribuía a indústria na cidade de Lisboa? Como se caracterizava globalmente o operariado e se distribuía pelos estabelecimentos industriais? Em que zonas da cidade estavam os locais de trabalho, de habitação e de associação destas pessoas? O que qualificava estes três eixos da vida operária?

Privilegiando a perspetiva espacial, utilizaram-se fontes que permitem não só uma análise de dados quantitativos e qualitativos relativos à vivência operária, mas também possibilitam a localização geográfica dessas mesmas vivências. É, portanto, essencial e fundamental para a realização desta investigação a frutífera produção de estatísticas, inquéritos, relatórios e estudos neste breve período que se debruçaram sobre a situação social e industrial do país e, mais pormenorizadamente, da sua capital e dos seus habitantes pertencentes às “classes laboriosas”. Usei também, fontes primárias, incluídas em fundos de arquivos documentais, como as atas de fundação de diferentes tipos de associações de cariz operário e os processos de obras e de arruamentos de bairros e vilas operárias.⁵

⁵ Ver os Fundos arquivísticos e, na lista de referências, as publicações oficiais e os relatórios e estatísticas consultados no âmbito desta investigação.

A preponderância que pretendo dar à perspetiva espacial ditou a necessidade da criação de um Sistema de Informação Geográfico (SIG) que inclui toda a informação depois de georreferenciada. Isto significa que, sempre que possível, foi atribuída uma localização precisa, nomeadamente ao nível da rua e número da porta, a cada estabelecimento industrial, núcleo de habitação e associação operária. Deste modo, foi possível relacionar diferentes tipos e camadas de informação – qualitativas ou quantitativas; geográficas, económicas ou históricas – referentes a fábricas, associações ou aglomerados de habitação operária.

A aplicação desta ferramenta digital ao trabalho do historiador, tanto na vertente de investigação como na de divulgação, é uma prática que começou a surgir no final da década de 90 do século XX, sendo internacionalmente reconhecida como *Historical GIS* (HGIS). O emprego dos SIG na indagação do passado consiste, grosso modo, na construção de bases de dados que integrem informação cartográfica georreferenciada, sobre a localização de determinadas entidades com informação quantitativa e/ou qualitativa associada. Os vários níveis de dados, resultado da compilação crítica e fundamentada de informação histórica, e a cartografia digital daí resultante poderão ser comparados, sobrepostos, “questionados” de forma a serem exploradas alterações que ocorreram tanto a nível espacial como temporal. Tal como apontado por Knowles (2000), em seu artigo considerado fundador da conceptualização dos HGIS, a utilização das ferramentas e metodologias SIG é mais frutífera em investigações históricas que, para além de se usarem dados passíveis de serem representados através de pontos, linhas ou áreas, se pretende analisar e representar grandes volumes de dados. E, assim, testar hipóteses que relacionem entidades representáveis no espaço e/ou em pesquisas inerentes às alterações de padrões espaciais.

De facto, a importância e a mais-valia da utilização de um SIG numa investigação como esta que aqui apresentamos advém do facto de ser uma infraestrutura capaz de cruzar e gerir uma enorme quantidade e diversidade de dados. O grande desafio consiste em ir além das suas capacidades de representação cartográfica e de o explorar enquanto “ferramenta de gestão [e análise] de informação” (ELL, 2010, p. 148, tradução nossa)⁶, geradora de novas

⁶ Língua original: Inglês.

hipóteses e explicações históricas. Procurando, deste modo, “chegar mais perto da complexidade das alterações e da realidade histórica” (SCHWARTZ; GREGORY; MARTI-HENNEBERG, 2011, p. 252)⁷ lisboeta da década final do século XIX e “apontar uma ligação contingente entre alterações na estrutura da sociedade” (MOORE, 2010, p. 40).

A construção de um SIG permite associar, relacionar e explorar espacialmente múltipla informação relativa às vertentes do quotidiano operário analisadas – laboral, habitacional e de associação – proveniente das várias fontes primárias e secundárias assim como de diferentes fundos arquivísticos⁸. A georreferenciação destes dados possibilita a criação de cartografia digital detalhada da localização e caracterização das fábricas, das aglomerações habitacionais dos operários e dos seus espaços de associação e intervenção política. Permite, deste modo, deslindar e estudar a implantação industrial e operária no espaço urbano da capital deste período, possibilitando novas perspetivas de conhecimento deste processo histórico ao relacioná-lo com a sua localização. Por outro lado, das sobreposições e hiatos territoriais destes aspetos do quotidiano surge a clarificação do papel e da importância da indústria e do operariado na Lisboa do final do século XIX.

LOCAIS DE TRABALHO – Fábricas e oficinas

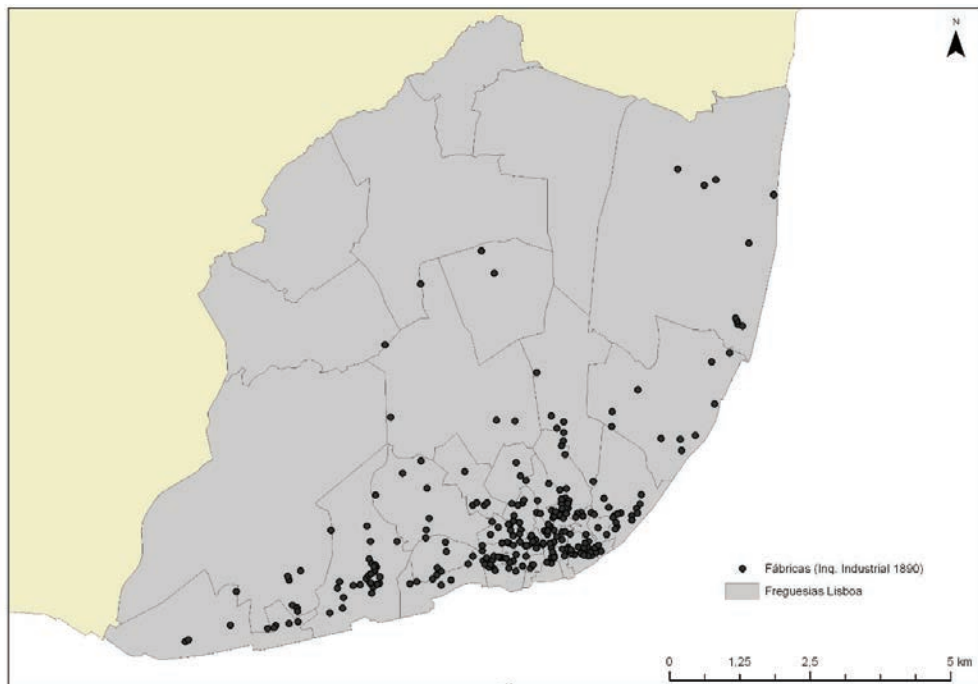
A análise da indústria e dos operários, particularmente da implantação e caracterização dos estabelecimentos industriais e do trabalho operário que neles se operava, apoia-se fundamentalmente na informação contida no *Inquérito Industrial de 1890*. As informações sobre as fábricas de Lisboa, com mais de cinco operários, são de tal forma detalhadas que permitem a localização dos estabelecimentos industriais ao nível da rua, mas também obter informação

⁷ Língua original: Inglês.

⁸ Como por exemplo: Fundo das Associações de Classe (Arquivo do Gabinete de Estratégia e Planeamento do Ministério da Solidariedade e Segurança Social); Cadastro de Associações Socioprofissionais e Mútuas; Estatutos das Associações de Classe e Recreativas (Arquivo do Governo Civil de Lisboa); Registo de licenças para estabelecimentos de comércio e indústria, 1890-1900; Urbanismo e Obras / Planeamento Urbanístico / Projectos de Arruamentos (Arquivo Municipal de Lisboa).

relativa à sua caracterização (como número de trabalhadores, sector de produção, nº de máquinas a vapor e quantidade de energia utilizada), e à dos operários que lá trabalhavam (através do sexo, faixa etária, alfabetização, rendimento obtido, entre outros).

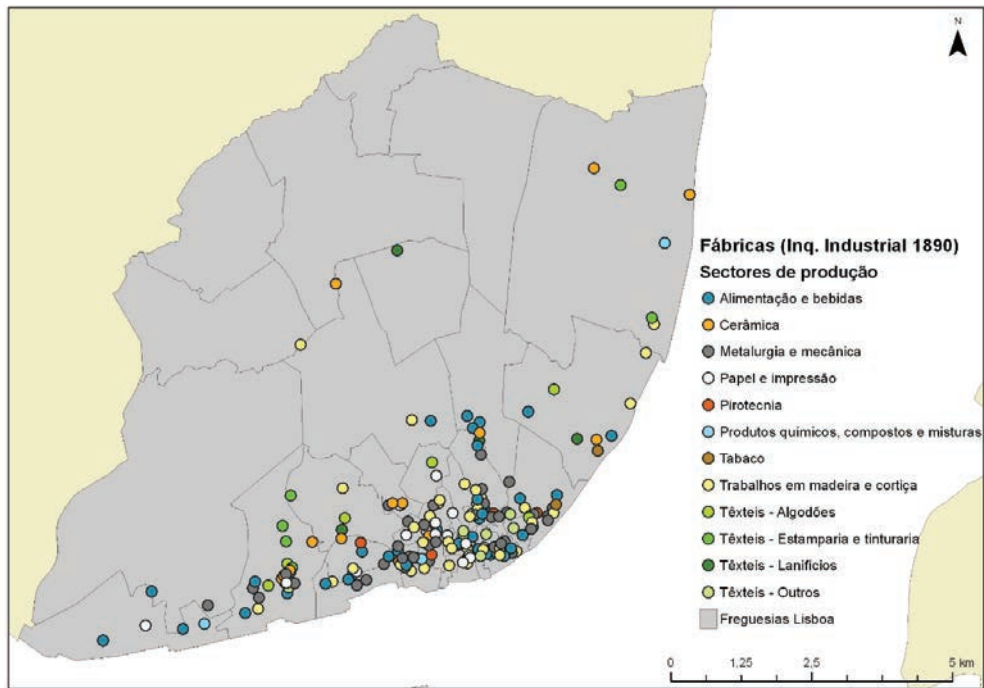
FIGURA 1 - ESTABELECIMENTOS INDUSTRIAS (1890)



Fonte: Elaborada pela autora, 2020.

Da análise espacial da localização das fábricas e oficinas (Figura 1) reconhece-se a grande dimensão do polo fabril de Alcântara onde, pela localização dos estabelecimentos industriais, se “lê” o traçado e a importância da ribeira de Alcântara no desenvolvimento industrial desta zona da cidade e o surgimento do segundo polo industrial, na zona oriental da cidade. Mas o que salta à vista é a grande densidade de ocupação industrial no centro da cidade. O centro tradicional e comercial da cidade era também um centro industrial.

FIGURA 2 - SETORES INDUSTRIAIS (1890)



Fonte: Alcântara, 2016, p. 48.

Observando fábricas e oficinas de Lisboa dos sectores de produção industrial com mais operários na Lisboa desta época (Figura 2),⁹ tendo em conta a dispersão/concentração dos diferentes sectores produtivos, identificam-se diferentes bolsas industriais. Embora a diversidade dos sectores de produção fosse uma constante no território lisboeta, reconhecem-se algumas especificidades e diferenças que denotam três tipos de padrões: o primeiro marcado por uma dispersão pelo tecido urbano; outro que se distingue nos limites urbanos da cidade e, por fim, um padrão que denota uma certa centralidade.

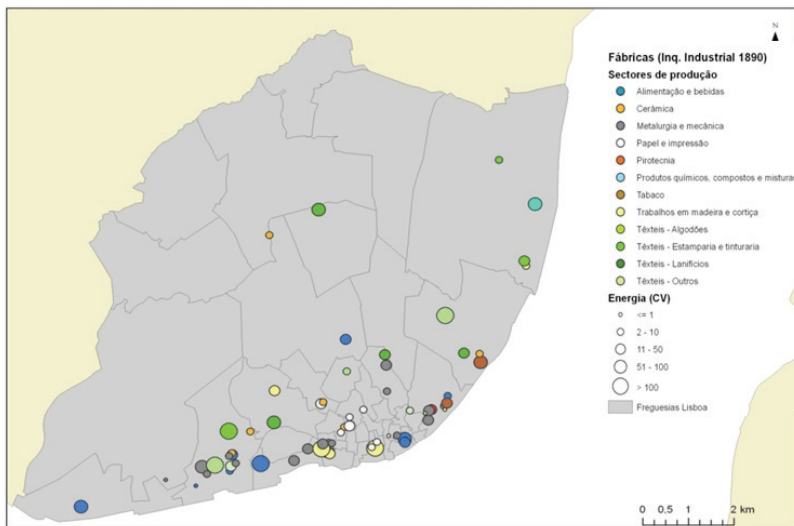
⁹ Foram considerados os nove sectores de produção industrial que empregavam mais de 2% do total de operário/as registados no *Inquérito Industrial de 1890*. Assim, quanto à percentagem de operários a trabalhar em cada um dos sectores de produção, a hierarquia de importância era a seguinte: o dos *Têxteis* empregava 25% dos trabalhadores; o dos *Tabacos* 17%; o dos *Trabalhos em madeiras e cortiça* 15%; o da *Metalurgia e Mecânica* 14%; o da *Alimentação e Bebidas* 7%; o do *Papel e Impressão* 6%; o da *Produção de Químicos* 3%; o da *Pirotecnia* 2,6% e o da *Cerâmica* 2,6%.

Os sectores de produção que estavam relacionados com as necessidades diárias das populações – “Alimentação e Bebidas”, “Trabalhos em Madeira e Cortiça” e “Metalurgia e Mecânica” – estavam naturalmente bastante dispersos pelo território lisboeta. No segundo padrão enunciado, identifica-se na zona ocidental da cidade as grandes fábricas de têxteis de “Algodão” e as “Estamparias e Tinturarias”, em Alcântara, e a produção de “Cerâmica”. E na zona oriente do centro tradicional da cidade encontrava-se o mesmo tipo de setores produtivos, acrescentando-se ainda as fábricas de “*Tabacos*” que se concentravam em Xabregas. Todos estes são tipos de produção que por questões relacionadas com a disponibilidade de espaço, matérias-primas e/ou salubridade tendiam a implantar-se nas zonas periféricas da cidade.

A zona central e a “elegante Baixa” de Lisboa, onde se observa o terceiro tipo de padronização sectorial, era povoada pelas tipografias e fábricas de produção de “Papel”, para além das confeções de vestuário, fabricas de chapéus e de luvas – “Têxteis – outros”.

O *Inquérito Industrial de 1890* constitui também uma fonte importante para a caracterização tecnológica da indústria da cidade de Lisboa neste final de Oitocentos. Permite-nos examinar não só que fábricas utilizavam esta energia na sua produção, mas também as máquinas a vapor que usavam e a quantidade de cavalos-vapor gastos.

FIGURA 3 - UTILIZAÇÃO DA ENERGIA A VAPOR POR SETOR INDUSTRIAL (1890)



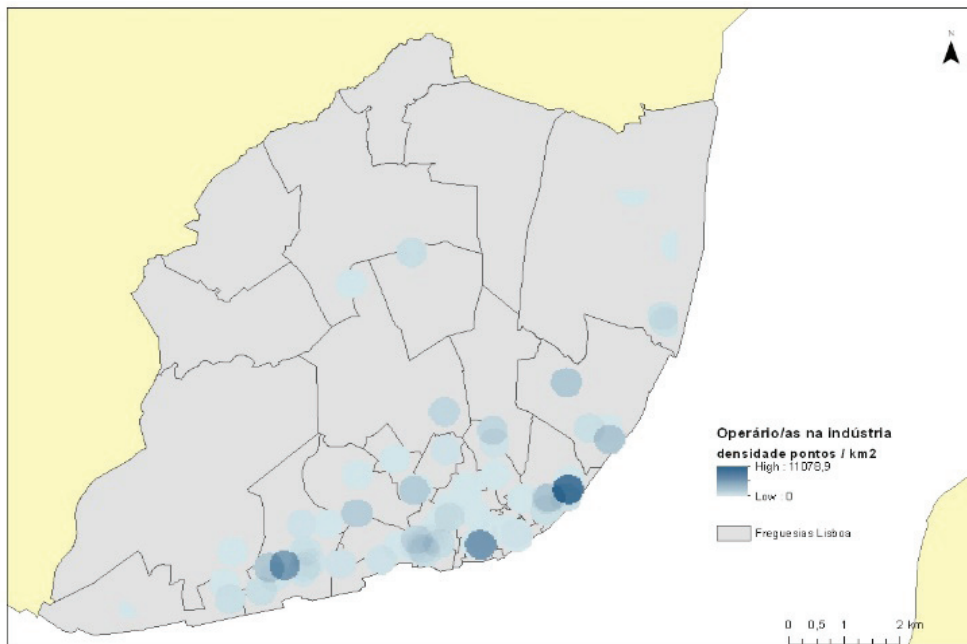
Fonte: Alcântara, 2016, p. 51

A análise espacial da dispersão da energia do vapor (Figura 3) mostra que o seu uso não se circunscrevia aos novos polos industriais ou a grandes fábricas, estando de algum modo difundida por toda a cidade, existindo mesmo nas zonas mais centrais. Os “Têxteis” no seu conjunto eram o setor industrial onde o uso do vapor estava mais difundido. No entanto, muitas fábricas e mesmo pequenas oficinas com outro tipo de produção, como nos “Tabacos”, na “Metalurgia”, na “Alimentação” e no “Papel e Impressão”, também se empregava esta energia. Assim, o uso do vapor não se circunscrevia aos polos industriais de Alcântara e Xabregas ou às grandes fábricas, mas estava de algum modo difundido pela cidade, existindo inclusive na zona histórica e central.

A análise da dimensão das fábricas e oficinas em número de operários permite estudar a dispersão da densidade do trabalho operário (Figura 4)¹⁰ e perceber em que zonas da cidade trabalhavam mais operários e operárias. Embora a densidade das unidades de produção fosse maior no centro (Figura 1), com uma maior concentração de pequenas fábricas e oficinas, os estabelecimentos com mais trabalhadores encontravam-se essencialmente nas novas zonas industriais em afirmação neste período: a zona ocidental e a zona oriental. No entanto, não é de desprezar a concentração de trabalhadores ligados à indústria na zona histórica e mais antiga da cidade.

¹⁰ Para cartografar o trabalho industrial optou-se por calcular a densidade de pontos por quilómetro quadrado, sendo cada ponto valorizado conforme o número total de operários e operárias que trabalhavam em cada fábrica ou oficina referenciada no Inquérito Industrial de 1890. Assim, na Figura 4, as zonas com manchas mais escuras representam áreas onde se concentravam (nos seus locais de trabalho) um maior número de trabalhadores industriais.

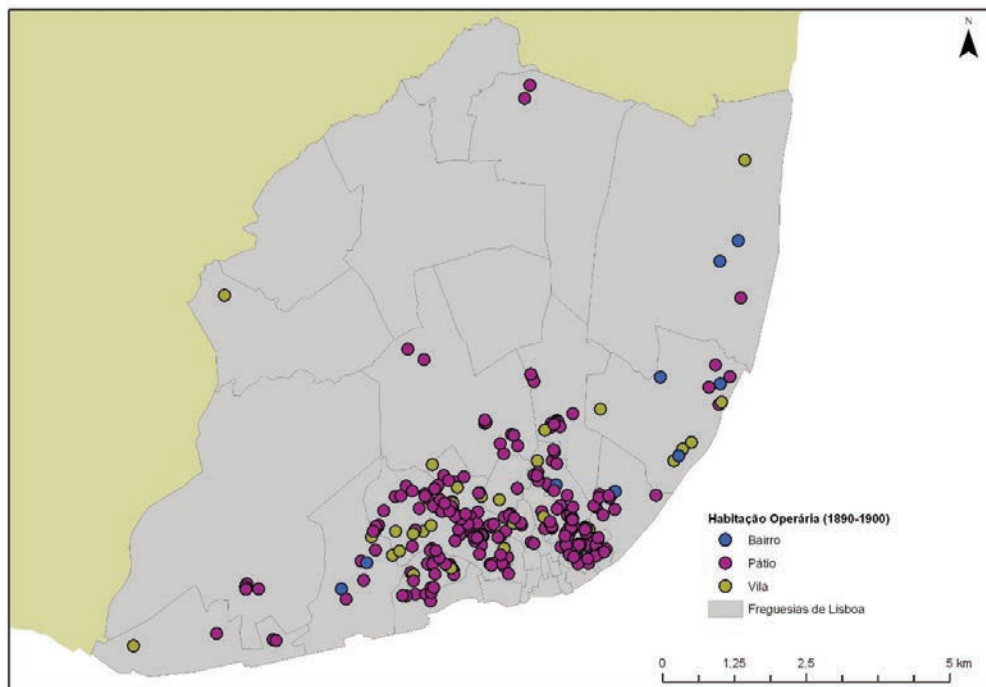
FIGURA 4 - OPERÁRIOS/AS NA INDÚSTRIA (1890)



Fonte: Elaborada pela autora (2020).

LOCAIS DE HABITAÇÃO – PÁTIOS, VILAS E BAIRROS OPERÁRIOS

Sabendo-se do aumento demográfico de Lisboa neste período (RODRIGUES, 1995), importa perceber onde e como habitavam os novos lisboetas. A habitação operária não foi enquadrada, a nível estatal, até aos anos 20 do século XX (BARATA, 2010; FERREIRA, 1987), deixando a busca de soluções para privados. Foi dos próprios operários, dos donos de fábricas e de outros proprietários que surgiram iniciativas de construção e/ou adaptação de espaços para habitação dos novos efetivos populacionais que chegavam à capital.

FIGURA 5 - HABITAÇÃO OPERÁRIA (1890-1900)

Fonte: Elaborada pela autora (2020).

Com base em dados provenientes de diversas fontes da época, que permitem a localização e caracterização da habitação operária – *Inquérito aos pátios de Lisboa; Fundo de Urbanismo e Obras e Processos de Obras* do Arquivo Municipal de Lisboa; relatórios produzidos por instâncias estatais (FUSCHINI, 1884; AZEVEDO, 1905) – cujas falhas e imprecisões foram colmatadas pelo recurso a investigações atuais (DIVISÃO DE REABILITAÇÃO URBANA DOS PÁTIOS E VILAS, 1993; PEREIRA, BUARQUE, 1995), foi construída cartografia digital representando as diferentes formas e núcleos de habitação operária cuja ocupação pode ser confirmada na década de 1890-1900.

A análise espacial da distribuição territorial e a caracterização dos aglomerados de habitação operária e a análise espacial da distribuição destas (Figura 5) revelam uma dispersão muito semelhante à dos estabelecimentos industriais (Figura 1), indicando que o operariado viveria perto dos locais de trabalho. A exceção seria

quem trabalhava mesmo no centro tradicional da cidade. Por imperativos que se prendiam com o preço das rendas, essas pessoas viveriam mais afastados dos seus locais de trabalho, ainda assim não longe do velho centro económico e político da cidade. Conclui-se, portanto, que os espaços industriais e habitacionais se interligavam na geografia urbana da capital neste período.

O estudo da localização das diferentes tipologias da habitação operária (Figura 5) identificadas para esta época – bairros, pátios e vilas operárias – permite afirmar que aquelas que eram construídas de raiz com o propósito de servirem comunidades operárias eram em menor número e estavam, essencialmente, nas novas zonas de crescimento industrial. Este é o caso dos bairros operários e das vilas que, embora construídas também por iniciativa privada de alguns proprietários industriais, por terem menores dimensões, existiam em maior quantidade e estavam mais dispersas pelo tecido urbano.

No entanto, a tipologia de habitação operária que dominava o espaço urbano operário eram os pátios. Estes resultavam, na sua maioria, de construções mais ou menos precárias edificadas em espaços livres nas traseiras de prédios e quarteirões ou do reaproveitamento de espaços e/ou construções anteriores.

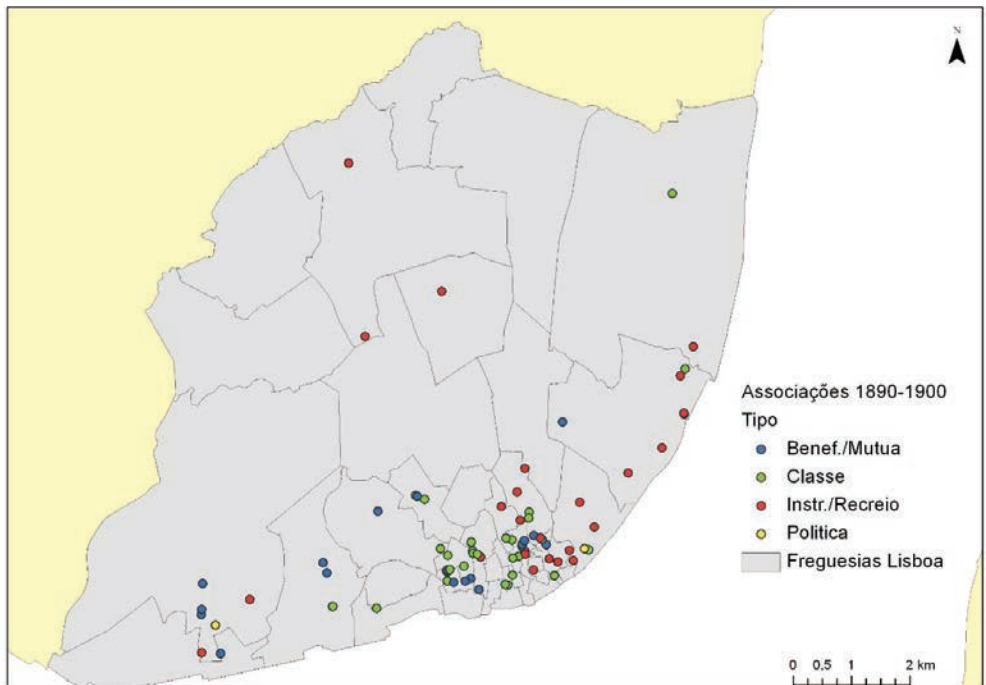
Para entender onde realmente viviam as pessoas é necessário analisar a concentração de habitantes em cada polo de habitação operária. A densidade de moradores por aglomerado habitacional (Figura 6)¹¹ intensifica a leitura dada pela dispersão da habitação per si abordada anteriormente (Figura 5). Ou seja, os aglomerados habitacionais com mais moradores encontravam-se em redor do centro histórico e não nas novas áreas industriais. Existindo uma clara intensificação da densidade de moradores nas zonas de Alfama e Sapadores – a oriente do centro – e da Rua de S. Bento e do Rato/Amoreiras – a norte/ocidente –, onde viveriam a maior parte das pessoas da “classe laboriosa”.

¹¹ Para cartografar a quantidade de moradores em habitações operárias optou-se por calcular a densidade de pontos por quilómetro quadrado, sendo cada ponto valorizado conforme o número total de moradores que viviam em cada núcleo de habitação operária incluída na base-de-dados por nós construída (com base em diversas fontes já referidas). Assim, na Figura 6, as zonas com manchas mais escuras representam áreas onde se concentravam um maior número de habitantes registados em aglomerados habitacionais operários.

as mutualistas e, posteriormente, com a legalização das Associações de Classe em 1891, as ligadas ao trabalho.

Os dados relativos a associações de classe, grêmios ou coletividades, com atividade entre 1890 e 1900, cujos sócios fundadores fossem operários e/ou os seus objetivos estatutários estivessem ligadas à vivência social, cultural e/ou política da população operária permitiram analisar esta temática. Informações estas que estavam espalhadas por diversas fontes: nos fundos Cadastro de Associações Socioprofissionais e Mútuas e Estatutos das Associações de Classe e Recreativas do Arquivo do Governo Civil de Lisboa; no Fundo das Associações de Classe do Ministério das Corporações e nos *Anuários Comerciais* da cidade publicados ao longo deste período.

FIGURA 7 - ASSOCIAÇÕES OPERÁRIAS (1890-1900)



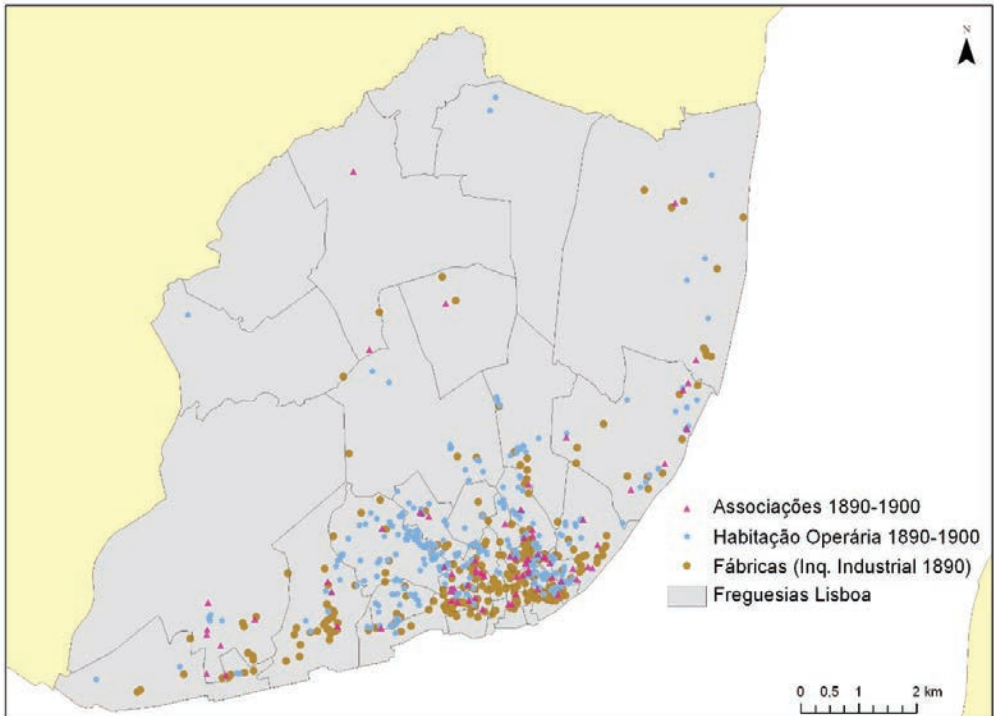
Fonte: Elaborada pela autora (2020).

A cartografia digital construída a partir dos dados referidos anteriormente permite a análise da disseminação dos diferentes tipos de associações operárias (Figura 7), revelando uma dispersão em grande parte próxima dos locais de habitação operária (Figura 5). O que aponta para que estas organizações – que tinham variadas funções tanto a nível político e reivindicativo como de cariz cultural e de instrução e, até, relacionadas com apoio social e económico – tenham representado um polo agregador das, ou de algumas, comunidades operárias. Esta hipótese é ainda suportada pelo facto das associações de carácter social (cooperativas, mutualistas, sociedades recreativas e musicais) estarem maioritariamente perto das zonas onde se concentrava a habitação das “classes laboriosas” e, por terem propósitos diferentes, as associações de classe e políticas estarem naturalmente mais próximas do centro e dos pontos de decisão económica e política. Destas últimas, as que não estão no centro da cidade situam-se junto às fábricas com muitos operários. Ou seja, as associações de classe concentravam-se perto do poder, junto dos políticos ou dos patrões.

ESPAÇOS DA LISBOA OPERÁRIA E INDUSTRIAL

Das sobreposições e hiatos territoriais destas três vertentes da vivência operária (Figura 8) podemos entender que nesta última década do século XIX existia em Lisboa uma dualidade de contextos que ocupavam diferentes espaços do território da cidade. Duas realidades paralelas numa mesma cidade – por um lado, o centro histórico e, por outro, as periferias industriais.

FIGURA 8 - ESTABELECIMENTOS INDUSTRIAIS, HABITACIONAIS E ASSOCIAÇÕES OPERÁRIAS (1890-1900)



Fonte: Elaborada pela autora (2020).

A zona central, representante de uma realidade industrial “em extinção”, mostrava-se como um “centro industrial” que ainda não se tinha transferido para as periferias industriais em afirmação. Caracterizado por uma heterogeneidade social e económica, com uma indústria muito diversificada em termos setoriais incluída no tecido urbano, na qual a habitação operária se misturava com os palácios e as casas burguesas.

Nas zonas periféricas, a ocidente e oriente da cidade, desponta a “nova cidade industrial”. Caracterizada por uma maior homogeneidade social e de produção fabril, onde a existência de grandes fábricas com tecnologia moderna e alta grande concentração de operários e operárias provocou uma aproximação entre locais de trabalho e de habitação, potenciando uma crescente segregação do espaço urbano.

CONCLUSÃO

A leitura espacial da Lisboa industrial e operária na última década do século XIX permite concluir que população operária não existia só na “Lisboa Industrial” das zonas periféricas, a ocidente e oriente do centro tradicional, em afirmação nesta época. A classe operária também estava, porque lá trabalhava, na “Lisboa Elegante” da Baixa, das lojas, dos cafés e da política. A habitação operária ocupava fundamentalmente os espaços permitidos pela pressão da construção da cidade burguesa, que também se está a instituir nesta altura: nas áreas onde havia espaço livre, com o reaproveitamento das traseiras de prédios e quarteirões, antigos conventos e outros edifícios disponíveis. nas zonas mais centrais da cidade, e/ou onde os terrenos disponíveis eram mais baratos e permitiam a construção por parte de donos de fábricas ou proprietários industriais.

Já o emergente associativismo operário afirmava-se por uma dupla intervenção, com características territoriais próprias. A intervenção política, por parte das associações de classe e políticas, aproxima-se dos centros de decisão política – o centro da cidade e os locais de trabalho - enquanto que a intervenção social e cultural se aproxima das comunidades – dos seus locais de habitação.

A utilização do espaço como variável fundamental desta investigação e a associação entre as metodologias usuais na ciência histórica e as ferramentas SIG possibilitou o abarcar de uma grande diversidade de dados e, assim, fazer novas análises, equacionar de novas hipóteses e reexaminar questões antigas. Tornou-se, portanto, possível uma “visualização” e reinterpretação da Lisboa industrial e operária do final do século XIX, onde as alterações de carácter urbano, social e económico da cidade potenciaram a segregação das diferentes classes sociais. Consubstanciando um momento de transição fundamental para a formação da classe operária lisboeta. Tal como escreveu Katznelson (1986, p. 6), “A cidade interclassista quebra-se [...], com esta separação entre trabalho e casa e entre as classes sociais no espaço, as relações de classe são vividas não só no local de trabalho, mas também fora do local de trabalho, nas comunidades residenciais”.

FUNDOS ARQUIVÍSTICOS CONSULTADOS

- Arquivo do Gabinete de Estratégia e Planeamento do Ministério da Solidariedade e Segurança Social
 - *Fundo das Associações de Classe*
- Arquivo do Governo Civil de Lisboa
 - *Cadastro de Associações Socioprofissionais e Mútuas*
 - *Estatutos das Associações de Classe e Recreativas*
- Arquivo Municipal de Lisboa
 - Núcleo do Arco do Cego
 - *Urbanismo e Obras / Planeamento Urbanístico / Projectos de Arruamentos*
 - Núcleo Histórico
 - *Licenças e Visitas sanitárias (1890-1900)*
 - *Processos de obras*
- Arquivo Nacional Torre do Tombo (IANTT)
 - *Governo Civil de Lisboa (1ª incorporação)*
 - Processos de instalações industriais

REFERÊNCIAS

- Publicações Oficiais

REPARTIÇÃO DE ESTATÍSTICA DO MINISTÉRIO DAS OBRAS PUBLICAS, COMMERCIO E INDUSTRIA. *Anuário estatístico do reino de Portugal*. Lisboa: Imprensa Nacional, 1890-1900.

CALDEIRA, P (Coord.). *Anuário comercial de Portugal, Ilhas e Ultramar*. Lisboa, 1890-1900.

DIRECÇÃO DA ESTATÍSTICA GERAL E COMMERCIO. *Censo da população do Reino de Portugal a 1 de Dezembro de 1890*. Lisboa: IN, 1896-1901.

DIVISÃO DE REABILITAÇÃO URBANA DOS PÁTIOS E VILAS. *Câmara Municipal de Lisboa. Plano de Pormenor de Salvaguarda de Pátios e Vilas*. Lisboa: CML, 1993.

FINO, G. C. C. *Collecção de Legislação Industrial*. Lisboa: IN, 1893.

- BODENHAMER, D. J.; CORRIGAN, J.; HARRIS, T. M. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010.
- ELL, P. GIS, e-Science, and the Humanities Grid. In: BODENHAMER, D. J.; CORRIGAN, J.; HARRIS, T. (Ed.). *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010. p. 143-166.
- ETHINGTON, P. J. Placing the past: "Groundwork" for a Spatial Theory of History. *Rethinking History*, v. 11, n. 4, p. 465-493, 2007.
- FERREIRA, V. M. *A cidade de Lisboa: de capital do Império a centro da Metrópole*. Lisboa: Publicações Dom Quixote, 1987.
- FRANK, Z. Layers, Fows and Intersections: Jeronymo José de Mello and artisan life in Rio de Janeiro, 1840's-1880's. *Journal of Social History*, v. 41, n. 2, p.307-328, 2007.
- GREGORY, I.; ELL, P. S. *Historical GIS: Technologies, Methodologies and Scholarship*. Cambridge: Cambridge University Press, 2007.
- HEALEY, R. G.; STAMP, T. R. Historical GIS as a Foundation for the Analysis of Regional Economic Growth. *Social Science History*, v. 24, n. 3, p. 575-612, 2000.
- KATZNELSON, I. *Working-class formation: Nineteenth-Century Patterns in Western Europe and the United States*. Princeton: Princeton University Press, 1986.
- KNOWLES, A. K. Historical GIS: The Spatial Turn in Social Science History. *Social Science History*, v. 24, n. 3, p. 451-470, 2000.
- KNOWLES, A. K. *Past Time, Past Place: GIS for History*. California: Esri Press, 2002.
- LEFEBVRE, H. *The Production of Space*. Maiden: Blackwell Publishing, 2007.
- MOORE, B. *As origens sociais da ditadura e da democracia: senhores e camponeses na construção do mundo moderno*. Lisboa: Edições 70, 2010.
- OYÓN, J. L. Obreros em la ciudad: líneas de un proyecto de investigación en historia Urbana. *Historia Contemporánea*, n. 18, p. 317-345, 1999.
- OYÓN, J. L.; PERMANYER, M. S. Historia Urbana: el espacio no es inocente. *Historia Contemporánea*, n. 39, p. 387-401, 2009.
- PEREIRA, N. T.; BUARQUE, I. (Fotog.). *Prédios e vilas de Lisboa*. Lisboa: Livros Horizonte, 1995.
- RODRIGUES, T. *Nascer e morrer na Lisboa Oitocentista: migrações, mortalidade e desenvolvimento*. Lisboa: Edições Cosmos, 1995.

SCHWARTZ, R. M.; GREGORY, I.; MARTI-HENNEBERG, J. History and GIS: Railways, Population change, and Agricultural Development in Late Nineteenth Century Wales. In: DEAR M.; KETCHUM J.; LURIA, S.; RICHARDSON, D. (Ed.). *GeoHumanities: Art, History, Text at the Edge of Place*. Abingdon: Routledge, 2011. p. 251-266

SILVEIRA, L. N. E.; ALVES, D.; LIMA, N. M.; ALCÂNTARA, A.; PUIG-FARRÉ, J. Population and Railways in Portugal, 1801-1930. *Journal of Interdisciplinary History*, v. 42, n. 1, p. 29-52, 2011.



Novos desafios para antigas fontes: a experiência DOVIC na nova linguística histórica

*Cristiane Namiuti
Jorge Viana Santos*

INTRODUÇÃO

As fontes documentais que fundamentaram os estudos em humanidades em diferentes momentos da história possuem uma materialidade restrita. O suporte material dessas fontes caracteriza um tipo de documento que limita seu acesso a um tempo e espaço: trata-se de um documento físico que, por ser físico, traz algumas complexidades¹ que marcaram e marcam o modo de fazer humanidades desde a antiguidade.

Uma vez que a investigação na área das humanidades deparou-se e depara-se com os limites e as possibilidades do suporte material do documento histórico, o estudioso interessado em consultar as fontes, além de ter permissão, necessita estar no mesmo espaço físico do documento. Em relação à forma, o documento, por ser físico, tridimensional, não mutável e não padronizado, carece de meios e técnicas que garantam o resgate das informações nele contidas. O suporte material do objeto, portanto, caracteriza um modo de fazer humanidades. Prova disso é que, com o advento das tecnologias, surgem novos suportes para as fontes documentais, a exemplo do digital, trazendo com ele não só novas possibilidades e limites, mas uma nova forma de fazer humanidades, hoje denominada como Humanidades Digitais.

Com o suporte digital, a complexidade em relação ao acesso muda, pois o estudioso interessado em consultar as fontes, além de ter permissão, não necessita estar no mesmo espaço físico do documento. Em relação à forma, o documento, por

¹ Ao tratar do objeto livro manuscrito, Santos (2010) postula que as principais complexidades do documento físico são: o acesso, a forma e a fragilidade e/ou raridade .

ser digital, é uma visão mutável que possibilita o controle sobre sua apresentação, ao passo que os meios e técnicas que garantem o resgate das informações contidas em cada documento se beneficiam das tecnologias digitais. Quanto à fragilidade e/ou raridade dos documentos, o suporte digital amplia as possibilidades de manuseio graças, por exemplo, à possibilidade de duplicação.

A reprodução de documentos históricos para a pesquisa científica necessita da garantia de fidedignidade. No caso de *corpora* eletrônicos, esse pressuposto fundamental precisa ser integrado com as exigências impostas pelas vertentes tecnológica, computacional e linguística, tais como: o arquivo digital, a confiabilidade do código, a necessidade de quantidade e de automação no processamento de dados. Nesse sentido, aqui questionamos: como se beneficiar das vantagens do suporte digital sem dispensar a autenticidade do documento original físico?

Santos (2010), buscando solucionar o problema relativo à fidedignidade entre o documento físico (DF) e sua versão digital, para o *corpus* de Documentos Oitocentistas de Vitória da Conquista e região (DOViC)², planeja uma forma de transposição de documentos manuscritos históricos para o meio digital através da fotografia cientificamente controlada. Em Namiuti e Santos (2015) denominamos o alvo desta transposição material cientificamente controlada de documento digital imagem (DDI). Como sequência da etapa de transposição, há a necessidade de transcrição para transformar o DDI em documento digital texto (DDT) para posterior tratamento computacional do *corpus* DOViC. Assim, utilizando-se das vantagens da tecnologia, da computação e da linguística, como pressupõem as Humanidades Digitais, juntamente com o grupo de pesquisadores do Laboratório de Pesquisa em Linguística de *Corpus* (LAPELINC) da Universidade Estadual do Sudoeste da Bahia (UESB), *campus* de Vitória da Conquista, temos desenvolvido e aplicado o método LAPELINC com um fluxo

² *Corpus* compilado no âmbito do projeto financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq): Cristiane Namiuti e Jorge Viana Santos, *Memória conquistense: implementação de um corpus digital* (2013), que dá continuidade ao trabalho iniciado pelos autores nos projetos *Memória Conquistense: recuperação de documentos oitocentistas na implementação de um corpus digital* (2009) e *Corpora Digitais Para a História do Português Brasileiro – região Sudoeste da Bahia: Aliança PHPB – Tycho Brahe* (2010), financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESP).

de trabalho (*workflow*) que compreende três etapas para a construção de *corpora* eletrônicos anotados, cientificamente controlados (NAMIUTI; SANTOS, 2015; NAMIUTI; SANTOS; COSTA, 2016): (i) transposição; (ii) transcrição; (iii) compilação de *corpora*. Em tais etapas, utilizam-se tecnologias para desenvolver ferramentas e instrumentos que auxiliem na coleta de material, levantamento, organização, armazenamento e divulgação de dados, além de se promover o acesso para distribuição do material coletado a pesquisadores interessados. Dessa forma, garante-se a possibilidade de se beneficiar das vantagens do suporte digital sem dispensar a autenticidade do documento original físico.

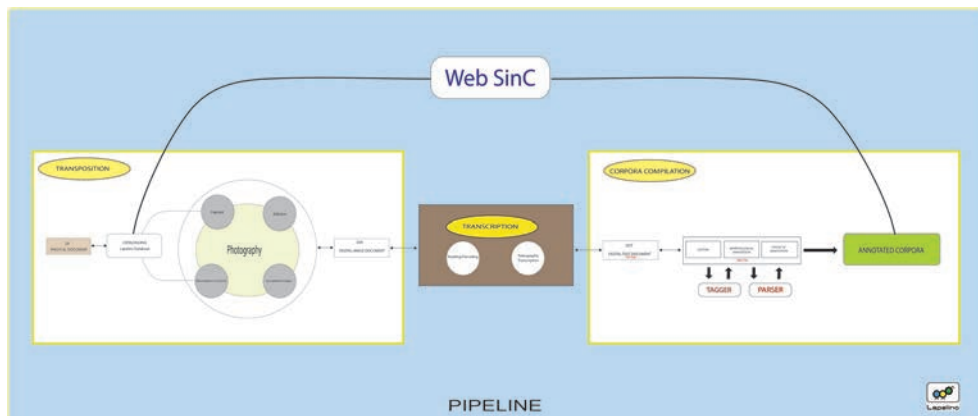
Nesta perspectiva, o nosso propósito neste trabalho é apresentar descritivamente as três etapas do método LAPELINC.

DO PAPEL AO TEXTO DIGITAL: O método LAPELINC

No desenvolvimento do método LAPELINC, procuramos enfatizar a necessidade de se colocar na posição de um Pesquisador Formador de *Corpora* (PFC)³ que tem como objetivo central a reflexão sobre a complexidade das fontes documentais em papel a se tornar documento digital. No processo cientificamente controlado de transposição material do papel para o suporte digital, primeiramente, partindo do DF, constrói-se o DDI que servirá de fonte original no meio digital para os processos de constituição de *corpora* eletrônicos anotados que terá como resultado o DDT com camadas de anotação que registram a memória do processamento na construção de *corpora* anotados, como demonstra a Figura 1:

³ Santos (2010) opõe duas categorias de pesquisador que trabalha com *corpora* para pesquisa: o Pesquisador Pragmático (PP) que constrói um *corpus* para uma pesquisa específica sem necessariamente se preocupar com possibilidades de abertura e reutilização do material gerado e o Pesquisador Formador de *Corpora* (PFC) que constrói *corpus* buscando cientificidade e abertura do *corpus* para uso e reuso.

FIGURA 1 - FLUXO DE TRABALHO (*WORKFLOW*) DO MÉTODO LAPELINC



Fonte: Namiuti; Santos (2015).

Assim, nas subseções seguintes, apresentaremos as três grandes etapas do fluxo de trabalho (*workflow*) do método LAPELINC: (i) transposição; (ii) transcrição; (iii) compilação de *corpora*.

TRANSPOSIÇÃO

A Transposição material do DF para o DDI, no método LAPELINC, emprega um método de fotografia cientificamente controlada⁴ que desenvolvemos e temos aplicado e aperfeiçoado desde 2008 (BRITO, 2015; SANTOS, 2008, 2013; SANTOS; BRITO, 2014; NAMIUTI; SANTOS; LEITE, 2011; NAMIUTI; SANTOS; COSTA; FARIAS, 2013).

A transposição enquanto passagem de um documento físico para um documento digital possibilita uma nova forma de acesso aos documentos, a visual fotográfica, que dentre algumas vantagens permite o acesso a novos formatos do documento, como por exemplo, os arquivos em formato Raw (LONG, 2004) e JPEG (LONG, 2004), e a visualização eletrônica do documento a partir de ferramentas como: ampliação, recorte, contraste, brilho, cor, tons, bem como

⁴ Para detalhes sobre a fotografia como componente de um método cientificamente controlado de formação de *corpora*, ver Santos e Brito – A fotografia digital na transposição de documentos manuscritos históricos–, neste eBook.

uso de máscaras, filtros, *layers* (camadas), dentre outras. Com vistas a garantir a reprodução fidedigna de documentos, tal etapa do método apresenta cinco subetapas principais: (1) catalogação (controle); (2) captura fotográfica da imagem do original; (3) edição fotográfica; (4) criação de imagens de uso coindexadas à imagem original; (5) análise descritiva/topográfica.

CATALOGAÇÃO (CONTROLE)

O processo de catalogação (controle) envolve a captura de informações das fontes, tais como a catalogação de dados de um livro a ser fotografado e a catalogação de dados das folhas-imagem componentes de um livro já fotografado. Tal processo constitui a primeira subetapa da etapa de transposição do Método LAPELINC, englobando duas ações: 1) elaboração do Catálogo Visual e 2) elaboração de Dossiê de Observações Pertinentes (DOP).

O Catálogo Visual de dados dos livros, desenvolvido por nós para o método LAPELINC, conforme descreve Santos e Namiuti (2014, p. 8)⁵,

[...] agrega dupla funcionalidade: tanto funciona como catálogo físico descritivo de cada documento em termos de suas características físicas fundamentais, quanto possibilita que o pesquisador/consulente visualize o documento, especificamente em suas cinco imagens-chave: capa frontal, termo de abertura, termo de encerramento, capa final.

O Catálogo Visual serve de base para a criação de um banco de dados sistematizados que permite realizar buscas automáticas de metadados e possui duas partes complementares: a) parte descritiva e b) parte imagética.

A parte descritiva do Catálogo Visual envolve a criação de um catálogo físico descritivo das características fundamentais de cada livro, que são essenciais para a alimentação de um banco de dados, conforme se observa no exemplo a seguir (Figura 2).

⁵ SANTOS, J. V.; NAMIUTI, C. (Coords.). *Corpora digitais para a história do português brasileiro: região Sudoeste da Bahia; Aliança PHPB - Tycho Brahe*, FAPESB 6171/2010, Vitória da Conquista, 2014. (Relatório Final).

FIGURA 2 - RECORTE DE TABELA DE DADOS DO CATÁLOGO FÍSICO DESCRITIVO REFERENTE AOS LIVROS DE NOTAS 1 A 6

	LIVRO No.	ANO	TIPO	TAMANHO (cm)			CAPA	FOTO	Observação	Data catalogação	Catalogador	Conférence
				Altura	Largura	Profundidade						
			ESCRITURAS									
1	1	1841-1848	Escrituras	31,5	22,3	:3	Marron	N.	Fita adesiva na lombada	22.07.13	Giovane, Adilson e Jorje	OK
2	2	1841-1850	Escrituras	22	22	:3	Bigo	N.	Fita adesiva na lombada	22.07.13	Giovane, Adilson e Jorje	OK
3	3	1849-1858	Escrituras	31	23	:5	Capa couro	S.		27 e 29/08/2012	Vanessa, Giovane, Jorje	OK
4	4	1856-1863	Escrituras	24	34	:2	Livro sem capa original	N.	Com partes das folhas traseira e final soltas. Livro ligado sem encadernação, os pedaços de pedras amarelas. Livro sem número inventariado como 4.	22.07.13	Giovane, Adilson e Jorje	OK
5	5	1852-1860	Escrituras	34	23	:2	Capa Papelão	S.		27 e 29/08/2012	Vanessa, Giovane, Jorje	OK
6	6	1866-1870	Escrituras	33	22	:1	Ampl.	N.	Fita de tecido na lombada	22.07.13	Giovane, Adilson e Jorje	OK

Leitura_Catálogo_físico_notas_notas_publicadas_181213_fornecido_292014_1316.xlsx



Fonte: Elaboração própria, a partir do Corpus DOViC (2014)⁶.

Dessa tabela de dados são extraídas informações que possibilitarão/comporão o vínculo entre o DF e o DDI, que se forma inicialmente a partir das imagens-chave da parte imagética do Catálogo.

Por sua vez, a parte imagética do Catálogo Visual envolve a coleta de cinco imagens-chave do objeto (livro) enquanto documento manuscrito histórico jurídico, a saber: capa frontal, termo de abertura, termo de encerramento, capa final. É o que se vê na Figura 3:

⁶ SANTOS, J. V.; NAMIUTI, C. *DOViC* (Documentos Oitocentistas de Vitória da Conquista e região). *Corpus* eletrônico. Vitória da Conquista: Universidade Estadual do Sudoeste da Bahia, 2014.

FIGURA 3 - PARTE IMAGÉTICA DO CATÁLOGO VISUAL DO DOCUMENTO MACRO E1-C11 REFERENTE AO *LIVRO DE NOTAS 1 (1841 A 1848)* DO ARQUIVO DO FORUM DE VITÓRIA DA CONQUISTA-BA CONTENDO 5 IMAGENS-CHAVE



Fonte: *Corpus DOVIC* (2014).

Estabelecido o vínculo inicial entre o DF e o DDI através do Catálogo Visual do LAPELINC (parte descritiva e parte imagética), passa-se à segunda ação da etapa catalogação (controle), a elaboração do DOP como parte preparatória para a captura posterior das folhas-imagem dos DDIs.

No DOP é realizada a pré-análise de cada folha dos documentos constantes em um livro, anotando-se as observações de ordens fotográficas, filológicas, de edição, dentre outras. A Figura 4 ilustra este objeto.

FIGURA 4 - TRECHO DE UM DOSSIÊ DE OBSERVAÇÕES PERTINENTES PREPARATÓRIO PARA A CAPTURA

LIVRO No.	ANO	TIPO	TAMANHO			CAPA	CÓDIGO DE OBSERVAÇÃO			FOLHA-IMAGEM	Tipo de Observações	OBSERVAÇÕES PARTICULARES (OP-XX)
			A	L	P		CÓDIGO LAPELINC	Imagens				
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	34	Contracapa frente	Filológica	Folha anexada posteriormente. Transcrição do termo de abertura.
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	35	SN	Filológica	Folha solta.
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	27	Termo de abertura	Filológica	
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	10	198	Filológica	Seta de 198 para 201
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	22	203	Filológica	Termo de encerramento com nome De: "Termo de Apresentação"
1	1841-1848	Escrifuras (Notas)	31,5	22,3	3	Marrom	C11-E01	412	22	204	Filológica	Termo de encerramento com nome De: "Termo de Apresentação"

Fonte: Elaboração própria, a partir do *Corpus DOVIC* (2014).

Esta última ação da subetapa catalogação (controle) é de suma importância para o desenvolvimento da fotografia sequenciada, na medida em que permite a anotação de informações fotográficas, filológicas, dentre outras, que otimizam o processo de captura de imagens.

CAPTURE FOTOGRAFICA DA IMAGEM DO ORIGINAL

A captura fotográfica da imagem do original envolve a fotografia sequenciada dos documentos utilizando equipamentos adequados, inserindo na imagem a quantidade necessária de dados que garanta a sua relação com o objeto que a originou. Ou seja: fotografa-se o DF para se formar o DDI. No que diz respeito à captura fotográfica em si optamos pela captura digital, com câmera e lente⁷.

A câmera fotográfica digital captura a imagem via CCD⁸ (Sensor Digital) e, no que concerne ao formato das imagens, a captura fotográfica no LAPELINC atende ao padrão recomendado pelo Conselho Nacional de Arquivos (CONARQ, 2010), que sugere a captura de uma matriz no formato Raw que, em nosso caso, tem a dimensão de 4256 x 2832 pixels, com profundidade de 14 bits por canal (RGB⁹), gerando um arquivo não comprimido (*uncompressed*) de 12.1 megapixels. Juntamente com esse arquivo, a câmera gera um arquivo em JPEG, com resolução mínima de 300 dpi. O mesmo arquivo em Raw possibilita, ainda, a criação de arquivos de alta resolução para armazenamento no formato TIFF¹⁰, outra recomendação do CONARQ (SANTOS; BRITO, 2014).

Para a fotografia sequenciada dos documentos, no processo de captura fotográfica da imagem do original, desenvolvemos e temos aperfeiçoado a Mesa Cartesiana¹¹, instrumento com vistas a garantir a inserção dos dados essenciais do DF na imagem a ser capturada de forma a viabilizar de modo controlado e

⁷ Santos e Brito, neste eBook, abordam a fotografia no método LAPELINC.

⁸ As câmeras digitais não utilizam filme, elas capturam a luz da cena a ser fotografada por meio de células fotossensíveis chamadas *Charged Coupled Device* (CCD) (MELO; ALMÉRI, 2013, p. 18).

⁹ RGB: *Red Green Blue* – espaço de cores primárias considerado em imagens (LONG, 2004).

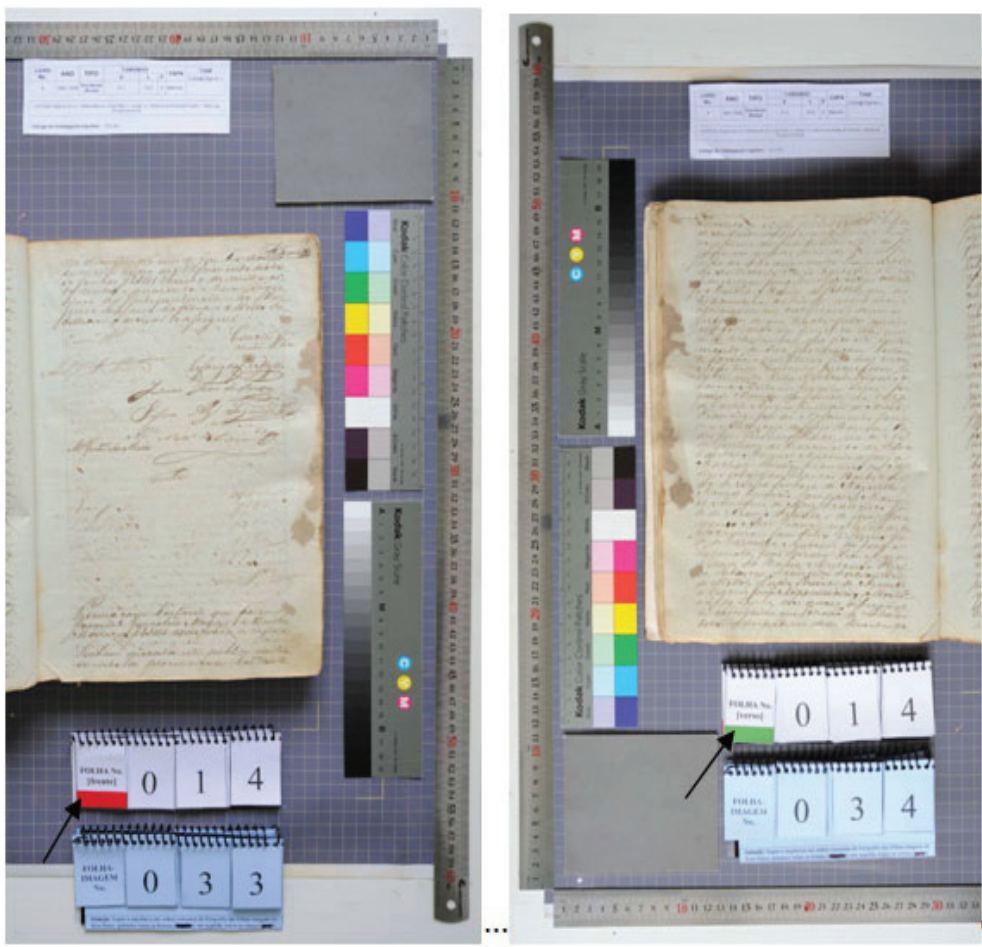
¹⁰ TIFF, arquivo digital não comprimido, o que, em comparação, por exemplo, ao JPEG, o torna superior em qualidade de imagem (LONG, 2004).

¹¹ Para detalhes sobre a Mesa Cartesiana, ver Santos e Brito (2014).



cientificamente padronizado a fotografia técnica de documentos, como podemos ver exemplificado na Figura 5.

FIGURA 5 - FOTOGRAFIA SEQUENCIADA DE UM DOCUMENTO DO *CORPUS* DOVIC SOBRE A MESA CARTESIANA: FOLHA FRENTE COM INDICADOR VERMELHO (ESQUERDA); FOLHA VERSO COM INDICADOR VERDE (DIREITA)

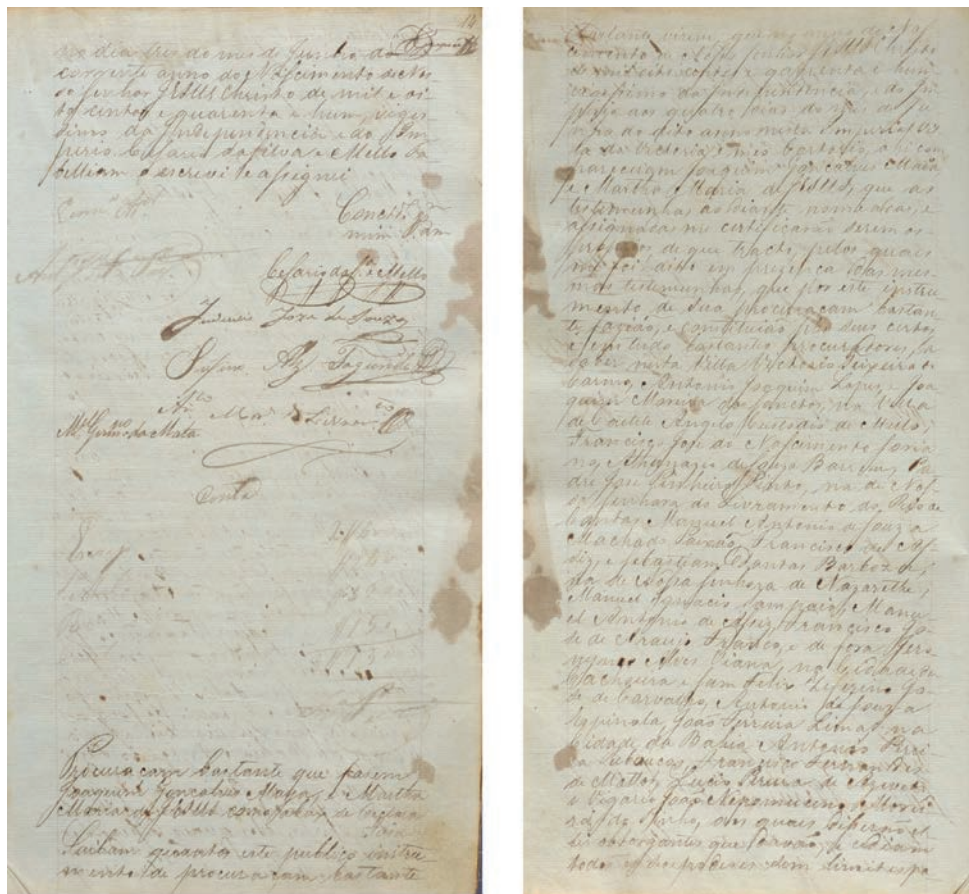


Fonte: *Corpus* DOVIC – folhas-imagem 33 e 34 do Documento Macro E1-C11 referente ao *Livro de Notas 1 (1841 a 1848)* do Arquivo do Fórum de Vitória da Conquista-BA.

EDIÇÃO FOTOGRÁFICA

As imagens capturadas são editadas, em termos de *crop*, corte para deixar visível apenas o documento, sem as informações visuais complementares integrantes da Mesa Cartesiana, ficando pronto, assim, o DDI original, do qual podem ser geradas outras imagens/arquivos, conforme exemplificado na Figura 6:

FIGURA 6 - FOLHA FRENTE EDITADA (ESQUERDA) E FOLHA VERSO EDITADA (DIREITA)

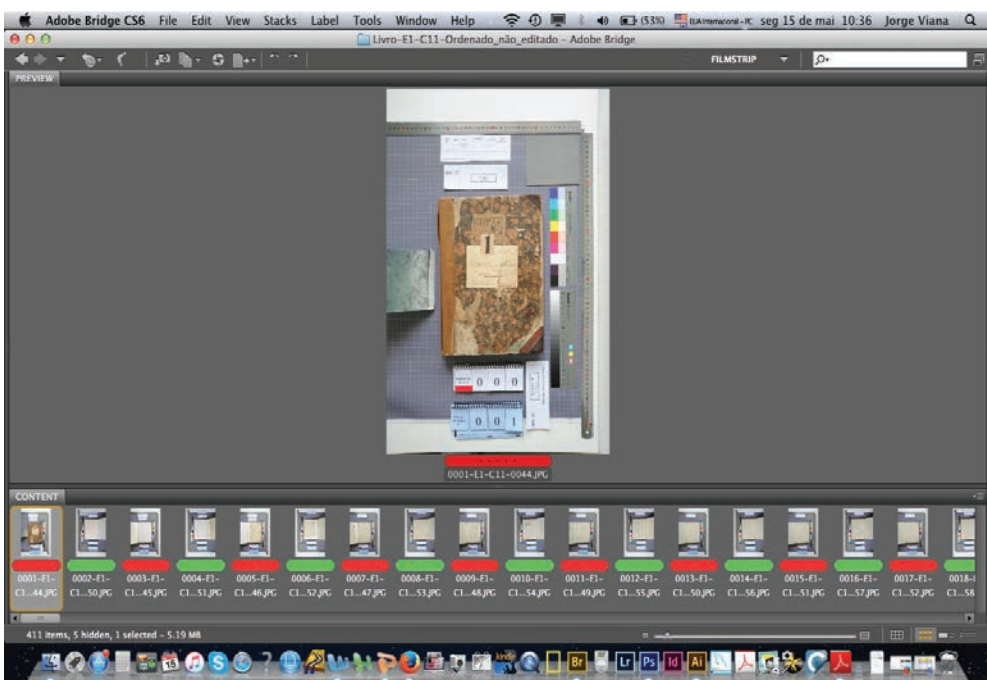


Fonte: Corpus DOVIC – Folhas Imagens 33 e 34 do Documento Macro E1-C11 referente ao Livro de Notas 1 (1841 a 1848) do Arquivo do Forum de Vitória da Conquista-BA.

CRIAÇÃO DE IMAGENS DE USO COINDEXADAS À IMAGEM ORIGINAL

A partir do DDI original, são criados, conforme a demanda dos pesquisadores/consulentes, outros DDIs, por exemplo: a) Livro ordenado não editado; b) Livro ordenado editado – JPEG de alta resolução; c) Livro ordenado editado – JPEG de baixa resolução; d) Livro ordenado editado – PDF¹². Um exemplo de visualização de um *livro ordenado não editado* pode se ver na Figura 7:

FIGURA 7 - VISUALIZAÇÃO DO DOCUMENTO MACRO E1-C11 ORDENADO NÃO EDITADO REFERENTE AO LIVRO DE NOTAS 1 (1841 A 1848) DO ARQUIVO DO FORUM DE VITÓRIA DA CONQUISTA-BA



Fonte: *Corpus DOViC* (2014).

¹² PDF: *Portable Document Format* (Formato Portátil de Documento), um formato de arquivo criado pela empresa Adobe Systems.

FIGURA 9 - DESCRIÇÃO LINGUÍSTICO-JURÍDICA (PARCIAL) DOS DOCUMENTOS INTERNOS DE UM LIVRO DO CORPUS DOVIC



Lapelino
Laboratório de Pesquisa em Linguística de Corpus

Pesquisador: Celina Marília Melo Maurício Roque – Livro E21 – C11
Bolsa: [] Cnpq [x] Fapesb [] Uesb
Orientador: Prof. Dr. Jorge Viana
Data início: 29/04/2015
Data término: 08/06/2015

TABELA COMPARATIVA DE DEFIÇÕES DE TERMOS

	Dicionários históricos (sec. XVII/XXI)			Dicionários atuais (sec. XXI)		
	Dicionário 1 Dicionário da língua portuguesa composto pelo padre D. Rafael Bluteau, reformado, e acrescentado por António de Moraes Silva natural do Rio de Janeiro, Silva, António de Moraes, 1755-1824.	Dicionário 2 Vocabulário Portuguez e Latino Bluteau, Rafael, 1638-1734	Dicionário 3 Dicionário da língua brasileira. Pint o, Luiz Maria da Silva, 1775-1869 Obs: (Esse dicionário não tem o rodapé indicando o número da página. Por isso não há como colocar a referência da página como nos outros.)	Dicionário 1 FERREIRA, Aurélio Buarque de Holanda. Novo Aurélio: O Dicionário da Língua Portuguesa Século XXI. Editora Nova Fronteira; 3ª edição, Rio de Janeiro, 1999	Dicionário 2 HOUAISS, Antônio. Dicionário Houaiss da Língua Portuguesa. Editora Objetiva; 1ª edição, Rio de Janeiro, 2009.	Dicionário 3 Dicionário AULETE. Caldas. Dicionário contemporâneo da Língua Portuguesa. Editora Delta S. A.; 2ª edição, 5 volumes, 1964.
Palavra						
ACTA	1-Resoluções, determinações dos Concílios, Parlamentos, e semelhantes corporações. 2-Actas dos Sanctosm escrituras,	Assentos, ou determinações registadas em escrituras publicas. (pág 11, volume 01, Letra A).	Determinações, sanções dos Concílios, etc. de quaisquer Corporações, Escripturas, Memoriaes, etc,	1-registro escrito no qual se relata o que se passou numa sessão, convenção, congresso, etc.	1-registro ou resenha de fatos ou ocorrências verificadas e resoluções tomadas	1. Narração por escrito do que se passou em uma sessão, ou em uma cerimônia; resumo dos atos deliberativos de um corpo coletivo; lavrou-se uma ata. (p. 426)

Fonte: Elaboração própria, a partir do Corpus DOVIC (2014).

Como podemos ver, a etapa de transposição envolve um método fotográfico específico que pressupõe domínio da fotografia (*photography*) enquanto linguagem e enquanto técnica, e necessita de equipamentos e aparato técnico auxiliar específicos.

TRANSCRIÇÃO

O Método LAPELINC estabelece, para a transcrição, uma estreita relação entre a fotografia e a Paleografia nas suas duas etapas fundamentais: (i) leitura e decodificação do DDI, e (ii) transcrição paleográfica. Na paleografia praticada lendo e decodificando um DDI, não um DF, o paleógrafo pode se beneficiar das vantagens do suporte digital¹.

¹ Sobre a relação entre Fotografia e Paleografia, ver Brito (2015).

A etapa de Transcrição envolve as soluções técnicas para a edição especializada de textos antigos em meio eletrônico. Compreende uma etapa inicial de leitura e transcrição paleográfica dos DDIs, gerando um documento texto que servirá de entrada para a compilação do DDT.

Como aponta Paixão de Sousa (2006), já se destacam para o processo de trabalho as singularidades técnicas do meio eletrônico. Ao se transcrever ou digitalizar um texto, ou seja: na passagem do meio físico para o meio digital, está-se alterando substantivamente o sistema de codificação da informação, de visual para computacional-matemático, isto é, de DF para DDI para DDT.


A transcrição, se realizada da forma da paleografia tradicional, encerra potencial de perda de informações. Em textos manuscritos ou impressos, a sequência de caracteres que forma o texto, bem como diversas informações estruturais importantes (por exemplo, a paragrafação), são codificadas de modo direto e visual. Em textos processados eletronicamente, essas informações precisam ser codificadas indiretamente por programas de processamento de texto. Assim, a transcrição no método LAPELINC segue a recomendação de Paixão de Sousa (2006) de que, na produção de textos em meio eletrônico com a finalidade específica de construção de *corpora* de língua, se deve fazer uso de um processamento controlado que permita a codificação de uma grande variedade de informações, de modo confiável e transportável. De acordo com esse pensamento, as estruturas no processamento eletrônico de textos precisam ser anotadas em alguma linguagem de anotação, e depois traduzidas ou lidas por uma programação que gera a apresentação final do texto.

Na Figura 10 temos um exemplo de visão associando o DDI ao DDT na sua forma fiel ao texto original, visão esta possível apenas graças ao processo de compilação de *corpora* eletrônico que será apresentado a seguir.

FIGURA 10 - VISÃO PARCIAL DO DDI E DO RESPECTIVO DDT DA CARTA DE LIBERDADE DA CABRINHA DE NOME SOFIA

[:| 1.29
[:| Cesario da Silva e Melo : "Carta de liberdade da Cabra de nome Sofia"
[:| Versão Texto original [DO_BREAKLINES]
[:| Arquivo gerado pela ferramenta E-Dictor

"Carta de liberdade da Cabra de nome Sofia" (author: Cesario da Silva e Melo,) [extensão: full, 440 palavras]



Livros: folha 40 verso

Carta de liberdade da Cabra de nome Sofia passada por Antonio Jose de Souza Paes, outrora Senhor daquela

Eu Antonio Jose de Souza Paes abaixo **assi-**
gnado, sou possuidor da Cabrinha Sofia
sem embargo algum, e por que **he** minha
vontade, e lhe tenho grande amor, de **hoji**
em diante lhe confiro a liberdade, e **fi**
ca forra, como **si** tal nascesse: podendo
seguir o destino, que lhe parecer como
arbitra de si mesma, e para **seo** titulo
lhe passo a **prezente** carta por mim **escri**
pta, e assignada, que quero tenha **va**
lidade, como **si** fosse verba de **titulo**, **pe**
dindo as Justiças do **Imperio** lhe deem
toda a validade que o Direito outorga. São
Felipo cinco de abril de mil oito centos e quatro
digo mil oito centos e trinta
e quatro = **Antonio Jose de Souza Paes**
= Reconheço verdadeiras e dou fé. **Caetite**

Fonte: *Corpus DOVIC* (2014).

COMPILAÇÃO DE *CORPORA*

Os textos antigos possuem características gráficas e grafemáticas que dificultam o processamento computacional posterior à etapa de transcrição. Por essa razão, os textos precisam ser editados, mas as características do texto original devem ser preservadas, devido à sua importância para estudos linguísticos e filológicos.

No método LAPELINC, as fontes documentais após passarem pelas etapas de transposição e transcrição ganham o formato digital de texto simples (TXT), um requisito necessário para se iniciar o processo de compilação de *corpora*. Esta etapa envolve algumas ferramentas computacionais, tais como o eDictor² – editor

² Ver Paixão de Sousa, Kepler e Faria (2012).

de marcação extensível XML (*Extensible Markup Language*)³. e o WebSinC⁴ – aplicativo web para o trabalho de registro, armazenamento, disponibilização, visão e busca de dados em *corpora* cientificamente controlados.

Os textos transcritos passam pelo processo de edição. As edições dos textos, portanto, são anotadas, segundo esquema de anotação proposto por Paixão de Sousa (2006), o mesmo esquema de anotação do *corpus* Tycho Brahe⁵, mantendo as informações sobre a interferência realizada e sobre o texto original no mesmo arquivo de anotação morfosintática em formato XML.

Incluem-se, neste plano, diferentes graus de interferências de edição – das mais restritas, próprias das edições paleográficas (desdobramento de abreviaturas; decisões de leitura), às mais amplas, próprias das edições modernizadas (atualização de grafia). O DDT no método LAPELINC traz, num único arquivo, todas as informações referentes, ao processamento do documento – metadados, transcrição, informação de edição encaixadas, anotação linguística – em camadas para possibilitar gerar diferentes visões do texto.

A Figura 11 apresenta um trecho da marcação XML de um DDT compilado utilizando-se da ferramenta eDictor:

FIGURA 11 - PARTE DA ANOTAÇÃO XML DA CARTA DE LIBERDADE DA CABRINHA DE NOME SOFIA

<pre>[...] <n>5. Original Text Editor:</n> <v> texto manuscrito</v> </meta> <meta> <n>6. Original Text Reference:</n> <v>Carta de Liberdade da Cabra de nome Sofia passada por Antonio Jose de Souza Paes, outrora Senhor daquela, Livro 1, folha 101 verso e 102 frente e verso, 1845. Arquivo do 1º Tabelionato de Notas Paes (antigo 1º Cartório de Notas do Forum). Vitória da Conquista, Bahia. In: SANTOS, Jorge Viana; NAMIUTI-TEMPONI Cristiane (2014), Corpus de Documentos Oitocentistas de Vitória da Conquista - DOVIC (versão BETA)</v> </meta> [...]</pre>	<pre>[...] <w id="13"> <o>Jose</o> <e t="mod">José</e> <m v="NPR"/> </w> [...]</pre>
--	--

Fonte: *Corpus* DOViC (2014).

³ XML: uma linguagem de editoração proposta pelo World Wide Web Consortium (W3C, XML Technology, 2010, <http://www.w3.org/standards/xml/>) como uma nova alternativa à linguagem HTML.

⁴ Namiuti, Santos e Costa (2015); Costa e Namiuti (2015); Costa (2015).

⁵ Disponível em: <<http://www.tycho.iel.unicamp.br/corpus/>>. Acesso em: 08 set. 2020.

A Figura 11 ilustra, à esquerda, a marcação XML para metadados do documento do *corpus* DOViC, identificado pelo cabeçalho “Carta de Liberdade da Cabra de nome Sofia passada por Antonio Jose de Souza Paes, outrora Senhor daquela”, e, à direita, a marcação para edição e anotação linguística da palavra <w> “José”, no mesmo documento, transcrita em sua forma original <o> “Jose”, modernizada <e t=“mod”> para “José” com anotação linguística morfológica <m V=“NPR”> para nome próprio.

O DDT no método LAPELINC apresenta uma singularidade crucial do trabalho de edição eletrônica que explora as possibilidades próprias do suporte informático de modo a permitir a manutenção do texto original no mesmo plano em que se realizam as interferências editoriais. Assim, o documento eletrônico usado pelo editor contém todas as informações de transcrição e de edições, devidamente codificadas, de forma a garantir a integridade do texto transcrito do início ao fim do processo. Dito de outra maneira, as palavras (e todo o texto nas suas respectivas versões e graus de interferências) são mapeadas, e, por isso, podemos transitar pelas edições e recuperar as informações da palavra original no texto modernizado. É esta a característica que confere controle e confiabilidade às edições eletrônicas assim desenvolvidas.

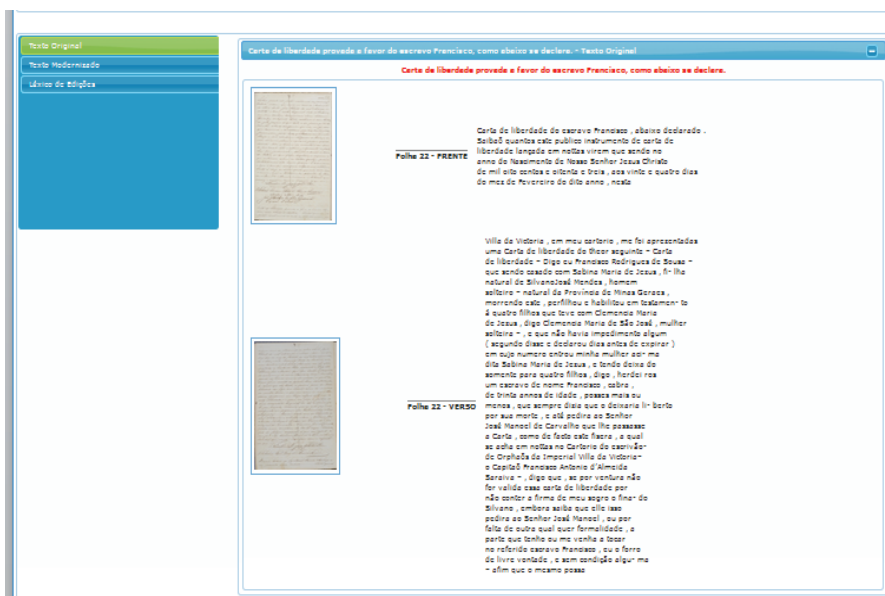
Para gerenciar e disponibilizar os textos do *corpus* digital DOViC criamos a ferramenta WebSinC, pois necessitávamos de recursos específicos, a exemplo de um sistema de gerenciamento de bancos de dados (SGDB), requisito crucial no método LAPELINC para buscar o controle e a cientificidade no processo de construção de *corpora*, garantindo, assim, a integração e relação entre as três etapas do método e conseqüentemente entre os objetos DF, DDI e DDT. Além do sistema SGDB para controle e gerenciamento de informações de dados referente ao workflow do LAPELINC, o WebSinC integra ferramentas de disponibilização e busca de informações em estrutura XML⁶. Uma outra motivação para o desenvolvimento do software WebSinC é o fato de os recursos desenvolvidos e aplicados para a compilação, anotação e busca de dados no *corpus* Tycho Brahe não seguirem um padrão de linguagem: a linguagem utilizada para a anotação de edição e para a anotação morfossintática é XML enquanto a anotação sintática segue o

⁶ Para detalhes sobre o WebSinC e suas funcionalidades, ver Costa (2015).

formato Penn Treebank⁷, um outro formato, que, devido a essa heterogeneidade de linguagem, implica a reduplicação do texto e perda de informação. A homogeneidade na linguagem de edição e buscas favorece a criação de recursos padronizados, permitindo reuso de tecnologia, oferecendo mais flexibilidade para as buscas e exibição dos resultados, além de independência tecnológica para grupos de pesquisa interessados no *corpus*. A ferramenta WebSinC está em busca da homogeneidade de linguagem, uma vez que permite a conversão do formato Penn Treebank para o XML, assim possibilitando a exploração das potencialidades do XML para todo tipo de busca automática.

A Figura 12 é uma tela do WebSinC apresentando uma visualização de um DDT em sua versão editada:

FIGURA 12 - TELA DE VISUALIZAÇÃO DA APLICAÇÃO WEBSINC EXIBINDO A CARTA DE LIBERDADE DO ESCRAVO FRANCISCO – 1883



Fonte: *Corpus DOVIC* (2014).

7 O *Penn Treebank format* (Formato Penn Treebank) é um esquema de anotação sintática de *corpora* desenvolvido pela Universidade da Pensilvânia. O esquema utiliza uma representação arbórea delimitada por parênteses etiquetados, ver Santorini (2010). Disponível em: <<http://www.ling.upenn.edu/hist-corpora/annotation/index.html>>. Acesso em: 02 set. 2015.

CONSIDERAÇÕES FINAIS

O método LAPELINC estabelece pontes entre as antigas fontes e o novo, o suporte digital, na medida em que se utiliza das vantagens da tecnologia, da computação e da linguística, como pressupõem as Humanidades Digitais, e apresenta um fluxo de trabalho (*workflow*) que compreende três etapas integradas para a construção de *corpora* eletrônicos anotados, cientificamente controlados: (i) transposição; (ii) transcrição (iii) compilação de *corpora*. Como demonstrado, o método garante a integração entre as etapas do processo de construção de *corpora*, a exemplo do DOViC, desenvolvendo e aplicando sistemas de gerenciamento de informação que perpassam todas as etapas do fluxo do trabalho, a exemplo do WebSinC, bem como desenvolvendo e aplicando métodos cientificamente controlados, a exemplo do método da fotografia científica, da transcrição a partir do DDI e da anotação em camadas XML a partir do programa eDictor. Tais tecnologias e ferramentas auxiliam na coleta de material, levantamento, organização, armazenamento e divulgação de dados, e igualmente promovem o acesso para distribuição do material coletado a pesquisadores interessados.

A reprodução de documentos históricos para a pesquisa científica feita através do método LAPELINC garante, de fato, a possibilidade de se beneficiar das vantagens do suporte digital sem dispensar a autenticidade do documento original físico, respondendo assim a questão-chave dos estudos diacrônicos na era das Humanidades Digitais.

REFERÊNCIAS

BRITO, G. S. *Do texto ao documento digital: transposição fotográfica de documentos manuscritos históricos para formação de corpora linguísticos eletrônicos*. 2015. Dissertação (Mestrado) - Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista, 2015.

CONARQ. Conselho Nacional de Arquivos. *Recomendações para digitalização de documentos arquivísticos permanentes*. Rio de Janeiro: Arquivo Nacional, 2010.

COSTA, A. S.; NAMIUTI, C. WebSinC: buscas online em corpora sintaticamente anotados. In: CONGRESSO DE HUMANIDADES DIGITAIS EM PORTUGAL: CONSTRUINDO PONTES E QUEBRANDO BARREIRAS NA ERA DIGITAL. Lisboa: Universidade Nova de Lisboa, 2015. (Comunicação oral).

SANTORINI, B. Annotation Manual for the Penn Historical Corpora and the PCEEC. 2010. Disponível em: <<http://www.ling.upenn.edu/hist-corpora/annotation/index.html>>. Acesso em: 03 set. 2015.

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

SANTOS, J. V. Técnicas de transporte do texto manuscrito para o meio digital. In: OFICINA DE LINGUÍSTICA DE CORPUS DA BAHIA (UEFS, UESB, UFBA), 1., 2010, Feira de Santana *Anais*. Feira de Santana: UEFS, 2010. (Comunicação oral).

SANTOS, J. V. *Liberdade na escravidão: uma abordagem semântica do conceito de liberdade em cartas de alforria*. 2008. 274 f. Tese (Doutorado) – Universidade Estadual de Campinas, Campinas, 2008.

SANTOS, J. V. Um método de fotografia técnica documental para formação de corpora digitais de documentos históricos manuscritos. Vitória da Conquista: Universidade Estadual do Sudoeste da Bahia, 2013. (Curso ministrado).

Narrativas transmídia como *performances* de leitura

Fernanda Bonacho

Para exemplificar o modo como uma narrativa transmídia implica uma experiência de leitura diferente, proponho a hiperficção de Kate Pullinger.

FIGURA 1 - *INANIMATE ALICE*, EPISODE 5. HOMETOWN2



Fonte: Disponível em: <www.inanimatealice.com>. Acesso em: 08 set. 2020.

A hiperficção *Inanimate Alice* (IA) começou por ser uma narrativa multimídia e interativa disponível na internet. Porém, ao aplicar algumas técnicas narrativas multimídia combinadas com a utilização de outras plataformas acabou por se tornar uma experiência transmídia numa *paisagem imersiva*, em que, dependendo da performance de leitura, *abre e fecha* diferentes pontos de *entrada e saída* para o universo narrativo (FLEMING, 2013).

A narrativa multimídia resultou da colaboração entre a escrita de Kate Pullinger, a arte digital de Chris Joseph e a produção de Ian Harper. O trabalho, apresentado como um [*Romance Digital*] “*Digital Novel*” – sugerindo claramente uma proposta de romance diferente, é contado em prosa, mas de uma forma cinematográfica e multimídia, reinventando um género no meio digital, particularmente apelativo para a geração contemporânea.

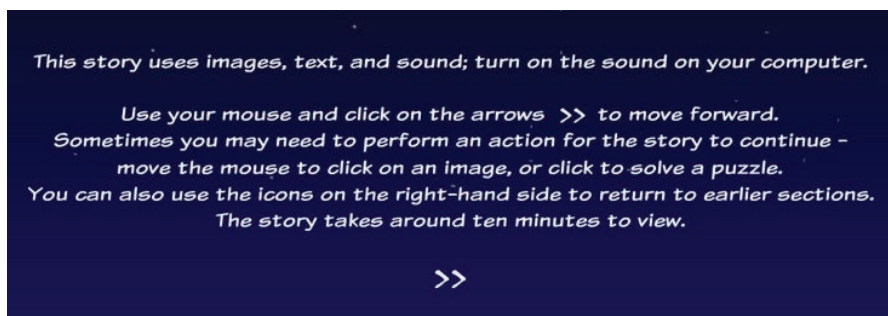
A maior parte dos elementos da série de aventuras que compõem esta história têm origem na cultura moderna e digital (a internet, o jogo, o design interativo,



a composição musical eletrónica), mas os criadores de IA reconhecem que o seu objetivo seria imitar e, simultaneamente, construir uma nova experiência de leitura de um livro.

Depois de mais quinhentos anos de literatura impressa, nas últimas décadas, as tecnologias da informação e comunicação estimularam uma diáspora textual do suporte impresso para o suporte digital e, pela primeira vez, a literatura passou a estar presente nas várias plataformas digitais (computador, *e-readers*, smartphones, *tablets* etc). Os meios audiovisuais e multimídia tornaram-se veículos de uma literatura digitalizada (obras impressas e digitalizadas eletronicamente), mas também parceiros de uma outra literatura, que inclui, de uma forma completamente inovadora, criações originalmente criadas com o computador para serem lidas, quase, exclusivamente em ambiente digital.

FIGURA 2 - *INANIMATE ALICE*, EPISODE 2. ITÁLIA



Fonte: Disponível em: <www.inanimatealice.com>. Acesso em: 08 set. 2020.

Os vários elementos audiovisuais são, de facto, incorporados e articulados nesta narrativa de Alice, existindo, todavia, alguns aspetos que aproximam este trabalho de qualquer história impressa, tais como a própria narração, a densidade da personagem, o ambiente, o tempo e o espaço: nos cinco episódios disponíveis até ao momento (dos 10 previstos) assiste-se ao crescimento de Alice – uma menina que, devido à profissão do pai na indústria petrolífera, passa pela experiência de viver em várias cidades do mundo.

A história inicia-se numa zona remota da China, passa pelos desertos da Arábia Saudita, Alpes, Moscovo, e vai até uma pequena cidade do interior de Inglaterra, no quarto e quinto episódios. Cada episódio mostra uma aventura

completa da protagonista, que acaba por se tornar uma especialista em design e animação de jogos eletrônicos.

No primeiro episódio começamos por ouvir um som estridente e contínuo, como se fosse um rádio não sintonizado. Este indício sonoro funciona de uma forma envolvente para o início de uma série de aventuras em que a protagonista vive perdida e assustada. Quer os indícios sonoros quer visuais estão sempre presentes e evoluem nos cinco episódios para ajudar a contar a história de uma menina, a quem nunca vemos a cara nem ouvimos a voz, mas que está sempre acompanhada deste som eletrônico, que se torna cada vez mais intenso. Nunca ouvimos a voz de Alice, nem de nenhuma das personagens em IA, à exceção de Brad – o seu único amigo imaginário e virtual. A experiência de vida de Alice, muito ligada aos meios eletrônicos, levam-na a criar Brad que vive dentro do *báxi*.

FIGURA 3 - *INANIMATE ALICE*, EPISODE 1: CHINA (ALLUMER CONSOLE)



Fonte: Disponível em: <www.inanimatealice.com>. Acesso em: 08 set. 2020.

O *báxi* é um jogo eletrônico que Alice leva para todo o lado e que lhe permite tirar fotografias, fazer vídeos, jogar, telefonar, aceder à Internet etc. O *báxi* é, de uma forma simbólica, uma parte de Alice, pois possui parte da sua memória e da sua identidade. Aliás, este *báxi* é também o fio condutor da própria narrativa, porque, por vezes, a história só continua se interagirmos com ele. O dispositivo eletrônico é a versão ficcionada de um qualquer dispositivo eletrônico atual, em que convergem diferentes meios de comunicação. É, afinal, o espaço onde Alice descreve os acontecimentos, pinta, desenha, faz rascunhos ou procura informação.

Ora, não será de estranhar que Alice seja muito cuidadosa com o seu *baxi*: é uma parte integrante de si.

À medida que o tempo passa, vamos conhecendo uma Alice cada vez mais adulta, independente e com mais poder de decisão perante as experiências. Quando, no início, Alice é uma menina de oito anos à procura do pai, o desespero dessa procura é representado pelas imagens apressadas e frenéticas, o texto curto e inquieto assistido por sons orientais e insistência da música eletrónica. Este quadro repete-se nos 5 episódios. Cada episódio mantém o mesmo formato (um país diferente com um novo problema que se resolve) e a característica das narrativas em série, em que acompanhamos o crescimento da personagem, com grande enfoque nas suas mudanças – aqui realçado através de todos os elementos da narrativa: texto verbal, imagens, animação e som. É interessante notar que apesar de nunca vermos ou ouvirmos Alice, a sua presença é muito intensa, pois temos constantemente a representação de tudo o que faz, diz, vê, ouve ou sente na concretização de todos os elementos audiovisuais e multimédia.

Convém salientar ainda o facto de que em nenhum episódio existe muito texto verbal. Pelo contrário, a história baseia-se sobretudo nessa atividade audiovisual e multimédia que reproduz toda a experiência e sentimentos de Alice. Desta forma, o leitor é levado a ler de uma forma sensível e criativa, recorrendo à sua capacidade imaginativa e interpretativa. Mais do que uma personagem da história, Alice potencia um processo de identificação com o leitor através da interação medial requerida pela própria narrativa, que pode levar o leitor a ser (ou pelo menos a produzir) a própria história (AARSETH, 2005).

Desta forma, a natureza desta textualidade eletrónica não é estática, mas profundamente dinâmica e multimodal, na medida em que gera um movimento contínuo de novos meios que simulam constantemente outros meios, em um contexto que não está à superfície (como o texto impresso) mas atrás daquilo que se vê. Para lá do plástico, metal ou silicone que compõem o *hardware* do computador, do *tablet* ou do telemóvel, neste espaço sem limites físicos, a tecnologia não é mais do que uma máquina de construção e decodificação contínua de várias linguagens que materializam os sentidos da história de Alice, ou seja, é o *software* que dá ao computador a sua funcionalidade. Toda a interface de ícones, botões e aplicações funciona como metáfora, uma construção ao nível da programação, para, assim, tornar *visível e legível* o que na linguagem da máquina são números (zeros e uns)

codificados (algoritmos). Segundo Christin (2010), as noções de “visível” e de “legível” são muito próximas, apesar de serem diferentes. A autora distingue os dois conceitos da seguinte forma:

O “visível” remete para o modo geral e espectacular do “aparecer”. É, antes de mais, uma revelação. [...] O “legível” releva, também ele, da revelação, mas de um modo singular, combinando o visível com um dado que lhe é exterior, na medida em que implica uma forma de comunicação específica: a escrita. Quer surja em sentido literal ou metafórico, o legível implica sempre “o escrito”. Por isso, a revelação que propõe não é da ordem do fascínio ou sequer da surpresa, mas da reminiscência.

Ou seja, o computador funciona como uma máquina de conversão e reminiscência que recria o som, a imagem ou o texto a partir de um código informático necessário para a materialização de uma fisiologia “visível” e – talvez – “legível”. A legibilidade daquilo que é visível torna-se um problema devido exatamente à fluidez e instabilidade das diferentes linguagens que se cruzam no interior de qualquer obra e se exteriorizam agora no discurso hipermídia de IA, por exemplo.

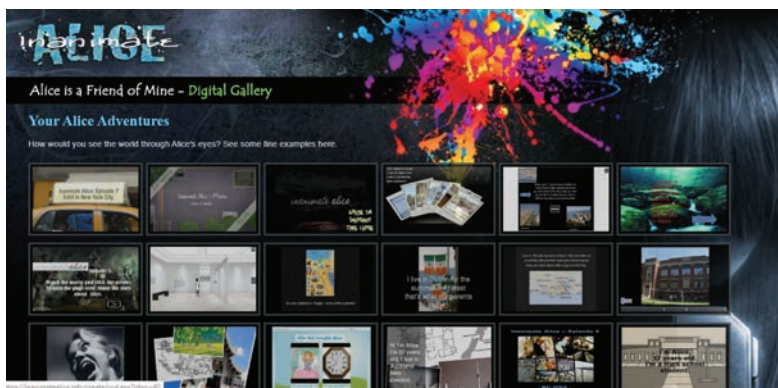
Um texto eletrónico como este resulta de um processo de transformação de algo estático (informação) para algo variável (interpretação). Hayles (2002) defende, por exemplo, que até a transformação de um documento impresso num documento digital acontece como um “acto de interpretação” (HAYLES, 2002, p. 263), pois implica sempre escolhas de diversa natureza.

Se a literatura também é tornar estranho o que é conhecido e fazer-nos olhar para as coisas de maneira diferente; se sempre existiram formas diferentes de escrever e de ler, as singularidades desta nova forma de leitura decorrem, principalmente, do nível elevado de experimentação, a que o leitor é obrigado, devido, por um lado, à hibridez das formas e inovação tecnológica que os artistas trazem para este trabalho de escrita hipermídia e, por outro, a complexidade que toda a experiência imersiva acarreta.

Desse modo, a IA é um projeto ambicioso e transversal: uma história que é não só uma narrativa digital (criada no meio digital e para aí ser lida); uma narrativa interativa (na qual o leitor é participante na própria narrativa); uma narrativa multimídia (na qual se faz uso de fotografia, vídeo, efeitos, música, som, texto,

jogo); um romance¹ (com a história complexa de uma personagem); uma novela (constituída por uma série de episódios, que se vão complexificando à medida que a personagem principal se vai tornando mais velha); mas também um instrumento transmídia didático. A obra tem tido sucesso no meio escolar de língua inglesa através da sua dimensão colaborativa, que acrescenta uma componente de “fan-fiction” ao seu universo, pois os alunos podem participar ativamente, construindo novas histórias adaptadas aos seus países. *Alice’s School Report* da equipa de Jessica Lassetti (2011), por exemplo, é um dos primeiros trabalhos desenvolvidos na Austrália, tal como outros que seguiram a proposta do projeto educativo e transversal de IA – *Create with Alice* –, em que alunos e jovens do mundo inteiro imaginam novas aventuras para Alice.

FIGURA 4 - INANIMATE ALICE, ITEACH



Fonte: Disponível em: <<https://inanimatealice.com/create-with-me/>>. Acesso em:08 set. 2020.

No âmbito desta dimensão mais juvenil de IA, o elemento jogo é parte integrante e determinante da narrativa. Porém, se normalmente no jogo é suposto saber as regras antes de poder jogar, neste trabalho eletrónico, temos, por vezes, de jogar primeiro para conseguir continuar a ler e perceber a história. Não podemos esquecer que a função configurativa e interpretativa deste tipo de jogos narrativos dependerá do utilizador, que será sempre responsável pelo desenrolar

¹ Do ponto de vista do género, “IA” pode considerar-se um romance na medida em que apresenta um exemplo digital de um romance serializado, em que a complexidade narrativa se relaciona com a complexidade medial em cada pormenor escolhido pela produção da história.

A natureza multimídia desta textualidade eletrônica obriga a repensar como se articula a receção do texto e da imagem, do som, do corpo e do movimento, que devido à sua fluidez, dificilmente podemos limitar em determinado conceito de leitura. Na verdade, o conceito de texto-tecido personifica o significado latino original do termo: “textus” como “urdidura”, “encadeamento” ou “entrelaçamento” de vários planos de expressão e conteúdo.

Estes novos textos são o resultado da construção de um código que, em estado latente ou virtual, permite diferentes vias de sentido, ou seja, múltiplas possibilidades operacionalizadas com a leitura. Sabemos que uma obra só pode ser obra aberta, segundo Umberto Eco, no momento da interpretação. Aqui, porém, a abertura é anterior à percepção, pois encontra-se ao nível do próprio objeto literário, no momento em que ele possibilita a abertura a essa interpretação, quando inscreve as potencialidades das virtualizações combinatórias que obrigam a selecionar caminhos perante rotas de navegação e de sentidos diferentes. É assim que se torna pertinente pensar a leitura dentro do âmbito de uma “performance”, ou seja, uma interação explícita com um leque de possibilidades combinatórias materiais.

À partida, o conceito do *novo* deverá ser relativizado, porque os novos meios são normalmente caracterizados por uma síntese de inovações e, por vezes, esquece-se o facto de que, como diz Manovich, *os novos média nunca deixaram de ser uma recodificação de técnicas e estratégias formais modernistas* (como o futurismo, o dadaísmo, Bauhaus) (MANOVICH, 2001). Contudo, desde os livros eletrónicos (livros digitalizados) até as obras hipermídia (como IA), o fator diferenciador vai além da construção multimídia com vídeos, som, imagens. Todos estes aspetos multimodais participam no conceito de uma nova textualidade transmediada.

A principal diferença está, não só, na intervenção obrigatória da tecnologia dos meios como instrumentos de participação, mas também na interferência do “público” ou mesmo na participação das redes sociais (*Twitter e Facebook*) para a construção das narrativas. Na verdade, o facto de IA se encontrar disponível na internet permite, por exemplo, no episódio 4, a autorreferencialidade do texto com o “iStori.es” – uma aplicação que ajuda os amigos de Alice (e nós leitores) a criar novas histórias personalizadas.

FIGURA 6 - INANIMATE ALICE, EPISODE 4: HOMETOWN (HOMETOWN)



Fonte: Disponível em: <www.inanimatealice.com>. Acesso em: 08 set. 2020.

A internet, enquanto plataforma audiovisual e multimídia, permite uma realidade amplificada ou alternativa (jogos/aplicações) que possibilita ao recetor converter-se em um espectador/ator/leitor/jogador/fã/criador. Na plataforma livro impresso, conta-se uma história por *meio* da escrita; no cinema, a história conta-se por *meio* de imagens em movimento e por *meio* do som das vozes, do espaço e da música; a Rede, por sua vez, trouxe outros rasgos determinantes à narrativa, nomeadamente, a capacidade de integrar outras plataformas e meios correspondentes num só espaço e sobretudo uma interatividade com capacidade de implicar uma variedade de níveis de participação ao leitor internauta.

Importa dar atenção à multimodalidade e abertura destas novas formas textuais, principalmente, devido à necessidade de afastar uma análise exclusivamente linguística, em prol de uma análise semiótica; ou seja, passar de uma teoria que explica uma única linguagem (escrita), para uma abordagem que explique o gesto, discurso, imagem, escrita, objetos 3D, cor, música e outros elementos que agora fazem parte deste texto. Esta análise não deverá privilegiar uma abordagem linguística (nem negá-la), mas focar a importância do trabalho desenvolvido por estes processos para alcançar um propósito interpretativo. Assim, a partir daqui a linguagem verbal torna-se uma entre muitas outras linguagens utilizadas para comunicar e que precisam de ser tratadas tanto semioticamente como linguisticamente.

ler sentidos/ler com sentido. Será este depois, também, o denominador comum na construção de um perfil de leitura multimídia.

FIGURA - ELECTRONIC LITERATURE COLLECTION. VOLUME 3



Fonte: Disponível em: <<http://collection.eliterature.org/3/>>. Acesso em: 08 set. 2020.

Agora, IA, ou muitos dos trabalhos de literatura eletrônica disponíveis na *Electronic Literature Organization*, podem ser não só lidos, vistos, ou jogados – mas têm também potencial para serem experimentados, enquanto *performances* de leitura ou processos fenomenológicos inclusivos. Se a transmedialidade deste tipo de trabalhos pressupõe a sua distribuição por múltiplas plataformas, em versões adaptadas das suas narrativas a essas plataformas, a natureza hipermídia e transmídia da literacia necessária para a leitura deste gênero de obras colocamos, sem dúvida, novos desafios. Como experiência narrativa criada no intervalo entre a realidade e a ficção, IA é um exemplo de uma narrativa multimídia, que desencadeia uma experiência de alteridade entre a leitura e a representação a partir do momento em que encontra no espaço transmídia o ambiente perfeito para uma autêntica *performance* de leitura.

REFERÊNCIAS

AARSETH, E. J. *Cibertexto: perspectivas sobre a literatura ergódica*. Tradução Maria Leonor Têlles e José Augusto Mourão. Lisboa: Pedra de Roseta, 2005.

BENJAMIN, W. *Illuminations*. London: Fontana Press, 1992.

CHRISTIN, A-M. Legível/Visível. Dicionário Crítico de Arte, Imagem, Linguagem e Cultura. *Centro de Estudos de Comunicação e Linguagem*, 2010. Disponível em: <<http://www.artecoa.pt/index.php?Language=pt&Page=Saberes&SubPage=ComunicacaoELinguagemLinguagem&Menu2=Legivel>>. Acesso em: 11 jul. 2013.

CICCORICCO, D. *Reading Network Fiction*. Tuscaloosa, AL: University of Alabama Press, 2007.

ELECTRONIC LITERATURE ORGANIZATION. Disponível em: <<http://collection.eliterature.org/3/>>. Acesso em: 22 maio 2020.

FLEMING, L. Expanding Learning Opportunities with Transmedia Practices: Inanimate Alice as an Exemplar. *Journal of Media Literacy Education*, v. 5, n. 2, p. 370-377, 2013. Disponível em: <<http://digitalcommons.uri.edu/jmle/vol5/iss2/3>>. Acesso em: 22 maio 2020.

HAYLES, N. K. *Electronic Literature: New Horizons for the Literary*. Notre Dame, Indiana: University of Notre Dame, 2008.

HAYLES, N. K. *Writing Machines*. Cambridge, Mass: MIT Press, 2002.

JENKINS, H. "Transmedia 202: Further Reflections." *The Official Weblog of Henry Jenkins*, 2011. Disponível em: <http://henryjenkins.org/2011/08/defining_transmedia_further_re.html>. Acesso em: 22 maio 2020.

LACCETTI, J. Teacher Training. *Inanimate Alice: School Report 3*, p. 6, 2011. Disponível em: <http://issuu.com/inanimatealice/docs/school_report_3>. Acesso em: 22 maio 2020.

MANOVICH, L. *The Language of New Media*. Cambridge: MIT Press, 2001.

PULLINGER, K. *Inanimate Alice*. 2005. Disponível em: <<http://www.inanimatealice.com/>>. Acesso em: 22 maio 2020.

PULLINGER, K. Live webchat: Kate Pullinger. *The Guardian*, 2011. Disponível em: <<http://www.theguardian.com/books/booksblog/2011/jun/29/live-webchat-kate-pullinger>>. Acesso em: 22 maio 2020.

ZUERN, J. D. Minds, Messages, and the Moral Imagination in the Media of Fiction: Inanimate Alice between Cognitive and Rhetorical Paradigms. *Journal Studies Neophilologica*, v. 87, p. 8-28, 2015.

A fotografia digital na transposição de documentos manuscritos históricos

Jorge Viana Santos
Giovane Santos Brito

INTRODUÇÃO

O mundo digital “[...] é o universo dos equipamentos que utilizam a linguagem digital para a captação, armazenamento, processamento e transmissão de informações” (CESAR; PIOVAN, 2003, p. 156). Nesse contexto, estão as construções e os tratamentos de imagens técnicas geradas por aparelhos fotográficos, tais como câmeras e escâneres que, em sua grande maioria, se constituem enquanto imagens fotográficas codificadas digitalmente.

Visualizáveis através de telas, as imagens fotográficas, assim codificadas, constituem um dos meios mais democráticos de veiculação e transmissão de informações, desde a invenção da fotografia¹ em 1829 por Joseph Niépce e Louis Jacques Daguerre (CURT E ALEX ASSOCIADOS LABORATÓRIO CINEMATOGRAFICO LTDA, 1981) e da câmera fotográfica de 35mm². O caráter popular da fotografia fez com que uma quantidade cada vez maior de pessoas explorasse o potencial da técnica fotográfica.

Por recorte metodológico, interessa-nos aqui explorar o potencial científico da fotografia, uma vez que

Para os estudiosos da história social, da história das mentalidades e dos diferentes gêneros de história, assim como para **os pesquisadores de outros ramos do conhecimento**, são as imagens documentos insubstituíveis cujo **potencial** deve ser explorado. [...] (KOSSOY, 1989, p. 20, grifo nosso).

¹ Sobre a história da invenção simultânea tríplice da Fotografia por Louis Jacques Daguerre, Joseph Niépce e Hercules Florence, ver Kossoy (1989).

² Para mais detalhes a respeito da invenção da câmera fotográfica de 35mm, consultar Benjamin (1985), Newhall (2002) e Frizot (1989).

Para tanto, abordamos neste trabalho algumas noções referentes à fotografia no tocante ao processo de captura de imagens e ao resultado final desse processo. Assim, destacamos o uso científico da técnica fotográfica, fundamentado no método de transposição de documentos manuscritos históricos (SANTOS; BRITO, 2014) para o meio digital integrado ao método cientificamente controlado de construção de *corpora* anotados: o método Lapelinc³.

FOTOGRAFIA: PROCESSO E RESULTADO

Desde a sua origem, no início do século XIX, a fotografia tem surpreendido muitas pessoas graças à sua capacidade de representar, de forma fidedigna, os objetos que estão diante da câmera, “[...] já que a informação visual do fato representado na imagem fotográfica nunca é posta em dúvida” (KOSSOY, 1989, p. 69). Esse caráter fidedigno de representação da Fotografia se dá pela natureza de sua linguagem. Assim, compreender o que constitui a fotografia enquanto tal é entender o que faz com que sua técnica garanta a fidedignidade aos objetos que ela representa e, por conseguinte, as características da linguagem fotográfica.⁴

De acordo com Stroebel e Zakia (1993), a Fotografia pode ser compreendida de duas maneiras e, para tanto, apresenta dois vocábulos distintos: *photography* e *picture*. O primeiro, refere-se ao fato de que “A palavra fotografia, derivada da raiz grega, significando escrever com a luz, foi posteriormente substituída por Niépce pela palavra Heliografia (escrever com o sol) e por Talbot pela frase desenho fotogênico” (p. 598, tradução nossa)⁵. O segundo, diz respeito “A representação visual de um objeto ou cena, a partir de uma fotografia, pintura ou desenho” (p. 634, tradução nossa)⁶.

³ Para detalhes sobre o método Lapelinc em sua etapa de transposição, ver o capítulo de Cristiane Namiuti e Jorge Viana Santos – *Novos desafios para antigas fontes: a experiência DOViC na nova Linguística Histórica* – neste livro.

⁴ Para outras visões acerca desse tema, ver: Shaeffer (1996) e Machado (1984).

⁵ “The word Photography was derived from the greek roots meaning to write with light, and it immediately replaced Niépce’s word, heliography (sun writing), and Talbot’s phrase, photogenic drawing”

⁶ “A visual representation of an object or scene, such as a photograph, painting or drawing”.

Percebe-se, assim, que o vocábulo *photography* em inglês descreve um percurso de técnicas fotográficas que abrange, por um lado, tanto a linguagem e o seu respectivo processo de captura de imagens por um aparelho com, no mínimo, uma superfície sensível à luz, a exemplo da câmera, e, por outro, o termo *picture* que indica o resultado final dessa captura, a imagem fotográfica⁷. Para fins de esclarecimento, utilizamos ao longo deste trabalho “Fotografia” com inicial em maiúscula, para nos referir ao processo, e “fotografia” com inicial em minúscula, para nos referir ao resultado, tendo em vista que em português uma única palavra é utilizada para exprimir os dois conceitos.

Assim, em relação à fotografia enquanto resultado, Flusser (2002, p. 7) afirma que “[...] As imagens são, portanto, resultado do esforço de abstrair-se duas das quatro dimensões do espaço-tempo para que se conservem apenas as dimensões do plano”. Percebe-se assim que, enquanto resultado, a fotografia sugere a transformação de uma realidade tridimensional para outra bidimensional. A partir disso, percebemos a necessidade de estabelecer uma correlação entre a realidade bidimensional da fotografia e a realidade tridimensional dos manuscritos históricos, demonstrando, sobremaneira, a relevância de que a fotografia digital, para estudos no âmbito da Linguística e outras áreas, se faça de modo cientificamente controlado.

No que tange à Fotografia enquanto processo, muitos elementos entram em cena no momento da sua realização. Machado (1984, p. 39, grifo nosso), ao tratar de alguns desses elementos, afirma que

Se não existir a câmera escura, a lente com seu poder organizador dos raios luminosos, um diafragma rigorosamente aberto como manda a análise da luz operada pelo fotômetro, um obturador com velocidade compatível com a abertura do diafragma e a sensibilidade da película, se não houver ainda uma fonte de luz natural ou artificial modelando o referente e um operador regendo tudo isso, também não haverá **fotografia** [...]

Conforme o autor, durante o processo da Fotografia, um conjunto de elementos próprios da linguagem fotográfica corroboram para a formação de uma imagem fotográfica enquanto resultado, ou seja, o produto dessa operação. Nesse

⁷ Santos (2013) cita pelo menos a existência de quatro imagens fotográficas distintas capturadas pelo processo óptico da câmera fotográfica com lente.

sentido, Kossoy (1989, p. 23) estabelece os três elementos essenciais da Fotografia: o assunto, o fotógrafo e a tecnologia, salientando que

[...] são estes os três elementos constitutivos que lhe deram origem através de um *processo*, de um ciclo que se completou no momento em que o objeto teve sua imagem cristalizada na bidimensão do material sensível, num preciso e definido *espaço e tempo*.

Nesse sentido, para o mesmo autor,

O produto final, a **fotografia**, é portanto resultante da ação do homem, o fotógrafo, que em um determinado espaço e tempo optou por um assunto em especial e que, para seu devido registro, empregou os recursos oferecidos pela tecnologia em cada época. (KOSSOY, 1989, p. 23, grifo nosso)

A imagem fotográfica que se alcança através da conjunção dos elementos da Fotografia citados por Kossoy se constituiu na utilização de uma linguagem que lhe é própria, a linguagem fotográfica, e dela falaremos no tópico a seguir.

FOTOGRAFIA: LINGUAGEM

Santos (2010) considera que os elementos fundamentais da Fotografia são ao mesmo tempo técnicos, teóricos e práticos. O autor, em sua exposição, aborda alguns dos seus principais elementos, dando ênfase para os quatro seguintes: o enquadramento, a composição, a focalização e a iluminação.

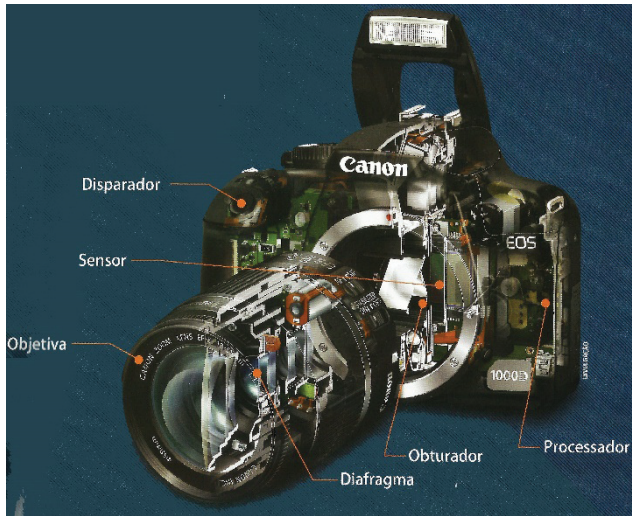
Em relação ao enquadramento, o autor esclarece que “Diz respeito a quanto e o que aparecerá na imagem; envolve a definição do quadro, podendo ser aberto ou fechado, o qual se relaciona diretamente com a proximidade do objeto e com o uso de tal ou qual lente” (p. 13). Nesse sentido,

[...] A **lente** objetiva de uma máquina fotográfica é uma lente convergente que, de um objeto real, conjuga uma imagem real e invertida. Essa imagem deverá recair sobre o filme, para que possamos obter uma fotografia nítida. A máquina fotográfica possui dispositivo que altera a distância focal da lente ao filme para que, a cada exposição do objeto, a imagem possa se projetar sobre o filme. A lente objetiva em geral, é a lente resultante da associação de lentes; o diafragma

da máquina fotográfica é o dispositivo que regula a quantidade de luz, nela incidente (TASHIBANA; FERREIRA; ARRUDA, 1996, p. 176, grifo nosso).

Tashibana, Ferreira e Arruda (1996) elencam alguns dos componentes da linguagem fotográfica que entram em conjunto na constituição de uma imagem pela câmera fotográfica com lente, fato que pode se observar na Figura 1.

FIGURA 1 - CÂMERA FOTOGRÁFICA COM LENTE (OBJETIVA)



Fonte: Melo e Alméri (2013, p. 23).

Para complementar, Aumont (1993, p. 153, grifo nosso), ao comentar os elementos da Fotografia, postula que

[...] O **enquadramento** é pois a atividade da moldura, sua mobilidade potencial, o deslize interminável da janela à qual a moldura equivale em todos os modos da imagem representativa baseados numa referência, primeira ou última, a um olho genérico, a um olhar, ainda que perfeitamente anônimo e desencarnado, cuja imagem é o traço.

No que tange à composição, Santos (2010, p. 19, grifo nosso) assinala que “Refere-se a **como** o *objeto* será disposto no quadro, isto é, como aparecerá na fotografia (*photograph, picture*), relacionando-se com outros objetos e com o

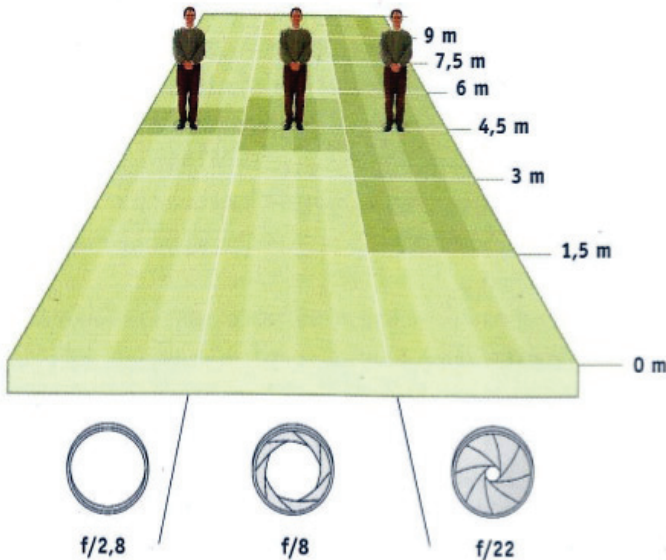
fundo.” O autor destaca ainda alguns elementos de composição, tais como o ponto de vista/ ângulos e as perspectivas.

Quanto à focalização, o autor considera que “Refere-se a **que plano(s) aparecerá(ão) e/ou será(ão) destacado(s)**. Pressupõe um fenômeno óptico fundamental: a profundidade de campo, também responsável pela textura” (p. 20, grifo nosso).Tendo em vista que

[...] Cada vez que focamos um ponto de uma determinada cena, temos uma zona de foco nítido, tanto na frente, como atrás desse ponto. Se dividirmos esta área em três partes iguais e partirmos do ponto onde colocamos o foco na câmera, teremos foco em 1/3 antes do ponto focado da cena e 2/3 na parte posterior da área. Essa área de foco nítido é conhecida como **profundidade de campo**. Esta **profundidade de campo** é determinada pela abertura da objetiva (f), pela distância focal da objetiva (50mm, 135mm) e a distância do objeto à objetiva (em centímetros ou metros) (CESAR; PIOVAN, 2003, p. 42, grifo nosso).

O fenômeno da profundidade de campo pode ser verificado na Figura 2.

FIGURA 2 - PROFUNDIDADE DE CAMPO: ABERTURAS (F) E DISTÂNCIA FOCAL



Fonte: Google Imagens.

Por fim, no tocante à iluminação, vimos que a luz é a condição essencial da Fotografia e, por definição,

Luz é energia que se propaga através de ondas eletromagnéticas, cujas frequências sensibilizam nossos olhos. [...] Para enxergarmos um objeto qualquer é necessário que a luz proveniente dele atinja nossos olhos. Este objeto é uma fonte de luz. As fontes de luz podem ser: **Primária** é aquela que produz a luz que emite. As estrelas, e em particular o Sol, são exemplos de fontes primárias, pois a luz emitida é produzida por reações nucleares ocorridas na própria estrela [...] (TASHIBANA; FERREIRA; ARRUDA, 1996, p. 145-146, grifo nosso).

A luz do sol é uma das luzes mais utilizadas na Fotografia, servindo também como parâmetro para outros tipos de iluminação, como o *flash*, as luzes frias (artificiais), luzes incandescentes (lâmpadas comuns), fluorescentes ou de tungstênio.

Desse modo, percebe-se que o conhecimento dos aspectos técnicos (enquadramento, composição, focalização e iluminação) vistos nesta seção, constituem a linguagem fotográfica, e que o domínio das técnicas que a constituem pressupõe características indispensáveis para o controle da Fotografia enquanto processo e resultado, pois tal técnica pode ser utilizada para vários fins. No nosso caso, cabe ressaltar a aplicação científica da Fotografia no tocante ao caráter fidedigno de representação que lhe é característico, principalmente no que tange ao processo de transposição de documentos manuscritos históricos, conforme apresentado por Santos e Brito (2014).

Assim,

Seja em função de um desejo individual de expressão de seu autor, seja de comissionamentos específicos que visam uma determinada *aplicação* (**científica**, comercial, educacional, policial, jornalística, etc.) existe sempre uma *motivação* interior ou exterior, pessoal ou profissional, para a *criação* de uma fotografia e aí reside a primeira opção de um fotógrafo, quando este seleciona o assunto em função de uma determinada *finalidade/intencionalidade*. Esta motivação influirá decisivamente na concepção e construção da imagem final (KOSSOY, 1999, p. 27, grifo nosso).

Kossoy (1999) estabelece que a Fotografia pode, por um lado, ter uma função subjetiva e, por outro, uma função objetiva. Na primeira função, enquanto desejo

individual de expressão, a Fotografia pode ser vista como arte, como expressão subjetiva. Na segunda função, destacam-se algumas aplicações, tais como científica, comercial, educacional, policial, jornalística etc. Em ambas as funções, o fotógrafo é motivado interior ou exteriormente, pessoal ou profissionalmente para criar uma fotografia, tendo em vista a sua finalidade. Para se alcançar um determinado fim, o fotógrafo faz determinadas escolhas, utiliza-se deste ou daquele equipamento, e no caso da era digital, deste ou daquele *hardware* e/ou *software*, caracterizando, assim, seu *modus operandi* de fotografar para alcançar seus objetivos.

Tendo em vista o que foi exposto a respeito da linguagem fotográfica, o objetivo geral deste trabalho é demonstrar a aplicação científica da Fotografia digital, principalmente no que tange ao processo de transposição de documentos manuscritos históricos e sua relevância para a Linguística. Santos e Brito (2014) descrevem etapas de um método que vem sendo desenvolvido e aplicado na transposição de documentos manuscritos históricos para o meio digital, através da Fotografia cientificamente controlada: o Método Lapelinc. Tal método utiliza uma Fotografia que prioriza a informação ao invés da estética, levando em conta alguns dos aspectos abordados neste tópico, mais precisamente os aspectos técnicos de enquadramento, composição, focalização e iluminação, que permitem aos pesquisadores detectar um maior número de informações visuais que só podem ser colhidas através da Fotografia enquanto linguagem técnica.

Nesse sentido, ao destacar alguns dos aspectos da Fotografia enquanto linguagem técnica e sua relevância para a Linguística, faz-se necessária a exposição pormenorizada da aplicação desta modalidade de Fotografia dentro do método que está sendo desenvolvido no Lapelinc.

FOTOGRAFIA CIENTÍFICA NO MÉTODO DESENVOLVIDO NO LAPELINC

A Fotografia cientificamente controlada e aplicada no processo de transposição de documentos manuscritos históricos originais em papel para o formato digital, com vistas a integrar *corpora* eletrônicos anotados, apresenta três propriedades:

- a) Pressupõe domínio da Fotografia (*photography*) enquanto linguagem e enquanto técnica;
- b) Necessita de equipamento e aparato técnico auxiliar específicos, a exemplo da Mesa Cartesiana;
- c) Visa à construção, sobretudo, de *corpora* manuscritos para uso científico: Linguística, História, Direito, Memória, dentre outras.

Tais características aliadas ao controle de (meta)informações previstas no método Lapelinc, como apontam Santos e Namiuti (2018), permitem transformar a Fotografia não apenas em um simples meio de reprodução de um documento ou em uma fotografia pragmática, que serve apenas a uma pesquisa, sem necessariamente ter compromisso de futuro, mas sim praticá-la como método científico de reprodução digital.

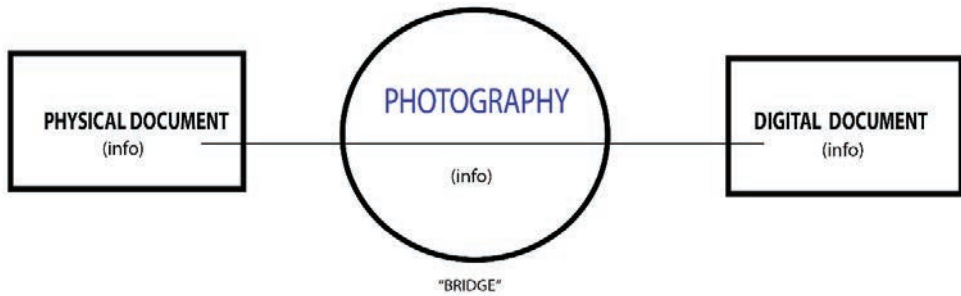
Em relação à captura fotográfica em si, optamos pela captura digital, com câmera e lente, e dentro deste processo, pode-se optar pela utilização de câmeras digitais ou escâneres. Neste sentido,

[...] o Método Lapelinc se propõe como método que visa à digitalização por **Fotografia** e não por escaneamento. Isto devido à natureza do objeto que compõe o acervo com que lidamos: livros notariais manuscritos que, dada a sua idade/datação (Século XIX) e grande tempo de manuseio (muitos ainda estão em uso), apresentam-se hoje em estado de extrema fragilidade, sendo impraticável a sua digitalização por *scanner*. O processo de escaneamento por mais moderno que seja é fixo e exige que o objeto/documento se adeque a ele em termos de tamanho, iluminação, etc. Diferentemente, a **Fotografia**, enquanto **linguagem** e enquanto **técnica**, dada à sua flexibilidade, devida a quase infinita possibilidade de variação de parâmetros, se adequa à realidade do objeto e não o inverso(SANTOS; BRITO, 2014, p. 8, grifo nosso).

A máquina digital captura a imagem via CCD⁸ (Sensor Digital) e a transfere para o cartão de memória ou para um computador. O processo digital, enquanto transposição de um documento físico para o meio digital, pode ser melhor compreendido a partir do que informa e demonstra o Gráfico 1:

⁸ De acordo com Melo e Alméri (2013, p. 18), “as câmeras digitais não utilizam filme, elas captura a luz da cena a ser fotografada por meio de células fotossensíveis chamadas *Charged Coupled Device* (CCD)”.

GRÁFICO 1 - FOTOGRAFIA PRATICADA COM MÉTODO CIENTÍFICO DE REPRODUÇÃO DIGITAL: A PONTE ENTRE DOCUMENTO FÍSICO E DOCUMENTO DIGITAL



Fonte: Namiuti *et al.* (2013, p. 15).

Assim,

[...] a Fotografia funciona como uma espécie de *ponte* entre DF (documento físico) e o DD (documento digital). Mas para isso, defendemos que é preciso que ela registre, na própria imagem, dados/informações que façam com que a imagem gerada não perca o vínculo com o documento que lhe deu origem (NAMIUTI *et al.* 2013, p. 15).

A Figura 3 ilustra esse processo:

FIGURA 3 - A FOTOGRAFIA CIENTIFICAMENTE CONTROLADA DO MÉTODO LAPELINC



Fonte: *Corpus DOViC – Livro de Notas C11-E3.*

À esquerda, um instrumento desenvolvido para o método aplicado, “Mesa Cartesiana” (leiaute), com o DDI referente à folha 182 do Livro de Notas C11-E3 exemplificando metainformações da fotografia cientificamente controlada: (1) cartão cinza; (2, 3) escalas de tom e cores; (4, 5 e 6) instrumentos de medição calibrados; (7) informações catalográficas; (8) paginação; (9) sequenciação; (10) observações filológicas; e ao centro e à direita exemplos de possibilidade inerente ao suporte digital: possibilidade de reprodução para fins específicos, por exemplo, leitura e visão fractal fidedigna ao documento físico.

Para que possa operar de forma produtora, o método de Fotografia que vem sendo desenvolvido e praticado no Lapelinc tem como base a passagem do documento físico para o documento digital, e leva em conta alguns aspectos que dizem respeito ao processo denominado “transposição” no método Lapelinc. Propõe-se, assim, o uso da Fotografia enquanto método de transposição do texto em papel para o meio digital. Nas práticas do Laboratório, considera-se o uso da Fotografia cientificamente controlada enquanto um método de transposição do texto em papel para o meio digital, visando a construção do *Corpus* DOViC que tem como objetivo contribuir para recuperar a memória histórica e linguística da região e permitir à Linguística e outras ciências acessar imageticamente, de modo confiável, um objeto físico (o documento), muitas vezes, indisponível ao pesquisador (NAMIUTI; SANTOS; LEITE, 2011).

CONSIDERAÇÕES FINAIS

A compreensão do que constitui a Fotografia enquanto processo (*photography*) e enquanto resultado (*Picture*) como vimos, difere da compreensão da mera reprodução de objetos através de uma fotografia pragmática. A compreensão de que a Fotografia pode ser praticada cientificamente é condição primordial para o uso e desenvolvimento das técnicas e equipamentos que envolvem a transposição de documentos manuscritos históricos para o meio digital. A Fotografia científica, tal qual praticada no Lapelinc, tendo em vista os aspectos destacados neste trabalho, principalmente no que tange à sua linguagem técnica, depende dos conhecimentos técnicos, teóricos e práticos para se tornar produtora, e demonstra assim a sua relevância para estudos no âmbito da Linguística.

REFERÊNCIAS

- AUMONT, J. *A imagem*. Campinas: Papirus, 1993.
- BENJAMIN, W. Pequena história da fotografia. In: _____. *Magia, arte e técnica: ensaios sobre a literatura e história da cultura*. São Paulo: Brasiliense, 1985. p. 91-108.
- CESAR, N.; PIOVAN, M. *Making of: revelações sobre o dia-a-dia da fotografia*. São Paulo: Futura, 2003.
- CURT E ALEX ASSOCIADOS LABORATÓRIO CINEMATOGRAFICO LTDA. *O laboratório por dentro: curso completo de Fotografia*. Rio de Janeiro: Rio Gráfica e Editora, 1981.
- FLUSSER, V. *Filosofia da caixa preta*. Rio de Janeiro: Relume Dumará, 2002.
- FRIZOT, M. *Histoire de Voir*. Paris: Photo Poche, 1989.
- KOSSOY, B. *Fotografia e história*. São Paulo: Editora Ática, 1989.
- KOSSOY, B. *Realidades e ficções na trama fotográfica*. Cotia: Ateliê Editorial, 1999.
- MACHADO, A. *A ilusão especular: introdução à fotografia*. São Paulo: Brasiliense, 1984.
- MELO, E.; ALMÉRI, K. (Org.). *Fotografia digital: uma arte sem mistérios*. São Paulo: Escala, 2013.
- NAMIUTI, C.; SANTOS, J. V.; COSTA, A. S.; FARIAS, I. S. Computação e Linguística: importante diálogo para pesquisas e preservação da memória nos novos meios das antigas fontes. *RBBA: Diálogo entre as ciências*, v. 2, n. 1, p. 9-34, 2013.
- NAMIUTI, C.; SANTOS, J. V.; LEITE, C. M. B. Propostas e desafios dos novos meios das antigas fontes: a preservação da memória pela linguística de corpus. In: COLÓQUIO NACIONAL, 10.; COLÓQUIO INTERNACIONAL DO MUSEU PEDAGÓGICO, 2., 2011, Vitória da Conquista. *Anais...* Vitória da Conquista: UESB, 2011. p. 1-11.
- NAMIUTI, C.; SILVA, A. C. Reflexões sobre anotação sintática e ferramentas de busca: uso da linguagem XML para anotação sintática no corpus digital DOViC. *Letras & Letras*, v. 30, n. 2, p. 82-103, 2014.
- NEWHALL, B. *Historia de la Fotografía*. Barcelona: Gustavo Gili, 2002.
- SANTOS, J. V. Técnicas de transporte do texto manuscrito para o meio digital. In: OFICINA DE LINGUÍSTICA DE CORPUS DA BAHIA, 2010, Feira de Santana. Feira de Santana: UEFS; UESB; UFBA, 2010. p. 15-17.

SANTOS, J. V. *Um método de Fotografia técnica documental para formação de corpora digitais de documentos históricos manuscritos*. Vitória da Conquista: Universidade Estadual do Sudoeste da Bahia, 2013. (Curso).

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

SANTOS, J. V.; NAMIUTI, C. De manuscritos históricos a corpora anotados: do Documento Físico (DF) ao Documento Digital Imagem (DDI). In: BARREIROS, P. N.; ALMEIDA, I. S.; SANTOS, R. B. (Org.). *Filologia e Humanidades Digitais*. Feira de Santana: Editora UEFS, 2018. p. 120-145.

SANTOS, J. V.; NAMIUTI, C. *DOVIC (Documentos Oitocentistas de Vitória da Conquista e região)*. *Corpus* Eletrônico. Vitória da Conquista: Universidade Estadual do Sudoeste da Bahia, 2014.

SHAEFFER, J-M. *A imagem precária: sobre o dispositivo fotográfico*. Campinas: Papirus, 1996.

STROEBEL, L.; ZAKIA, R.D. *The Focal Encyclopedia of Photography*. Boston: Focal Press, 1993.

TASHIBANA, A. T.; FERREIRA, G. M.; ARRUDA, M. *Novo manual nova cultural: Física*. São Paulo: Nova Cultural, 1996.

Métodos não supervisionados de *clustering* para a análise de textos literários: a seleção de materiais sobre grandes volumes de *corpus*¹

María Luisa Fernández Rodríguez
Roberto Samartim

A análise da totalidade de materiais através dos quais é possível atingir um dado objeto de estudo contribui para a rigurosidade científica, ao permitir atender todo o universo dos possíveis. Contudo, para além de eventuais limitações no acesso ao *corpus*, o volume excessivo de materiais coloca reptos em relação ao seu manuseamento e análise, que nem sempre procedimentos manuais ou uma abordagem a olho nu permite superar. Neste sentido, o recurso a ferramentas ou tecnologias específicas pode contribuir para superar estas dificuldades. Assim acontece com as técnicas de *clustering* possibilitadas por *software* como os utilizados neste trabalho que, para além de precisarem de *corpora* volumosos para funcionar, possibilitam a redução de um universo volumoso a uma população mais restrita, representativa e significativa dos elementos em jogo, e permitem escolher aquelas partes do *corpus* (centroídes) com uma alta fiabilidade de que o dito para a mostra selecionada pode ser dito também para o conjunto da coleção. O resultado desta redução poderá ser usada para, por exemplo, realizar explorações contrastivas com outras tipologias de *corpus* (no nosso caso, de inquéritos) ou para validar hipóteses (como a opacidade da cidade ou da cultura local nos produtos e nas práticas ligadas aos Caminhos de Santiago), objetivos ambos presentes no projeto de investigação em que se integra o contributo (TORRES FEIJÓ, 2011).

Este projeto pretende analisar as imagens, os usos e os consumos relacionados com Compostela e os Caminhos de Santiago presentes quer em produtos culturais

¹ O presente trabalho é resultado do projeto de investigação *Discursos, imagens e práticas culturais sobre Santiago de Compostela como meta dos Caminhos de Santiago*, financiado pelo Ministério de Economía y Competitividad do Governo da Espanha entre 2012 e 2015 [FFI2012-35521]. Disponível em: <<https://redegalabra.org/impactos-caminho-comunidade-local-santiago-compostela/>>. Acesso em: 25 mar. 2020. Este texto utiliza soluções linguísticas características do galego-português da Galiza.

quer em discursos e práticas de visitantes. No âmbito deste projeto – para além de serem realizados inquéritos a visitantes, comerciantes e outros agentes locais – foram registados em um catalogador construído ao efeito (SAMARTIM, 2015) 546 livros, 211 sites e 90 produtos audiovisuais sobre as rotas jacobeanas ou a cidade compostelana produzidos entre 2008 e 2014 na Galiza, a Espanha, Portugal e o Brasil. Dentre todos estes materiais selecionámos para este trabalho os 19 textos narrativos (18 romances e uma novela) publicados na Espanha em castelhano no ano 2010 a) porque esta é a língua e o ano (único Ano Santo Jacobeu do período) onde se concentra uma maior produtividade de livros nos quatro espaços culturais, b) porque a narrativa é por sua vez o género também mais produtivo, e c) porque uma das macronarrativas detetadas no âmbito do projeto é também um romance (*O Diário de um mago* de Paulo Coelho; 1988). Nesse romance do escritor brasileiro, nos discursos do Papa João Paulo II durante as suas visitas a Compostela (1982 e 1989) e nas declarações de Santiago de Compostela e o Caminho de Santiago como património da Humanidade pela UNESCO (1995/ 1993) foram detetados os elementos repertoriais mais importantes elaborados na contemporaneidade sobre os Caminhos de Santiago (TORRES FEIJÓ, 2011).

Essa coleção de 2010 está composta por um número de textos e um volume de informação textual relativamente grande, pelo qual possui umas características favoráveis para a exploração de técnicas de agrupamento e categorização de conteúdos; isto é, para testar meios informáticos capazes de identificar com rapidez e precisão padrões de relações intratextuais não (ou dificilmente) perceptíveis através de uma leitura humana e agrupar os resultados. Precisamente, este teste é o objetivo do presente contributo: operar sobre os textos de 2010 para os agrupar e reduzir assim, de maneira não-supervisionada, a um conjunto significativo e representativo do conjunto, utilizando para isso métodos e ferramentas estatístico-computacionais próprios da técnica de análise de *cluster*. Dentro desta análise testaremos dois procedimentos: a análise de frequências e a medida de distância de compressão. Para selecionarmos entre os vários *software* que operam com estes procedimentos os programas informáticos utilizados acompanhamos os seguintes critérios: a) que o *software* esteja recomendado por catálogos *on-line* especializados em Humanidades Digitais (como TAPOR²; ou o *Laboratorio de Innovación en Humanidades Digitales*

² Disponível em: <<http://www.tapor.ca/>>. Acesso em: 25 mar. 2020.

da UNED³); b) que fosse criado ou já testado com textos literários; e c) que tenha uma interface relativamente intuitiva e com suporte técnico ou de manuais. De acordo com estes critérios básicos foram escolhidos para realizar a análise de frequências o programa *Lexos* e o pacote *Stylo*, fortes respetivamente em métodos de tipo quantitativo e qualitativo, e o programa *CompLearn* para o procedimento de medida de distância normalizada por compressão.

Poupamos a descrição das ferramentas selecionadas e as explicações procedimentais relativas tanto ao pré-processamento dos materiais como ao próprio uso do *software*, e apresentamos neste trabalho apenas as fortalezas e debilidades de cada um destes procedimentos em função do nosso objetivo e os resultados da nossa análise (mostramos apenas os resultados de operar com os valores que se verificaram ótimos após o processo de ensaio-erro acompanhado).

ANÁLISE DE FREQUÊNCIAS: PROGRAMAS *LEXOS* E *STYLO*

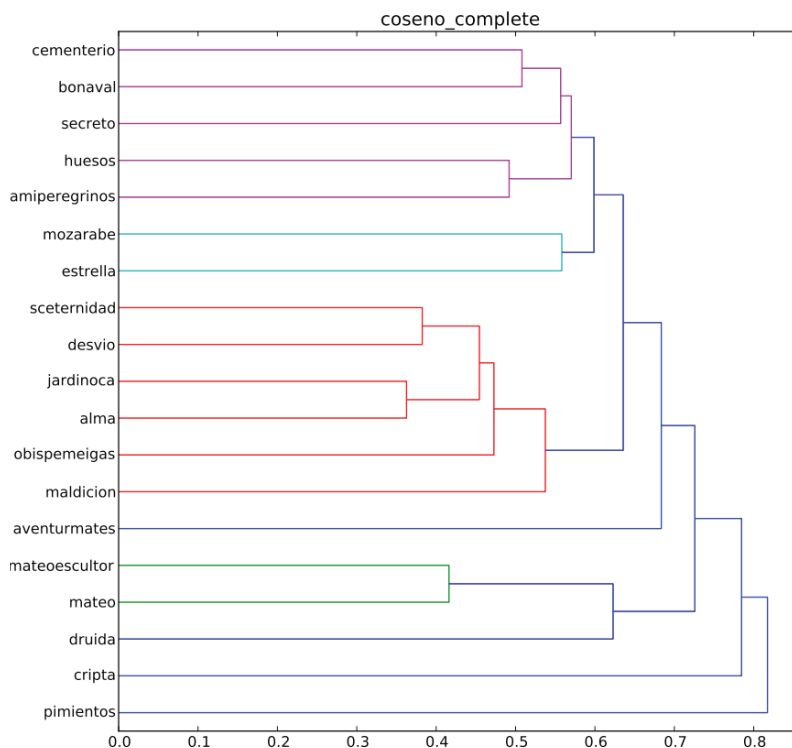
O primeiro *software* escolhido para agrupar a coleção de textos narrativos em espanhol de 2010 através da análise de frequências é o programa *Lexos*⁴. Utilizamos na nossa análise o algoritmo hierárquico aglomerativo já que não exige à partida a indicação de número de grupos, pelo qual supõe menos intervenção e condicionalismo humano, e medimos a similitude entre os textos com a fórmula TF/IDF, porque é uma das medidas de ponderação frequencial mais utilizadas em técnicas de recuperação de informação para atribuir os pesos das palavras (TOMÁS DÍAZ, 2009). Depois de operarmos com o conjunto das métricas e métodos disponíveis no programa, determinamos que *Canberra* e *Coseno*, combinadas com os métodos de distância *complete* e *average*, permitem obter os grupos mais definidos. Otimizamos a qualidade dos resultados modificando a prova de normalização de frequências desde TF/IDF por *proportional counts*, derivando-se assim dendogramas que oferecem maior robustez na definição de grupos, constroem mais subgrupos e conformam ramos a níveis mais baixos. Com a mudança neste parâmetro, as métricas *Canberra* e *Coseno* com o método *complete* definiram-se como as mais produtivas, especialmente a segunda, com dendogramas

³ Disponível em: <<http://linhd.uned.es/>>. Acesso em: 25 mar. 2020.

⁴ Disponível em: <<http://wheatoncollege.edu/lexomics/tools/>>. Acesso em: 25 mar. 2020.

que mostram relações compactas e agrupamentos tendentes a se localizarem entre 0 e 0,5.

FIGURA 1 - DENDOGRAMA DE ANÁLISE DE *CLUSTER* HIERÁRQUICO AGLOMERATIVO (PROGRAMA *LEXOS*. MEDIDA DE DISTÂNCIA *COSENO*, PONDERADO POR *PROPORCIONAL COUNTS*)



Fonte: Elaboração própria, 2015.

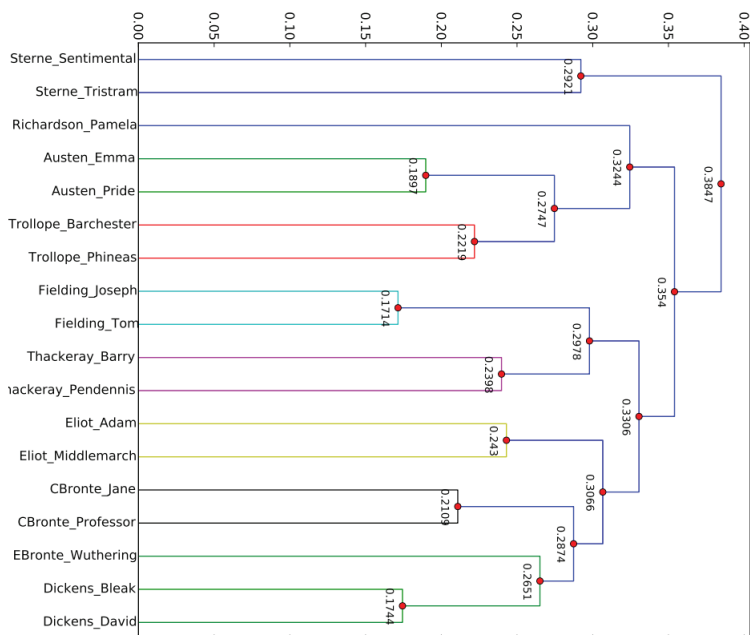
Das várias análises realizadas derivam-se duas conclusões básicas. A primeira é que as medidas de distância estándar (nomeadamente *Euclidean* e *Manhattan*) não funcionam bem com o nosso *corpus*, porque apenas podemos derivar uma abordagem ajustada com a aplicação de duas das métricas disponíveis combinadas com um único método dos quatro existentes; estas métricas são, precisamente, as mais utilizadas no âmbito da linguística (*Coseno* e *Canberra*), o que pode ser devido tanto à própria natureza literária dos textos como a que este conjunto em

concreto apresenta singularidades que será preciso detetar com uma análise mais apropriada. A segunda conclusão é que, em geral, há uma alta coincidência na formação de quatro/ cinco macroagrupamentos; esta mínima variação origina-se em função de por onde seja estabelecido o corte e de se são contados como grupo os textos com um comportamento individualizado. Também parece haver coincidência na extração de um grupo de textos com tendência a se agruparem em posição central (*Alma, Obisposmeigas, Maldición, Jardínoca, Estrella, Mozárabe*), outro conformado por aqueles que mostram um funcionamento intermedial (*Mateo, Secreto, CamiPeregrinos, Huesos*) e um terceiro agrupamento formado pelos textos tendentes a se individualizarem e apresentam resistência a serem classificados (nomeadamente *Pimientos*).

Ao lado destas conclusões devem ser indicados igualmente alguns problemas que não permitem garantir a fiabilidade dos agrupamentos derivados desta análise de *cluster*, assim como também não contribuem para atingir o objetivo de selecionar uma amostra representativa e significativa do *corpus* de narrativa em espanhol de 2010 relativo à Compostela ou aos Caminhos de Santiago:

1) Ainda levando em conta os resultados das agrupações mais compactas derivadas, as conexões tendem a se fazer em valores altos, há muita ramificação, o que pode significar que a coleção é muito diversa ou que a visualização é pouco precisa. Parece oportuno, então, realizar um estudo de controlo, para detetar se estamos perante um problema do *software* ou uma característica da coleção. Escolhemos então como *corpus* de controlo um conjunto de romances da literatura inglesa que já foram utilizados para testar a identificação automática de autoria a partir do vocabulário presente neles. Como podemos observar no dendograma da Figura 2, o resultado não só confirma a relação entre textos da mesma autoria, mas também define esses grupos com muita clareza e em categorias baixas. Isto demonstra que o nosso *corpus* está composto por textos muito singulares e que, se utilizarmos apenas a análise de frequências, poderemos concluir unicamente que possuem um vocabulário diferente, mas serão precisos outros métodos ou procedimentos para sabermos se este contribui ou não para compor assuntos ou linhas temáticas comuns; para isto haverá que trabalhar com a distribuição, a combinação, a intensidade ou o uso em espaços concretos do texto dessas palavras mais frequentes.

FIGURA 2 - DENDOGRAMA APLICANDO A MÉTRICA *CANBERRA* COMBINADA COM O MÉTODO *COMPLETE* E COM O PONDERADO *PROPORTIONAL COUNTS* (PROGRAMA *LEXOS*)



Fonte: Elaboração própria, 2015.

2) Os textos são vistos como sacos de palavras (*bag-of-words*⁵) e o seu conteúdo representado pela frequência delas, aplicando este método de maneira geral ao conjunto do *corpus* ainda quando não todos os textos têm o mesmo volume. Ao ser tomado o *corpus* na sua totalidade sem levar em conta a proporcionalidade, a análise não é estatisticamente válida, e tampouco dispomos da possibilidade de fazer comprovações apropriadas por causa da natureza do nosso *corpus*. A única

⁵ Considerar um documento como *bag-of-words* significa limitar o foco da análise a contar as palavras individuais que ele contém e calcular a frequência com que elas aparecem, sem levar em conta nem as suas combinações nem qualquer outra estrutura semântica de ordem superior. O conteúdo de um documento primeiro é quantificado calculando as frequências dos vocábulos individuais que contém e, depois, é representado como um conjunto ponderado destes termos, uma listagem de números cuja longitude é igual à quantidade de termos diferentes presentes no texto, e as entradas são os pesos frequenciais de cada termo. Tomando por junto essas listagens para todos os documentos de uma coleção, o programa constrói uma matriz de termo-documento, a qual exhibe como os diferentes termos estão distribuídos nos documentos constitutivos dessa coleção.

opção existente para evitar esta desigualdade é o módulo de corte (*chucking*) no nível de pré-processamento, escolha que não é operativa para o nosso *corpus* porque não podemos segmentar por capítulos (há uma novela com um único capítulo) e também não é possível fazer cortes com base no número de páginas por causa da grande diferença de volume entre os elementos da coleção de 2010 (escolhermos, por exemplo, as primeiras 100 páginas significa analisarmos uma novela na sua totalidade e apenas 10% ou 20% de uma dúzia de romances, com 500-700 páginas).

3) A própria técnica hierárquica aglomerativa tem alguma limitação, já que agrupa apesar dos eventuais erros de medição em algum conjunto sem realizar revisões do cálculo e também não dá opção de validação, com o qual continua a construir a árvore até ter todos os elementos unificados em um único elemento. A isto há que somar que as métricas disponíveis são as próprias das análises com dados numéricos, as quais, quiçá, não resultam as mais adequadas para operar com dados de natureza qualitativa (palavras). Ainda que há ótimas análises com textos literários utilizando a métrica *Euclídia* (estándar para operar com cifras), também é certo que outras medidas, como a métrica *Ward* a que alude Tsatsoulis (2013), não se encontram entre as opções deste *software*, pelo qual achamos que também deve ser levada em conta esta questão das métricas disponíveis como variável analítica.

Destas debilidades do *Lexos* deriva-se que é preciso contrastar estes resultados com outro *software* que aplique a análise de *cluster* a partir de um procedimento diferente. Contudo, antes de recorrermos à tipologia de formação de conglomerados por distância de compressão normalizada (NCD), vamos testar uma segunda modalidade de *clustering* baseada ainda no método de frequências, mas que permite a aplicação de outro tipo de métricas adaptadas a dados de natureza qualitativa (palavras), escolher o número de palavras que intervêm na análise e realizar provas complementares de validação.

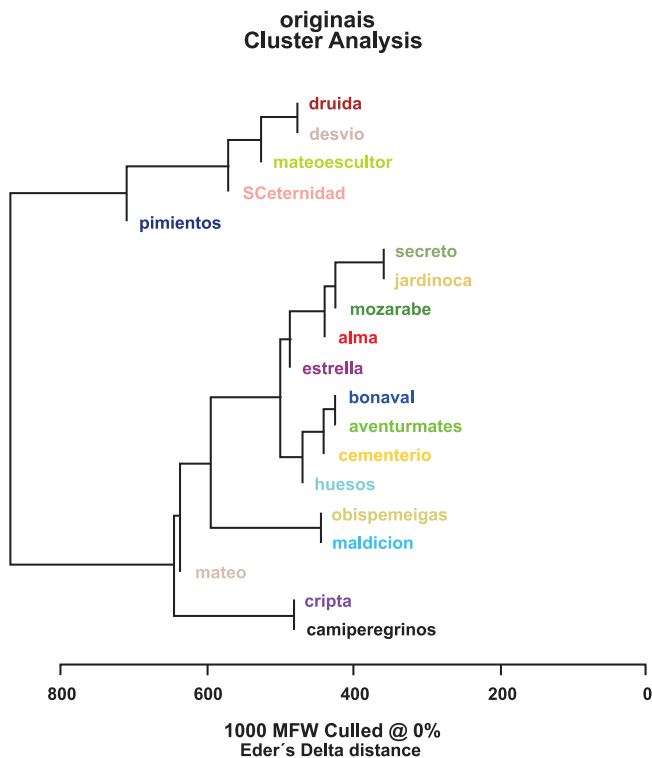
Aplicamos esta segunda modalidade por meio de um pacote do programa estatístico R chamado *Stylo*⁶. Este *software* permite trabalhar apenas com o

⁶ Disponível em: <<https://sites.google.com/site/computationalstylistics/home>>. Acesso em: 25 mar. 2020.

algoritmo hierárquico aglomerativo (com o qual representa os resultados em dendogramas, igual que *Lexos*, mas neste caso a visualização é mais esquemática e fácil de interpretar, porque a ramificação é menor) ora, para além da especialização qualitativa do tipo de métrica, *Stylo* estabelece também como critério analítico o número de palavras mais frequentes.

As nossas análises com *Stylo* foram realizadas alterando os valores em uma escala de 100 a 1000 palavras (de acordo com Jockers [2013] para a aplicação do método *topic modelling* com *corpora* literários) e utilizando as métricas *Canberra* e *Delta* (nas suas tipologias). A primeira já foi testada com o *Lexos*, o qual facilita o contraste, e as *Delta* são métricas especificamente criadas para trabalhar com dados textuais e línguas flexivas como o espanhol (BURROWS, 2002). Os agrupamentos definidos com maior robustez são os derivados das métricas *Canberra* e *Eder's Delta* e o resultado derivado dessas provas com a escala de 100 a 1000 palavras mais frequentes aponta para variações não tanto nos títulos dos agrupamentos como no grau de concreção dos relacionamentos entre os textos dos *clusters* maiores. Trata-se então de mudanças no nível micro, na construção de mais binómios ou, por outras palavras, de que as duas macroagrupações resultantes conformam mais pares no seu interior do que com o *Lexos*. Esta evidência permite afirmar que o tamanho do texto (isto é, ajustarmos a uma mesma quantidade de termos todos os documentos analisados), sim pode ser uma variável que cause interferências, uma questão que o programa *Lexos* não permitia esclarecer.

FIGURA 3 - DENDOGRAMA DE *CLUSTER* HIERÁRQUICO AGLOMERATIVO SOBRE AS 1000 PALAVRAS MAIS FREQUENTES E COM DISTÂNCIA MÉTRICA *EDER'S DELTA* (PROGRAMA *STYLO*)

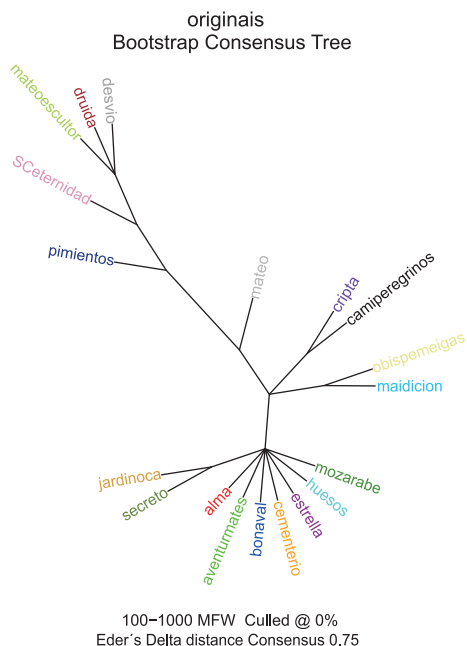


Fonte: Elaboração própria, 2015.

De acordo com o dendrograma mostrado na Figura 3, a seleção da métrica *Eder's Delta* combinada com o valor das 1.000 palavras mais frequentes organiza a coleção de textos de 2010 em dois grandes agrupamentos. O primeiro está constituído por cinco títulos coincidentes com os resultados da aplicação da distância *Canberra*: *Pimientos*, que atua como um grupo em si próprio, e um núcleo mais próximo criado entre *SCeternidad*, *Mateoescultor* e o par *Druida-Desvío*. O segundo grupo é de maior tamanho e consta dos outros 14 documentos, organizados, por sua vez, em quatro blocos: o par mais externo *Cripta-CamiPeregrinos*, um elemento intermedial solitário (*Mateo*), o par *Obispomeigas-Maldición* e o conjunto de *Huesos*, *Cementerio*, *Aventurmates-Bonaval* por um lado, e *Estrella*, *Alma*, *Mozárabe* y *Jardínoca-Secreto* por outro lado.

Ora, tal como foi indicado acima, o *Stylo* realiza a validação da robustez dos grupos e da sua formação utilizando para isto um método chamado árvore de consenso ou *bootstrap consensus tree*, que recalcula de maneira aleatória os grupos e oferece um agrupamento próprio dos textos que não se visualiza em distribuição hierárquica, mas em forma de árvore sem raiz; os elementos são colocados assim de maneira associativa, com o qual quanto mais próximos estejam nas ramificações mais similitude de vocabulário compartilharão entre eles. Este método da árvore de consenso exige definir a força desse consenso, estabelecido em 0,5 como medida predefinida (quer dizer que um determinado vínculo será formalizado se aparecer em 50% dos *clusters* virtuais completados), e fixar também o tipo de mostras sobre as quais vai operar (*sampling*), recomendando-se uma mostragem aleatória e a exportação por meio das opções *GRAPH* e horizontal *CA tree*. A Figura 4 representa a árvore de consenso correspondente com a métrica *Eder's Delta* e o valor de 1000 palavras mais frequentes, e confirma a descrição colocada acima.

FIGURA 4 - ÁRVORE DE CONSENSO DE ANÁLISE DE CLUSTER COM SELEÇÃO DE 100 A 1000 PALAVRAS MAIS FREQUENTES E MÉTRICA *EDER'S DELTA* (PROGRAMA *STYLO*)



Fonte: Elaboração própria, 2015.

Em geral, os resultados com o *stylo* evidenciam que, segundo o critério de formação de pares, *Canberra* é a métrica que melhor facilita a obtenção de mais binómios entre textos (7, frente aos 5 de *Eder's Delta*) e, deste ponto de vista, parece a métrica que melhor se ajusta ao nosso *corpus*. Porém, a capacidade de formação de pares, isto é, a maior definição de relacionamentos entre textos por MFW (*more frequency words*) que permite *Stylo*, permite afirmar que a escolha de um número similar de palavras mais frequentes para todos os textos é uma variável que favorece a deteção de padrões comuns. Para além disso, ao efetuarmos o contraste dos grupos por análise de cluster com a prova da árvore de consenso, os agrupamentos extraídos de *Canberra* não coincidem com o dendograma, o que sim acontece com *Eder's Delta*, facto que valida os resultados atingidos com esta última técnica como os mais fiáveis. Ora, tanto com uma métrica mais ajustada como com uma mais geral, e ainda resultando uma maior extração de pares que com o *Lexos*, a unificação dos textos continua a se produzir a escalas altas, característica que coloca de novo o foco na singularidade do *corpus* e nas limitações do simples reconto de frequências.

Assim, podemos concluir que os principais problemas deste procedimento de análise de frequências continuam a ser a impossibilidade de seleção de um dos textos desses pares extraídos, porque o programa não permite saber qual deles contribui em maior medida para a sua definição e, em especial, não poderemos aceder a uma mostragem dos termos característicos de cada grupo, já que as matrizes de vocabulário são demasiado extensas para serem revistas e sistematizadas a olho nu (e nos pretendemos agrupar a partir dos padrões léxico-semânticos do conteúdo textual). A isto será preciso acrescentar a dificuldade de operar com propostas de classificação derivadas, no sentido de que o contraste entre métodos dá um resultado que pode não coincidir com o da prova de validação (tal como apontado acima). Tudo isto aconselha testar outro procedimento de *cluster* diferente da análise de frequências.

MÉTODO DE COMPRESSÃO. PROGRAMA *COMPLEARN*

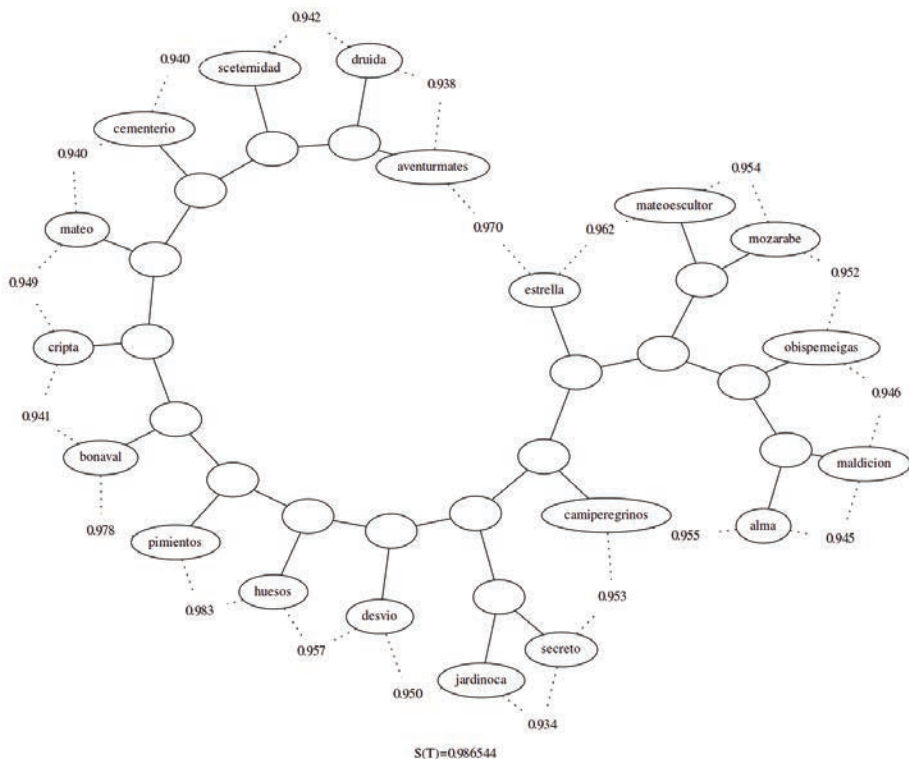
O *software CompLearn*⁷ oferece o procedimento de análise de *cluster* denominado *normalized compression distance* (NCD) para medir a similitude entre

⁷ Disponível em: <<http://complearn.org/>>. Acesso em: 25 mar. 2020.

dous objetos. Contudo, é possível também utilizar esta distância para agrupar objetos, tal como pretendemos neste trabalho. Os resultados desta operação são visualizados em uma árvore sem raiz baseada no *quarter method*, uma modalidade em que as relações entre os elementos são dispostas em forma de nodos e subnodos em função da sua maior ou menor proximidade, oferecendo-se assim, para cada texto, apenas a relação com aquele com o qual possui a similitude mais alta, quantificada em uma escala de 0 a 1. Esses valores aparecem expressados ao lado de cada ramificação que une dous textos e, a partir deles, é ponderado o valor geral colocado como nota de rodapé depois das siglas ST.

A partir, então, de uma mesma semente aleatória geram-se varias visualizações entre as quais é possível escolher o grafo com menor ST e com melhor disposição espacial dos elementos (alguns ramos podem aparecer mais ou menos solapados) ainda que a colocação no plano é aleatória e não é indicador de qualquer significação; de qualquer maneira, não é recomendável utilizar *corpus* de mais de 40 elementos, porque a partir desse volume a árvore tende a se degradar. Deste modo, a visualização dos resultados pelo método *Quarter* permite contrastar ou até resolver o eventual problema dos dendogramas apresentados pela classificação hierárquica, já que no caso do *Quarter* o algoritmo sim recalcula as operações de agrupamento até se deter quando deteta a melhor representação possível, no nosso caso a mostrada na Figura 5.

FIGURA 5 - ANÁLISE DE *CLUSTER* POR DISTÂNCIA DE COMPRESSÃO NORMALIZADA (NCD) COM *SOFTWARE COMPLEARN*



Fonte: Elaboração própria, 2015.

Na Figura 5 vemos como o segundo grupo está localizado na direita e formado por cinco textos muito próximos entre si (em cinco elementos são estabelecidos dois pares). Entre este segundo bloco e o primeiro localizam-se dois títulos, *Estrella* e *Camino de peregrinos*, cuja posição pode ser entendida quer como espaço de fronteira – isto é, como textos de natureza intermedial que partilham características de ambos os grupos – quer como extremos da cadeia do grupo maior (*CamiPeregrinos*) e do pequeno (*Estrella*). O primeiro agrupamento é o ramo localizado na esquerda da Figura 5 e une 12 textos (ou 13 se incluirmos *CamiPeregrinos*). Neste grupo também achamos dois binómios: *Jardínoca-Secretro* e *Druida-Aventurmates*. Quer o par *Jardínoca-Secretro* quer a proximidade entre *Maldición-Alma* e *Obispomeigas*

são resultados coincidentes com as análises de *cluster* por frequências, enquanto que as principais diferenças com esse procedimento têm a ver, por um lado, com a inclusão agora de *Pimientos* na macroagrupação maior, quando antes funcionava sempre de maneira individual, e por outro lado, com o nexos entre *Mateoescultor* e *Mozárabe*, quando nunca antes estiveram próximos (de facto, o primeiro deles formava o grupo mais radial com *Desvío*, *Druida* e *SCeternidad* ...).

As distâncias entre os textos situam-se entre 0,934 e 0,983, o que resulta em um valor ST para a coleção de 2010 de 0,986. Como o valor de confiança é aquele situado mais perto de 0 e aqui as quantidades aproximam-se de 1, devemos concluir que existem poucos elementos comuns entre os textos do nosso *corpus*, sendo portanto a heterogeneidade e a singularidade as suas principais características, tal como já hipotetizámos na análise com o *corpus* de controlo no *Lexos*⁸. Ao contrário da classificação hierárquica, a visualização do *cluster* por compressão tirada do *quarter method* não permite inicialmente derivar uma leitura tão intuitiva, mas sim serve de contraste frente às limitações de validação que oferece a primeira. O facto de que a compressão estabeleça um macrogrupo conformado por muitos mais títulos que nas análises de conglomerados por contagem de frequências coloca-nos novamente perante a dúvida de se este procedimento permite um ajustamento ótimo à informação textual da coleção de 2010, para além de que, outra vez, praticamente não foram derivados pares sobre os quais aplicarmos reduções, com o qual também não podemos conformar uma mostragem representativa e significativa da totalidade do nosso *corpus*.

BALANÇO E SÍNTESE CONCLUSIVA

A análise de *cluster* realizada a partir dos dous procedimentos de conglomerados explorados (frequências e compressão) permite estabelecer agrupações dos textos reduzindo empiricamente os materiais. Contudo, o problema

⁸ Ao contrario doutros exemplos observados na bibliografia disponível (CILIBRASI e VITÁNYI, 2005), chama a atenção a forma que resulta da disposição dos objetos na representação gráfica, já que tendem a se fecharem sobre si próprios e formarem uma espécie de oito. Com *corpus* doutra natureza (amostras musicais, estudos de ADN ou animais) isto não acontece e o desenho da árvore é aberto e linear. Neste sentido, colocamos apenas como hipótese que esta tendência se calhar tem a ver com uma maior criatividade da linguagem literária, que contribui para que não haja tantas estruturas repetidas.

para atingir através destes procedimentos o nosso objeto de estudo reside em que a tendência geral é à constituição de macroagrupamentos e não de subdivisões, assim como que esses agrupamentos tampouco coincidem exatamente nos objetos que os conformam nos três casos analisados. Fora isso, também não podemos conhecer qual é o contributo de cada texto para a definição de cada grupo, com o qual não é possível selecionar os textos mais representativos de cada agrupamento. Apesar disto, sempre é possível identificar alguns pares textuais, ainda que nenhum deles goze de suficiente robustez ou validação. De *Lexos* resultam *Mateoescultor-Mateo*, *Alma-Jardínoca*, *SCeternidad-Desvío*, *Mozárabe-Estrella*, *Huesos-CamiPeregrinos* e *Cementerio-Bonaval*; em *Stylo* vinculam-se *Druida-Desvío*, *Secreto-Jardínoca*, *Bonaval-Aventuras*, *Obispos-Maldición* e *Cripta-CamiPeregrinos*; e em *CompLearn* identificamos a maior proximidade entre *Mozárabe-Mateo*, *Maldición-Alma*, *Jardínoca-Secreto* e *Druida-Aventurmates*. De todos eles, apenas três são extraídos simultaneamente em dois dos três procedimentos: *Secreto-Jardínoca*, *Cripta-CamiPeregrinos* e *Obispos-Maldición*. Isto reforça a conclusão de que a capacidade de redução da coleção de 19 textos de 2010 a um grupo representativo e quantitativamente baixa e pouco operativa, pois os pares gerados permitem apenas reduzir em três elementos o conjunto dos objetos do *corpus* (ficando conformado este ainda por 16).

Por outro lado, ainda que sim é possível agrupar os textos, os métodos testados não permitem saber em que se baseia a similitude ou proximidade entre eles (que é o que determina as uniões), porque as matrizes de dados léxicos geradas são muito grandes e praticamente impossibilitam o seu processamento por meio de uma leitura cercana. Em consequência, a técnica de *cluster* não permite observar o conteúdo característico das agrupações extraídas, quer dizer, impede aceder ao vocabulário e, portanto, avançar noutros objetivos propostos porque, em última instância, não possibilita a categorização dos grupos nem a delimitação dos repertórios culturais que os configuram. Estas limitações fazem com que tenhamos que descartar a análise de conglomerados como técnica única de processamento e sistematização do *corpus* textual da nossa coleção de 2010.

Os testes feitos neste trabalho representam, enfim, um passo intermédio no caminho para conseguirmos os nossos objetivos, um percurso no qual avançamos por ensaio-erro e por descarte, e um processo crucial para irmos definindo as técnicas e procedimentos mais apropriados para as nossas necessidades. Nesse

sentido, ao detetarmos na aplicação da técnica de *cluster* as problemáticas e insuficiências descritas nesta achega, optamos por recorrer a uma modalidade de classificação semântica que incorpora técnicas multivariantes, a técnica textométrica de classificação semântica, uma tipologia que une ao método de contagem de frequências as coocorrências e permite determinar as relações léxico-temáticas, superando assim todas as debilidades apontadas para as técnicas aqui analisadas. Os resultados derivados da análise de *cluster* serviram, de qualquer maneira, como contraste e reforço dos grupos temáticos estabelecidos com o recurso à textometria por frequências-coocorrências, uma técnica aplicada satisfatoriamente com o programa informático *Iramuteq*⁹, tal como foi desenvolvido por extenso na tese de doutoramento de Fernández Rodríguez (2016).

REFERÊNCIAS

BURROWS, J. 'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, v. 17, n. 3, p. 267-287, 2002.

CILIBRASI, R.; VITÁNYI, P. M. B. Clustering by compression. *IEEE Transactions on Information Theory*, v. 51, n. 4, p. 1523-1545, 2005.

FERNÁNDEZ RODRÍGUEZ, M. L. *Discursos sobre Santiago de Compostela y el/ los Camino(s) de Santiago en la novela española actual (2010) a través de técnicas analíticas digitales: posibilidades y valor del conocimiento generado*. 2016. 493 f. Tese (Doutorado). Santiago de Compostela: Universidade de Santiago de Compostela, 2016.

JOCKERS, M. L. *Macroanalysis: Digital Methods and Literary History*. Chicago, Springfield: University of Illinois Press, 2013.

SAMARTIM, Roberto. Bases de dados para o estudo da cultura: apresentação do catalogador e possibilidades de abordagem sobre o *corpus* documental do Projeto Caminho de Santiago. In: TORRES FEIJÓ, Elias José; BELLO VÁZQUEZ, Raquel; SAMARTIN, Roberto; BRITO-SEMEDO, Manuel. (Ed.). *Estudos da AIL em Teoria e Metodologia: Relacionamento nas Lusofonias II*. Santiago de Compostela; Coimbra: Associação Internacional de Lusitanistas; AIL Editora, 2015. p. 115-125.

TOMÁS DÍAZ, D. *Sistemas de clasificación de preguntas basados en corpus para la búsqueda de respuestas*. 2015. 204 f. Tese (Doutorado) - Universitat d'Alacant, 2009.

⁹ Disponível em: <<http://iramuteq.org/>>. Acesso em: 25 mar. 2020.

TORRES FEIJÓ, Elias José. *Discursos contemporâneos e práticas culturais dominantes sobre Santiago e o Caminho*: a invisibilidade da cultura como hipótese. In: LOURENÇO, A. A.; SILVESTRE, D. M. (Coord.). *Literatura, espaço, cartografias*. Coimbra: Centro de Literatura Portuguesa, 2011. p. 93-151.

TSATSOULIS, C. Unsupervised Text Mining Methods for Literature Analysis: a Case Study for Thomas Pynchon's *V. Orbit: Writing Around Pynchon*, v. 1, n. 2, p. 1-33, 2013. Disponível em: <<https://orbit.openlibhums.org/article/id/408/>> Acesso em: 25 mar. 2020.

Comunicação digital em museus e instituições de património cultural do Alentejo: estado da arte¹

Nicola Schiavottiello

João Brigola

INTRODUÇÃO E DADOS PRELIMINARES

O Alentejo é, em diversos aspetos, uma das regiões mais interessantes de Portugal e do Sul da Europa. Sendo a região menos populosa, com a taxa de natalidade mais baixa e mais envelhecida de Portugal, é também a mais extensa, com uma área de 31 551,2 km² e uma população de 757 302 habitantes cuja média etária se situa na faixa dos 43 aos 50 anos de idade (INSTITUTO NACIONAL DE ESTATÍSTICA, 2012).

É indubitavelmente notável pelos seus recursos culturais, que parecem fundir-se de forma simbiótica em quatro grandes áreas: Histórica-Monumental-Industrial, Sacro-Artística, Etnografia e Arte Popular e Natureza. A riqueza de culturas que se foi consolidando nesta zona ao longo dos tempos contribuiu para aumentar a importância da região, não só a nível espacial, mas também a nível histórico.

As cidades mais emblemáticas, Évora (núcleo urbano antigo) e Elvas (arquitetura militar fortificada), foram classificadas como Património da Humanidade e, na última década, surgiram várias investigações e projetos públicos premiados na área da museologia e património cultural (ASSOCIAÇÃO PORTUGUESA DE MUSEOLOGIA, 2011).

Com base nos valores essenciais elencados no PENT 2013-2015 (PORTUGAL, 2013) – clima e luz favoráveis, história, cultura e tradição, hospitalidade e diversidade concentrada – e através da implementação do novo programa Alentejo 2020 (CCDR Alentejo, 2014), que define o Património e o

¹ A pesquisa para este trabalho foi feita no ano de 2016.

Ambiente como temas centrais, espera-se que a utilização de novas tecnologias (que visam criar, não só metodologias de investigação inovadoras, como também novos dispositivos de comunicação) possa vir a abranger a área do Património Cultural ao longo dos próximos anos.

Apesar de a riqueza de recursos nos permitir considerar a situação atual como sendo bastante promissora, o panorama da Comunicação Digital do Património parece estar, no mínimo, fragmentado, com a existência de projetos que são implementados pelas entidades locais sem que sejam estudados, organizados e relacionados entre si de acordo com uma abordagem colaborativa. Algumas propostas foram apresentadas no âmbito do PENT 2013-2015, cuja abordagem está mais centrada nos aspetos turísticos do que nos aspetos culturais. Muitas das atividades referidas neste documento, tais como a produção de narrativas audiovisuais e jogos, estão orientadas para a internet e para os dispositivos móveis. Destinam-se, no entanto, ao setor privado, não existindo qualquer referência a instituições culturais públicas. De um modo geral, o Documento Estratégico Turismo do Alentejo 2014-2020 (TURISMO DO ALENTEJO, 2014) encontra-se novamente orientado para a promoção da utilização de aplicações móveis juntamente com plataformas digitais *on-line* destinadas à consulta de recursos culturais por parte dos turistas.

Existem ainda outros protocolos que visam abordar esta questão, tais como o antigo ADI/ITP (PRESIDÊNCIA DO CONSELHO DE MINISTROS 2005) do 2006, que propõe a introdução de novas tecnologias no processo produtivo das atividades turísticas, ou o EREI 2020 (CCDR Alentejo, 2014), que reflete sobre as Key Enabling Technologies (KEY) e a importância das tecnologias móveis emergentes para o setor turístico.

Contudo, parece não existir, no âmbito desta questão, qualquer plano específico para as instituições ligadas ao Património Cultural, pelo que consideramos que deveria ser dado um importante primeiro passo no sentido de mapear a situação atual.

METODOLOGIA DE INVESTIGAÇÃO

A metodologia de investigação adotada considerou a existência de duas fases distintas. Em primeiro lugar, as instituições foram divididas em duas grandes categorias: aquelas que não oferecem meios de comunicação baseados em novas

tecnologias e as que apresentam algum tipo de dispositivos de comunicação digital. Procedeu-se a uma verificação do tipo de recursos existentes, bem como da sua efetiva utilização, através do contacto telefónico com todas as instituições. Este foi um processo bastante moroso, pois nem todas as instituições se mostraram disponíveis numa primeira abordagem. Em certos casos foram necessárias diversas tentativas e, como veremos adiante, existem falhas de informação, ou porque não foi possível contactar a instituição em causa, ou porque continuamos a aguardar a resposta ao pedido de contacto.

Numa segunda fase foi enviado um questionário para as instituições que possuem pelo menos um dispositivo de comunicação digital, independentemente de este ser um simples vídeo explicativo ou um ambiente imersivo de realidade virtual ou aumentada. Apesar de também levantar diversos problemas, tais como a má interpretação das perguntas, uma baixa taxa de respostas e a existência de diferentes respostas dadas por diferentes indivíduos da mesma instituição, este tipo de metodologia foi escolhido pelo seu baixo custo e pelo facto de permitir recolher uma grande quantidade de informação num período de tempo relativamente curto (BERGER, 2000).

O questionário fornecido teve em consideração diversos aspetos de questões específicas associadas à utilização de diferentes tipos de tecnologias, bem como aos respetivos conteúdos. Incluímos ainda outros fatores que, à partida, parecem não estar especificamente relacionados com os recursos digitais, mas que, tal como veremos na discussão, assumem alguma relevância. Referimos, como exemplos, o tipo de coleções detidas por cada entidade, os meios gerais de comunicação (dentro e fora das instituições) e as formas mais tradicionais de comunicação.

DADOS DESCRITIVOS GERAIS

Além da informação específica acerca dos recursos digitais, que exploraremos abaixo, considerámos, em primeiro lugar, a localização das instituições. Este é um aspeto extremamente importante para a proposta de estudos subsequentes sobre a acessibilidade de cada instituição, bem como para a análise do ambiente cultural local. Assim, a escolha de uma abordagem georreferenciada resulta do facto de a

base de dados incluir, por natureza, instituições físicas que precisam de ser estudadas no contexto do seu território e também porque, no futuro, isso permitirá analisar individualmente cada instituição mediante a recolha fácil de informação através de um simples mapa *on-line*.

A informação final incluída na ficha de cada entidade resume os campos mais importantes a apresentar nesta fase, tais como o nome, a classificação das tecnologias (ver classificação específica abaixo), o tipo de instituição, a morada, o código postal, a localidade, a freguesia, o número de telefone, o e-mail, o URL, a latitude, a longitude, a referência no sítio web dos museus nacionais, a referência no sítio web de turismo da região, a referência no sítio web de cultura da região e notas adicionais. Nem todos os campos estão completamente preenchidos, mas esse facto não compromete este trabalho específico. Numa fase posterior, todos estes campos serão preenchidos para a publicação *on-line* de um mapa de pesquisa interativo. Segue-se um exemplo de uma das fichas que foram criadas (Tabela 1).

TABELA 1 - EXEMPLO FICHA DE INFORMAÇÃO GERAL DE UMA DAS ENTIDADES

<p>Nome: Museu de Arqueologia de Montemor-o-Novo Classificação das tecnologias: 0 Classificação: Museu Morada: Rua de S. Domingos (Convento de S. Domingos)- Largo Prof. Dr. Banha de Andrade Código Postal: 7050-111 Localidade: Montemor-o-novo Concelho: Montemor o Novo Telefone: 266 890296 / 266 890235 Email: g.amigos.montemor@gmail.com Url: http://museumontemor.com.sapo.pt Latitude: 38.643406 Longitude: -8.212603 Museus De Portugal: http://www.museusportugal.org/museus.aspx?modo=detalhe&menu=125&id=126&start=0&idi=209 Turismo Alentejo: Cultura Alentejo: Note:</p>

Fonte: Museu de Portugal (2016).

INSTITUIÇÕES CONTACTADAS

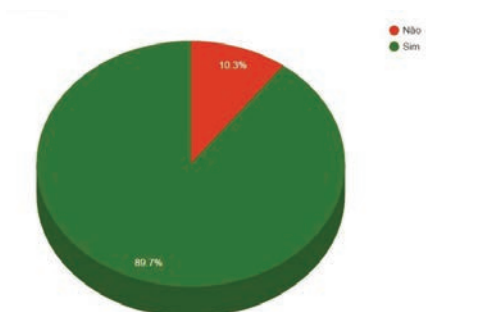
Existe, à data de outubro 2015, um total de 195 instituições no território em estudo² (Figura 1). Destas, 175 foram contactadas, restando apenas 20 para contactar no futuro, como se pode ver no Gráfico 1 abaixo. Assim que esse objetivo for atingido, este gráfico será eliminado e a amostra será representativa da totalidade da população museológica do Alentejo. As novas instituições que forem surgindo poderão facilmente ser incluídas na investigação.

Figura 1 - Mapa de Instituições contactadas e não contactadas⁹



Fonte: Schiavottiello (2016).

Gráfico 1 - Instituições contactadas e não contactadas 2016



Contatados 175 unidades - 89,7%
 Não Contatados 20 unidades - 10,3%

Fonte: Schiavottiello (2016).

INSTITUIÇÕES COM E SEM RECURSOS DIGITAIS

Podemos dividir a totalidade das instituições contactadas em duas categorias: museus que possuem recursos digitais (70 instituições em verde) e museus que não os possuem (105 instituições em vermelho) (Figura 2). O Gráfico 2 permite ver claramente que o número de museus que apresentam recursos digitais é apenas 20% inferior ao dos que não possuem. No entanto, como veremos adiante, precisamos de ter algum cuidado na leitura deste gráfico, pois a sua interpretação é influenciada

² A lista de instituições estudadas foi compilada através de um conjunto de dados fornecido de entidades regionais e nacionais, estas são: Turismo do Alentejo, Direção Regional da Cultura do Alentejo (DRCA) e Museus de Portugal.

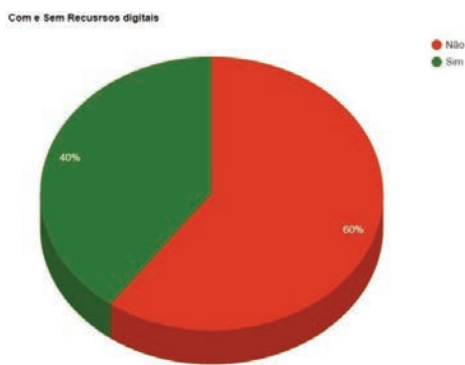
peelo tipo de recursos em causa. Por exemplo, muitos museus dispõem de apenas um pequeno vídeo introdutório ou um computador na entrada, mas, neste gráfico, todos os recursos são considerados ao mesmo nível. Por esse motivo, foi necessário subdividir esta categoria atribuindo diferentes níveis de importância às instituições em função dos respetivos recursos digitais. Isso foi conseguido através da criação de um fator de ponderação do recurso em causa, fator esse que será explicado adiante. No mapa (Figura 2) incluem-se também as instituições (na cor laranja) que não foram consideradas nesta fase, pois precisam ainda de ser contactadas.

Figura 2 - Mapa de Instituições contactadas e não contactadas



Fonte: Schiavottiello (2015).

Gráfico 2 - Instituições com e sem recursos 2016



Com recursos 105 unidades - 60%
Sem recursos 70 unidades - 40%

Fonte: Schiavottiello (2016).

INSTITUIÇÕES COM RESPOSTAS

O mapa (Figura 3) abaixo apresenta a categoria dos museus que possuem recursos de comunicação digital subdividida em duas subcategorias, uma correspondendo aos que responderam ao questionário (56 instituições em verde) e outra aos que ainda não o fizeram (14 instituições em vermelho). Esta secção será eliminada num futuro próximo, pois espera-se que todos os museus abrangidos enviem as suas respostas. No Gráfico 3, podemos ver que, apesar de serem poucos

os museus que ainda não responderam ao questionário, eles representam uma fatia significativa da amostra. Devemos ter em consideração que este não é apenas um levantamento quantitativo de recursos semelhantes e que cada instituição pode ter uma grande influência no resultado final do mapeamento dos recursos por tipo. No mapa, em cor cinzenta estão representadas as instituições que não possuem recursos digitais (não incluídas no Gráfico 3), e de cor laranja aquelas que precisam ainda de ser contactadas.

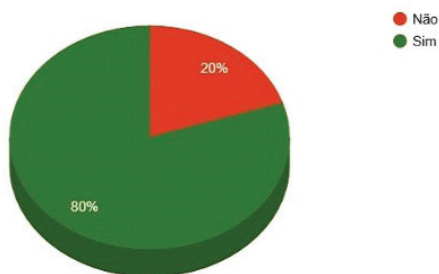
Figura 3 - Mapa de Instituições com recurso e sem recursos



Fonte: Schiavottiello (2016).

Gráfico 3 - Instituições com e sem recursos

Respostas no tipo de recursos

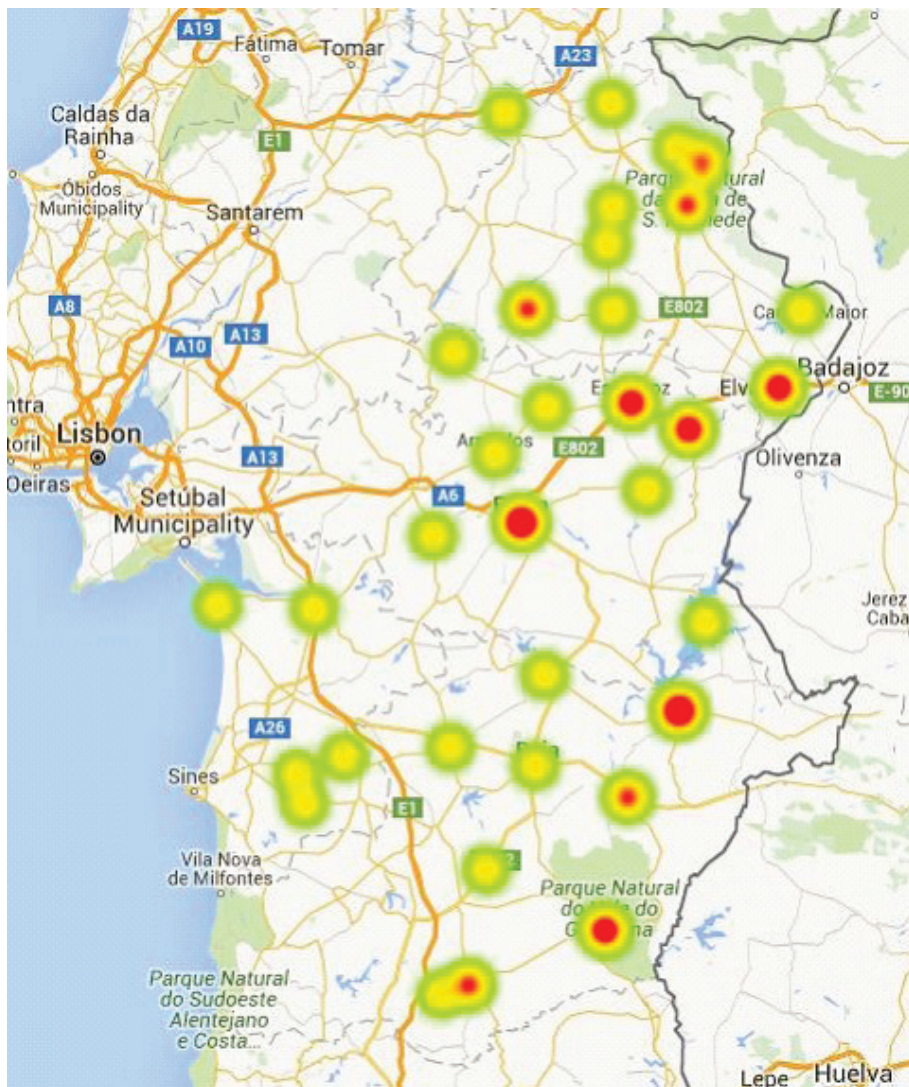


Com recursos 105 unidades - 80%
Sem recursos 70 unidades - 20%

Fonte: Schiavottiello (2016).

Finalmente, apresentamos um mapa de pontos quentes (Figura 4), que mostra a distribuição espacial dos museus com recursos digitais. As áreas são representadas com cores que variam do vermelho ao verde, sendo que o vermelho corresponde à maior concentração de instituições com recursos digitais. O mapa mostra-nos que, em geral, se não especificarmos o tipo de tecnologias utilizadas para comunicar com o público, elas estarão bem distribuídas nas regiões de maior concentração, Évora e Elvas. Contudo, devemos ter em conta o número de instituições existentes em cada localização, que é aquilo que este mapa realmente representa.

FIGURA 4 - MAPA DE PONTOS QUENTES (DISTRIBUIÇÃO DO NUMERO DE INSTITUIÇÕES COM RECURSOS DIGITAIS)



Fonte: Schiavottiello (2016).

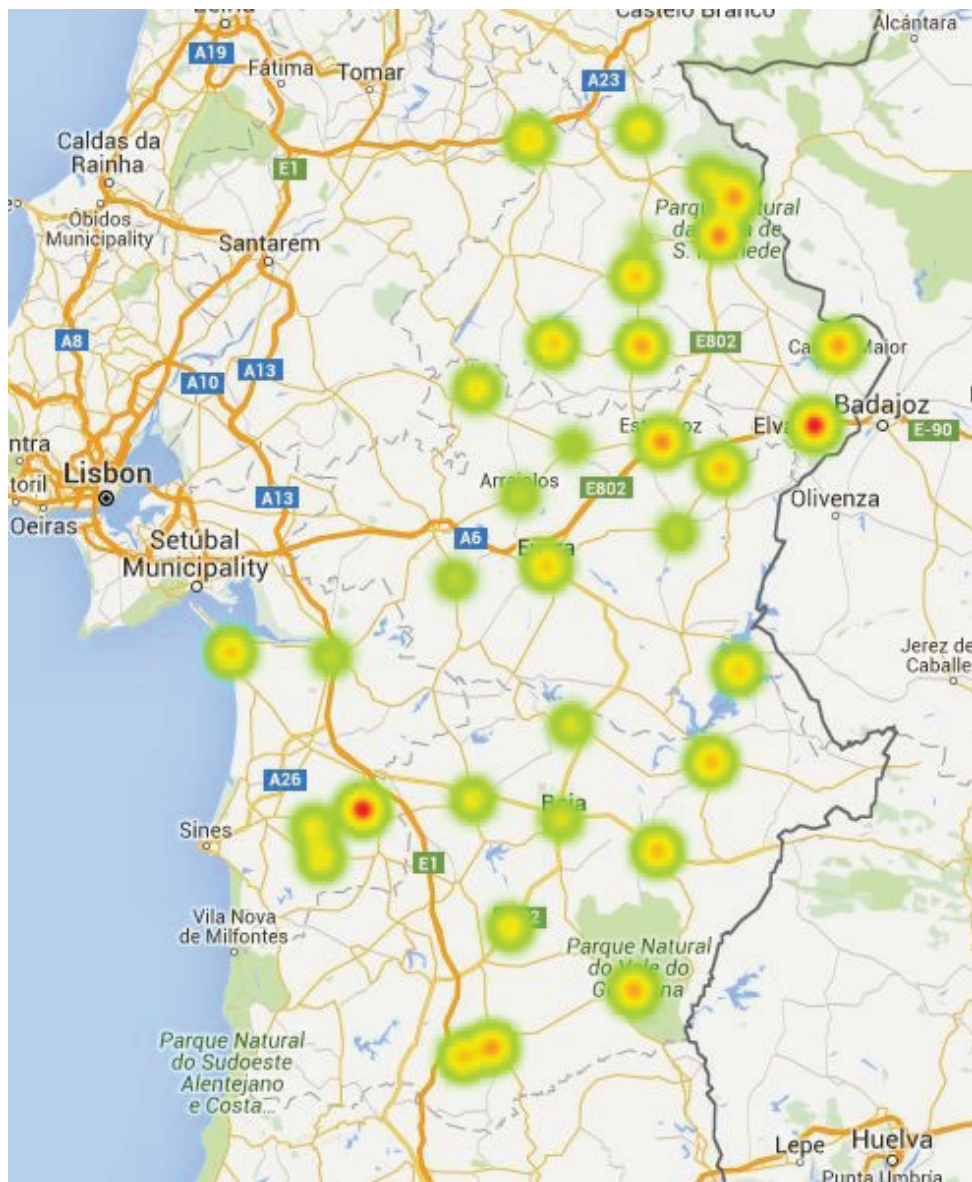
DADOS QUANTITATIVOS ESPECÍFICOS

O objetivo do presente estudo não se resume ao mapeamento das diversas instituições e dos respetivos conteúdos, pois isso apenas nos forneceria uma representação dos maiores núcleos de instituições com recursos digitais existentes no território em estudo (Figura 7). Na realidade, este estudo pretende ampliar a relevância de cada instituição com base em dois conceitos principais: imersividade e interatividade. Para isso, procuramos quantificar não apenas as instituições, como também, sobretudo, os próprios recursos. Neste sentido, criaremos um fator de ponderação calculado com base na soma dos recursos existentes em cada instituição, de forma a gerar um mapa que descreva os locais onde existem mais recursos tecnológicos, bem como os conteúdos associados.

Neste mapa final (Figura 5), cada instituição foi multiplicada por um fator de ponderação que corresponde, tal como referimos, à soma dos recursos digitais existentes em cada instituição. O cenário resultante é diferente do mapa anterior. Se compararmos os dois mapas podemos observar, por exemplo, que enquanto Évora dispõe de um grande número de museus com recursos digitais, o distrito de Elvas parece ter muito mais recursos de base tecnológica, bem como o concelho de Grândola, neste caso devido à presença do Centro Ciência Viva do Lousal, que conta com uma forte implementação de recursos de base tecnológica.

Na realidade, podem ser criados diferentes fatores de ponderação, de acordo com o objetivo específico do trabalho de investigação final (tais como, por exemplo, a soma dos recursos digitais dedicados apenas à comunicação externa ou a soma de meios de comunicação analógicos mais tradicionais etc.). Neste caso, calculámos apenas a soma dos recursos dedicados à comunicação visual e interativa interna, incluindo dispositivos, métodos, técnicas e interfaces.

FIGURA 5 - MAPA DE PONTOS QUENTES (DISTRIBUIÇÃO DOS RECURSOS DIGITAIS)



Fonte: Schiavottiello (2016).

Exemplo de uma instituição com poucos recursos digitais inserida numa zona onde existem outras instituições com recursos digitais (Tabela 2):

TABELA 2 - FICHA DE EXEMPLO PARA MAPA FINAL DE PONTOS QUENTES

<p>Designação/ões: Fórum Eugénio de Almeida</p> <p>Weight: 3</p> <p>Latitude: 38.572011</p> <p>Longitude: -7.908086</p> <p>Loc: Évora</p> <p>Tipologia do Museu:: Centro de Arte Contemporânea</p> <p>Tipologia das colecções::</p> <p>Meios de comunicação com os diferente públicos:: Sitio web de apresentação, Email, Newsletters, Redes sociais, Comunicação social (Jornais/Revistas/TV/Radio), brochuras, catálogos, pendoes e posters</p> <p>Meios com diferente públicos valor:: 7</p> <p>Meios de comunicação analógicos: Atendimento, Visitas guiadas, Tabelas, Panfletos, Folhas de sala</p> <p>Meios analógicos valor: 5</p> <p>Meios de comunicação menos usuais:: o plano de actividades anual é apresentado publicamente uma vez por ano. fizemos várias campanhas de divulgação menos usuais, por exemplo, uma no dia internacional dos museus que juntou várias instituições da cidade de Évora e que numa arruada, ao som dos gigabombos, foram distribuídos vouchers gratuitos para uma entrada em todas as instituições que integraram o evento.</p> <p>Dispositivos de comunicação::</p> <p>Dispositivos valor: 0</p> <p>Métodos de comunicação digital::</p> <p>Métodos valor:: 0</p> <p>Temáticas:: Históricas, Artísticas</p> <p>Temáticas valor:: 2</p> <p>Técnicas digitais usados para comunicar o conteúdos ao público:: Imagens digitais (no monitor, projetadas ou outras formas digitais), Texto digital (no monitor, projetadas ou outras formas digitais)</p> <p>Técnicas valor:: 2</p> <p>Interfaces interativas:: Touch screens</p> <p>Interfaces valor:: 1</p> <p>Usa ou usaria ambientes virtuais para mostrar os conteúdos:: Não</p> <p>Caso tenha respondido não, justifique:: por um lado por serem muito dispendiosos, e por outro porque não se coaduna com os conteúdos artísticos da programação do Fórum. Também seriam necessários espaços especiais para estes ambientes que neste momento estão destinados a exposições temporárias.</p> <p>Utilizaria ambientes virtuais se a atualização dos respetivos conteúdos fosse simplificada:: Não</p> <p>Caso tenha respondido, não justifique:: pelas mesmas razões anteriores</p> <p>Observações partes em falta:: há vários mídias digitais que ainda não temos disponíveis mas nos quais estamos a trabalhar como vídeos e audioguias.</p>

Fonte: Schiavottello (2016).

Exemplo de uma instituição com muitos recursos digitais, mas isolada de outras instituições (Tabela 3):

TABELA 3 - FICHA DE EXEMPLO PARA MAPA FINAL DE PONTOS QUENTES

Designação/ões: Centro Ciência Viva do Lousal
Weight: 15
Latitude: 38.036231
Longitude: -8.426506
Loc: Lousal
Tipologia do Museu:: Museu de Ciência e de Técnica
Tipologia das coleções:: Património Industrial, Ciência
Meios de comunicação com os diferente públicos:: Sitio web de apresentação, Email, Redes sociais
Meios com diferente públicos valor:: 3
Meios de comunicação analógicos: Atendimento, Visitas guiadas, Projeções, Painéis, Panfletos
Meios analógicos valor: 5
Meios de comunicação menos usuais::
Dispositivos de comunicação:: Ecrãs, Projetores, Hologramas, Ambientes imersivos (360 panoramas, cave, etc.. para gerar uma sensação de imersão), Óculos VR/AR, Dispositivos de áudio imersivo
Dispositivos valor: 6
Métodos de comunicação digital:: Filmes (não interativos, imagens reais ou geradas por computador), Videojogos
Métodos valor:: 2
Temáticas::
Temáticas valor:: 0
Técnicas digitais usados para comunicar o conteúdos ao público:: Documentários filmados, Imagens digitais (no monitor, projetadas ou outras formas digitais), Texto digital (no monitor, projetadas ou outras formas digitais), Ambientes virtuais, Audio digital, Multimédia, Impressora 3 D e drone
Técnicas valor:: 6
Interfaces interativas:: Touch screens
Interfaces valor:: 1
Usa ou usaria ambientes virtuais para mostrar os conteúdos:: Sim
Caso tenha respondido não, justifique::
Utilizaria ambientes virtuais se a atualização dos respetivos conteúdos fosse simplificada:: Sim
Caso tenha respondido, não justifique::
Observações partes em falta::

Fonte: Schiavottiello (2016).

IMERSIVIDADE E INTERATIVIDADE

Devemos ter em mente que, uma vez que o fator de ponderação consiste apenas na soma dos recursos totais de uma dada instituição, para ser possível perceber o tipo de tecnologia presente no território precisamos de aprofundar o estudo da tipologia geral dos recursos digitais utilizados e, por fim, a qualidade e eficácia dos recursos existentes em cada instituição. O primeiro pode facilmente ser feito com base nos dados recolhidos (apresentados abaixo), que nos apresentam os tipos de dispositivos imersivos e interativos existentes, bem como os seus conteúdos. O segundo exige um estudo mais aprofundado, caso a caso, que permita perceber a importância de cada uma das instituições que recorrem a avanços na área da imersividade e interatividade. Apesar da sua relevância, este estudo não é abordado neste capítulo, constituindo matéria para uma publicação futura.

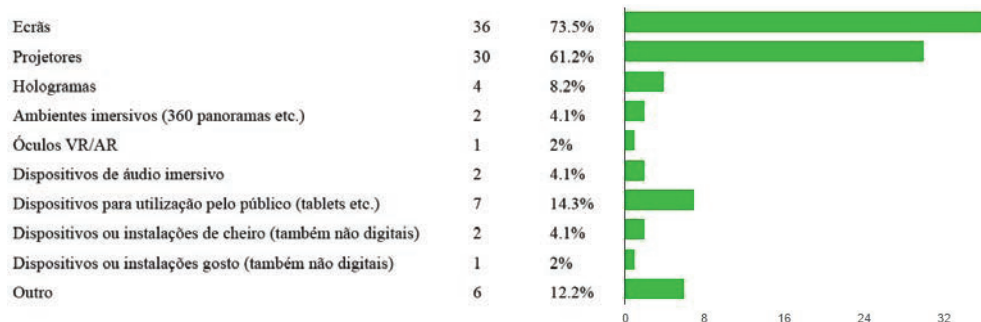
INSTRUMENTOS DE COMUNICAÇÃO DIGITAL

IMERSIVIDADE ESPACIAL BASEADA EM DISPOSITIVOS

Neste contexto, a imersividade espacial baseada em dispositivos descreve a forma como um dispositivo pode criar um maior ou menor grau de envolvimento do público com o conteúdo. O Gráfico 4 mostra que os monitores ainda são bastante utilizados em diversas instituições, seguidos dos projetores 2D. Podemos ver que são muito poucas as instituições que adotam novos tipos de representações imersivas. Contudo, existe uma pequena percentagem de instituições que utilizam tablets e outros dispositivos móveis fornecidos ao público. Estes dispositivos de carácter imersivo podem ser considerados dispositivos individuais, tais como os smartphones.

No Gráfico 4 considerámos ainda, para comparação, outros tipos de estimulação dos sentidos, cuja utilização contínua a ser bastante reduzida. É importante sublinhar que a imersividade espacial baseada em dispositivos não coincide automaticamente ao completo envolvimento com os conteúdos. Conforme Tilden (2007) tem argumentado, a interpretação baseia-se na experiência, envolvendo os níveis cognitivo e corporal (ALMEIDA, 2004, p. 1-25).

GRÁFICO 4 - DISPOSITIVOS DE COMUNICAÇÃO COM PÚBLICO NAS INSTITUIÇÕES, NICOLA SCHIAVOTTIELLO 2016

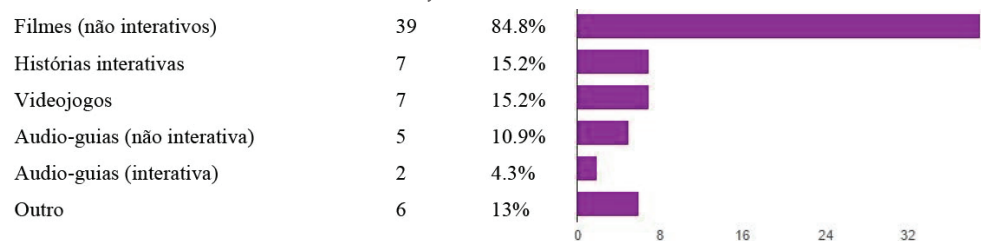


Fonte: Schiavottiello (2016).

MÉTODOS INTERATIVOS E NÃO INTERATIVOS

Torna-se evidente no Gráfico 5 que os conteúdos visuais não interativos, como os filmes, estão presentes na maioria das instituições, sendo poucas aquelas que apresentam formas de comunicação diferentes, nomeadamente videojogos e histórias interativas. Neste caso, será necessário explorar em maior detalhe as motivações que estão por trás deste tipo de escolha, que podem ir desde a falta de recursos, à falta de interesse, passando por uma escolha fundamentada ou por uma combinação de todos estes fatores. Esta questão será aprofundada mais à frente, na discussão dos resultados.

GRÁFICO 5 - MÉTODOS DE COMUNICAÇÃO DIGITAL, NICOLA SCHIAVOTTIELLO 2016



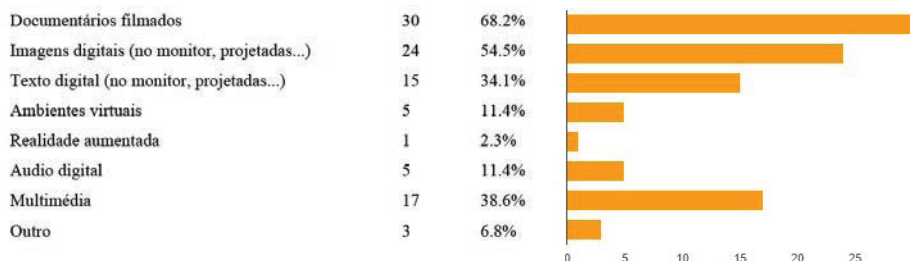
Fonte: Schiavottiello (2016).

FORMAS DE COMUNICAÇÃO INTERATIVAS E NÃO INTERATIVAS

O Gráfico 6 apresenta as técnicas utilizadas para transmitir os conteúdos ao público, diferenciando do anterior no sentido em que se centra apenas nas técnicas

utilizadas pelos diferentes dispositivos. Por exemplo, uma determinada história pode ser contada através de qualquer uma das técnicas indicadas, o que alterará a forma como o público final a irá receber. Os documentários filmados são, mais uma vez, a técnica mais utilizada, seguida das imagens estáticas e, finalmente, das técnicas multimédia que, neste caso, são considerados como uma combinação das duas anteriores. O texto digital é bastante utilizado, passando a ideia de que ainda não é uma escolha obsoleta.

GRÁFICO 6 - TÉCNICAS DIGITAIS USADAS PARA COMUNICAR OS CONTEÚDOS AO PÚBLICO, NICOLA SCHIAVOTTIELLO 2016

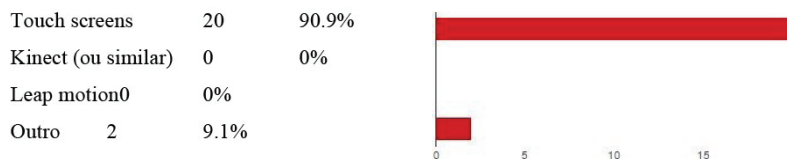


Fonte: Schiavottiello (2016).

INTERFACES INTERATIVAS

As interfaces interativas permitem-nos interagir com conteúdos digitais a partir do mundo real. O Gráfico 7 apresenta as diferentes opções utilizadas para criar interações com os conteúdos apresentados. Como podemos observar, o ecrã táctil é quase exclusivamente a única interface utilizada. Apesar de o ecrã táctil poder ser visto como uma solução inovadora, convém não esquecer que foi implementados nos museus há quase 20 anos como *interactive kiosk* (GRIFFITHS, 1999). As versões mais recentes incluem novos recursos, tais como as funcionalidades multi-toque e a sensibilidade à pressão. Será necessário averiguar os motivos pelos quais não são utilizadas outras soluções e se os conteúdos propostos apenas se adaptam a este tipo de interface.

GRÁFICO 7 - INTERFACES INTERATIVAS



Fonte: Schiavottiello (2016).

PREDISPOSIÇÃO PARA AMBIENTES VIRTUAIS

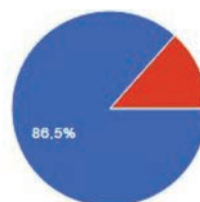
O Gráfico 8 apresenta as respostas a uma pergunta orientada que pretendia avaliar a propensão das instituições para a utilização de técnicas virtuais de transmissão de conteúdos. Como podemos ver, a percentagem é bastante elevada. Convém não esquecer que, em termos de património cultural, a região na qual estamos a trabalhar está pode ser bastante centrada em formas mais tradicionais e convencionais de comunicação (como pode-se observar mais abaixo non levantamento de outra soluções como meio de comunicação).

GRÁFICO 8 - INTERFACES INTERATIVAS

Sim	45	86.5%
Não	7	13.5%

Se não, justifique

- Preciso de conhecer melhor as possibilidades e impacto na mensagem
- Não creio que se adequa ao espaço físico existente.
- As salas que constituem o museu são muito pequenas para que seja possível fazer uma apresentação com qualidade.
- Neste momento não nos parece adequado. Será mais importante absorver e conhecer a realidade.
- Não está previsto a apresentação em ambientes virtuais, o conteúdo é apresentado em PowerPoint.
- Por um lado, por serem muito dispendiosos e, por outro, porque não se coaduna com os conteúdos artísticos da programação do Fórum. Também seriam necessários espaços especiais para estes ambientes que, neste momento, estão destinados a exposições temporárias.



Fonte: Schiavottiello (2016).

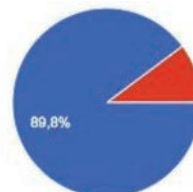
O Gráfico 9 mostra-nos a disponibilidade das instituições para a utilização de ambientes virtuais na criação de conteúdos de forma simples, sem recurso a serviços especializados. Mais do que o número de respostas positivas, que ultrapassou o do gráfico anterior, será interessante analisar cada um dos casos individualmente para averiguar as razões pelas quais as instituições estão indisponíveis para utilizar estas técnicas.

GRÁFICO 9 - UTILIZAÇÃO DE AMBIENTES VIRTUAIS COM ATUALIZAÇÃO DOS RESPETIVOS CONTEÚDOS SIMPLIFICADA: SIM OU NÃO

Sim	44	89.8%
Não	5	10.2%

Se não, justifique

- Preciso de conhecer melhor as possibilidades e o impacto na mensagem.
- Pelas razões apresentadas na resposta anterior.
- Mais uma vez, julgo que não seria eficiente nesta altura.
- Pelas razões apresentadas na pergunta anterior.
- Pelas mesmas razões anteriores.



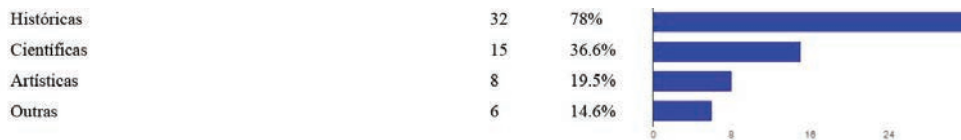
Fonte: Schiavottiello (2016).

Estas duas últimas perguntas foram incluídas de forma a ligar esta investigação a outra mais específica, a propor futuramente, com o objetivo de determinar se os ambientes digitais e virtuais são essenciais e bem-aceites pelas instituições culturais como forma de comunicação com os respetivos públicos.

INFORMAÇÕES DE SUPORTE COMPLEMENTARES

O Gráfico 10 mostra que as temáticas abordadas pelas instituições, com novas tecnologias digitais, nas suas coleções e exposições são essencialmente as históricas, seguidas das científicas e artísticas. Contudo, neste caso, precisamos de ser bastante cuidadosos no que toca à definição destas disciplinas abrangentes, pois muitas subdisciplinas são transversais a duas ou três delas. Por exemplo, seria básico integrar a arqueologia na história, quando é de conhecimento geral que se situa na interseção das três. Assim, cada instituição será responsável por destacar um ou outro aspeto da disciplina em causa.

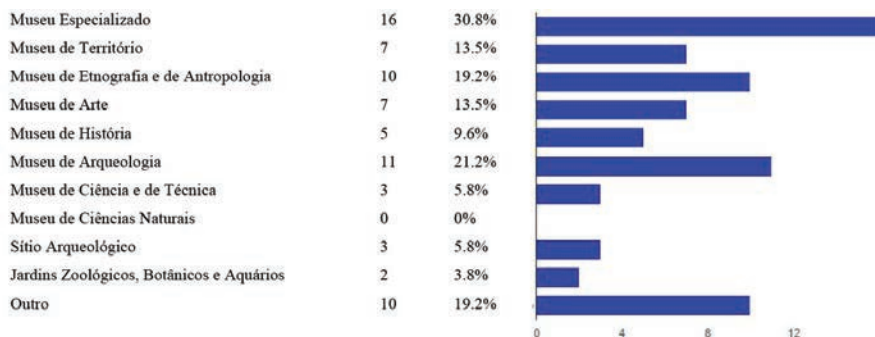
GRÁFICO 10 - TEMÁTICAS



Fonte: Schiavottiello (2016).

O Gráfico 11 apresenta os tipos de museus existentes no território em estudo com novas tecnologias de comunicação digital. As categorias utilizadas seguem as recomendações da Direção Regional da Cultura do Alentejo (DRCA). A primeira categoria considerada foi a dos Museus Especializados (p. ex.: Museu do Brinquedo, Évora; Museu do Vinho, Redondo; Museu do Relógio, Serpa e Évora), sendo também aquela que abrange um maior número de instituições, talvez devido à diversidade de temáticas presentes na região do Alentejo. Contudo, se considerarmos a disciplina mais promovida, observamos que os museus de história e de arqueologia são bastante relevantes. Em último lugar encontramos um número significativo de instituições de tipo diverso, o que levanta a necessidade da realização de estudos que permitam clarificar o tipo de instituições em causa, criando, se necessário, novas categorias.

GRÁFICO 11 - TIPOLOGIA DO MUSEU: OACERVO



Fonte: Schiavottiello (2016).

TIPOLOGIAS DE COLEÇÕES

O Gráfico 12 apresenta os diferentes tipos de coleções promovidos pelas diversas instituições com novas tecnologias digitais. Mais uma vez, a etnografia e a arqueologia são as categorias mais relevantes, embora se verifique também uma relevante presença da fotografia, algo que pode estar relacionado com a ênfase dada à imagem da população e dos lugares da região. Seria muito interessante analisar este tipo de coleções em maior detalhe. O património industrial está também bastante bem representado, seguido das restantes categorias.

GRÁFICO 12 - TIPOLOGIA(S) COLEÇÃO(ÕES) E FORMAS DE INCORPORAÇÃO

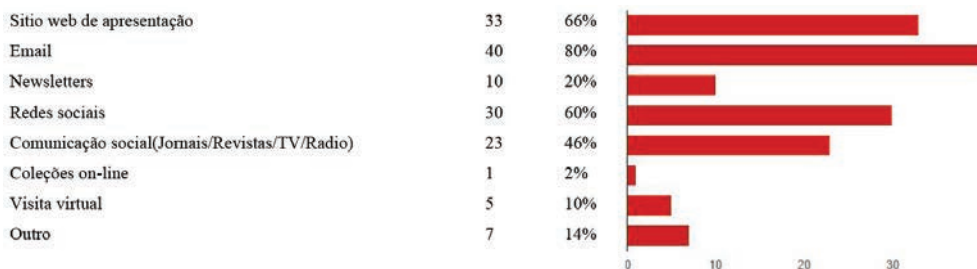


Fonte: Schiavottiello (2016).

O Gráfico 13 descreve os instrumentos externos de comunicação com o público, mostrando que, a par dos meios de comunicação tradicionais e não digitais, existe uma utilização generalizada de meios digitais: sítios web, redes sociais, newsletters etc.

É importante ter em conta que, apesar de o e-mail ser o instrumento mais utilizado, trata-se de uma categoria ambígua, pois pode ser incorretamente interpretada pelos inquiridos e incluir utilizações internas de natureza administrativa ou pessoal. Por esse motivo, deveria ser um objeto de estudo mais aprofundado.

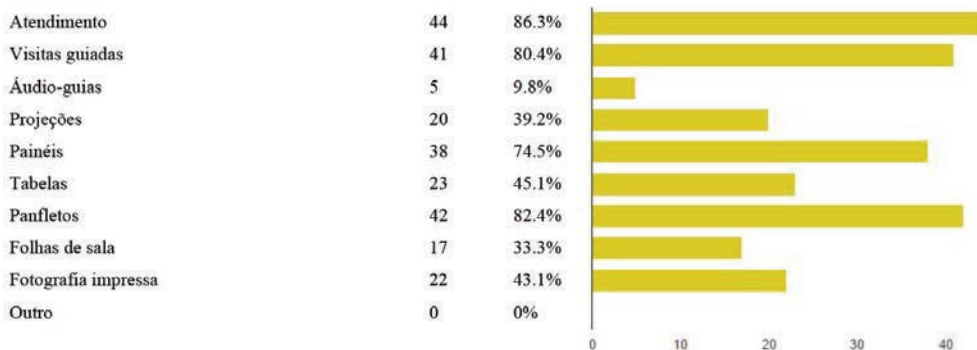
GRÁFICO 13 - INSTRUMENTOS EXTERNOS DE COMUNICAÇÃO COM OS DIFERENTES PÚBLICOS



Fonte: Schiavottiello (2016).

No Gráfico 14 estão todos os outros tipos de recursos, ou seja, os recursos não digitais. Neste caso, podemos observar que os mais relevantes são o atendimento pessoal por parte dos funcionários, seguido pelas visitas guiadas, folhetos e painéis, sendo todos os outros recursos bastante menos significativos.

GRÁFICO 14 - MAIOS DE COMUNICAÇÃO ANALÓGICOS



Fonte: Schiavottiello (2016).

OUTROS TIPOS DE TÉCNICAS DE COMUNICAÇÃO (MAIS TRADICIONAIS)

São várias as instituições que oferecem um grande número de tipos de comunicação não convencionais e atividades aparte de aquelas com métodos digitais. A descrição abaixo contempla a variedade das atividades existentes, revelando um grande interesse na promoção do Património Cultural de forma inovadora e inventiva. Neste caso, o objetivo seria de favorecer o envolvimento do público, não como recetor passivo, mas como participante interativo (GRAHAM, 2015).

- Meios de comunicação menos usuais
- Apresentação de livros; exposições temporárias coexistentes com a exposição permanente; recriações/evocações históricas; concertos; visitas temáticas; conferências; atividades diversas com objetivo educativo e lúdico, muitas das vezes a propósito de uma peça móvel, integrada ou imóvel, ou de um acontecimento – aspeto do património imaterial inerente ao museu. Peça do mês na Agenda Cultural do Município
- Atividades temáticas desenvolvidas pelo Serviço Educativo.
- Organização de concertos, exposições, conferências e debates, “Entre Agulhas e Dedais” (reunião bissemanal de costureiras e bordadeiras), “Mãos na Massa” (amassadura, cozimento e degustação de pizzas, bolos e bolachas em família ou grupo de crianças), Noites no Museu, Férias de Natal e Páscoa no Museu, recriação do ciclo do milho dos anos 30 do século passado (trabalhos do milho, com realização de documentário etnográfico) etc.
- Organização de eventos, como por exemplo: Atuação de Grupo de Cantares (Jornadas Europeias do Património 2014); Exposições Temporárias (Exposição Comemorativa do “Centenário do mestre Artur Azedo”; “Salgueiro Maia: um legado ainda não cumprido”; “40 Anos da Reforma Agrária”). Integração de Atividades em programas de outros serviços: exemplo disponibilidade de visitas guiadas no Programa Anima náutico;

Programa Municipal ACREDITA – Programa de estimulação à leitura. Redução de preços para Unidades de Alojamento da área do Município ou outras fora do Município, mas protocoladas para o efeito, as unidades hoteleiras distribuem (aquando da entrada) um voucher aos seus clientes.

- O núcleo de Arte Islâmica do Museu de Mértola dispõe de uma maleta pedagógica – o Baú de Aladino – destinada a crianças do Pré-Escolar e 1º Ciclo que complementa a visita.
- Público em geral – Visitas acompanhadas e distribuídas em quatro períodos: Manhã – 10h e 12h /Tarde – 15h e 17h; Grupos – visitas guiadas com marcação.
- No espaço museológico do Castelo de Serpa temos promovido algumas exposições temáticas e atividades pedagógicas com as Escolas. No designado Museu de Etnografia em parceria com o Musibéria realizámos representações de teatro com jovens artistas (no âmbito da Noite dos Museus). Atividades com a Academia Sénior. Atividades com Escolas e atividades no âmbito dos estágios integrados.
- Congressos de arquitetura militar, congressos de museologia militar, visitas guiadas temáticas, apresentação de livros.
- Organização de workshops; criação de músicas; documentários sobre determinados temas.
- Organização de eventos especiais de natureza variada – música, teatro e outros.
- Organização de Exposições temporárias ligadas ao contexto do Museu.
- Exposições temporárias com parceiros eventos de charme jantares de solidariedade.

- Temporadas de concertos, jornadas temáticas, apresentação de livros.
- Organização de eventos ligados a dias comemorativos, como o Dia da Mãe. Tertúlias com concerto no Museu. Representações. Abertura do Museu à noite.
- Demonstrações de sabonetes de glicerina. Teatros de fantoches.
- Organização de eventos: Realização de ações de formação “Livro Antigo -Restauro e Encadernação”; “Conservação preventiva de documentos”: Atividades de divulgação: “Scriptorium do Convento e mesteres de ofícios cristãos” (integrado na Feira Medieval Ibérica – Avis 2015). Redução de preços de entrada: atribuição de voucher aos clientes das unidades hoteleiras da área do Município ou outras entidades protocoladas (fora da área do Município). Participação em Programas Municipais.
- O plano de atividades anual é apresentado publicamente uma vez por ano. Foram feitas várias campanhas de divulgação menos usuais, por exemplo, uma no dia internacional dos museus que juntou várias instituições da cidade de Évora e que numa arruada, ao som da giga bombos, foram distribuídos vouchers gratuitos para uma entrada em todas as instituições que integraram o evento.
- Serviço Educativo dos Museus – atividades diversas no âmbito da temática do museu.
- Representações, teatro.
- Sessões de contos e atividades específicas desenvolvidas pelo Serviço Educativo.
- Mesa interativa com aplicação multimédia.

DISCUSSÃO, CONCLUSÕES E ABORDAGENS FUTURAS

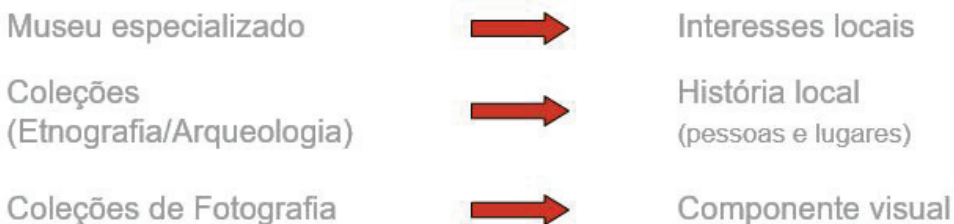
Este trabalho é uma pesquisa das estratégias de comunicação digital utilizadas pelos museus e instituições culturais da região do Alentejo (Portugal).

O mapeamento de todas as instituições culturais existentes no território em estudo constituiu-se como um exercício bastante importante. A localização dos museus e outras instituições é especialmente importante, pois permite-nos determinar se são facilmente acessíveis ou se estão inseridos em ambientes mais isolados. É interessante visualizar, através do mapa (Figuras 4 e 5) que, apesar de a maioria das instituições se localizar em zonas acessíveis, os recursos digitais estão distribuídos uniformemente pela região, confirmando a noção de que as cidades maiores contam com um maior número de instituições com recursos digitais, mas que estes também estão bem representados noutros locais. Neste momento, este aspeto apenas pode ser visualizado num mapa geral de pontos quentes. Contudo, seria significativo, no futuro, analisar de forma mais aprofundada, com base em métodos analíticos de (Sistema de Informação Geográfica (SIG), as distâncias e a acessibilidade dos referidos recursos. Um exercício ainda mais interessante é o de comparar o resultado obtidos em termos de recursos digitais com o tipo de coleções existentes no território.

Esta análise sugere que as coleções etnográficas são simultaneamente as mais exploradas e apresentadas ao público de forma mais tradicional, através da utilização de filmes em ecrãs básicos, oferecendo assim um grau de interatividade bastante reduzido. Em geral, o recurso à interatividade resume-se à presença de ecrãs tácteis, sendo que outras possibilidades, tais como interações visuais mais imersivas ou outras interações sensoriais, raramente são exploradas. O facto de o método de comunicação mais utilizado ser o filme não interativo pode ainda indicar que talvez o público tenha mais disponibilidade para aderir a este tipo de meio do que a outros tipos de técnicas interativas. Neste momento, podemos afirmar que a museografia alentejana se caracteriza essencialmente por exposições visuais vocacionadas para a promoção e divulgação de histórias locais. Apesar de esta estratégia ter sido bem-sucedida no passado, no contexto da comunicação na área do património, consideramos que a narrativa digital pode constituir-se como o próximo passo no sentido de um novo tipo de comunicação (YANNIS, 2013). A Figura 6 apresenta os dados de forma linear.

FIGURA 6 - DESCRIÇÃO DA TIPOLOGIA DAS INSTITUIÇÕES APRESENTADAS

Interpretação da tipologia dos museus



Fonte: Schiavorttiello (2016).

Tendo em conta o facto de a idade média da população alentejana se situar entre os 43 e os 50 anos de idade (seria também interessante analisar a idade média dos turistas), talvez a utilização de métodos de comunicação e narração não interativos seja uma opção ajustada (partindo do pressuposto que os meios interativos são mais ajustados a públicos mais jovens). No entanto, os dados podem também indicar a necessidade da implementação de métodos interativos imersivos, com apenas alguns exemplos pontuais distribuídos por todo o território, vocacionados para a comunicação de assuntos de carácter essencialmente científico. É o caso dos diversos centros de Ciência Viva, como o do Lousal³, de Extremoz⁴ e o de Campo Maior⁵. Neste caso, seria necessário analisar detalhadamente as motivações subjacentes a este tipo de escolha, que podem ir da falta de recursos à falta de interesse, passando por uma escolha fundamentada ou por uma combinação de outros fatores. Considerando as respostas à última pergunta do questionário, pensamos poder excluir a falta de interesse. As tecnologias 3D implicam a reconstrução de uma estória a partir de uma perspetiva exclusivamente digital (embora seja possível inserir imagens e vídeos reais).

Por fim, se analisarmos o esforço que foi feito pelas várias instituições, podemos afirmar que a tecnologia digital é utilizada sobretudo em meios de promoção (e-mail, sítios web, redes sociais). De facto, as tecnologias digitais que

³ Disponível em: <<http://lousal.cienciaviva.pt/home/>>. Acesso em: 15 jun. 2016.

⁴ Disponível em: <<http://www.ccvextremoz.uevora.pt/home/>>. Acesso em: 15 jun. 2016.

⁵ Disponível em: <<http://www.delta-cafes.pt/pt/empresa/centro-de-ciencia-do-cafe/>>. Acesso em: 15 jun. 2016.

fomentam a fruição e a vivência do museu escasseiam. As técnicas existentes são mais representativas das que eram utilizadas em 1995/2005 do que das de tipo imersivo. A Figura 7 apresenta os diversos recursos por ordem de utilização.

FIGURA 7 - DESCRIÇÃO DOS INSTRUMENTOS DE COMUNICAÇÃO DAS INSTITUIÇÕES APRESENTADAS

Interpretação dos instrumentos de comunicação (Ordem de utilização)

1 – Email	→	Comunicação bidirecional
2 – Sítio web	→	Comunicação passiva
3 – Rede sociais	→	Comunicação ativa
4 – Meios tradicionais	→	Comunicação unidirecional
5 – Visitas virtuais	→	Comunicação imersiva
6 – Coleções online	→	Comunicação visual

Fonte: Schiavorttiello (2016).

Esta visão geral dos recursos digitais utilizados na divulgação do Património Cultural na região do Alentejo, no sul de Portugal, corresponde apenas à fase preliminar de um estudo aprofundado que será realizado, num futuro próximo, pela DRC do Alentejo. Essa análise irá seguramente contribuir para uma melhor discussão destes tópicos ao introduzir uma comparação entre muitos outros fatores, tais como a tipologia específica das entidades em estudo, os seus públicos e, de forma mais geral, toda a informação necessária acerca dos recursos não digitais.

Além disso, o estudo tornar-se-á mais abrangente através da comparação entre os métodos de comunicação digital existentes nos museus do Alentejo e alguns dos projetos mais avançados implementados noutras regiões do país (como o World of Discoveries⁶ no Porto) e da Europa.

Será ainda necessário criar um questionário para o público de uma amostra de museus e instituições culturais alentejanas que utilizam técnicas de comunicação

⁶ Disponível em: <<https://www.worldofdiscoveries.com/>>. Acesso em: 15 jun. 2016.

digitais, com o objetivo de perceber se a tecnologia implementada está operacional e se é útil e apelativa para os visitantes.

Está ainda previsto o estudo da implementação de uma plataforma simples para a criação e atualização de conteúdos digitais, que permitirá perceber se as instituições culturais são propensas à utilização deste tipo de comunicação digital (SCHIAVOTTIELLO, 2015).

Em suma, sentimos a necessidade de analisar o estado da arte das metodologias de comunicação digital nos museus e instituições de preservação do Património Cultural do Alentejo, em Portugal, com o objetivo de disponibilizar uma base de dados georreferenciada alargada a todas as instituições da região, distinguindo as que promovem a interpretação e comunicação através de meios digitais das que não o fazem. A partir deste estudo, tentámos perceber, através de uma análise quantitativa, o modo como os dispositivos de comunicação digital são explorados neste território.

Esperamos ter colaborado para caracterizar a situação atual das tecnologias de visualização cujas metodologias são adequadas ou cujas necessidades precisam de ser reavaliadas. Gostaríamos que este trabalho também pudesse contribuir para que os recursos tecnológicos venham a ser distribuídos pelos locais que mais deles necessitam, para que as comunidades locais possam reviver as suas próprias tradições através de novas abordagens e para que os visitantes da região possam apreciar o seu significado através de novas experiências intelectuais e emocionais.

REFERÊNCIAS

ALISON, G. *Media Technology and Museum Display: A Century of Accommodation and Conflict*. Disponível em: <<http://web.mit.edu/comm-forum/papers/griffiths.html>>. Acesso em: 15 jun. 2016.

ALMEIDA, M. V. de O corpo na teoria antropológica. *Revista de Comunicação e Linguagens*, v. 33, p. 1-25, 2004.

ASSOCIAÇÃO PORTUGUESA DE MUSEOLOGIA. *Associação Portuguesa de Museologia*. 2011. Disponível em: <<http://www.apom.pt/pagina,47,73.aspx>>. Acesso em: 15 jun. 2016.

BERGER, A. A. *Media and Communication Research Methods: An Introduction to Qualitative and Quantitative Approaches*. Thousand Oaks, Calif: Sage Publications, 2000.

BLACK, G. *The Engaging Museum: Developing Museums for Visitor Involvement*. Heritage. London; New York: Routledge, 2005.

CCDR Alentejo. *Estratégia Regional de Especialização Inteligente do Alentejo*. 2014. Disponível em: <http://www.ccdr-a.gov.pt/docs/ccdra/alentejo2020/EREI_Alentejo_vf.pdf>. Acesso em: 15 jun. 2016.

GRAHAM, B. *The Engaging Museum: Developing Museums for Visitor Involvement*, Heritage. London; New York: Routledge, 2005. p. 2-3.

GRIFFITHS, A. *Media Technology and Museum Display: A Century of Accommodation and Conflict*. 1999. Disponível em: <<http://web.mit.edu/commforum/legacy/papers/griffiths.html>>. Acesso em: 15 jun. 2016.

INSTITUTO NACIONAL DE ESTATÍSTICA. *CENSOS 2011 Alentejo*. 2012. Disponível em: <http://censos.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=156656957&att_display=n&att_download=y>. Acesso em: 15 jun. 2016.

IOANNIDIS, Y.; EL RAHEB, K.; TOLI, E. KATIFORI, A.; BOILE, M.; MAZURA, M. One Object Many Stories: Introducing ICT in Museums and Collections through Digital Storytelling. *IEEE*, p. 421-424, 2013. <http://dx.doi.org/10.1109/DigitalHeritage.2013.6743772>.

MINISTÉRIO DA ECONOMIA E DO EMPREGO. Turismo de Portugal. *Plano Estratégico Nacional do Turismo 2013-2015*. 2013. Disponível em: <http://www.turismodeportugal.pt/Português/turismodeportugal/publicacoes/Documents/PENTurismo_07out14.pdf>. Acesso em: 15 jun. 2016.

PROGRAMA OPERACIONAL REGIONAL ALENTEJO 2020. 2014. Disponível em: <http://www.ccdr-a.gov.pt/docs/ccdra/alentejo2020/PORALENTEJO_2020_vf.pdf>. Acesso em: 15 jun. 2016.

PRESIDÊNCIA DO CONSELHO DE MINISTROS. *Resolução do Conselho de Ministros nº 190/2005*. 2005. Disponível em: <http://www.iapmei.pt/resources/download/RCM_190_2005.pdf>. Acesso em: 15 jun. 2016.

SCHIAVOTTIELLO, N. ArkTeller: A New 3D Real-Time Storytelling Platform for Cultural Heritage Interpretation. *IEEE*, p. 753-754, 2015. Disponível em: <<https://ieeexplore.ieee.org/document/7419620>>. Acesso em: 01 out. 2020.

TILDEN, F. *Interpreting Our Heritage*. 4th ed. Chapel Hill: University of North Carolina Press, 2007.

TURISMO DO ALENTEJO. *Documento Estratégico Turismo do Alentejo 2014-2020*. 2014. Disponível em: <http://www.visitalentejo.pt/fotos/editor2/Quem%20somos/alentejo2014_2020_documento_estrategico_turismo.pdf>. Acesso em: 15 jun. 2016.

YANNIS, I. *et al.* One Object Many Stories: Introducing ICT in Museums and Collections through Digital Storytelling. *IEEE*, p. 421-424, 2013. Disponível em: <<https://ieeexplore.ieee.org/document/6743772>>. Acesso em: 01 out. 2020.

A obra potencial – para uma edição digital dos projetos editoriais de Fernando Pessoa

Pedro Sepúlveda
Ulrike Henny-Krahmer

LISTAS DE PROJETOS EDITORIAIS

Para além do propósito de projetar uma futura publicação da obra, as inúmeras listas de projetos editoriais concebidas por Fernando Pessoa, reveladoras de uma obsessão editorial do poeta, conferiam à obra do autor um sentido de conjunto que sem estas não possuiria. Leitores que julguem encontrar em Pessoa a figura do poeta que se recolhe no anonimato e opta por não publicar a sua obra são porventura surpreendidos por tal afirmação. No entanto, a elaboração de listas de projetos acompanha em todos os momentos a escrita de Pessoa, sendo constitutiva do seu modo de escrita. Paralelamente à elaboração das principais obras, Pessoa agrupava nestas listas projetos de edição e publicação, que conferem à sua obra uma dimensão potencial, excedendo o que foi efetivamente concretizado¹.

As listas de projetos editoriais permitem acompanhar o desenvolvimento de títulos, as atribuições de obras a figuras autorais e o estabelecimento de conjuntos ou coleções ao longo do tempo, traçando uma cronologia a que não temos acesso apenas com base nos textos. Neste sentido, estes documentos possuem um significado que vai muito além da mera dimensão editorial, possuindo um forte sentido literário, neles se definindo afinal os fundamentos da obra de Fernando Pessoa. Importa, neste âmbito, distinguir entre as listas de projetos editoriais, que elencam vários projetos distintos, e planos que estruturam uma determinada obra. O *corpus* das listas de projetos editoriais distingue-se assim do dos planos, que lhe é próximo, nomeando projetos de índole distinta e estabelecendo relações entre eles. Estas listas de projetos são decisivas para um entendimento da estruturação da

¹ Cf. SEPÚLVEDA, 2013, p. 67-76.

obra enquanto conjunto, ao longo do tempo, entendimento este que os planos, ao focarem uma só obra, não permitem.

Apesar de as listas de projetos editoriais serem decisivas para uma compreensão da obra de Fernando Pessoa, estas nem sempre têm sido privilegiadas pela tradição crítica e editorial. O trabalho pioneiro de recolha e estudo destes documentos foi realizado por Jorge Nemésio, em *A obra poética de Fernando Pessoa: estrutura de futuras edições* (1958). Nemésio tinha primeiramente em vista, como indica o título, uma “estrutura das futuras edições”, que se deveria basear, no seu entender, nos “esboços, planos e apontamentos temáticos de livros, em português, existentes no espólio de Fernando Pessoa”². O seu volume reúne assim um conjunto destes esboços, planos e listas de projetos, apenas relativos à poesia em português de Pessoa, vindo criticar as edições publicadas anteriormente por procederem a uma seleção e ordenação subjetiva dos textos, e terá tido impacto no entendimento editorial posterior da obra. No final dos anos oitenta, Cunha (1987) referia-se aos planos e projetos editoriais de Pessoa como tratando-se de “uma velha questão”, concebendo uma edição de conjunto destes documentos, nunca concretizada. Na linha de Nemésio, Cunha (1987) sublinhava a sua importância enquanto fundamento para uma adequada edição da obra, exemplificando o seu potencial informativo a propósito de poemas de Alberto Caeiro. Seabra (1993) destacou igualmente a sua relevância no âmbito de uma edição crítica da obra, capaz de descrever a sua génese e as transformações de que foi alvo ao longo do processo de escrita. Qualquer destas posições críticas foca apenas uma, decerto importante, dimensão funcional dos projetos editoriais, na medida em que estes contribuem para um adequado entendimento editorial da obra e assim para melhores edições da mesma. No entanto, a sua importância não se esgota nesta dimensão, sendo absolutamente decisivo o seu carácter fundacional, o modo como neles se desenhava e dispunha um sentido de conjunto da obra que, ainda que pontual e mutável, determina o posicionamento das diferentes obras e as suas relações entre si.

PROPÓSITOS DE UMA EDIÇÃO DIGITAL

As edições de que hoje dispomos da obra de Pessoa seguem em maior ou menor grau as intenções editoriais expressas pelo poeta nos projetos e planos,

² Cf. NEMÉSIO, 1958, p. 65-88.

distinguindo-se entre si pelo modo como interpretam e privilegiam alguns dos projetos. As listas de projetos têm, pelo menos desde os anos 60, vindo a ser incluídas nas edições temáticas, no próprio corpo das edições ou em anexos. Falta ainda, no entanto, uma edição de referência destes documentos, que será disponibilizada em formato digital, no âmbito de uma colaboração entre o projeto *Estranhar Pessoa*, sediado na Universidade Nova de Lisboa (IELT), e financiado pela Fundação para a Ciência e Tecnologia (PTDC/CPC-ELT/4587/2012), e o Cologne Center for eHumanities (CCeH), centro de investigação da Universidade de Colónia. De um projeto global de revisão dos fundamentos da obra de Fernando Pessoa – *Estranhar Pessoa: um escrutínio das pretensões heteronímicas*³ –, nasce o projeto de edição das listas de projetos editoriais, enquanto documentos decisivos para o entendimento das categorias de autor, obra ou livro em Fernando Pessoa.

O *corpus* da edição é ainda estendido, já numa primeira fase⁴, à poesia publicada em vida em jornais e revistas, entre 1914 e 1935, de autoria de Alberto Caeiro, Álvaro de Campos, Ricardo Reis e Fernando Pessoa. Relativamente às listas de projetos, uma primeira fase reúne todas as listas que foi possível localizar no espólio de Fernando Pessoa à guarda da Biblioteca Nacional, assim como ainda em posse dos herdeiros, situadas entre 1913 e 1935. A extensão do *corpus* à poesia publicada em vida proporciona uma interação entre a dimensão potencial dos projetos e o que foi concretizado em termos de publicação, o que permite estabelecer ligações importantes entre os documentos, como será clarificado através de alguns exemplos em seguida.

Nesta edição digital das listas de projetos, até agora editadas de forma muito parcial e dispersa em alguns volumes, e reunindo pela primeira vez uma parte

³ Disponível em: <<http://estranharpessoa.com/>>. Acesso em: 20 abr. 2020.

⁴ A edição foi publicada, em finais de 2017, com o título *Edição Digital de Fernando Pessoa: Projetos e Publicações*. Disponível em: <<http://pessoadigital.pt>>. Acesso em: 20 abr. 2020. Inclui, tal como aqui anunciado, o *corpus* das listas de projetos de Fernando Pessoa, redigidas entre 1913 e 1935, assim como da totalidade da poesia publicada em vida pelo autor, entre 1914 e 1935. A indicação “BETA”, inserida na página de abertura, aponta para aspetos a rever relativamente ao *corpus* disponibilizado das listas, que ainda só se encontra parcialmente disponível, estando prevista para breve a sua disponibilização total, paralelamente ao *corpus* da prosa publicada em vida pelo autor. A disponibilização completa de ambos está prevista para o final de 2020.

substancial de material inédito, colaboram vários membros das respectivas equipas, estando a coordenação editorial a cargo de Pedro Sepúlveda e a coordenação técnica de Ulrike Henny-Krahmer. Tratando-se de escritos que nunca foram pensados para integrar livros, mas se apresentam enquanto instrumentos de planeamento da obra, a edição beneficia amplamente das possibilidades oferecidas por um formato que em diversa medida se aproxima do propósito original destes documentos.

Esta edição oferece ao leitor uma combinação de diferentes modos de edição dos textos e propõe uma nova forma de crítica textual, no âmbito do estudo e edição da obra pessoana. São apresentadas, para cada documento do espólio, quatro formas de edição:

1. Transcrição diplomática, incluindo todas as variantes, hesitações e trechos posteriormente rejeitados pelo autor;
2. Primeira versão do texto, tal como estabelecida pelo autor e contemplando o desenvolvimento de abreviaturas;
3. Última versão do texto não rejeitada pelo autor, incluindo igualmente abreviaturas desenvolvidas;
4. Versão pessoal do texto, que permite ao leitor estabelecer a sua própria versão, com base numa escolha de uma das três versões anteriores, combinadas com as possibilidades de desenvolver ou não as abreviaturas e de introduzir ou não quebras de linha.

Dada a existência de variantes textuais nos documentos do espólio, anteriores edições em formato de livro definem-se com base na lição escolhida e na versão estabelecida de cada texto. Como é amplamente conhecido, algumas edições seguem a primeira lição do texto, outras a última, existindo ainda casos em que a escolha entre as variantes existentes é realizada com base num critério hermenêutico⁵. O presente trabalho proporciona diversas modalidades de acesso a cada texto, superando anteriores oposições entre critérios editoriais e dando conta do dinamismo da escrita do autor e das suas diferentes etapas.

As variantes textuais, os sublinhados e os trechos riscados são apresentados na sua totalidade na transcrição diplomática de cada texto, de um modo graficamente legível e prescindindo de qualquer simbologia suplementar. Em particular a

⁵ Cf. a este propósito LOPES (1992), GALHOZ (1993), CABRAL (2011), CASTRO (2013).

última versão, as omissões e abreviaturas são expandidas. No portal, a apresentação é feita através de uma estrutura de tipo *tab*, na qual é possível optar entre uma versão e outra, surgindo o fac-símile sempre ao lado da transcrição selecionada.

FIGURA 2 - EXEMPLO DE DIFERENTES TIPOS DE TRANSCRIÇÃO (DIPLOMÁTICA, PRIMEIRA VERSÃO, ÚLTIMA VERSÃO)

Documentos do ^{Neo-symbolismo, do} Futurismo ~~portuguez~~ e do Sensacionismo ~~portu-~~ guezes.

- / 1. Prefacio de Orpheu (1915) – Luiz de Montalvor.
- / 2. " " Centauro (191) – " " "
- / 3. Ultimatum ás Gerações portuguesas do seculo XX – Almada.
- / 4. Ultimatum de Alvaro de Campos.

Documentos do Futurismo e do Sensacionismo ~~portuguezes.~~

- / 1. Prefacio de Orpheu (1915) – Luiz de Montalvor.
- / 2. [Prefacio][de]Centauro (191□) – [Luiz][de][Montalvor]
- / 3. Ultimatum ás Gerações portuguesas do seculo XX – Almada.
- / 4. Ultimatum de Alvaro de Campos.

Documentos do Neo-symbolismo, do Futurismo e do Sensacionismo ~~portuguezes.~~

- / 1. Prefacio de Orpheu (1915) – Luiz de Montalvor.
- / 2. [Prefacio][de]Centauro (191□) – [Luiz][de][Montalvor]
- / 3. Ultimatum ás Gerações portuguesas do seculo XX – Almada.
- / 4. Ultimatum de Alvaro de Campos.

Fonte: Disponível em: <<http://www.pessoadigital.pt/pt/doc/CP602>>. Acesso em: 08 set. 2020.

A concretização de várias formas de transcrição é facilitada pelo modelo de edição digital escolhido, definido por Pierazzo (2015, p. 25) como *The-Source-and-the-Output-Model*. Na base da edição digital está a codificação no formato XML-TEI. Para esse efeito, a transcrição dos documentos originais é realizada, numa primeira fase, pelos editores, através de um programa de processamento de texto, e é depois transferida para o formato TEI-P5. Nos ficheiros TEI, os metadados referentes a cada documento (como a cota do arquivo, a data de criação, o género) estão armazenados no *TEI Header*, as ligações com os respectivos fac-símiles e a transcrição dos próprios textos encontram-se no *TEI Body*. O resultado do processo de codificação é a representação digital das fontes.

Contrariamente à edição preparada exclusivamente para a impressão em formato de livro, o foco da edição digital desloca-se da apresentação para a representação. Em lugar de o editor se ver forçado a optar por uma forma específica de transcrição e por uma forma tipográfica do texto, a edição digital do modelo *source-and-output* permite codificar as características dos documentos e textos em vários níveis e a partir de diversas perspetivas. A transformação do método e do trabalho editorial através da alteração dos suportes foi descrita por Sahle (2013, p. 126, tradução nossa do original alemão) da seguinte forma:

A discussão sobre os ‘conteúdos’ da edição foi sempre uma discussão sobre a versão tipográfica de um texto, sobre o *layout* de uma impressão e sobre a estrutura e a ordem de um livro. O livro sugeriu sempre a produção do texto final único, estático, e não a reflexão sobre a diversidade das formas da tradição e a sua possível multiplicidade de processamento e apresentação. Assim, a edição tradicional teve sempre de apontar para a construção e a produção mais do que para a reconstrução e a representação.⁶

Procurando representações distintas de uma mesma fonte, a edição digital é uma forma particularmente conveniente para o trato do espólio de Fernando Pessoa, em que os textos não assumem uma forma fixa e finalizada. A apresentação dos documentos no portal, seja numa forma única ou em várias formas diferentes e complementares, como neste caso, é feita numa segunda fase de transformação da codificação de base para um formato digital previsto para a publicação em linha. O

6

resultado não está constringido por uma única forma de apresentação: emergindo de uma mesma base, a edição digital proporciona múltiplas possibilidades de processamento e apresentação. A edição digital dos projetos editoriais de Fernando Pessoa apresenta, como resultado de todo este processo, as seguintes possibilidades de acesso aos textos:

1. Apresentação de diferentes versões do texto (diplomática, primeira, final, pessoal);
2. Apresentação sinóptica (texto – fac-símile);
3. Criação de múltiplos acessos à edição: navegação, hiperligação dos conteúdos, pesquisa;
4. Preparação analítica dos conteúdos: índices e visualizações.

Os ficheiros dos documentos TEI e das imagens são guardados numa base de dados XML nativo, neste caso a *eXist-db*⁷, para facilitar a sua administração e o seu processamento com tecnologias X. Os dados são transformados dinamicamente mediante scripts XQuery e XSLT, para produzir o portal em formatos habituais da Web: XHTML, CSS e JavaScript. Adicionalmente, as revisões dos dados e scripts estão administrados e preservados em *GitHub*, como forma de apoio a um trabalho inerentemente colaborativo.⁸

Como foi sugerido, o portal permite ainda outras modalidades de leitura, distintas daquelas que oferece o formato do livro, como sejam a pesquisa por um nome de autor, uma cota ou uma passagem específica. Motores de busca facilitam a pesquisa por conteúdos específicos e proporcionam distintos modos de acesso à informação contida nos textos. Um motor de busca simples presente em todas as páginas do portal permite procurar em todo o texto (*full-text search*).

Para além disso, pode-se fazer uma pesquisa avançada mais específica. A Figura 3 apresenta o exemplo de uma pesquisa por todos os documentos do género “lista editorial” que contém a menção do nome de autor “Álvaro de Campos”. Por detrás do motor de busca avançado está a codificação TEI. Combinam-se informações contidas nos metadados – como a data do documento, se o texto foi publicado ou não – com informações marcadas na transcrição do texto. No caso dos

⁷ Cf. SIEGEL; RETTER (2014).

⁸ Disponível em: <<https://github.com/cceh/pessoa>>. Acesso em: 20 abr. 2020.

nomes de autor, acrescenta-se a cada menção de Fernando Pessoa e dos heterónimos Alberto Caeiro, Álvaro de Campos e Ricardo Reis a função que o nome tem na lista editorial – função de autor, editor, tradutor ou tema de uma obra mencionada – o que permite posteriores pesquisas com focos particulares nos documentos do portal.

FIGURA 3 - MOTOR DE BUSCA

Pesquisa

autores

Fernando Pessoa
Alberto Caeiro
Álvaro de Campos
Ricardo Reis

prima a tecla Ctrl/CMD para seleccionar várias entradas

mencionado como:

- autor
- editor
- tradutor
- tema

A sua pesquisa foi:

Álvaro de Campos:
mencionado como: autor

gênero: Lista editorial

Resultados: 28

Ordem: cronológico alfabético

publicado & não publicado

gênero

Lista editorial
Nota editorial
Plano editorial
Poesia

prima a tecla Ctrl/CMD para seleccionar várias entradas

data

língua

pesquisa aberta

BNP/E3 133F-28v

BNP/E3 133F-36v

BNP/E3 144C-11r

BNP/E3 144Y-20r

BNP/E3 144Y-51V

BNP/E3 14E-5v

BNP/E3 15B4-48v

BNP/E3 180r

BNP/E3 22-77v

BNP/E3 44-47r

BNP/E3 48B-11r

BNP/E3 48B-34r

BNP/E3 48B-35r

BNP/E3 48B-64r

BNP/E3 48B-8r

BNP/E3 48C-22

BNP/E3 48C-24

BNP/E3 48C-28r

BNP/E3 48C-29r

BNP/E3 48D-13

BNP/E3 48D-5r

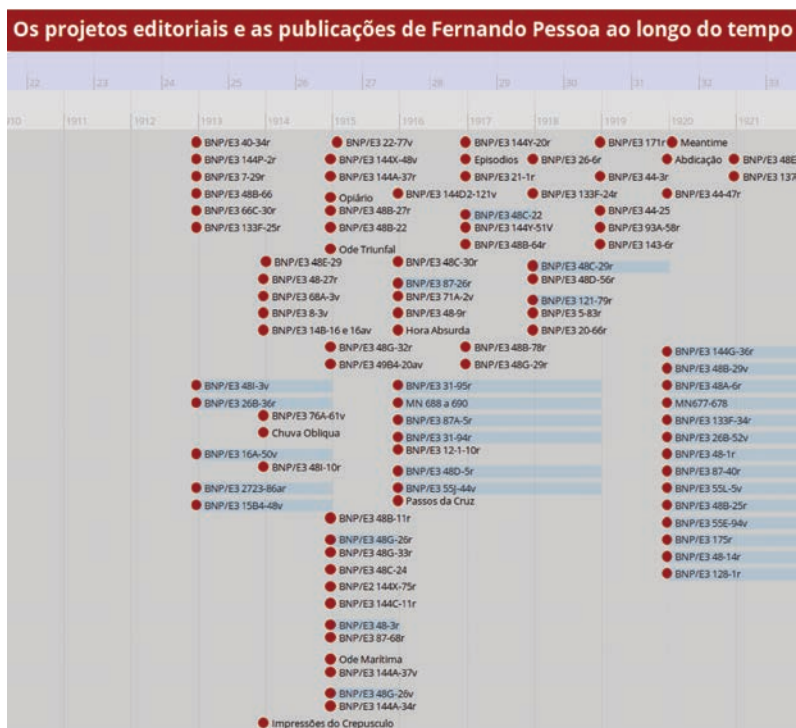
BNP/E3 48G-29r

BNP/E3 71A-2v

Fonte: Disponível em: <<http://www.pessoadigital.pt/pt/search>>. Acesso em: 30 out. 2020.

Uma forma de apresentação visual das informações marcadas na codificação de base é a linha do tempo que foi integrada ao portal. A Figura 4 mostra só uma parte dessa visualização. Na linha do tempo, é possível avançar cronologicamente e ver o número e o tipo de textos produzidos por Pessoa num determinado período. Clicando numa cota ou num título aparece uma caixa com informações adicionais que, seguindo o *link* na caixa, permite chegar ao respetivo texto. Já um recorte de toda a linha do tempo, aqui desde o ano 1913 até ao ano 1921, revela a relação entre o grande número de projetos editoriais do poeta e as poucas publicações em vida. A sua cronologia é estabelecida com base em elementos materiais e de conteúdo, procurando mapear a história dos projetos editoriais e das publicações de Fernando Pessoa. Notas concisas relativas ao lugar de publicação e características materiais dos escritos acompanham cada documento.

FIGURA 4 - LINHA DO TEMPO



Fonte: Disponível em: <<http://www.pessoadigital.pt/pt/timeline>>. Acesso em: 08 set. 2020.

Outra parte fundamental da edição digital e que vai além das possibilidades de uma edição impressa é a marcação semântica de entidades nos documentos, para possibilitar a exploração de relações entre diferentes documentos, por um lado, e entre documentos e publicações, por outro. Nos documentos são marcadas as ocorrências de nomes de pessoas, títulos de textos e nomes de jornais. Foi estabelecido um índice central de pessoas, obras e jornais, em que cada entidade obtém um identificador, uma chave única no âmbito da edição. As ocorrências e referências nos documentos estão ligadas às entradas no índice central, que representam entidades num sentido semântico. Assim, diferentes títulos de textos nas listas editoriais podem apontar para uma mesma obra no índice central. Por outro lado, as obras no índice servem de âncora para as publicações: a publicação de um poema é considerada uma referência a uma obra determinada, do mesmo modo que a entrada numa lista editorial pode referir-se a essa mesma obra. Deste modo, é possível estabelecer ligações entre os projetos editoriais e as publicações e traçar a evolução da obra concebida por Fernando Pessoa, sem que esta esteja limitada aos projetos de obra efetivamente realizados e publicados.

As ligações feitas entre os documentos e as entradas do índice central são também relevantes na apresentação das listas editoriais no portal. Na apresentação pré-definida, a transcrição fica à esquerda e o fac-símile à direita. A parte direita da sinopse pode ser substituída por um índice de pessoas, textos e jornais, que lista as referências que se encontram no documento selecionado, o que é demonstrado através de um exemplo na Figura 6:

FIGURA 6 - EXEMPLO DE UMA APRESENTAÇÃO SINÓPTICA DE TRANSCRIÇÃO E ÍNDICE (BNP/ E3 120-23^R)

The image shows a digital interface with a list of documents on the left and a semantic index on the right. At the top, there are four tabs: 'transcrição diplomática' (highlighted in red), 'primeira versão', 'última versão', and 'versão pessoal'. On the right side, there are two more tabs: 'Facsimile' and 'Índices' (highlighted in red). The left side displays a numbered list of six items:

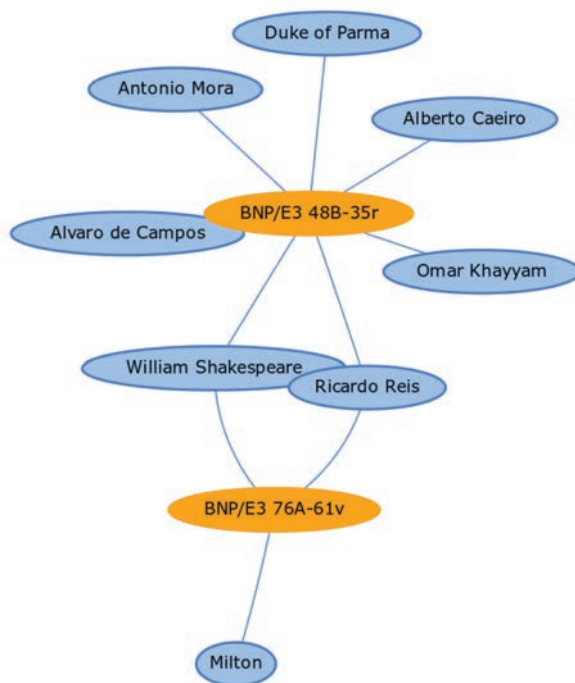
1. O Homem e a Cobra.
2. O Valle do Desasocego. (?)
3. De como Napoleão...
4. A O Segredo de Tse-i-la.
5. Carta do M² de Pombal.
6. Miscellanea - I, II, etc.

The right side is divided into two sections: 'Pessoas' and 'Textos'. Under 'Pessoas', there are two entries: 'Marquez de Pombal' and 'Napoleon', each with a small person icon. Under 'Textos', there are six entries, each with a small document icon: 'A Agua de Tse-i-la', 'O Segredo de Tse-i-la', 'Carta do Marquez de Pombal', 'De como Napoleão', 'Miscellanea', and 'O Homem e a Cobra'. The last entry, 'O Valle do Desasocego', is highlighted in red.

Fonte: Disponível em: <http://www.pessoadigital.pt/pt/doc/BNP_E3_120-23r>. Acesso em: 30 out. 2020.

No caso das pessoas mencionadas, a marcação semântica refere ainda outra dimensão da obra potencial. Além dos quatro nomes de autores principais que Pessoa utilizou e que estão no foco da edição (Caeiro, Reis, Campos e Pessoa) surgem mais nomes nas listas editoriais, de autores, figuras imaginadas, e pessoas reais, que são tema de determinado projeto editorial. Mediante o índice central é possível vislumbrar todos estes nomes, associando-os às listas em que foram referidos. Através das ligações entre os documentos revela-se uma rede de referências que, analisada no seu conjunto, pode estabelecer relações entre conjuntos de projetos editoriais que têm como referência as mesmas pessoas ou figuras. A Figura 7 mostra um exemplo de referências comuns a duas listas editoriais.

FIGURA 7 - EXEMPLO DE UMA REDE DE DOCUMENTOS E PESSOAS MENCIONADAS



Fonte: Elaborado pelos autores (2020)

A história das listas de projetos é também a de um desenvolvimento dinâmico da escrita e da sua fixação numa publicação ou através de uma proposta de organização

editorial. Estas fixações distintas são, ainda que mutáveis ao longo do tempo, pontualmente significativas, contribuindo para uma definição da obra de um ponto de vista não só editorial como poético e sistémico. Traçar a história de projetos e publicações pessoais com base numa edição digital significa assim analisar uma obra em construção, mas que alude a um propósito maior, o de um projetado estabelecimento do seu sentido enquanto todo plenamente constituído.

REFERÊNCIAS

- BNP/E3: Espólio de Fernando Pessoa à guarda da Biblioteca Nacional de Portugal.
- CABRAL, F. M. Fernando Pessoa e o original perdido. *Tágides, Revista de Literatura, Cultura e Arte Portuguesas*, São Paulo, v. 1, n. 1, p. 89-100, 2011.
- CASTRO, I. *Editar Pessoa*. Lisboa: Imprensa Nacional-Casa da Moeda, 2013.
- CUNHA, T. S. Planos e projectos editoriais de Fernando Pessoa: uma velha questão. *Revista da Biblioteca Nacional*, v. 2, n. 1, p. 93-107, 1987.
- GALHOZ, M. A. A fortuna editorial pessoana e seus problemas. In: SEABRA, J. A. (Ed.). *Mensagem: poemas esotéricos*. Madrid, Espanha: Scipione, 1993. p. 216-226. (Coleção Arquivos).
- LOPES, T. R. A crítica da edição crítica. *Revista Colóquio/Letras*, n. 125/126, p. 199-218, 1992.
- NEMÉSIO, J. *A obra poética de Fernando Pessoa: estrutura de futuras edições*. Salvador da Bahia: Progresso Editora, 1958.
- PIERAZZO, E. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham: Ashgate, 2015.
- SAHLE, P. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*. Teil 2: Befunde, Theorie und Methodik. Norderstedt: BoD, 2013.
- SEABRA, J. A. Introdução. In: _____. (Ed.). *Mensagem: poemas esotéricos*. Madrid [etc.]: CSIC, 1993. p. XXV-XL. (Coleção Arquivos).
- SEPÚLVEDA, P. *Os livros de Fernando Pessoa*. Lisboa: Ática, 2013.
- SEPÚLVEDA, P.; HENNY-KRAHMER, U. (Ed.). *Edição Digital de Fernando Pessoa. Projetos e Publicações*. Coordenação editorial por Pedro Sepúlveda, coordenação técnica por Ulrike Henny-Krahmer. Lisboa e Colónia: IELT, Universidade Nova de Lisboa

e CCeH, Universidade de Colónia 2017. Disponível em: <<http://www.pessoadigital.pt>>
. Acesso em: 08 set. 2020.

SIEGEL, E.; RETTER, A. *eXist*. A NoSQL Document Database and Application Platform. Sebastopol, Calif.: O'Reilly Media, 2014.

V. A. *Projeto Estranhar Pessoa*. 2011. Disponível em: <<http://estranharpessoa.com/>>. Acesso em: 20 abr. 2020.

O uso de mapas interativos como ferramenta de aprendizagem colaborativa e multimodal

Silvia Araújo

Diogo Cunha

Adriana Silvério

Francisco Abrunhosa

INTRODUÇÃO

É inquestionável a importância da integração das Tecnologias de Informação e Comunicação na formação dos profissionais ao serviço da língua, que se pretende atual e em linha com as exigências do mercado e da sociedade moderna (MAIA; HALLER; ULYRCH, 2002). A tecnologia veio para ficar e está a mudar a forma como vivemos. Como referem vários autores (ANANIADOU; CLARO, 2009; GERBAULT, 2012, entre outros), a literacia digital ocupa um lugar importante na construção das competências a adquirir pelos adultos do século XXI, competências essas que envolvem a capacidade de compreender e produzir textos multimodais em contexto monolíngue ou plurilíngue (LEBRUN; LACELLE; BOUTIN, 2012; LEBRUN; LACELLE, 2014). Esta capacidade de ler e comunicar de forma eficaz, combinando escrita, imagens e efeitos sonoros em diferentes tipos de suporte, pressupõe competências subjacentes de ordem cognitiva, emocional, pragmática, semiótica e textual (THIBERT, 2012). É precisamente neste contexto de integração criativa das tecnologias que se inscreve o projeto *Story@* que nos propomos apresentar na secção seguinte.

PROJETO *STORY@*: DA ESCRITA COLABORATIVA À ESCRITA MULTIMODAL

Com base na experiência adquirida quer ao nível da compilação de corpora no âmbito do projeto Per-Fide (ALMEIDA *et al.*, 2004) financiado pela FCT

(PTDC/CLE-LLI/108948/2008), quer ao nível da formação especializada em TIC de professores de línguas (assegurada, desde 2010, no âmbito do Congresso Internacional ticLÍNGUAS1), surgiu a ideia de criarmos uma plataforma online (SILVA; ARAÚJO, 2016) que possa ajudar os professores a dotar os seus alunos de competências de produção multimodal (produções que combinam som, imagens fixas ou animadas em suporte digital etc.):

FIGURA 1 - GALERIA DAS PRODUÇÕES DOS ALUNOS NA PLATAFORMA *Story@*



Fonte: Captura de ecrã da plataforma *Story@* (2020).

¹ Organizado, desde 2010, em parceria com o Centro *e-learning* da TecMinho/Gabinete de Formação Contínua da Universidade do Minho, o Congresso Internacional *ticLÍNGUAS* (<https://sites.google.com/site/ticlinguas2010/inicio>) tinha como principal finalidade apresentar e partilhar experiências e práticas de utilização pedagógica das TIC nas aulas de língua. Este congresso (doravante designado por *techLING*, <http://ceh.ilch.uminho.pt/techling2019/>) pretende ir um pouco mais além da aplicação das tecnologias da informação e comunicação em contexto educativo, dado que passa a reunir trabalhos que refletem diferentes abordagens, académicas e não académicas, que demonstrem o impacto das ferramentas de base tecnológica no tratamento das línguas.

No âmbito deste projeto de escrita multimodal, trata-se de propor atividades pedagógicas a partir de formas narrativas curtas, com vista ao desenvolvimento de competências escritas e orais em língua materna e/ou estrangeira com recurso, se possível, a tecnologias emergentes. Como se pode ver na Figura 2 abaixo, estas atividades são organizadas a partir de um determinado elemento desencadeador (etapa 1) passível de servir de origem à escrita colaborativa (HULIN, 2013)² de um texto monomodal (etapa 2) que pode ser posteriormente convertido num texto multimodal³ (etapa 3):

FIGURA 2 - PROJETO *STORY@*: PERCURSO FLEXÍVEL EM TRÊS ETAPAS



Fonte: Elaborada pelos autores (2020).

Esses conteúdos multimodais dão expressão à antiga arte de contar histórias, mas de um modo moderno, dado que incorporam, pelo menos, duas formas de expressão (texto, som e/ou imagem) (LOURENÇO, 2012; ROBIN; McNEIL, 2012; LEBRUN; LACELLE, 2014). Esses conteúdos podem obviamente ser mais ou menos complexas (e criativas) conforme o nível dos alunos e o tempo/meios

² Existe na Internet inúmeras ferramentas de escrita colaborativa: *Google docs* (<https://www.google.com/docs/about/>), *etherpad* (<http://etherpad.org/>), *Framapad* (<https://framapad.org/>) entre outras.

³ Vejam-se, por exemplo, algumas das produções multimodais (SILVA; ARAÚJO, 2016, p. 5-11) criadas por alunos de Humanidades da Universidade do Minho (a partir de uma imagem), com recurso a ferramentas de *storytelling* (*Storybird*, *PowToon*, *ToonDoo*, entre outras).

disponíveis para o projeto: leitura “encenada” da narrativa com música e ruídos de fundo, adaptação da narrativa em banda desenhada, em filme de animação ou em videojogo etc. Na plataforma *Story@*, os professores podem registar-se para poder carregar quer os textos monomodais redigidos pelos seus alunos, quer eventualmente (se tiverem optado por avançar para a terceira etapa acima mencionada) as respetivas adaptações multimodais.

Importa referir que esse desencadeador pode ser proposto pelo próprio professor, dependendo do aspeto programático que pretende abordar. Os conteúdos elaborados a partir desse desencadeador podem mobilizar a língua materna e/ou língua(s) estrangeira(s).

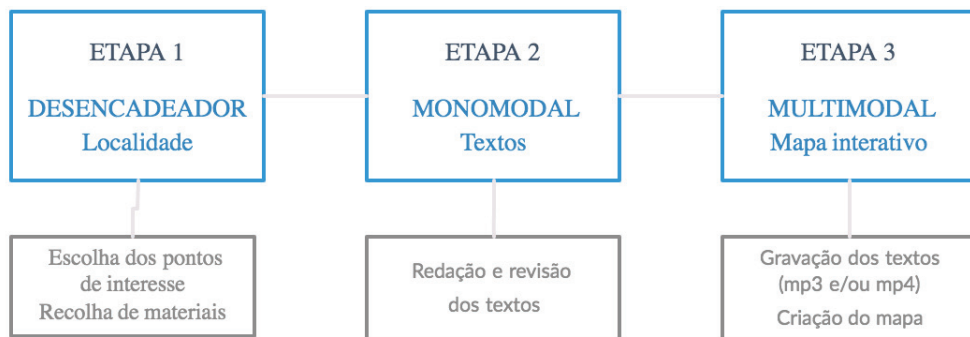
UM CONVITE À VIAGEM PARA DESENVOLVER AS COMPETÊNCIAS LINGUÍSTICAS E DIGITAIS

As crianças e os jovens de hoje pensam e processam as informações de forma diferente das gerações anteriores. Por crescerem rodeados pelas tecnologias, estes nativos digitais necessitam de novas formas de aprendizagem mais interativas e participativas. Não é, pois, de estranhar que sintam momentos de total desânimo (RAVEAUD, 2011) quando confinados numa sala de aulas que tende a privilegiar a transmissão de tópicos programáticos de forma dirigida e unidirecional. Apesar de todas as limitações que conhecem os professores (excesso de alunos por turma, programas densos, entre outras), parece-nos importante envolver os alunos em tarefas intelectualmente desafiadoras que os possam dotar de competências para procurar, avaliar, produzir e partilhar informação na Web (PUREN, 2002; NISSEN, 2004) de forma crítica, ética e segura. Num mundo em que o fluxo de informações é intenso e em permanente mudança, é, de facto, fulcral que os alunos adquiram competências ao nível da literacia da informação e da literacia digital (LOURENÇO, 2012) para que possam participar e interagir numa sociedade em rede (CASTELLS, 2009), altamente competitiva, que valoriza o ser-se flexível e criativo (OHLER, 2008; COUTINHO; LISBÔA, 2011). Ou seja, parece-nos importante que o aluno deixe de ser um mero consumidor passivo e se torne um produtor ativo de conteúdos sensível às questões do respeito dos direitos de autor.

Nestes novos cenários, a integração curricular das TIC pode contribuir significativamente para que sejam implementadas, nos espaços formais de educação,

estratégias pedagógicas inovadoras e significativas que se inspiram em metodologias ativas tais como o *Project-Based Learning* que estimulam a flexibilidade, a autonomia e autoformação (ARPIN; CAPRA, 2001; PROULX, 2004). Perante a constelação de ferramentas e recursos digitais que são disponibilizados diariamente na internet, não é de estranhar que a comunidade escolar sinta dificuldades em testar, avaliar e usar de forma produtiva todo este manancial tecnológico. Nas linhas que se seguem, é nosso propósito propor uma experiência pedagógica que leve alunos de diferentes níveis de escolaridade (ensino básico e superior) a realizar, em modo colaborativo, um mapa turístico interativo. Retomando o percurso flexível das três etapas acima descrito (ver acima, Figura 2), podemos então dizer que se trata aqui de partir de uma localidade (etapa 1) para desenvolver as competências linguísticas e digitais dos alunos (etapa 2), através da criação de um mapa turístico (etapa 3) que incorpora, como veremos, outros modos de representação além da escrita:

FIGURA 3 - PERCURSO_STORY@ APLICADO À CRIAÇÃO DE MAPAS INTERATIVOS



Fonte: Elaborada pelos autores (2020).

Para criar estes roteiros turísticos, os alunos precisam, como sugeríamos acima, de procurar e selecionar os conteúdos (textos, imagens e/ou vídeos) relativos a cada um dos pontos de interesse que farão parte do mapa; algumas das informações encontradas em língua estrangeira sobre determinadas cidades europeias terão de ser traduzidas para a língua materna e todo este material multilíngue compilado terá de ser posteriormente gravado, recorrendo a programas de edição de áudio. Estes mapas constituem, como vemos, um verdadeiro convite à viagem, uma viagem virtual que pode perfeitamente transformar-se numa viagem real. Esta viagem permite partir

à descoberta da cultura dos outros, mas também da nossa (se optarmos por um mapa sobre uma cidade portuguesa), mas também promove o desenvolvimento das competências orais e escritas de forma integrada e contextualizada (STOLLER, 2006).

DESCRIÇÃO DO PROJETO *VIAJAR NA AULA*_ALUNOS DE LÍNGUA ESTRANGEIRA DO ENSINO BÁSICO

O projeto *Viajar na Aula* foi integrado numa ação de formação acreditada para professores, levada a cabo pela Porto Editora e pela Casa do Professor⁴. Esta ação de formação decorreu 22 vezes, entre dezembro de 2014 e junho de 2015, em 20 locais, chegando a centenas de docentes do ensino básico das disciplinas de língua espanhola e francesa:

FIGURA 4 - LOCAIS DE REALIZAÇÃO DA FORMAÇÃO *VIAJAR NA AULA*



Fonte: Porto Editora (2015).

OBJETIVOS PRETENDIDOS COM O PROJETO

Este projeto teve como objetivo principal familiarizar os docentes com a ferramenta *Google My Maps*⁵ que lhes permite criar um mapa interativo *online* com

⁴ Disponível em: <<http://www.casadoprofessor.pt>>. Acesso em: 23 abr. 2020.

⁵ Esta aplicação está disponível em: <<https://www.google.com/maps/d/>>. Acesso em: 24 abr. 2020.

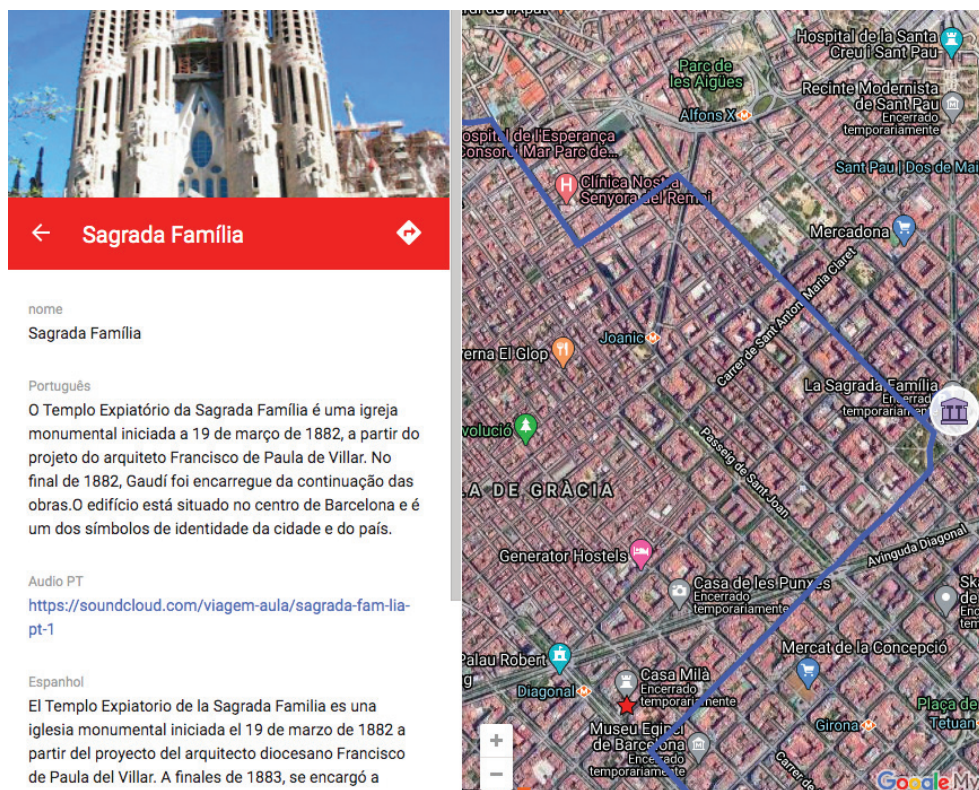
os seus alunos em ambiente de aula. Foi criada e disponibilizada uma plataforma *online* para que alunos e docentes pudessem não só partilhar os seus mapas, mas pudessem também consultar os mapas realizados por outras turmas do país. Nesta plataforma, encontram-se vários tutoriais destinados a apoiar a criação dos mapas interativos pela rede de escolas interessadas em integrar o projeto.

DESCRIÇÃO DA FORMAÇÃO ASSEGURADA NO ÂMBITO DO PROJETO

Proposta, como vimos, em 22 localidades do país, a ação de formação dividia-se numa sessão teórico-prática (com os docentes de ambas as disciplinas, francês e espanhol) e numa sessão prática (para cada uma das disciplinas em separado, devido a limitações logísticas e para personalizar os conteúdos do workshop a cada uma das disciplinas em questão). Na sessão teórico-prática, depois de um breve enquadramento teórico, em que eram destacados conceitos como pedagogia de projetos, conectivismo, aprendizagem colaborativa (assente nas TIC), entre outros, propunha-se, em tempo real, uma demonstração de todos os passos necessários para a construção de um mapa interativo; com um manual impresso, os docentes acompanhavam passo a passo as explicações sobre a utilização das ferramentas. Na sessão prática, para simular um ambiente de aula, a turma de docentes era dividida em grupos de três ou quatro pessoas por forma a que cada grupo pudesse escolher um ponto de interesse referente ao mapa da cidade em que a turma iria trabalhar colaborativamente. Numa primeira etapa, os docentes procediam à recolha de imagens/vídeos e informação sobre o ponto de interesse escolhido e redigiam uma pequena síntese descritiva entre 40 a 60 palavras em português e na língua da sua disciplina. Na segunda etapa, os docentes exploravam a ferramenta *online SoundCloud*⁶ para fazer uma gravação áudio dos dois textos redigidos anteriormente. Na etapa final, os docentes recorriam à ferramenta *Google My Maps* para criar o mapa interativo da cidade proposta, preenchendo-o com os conteúdos (textos e respetivas gravações áudio) produzidos nas duas etapas anteriores:

⁶ *SoundCloud* (<https://soundcloud.com/logout>) é uma plataforma *online* de publicação de *podcasts*. Existem, na web, outros serviços do tipo: por exemplo, *audioboom* (<https://audioboom.com/>) ou *podomatic* (<https://www.podomatic.com/>).

FIGURA 5 - EXEMPLO DE MAPA CRIADO COLABORATIVAMENTE DURANTE A FORMAÇÃO



Fonte: Mapa personalizado criado Google My Maps (2015).

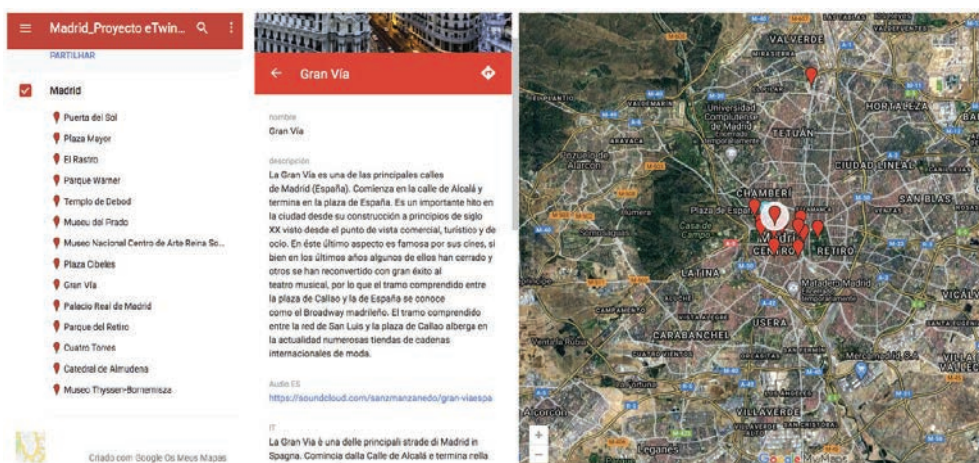
Caso completassem as três etapas antes do tempo disponibilizado, os grupos dispunham de uma quarta etapa mais livre que lhes permitia adicionar outras características ao mapa, tais como criação de rotas personalizadas e outras alterações de ordem estética (ver Figura 5 acima).

APRESENTAÇÃO DOS MAPAS INTERATIVOS PRODUZIDOS

Dos trabalhos realizados por escolas e submetidos no site *Viajar na Aula*, destacam-se dois: o projeto *Madrid eTwinning Connecting Classrooms* e o projeto *Forjães à la carte*.

Madrid_Proyecto eTwinning Connecting Classrooms retrata o sucesso de um trabalho colaborativo *online* entre duas turmas de dois países diferentes (Itália e Portugal). Tratou-se de um projeto de *eTwinning*, um programa europeu que promove a cooperação entre escolas e parcerias entre escolas da Europa, que levou os alunos a criar um mapa interativo em três línguas (espanhol, italiano e português) da cidade de Madrid. Este trabalho foi coordenado pelas professoras Marta Sanz Manzanedo (da escola IIS Vespucci-Colombo – ITC Amerigo Vespucci, Livorno, Itália) e Isabel Costa (do Agrupamento de Escolas José Relvas, Alpiarça, Portugal):

FIGURA 6 - MADRID eTWINNING CONNECTING CLASSROOMS COM ES_IT_PT



Fonte: Roteiro personalizado criado no *Google My Maps* (2015).

Ao clicar em cada um dos apontadores, abre-se uma janela interativa que agrega imagens/vídeos sobre o ponto de interesse escolhido (neste caso, a Gran Vía em Madrid, ver Figura 6 acima), um pequeno texto descritivo e um link que corresponde à gravação deste texto. A esta junção de texto, imagem e áudio corresponde uma abordagem multimodal da língua que permite desenvolver, de forma contextualizada, as diferentes competências de escrita e de oralidade. Esta multimodalidade é realmente muito importante, pois promove uma educação inclusiva: nenhum aluno com deficiências visuais ou auditivas se sente excluído

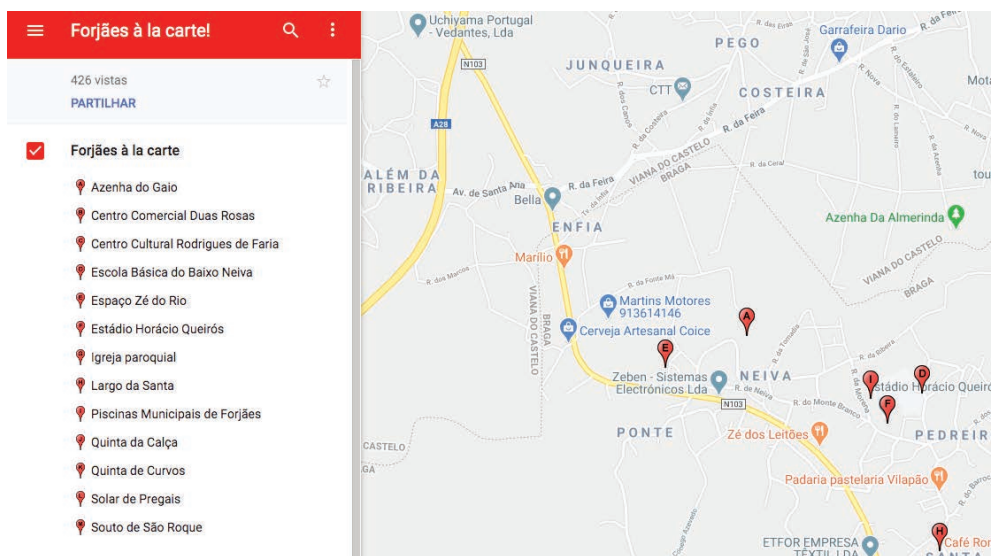
do projeto (quem não vê, pode ouvir; quem não ouve, pode ver ou ler). Por outro lado, importa referir que não se trata aqui de chegar a mapas interativos⁷ perfeitos, recusando, por exemplo, que alunos com maiores problemas de pronúncia possam gravar os textos. O objetivo é precisamente o inverso: levar os alunos com mais dificuldades a participar e suprir as suas lacunas.

Uma outra vantagem destes mapas prende-se com o seu carácter multilíngue. Como acabámos de ver, para além de estar em português, cada ponto de interesse do mapa de Madrid (ver Figura 7) surge igualmente em espanhol e em italiano. Outras línguas podem obviamente integrar este projeto que coloca em rede as escolas portuguesas entre si, mas também com outras escolas estrangeiras. Se uma língua veicula uma cultura, então ao introduzirmos neste projeto várias línguas, estamos a promover um diálogo intercultural que fomenta o respeito pela diferença. Não há línguas maiores ou menores, há línguas diferentes, com identidade própria. Todas elas merecem ser respeitadas. Esta abordagem multilíngue é importante quando se pretende educar para a cidadania. Por outro lado, estes mapas podem e devem dar origem a sinergias entre as línguas e outras disciplinas como a História ou a Geografia. Vemos a importância de propor estes mapas em cursos profissionais que lidam com questões de gastronomia, de património, i.e., com aspetos relacionados com o turismo.

Forjães à la carte foi um projeto transdisciplinar realizado por uma turma do 7º ano da escola de Forjães que levou os alunos a explorar e catalogar pontos de interesse da sua freguesia. O mapa interativo que se segue integra os aspetos turísticos considerados mais relevantes:

⁷ Estes mapas podem obviamente ser recheados com material recolhido *in loco*, pois se tiverem a oportunidade de visitar a localidade, os alunos poderão com seus telemóveis tirar fotografias e publicá-las a posteriori nestes mapas. Ou seja, estes mapas podem servir para dar conta de uma viagem que se viveu ou para preparar uma viagem que se vai fazer.

FIGURA 7 - *FORJÃES À LA CARTE* COM FR_EN

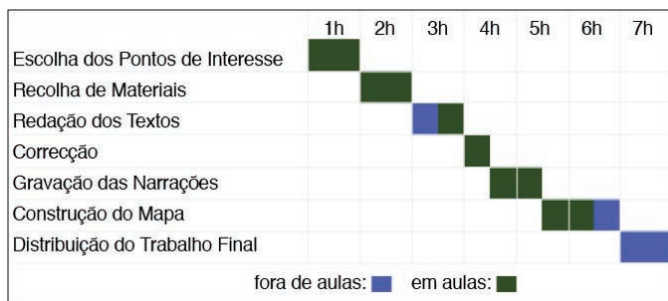


Fonte: Roteiro personalizado criado na *Google My Maps* (2015).

Este exercício de cidadania ativa por parte dos alunos suscitou o interesse da própria junta de freguesia. Esta sugeriu que se criasse, numa fase futura deste projeto, um postal com as informações sobre os pontos de interesse criados pelos alunos. Este postal poderá ser distribuído no posto de turismo e permitirá que o turista aceda diretamente ao mapa interativo através de um *QR code*. Também se equacionou a possibilidade de inserir um *QR code* nas placas informativas dos respetivos locais turísticos com ligação ao mapa criado pelos alunos.

No caso do projeto *Forjães à la carte*, foi possível acompanhar e dar algum apoio técnico aos alunos e docentes da escola de Forjães. Através dos dados recolhidos aquando da implementação deste projeto, traçou-se o plano de trabalho abaixo para propor uma representação visual das etapas e tempo necessário para a criação de um mapa interativo em contexto escolar:

FIGURA 8 - TEMPO DESPENDIDO POR TURMA PARA A CRIAÇÃO DO MAPA INTERATIVO



Fonte: Elaborada pelos autores (2015).

DESCRIÇÃO DO PROJETO *O MEU ROTEIRO DE BRAGA*_ALUNOS DA LICENCIATURA EM ESTUDOS CULTURAIS

A partir de um desencadeador⁸ (ver Figura 9), criado e partilhado com a ferramenta *Storify*, propôs-se a alunos (de 3º ano) da Licenciatura em Estudos Culturais da Universidade do Minho que elaborassem um roteiro passível de incentivar turistas nacionais e estrangeiros a (re)visitar Braga, a cidade mais antiga de Portugal:

FIGURA 9 - DESENCADEADOR *O MEU ROTEIRO DE BRAGA*



Fonte: Página web criada no *Storify*⁹.

⁸ Trata-se da Praça da República, popularmente referida apenas como Arcada, onde se encontram a Igreja da Lapa e dois cafés emblemáticos da cidade, ambos centenários: o *Café Vianna* e o *Astória*.

⁹ A plataforma *Storify* foi descontinuada em maio de 2018.

Mas, neste caso, os alunos não se limitaram a explorar a ferramenta *Google My Maps*; reconhecendo que a capacidade de resolução de problemas e de tomada de decisões é importante para o contexto empresarial atual (JONASSEN, 2007), optou-se por convidar os próprios alunos a procurar, selecionar e manusear uma série de outras ferramentas¹⁰ que permitem criar mapas personalizados. Alguns dos mapas produzidos com base nessas ferramentas gratuitas encontradas na internet são apresentados a seguir.

➤ **Esplanadas para Amar** – criado com a ferramenta *Jauntful*¹¹, o folheto apresentado abaixo pretende dar a conhecer algumas das mais agradáveis esplanadas¹² da cidade de Braga, através da criação de micro narrativas românticas¹³:

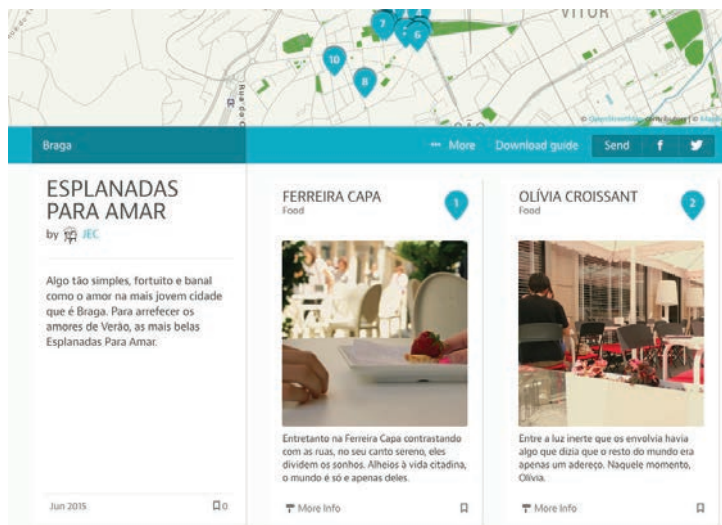
¹⁰ Na Unidade curricular de *Tecnologias de Comunicação em Humanidades*, que consta do plano de estudos da Licenciatura de Estudos Culturais da Universidade do Minho, os alunos de 1º ano foram incentivados a dar conta da sua cidade/freguesia, mas, neste caso, tiveram também de pesquisar e incorporar as lendas que estão associadas a esta localidade. Vejam-se, por exemplo, os dois trabalhos realizados sobre a lenda do galo de Barcelos: o vídeo coloca em cena uma habitante de Barcelos (<https://goo.gl/yJ3Eqa>) e a narrativa digital, elaborada com a ferramenta *adobe slate*, esforça-se por conseguir um aspeto profissional (<https://goo.gl/BRki79>).

¹¹ Para além do roteiro, esta ferramenta (<https://jauntful.com>) permite criar na parte inferior separadores que descrevem de forma mais detalhada cada apontador visível no roteiro.

¹² O folheto apresentado na figura 10 mostra as esplanadas ideais para amar de manhã (*Brasileira; Lusitana; Bolacha Maria*), para amar à tarde (*De Gemma Hamburgueria; Cantinho das frigideiras; Spirito; Café da livraria centésima Página*) e à noite (*Barhaus; Velha a Branca; Ma vie*).

¹³ Trata-se de mostrar o lado romântico do centro de Braga, i.e., o outro lado de uma cidade tida como conservadora.

FIGURA 10 – *ESPLANADAS PARA AMAR*



Fonte: Folheto interativo criado no *Jauntful*¹⁴.

➤ **Ecoss da noite** – aliando desta vez elementos textuais a vídeos produzidos de raiz, os alunos procuram oferecer, recorrendo à ferramenta *Scribble Maps*¹⁵, uma viagem pelos bares que animam a vida noturna da cidade:

FIGURA 11 – PROJETO_ECO DA NOITE



Fonte: Roteiro dos bares bracarenses criado no *Scribble Maps*¹⁶.

¹⁴ Disponível em <<https://goo.gl/SMyMq9>>. Acesso em: 18 abr. 2020.

¹⁵ A vantagem desta ferramenta (<http://www.scribblemaps.com>) relativamente ao *Google Maps* prende-se com o facto de se poder customizar os apontadores, inserindo códigos html.

¹⁶ Disponível em <<http://goo.gl/XKgjgY>>. Acesso em: 23 abr. 2020.

Importa salientar que, em ambos os casos (ver Figuras 10 e 11), os alunos optaram por dar a conhecer a cidade dos Arcebispos dando corpo aos atores que figuram nas fotografias e vídeos por eles concebidos. Para realizar estes conteúdos (audio)visuais, os alunos tiveram de lidar diretamente com os gerentes dos diferentes comércios envolvidos para obter as devidas autorizações de cedência de imagem (fotografia e/ou filmagem).

➤ **Roteiro do Barroco de André Soares** – a partir de um mapa estático de divulgação do Barroco publicado na página web da Câmara Municipal de Braga¹⁷, os alunos optaram, neste caso, por desenvolver duas versões de um mapa interativo sobre um roteiro de André Soares, um artista bracarense do século XVIII, famoso pela excelência das suas obras e criador de uma versão muito pessoal do estilo barroco-rococó.

Um dos roteiros (criado com a ferramenta *Google Earth Pro*¹⁸) inclui comentários áudio redigidos com base em documentação especializada sobre a cidade de Braga (OLIVEIRA, 2014):

¹⁷ Disponível em: <https://www.cm-braga.pt/archive/doc/Mapa_Barroco_PT.pdf>. Acesso em: 23/04/2020. Note-se que na unidade curricular de *Localização e Produção Digital Multilíngue* do Mestrado de Tradução e Comunicação Multilíngue da Universidade do Minho, os alunos foram levados a atualizar o website da *Erasmus Student Network Minho* (ESN Minho) convertendo, entre outras coisas, os mapas dos *campi* da universidade do Minho disponíveis em formato PDF em mapas interativos (em português e em inglês) destinados à integração dos alunos estrangeiros (veja-se, por exemplo, o campus de Gualtar: <<https://goo.gl/87drFV>> ou o campus de Azurém: <<https://goo.gl/y7Tpmy>>). Esta experiência pedagógica permitiu-lhes experimentar a real dinâmica entre o cliente e o prestador de serviços.

¹⁸ Para a construção do roteiro do Barroco de André Soares, os estudantes recorreram à versão Pro do *Google Earth* (Disponível em: <<http://www.google.com/earth/download/gep/agree.html>>. Acesso em: 23 abr. 2020, com funcionalidades adicionais em relação à versão básica.

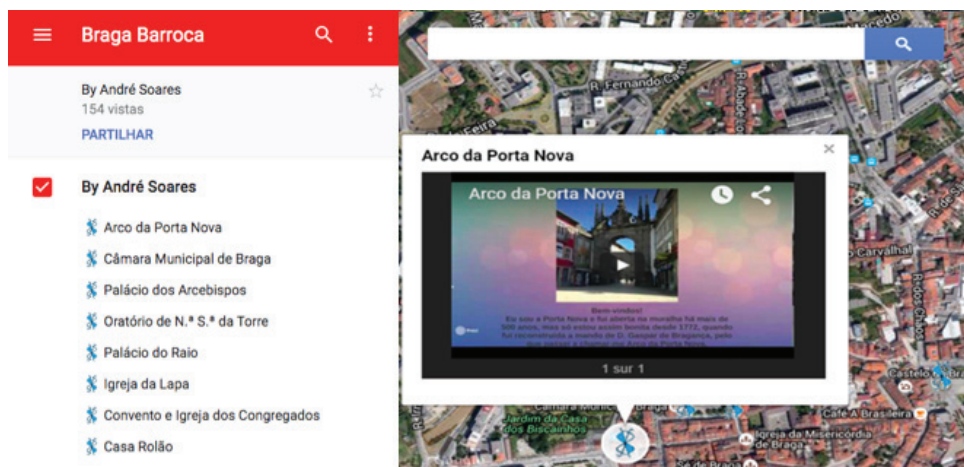
FIGURA 12A - ROTEIRO DO BARROCO DE ANDRÉ SOARES



Fonte: Roteiro criado na *Google Earth Pro*¹⁹.

O segundo roteiro (elaborado com a ferramenta *Google My Maps*) integra pequenos vídeos acompanhados de elementos textuais e áudio, para que pessoas com algum tipo de deficiência (visual ou auditiva) possam aceder ao conteúdo:

FIGURA 12B - ROTEIRO DO BARROCO DE ANDRÉ SOARES



Fonte: Roteiro criado no *Google My Maps*²⁰.

¹⁹ Disponível em <<https://goo.gl/4vsB0D>>. Acesso em: 23 abr. 2020.

²⁰ Disponível em <<https://goo.gl/j0K6ps>>. Acesso em: 23 abr. 2020.

Para representar os marcadores embutidos como ícone dos locais a visitar nos dois roteiros, desenhou-se um símbolo “borboleteante”, que aponta para a “Braga Barroca”, num feliz encontro entre as ideias de liberdade e de viagem. Recorrendo ao programa *online pilxr*²¹, chegou-se a uma solução que se considerou agradável, coerente e reconhecível, i.e., ao símbolo do projeto que assume a cor azul dos elementos constituintes da Heráldica de Braga²²:

FIGURA 13 - MARCADOR “BRAGA BARROCA” EMBUTIDO NO ROTEIRO BARROCO DE ANDRÉ SOARES



Fonte: Elaborada por Francisco Abrunhosa (2015).

CONSIDERAÇÕES FINAIS

Hoje em dia, é indiscutível a relação entre as tecnologias digitais e a aprendizagem (formal e informal) em áreas como a aprendizagem/aquisição de línguas. Já não é possível ignorar a tecnologia que nos rodeia e são notórias as oportunidades de aprendizagem que as tecnologias proporcionam aos alunos. Ao percorrer vários pontos do país para apresentar (a professores do ensino básico) este projeto de criação de mapas interativos turísticos e ao implementar este mesmo projeto em contexto universitário (para divulgar a cidade de Braga), foi nosso objetivo propor um projeto que fomente a integração pedagogicamente orientada das tecnologias, isto é, não apenas como mero fator lúdico, mas que influencie igualmente os alunos à produção espontânea e criativa em língua materna e estrangeira.

²¹ Disponível em: <<https://pixlr.com>>. Acesso em: 24 abr. 2020.

²² Disponível em: <<https://www.cm-braga.pt/pt/0101/municipio/heraldica>>. Acesso em: 24 abr. 2020.

Esta experiência teve como principal objetivo gerar uma reflexão em torno das vantagens pedagógicas associadas à Metodologia de Projeto e ainda compreender os efeitos associados à sua aplicação em contexto educativo, sobretudo quando mobilizada com estratégias de ensino-aprendizagem orientadas para o trabalho colaborativo. Ao levar os alunos a participar ativamente na sua aprendizagem, esta metodologia de aprendizagem baseada em Projetos constitui, sem dúvida, uma boa estratégia de formação na medida em que fomenta um uso crítico e criativo das tecnologias e promove uma mudança de atitudes. Importa, contudo, certificarmos de que quando pensamos no uso das tecnologias na educação, não estamos a falar no uso meramente instrumental das ferramentas de tecnologia da informação e comunicação, mas sim no conjunto de metodologias e estratégias específicas que as utilizam como ferramentas de apoio ao ensino/aprendizagem. Torna-se, pois, relevante compreender as vantagens da utilização das tecnologias da informação e comunicação como meio para inovar as práticas pedagógicas e melhorar as competências quer linguísticas, quer transversais (criatividade, trabalho em equipa, capacidade de resolver problemas, sentido de responsabilidade e autonomia) dos alunos.

No âmbito das experiências pedagógicas acima descritas e que envolveram, como vimos, alunos de diferentes níveis escolares, quisemos mostrar que é possível recorrer a ferramentas tecnológicas, gratuitas e intuitivas, para implementar, em ambiente normal de sala de aula, projetos pedagógicos interdisciplinares que se podem expandir para fora das paredes da escola. Ao propor a combinação de diferentes sistemas de representação e significação em contexto monolíngue ou multilíngue, este tipo de projetos viabiliza, como tentámos mostrar, a construção de ambientes multimodais de aprendizagem colaborativa que impulsionam novas formas de comunicação, de interação, de organização e produção do conhecimento. Numa relação direta com questões de motivação, autonomia e desenvolvimento de competências cognitivas, linguísticas, culturais e tecnológicas, esta abordagem multimodal (que subjaz ao projeto *Story@*) pode, de facto, ajudar-nos a promover uma aprendizagem colaborativa, lúdica e criativa, que procura equilibrar a prática da escrita e da oralidade e respeitar os diferentes estilos de aprendizagem.

REFERÊNCIAS

ALMEIDA, J. J.; ARAÚJO, S.; CARVALHO, N.; DIAS, I.; OLIVEIRA, A.; SANTOS, A.; SIMÕES, A. The Per-Fide Corpus: A New Resource for Corpus-Based Terminology, Contrastive Linguistics and Translation Studies. In: SARDINHA, T. B.; FERREIRA, T. (Ed.) *Working with Portuguese Corpora*. London: Bloomsbury Academic, 2004. p. 177-200.

ANANIADOU, K.; CLARO, M. 21st Century skills and competences for New Millenium learners in OECD countries. *OECD Education Working Papers*, n. 41, 2009. Disponível em: <<http://goo.gl/a6rHnO>>. Acesso em: 24 abr. 2020.

ARPIN, L.; CAPRA, L. *L'apprentissage par projets*. Montréal: Chenelière/McGraw-Hill, 2001.

CASTELLS, M. *Communication Power*. Oxford: Oxford University Press, 2009.

COUTINHO, C.; LISBÔA, E. Sociedade da Informação, do conhecimento e da aprendizagem: desafios para educação no Século XXI. *Revista de Educação*, v. XVIII, p. 5-22, 2011.

GERBAULT, J. Littératie numérique Les nouvelles dimensions de l'écrit au 21e siècle. *Les Cahiers de l'Acedle*, v. 9, n. 2, p. 109-128, 2012.

HULIN, T. Enseigner l'activité 'écriture collaborative'. *Tic&Société*, v. 7, n. 1, p. X-X, 2013. Disponível em: <<http://ticetsociete.revues.org/1314>>. Acesso em: 24 abr. 2020.

JONASSEN, D. *Computadores, ferramentas cognitivas: desenvolver o pensamento crítico nas escolas*. Porto: Porto Editora, 2007.

LEBRUN, M.; LACELLE, N. L'ère numérique: un défi pour la didactique du FLE. *Synergies Portugal*, n. 2, p. 107-117, 2014.

LEBRUN, M.; LACELLE, N. ; BOUTIN, J-F. (Dir.). *La littératie médiatique multimodale: De nouvelles approches en lecture-écriture à l'école et hors de l'école*. Québec. Université du Québec, 2012.

LOURENÇO, M. C. *Da narrativa à narrativa digital: o texto multimodal no estudo da narrativa*. 2012. 141 f. Dissertação (Mestrado em Ciências da Educação) - Universidade do Minho, 2012.

MAIA, B.; HALLER, J.; ULYRCH, M. (Ed.). *Training the Language Services Provider for the New Millenium*. Porto: Faculdade de Letras da Universidade do Porto, 2002.

NISSEN, E. Importance du scénario pédagogique dans l'apprentissage d'une langue étrangère en ligne. *Les Langues Modernes*, p. 14-24, 2004. Disponível em : <<https://hal.archives-ouvertes.fr/edutice-00001446/document>>. Acesso em: 24 abr. 2020. Disponível em: <<https://hal.archives-ouvertes.fr/edutice-00001446/document>>. Acesso em: 24 abr. 2020.

OHLER, J. *Digital Storytelling in the Classroom: New Media Pathways to Literacy, Learning and Creativity*. Thousand Oaks, CA: Corwin Press, 2008.

OLIVEIRA, E. *Braga Top-Secret*. V. N. Famalicão: Centro Atlântico, Lda, 2014.

PROULX, J. *Apprentissage par projet*. Canada: Presses de l'Université du Québec, 2004.

PUREN, C. Innovation et cohérence didactique en langue. *New Standpoints*, v. 12, p. 2-7, 2002.

RAVEAUD, M. Le plaisir et l'ennui comme choix pédagogiques. *Revue Internationale d'éducation de Sèvres*, n. 57, p. 89-98, 2011.

ROBIN, B. Digital Storytelling: A Powerful Technology Tool for the 21st Century Classroom. *Theory into Practice*, v. 47, p. 220-228, 2008.

ROBIN, B.; McNEIL, S. What Educators should Know about Teaching Digital Storytelling. *Digital Education Review*, v. 22, p. 37-51, 2012.

SILVA, M. M. da; e ARAÚJO, S. La subjectivité du lecteur-scripteur et l'écriture collaborative multimodale avec la plateforme Story@: une étude de cas. *Myriades, Revue d'Études Francophones*, n. 2, p. 58-72, 2016.

STOLLER, F. Establishing a Theoretical Foundation for Project-Based Learning in Second and Foreign Language Contexts. In: BECKETT, G. H.; MILLER, P. C. (Ed.). *Project-Based Second and Foreign Language Education: Past, Present, and Future*. Greenwich. Connecticut: Information Age Publishing, 2006. p. 19-40.

THIBERT, R. Pédagogie + Numérique = Apprentissages 2.0. *Dossier d'actualité Veille & Analyses*, n.79, p. 1-22, 2012.

A Digital Platform for Literary Translation: collaborative translation and teaching¹

Susana Valdez
Isabel Oliveira Martins

GENERAL OVERVIEW OF THE PROJECT

PEnPAL in Trans – Portuguese-English Platform for Anthologies of Literature in Translation is an interinstitutional literary translation project comprising two components: research and teaching. This venture brings together teachers and researchers in literature, translation studies and linguistics from various institutions (the University of Lisbon, Nova University, the University of Minho, the Portuguese Catholic University, and, until 2014, Lusófona University). One of the main goals is to offer support for literary translation in higher education, while contributing to innovative research in other areas, such as Comparative Literature, American Studies, Diaspora Studies, Contrastive Linguistics, Literary Translation and Digital Humanities.

Drawing on the concept of process-oriented didactics and taking its cue from an extensive body of literature (KIRALY, 1995, 2013) which attempts to bridge the gap between translation teaching and translation practice, *PEnPAL in Trans* seeks to establish collaborative translation stemming from a teaching environment that extends beyond academia and the classroom, thus reaching the community at large.

The literary translator is still, to this day (and at least at first sight), not conceived as part of the online professional network, capable of taking advantage of collaborative online work by fostering translation competence through social engagement and peer empowerment: literary translators “are still absent from the

¹ Special thanks are due to Margarida Vale de Gato (the project’s PI) for her feedback and support, Fernando Ferreira Alves, Rui Azevedo, Ana Maria Chaves, Nancy Vieira Couto, Katherine Vaz and our students for their participation in the project.

game” (MOGHADDAM, 2013, p. 205). Literary translation mirrors the romantic ideal of the author that works in isolation, oblivious to the impact of collaborative work/interaction. As Austermühl (2001, p. 11) has pointed out: “the antiquated image of a lone translator, armed only with a pencil or typewriter and surrounded by dusty books, is no longer realistic”. However, even though the scientific-technical translator is closely associated with a growing mechanization or automation of the translation process within a complex multi-agent network (VALDEZ, 2019), the image of the literary translator is to some extent still crystalized. When collaborative “construction” is considered, it is not among translators but arises as a result of intervention by the publisher, editor, and sometimes the author (PAUL, 2009, p. 39).

However, *PEnPal* assumes that literary translation can be perfected in collaborative environments, and thus we agree with Moghaddam (2013, p. 200) when he supports the use of weblogs in literary translation:

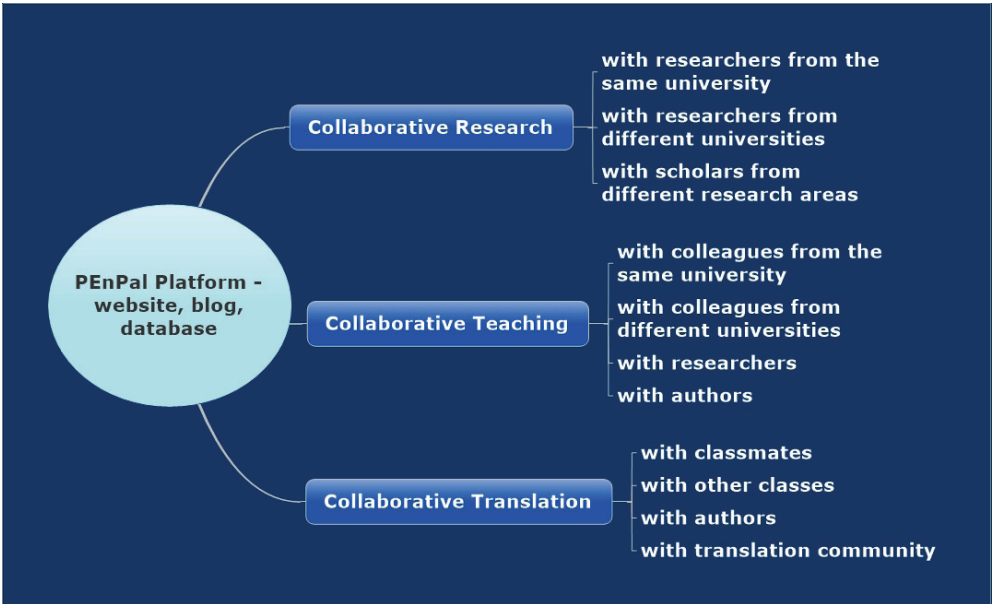
It can act as a ‘mediator’ between the author of the work under translation, the translator, and the prospective readers. It also provides facilitated ‘collaborative possibilities’ for the translation itself, by creating a virtual world around the translation which is bigger than the translator’s small, isolated space... When literary translators play the role of bloggers in literary translation, they are no longer lonely and isolated translators, surrounded by different kinds of dusty dictionaries. The weblog gives them the opportunity to enlarge the world around their translation.

Thus, the didactic and research components of the *PEnPAL in Trans* project are supported by a range of digital resources developed in order to contribute to the acquisition of literary translation competence, namely:

- (1) a website that connects all the digital resources and content online where all those involved can follow the project;
- (2) a blog that aims to extend the “network” of literary translation training (students, authors, teachers, translators, among others), providing a forum and a record of intuitively formulated translation problems;
- (3) a digital database where teachers and researchers can add translation problems, alternative solutions and discussions of their strategies.

Figure 1 provides an overview of this collaborative construction within and fostered by the PEnPal project. The combination of the website, blog and database encourages three levels of collaboration, namely: (1) collaborative research; (2) collaborative teaching; and (3) collaborative translation.

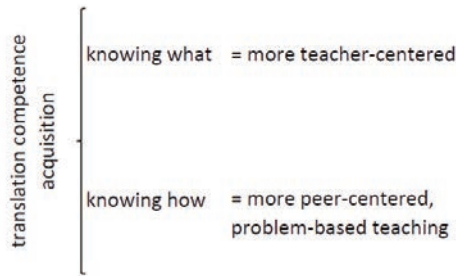
FIGURE 1 – COLLABORATIVE CONSTRUCTION WITHIN AND FOSTERED BY *PENPAL IN TRANS*



Source: Elaborated by the authors (2020).

Through the use of an online platform, *PEnPal* creates a “landscape which is not mapped by conventional geographies” (BUSH, 2001, p. 127-130). Researchers, teachers and translators from different backgrounds, institutions and geographical areas interact and collaborate towards a common goal: the acquisition and development of procedural knowledge (“knowing how”) based on growing declarative knowledge (“knowing what”) and including learning-by-doing in an extended notion of the collaborative classroom environment. The classroom in this setting is not limited to the presence of a traditional teacher. Looking at Figure 2, it is clear that procedural knowledge is not solely centered on the teacher’s intervention. “Knowing how” is stimulated by peer feedback and is based on problem solving, whilst “knowing what” is more teacher-centered (KIRALY, 1995, p. 197-224).

FIGURE 2 – TRANSLATION COMPETENCE ACQUISITION



Source: Elaborated by the authors (2020).

As the PATT group stated (OLVERA-LOBO *et al.*, 2007, p. 517), and Gonzáles Davies (2004) corroborated: “teacher-centered teaching is obsolete and has largely been replaced by a learner-centered alternative which favors student autonomy and eliminates the figure of the all-knowing teacher capable of resolving all problems” (OLVERA-LOBO *et al.*, 2007). Instead, feedback from different sources feeds the organic development of translation competence. In fact, a number of studies (BARROS, 2011; DESJARDINS, 2011; GAMBIER, 2012; GATO *et al.*, 2016; KENNY, 2008; KIRALY, 2001; LISAITÉ *et al.*, 2016; O’BRIEN, 2011) have argued that collaborative translation can be advantageous for translation training.

FIRST OUTPUT

The first output of this venture is the already published anthology *Nem cá nem lá: Portugal e América do Norte entre escritas (Neither Here nor There: Writings Across Portugal and North America)*.² It includes works by 29 authors, comprising Americans and Canadians who write in English but maintain some sort of connection to their Portuguese heritage and American authors who live in Portugal and write in English about their Portuguese experiences. In total, the anthology contains 56 source texts divided into four categories: fiction; poetry; autobiography/memoir;

² The anthology was funded by the Luso-American Foundation, by means of the Alberto de Lacerda Translation Award 2013; see Alves *et al.* (2016).

and historical discourse, children's literature and drama. All the translated texts involve cross-cultural transit between Portugal and North America; they are mostly narratives of displacement, cultural and interlingual exchange between English and Portuguese, and display heterolingualism, mutilated or corrupted speech, intertextuality from multiple literary traditions, lyrical evasions and dis-order, calling for different functional and hermeneutic considerations in the translation strategies employed. The anthology also contains entries on each of the 29 authors which offer not only biographical information but also some considerations on the translation process and main translation strategies used, such as selective non-translation as compensation for heterolingualism, explicitation, italicization of passages marked as foreign in the source text, replacement of disruptive language with non-standard target varieties, reliance on mixed literary repertoires, and the systematized replication of orality markers in the target language, among others.

CASE STUDIES

The project began in 2011, and some of the translations were done when the digital platform was taking its first steps and thus it was used rather unevenly. The focus of this article is to describe and discuss how the collaborative practice was performed during the translation and teaching processes, using two specific cases with different approaches:

1. Collaborative translation of Katherine Vaz's short story "Lisbon Story" in the classroom.
2. Collaborative translation of Nancy Vieira Couto's poem "They Double Up Around the Absence of Campfires" in an experimental setting³.

"LISBON STORY"

The short story (30 pages long) was first translated in a master's seminar in 2011/2012 at NOVA University. The translation involved twenty students aged between 20 and 50 from a wide variety of academic backgrounds (former translation

³ For more information about these authors in the Portuguese-American context and the collaborative translation of selected works, see Martins (2016a, 2016b).

undergraduate students, but also students from other academic areas) and mostly did not include contact with digital tools. The text was divided and distributed between 5 groups of 4 students. They were alerted to the specific problems connected to this type of text, namely interference by the target language in the English text, and other more general problems related to the translation of literary texts – polysemy, connotation, and style, among others. Additionally, the groups were also encouraged to collaborate with each other using the blog and to contact the author.

Some of the students did use the blog and one of the first conclusions was that their interventions were generally about semantic problems or lexical choices and had a very impressionistic character which indicated personal limitations, including a slightly limited level of competence in both languages. Moreover, it was also possible to observe that the exchange of opinions helped some of them to improve their translations, but although students were also encouraged to reflect on the strategies they should use whenever confronted with a certain type of problem (for example, corrupted Portuguese words or expressions used by the author), these strategies were not aspects the students felt compelled to discuss. This type of discussion only occurred in the classroom when prompted by the teacher and when each translation was being revised by a different group. Similarly, aspects such as coherence, namely in the use of forms of address – so very different in the two languages and particularly important in this source text – were not even considered by the students. Once again, this was an aspect which was only approached after the teacher intervened. Contact with the author – made by email – was, on the other hand, rather productive, since the students were able to understand the reasons behind the use of certain expressions in Portuguese which sounded rather unfamiliar. On the whole, this first experience showed the teacher what kind of strategies should be used in order to avoid some of the above-mentioned situations.

Thus, in 2013-2014, the same text was used in another master's seminar on literary translation. The class was composed of 14 students, again with ages ranging between 20 and 50, and once again with rather varied academic backgrounds and different levels of competence in the use of digital tools. The students were divided into groups and their task was to revise the translation produced by their classmates. The version handed to the students contained some minor or subtle

errors. They were alerted to the different aspects already mentioned and were even given a list of what kind of specific problems they should address, as well as an indication that they should use the blog both to collaborate among themselves and to see what difficulties previous students had pinpointed about that specific text or about other texts that were already online.

The general result was productive, in view of the almost final version of the translation. Some students, once again, found it difficult or were reluctant to use the digital platform. The difficulty was mainly due to some of them not being familiar with the use of digital tools, but the reluctance was related to an aspect that arises from the collaborative environment, one that was not expected: the students were willing to put forward their doubts and difficulties in the classroom, accepting suggestions and discussing possible problems with both the teacher and their classmates, but some of them did not feel comfortable revealing their own doubts and problems online. In other words, openly posting what they considered to be their own lack of knowledge was a drawback and thus peer pressure may be a factor to take into consideration.

Furthermore, we must bear in mind that the text they were dealing with was being revised and therefore some of the problems had already been dealt with. Nevertheless, the students still used the blog mostly to consider alternative lexical choices, and were not aware of some of the mistakes. Again, it was only in the classroom that they were able to work with each other more effectively and come to an acceptable version of the translation that took into consideration the list of specific problems to be dealt with, and once more this was under the teacher's supervision.

On the whole, it may seem that translating in a collaborative environment was not very helpful but actually the process as a whole was indeed quite useful, having furnished the teacher – and the final revisers – with some options and insights that might not have been considered if the translation had been done by one solitary translator. On the other hand, these two experiences, even if they are only supported by empirical observation, showed some of the aspects that the overall project should take into account and these will be suggested after the second case has been presented.

“THEY DOUBLE UP AROUND THE ABSENCE OF CAMPFIRES”

Turning now to the second case study — “They Double Up Around the Absence of Campfires” — our starting point was the evaluative research question “what impact does using a blog as a collaborative online tool have on literary translation practice?” A pilot study was conducted in May 2013 to explore this question. The experiment aimed to record the translation product and the students’ translation process and observe interactions between the students and the translation teachers, the author and the professional literary translator. We hypothesized that the use of the collaborative online tool would lead to a better-quality product.

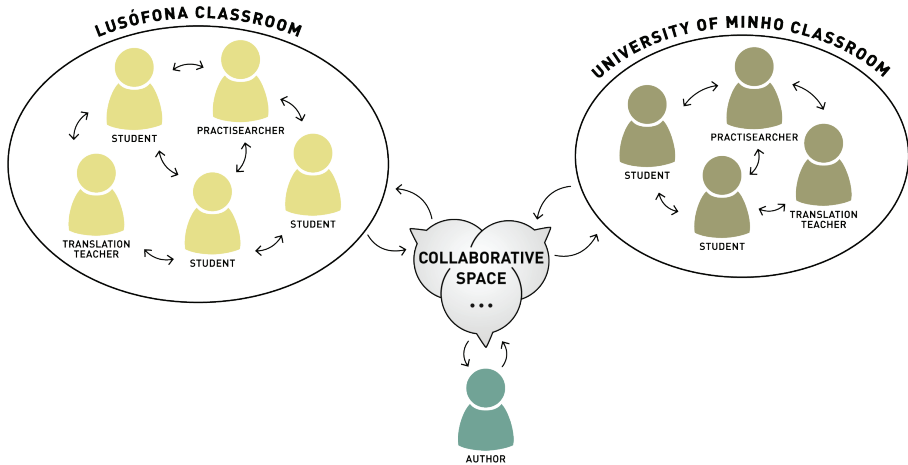
In order to ensure triangulation, five types of data were collected: (1) translation product; (2) screen recording data; (3) comments posted on the blog; (4) pre-interviews and post-interviews; and (5) the researchers’ observational notes. BB Flashback (<https://www.flashbackrecorder.com/>) was used to capture the screen recording data.

EXPERIMENTAL DESIGN

The experiment took place simultaneously at two different universities in two different cities – at Lusófona University in Lisbon and at the University of Minho in Braga – in a classroom setting. The translation students were asked to translate a poem at the same time and use the blog to communicate with each other and with the teachers, author and professional translator by posting their translation problems and difficulties. Part of their translation task was to help solve their classmates’ problems and difficulties and share their own problems and difficulties. The author of the source text could also be contacted through the blog to give feedback on the students’ problems and answer specific queries. Participants were supported in person by two specialists – their literary translation teacher at their university and one *practisearcher*⁴ who monitored the entire process. The collaborative work done during the pilot study is shown in Figure 3. The students were informed that they should try to finish the translation within two hours.

⁴ The concept of *practisearcher* was first put forward by Gile (1998).

FIGURE 3 - VISUAL REPRESENTATION OF THE COLLABORATIVE WORK DURING THE PILOT STUDY



Source: Elaborated by the authors (2020).

All translations were carried out using computers and the students were instructed to use all their usual online resources, focusing on the blog in particular. Screen recording software was installed on the computers before the experiment started and all participants were informed that their translation process was going to be recorded. At the beginning of the experiment, each participant received a paper copy and a Word file of the source text. All instructions were given orally before the task.

PARTICIPANTS

Five students studying translation were recruited from three different Portuguese universities. The participants were all Portuguese: three undergraduate students and two MA students. Table 1 below shows information on the translators who participated in the pilot experiment. As it indicates, two of the students came from the University of Minho – one an undergraduate studying Applied Languages with no professional experience and the other a student on an MA course in Translation and Multilingual Communication with (little) professional

experience, both familiar with IT tools; two were from Lusófona University – two undergraduate students studying Translation and Creative Writing with no professional experience, one familiar with IT tools and the other not; and one student on an MA course in Translation at Nova University with no professional experience but familiar with IT tools.

TABLE 1 - INFORMATION ON THE TRANSLATORS WHO PARTICIPATED IN THE PILOT EXPERIMENT

Participant1	F	Undergraduate	Applied Languages	University of Minho	No professional experience	Familiar
Participant2	F	MA student	Translation and Multilingual communication	University of Minho	Professional experience	Familiar
Participant3	F	Undergraduate	Translation and Creative Writing	Lusófona University	No professional experience	Not familiar
Participant4	F	Undergraduate	Translation and Creative Writing	Lusófona University	No professional experience	Familiar
Participant5	F	MA student	Translation	Nova University	No professional experience	Familiar

Source: Elaborated by the authors (2020).

SOURCE TEXT

The five students were asked to translate the poem “They Double Up Around the Absence of Campfires” by Nancy Vieira Couto. The poem was selected due its representativeness of Couto’s body of work. Additionally, its difficulty and length was taken into consideration, given the students’ translation competence level and the time allocated for the task⁵. The author’s availability for the experiment was also taken into account.

⁵ During the process to select the source text, we asked a professional literary translator – Ana Maria Chaves – to translate the poem in order to determine the difficulty and time needed for it to be translated.

RESULTS

The main purpose of the pilot study was to determine (1) the time spent on each of the tasks during the translation process; (2) each participant's interaction using the collaborative space; and (3) the correlation (if any) between (1) and (2) and the quality of translation.

TIME

The screen recording data revealed five clear translation phases: reading the source text, researching the author, language and terminology problems, blog-related tasks (i.e. posting questions, answering questions, reading peers' questions and answers), translation *per se* and revision. Table 2 presents the experimental data regarding the time spent in minutes by each participant. The time presented in the table for each one of the phases results from the sum of the different blocks of time spent on each task throughout the translation process, i.e. in most cases the participant did not perform each task uninterruptedly. For instance, Participant3 started by reading part of the text, researched a translation problem, and then carried on reading.

TABLE 2 - TIME SPENT IN MINUTES ON EACH TASK DURING THE TRANSLATION PROCESS

	Reading	Research	Blog	Translation	Revision	Total
Participant1	04:88	23:65	14:57	56:00	00:00	01:40:30
Participant2	25:00	24:24	32:00	29:00	15:00	02:05:24
Participant3	03:02	10:48	36:53	17:28	39:41	01:47:52
Participant4	02:42	23:00	41:29	19:31	27:03	01:53:45
Participant5	05:16	23:11	28:14	28:19	35:00	02:00:00

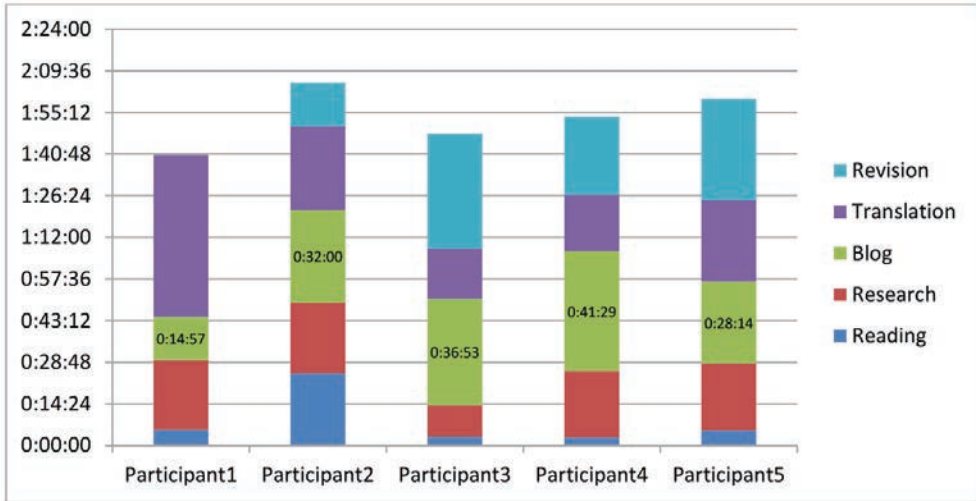
Source: Elaborated by the authors (2020).

As can be seen in the Table 2, there is a big difference in the time each participant spent on each task. For example, the time spent reading varied from just under five minutes to twenty-five minutes.

Figure 5 below displays an overview of the time each participant spent on each task, with particular emphasis on the time spent on the collaborative space, i.e. the *PEnPal* blog. What is interesting about the data in this table is that it becomes

clear that all participants spent an average of 31 minutes working collaboratively. Also, with the exception of Participant1, who only spent just under 15 minutes on the collaborative space, most students spent about the same amount of time working with each other.

FIGURE 4 - TIME EACH PARTICIPANT SPENT ON EACH TASK



Source: Elaborated by the authors (2020).

COLLABORATIVE WORK

The analysis of the comments posted on the blog revealed three types of comments:

- (1) questions related to translation problems and difficulties, such as when a translator asked what “They sing ‘Oh bury me not’” refers to;
- (2) answers to peers’ problems and difficulties, such as when one of the translators answered the above question with a link to a Wikipedia page: “I think it is a song http://en.wikipedia.org/wiki/Bury_Me_Not_on_the_Lone_Prairie”; and
- (3) observations not immediately related to problems or difficulties, like, for instance when the translators replied to a peer’s answer with “Thank you!”.

It should be noted that the interactions only began after one of the *practisearchers* posted the first question. Until that time, no student took the initiative to post any type of comment even though the screen recording data showed that most were visiting the blog frequently. Table 3 presents a breakdown of the interaction by each participant on the *PEnPal* blog. The total number of questions (11), answers (13) and comments (11) are not evenly distributed. As can be seen in the table below, Participant2 and Participant5, both master's students, clearly interacted more than the undergraduate participants.

TABLE 3 - EACH PARTICIPANT'S INTERACTION IN THE COLLABORATIVE SPACE

	Questions	Answers	Observations	Total
Participant1	1	0	1	2
Participant2	2	3	4	9
Participant3	1	4	0	5
Participant4	5	0	0	5
Participant5	2	6	6	14

Source: Elaborated by the authors (2020).

Of the 11 questions raised, the majority (7 questions) were related to meaning. Two of the translators, for example, asked about the meaning of “double up”. The remaining questions (4 questions) were related to context, culture and the author's intentions, as in these examples: “Did you imagine them [the hats] to be made of straw or more like felt?”; “Can you please tell me if the average American would immediately know what the flappers' clothes are?” Interestingly, this is also our experience in the classroom as stated above in the “Lisbon Story” case study. Most of the students' translation difficulties were related to semantics and lexicon and therefore focused on the micro-level. Macro-level problems were rarely verbalized, as we have seen in the previous case study, unless the teacher draws the students' attention to them.

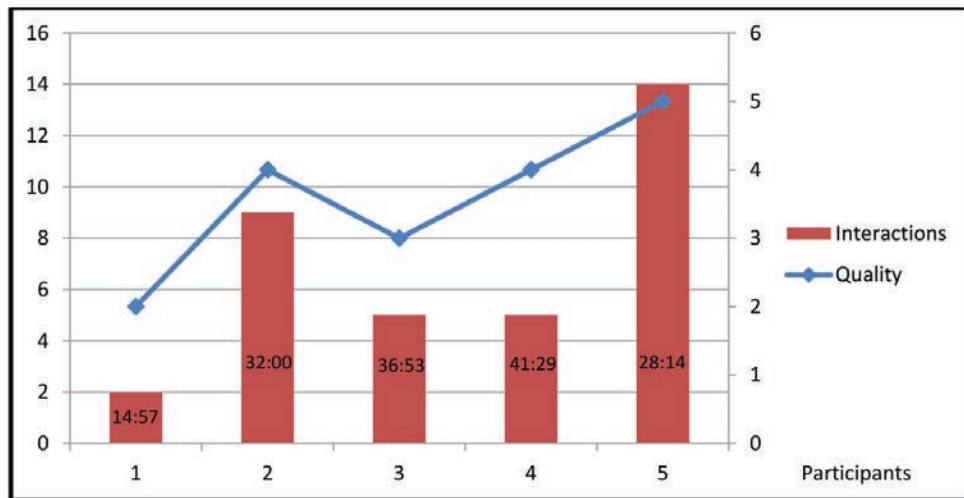
Regarding the interaction between the author and the students on the *PEnPal* blog, one participant commented in the post-interview: “I liked having direct input from the author. It's something that I'm not used to when I translate. Usually we have to guess what the author intended in the original and in this case we did not have that barrier” (our translation). One of the advantages of using a

blog is precisely the opportunity to connect translators with living authors with the view of active cooperation as Moghaddam (2013, p. 206) highlights: “the author helps the translation and reflects upon the feedback received by the translator, all of which are available to the public.” The author – Nancy Vieira Couto – commented to this effect in the post-interview: “As a poet, I found the experiment very valuable. It illustrated just how difficult it is to translate a poem, even a short one, not only from one language to another but also from one culture to another.”

COMPARISON BETWEEN QUALITY AND COLLABORATIVE WORK

Before considering a possible connection between quality and collaborative work, it was necessary to analyze the translated poems and grade them in terms of quality. The researchers and teachers came to an agreement regarding the evaluation of each translated poem on a scale of 1 (major revision needed before publication) to 5 (no changes or minor changes needed before publication). A comparison was then established between the quality of the translation and collaborative work by means of the number of interactions and the time spent on the *PEnPal* blog. The results obtained from the analysis of this comparison can be seen in Figure 5.

FIGURE 5 - COMPARISON BETWEEN THE QUALITY OF THE TRANSLATION (IN BLUE), THE NUMBER OF INTERACTIONS (IN RED) AND THE TIME IN MINUTES SPENT ON THE PEnPAL BLOG (IN BLACK)



Source: Elaborated by the authors (2020).

From the graph above we can clearly see that in two of the cases – Participant2 and Participant5 – there seems to be a correlation between the collaborative work (both time and interactions) and the quality of the translation. In fact, one participant, when asked about this experiment in the post-interview, gave statements that corroborated this. She explicitly referred to the impact on the quality of the translation:

I appreciated the interaction with classmates, and it was very useful to discuss solutions for the most difficult terms. In this respect, the blog was a contributing factor to the final quality of my translation because it accelerated the problem-solving process. (our translation)

The correlation between collaborative work (both time and number of interactions) and the (high) quality of the translation in the case of Participant4 is not self-evident. The number of interactions is lower in comparison with Participant2 and Participant5 but the time spent on the blog is the highest (41 minutes and 29 seconds). She also rates highly on the quality scale (4). This suggests that although she did not interact explicitly on the blog, she used it as a reference. In fact, she spent more time using the blog than translating *per se*: 41 minutes and 29 seconds on the blog and 19 minutes and 31 seconds translating. When asked about the time spent on the blog, the translator said: “I think that the time I spent on the blog made me translate faster because it helped solve the translation problems.” (our translation)

The correlation between low quality and little collaborative work can be seen in Participant1’s case. Even though she was familiar with IT tools, she reported difficulties using the blog during the experiment (e.g. logging questions and answers, using it as a reference tool). The student’s poor performance may be the result of multiple factors: difficulties in using the blog, pressure from the experiment itself, time pressure, peer pressure (as in the “Lisbon Story” case), to name just a few.

Taken together, these results suggest that there is a connection between collaborative work, the proficient use of collaborative tools, and translation quality. However, with a small sample size, caution must be applied, as the findings may not be representative.

PRELIMINARY CONCLUSIONS

This study set out with the aim of assessing the impact of using collaborative online tools on translation competence acquisition and literary translation teaching and practice. We looked at two case studies for this purpose. In the first case, we dealt empirically with both the teacher’s and the students’ experiences using the digital platform as a complement to the classroom. In the second case, we hypothesized the existence of a correlation between the use of collaborative tools and translation quality. The data from these two cases studies revealed important insights.

The collaborative environment created by our project was shown to be productive, since generally the final translation(s) benefited from that collaboration, particularly if we bear in mind factors such as the possibility of contact with the authors and students from other universities.

FROM A TEACHER-CENTERED TO A STUDENT-CENTERED APPROACH

Both cases studies revealed that teachers and particularly students are used to a teacher-centered approach. The students’ lack of autonomy was visible. In “Lisbon Story” and “They Double Up Around the Absence of Campfires” students did not make the first step to verbalize and problematize translation problems and difficulties. Most cues came from facilitators – teachers and/or *practisearchers*. In “Lisbon Story”, verbalization and problematization at micro level was clearly more accessible for students than problematization at macro level, and thus students’ online behavior mirrored what happened in the classroom. Only when the teacher called their attention to macro-level issues, were (some of) the students able to verbalize and problematize at a different level. In the second case study — “They Double Up Around the Absence of Campfires” — the participants’ interactions were also mainly focused on micro-level translation problems and difficulties (both in their questions and answers), and even those type of interactions only began after one of the *practisearchers* initiated the dialogue. In other words, the facilitators had/have to be the catalyst and this raises another problem: the participants’ technological competence.

DEVELOPMENT OF TECHNOLOGICAL COMPETENCE

An underlying assumption of our approach with *PEnPal in Trans* was that technological competence was one part of the students' and teachers' acquired competences. However, as we found out throughout the different stages of the project, teachers, researchers and students showed difficulties in using online tools effectively. It was also assumed that the potential difficulties did not have an impact on translation practice and translation competence acquisition.

Yet it became clear that without technological competence (1) teachers and researchers may not be able to understand the advantages of the online tools and teach correct and effective use of the tools to students; (2) teachers and researchers that do not master technological competence may avoid the use of the tools or scarcely encourage their use; and (3) consequently only a small number of students will use the tools. As pointed out by Moghaddam (2013, p. 207), “we need to be aware of the fact that some of the most celebrated literary translators may still prefer old habits”, and this can also be said for teachers, researchers and students.

Therefore, this reflection has enabled us to suggest the creation of two advanced courses for teachers and researchers: (1) “Digital Humanities” by Diogo Queiroz de Andrade (June 2015) and (2) “Digital Humanities and Literary Translation” by Maarten Janssen, Manuel Portela and Margarida Vale de Gato (June 2016).

Additionally, a redefinition of technological competence is suggested based on this experience. Widely varying definitions of the term technological competence have emerged. Technological competence, sometimes referred to as “instrumental subcompetence” (PACTE Research Group, 2003), is one of the competences needed to successfully execute the translation process. This term has come to be used to refer to the knowledge associated with the use of translation tools. The EMT expert group (2009, online) specifies the procedural expertise that constitutes technological competence in greater detail:

Knowing how to use effectively and rapidly and to integrate a range of software to assist in correction, translation, terminology, layout, documentary research (for example text processing, spell and grammar check, the internet, translation memory, terminology database, voice recognition software); knowing how to create and manage a database and files; knowing how to adapt to and familiarise oneself with new tools, particularly for the translation of multimedia and audiovisual material; knowing how to prepare and produce a translation in different formats and for different technical media and knowing the possibilities and limits of MT.

As far as it was possible to ascertain, technological competence is restricted in previous literature to the tools that are unique to the translation task (such as, for instance, CAT tools). Although no definition can be completely comprehensive, up to date, and comprise all possible scenarios, these definitions may be problematic. Professional translation is an increasingly collaborative effort. Translators not only directly resort to input from peers by asking for feedback through different technological resources such as professional forums, *Facebook* profiles, *Facebook* groups or messages, email and *Skype* calls, but also do so indirectly by looking up previously posted translation problems on blogs, professional groups and websites (*Kudoz*, *Ciberdúvidas*, and *WordReference* are among the most recurrent resources used and mentioned by Portuguese translators and translation agencies in style guides and QA reports). Therefore, this definition has been broadened to include all tools used by translators and is not restricted to those that are unique to the task of translation, such as CAT tools and quality control tools. Therefore, we propose a redefinition of technological competence as knowing how to use online and offline technological resources productively in order to effectively perform all tasks involved in the translation process (including translation, revision, terminology, layout, documentary research, and others).

Despite these insights, there are still many unanswered questions about the use of collaborative tools in collaborative translation in general and collaborative teaching and learning in particular. Further research should be undertaken to investigate the correlation between collaborative work and quality, and also to document and foster multi-agent networking in literary translation practice, teaching and research.

REFERENCES

ALVES, T. F. A. *et al.* (Ed.). *Nem cá nem lá: Portugal e América do Norte entre escritas*. Lisboa: Edlp, 2016.

AUSTERMÜHL, F. *Electronic Tools for Translators*. Manchester: St. Jerome, 2001.

BUSH, P. Literary Translation, Practices. In: BAKER, M. (Ed.). *Routledge Encyclopedia of Translation Studies*. London and New York: Routledge, 2001. p. 127-130.

DESJARDINS, R. Facebook Me!: Initial Insights in Favour of Using Social Networking as a Tool for Translator Training. *Linguistica Antverpiensia, New Series—Themes in Translation Studies*, v. 10, p. 175-196, 2011.

DONATA, L. *et al.* Negotiating Meaning at a Distance: Peer Feedback in Electronic Learning Translation Environments. *Translation and Meaning – New Series*, v. 41, n. 1, p. 99-116, 2016.

EMT expert group. *Competences for Professional Translators, Experts in Multilingual and Multimedia Communication*. 2009. Disponível em: <http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf>. Acesso em: 08 Sep. 2015.

GAMBIER, Y. Teaching Translation/Training Translators. In: GAMBIER, Y.; DOORSLAER, L. v. (Ed.). *Handbook of Translation Studies*. Amsterdam and Philadelphia: John Benjamins, 2012.

GATO, M. de V.; JANSSEN, M.; BARROS, R.; VALDEZ, S. Teaching and Researching Literary Translation in the Digital Context: PENPAL in Trans as a Case-Study. *MATLIT: Materialities of Literature*, v. 4, n. 1, p. 63-80, 2016. Disponível em: <<http://iduc.uc.pt/index.php/matlit/article/view/2377/1940>>. Acesso em: 08 Sep. 2015.

GILE, D. Observational Studies and Experimental Studies in the Investigation of Conference Interpreting. *Target*, v. 10, n. 1, p. 69-93, 1998,

GONZÁLEZ DAVIES, M. *Multiple Voices in the Translation Classroom*. Amsterdam: John Benjamins Publishing Company, 2004. (v. 54).

HUERTAS BARROS, E. Collaborative Learning in the Translation Classroom: Preliminary Survey Results. *JosTRANS Journal of Specialized Translation*, v. 2, p. 42-60, 2011. Disponível em: <http://www.jostrans.org/issue16/art_barros.php>. Acesso em: 07 Sep. 2015.

KENNY, M. A. Discussion, Cooperation, Collaboration: The Impact of Task Structure on Student Interaction in a Web-Based Translation Exercise Module. *The Interpreter and Translator Trainer*, v. 2, n. 2, p. 139-164, 2008.

KIRALY, D. *Pathways to Translation: Pedagogy and Process Translation Studies*. Kent, Ohio, & London: Kent State University Press, 1995.

KIRALY, D. Towards a Constructivist Approach to Translator Education. *Quaderns: Revista de Traducció*, v. 6, p. 50-53, 2001.

KIRALY, D. Towards a View of Translator Competence as an Emergent Phenomenon: Thinking Outside the Box(es) in Translator Education. In: KIRALY, D.; HANSEN-SCHIRRA, S.; MAKSYMYSKI, K. (Ed.). *New Prospects and Perspectives for Educating Language Mediators*. Tübingen: Narr Francke Attempto, 2013. p. 197-224.

LISAITÈ, D.; VANDEPITTE, S.; MAYLATH, B.; MOUSTEN, B.; VALDEZ, S. CASTEL-BRANCO, M.; MINACORI, P. Negotiating Meaning at a Distance: Peer

Feedback in Electronic Learning Translation Environments. *Translation and Meaning-New Series*, v. 1, n. 41, p. 99-116, 2016.

MARTINS, I. O. Katherine Vaz. In: ALVES, T. F. A. *et al.* (Ed.). *Nem cá nem lá: Portugal e América do Norte entre escritas*. Lisboa: Edlp, 2016a. p. 378-381.

MARTINS, I. O. Nancy Vieira Couto. In: ALVES, T. F. A. *et al.* (Ed.). *Nem cá nem lá: Portugal e América do Norte entre escritas*. Lisboa: Edlp, 2016b. p. 334-336.

MOGHADDAM, E. H. The Weblog: A Tool for Literary Translators. *Perspectives*, v. 21, n. 2, p. 200-208, 2013.

O'BRIEN, S. Collaborative Translation. In: GAMBIER, Y.; DOORSLAER, L. v. (Ed.). *Handbook of Translation Studies*. Amsterdam and Philadelphia: John Benjamins, 2011. p. 17-20.

OLVERA-LOBO, M. D.; ROBINSON, B.; CASTRO PRIETO, R. M.; QUERO GERVILLA, E.; MUÑOZ MARTÍN, R.; MUÑOZ RAYA, E.; MURILLO MELERO, M.; SENSO RUIZ, J. A.; VARGAS QUESADA, B.; DÍEZ LERMA, J. L. A Professional Approach to Translator Training (PATT). *Meta: Journal Des Traducteurs*, v. 52, n. 3, p. 517-528, 2007.

PACTE Research Group. Building a Translation Competence Model. In: _____. *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: John Benjamin, 2003.

PAUL, G. *Translation in Practice*. Champaign: Dalkey Archive Press, 2009.

VALDEZ, S. Perceived and Observed Translational Norms in Biomedical Translation in the Contemporary Portuguese Translation Market: A Quantitative and Qualitative Product- and Process-Oriented Study. 2019. (Phd diss.) - University of Lisbon and Ghent, 2019.

Projeto Corpus Eletrônico de Documentos Históricos do Sertão: A documentação epistolar

Zenaide de Oliveira Novais Carneiro
Mariana Fagundes de Oliveira Lacerda

INTRODUÇÃO

A tradição dos estudos de Linguística Histórica é marcada pela natureza atomística das análises feitas nesse campo de estudo da língua. Esse caráter atomístico das análises dos fatos linguísticos, que, inicialmente, refletia concepções igualmente atomizadas (pré-saussurianas) do fenômeno linguístico, manteve-se na Linguística Histórica, mesmo quando essas concepções que o fundamentavam já estavam superadas, em boa medida devido à dificuldade de se proceder a uma observação sistemática, e, na medida do possível, exaustiva, dos materiais disponíveis.

A constituição de banco de textos visa exatamente a romper com essa tendência nos estudos de história da língua, possibilitando, com a facilidade de um amplo acesso aos materiais, a aplicação das novas teorias que propugnam uma apreensão globalizante do objeto através de sua estrutura interna (linguística) e daquelas que, ainda mais globalizantes, propõem a apreensão dos fatos através da interação do sistema de relações linguísticas com as disposições e relações nas quais esse sistema se atualiza (as relações sociolinguísticas).

Essa constituição tem em mente, por outro lado, a dificuldade, já destacada por Labov (1972), em relação aos dados para o estudo da língua no *tempo real*. Um obstáculo irrefutável, diante do qual só resta à ciência buscar contorná-lo, através da maximização dos recursos existentes, dos textos remanescentes escritos em fases pretéritas da língua.

Hoje, contando com melhores recursos tecnológicos, no universo das Humanidades Digitais, os bancos de textos disponibilizam não somente edições semidiplomáticas, em PDF, mas também edições digitais – a partir do estabelecimento de redes entre projetos que desenvolvem a Linguística de Corpus e

a Linguística Computacional –, que servem como recurso eletrônico para estudos linguísticos, entre outros. Como se vê, “Do feliz conagraçamento entre as mais recentes tecnologias e a antiga Filologia, surgiu um novo universo de possibilidades para a preservação, disponibilização e análise de textos antigos, universo em que é possível oferecer ao leitor mais de uma edição do mesmo texto, permitindo que tenha ao seu dispor o texto editado, em diferentes versões, e o seu original.” (GONÇALVES; BANZA, 2013, p. 4)

Este trabalho discute o processo de constituição desses bancos de textos, no universo das Humanidades Digitais, apresentando a experiência na construção de *corpora* digitais e seus aspectos linguísticos e computacionais no âmbito do projeto Corpus Eletrônico de Documentos Históricos do Sertão (CE-DOHS)¹.

O objetivo do CE-DOHS é contribuir com o Projeto para a História do Português Brasileiro (PHPB) em diferentes perspectivas teóricas. Por meio da parceria tecnológica com o projeto Corpus Histórico do Português Tycho Brahe², ele possui um conjunto de documentos originados sobretudo da grande área do semiárido baiano. Esses documentos são editados em linguagem XML, com o uso do *eDictor*, um editor de texto desenvolvido por Paixão de Sousa, Kepler e Faria (2010). Essa ferramenta é especificamente voltada ao trabalho filológico e à análise linguística automática, e combina um editor de XML com um etiquetador morfossintático para permitir a geração automática de versões correspondentes a edições diplomáticas, semidiplomáticas e modernizadas (em HTML) e de versões com anotação morfossintática (em texto simples e XML).

Como resultado da primeira fase de pesquisa, o projeto CE-DOHS já disponibiliza diversos acervos, sobretudo de cartas manuscritas, organizando-as por grau de escolaridade e por grau de habilidade com a escrita; são 1.084 cartas particulares (1808-2000), num total de 350.850 palavras, escritas por 422 remetentes (nascidos entre 1724 e 1980), a maioria extraídas de Carneiro *et al.* (2011).

¹ O projeto é financiado pela Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB)/5566-2010 e é desenvolvido atualmente no Núcleo de Estudos em Língua Portuguesa (NELP), no Departamento de Letras e Artes da Universidade Estadual de Feira de Santana (UEFS). Disponível em: <www.uefs.br/cedoh>. Acesso em: 22 mar. 2020.

² Disponível em: <www.tycho.iel.unicamp.br>. Acesso em: 22 mar. 2020.

Na segunda fase do projeto, que está em andamento, o número de documentos tem sido ampliado, tanto manuscritos como impressos, com inserção, ainda, de amostras de fala, organizadas no CE-DOHS por comunidade, tipo de contato linguístico e vertente (popular e culta). Essa ampliação do *corpus*, segundo Bacelar do Nascimento (2004, p. 1),

[...] favorece essencialmente uma Linguística descritiva, fortemente apoiada pelas novas tecnologias, e permite tomar como ponto de partida da descrição a análise de quantidade significativa de dados autênticos, à semelhança do que se faz noutros domínios científicos. O uso de *corpora* permite a realização de descrições linguísticas de base empírica e promove, com isso, a discussão de questões teóricas solidamente fundamentadas. (BACELAR DO NASCIMENTO, 2004, p. 1).

Todo o material do CE-DOHS é representativo de variedades diacrônicas do português brasileiro (PB), de diferentes regiões do país e de graus de escolaridade distintos. Ele está sendo preparado para a anotação morfossintática, que manterá a maioria das características do padrão de anotação existente e permitirá a busca automática de dados, o que facilitará o estudo linguístico dos acervos, no que consiste o principal objetivo do CE-DOHS. O material disponível no Banco atende, entretanto, não somente a pesquisadores interessados em análises de aspectos linguísticos, mas em aspectos da difusão da escrita, da leitura, das transmissões textuais, históricas, políticas, econômico-sociais, entre outros.

CONSTITUIÇÃO DO BANCO

O CE-DOHS apresenta a versão digital de documentos em edição semidiplomática que compõem o banco Documentos Históricos do Sertão (DOHS), do projeto Vozes do Sertão em Dados: História, Povos e Formação do Português Brasileiro³, reunidos com o objetivo de estudar o processo de formação histórica do PB, especialmente na região do semiárido baiano.

Os documentos do Vozes mantém parceria com o PHPB por meio da prospecção e edição de documentos e da formação de *corpora* representativos de demandas histórico-sociais da região semiárida baiana. Devido às suas repercussões

³ Disponível em: <<http://www.uefs.br/nelp>>. Acesso em: 21 mar. 2020

no processo de formação histórica do PB e ao amplo contato linguístico de populações de origem portuguesa, indígena e africana, bem como com projetos temáticos de análise linguística, vêm servindo de base para a composição de uma Plataforma de Corpora do PHPB⁴, a cargo de Afrânio Barbosa, da Universidade Federal do Rio de Janeiro (UFRJ), e de Marcelo Módulo, da Universidade de São Paulo (USP).

A maior parte dos documentos do DOHS, datados e localizados – também disponíveis na página do CE-DOHS em formato digital – são cartas manuscritas pertencentes a nove acervos dos séculos XIX e XX (1.084 cartas, 418 remetentes). Essas cartas foram editadas principalmente por Carneiro (2005), que investiu na busca e na organização de acervos documentais que pudessem contribuir para o processo de reconstrução sócio-histórica do PB. Posteriormente, esses arquivos foram publicados na obra intitulada *Cartas brasileiras: coletânea de fontes para o estudo do português* (2011), organizada pela mesma autora.

Em 98% dos acervos, é possível determinar *onde, quando, por quem e para quem* as cartas foram escritas, atendendo à proposta de Petrucci (2003).

A DOCUMENTAÇÃO EPISTOLAR

Os nove acervos de cartas disponíveis no CE-DOHS representam as normas vernáculas e as normas cultas, de forma seriada, oferecendo um painel dos modelos de escrita e uma amostra da língua do período.

O Quadro 1 a seguir mostra a distribuição, por década, da documentação epistolar do CE-DOHS; são cartas dos séculos XIX e XX.

QUADRO 1 - DISTRIBUIÇÃO, POR DÉCADA, DA DOCUMENTAÇÃO EPISTOLAR DO CE-DOHS

Década	<i>Subcorpus</i> de cartas	Quant.
1800	Cartas para Vários Destinatários	1
1810	Cartas para Vários Destinatários	4
1820	Cartas para Vários Destinatários	9

continua

⁴ Disponível em: <<https://sites.google.com/site/corporaphpb>>. Acesso em: 21 mar. 2020

QUADRO 1 - DISTRIBUIÇÃO, POR DÉCADA, DA DOCUMENTAÇÃO EPISTOLAR DO CE-DOHS

continuação

Década	<i>Subcorpus de cartas</i>	Quant.
1830	Cartas para Vários Destinatários	2
1840	Cartas para Vários Destinatários	2
1850	Cartas para Vários Destinatários	7
1860	Cartas para Vários Destinatários	102
1870	Cartas para Vários Destinatários	34
1880	Cartas para Vários Destinatários, Cartas para Cícero Dantas Martins, Barão de Jeremoabo	24
1890	Cartas para Vários Destinatários, Cartas para Cícero Dantas Martins, Barão de Jeremoabo	164
1900	Cartas para Vários Destinatários, Cartas para Cícero Dantas Martins, Barão de Jeremoabo, Cartas para Severino Vieira, Governador da Bahia, Cartas do Acervo Dantas Jr.	158
Total de cartas do século XIX		502
1910	Cartas Baianas: o acervo de João da Costa Pinto Victoria, Cartas do Acervo Dantas Jr., Cartas em Sisal	20
1920	Cartas Baianas: o acervo de João da Costa Pinto Victoria, Cartas do Acervo Dantas Jr.	34
1930	Cartas Baianas: o acervo de João da Costa Pinto Victoria, Cartas do Acervo Dantas Jr., Acervo da família Freire	122

continua

QUADRO 1 - DISTRIBUIÇÃO, POR DÉCADA, DA DOCUMENTAÇÃO EPISTOLAR DO CE-DOHS

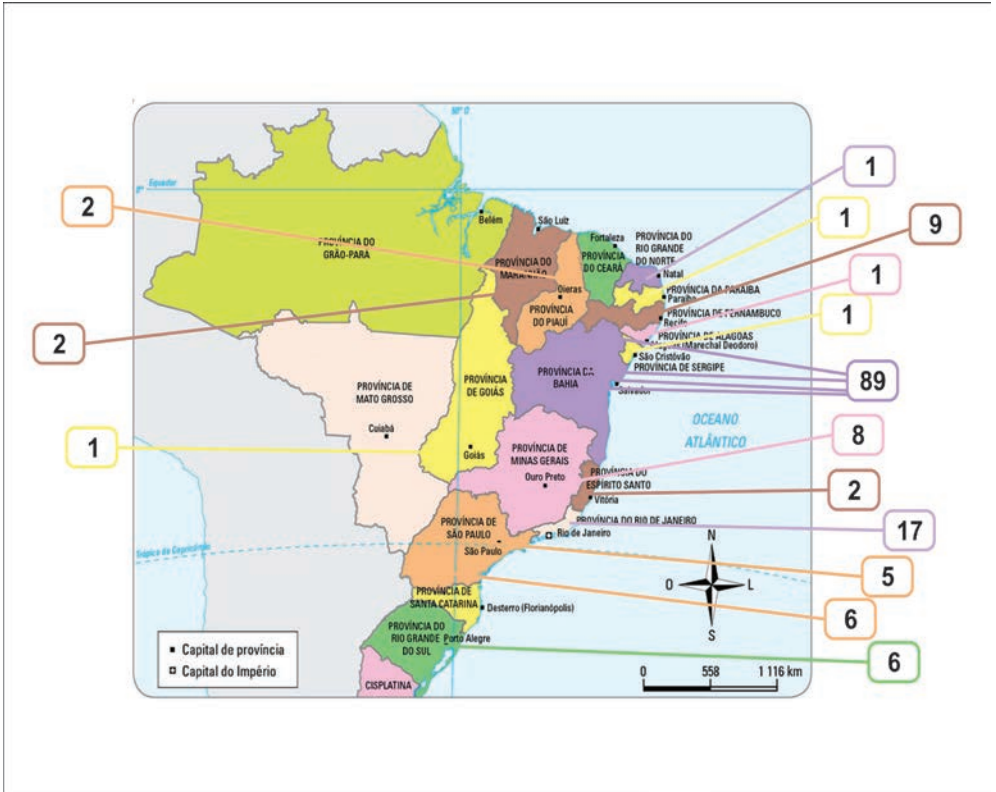
continuação

Década	<i>Subcorpus de cartas</i>	Quant.
1940	Cartas Baianas: o acervo de João da Costa Pinto Victoria, Cartas do Acervo Dantas Jr., Acervo da família Freire, Acervo particular da Família Soledade	120
1950	Cartas Baianas: o acervo de João da Costa Pinto Victoria, Cartas do Acervo Dantas Jr., Cartas em Sisal, Acervo particular da Família Soledade	96
1960	Cartas do Acervo Dantas Jr., Cartas em Sisal, Acervo da família Oliveira	61
1970	Cartas em Sisal, Acervo da Família Oliveira	30
1980	Cartas em Sisal, Correspondências Amigas, o acervo de Valente, Bahia	69
1990	Cartas em Sisal, Correspondências Amigas, o acervo de Valente, Bahia	17
2000	Cartas em Sisal	1
Sem data no século XX	Cartas do Acervo Dantas Jr., Correspondências Amigas, o acervo de Valente, Bahia e do Acervo da família Freire	12
Total de cartas do século XX		582
Total Geral de Cartas		1.084

Fonte: Elaborado pelas autoras (2019).

O Mapa 1 a seguir traz uma amostragem do local de nascimento dos remetentes dos acervos do século XIX, no qual se observa que há representante de praticamente todas as províncias:

MAPA 1 - MAPA DO BRASIL DO SÉCULO XIX, COM UMA AMOSTRAGEM DO LOCAL DE NASCIMENTO DOS REMETENTES



Fonte: Carneiro (2005).

O Mapa 2 a seguir mostra uma amostragem do local de nascimento dos remetentes dos acervos do século XX.

MAPA 2 - MAPA DO BRASIL DO SÉCULO XX, COM UMA AMOSTRAGEM DO LOCAL DE NASCIMENTO DOS REMETENTES

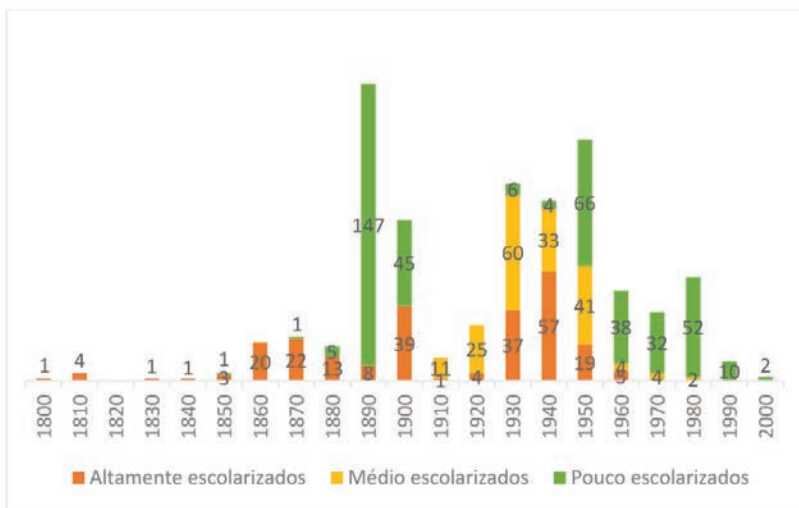


Fonte: Carneiro *et al.* (2011).

A seguir, apresenta-se – no conjunto das cartas brasileiras que integram o CE-DOHS – a distribuição das cartas baianas, por década de escrita e por escolaridade⁵:

⁵ Ver a pesquisa de Lacerda, Andrade e Carneiro (2016) sobre formas tratamentais nessas cartas, considerando as variáveis década de escrita da carta e escolaridade do remetente.

GRÁFICO 1 - DISTRIBUIÇÃO DAS CARTAS BAIANAS POR DÉCADA DE ESCRITA E POR ESCOLARIDADE



Fonte: Elaborado pelas autoras (2020).

Uma descrição completa de cada acervo do projeto e da ficha biográfica de cada remetente pode ser vista em <www.uefs.br/cedohs>; uma parte dos acervos referidos também consta em Carneiro *et al.* (2011).

Para além dos acervos de cartas, entre os quais cartas de inábeis⁶, o CE-DOHS trabalha com livros de fazenda manuscritos, dos séculos XVIII e XIX⁷, textos impressos, dos séculos XX e XXI⁸, e amostras de fala do século XX⁹. Todo esse

⁶ São 91 cartas escritas por 43 baianos “inábéis” do interior da Bahia, editadas por Santiago (2012). Esse acervo tem especial relevância para a Linguística Histórica, por ser uma amostra representativa da escrita por *mãos inábéis* – termo consagrado pela tradição paleográfica (MARQUILHAS, 2000) –, considerando-se a dificuldade de encontrar textos que refletem a escrita cotidiana, vernacular, produtos de indivíduos com baixo nível de letramento.

⁷ Ver Lacerda e Carneiro (2016), que apresentam o Livro do Gado e o Livro de Razão da Fazenda do Campo Seco-Bahia.

⁸ Os textos impressos estão disponíveis na obra *Publica-se em Feira de Santana: das cartas de leitores e redatores e dos anúncios em O Progresso e Na Folha do Norte (1901-2006)*, organizada por Carneiro e Oliveira (2012).

⁹ Esse *corpus* oral, representativo do português popular brasileiro da década de 90 do século XX, é produto do projeto A Língua Portuguesa no Semiárido Baiano, do NELP (www.uefs.br/nelp), e foi publicado, em 2008, por Almeida e Carneiro, na *Coleção amostras da língua falada no semiárido baiano*.

material foi selecionado tendo em vista os interesses de investigação do projeto: textos diacrônicos, manuscritos, impressos e orais, nas vertentes popular e culta, para constituição de banco de dados para estudo da história do PB.

A base documental aqui apresentada pode ajudar nos estudos sobre variação e mudança do PB; no que concerne à documentação epistolar, estudos do século XIX ao século XX, se se considerar a data de escrita das cartas; a partir do século XVIII, se a cronologia for feita por data de nascimento de seus remetentes. As informações sobre o grau de escolaridade dos escreventes, aliadas a possíveis análises do tipo de escrita, permitem também a separação das cartas representativas de variedades *standard* e não *standard* do PB, em uma perspectiva histórica.

EDIÇÃO FILOLÓGICA E EDIÇÃO DIGITAL

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

Concluídas as etapas do processo de constituição dos bancos de textos eletrônicos e de localização e seleção de documentos, os arquivos são transcritos segundo normas filológicas conservadoras e, a partir dessa transcrição, realiza-se, na etapa seguinte, a edição digital, em linguagem XML, finalizando com o preenchimento dos metadados.

No projeto CE-DOHS, os textos-fonte são apresentados em edição semidiplomática, segundo as normas de transcrição do PHPB, sendo oferecidas informações sobre os documentos, sua descrição extrínseca e intrínseca, e, sempre que possível, dados biográficos sobre os autores ou, no caso das cartas, sobre os remetentes e os destinatários, como nome, origem, idade, nível de escolaridade, profissão, estado civil etc. A codificação dos dados, textuais e extratextuais (ou metadados) é feita com o uso da ferramenta *eDictor*, o que possibilita a conversão

dos textos para diferentes formatos (TXT, XML, HTML) e evita problemas de processamento eletrônico.

As edições filológicas, fidedignas ao texto original, realizadas segundo critérios de transcrição bem definidos, ganham, nos *corpora* digitais, uma versão modernizada, com padronização da grafia, da acentuação e desenvolvimento de abreviaturas. Qualquer alteração realizada fica visível para o leitor, o que possibilita o controle e mapeamento das intervenções realizadas nos textos, garantindo a recuperabilidade das formas originais. Entretanto, na edição digital, as mudanças de parágrafo, de linha, as correções do autor, os acidentes do suporte, a orientação da escrita etc. são respeitadas. Com isso, oferece-se uma versão eletrônica de textos sem perder o rigor filológico.

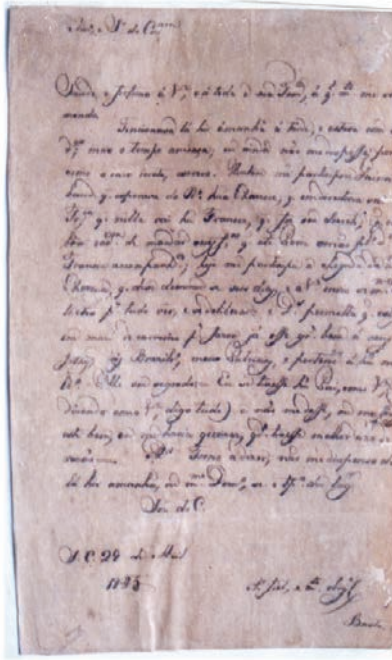
Procura-se, no âmbito do CE-DOHS, seguir os mesmos critérios de edição digital que seguem outros projetos de *corpora* eletrônicos, como o Corpus Histórico do Português Tycho Brahe (UNICAMP), o Labor Histórico, da UFRJ, o Post Scriptum: Archivo Digital de Escritura Cotidiana em Portugal e Espanha na Época Moderna¹⁰, do Centro Linguístico da Universidade de Lisboa (CLUL), o que garante maior praticidade no trabalho e nas consultas e maior integração entre os pesquisadores.¹¹

A seguir, será apresentado o passo a passo da edição digital de textos, segundo a metodologia utilizada no CE-DOHS, tomando, como exemplo, uma carta do acervo Cartas para Vários Destinatários (1809-1904).

¹⁰ Disponível em: < <http://www.clul.ul.pt/pt/recursos/462-post-scriptum-home>>. Acesso em: 21 mar. 2020.

¹¹ Por ocasião do Workshop Construction and Use of Large Annotated Corpora, realizado na UNICAMP, em 2013, pela equipe do projeto Corpus Histórico do Português Tycho Brahe, do qual pesquisadores de diversos projetos de *corpora* eletrônicos participaram – entre eles o CE-DOHS –, reafirmou-se a importância de esses projetos seguirem os mesmos padrões de edição digital e de anotação morfossintática, tendo em vista a praticidade do trabalho e a integração dos pesquisadores.

FIGURA 1 - EDIÇÃO SEMIDIPLOMÁTICA COM FAC-SÍMILE



Carta 7

AIGHBA. Ant. pasta 5. Documento contendo dois fólhos. Papel almaço sem pautas. Fólho protegido por papel manteiga. Carimbo do IGHB na margem superior esquerda sob marca d'água ilegível. No último fólho, as informações relativas ao destinatário foram escritas na vertical.

Amigo, e Sr. do Coraçam|

Saude e fortuna á *Vosmice* e á toda a sua *Familia*, a qual muito me reco|mando.|

Tencionava lá hir amanhã á tarde, e estava com saul|dades; mas o tempo amiaça; eu ainda não me dispesso; pore[m] como o caso insta, escrevo. Hontem me participou *Guinae|baud* que esperava do Rio hũ a *Charrua*, que endireitara vai á| *França*, que n'ella vai hũ *Francez*, que foi seu *secretario*; porque era| bõa ocaizem de mandar seus *filhos*, que ate *Paris* serão pelo *dito*;*Francez* acompanhados; hoje me participa a chegada da mesma| *charrua*, que deve demorar se seis dias, e a *Vosmice* envio os mesmos bilhetes para tudo vêr, e se deliberar: e *Deus* permitta que esteja| em maré de carvoeiro para fazer já esse grande bem á seus| *filhos* dois *Brazileiros*, meus *Patricios*, e pertecendo á hũ meu| *Amigo*. Olhe em segredo: = Eu se Tivesse hum *Pai*, como *Vosmice*, (e| dizendo como *Vosmice*, digo tudo) e não me desse, ou me fizesse| este bem eu não havia queixar, quando tivesse melhor uzo da| razão= Adeus. Torno a dizer, não me dispesso de| lá hir amanhã, ou mesmo *Domingo*, se o tempo der lugar.|

Sou do *Coraçam|*

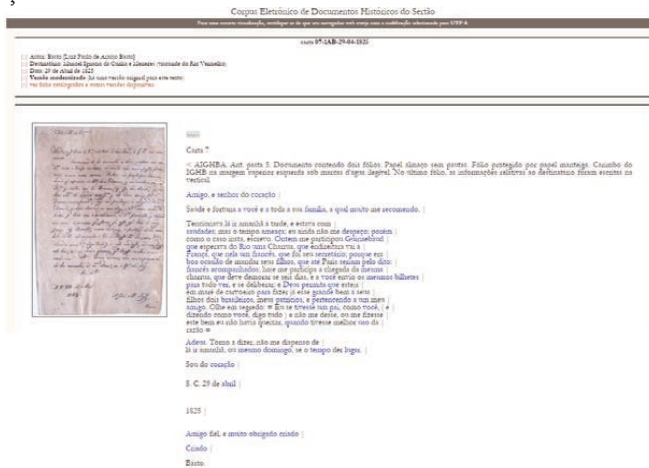
S. C. 29 de Abril|
1825|

Amigo fiel, e muito obrigado criado|

Basto.|

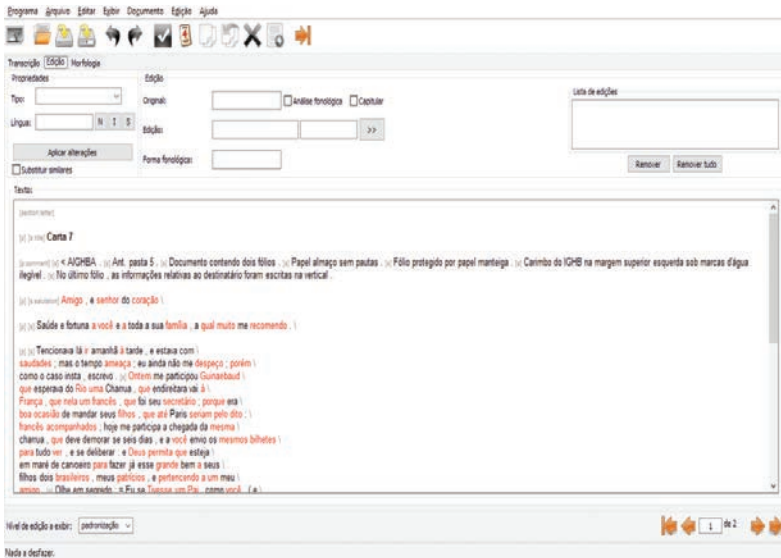
Fonte: Carneiro (2005).

FIGURA 2 - EDIÇÃO SEMIDIPLOMÁTICA COM FAC-SÍMILE NO EDICTOR



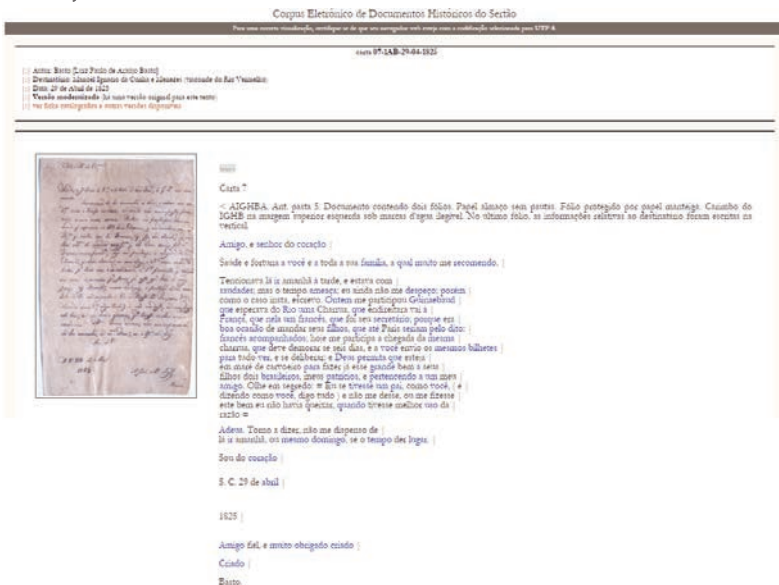
Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

FIGURA 3 - PROCESSO DE EDIÇÃO NO *EDICTOR*



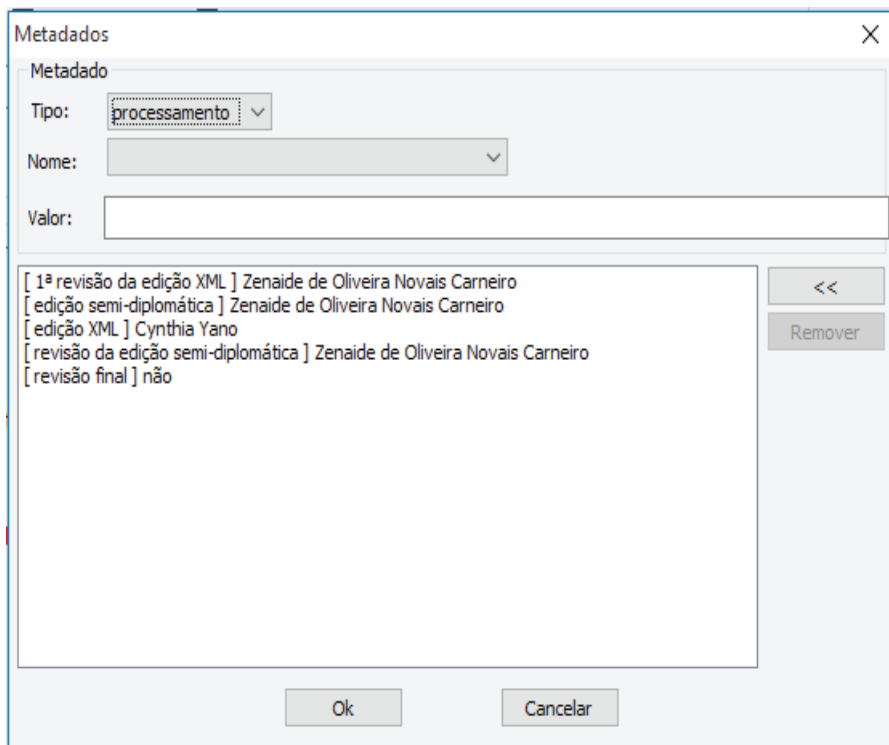
Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

FIGURA 4 - EDIÇÃO MODERNIZADA FINAL



Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

FIGURA 5 - METADADOS



Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

PADRONIZAÇÃO ORTOGRÁFICA DOS TEXTOS

A padronização ortográfica dos textos originais – que apresentam muitas variações ortográficas – é necessária na edição modernizada, que será anotada linguisticamente, porquanto dessa padronização depende uma maior eficiência de programas de etiquetagem automática. Não se interveio, todavia, nos regionalismos, arcaísmos lexicais e neologismos.¹²

Foram padronizadas palavras como: cete (sete); podião (podiam); acentarmos (acertarmos); emcomodos (incômodos); cabeça (cabeça); rezulvida (resolvida); intendo (entendo); avizarei (avisarei); instruçons (instruções); lógar

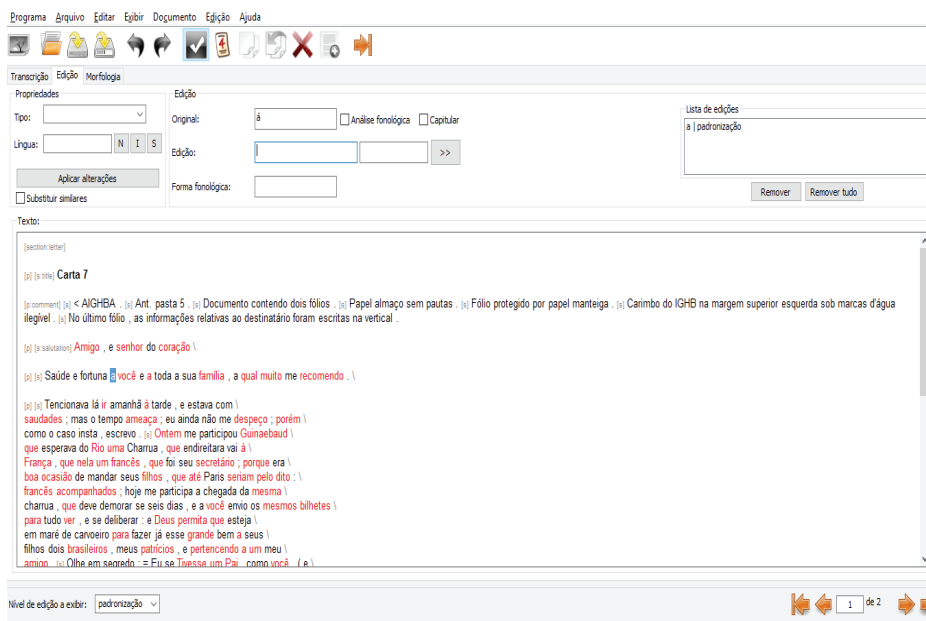
¹² No caso das amostras de fala, quase não foram feitas alterações.

(lugar); precipio (princípio); fis (fiz); conclusão (conclusão); comferir (conferir); conversou (conversou); podia (podia); piriodos (períodos); poderse (pudesse); munto (muito).

As seguintes palavras são exemplos de padronização da acentuação: debito (débito); tirar (tirar); nos (nós); más (mas); credito (crédito); contavamos (contávamos); ultima (última); negocio (negócio).

Com o *eDictor*, é possível selecionar a palavra original, que apresenta variação, e editá-la segundo a grafia padrão, ficando disponível a lista de alterações realizadas na edição modernizada. Seleciona-se, manualmente, palavra por palavra que se deseja alterar – substituir, separar, juntar, expandir, etc. (Ver as Figuras 6 e 7).¹³

FIGURA 6 - PADRONIZAÇÃO ORTOGRÁFICA NO *EDICTOR*



Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

¹³ Os criadores do *eDictor* têm trabalhado para sofisticar a ferramenta, tornando-a mais inteligente, a fim de que os processos de edição digital dos textos e sua anotação morfossintática e sintática sejam mais automatizados, facilitando a constituição dos bancos eletrônicos.

FIGURA 7 - LISTA DE DADOS ALTERADOS

Item	Juncao	Segmentacao	sobrescrito	ilegivel	rasurado	subscrito	tachado	pontuacao	expansão	padronizacao
Abril										
acompanhos									acompanhados	
Ads									Adeus	
amiaga										ameaça
Amo									Amigo	
Ao									amigo	
Ao									Amigo	
ate										até
bilhetes		bilhetes								
Brazilrs									Brazileiros	brasileiros
bõa										boa
C									Coraçam	coração
C									criado	
C									Criado	
C									Cunha	
Coram									coraçam	coração
dispeso										despeço
do									dito	
Domo									domingo	
Ds									Deus	
eMenezes		e Menezes								
Exmo									Excelentissimo	Excelentissimo
Fama									familia	familia
fõs									filhos	

Fonte: Site do projeto CE-DOHS: <<http://www.uefs.br/cedohs/>>.

No projeto CE-DOHS, no período inicial das edições, os pesquisadores enfrentaram dificuldades para decidir o que padronizar; por exemplo, casos em que o pronome está ligado ao verbo, em posição enclítica, sem hífen, o que contraria a norma-padrão atual (athe que eu poça pençar a respeito de *poderse* cuidar neste negocio; *tendome* medicado veio o Jejuino para me dar conta, e eu poder *dalla*; *Vossa Excelencia* esta rezulvida a *intregarlhe* a obra etc.). Nesses casos, decidiu-se manter a ortografia original, porque isso faz diferença no estudo da sintaxe dos clíticos no PB.

Pensou-se inicialmente em desenvolver apenas as abreviaturas desconhecidas do leitor atual; depois os pesquisadores decidiram expandir todas as abreviaturas, em razão de sua expansão evitar erros na aplicação do anotador automático, que é incapaz de reconhecer a forma correspondente. Como se vê, há uma relação

dinâmica nas etapas de constituição de *corpora* digitais: uma decisão tomada em determinado nível (por exemplo, na padronização dos textos) tem consequências no nível posterior (por exemplo, na anotação linguística). As formas abreviadas, que se mantêm na edição semidiplomática, aparecem expandidas na edição XML, usando-se a etiqueta <expand>.

CONSIDERAÇÕES FINAIS

Hoje, no banco do projeto CE-DOHS, estão à disposição a edição semidiplomática dos documentos, com fac-símile; as transcrições das amostras de fala; a edição modernizada dos textos; resumos com a contextualização sócio-histórica dos materiais e com informações sobre os autores e, no caso das cartas, sobre os destinatários, além da ficha técnica, com nome dos editores, revisores etc.

Os textos que compõem o CE-DOHS estão ainda na forma não anotada, ou seja, sem informações linguísticas. Essa tarefa será desempenhada no próximo ano, depois de realizadas oficinas, junto ao projeto Corpus Histórico do Português Tycho Brahe, para treinamento de bolsistas de mestrado e doutorado que iniciarão a etiquetagem morfosintática e anotação sintática do material.

O CE-DOHS possui um material extenso e rico, que oferece à comunidade científica diferentes possibilidades de pesquisa; e, para a história do PB – especialmente do português no interior da Bahia, através de um contínuo, do mais escolarizado para o menos escolarizado, *os inábeis* –, trata-se de um *corpus* extremamente significativo.

REFERÊNCIAS

ALMEIDA, N. L. F; CARNEIRO, Z. O. N. *Projeto A língua falada no semiárido baiano*. Disponível em: <http://www.uefs.br/nelp/sobre_projeto.htm>.

ALMEIDA, N. L. F. de; CARNEIRO, Z. de O. N. (Org.). *Coleção amostras da língua falada no semiárido baiano*. Feira de Santana, Bahia: UEFS, 2008.

BACELAR DO NASCIMENTO, M. F. *O lugar do corpus na investigação linguística*. 2004. Disponível em: <<http://www.clul.ul.pt/equipa/berlim-2000-nascimento.pdf>> . Acesso em: 09 set. 2020.

AUTORIA. A língua falada no semiárido baiano. Disponível em: <http://www.uefs.br/nelp/sobre_projeto.htm>. Acesso em: 09 set. 2020.

CARNEIRO, Z. O. N. *Cartas brasileiras (1809-1907): um estudo filológico-linguístico*. 2005. 2329 f. Tese (Doutorado) - Universidade Estadual de Campinas, Campinas, 2005.

CARNEIRO, Z. O. N. (Org). *Cartas brasileiras (1809-2000): coletânea de fontes para o estudo do português*. Feira de Santana, Bahia: UEFS, 2011.

CARNEIRO, Z. O. N.; LACERDA, M. F. O. *Publica-se em Feira de Santana: das cartas de leitores e redatores e dos anúncios em O Progresso e Na Folha do Norte (1901-2006)*. Feira de Santana, Bahia: UEFS, 2012.

CE-DOHS. Corpus Eletrônico de Documentos Históricos do Sertão. Disponível em: <<http://www.uefs.br/cedohs>>. Acesso em: 09 set. 2020.

CRANE, G. *ePhilology: when the boooks talk to their readers*. Blackwell Companion to Digital Literary Studies. Oxford: Blackwell, 2008.

GALVES, C.; BRITTO, H. *The Tycho Brahe Corpus of Historical Portuguese*. Technical Report, Department of Linguistics, University of Campinas. 2002.

GONÇALVES, M. F.; BANZA, A. P. Fontes de metalinguísticas para a história do português clássico. In: GONÇALVES, M. F.; BANZA, A. P. *Patrimônio Textual e Humanidades Digitais: da antiga à nova filologia*. Évora: CIDEHUS, 2013. p. 73-112.

LABOV, W. *Sociolinguistic Patterns*. Pennsylvania: University of Pennsylvania Press, 1972.

LACERDA, Mariana Fagundes de Oliveira; ANDRADE, Aroldo Leal de; CARNEIRO, Zenaide de Oliveira Novais. Formas tratamentais em cartas baianas: sujeito e outras funções. *Cadernos de Estudos Linguísticos*, v. 58, n. 2, p. 257-276, 2016.

LACERDA, M. F. de O.; CARNEIRO, Z. de O. N. Edição filológica e digital do Livro do Gado e do Livro de Razão do arquivo do Sobrado do Brejo (Bahia setecentista e oitocentista). *Labor Histórico*, v. 2, n. 1, p. 151-163, 2016. <http://dx.doi.org/10.17074/lh.v2i1>.

MARQUILHAS, R. *A faculdade das letras: leitura e escrita em Portugal no séc. XVII*. Lisboa: Imprensa Nacional-Casa da Moeda, 2000.

MATTOS E SILVA, R. V. De fontes sócio-históricas para a história social linguística do Brasil: em busca de indícios. In: MATTOS E SILVA, R. V. (Org.). *Para a história do*

português brasileiro: primeiros estudos. São Paulo: Humanitas/FFCHL/USP: FAPESP, 2001. p. 275-302. Tomos I e II.

PAIXÃO DE SOUSA, M. C.; KEPLER, F. N.; FARIA, P. P. F. E-Dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: SHEPHERD, T.; SARDINHA, T. B.; PINTO, M. V. (Org.). *Caminhos da linguística de corpus*. Campinas: Mercado de Letras, 2010.

PETRUCCI, A. *La ciencia de la escritura*: primera lección de paleografía. Buenos Aires: Fondo de Cultura Económica de Argentina, 2003.

SANTIAGO. H. S. *Um estudo do português popular brasileiro em cartas pessoais de “mãos cândidas” do sertão baiano*. 2012. 423 f. Dissertação (Mestrado) – Universidade Estadual de Feira de Santana, Feira de Santana, 2012.

SANTOS, J. V.; BRITO, G. S. Fotografia técnica de documentos para formação de corpora digitais eletrônicos: o método Lapelinc. *Revista LETRAS & LETRAS*, v. 30, n. 2, p. 24-36, 2014.

SCHREIBMAN, S. *et al. A Companion to Digital Humanities*. Oxford: Blackwell, 2004.

Humanidades Digitais e as instituições memoriais brasileiras: evolução em descompasso

Aquiles Alencar Brayner

Ainda que não haja um consenso estabelecido do que venham a ser as práticas e o escopo das Humanidades Digitais (HDs), o tema vem ganhando cada vez mais espaço nas publicações e debates acadêmicos. Via de regra, o termo HD é empregado em relação às novas metodologias de pesquisa surgidas nas últimas décadas, baseadas no uso das tecnologias de informação (TICs), ferramentas e técnicas computacionais para criação, seleção, extração, descrição, mapeamento e análise de dados apresentados em formato eletrônico. Todavia, e apesar da sua recente popularização no campo das Ciências Humanas, é incontestável o fato de que as HDs estão ampliando os horizontes das pesquisas realizadas no campo das Humanidades, permitindo-nos elaborar novos questionamentos e chegar a novas descobertas nunca antes imaginadas.

Por um lado, as HDs estão gerando novos *insights* sobre o conhecimento acumulado do passado e contextualizando as informações e dados que emergem no presente; por outro, seu campo de atuação abre um novo espaço de pesquisa cujo propósito abrange tanto o estudo da influência das TICs na formação do sujeito contemporâneo quanto a compreensão do crescente impacto social, político, econômico e cultural que o constante desenvolvimento tecnológico exerce nas relações que estabelecemos com a produção do conhecimento na sociedade informatizada. Alguns entusiastas chegam até mesmo a prever que o termo “Humanidades Digitais” se tornará, dentro de alguns anos, obsoleto, já que toda e qualquer pesquisa em Ciências Humanas se desenvolverá necessariamente sob o crivo da tecnologia, da utilização de dados massivos, mais comumente denominados como Big Data, e do acesso contínuo a um número exponencial de dados e objetos eletrônicos.

De fato, nos cabe questionar não apenas o emprego do termo “Digital” em referência à pesquisa no campo das Humanidades, mas, principalmente, repensarmos como a tecnologia irá transformar, como já o vem fazendo, as fronteiras de atuação

entre os pesquisadores das Ciências Humanas e os profissionais de outras áreas acadêmicas tradicionalmente entendidas como desvinculadas entre si por adotarem métodos de pesquisa e objeto de estudos nem sempre convergentes. A adoção de um modelo de trabalho colaborativo com pesquisadores de várias outras áreas como a Sociologia, Economia, Geografia, Ciência da Informação, Informática, para citar apenas algumas delas, faz com que os humanistas digitais estejam constantemente ampliando o seu campo de investigação para alcançar metas compartilhadas que atendam os objetivos de pesquisa específicos de cada uma destas diferentes áreas.

A perspectiva multidisciplinar é o que, de fato, caracteriza o trabalho das HDs, facilitando uma atuação conjunta a partir de uma amplitude de perspectivas distintas que possibilitam o entrelaçamento de ideias e metodologias de trabalho até então nunca exploradas de forma tão dinâmica. Como tal, as HDs estão cada vez mais presentes em universidades, centros de pesquisa e principalmente em instituições de memória cultural, como bibliotecas, arquivos e museus. Os inúmeros projetos de digitalização e disponibilização dos acervos memoriais vêm fornecendo novos modelos de trabalho que facilitam a melhoria na descrição dos objetos institucionais. A interação entre usuários e acervos em convite aberto à participação pública através de atividades de *crowdsourcing* tem contribuído de maneira significativa para o enriquecimento dos metadados catalográficos de coleções institucionais e possibilitando a criação de novos objetos memoriais. Conforme afirma Ridge (2014, p. 6, tradução nossa)¹

[...] as tarefas realizadas pelos participantes no *crowdsourcing* do patrimônio cultural envolvem a transformação de conteúdo de um formato para outro (por exemplo, transcrever texto ou notação musical), descrever artefatos (através de etiquetas, classificações, anotações estruturadas ou texto livre), sintetizar novos conhecimentos ou produzir novos objetos criativos (como fotografia ou design).

O envolvimento público na descrição e criação de conteúdos auxilia o trabalho dos pesquisadores na recuperação da informação e acesso a novos recursos que complementam os acervos presentes nas instituições memoriais.

¹ “the task performed by participants in cultural heritage crowdsourcing involve transforming content from one format to another (for example, transcribing text or musical notation), describing artefacts (through tags, classifications, structured annotations or free text), synthesising new knowledge, or producing creative artefacts (such as photography or design)”.

As atividades de folksonomia², principalmente os processos de etiquetagem, oferecem uma grande vantagem aos catálogos *on-line* permitindo que pesquisadores e usuários em geral possam enriquecer as informações sobre os objetos e dados institucionais, fazendo uso de terminologias mais pertinentes às suas comunidades, as quais muitas vezes não se encontram representadas nos vocabulários controlados criados e compartilhados por bibliotecários, arquivistas e museólogos. Estas atividades estabelecem uma mudança de paradigma da autoridade absoluta tipicamente representada pelos profissionais de instituições memoriais até o final do século XX, abrindo espaço para uma nova expressão de autoridade contextual em que os objetos culturais passam por um processo de relativização ao serem representados a partir das suas diferentes perspectivas de significância e usabilidade.

A contribuição dos usuários na descrição de objetos, além de enriquecer substancialmente a terminologia ou vocabulário controlado utilizado por indexadores no processo de catalogação de acervos, possibilita sobretudo uma mais eficiente recuperação de informação entre grupos de usuários que compartilham um mesmo campo semântico marcado por peculiaridades linguísticas (uso de sinonímias, gírias, termos regionais etc.), muitas vezes não contempladas no léxico empregado pelos profissionais da informação.

O enriquecimento de coleções digitais institucionais e a formação de novos acervos através de uma cultura participativa são, indubitavelmente, fatores importantes que ampliam o apoio das instituições de memória aos pesquisadores em HDs. Os programas de construção de acervos eletrônicos, em sua grande maioria financiados por recursos públicos, estão cada vez mais voltados a projetos que visam a complementação de coleções *on-line*, aumentando as possibilidades de acesso remoto a seus objetos e potencializando as fontes de pesquisas para acadêmicos, estudantes e público em geral. O projeto Europeia 1914-1918³, uma iniciativa pensada pela Comunidade Europeia como parte das celebrações do centenário da Primeira Guerra Mundial, ilustra de maneira clara como as instituições memoriais

² O termo folksonomia designa a contribuição dos usuários na descrição de objetos bibliográficos, museológicos e arquivísticos realizada em catálogos eletrônicos institucionais ou em plataformas abertas de disseminação de acervos.

³ Disponível em: <<https://www.europeana.eu/portal/pt/collections/world-war-I>>. Acesso em: 12 fev. 2020.

têm unido esforços para complementar as suas coleções temáticas de maneira colaborativa utilizando, para isso, plataformas digitais que possibilitem tanto a unificação de acervos espalhados em várias bibliotecas, arquivos e museus europeus, quanto a agregação de novos objetos a esta vasta coleção compilada. Muito mais do que coordenar um programa de digitalização interinstitucional, este projeto colaborativo contou, sobretudo, com a participação das comunidades de usuários que, ao digitalizar os seus acervos privados contendo fotos, cartas e outros objetos da época, possibilitaram redimensionar os acervos memoriais, normalmente composto por documentos e objetos oficiais e que pouco representam as percepções dos cidadãos comuns que vivenciaram a guerra no seu cotidiano. Estas narrativas pessoais, quando adicionadas às coleções institucionais, oferecem uma fonte inestimável para uma compreensão mais ampla e completa do contexto histórico em que os objetos memoriais se inserem.

Outra importante contribuição de bibliotecas e arquivos para as HDs tem sido a implementação e gestão de repositórios institucionais visando não somente a consulta de publicações acadêmicas, mas oferecendo, sobretudo, acesso a bancos de dados e objetos criados ou organizados por pesquisadores para que estes materiais possam ser reutilizados e/ou manipulados por outros usuários. O modelo tradicional de repositório institucional, usualmente entendido como espaço de acesso à informação e objetos eletrônicos, deve ser ampliado no ambiente das HDs de modo a contemplar uma interação mais dinâmica, e ao mesmo tempo mais integrada, entre pesquisadores e as fontes informacionais acessadas. O que realmente interessa ao usuário, na maior parte das vezes, não é a interpretação dos dados utilizados que corroborem as ideias de uma determinada publicação, mas sim o acesso direto a estes dados para se obter um melhor discernimento sobre o processo criativo do autor, ou mesmo a reutilização dos dados depositados para fins distintos daquele em que foram originalmente criados. Este novo papel dos repositórios digitais surge em resposta ao

[...] rápido crescimento da cultura participante que, incorporado pela Web 2.0, tem visto a produção criativa ultrapassar o acesso básico como o principal motivo para interação com bancos de dados, arquivos e mecanismos de busca por usuários públicos. A exploração intuitiva de diversos tipos de dados permite

que os usuários encontrem novos significados em vez de simplesmente acessar informações. (KENDERLINE, 2016, p. 24, tradução nossa⁴)

É neste novo contexto interativo de uma cultura participativa facilitada pelas plataformas Web 2.0 que a função dos profissionais da informação e curadoria de acervos digitais está se transformando. Antes entendidos como guardiões (*gatekeepers*) institucionais, bibliotecários, arquivistas e museólogos assumem hoje uma posição bem distinta daquela que lhes era tradicionalmente confiada. Sua atenção não se encontra mais centrada exclusivamente no gerenciamento e preservação das coleções em si, mas amplia-se de modo a explorar novos canais de comunicação que possam alcançar distintos públicos, aproximando os acervos institucionais a novas comunidades de usuários. Conforme afirmam Palmer e Stevenson (2011, p. 18), a “Web 2.0 nos dá a oportunidade de pensar não apenas em promover nossas coleções por meio de sistemas de busca tradicionais e online, mas também apresentá-las de forma mais criativa de modo a fomentar o diálogo e construir comunidades de usuários em torno dos arquivos” (PALMER; STEVENSON, 2014, p. 18, tradução nossa⁵).

Com as mudanças de paradigmas em instituições memoriais, seus profissionais assumem um papel cada vez mais ativo no processo de desenvolvimento e publicação de pesquisas. A participação dos profissionais da informação se dá principalmente pela reconfiguração constante dos objetos e dados digitais através do uso combinado de técnicas computacionais, como o georeferenciamento de conteúdos e a mineração de textos, para citar aqui apenas alguns exemplos, explorando, de forma criativa e original, as ferramentas disponíveis e desenvolvendo novos processos de manipulação de dados e objetos eletrônicos que possam apoiar as várias etapas de pesquisa.

⁴ “The rapid growth in participant culture embodied by Web 2.0 has seen creative production overtake basic access as the primary motive for interaction with databases, archives, and search engines by public users. Intuitive exploration of diverse bodies of data allows users to find new meanings rather than simply access information”.

⁵ “Web 2.0 gives us opportunity to think not just about promoting our collections through online and traditional finding aid but also about working to present them more imaginatively – to engage in dialogue and build communities in and around archives”.

O processo colaborativo entre pesquisadores, bibliotecários, arquivistas e museólogos, que tem definido o termo Humanidades 2.0, é uma consequência direta do progresso tecnológico que abriu novos canais, metodologias e práticas profissionais, influenciando a “criação, curadoria, colaboração, experimentação e multi-propósito ou multifuncionalidade do conhecimento humanístico” (PRESNER, 2010, p. 73, tradução nossa)⁶.

O caráter experimental das HDs, principalmente no contexto da Humanidades 2.0, vem aproximando pesquisadores e instituições memoriais em uma relação mutuamente benéfica para os agentes envolvidos. Estas experimentações ocorrem, via de regra, nas bibliotecas, arquivos e museus, como nos mostra o crescente número de laboratórios digitais mantidos pela British Library⁷, pelo Arquivo Nacional⁸, e pelo Museu da Ciência⁹. Além das atividades de processamento de dados por meio de técnicas computacionais, os novos experimentos com acervos digitais incluem a utilização de plataformas abertas como Flickr¹⁰ e Zooniverse¹¹, que facilitam um canal aberto de comunicação e interação com acervos públicos entre profissionais da informação, pesquisadores e público em geral.

Atualmente, grupos como a Aliança das Organizações Internacionais em Humanidades Digitais¹², a Rede de Humanidades Digitais¹³ e, no caso dos países lusófonos, a Associação das Humanidades Digitais¹⁴, vêm desenvolvendo programas de pesquisa, conferências e atividades interdisciplinares envolvendo pesquisadores das mais variadas áreas científicas. A participação conjunta destes

⁶ “creation, curation, collaboration, experimentation, and the multi-purposing or multi-channeling of humanistic knowledge.”

⁷ Disponível em: <<http://labs.bl.uk/>>. Acesso em: 8 dez. 2016.

⁸ Disponível em: <<http://www.nationalarchives.gov.uk/about/our-research-and-academic-collaboration/events-and-training/digital-experimentation-workshops/>>. Acesso em: 5 jan. 2019.

⁹ Disponível em: <<https://lab.sciencemuseum.org.uk>>. Acesso em: 16 dez 2018.

¹⁰ Disponível em: <<https://www.flickr.com>>. Acesso em: 16 dez. 2018.

¹¹ Disponível em: <<https://www.zooniverse.org>>. Acesso em: 16 dez. 2018.

¹² Disponível em: <<https://adho.org/>>. Acesso em: 7 jan. 2019.

¹³ Disponível em: <<http://www.humanidadesdigitales.net/>>. Acesso em: 20 dez. 2018.

¹⁴ Disponível em: <<https://ahdig.org/>>. Acesso em: 20 dez. 2018.

pesquisadores tem suscitado novos questionamentos e interpretações sobre a nossa história, língua, arte e cultura. No Brasil, várias iniciativas e projetos em HDs estão sendo desenvolvidos em laboratórios digitais por grupos de investigação vinculados a departamentos acadêmicos e centros de pesquisa, como é o caso do Laboratório Liber da Universidade Federal de Pernambuco (UFPE)¹⁵.

Apesar do aumento no número de atividades voltadas às HDs, o que se nota ainda por parte das instituições memoriais brasileiras é uma desvinculação generalizada na participação em projetos de pesquisa envolvendo conteúdos e dados eletrônicos. Apesar dos financiamentos obtidos para a digitalização de seus acervos através de recursos federais ou mesmo por iniciativas privadas, os museus, bibliotecas e arquivos do país têm se preocupado em desenvolver plataformas acesso aos seus conteúdos digitalizados sem, no entanto, promover o reuso destes objetos e seus respectivos metadados por parte de pesquisadores e público em geral. Como bem observa Valente ⁽²⁰¹⁷⁾, o grande desafio que se lança às instituições de memória brasileiras no ambiente digital é pensar em “como estimular o uso criativo dos acervos” já que “colocar os materiais online, pura e simplesmente, pode ser insignificante se não se criar interesse em torno deles” (p. 16). Ainda em relação aos programas nacionais de digitalização de acervos, a autora aponta que “a falta de planejamento quanto a passos futuros, pós-digitalização e disponibilização” (p. 15) põe em risco a sustentabilidade dos recursos digitalizados, principalmente pela ausência de uma política nacional de preservação digital que garanta o acesso deste material para pesquisadores no futuro.

Em se tratando do arquivamento e gestão de conteúdos nato-digitais, a situação brasileira ainda é mais alarmante, principalmente pelo fato de não termos, até o presente momento, um programa nacional que garanta a preservação de informações publicadas em rede através de páginas Web, *blogs* ou mesmo informações veiculadas por instituições e órgãos oficiais através de postagens em plataformas de mídias sociais. Há 16 anos, ainda nos primórdios das publicações em páginas Web, estimava-se haver então cerca de 170 terabytes de informação disponibilizada em rede; um volume 17 vezes maior que o número total do acervo de obras impressas na Biblioteca do Congresso nos Estados Unidos (LYMAN; VARIAN, 2003). Segundo estudos realizados sobre a efemeridade das páginas Web,

¹⁵ Disponível em: <<http://www.liber.ufpe.br>>. Acesso em: 7 jan. 2019.

se constatava então que “a vida útil média de uma página Web em 2003 era estimada em 100 dias e não seria desproposital afirmar que este prazo seja ainda mais curto na atualidade” (KELLY; PENNOCK, 2006, n. p., tradução nossa)¹⁶. O problema se agrava quando levamos em conta não somente a alarmante taxa de desaparecimento destas páginas mas, sobretudo, o alto índice de modificações nas informações publicadas durante o tempo em que as páginas se encontravam disponíveis para acesso *on-line*. Mesmo que não seja possível estabelecer uma estimativa mais completa sobre as modificações periódicas feitas nas informações publicadas em sítios Web, Brügger (2005) sugere que 80% das páginas que ainda estão disponíveis para acesso público sofreram algum tipo de alteração desde a sua publicação inicial. Essa alta taxa de volatilidade dos conteúdos Web, que vão desde a atualização da informação publicada até a reestruturação completa de suas páginas, incluindo mudanças de provedor e exclusão de endereços na Internet, gera uma situação extremamente adversa aos pesquisadores. Afinal, é impossível se obter um panorama histórico, político e cultural da sociedade atual sem se ter acesso aos recursos informacionais que lhe são mais importantes como forma de expressão pública. Sem um programa de preservação das informações publicadas na Internet, pesquisadores da área de Humanidades irão deparar-se com um incomensurável problema que afetará sobremaneira a sua compreensão sobre a era da comunicação eletrônica. Segundo nos prognostica Brindley (2009, *on-line*), “historiadores e cidadãos do futuro encontrarão um buraco negro na base de conhecimento do século 21” caso não preservemos a nossa memória digital.

Dentro do vasto universo informacional oferecido pelas publicações *on-line*, cabe aqui ressaltar o papel preponderante exercido pelos blogues como forma de expressão informal e subjetiva e que, por conseguinte, vêm sendo cada vez mais utilizados por pesquisadores e instituições para comunicar para uma audiência mais abrangente os novos questionamentos científicos ou fornecer uma contextualização mais ampla de seus acervos, complementando os “meios formais de comunicação científica e serv[indo], essencialmente, como repositório pessoal e coletivo de conteúdo e meio de difusão seletiva de informação” (GOMES; TORACI; FLORES, 2012, p. 7). No campo dos estudos literários, para citar aqui uma área de pesquisa

¹⁶ “the average lifespan of a Web page in 2003 was deemed to be 100 days and it is not unreasonable to suggest that it is even shorter today.”

mais tradicionalmente ligada às Humanidades, já não se pode mais ignorar a importância que os *blogs* exercem tanto na formação de escritores iniciantes quanto na carreira de autores já consagrados em nossa época. “Para aquele que aceite enfrentar o desafio de pensar o contemporâneo” afirma Azevedo (2008, p. 32), “é quase impossível escapar do fato de que os salões virtuais da web invadiram a cena literária” (p. 32). A omissão que ainda se observa por parte das instituições culturais na seleção, arquivamento e preservação de *blogs* e páginas Web referentes à cultura brasileira, mesmo os conteúdos produzidos pelas próprias instituições que os deveriam resguardar, resultará, indubitavelmente, em uma perda substancial da nossa memória cultural.

Diante deste grande desafio imposto à preservação das publicações em rede, vários países começaram a atuar a nível de cooperação internacional para garantir que os seus patrimônios digitais nacionais permaneçam acessíveis a gerações futuras. O Consórcio Internacional de Preservação da Internet¹⁷, constituído por mais de 45 países membros, é uma das associações pioneiras nas discussões sobre projetos e práticas de trabalho que garantam uma maior interoperabilidade entre sistemas de coleta e preservação de páginas Web. A maior parte dos programas de arquivamento de páginas Web está concentrada na Europa e América Norte, com pouca repercussão em outras áreas do planeta. Atualmente, o Chile é a única nação na América Latina a ter um programa oficial de arquivamento de páginas Web registradas no domínio *.cl*, mantido pela Biblioteca Nacional do país desde 2013 e em consonância com a sua legislação de depósito legal eletrônico (AGUIRRE, 2014).

No Brasil, ainda que a Lei de Depósito Legal 10.994/2004 estabeleça em seu artigo 2º a obrigatoriedade de se preservar como parte do patrimônio cultural e intelectual da nação “todas as publicações, produzidas por qualquer meio ou processo, para distribuição gratuita ou venda”(BRASIL, 2004), não existe, até o presente momento, nenhum programa por parte do governo para o arquivamento de páginas Web no domínio nacional.

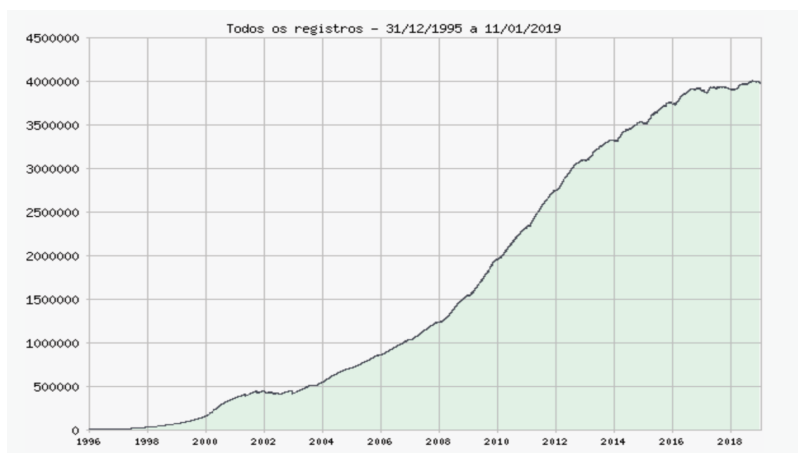
A Fundação Biblioteca Nacional, instituição responsável por assegurar a execução da lei de depósito legal no Brasil, tampouco tem apresentado propostas na implementação de um programa de arquivamento de páginas Web que garanta a

¹⁷ Disponível em: <<http://netpreserve.org>>. Acesso em: 7 dez. 2011.

custódia e preservação de conteúdos digitais publicados na Webosfera nacional de relevância científica, histórica e cultural para a nação.

Tomando como referência os dados estatísticos disponibilizados pelo site Registro.br, entidade responsável pela emissão de endereços no domínio .br da Web, foram cadastrados, desde 1996, cerca de 4 milhões de sítios Web no Brasil, atingindo um alto índice de crescimento principalmente nas duas últimas décadas.

FIGURA 1 - EVOLUÇÃO NO NÚMERO DE REGISTROS DE NOVOS SÍTIOS WEB NO DOMÍNIO .BR ENTRE 1996 E 2019



Fonte: Registro.br (2019). Disponível em: <<https://registro.br/estatisticas.html>>. Acesso em: 23 fev. 2019.

Se considerarmos que cada sítio brasileiro contém, em média, 85 páginas (MODESTO *et al.*, 2005), chegamos à conclusão de que o nosso domínio nacional possui algo em torno de 340 milhões de páginas Web. Algumas destas páginas, seguramente com informações de enorme valor para projetos acadêmicos, já desapareceram ou se encontram em perigo de extinção, representando uma perda irreparável de fontes informacionais para pesquisadores no futuro. Este quadro se mostra ainda mais preocupante quando comparamos a diferença entre as taxas de crescimento e os índices de remoção de sites no Brasil. Somente entre os meses de dezembro de 2018 e janeiro de 2019, o site Registro.br aponta uma diferença de 21.6% entre o número de sítios removidos (92.396) e o total de novos sítios registrados (70.714) no domínio nacional. A expansão no desaparecimento de páginas Web brasileiras mostra que, caso não sejam tomadas medidas urgentes

para a preservação dos conteúdos publicados on-line, os pesquisadores de HD interessados sobre o Brasil, em um futuro não muito longínquo, contarão com escassíssimas fontes digitais que possam apoiar os seus estudos.

As postagens publicadas em mídias sociais são, da mesma forma que as páginas Web, fontes que nos oferecem referências imprescindíveis sobre as atividades de comunicação entre usuários, órgãos públicos, partidos políticos, instituições de pesquisas e outros atores de relevância no atual cenário cultural e histórico brasileiros. Dados levantados pelo site Techcrunch apontam que em 2012 os usuários da plataforma *Facebook*, considerada a mídia social mais utilizada no mundo, geraram diariamente mais de 500TB de informação (CONSTINE, 2012). A importância em se arquivar pelo menos parte do que se é produzido nestas mídias se dá não tanto a nível das postagens individuais publicadas por instituições e personagens de relevância na atualidade, mas, sobretudo, pela possibilidade de se mapear os padrões comportamentais dos usuários nos espaços de interação Web 2.0. As redes sociais estão transformando as nossas sociedades ao inaugurar novos canais de expressão sócio-político e cultural que, não se restringindo a delimitações geográficas, favorece a comunicação direta e em tempo real entre bilhões de indivíduos. A ausência de uma política cultural de preservação do material produzido e publicado nos canais de mídia social resultará, seguramente, em uma perda irreversível de dados primordiais para pesquisas no campo das HDs.

O arquivamento do material nato-digital inclui, além da seleção e gestão de informações produzidas e disseminadas na Webosfera, atividades de extração e preservação dos chamados manuscritos digitais – conteúdos criados em computadores e dispositivos móveis. O processo de elaboração destes conteúdos, como os documentos produzidos em processadores de texto, permanece registrado nos discos rígidos de servidores, laptops, tablets e celulares, permitindo-nos, através da utilização de software forense, recompor a genealogia destes objetos partindo de sua origem e processos de alteração até chegar à sua versão final.

A análise forense de discos rígidos permite aos usuários acessar e revisar os arquivos nato-digitais em seu contexto e programas originais de criação, oferecendo, ainda, dados importantes sobre as atividades dos autores dos objetos digitais, tais como frequência de trabalho, conexão entre ideias apresentadas nos documentos, fontes referenciais obtidas por pesquisas em máquinas de busca, trocas de mensagens por correio eletrônico, consultas a bases de dados ou quaisquer outros

elementos metatextuais que tenham exercido alguma influência direta na produção do conteúdo analisado. Afinal, como apontam os profissionais que trabalham com arquivos digitais, “[p]reservar materiais nato-digitais significa preservar não apenas o objeto em si, mas também sua relação com outros objetos, ou sua posição como parte de um processo mais amplo. Esses relacionamentos [...] são o que torna cada arquivo único e insubstituível” (KIRSCHENBAUM; OVENDEN; REDWINE, 2010, p. 23, tradução nossa)¹⁸.

Tal como ocorre com os programas de arquivamento de páginas Web, muitas instituições internacionais de memória cultural vêm trabalhando conjuntamente com o objetivo de compartilhar experiências, desenvolver *softwares* abertos de extração e preservação de conteúdos nato-digitais e estabelecer novos parâmetros de trabalho e ferramentas computacionais que auxiliem pesquisadores. O consórcio BitCurator¹⁹, iniciado em 2016 e mantido com o apoio financeiro da fundação Andrew Mellon nos Estados Unidos, estabelece um sólido canal de comunicação e apoio a bibliotecários e arquivistas que trabalham no emergente campo da curadoria de acervos digitais.

No Brasil, ainda que não se conheça até o presente momento nenhum projeto desenvolvido ou implementado em instituições de memória cultural que faça uso de computação forense para o gerenciamento de recursos nato-digitais em apoio a pesquisas em HDs, acredita-se que, dada a crescente demanda informacional por parte dos usuários em acessar e analisar objetos e dados eletrônicos como parte essencial de suas pesquisas, bibliotecas e arquivos brasileiros deverão, dentro em breve, preparar-se para receber equipamentos digitais e computadores de uso pessoal para que estes aparelhos sejam incluídos nas suas políticas de formação e desenvolvimento de acervos.

Em face à ausência de uma atuação mais propiciamente voltada à gestão dos acervos nato-digitais no Brasil, cabe-nos alertar as instituições nacionais e gestores públicos sobre a urgência em se tomar as medidas cabíveis que garantam, para os

¹⁸ “Preserving born-digital materials means preserving not only the object itself but also its relationship to other objects, or its position as part of a larger process. Those relationships—how a file fits into a particular system, whether that system is actually the file system, a personal organizational strategy, or a much larger network—are what make each file unique and irreplaceable”.

¹⁹ Disponível em: <<https://bitcurator.net/bitcurator/>>. Acesso em: 8 jan. 2019.

pesquisadores do futuro, a preservação de uma parte importante da nossa memória digital contemporânea que, dado o seu alto índice de efemeridade, vem se tornando cada vez mais relegada ao esquecimento.

REFERÊNCIAS

AGUIRRE, R. Biblioteca Nacional Digital: un nuevo paso para la preservación de la memoria de Chile. *IFLA WLIC*, ago. 2014. Disponível em: <<http://library.ifla.org/1000/1/107-aguirre-es.pdf>>. Acesso em: 8 nov. 2018.

AZEVEDO, L. A. Autoficção e literatura contemporânea. *Revista Brasileira de Literatura Comparada*, v. 10, n. 8, p. 31-49, 2008. Disponível em: <<http://revista.abralic.org.br/index.php/revista/article/view/179>>. Acesso em 5 jan. 2019.

BRASIL. *Lei nº 10.994, de 14 de dezembro de 2004*. Dispõe sobre o depósito legal de publicações, na Biblioteca Nacional, e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10994.htm>. Acesso em: 6 dez. 2018.

BRINDLEY, L. Filling in a Digital Black Hole. *British Library*, 2009. Disponível em: <<http://www.bl.uk/pdf/digitalblackhole.pdf>>. Acesso em: 12 dez. 2018.

BRÜGGER, N. *Archiving Websites: General Considerations and Strategies*. Aarhus [Dinamarca]: The Centre for Internet Research, 2005.

CONSTINE, J. How Big is Facebook data? 2.5 billion Pieces of Content and 500+ Terabytes Ingested Every Day. *Techcrunch*, 2012. Disponível em: <<https://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/>>. Acesso em: 08 jan. 2019.

EDMOND, J. Collaboration and Infrastructure. In: SCHREIBMAN, S., SIEMENS, R., UNSWORTH, J. (Ed.). *A New Companion to Digital Humanities*. Oxford and Malden (MA): Wiley Blackwell, 2016. p. 54-65.

GOMES, D; MIRANDA, J.; COSTA, M. A Survey on Web Archiving Initiatives. In: PROCEEDINGS OF THE 15TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF DIGITAL LIBRARIES, 2011, Berlim. *Anais...* Berlim: Springer-Verlag, 2011. p. 408-420.

GOMES, I. M.; TORACI, V.; FLORES, N. Comunicação científica e cultura da participação: análise de blogs de ciência. In: CONGRESSO BRASILEIRO DE

CIÊNCIAS DA COMUNICAÇÃO, 35., 2012, Fortaleza. *Anais...* Fortaleza: Intercom, 2012. Disponível em: <http://www.intercom.org.br/sis/2012/resumos/R7-0689-1.pdf>. Acesso em: 22 dez. 2018.

KELLY, B.; PENNOCK, M. Archiving Web sites resources: a records management view. In: INTERNATIONAL WORLD WEB CONFERENCE, 15., 2006, Edinburgh. *Anais...* Disponível em: <https://purehost.bath.ac.uk/ws/portalfiles/portal/460441/wwwpp100-pennock.pdf>. Acesso em: 15 out. 2018.

KENDERLINE, S. Embodiment, Entanglement and Immersion in Digital Cultural Heritage. In: SCHREIBMAN, S., SIEMENS, R., UNSWORTH, J. (Ed.). *A New Companion to Digital Humanities*. Oxford and Malden (MA): Wiley Blackwell, 2016. p. 22-41.

KIRSCHENBAUM, M.; OVENDEN, R.; REDWINE, G. *Digital forensics and born-digital content in cultural heritage collections*. Washington D. C.: Council on Library and Information Resources, 2010. Disponível em: <https://www.clir.org/wp-content/uploads/sites/6/pub149.pdf>. Acesso em: 15 dez. 2018.

LYMAN, P.; VARIAN, H.R. How much Information. *Technical Report*, 2003. Berkley: University of Berkley, 2003. Disponível em: http://groups.ischool.berkeley.edu/archive/how-much-info-2003/printable_report.pdf. Acesso em: 08 dez. 2018.

MODESTO, M. Um novo retrato da web brasileira. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 32., 2005, São Leopoldo. *Anais...* São Leopoldo: SEMISH, 2005. p. 2005-2017. Disponível em: http://200.239.128.16/bitstream/123456789/1680/1/EVENTO_NovoRetratoWeb.pdf. Acesso em: 20 dez. 2018.

PALMER, J.; STEVENSON, J. Something worth sitting still for? Some implications of Web 2.0 for outreach. In: THEIMER, K. (Ed.). *A Different Kind of Web: New Connections between Archives and our Users*. Chicago: Society of American Archivists, 2011. p. 1-19.

PRESNER, T. Digital Humanities 2.0: A Report on Knowledge. In: BAILLAR, M. (Ed.). *Emerging Disciplines*. Houston: Rice University Press, 2010. p. 63-86. Disponível em: <http://cnx.org/content/m34246/latest/>. Acesso em: 15 dez. 2018.

RIDGE, M. RIDGE, Mia. Crowdsourcing our Cultural Heritage: an Introduction. In: RIDGE, Mia (Ed). *Crowdsourcing our Cultural Heritage*. London: Ashgate, 2014, p. 1-13.

VALENTE, M. G. Introdução: notas gerais sobre a digitalização de acervos no Brasil. In: FREITAS, B. C.; VALENTE, M. G. (Org.). *Memórias digitais: o estado da digitalização de acervos no Brasil*. Rio de Janeiro: FGV Editora, 2017. p. 7-55.

em SI. Inspirado na palestra de Dan Tapscott – *Four principles for the open world*³, um painel apresentou a visão de pesquisadores e profissionais de empresas sobre os princípios para viver e sobreviver no/ao mundo aberto – transparência, colaboração, compartilhamento e deslocamento de poder – tanto por empresas como por todo e qualquer cidadão comum.

Naquele painel, pesquisadores da comunidade de SI apresentaram artigos versando sobre os desafios da Lei do Acesso à Informação, sobretudo a disponibilização de dados em formato aberto, os avanços na análise de informações em redes sociais – como as sensações expressas por seus participantes – e como planejar e medir a maturidade das cidades na oferta de serviços cada vez mais inteligentes aos seus cidadãos. Refletimos sobre a privacidade, ou antes, nossas reações à exposição da informação, sobretudo pessoais, em ritmo e escala pouco controláveis. Nos defrontamos com a complexidade que nos oferece a construção de sistemas cada vez menos caracterizados como um artefato fechado e intraorganizacional e cada vez mais como um organismo de conexão intra, inter e sócio-organizacional, de comportamento imprevisível e emergente. Tomamos conhecimento dos esforços, progressos e desafios de empresas governamentais em sua estruturação interna para oferta de serviços e informação cada vez mais modernos e sintonizados às necessidades do cidadão brasileiro. O investimento é enorme, e o potencial e os desafios da ‘abertura’ no Brasil são tão grandes quanto seu tamanho, sua riqueza e suas mazelas, sobretudo suas competências e sua legislação.

Cada vez mais percebemos que estamos vivendo em um mundo rapidamente mais conectado e aberto, que abre novas oportunidades tanto para as organizações inovarem em seu negócio, como para indivíduos adquirirem mais poder, autonomia e satisfação no uso de serviços. Em uma época em que vivemos o forte poder disruptivo da tecnologia, potencializado pela convergência da colaboração, mobilidade e de grandes volumes de dados, um dos grandes desafios da comunidade de pesquisa em Sistemas de Informação para os próximos anos (BOSCARIOLI; ARAUJO; MACIEL, 2017) está em como promover a integração destas dimensões tecnológicas de forma a equilibrar tanto as necessidades de controle como as oportunidades de comportamento emergente e inovação.

Esse mundo em que a transparência e a prestação de contas se torna cada vez mais cobrada, e o entendimento da diversidade é considerado como prática social,

³ Disponível em: <https://www.ted.com/talks/don_tapscott_four_principles_for_the_open_world_1> . Acesso em: 09 set. 2020.

não somente um desejo, e a inovação e a multidisciplinaridade são peças-chave para a solução de problemas complexos que afetam a humanidade, precisa estar aberto, conectado, acessível, e seus principais atores (indivíduos e organizações) devem ser capazes de se organizar sem a necessidade específica de uma estrutura, controle ou ordem prevista (ARAUJO, 2017; SHIRKY, 1998), de forma democrática pelo uso da tecnologia disponível. Isto se tornará ainda mais complexo nas próximas décadas, uma vez que as organizações e a sociedade precisarão estar preparadas para uma abertura ainda maior do que tenham em algum momento imaginado, para resolver novos problemas a serem enfrentados no presente e no futuro.

E que problemas seriam esses? Muito além das questões enfrentadas por empresas, instituições ou nações específicas, estão os desafios a serem enfrentados (juntos) por todos os seres humanos neste planeta. A Organização das Nações Unidas (ONU) identificou alguns, denominando-os como os Objetivos de Desenvolvimento Sustentável (ONU, 2015), motivando as nações associadas a abordar soluções para estes problemas e acompanhar seus impactos até o ano de 2030. Estes objetivos incluem soluções para diversos problemas (Figura 1) que hoje afligem toda a humanidade, de solução complexa, multidisciplinar e cuja efetividade exigirá esforços conjuntos entre indivíduos, instituições, pesquisadores, governos e nações para sua solução. Neste mundo conectado e aberto, não parece mais haver espaço para soluções individuais. Os problemas são de todos.

FIGURA 1 - OBJETIVOS DE DESENVOLVIMENTO SUSTENTÁVEL



Fonte: ONU (2015).

No cenário brasileiro, alguns destes objetivos saltam ainda mais aos olhos: a necessidade do enfrentamento da corrupção (Paz, Justiça e Instituições Fortes), os desafios de enfrentar todos estes desafios em um país de dimensões continentais, a diferença econômica e intelectual abismal entre extratos de nossa sociedade, a insegurança urbana, a flagelação dos direitos humanos, os ônus e bônus de uma economia agrária, a sede de uma floresta que pode ser a salvação do mundo, entre vários outros.

Em absolutamente todas as dimensões apontadas como os principais problemas de nossa humanidade, a tecnologia, sobretudo as chamadas Tecnologias da Informação e Comunicação (TICs), aparecem como ator fundamental, como vetor de diálogo, colaboração e inovação para a criação, aplicação e avaliação de impacto de soluções. No entanto, apesar da tecnologia ser um ator fundamental neste cenário, no cerne destes problemas reside a oportunidade de exercitarmos os aspectos mais fundamentais de sermos humanos: o raciocínio consciente, a capacidade de diálogo, a resistência ao mundo pela colaboração, a inovação e a não conformidade. A conjugação destas capacidades com a flexibilidade tecnológica é o que gosto de entender como Humanidades Digitais.

CONHEÇA O MEU LUGAR DE FALA: SISTEMAS DE INFORMAÇÃO

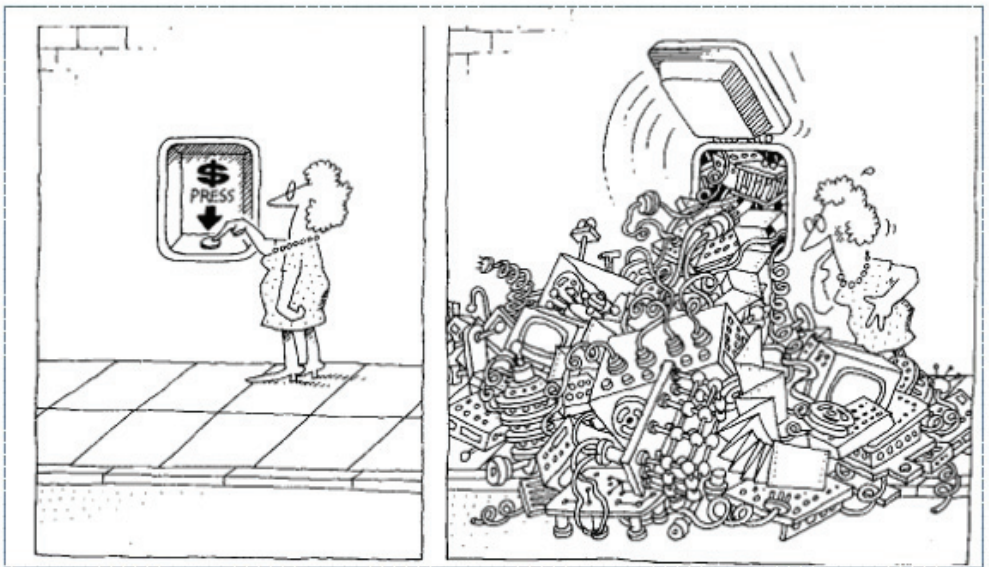
Compreender o impacto e as implicações do uso de tecnologia nas atividades humanas atualmente requer muitos olhares. Cada um deles sozinho só será capaz de compreendê-lo de forma muito limitada. O esforço de um olhar múltiplo e sistêmico é o que se entende principalmente ser um dos objetivos da área das Humanidades Digitais.

Há um lugar de observação neste cenário muito pouco compreendido, mas de fundamental importância para o estudo das Humanidades Digitais – a Computação, mais especificamente, a área que detém o conhecimento teórico e técnico sobre como construir a tecnologia que vemos e usamos hoje. Via de regra, as demais áreas que ocorrem ao estudo das Humanidades Digitais tendem a ver a tecnologia como uma “caixa-preta”, uma entidade a mais no ecossistema de relações humanas, pré-existente e pronta para ser discutida e criticada. Em muitas ocasiões escuto debates sobre “a Tecnologia” (sim, com “T” maiúscula!) como se fosse um ente consciente e

intencionado a provocar um determinado efeito em uma determinada coletividade. Em outras situações, e cada vez mais decorrente do volume de dados disponíveis sobre tudo, a tecnologia tem sido vista como ferramental de suporte para pesquisas em todas as áreas e, não raro, os pesquisadores da área de Computação são desejados como profissionais prestadores de serviço técnico e nem sempre reconhecidos como parceiros científicos.

A Computação se desdobra em muitas subáreas técnicas, mas também científicas, uma delas em particular, denominada Engenharia de Software, tem como objetivo principal, estudar e inovar em teorias, métodos e técnicas sobre como construir software, ou seja, como construir o elemento principal de TICs – seu funcionamento lógico, capaz de ser interpretado por computadores. O funcionamento lógico de um artefato tecnológico e sua execução por uma máquina trata-se de um universo bastante complexo que se tenta simplificar ao máximo para quem o utiliza. Como diria Booch (1994), a tarefa de uma equipe de desenvolvimento de software é criar a ilusão da simplicidade para quem o utiliza (Figura 2).

FIGURA 2 - A ILUSÃO DA SIMPLICIDADE DE UM SOFTWARE

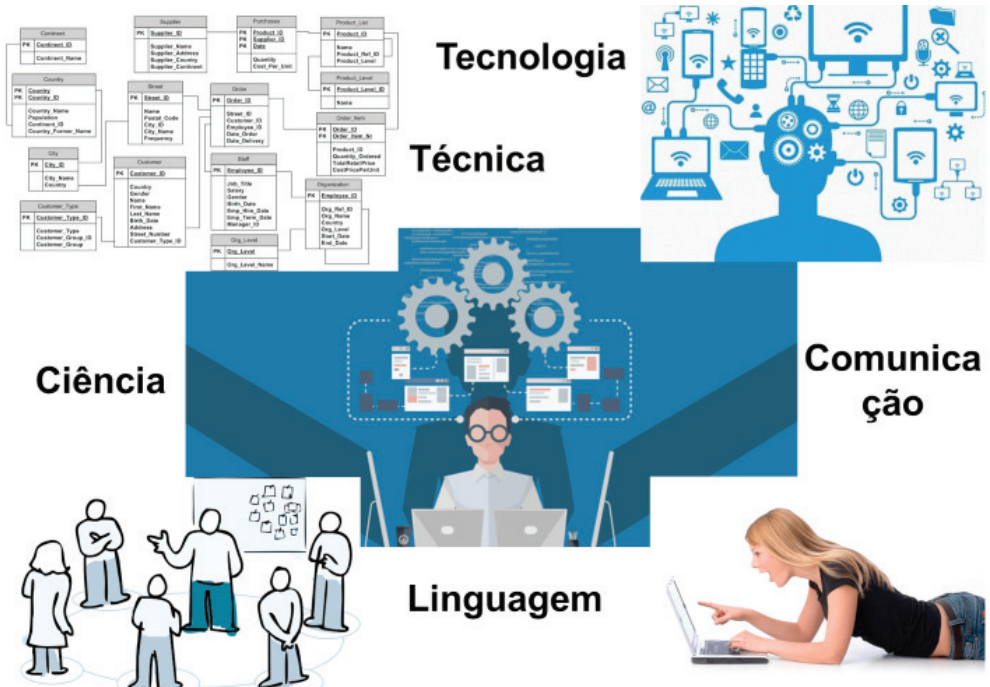


Fonte: Booch (1994).

A construção de um software exige o esforço de muitos profissionais e a aplicação de métodos e técnicas específicas desenvolvidos e aprimorados ao custo de muito debate em ciência nos últimos 50 anos. Além disso, é preciso muita tecnologia para construir tecnologia: os profissionais de desenvolvimento de software fazem uso de ferramentas tecnológicas – computadores, editores de código, ferramentas de testes, editores de interface, entre muitas outras. O profissional que constrói software, de forma muito resumida, identifica as necessidades de alguém, as abstrai em modelos técnicos e as implementa em um artefato tecnológico na tentativa de atender tais necessidades. Portanto, no limite, construir um software é um ato de linguagem, de comunicação de desejos, necessidades e possibilidades entre seres humanos. Quando interagimos com um sistema em nossos computadores, nossos celulares, nosso dia a dia, em qualquer situação, não estamos dialogando com uma máquina unicamente. Estamos dialogando também com quem a construiu (Figura 3).

Outras áreas da Computação também são desafiadas cientificamente para simplificar continuamente o uso de tecnologia pela sociedade. Os pesquisadores de banco de dados criam modelos e estruturas computacionais avançados para enfrentar o desafio do uso de dados e informação de forma eficiente e eficaz em um mundo aberto e conectado. A Inteligência Artificial promove o uso de técnicas que permitem o processamento de informação para o raciocínio, a dedução e a predição de fatos a fim de auxiliar e facilitar o mundo moderno. A área de Interface Humano-Computador é desafiada continuamente para permitir que a sociedade use de forma simples, dedutiva e rápida os dedos, a fala, a visão, para interagir com a tecnologia criada. A comunicação de dados avança também velozmente pela contribuição da área de Redes de Computadores.

FIGURA 3 - A TECNOLOGIA COMO LINGUAGEM



Fonte: Elaborado pela autora (2020).

Uma área dentro da Computação, em particular, é colocada no centro de minha contribuição: Sistemas de Informação. Esta área científica tem como objetivo principal compreender a Computação como meio ou instrumento para a solução de problemas do mundo real, da sociedade e das organizações. A área de Sistemas de Informação contribui de forma importante em diversos domínios, lida com sistemas complexos que requerem conhecimentos técnicos e organizacionais para serem projetados, desenvolvidos e gerenciados, que afetam as operações e as estratégias das organizações e as atividades em geral da sociedade.

Um Sistema de Informação pode ser definido como um conjunto de componentes inter-relacionados que trabalham de forma convergente para coletar (ou recuperar), processar, armazenar e distribuir informação para suporte à tomada de decisão, coordenação, controle, análise de problemas, visualização de situações complexas e criação de novos produtos em uma organização (GALLIERS;

CURRIE, 2011; LAUDON; LAUDON, 2016). Uma organização aqui pode ser entendida tanto por: i) seu aspecto técnico – uma estrutura social formal e estável durante um período de tempo, que utiliza e processa recursos do ambiente para a produção de novos produtos; como por ii) seu aspecto comportamental – uma coleção de direitos, privilégios, obrigações e responsabilidades balanceadas por meio de resolução de conflitos. Organização é um conceito mais abrangente do que uma empresa ou uma instituição, e pode compreender a relação entre indivíduos, empresas ou ambos.

Sistemas de Informação apoiam e automatizam processos nas organizações, construindo vantagem competitiva por possibilitar análise de cenários, apoio aos processos de decisão, além da definição e implementação de novas estratégias organizacionais. Por conseguinte, torna-se uma constante a preocupação com a coleta, armazenamento, processamento e transmissão da informação na medida em que a disponibilidade da informação certa, no momento certo, para o tomador de decisão responsável, é requisito fundamental para a melhoria contínua da qualidade e competitividade organizacionais, bem como para a melhoria da qualidade de vida humana.

Para além de sua importância nos contextos empresariais, Sistemas de Informação estão atualmente imbricados no funcionamento da sociedade. O acesso à informação e a aplicações em qualquer hora, qualquer lugar, amplia e reestrutura as relações e trabalho e sociais, sobretudo as formas de comunicação entre seus agentes, abrindo espaço para configurações sociais e modelos de negócio antes impensados. Neste sentido, Sistemas de Informação representam agentes fundamentais da transformação social contemporânea, requerendo entendimento dos contextos sociais e suas relações onde serão utilizados e da influência que as características do ambiente, bem como as características individuais e de comportamento dos agentes sociais humanos exercem no uso e no projeto destes sistemas.

Boa parte dos entendimentos teóricos na área advêm de teorias como a Teoria Geral de Sistemas (BERTALANFFY, 2008), que tenta explicar a dinâmica dos sistemas biológicos, sociais, organizacionais e tecnológicos. Ela se consolida no saber técnico da Engenharia (SOMMERVILLE, 2011), na busca por métodos mais eficazes para a construção de novas tecnologias. Mas é também desafiada pela necessidade de compreender seu impacto no mundo, onde torna-se indispensável

o pensamento sistêmico (CAPRA; LUISI, 2014) e saber abordar a complexidade (MORIN, 2015).

Meu lugar de fala, e portanto, minha contribuição para o debate das Humanidades Digitais, vem do ponto de vista de quem maneja a complexidade dos Sistemas de Informação, de quem procura dialogar com os ambientes e pessoas em seu contexto por meio da linguagem tecnológica, de quem pretende definitivamente interferir em suas práticas na expectativa de aprimorá-las, sem desconhecer a imprevisibilidade de seus efeitos.

CIBERDEMOCRACIA

Cibernética (WIENER, 1948) é o estudo interdisciplinar da estrutura de sistemas regulatórios. A Cibernética está fortemente relacionada à Teoria do Controle e à Teoria Geral dos Sistemas (BERTALANFFY, 2008). Tanto em sua origem como em seu desenvolvimento, as teorias da Cibernética podem ser aplicadas tanto para sistemas físicos como sistemas sociais. Sistemas complexos afetam seu ambiente externo e se adaptam a ele. Em termos técnicos, sistemas possuem foco em funções de controle e comunicação nos fenômenos de entrada/saída do sistema. Esta habilidade é natural em organismos vivos e tem sido imitada em máquinas e organizações. A Cibernética é a ciência do controle, a ciência da governança.

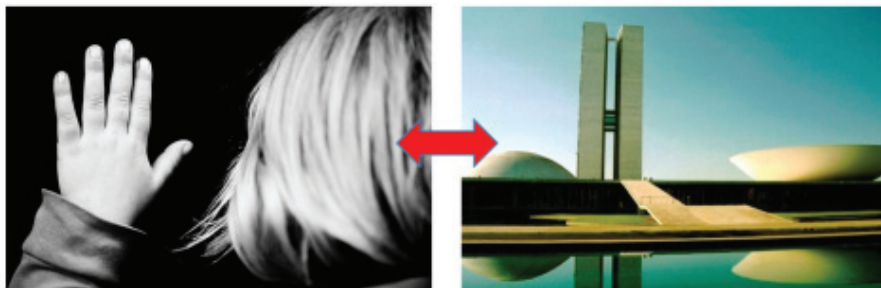
O conceito de Democracia (“demo+kratos”) é usualmente associado a um modelo de governo em que o poder de tomar decisões políticas importantes vem diretamente dos cidadãos, ou em sua forma mais usual, por representantes eleitos. A história da Democracia se refere a uma série de processos históricos e é um conceito de difícil definição, baseado na noção de uma comunidade em que todas as pessoas têm o direito de igualmente participar, debater e decidir a respeito de processos políticos e, em um sentido mais moderno, que certos direitos sejam universalizados como o princípio da liberdade de expressão e a dignidade humana. O conceito de Democracia, apesar de fortemente ligado à legislação e constitucionalismo, não está limitado à questão da igualdade legal, mas também depende do acesso democrático (ou seja, o mesmo para todos) de espaços e benefícios sociais.

Com base nestes dois conceitos, passo a usar o termo Ciberdemocracia – a suposição de que em um mundo aberto possibilitado pelas TICs, a transmissão

de informação, conexão e reconfiguração realizada pelos indivíduos proporciona colaboração, pluralidade, abertura, empoderamento e governança. A ideia de que quanto mais produzimos, entregamos, distribuímos e compartilhamos informação, mais inteligentes (no sentido de governança) e conscientes a sociedade pode se tornar. A Ciberdemocracia é, para mim, um termo que significa a inteligência coletiva que surge por meio das possibilidades de opinião pública e empoderamento com o uso da tecnologia, levando a melhores níveis de governança de sistemas sociais e organizacionais. De acordo com Lemos e Levy (2010), o relacionamento entre comunicação (poder social) e técnica (poder de ação) está na base para esta nova dimensão política, proporcionando a cada indivíduo conectado pela tecnologia um novo relacionamento com o espaço e o tempo, uma nova dimensão de viver coletivamente.

Também gosto de entender a Ciberdemocracia como a possibilidade de interação política e de gestão pública por meio da tecnologia, potencializada pela Internet. A Ciberdemocracia é um termo que reúne os grandes aspectos da sociedade atual – conectada, convergente, informada e colaborativa e os grandes desafios das instituições governamentais em acompanhar esta sociedade. A Ciberdemocracia aponta para os desafios destes novos ecossistemas que envolvem cidadãos, instituições públicas, tecnologia, informação, práticas, políticas e processos (Figura 4).

FIGURA 4 - CIBERDEMOCRACIA



CiberDemocracia é um tempo que reúne grandes aspectos da sociedade atual – conectada, convergente, informada e colaborativa.

CiberDemocracia é também o tempo que reúne os grandes desafios das instituições governamentais em acompanhar esta sociedade.

A CiberDemocracia aponta para os desafios destes novos **ecossistemas** que envolvem cidadãos, **instituições** públicas, **tecnologia**, **informação**, **práticas**, **políticas** e **processos**.

Fonte: Elaborado pela autora (2020).

SISTEMAS DE INFORMAÇÃO DE GOVERNO ABERTOS E COLABORATIVOS

No sentido de ampliar a rede de interação entre Governo, Administração Pública e cidadãos, e permitir novas formas de governança, muitas soluções têm sido exploradas tanto na academia como nos governos ao redor do mundo, para criar sistemas de informação que permitam diferentes formas de interação cidadão-governo e cidadão-cidadão no que se refere à prestação de serviços, coleta de informações, transparência e tomada de decisão colaborativa (ARAUJO *et al.*, 2011).

Estas soluções eu chamo de Sistemas de Informação de Governo Abertos e Colaborativos – sistemas de informação que apoiam o acesso à informação, participação, gestão de conhecimento e inovação social nas relações governo-cidadão. De forma muito geral, gosto de organizar estas soluções nas seguintes classes: i) soluções de Governo Eletrônico: sistemas de informação para oferta de serviços públicos em plataforma digital (ex. Portal de Serviços do Governo Brasileiro)⁴; ii) soluções de Participação Eletrônica: sistemas de informação para permitir a participação dos cidadãos nas atividades relacionadas à prestação de serviços públicos (ex. aplicativos municipais para comunicação pelos cidadãos de problemas na cidade de diversas naturezas à Prefeitura); iii) soluções de Democracia Digital: sistemas de informação que apoiam o debate e a deliberação sobre políticas públicas (ex. Participa.br⁵); iv) soluções de Governança Digital: sistemas de informação que permitem ao cidadão o poder de fiscalizar, acompanhar, governar as instituições públicas (ex. Operação Serenata de Amor⁶); e v) soluções de Ciberativismo: sistemas de informação que permitem aos cidadãos auto-organização e articulação entre si ações, movimentos e manifestações (ex. MeuRio⁷)

⁴ Disponível em: <<https://www.servicos.gov.br/>>. Acesso em: 09 set. 2020.

⁵ Disponível em: <<http://participa.br/>>. Acesso em: 09 set. 2020.

⁶ Disponível em: <<https://serenata.ai/>>. Acesso em: 09 set. 2020.

⁷ Disponível em: <<https://www.meurio.org.br/>>. Acesso em: 09 set. 2020.

FIGURA 5 - SISTEMAS DE INFORMAÇÃO DE GOVERNO ABERTOS E COLABORATIVOS



Fonte: Elaborado pela autora (2020).

Há quatro anos, eu assistia, admirada, nas salas de cinema, os contos tão inverossímeis, se não fossem tão reais, do filme *Relatos Selvagens*, de Damián Szifron. Um desses contos narra a história de um cidadão comum (Roberto Darín) que recebe uma multa por estacionar em lugar supostamente indevido depois de buscar por uma eternidade por uma vaga, pressionado pela necessidade de buscar uma encomenda da família em algum lugar da cidade. O fato lhe irrita tão intensamente – pois a multa não parecia justa – que ele resolve contestá-la junto à prefeitura. A saga que se segue é tão horripilante porque a gente se sente totalmente dentro dela. A burocracia, a ausência total de diálogo entre cidadão e coisa pública, mesmo que diante de outro ser humano (um atendente), as regras confusas, a total incapacidade de ser ouvido. O fim eu não conto.

Em outro filme, *Eu, Daniel Blake*, eu me sentia novamente chocada com o drama social vivido por um britânico desempregado, incapaz de conseguir seu

seguro-desemprego por baixo letramento digital – a incapacidade de se apropriar da tecnologia – sendo o meio digital o único possível para interação com o governo de seu país. O filme não comenta explicitamente isso, há outros temas mais relevantes para serem explorados na tela, mas o Reino Unido é um dos países mais avançados em transformação digital no governo⁸ e que investe pesadamente no conceito de *digital by default*, ou seja, todos os serviços novos prestados pelo governo já nascem no digital.

FIGURA 6 - SE NÃO VIU, ASSISTA. AJUDARÁ A ENTENDER MELHOR O PROBLEMA O QUAL ESTOU TENTANDO CONVENCER VOCÊ QUE EXISTE



Fonte: Elaborado pela autora (2020).

Por um certo tempo, os desenvolvedores de sistemas de informação voltados à interação entre cidadão e administração pública se comprometeram com o desafio de transformar serviços públicos em serviços digitais. Neste processo, um elemento fundamental foi negligenciado: o cidadão, suas necessidades e suas competências para lidar com o digital. Não é muito difícil ver problemas similares ao de Daniel Blake: cidadãos não letrados digitalmente precisando contratar os serviços de um

⁸ Visite: <Gov.uk>.

“despachante digital” capaz de se comunicar facilmente com o universo dos serviços eletrônicos. Embora a tramitação dos serviços, possa, em alguma escala, ter sido facilitada, ainda persistem os problemas de desinformação, de inconsistências, de diálogos surdos com ambientes dependentes da automação e, no caso do Brasil, o desafio da escalabilidade em um país de dimensões continentais e população de enorme diversidade.

DESIGN DE SOLUÇÕES

No contexto da prática da administração pública, são muitas as iniciativas relacionadas à melhoria e à inovação em matéria de sistemas de informação para governança digital e interação com o cidadão. As palavras de ordem neste contexto atualmente são: design, empatia e inovação. Vivemos a era do design (PINHEIRO; ALT; PONTES, 2011), sendo assim, as soluções não estão prontas e por isso precisam ser reinventadas a cada contexto. Assim também é para soluções de interação com o cidadão.

Se em algum momento no passado confiamos desproporcionalmente no poder da tecnologia para solucionar nossos problemas, sem contestar os obstáculos e implicações que ela proporciona em seu uso, hoje vivemos a era da empatia, na qual a satisfação do cliente (ou cidadão), suas personas, suas experiências individuais precisam ser cuidadosamente consideradas hoje, sob a força da concorrência do mercado (GARRETT, 2002).

No mundo aberto, negócios e organizações só se manterão competitivos se aprenderem como gerenciar seus processos neste novo cenário, aberto e conectado. Ao mesmo tempo, o ambiente organizacional interno precisará desenvolver a interação e a colaboração entre seus profissionais e conexões com o ambiente externo, de forma a assegurar o desempenho de seus processos de negócio/trabalho com menos burocracia, mais autonomia e qualidade. Como pesquisadores na área de Sistemas de Informação, precisamos dedicar nosso tempo a investigar como aumentar e fortalecer a ligação entre organizações e seus ambientes externos, entregando melhores serviços e estabelecendo um diálogo efetivo entre eles pelo uso da tecnologia. Devemos nos questionar sobre: i) como fazer com que indivíduos, sociedade e instituições cooperem e busquem maneiras de efetivamente compartilhar e convergir em objetivos comuns; ii) como integrar instituições e indivíduos como colaboradores na tarefa de gerenciar e aprimorar constantemente

seus processos; iii) como construir soluções que pode ajudar os indivíduos a obter acesso à forma como as organizações trabalham e se comportam; e iv) como apoiar os indivíduos e organizações para produzirem colaborativamente e permanecerem conectados neste espaço virtual, deixando de lado velhos relacionamentos – usualmente opostos.

Em nosso grupo de pesquisas, temos apostado na possibilidade de desenvolver novos sistemas de informação ainda mais capazes de promover a colaboração e a abertura, com foco no design e no cidadão. Nossa abordagem tem sido a de promover o entendimento, o diálogo, o aprendizado e a interação (ARAUJO; TAHER, 2014; ARAUJO; CAPPELLI, 2014; CLASSE; ARAUJO; XEXEO, 2017, 2018; DIIRR; ARAUJO; CAPPELLI, 2014; GOMES; ARAUJO, 2012; IGLESIAS; CAPPELLI; ARAUJO, 2014; ENGIEL; ARAUJO; CAPPELLI, 2014; PIRES; ARAUJO, 2017; SELL, 2016; TAVARES; PIMENTEL; ARAUJO, 2014). É um grande desafio que não podemos abordar sem a ajuda de outras disciplinas, principalmente as que dominam o conhecimento sobre as práticas políticas, as que dominam o conhecimento sobre as práticas sociais e todas as outras que possam nos ajudar a compreender este fenômeno dos novos ecossistemas onde coexistem pessoas e tecnologia.

CONCLUSÃO

Democracia é um esforço perseguido por nossa sociedade, não exatamente em busca de igualdade absoluta, o que a Natureza já demonstrou ser impossível, mas para encontrar equilíbrio e bem-estar. A democracia é não só uma questão política ou de administração pública, mas também um desafio para a gestão de organizações privadas e grupos sociais. O mundo aberto e as tecnologias associadas a esta abertura, podem nos indicar oportunidades de governança mútua e equilíbrio por meio da Ciberdemocracia.

Os sistemas de informação não podem ser vistos como artefatos fechados no estudo das Humanidades Digitais, pois seu design afeta diretamente seus efeitos. Uma visão sistêmica ampla precisa ser exercitada para o design, desenvolvimento, uso e avaliação de sistemas de informação no mundo aberto, facilitando processos, empoderando pessoas, gerando confiança e estabelecendo uma nova era de auto-organização e governança.

AGRADECIMENTOS

As ideias contidas neste capítulo são provenientes de diversas discussões e projetos desenvolvidos no âmbito do Grupo de Pesquisa e Inovação em Ciberdemocracia (CIBERDEM) nos últimos 10 anos. Agradeço a todos os pesquisadores acolhidos por este grupo neste período por terem me ajudado a refletir sobre os desafios que nossa comunidade de pesquisa necessita de enfrentar no futuro. Agradeço também ao CNPq pela concessão da Bolsa de Produtividade em Desenvolvimento Tecnológico e Extensão Inovadora.

REFERÊNCIAS

ARAUJO, R.M. Information Systems and the Open World Challenges. In: BOSCARIOLI, C.; ARAUJO, R. M.; MACIEL, R. S. P. *I GranDSI-BR: Grand Research Challenges in Information Systems in Brazil 2016-2026*. 1. ed. Porto Alegre: Special Committee on Information Systems (CE-SI); Brazilian Computer Society (SBC), 2017. p. 42-45.

ARAUJO, R. M.; CAPPELLI, C.; DIIRR, B.; ENGIEL, P.; TAVARES, R. L. Democracia Eletrônica. In: PIMENTEL, M.; FUKS, H. *Sistemas colaborativos*. 1. ed. Rio de Janeiro: Campus/SBC, 2011. p. 110-121.

ARAUJO, R. M.; TAHER, Y. Refining IT Requirements for Government-Citizen Co-participation Support in Public Service Design and Delivery. In: CONFERENCE FOR E-DEMOCRACY AND OPEN GOVERNMENT, 2014, Krems. *Anais...* Krems: Donau-Universität Krems, 2014. p. 61-72.

BERTALANFFY, L. *Teoria geral de sistemas*. Petrópolis: Editora Vozes, 2008.

BOOCH, Grady. *Object-Oriented Analysis and Design with Applications*. 2. ed. California: Addison-Wesley, 1994. 543 p.

BOSCARIOLI, C.; ARAÚJO, R. M.; MACIEL, R. S. P. *I GranDSI-BR: Grand Research Challenges in Information Systems in Brazil 2016-2026*. 1. ed. Porto Alegre: Special Committee on Information Systems (CE-SI); Brazilian Computer Society (SBC), 2017. 184 p.

CAPRA, F.; LUISI, P.L. *A visão sistêmica da vida: uma concepção unificada e suas implicações filosóficas, políticas, sociais e econômicas*. 1. ed. São Paulo: Cultrix, 2014. 616 p.

CLASSE, T.; ARAUJO, R. M.; XEXEO, G. De processos de negócio para jogos digitais: uma proposta de mapeamento. In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO, 14., 2018, Caxias do Sul. *Anais...* Porto Alegre: Sociedade Brasileira de Computação, 2018.

CLASSE, T. M.; ARAUJO, R. M.; XEXEO, G. Desaparecidos RJ: um jogo digital para o entendimento de processos de prestação de serviços públicos. In: SIMPÓSIO BRASILEIRO DE JOGOS E ENTRETENIMENTO DIGITAL, 16., 2017, Curitiba. *Anais...* Porto Alegre: Sociedade Brasileira de Computação, 2017.

DIIRR, B.; ARAUJO, R. M.; CAPPELLI, C. Encouraging Society Participation Through Conversations About Public Service Processes. *International Journal of Electronic Government Research*, v. 10, n. 2, p. 22-42, 2014.

ENGIEL, P.; ARAUJO, R. M.; CAPPELLI, C. Designing Public Service Process Models for Understandability. *Electronic Journal of e-Government*, v. 12, n. 1, p. 95-111, 2014.

GALLIERS, R.D.; CURRIE, W.L. *The Oxford Handbook of Management Information Systems: Critical Perspectives and New Directions*. 1. ed. Oxford University Press, 2011. 725 p.

GARRETT, J. J. *The Elements of User Experience: User-Centered Design for the Web*. 1. ed. São Francisco: New Riders, 2002. 208 p.

GOMES, J. O.; ARAUJO, R. M. Promovendo a Compreensão de Regras em Processos de Prestação de serviços Públicos Utilizando a Animação. In: WORKSHOP DE COMPUTAÇÃO APLICADA EM GOVERNO ELETRÔNICO, 4., 2012, São Paulo. *Anais...* São Paulo: Sociedade Brasileira de Computação 2012. p. 25-32.

IGLESIAS, C.; CAPPELLI, C.; ARAUJO, R. M. Formalização do catálogo de características de entendimento de modelos de processos de prestação de serviços públicos inseridos na carta de serviços. In: WORKSHOP DE TRANSPARÊNCIA EM SISTEMAS, 2., 2014, Londrina. *Anais...* Porto Alegre: Sociedade Brasileira de Computação, 2014.

LAUDON, K. C.; LAUDON, J. P. *Management Information Systems: Managing the Digital Firm*. 14. ed. New York: Pearson, 2016. 672 p.

LEMONS, A.; LÉVY, P. *O futuro da Internet: em direção a uma ciberdemocracia planetária*. 1. ed. São Paulo: Paulus, 2010. 264 p.

MORIN, E. *Introdução ao pensamento complexo*. 4. ed. Porto Alegre: Sulina, 2015. 120 p.

ONU. Organização das Nações Unidas. *Objetivos de Desenvolvimento Sustentável*. 2015. Disponível em: <<https://nacoesunidas.org/conheca-os-novos-17-objetivos-dedesenvolvimento-sustentavel-da-onu/>>. Acesso em: 20 dez. 2018.

PINHEIRO, T.; ALT, L.; PONTES, F. *Design Thinking Brasil: empatia, colaboração e experimentação para pessoas, negócios e sociedade*. São Paulo: Elsevier, 2011. 248 p.

PIRES, E.; ARAUJO, R. M. Mediações de conflitos no Poder Judiciário (MARC's): alternativas tecnológicas para aproximação cidadã. In: WORKSHOP DE TESES E DISSERTAÇÕES EM SISTEMAS DE INFORMAÇÃO, 2017, Lavras. *Anais...* Porto Alegre: Sociedade Brasileira de Computação, 2017. p. 68.

SELL, M. M. *Regra clara: uma proposta de design de artefato para compreensão de regras de negócios em processos de prestação de serviços públicos*. 2016. 97 f. Dissertação (Mestrado em Informática) - Programa de Pós-Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

SHIRKY, C. *Here Comes Everybody: The Power of Organizing Without Organization*. 1. ed. New York: The Penguin Press, 1998. 344 p.

SOMMERVILLE, I. *Engenharia de software*. 9. ed. Pearson, 2012. 529 p.

TAVARES, R. L.; PIMENTEL, M.; ARAUJO, R. M. Visualizing Clashes and Alliances in Social Networks of Political Discussions. *Social Networking*, v. 3, p. 94-101, 2014.

WIENER, N. *Cybernetics: Or Control and Communication in the Animal and the Machine*. 2. ed. Cambridge, Massachusetts: MIT Press, 1965. 212 p.

Uso de Mineração de Dados no contexto eleitoral brasileiro: um caso de pesquisa à luz das Humanidades Digitais

Josir C. Gomes
Marco Schneider
Ricardo M. Pimenta

Em tempos de contínuos avanços tecnológicos e aceleração social (ROSA, 2014) das formas de ser e de agir no espaço público, a pesquisa crítica da circulação da informação tem buscado novos métodos e ferramentas para conseguir processar e analisar o volume de dados cada vez maior que vem sendo gerado não só pelas redes sociais e pelos meios de comunicação tradicionais, mas também pelos próprios entes governamentais. Tais esforços encontram hoje nas chamadas Humanidades Digitais (HDs) uma “comunidade de práticas” (ALVES, 2016) que visa experimentar e comunicar técnicas, teorias e métodos mediados pela tecnologia digital aplicados às questões e problemáticas oriundas das ciências humanas, sociais, sociais aplicadas, letras e artes. Desse modo, a definição das HDs está subordinada às práticas e às ações de seus agentes em torno de um projeto comum e, não a este ou aquele “campo” ou “área” de predominância dos pesquisadores envolvidos, torna-se claro que ela é aberta, plástica e relacional. Constitui-se de maneira multidisciplinar, ou transdisciplinar, em seu escopo e interdisciplinar por meio de suas práticas, portanto, de fronteira” (KLEIN, 2015, p. 31-32).

Para Berry (2011, p. 1), “as humanidades digitais tentam levar em consideração a plasticidade das formas digitais e a maneira pela qual elas apontam para uma nova maneira de trabalhar com representação e mediação”. Mediar o acesso à informação, principalmente no espaço público, requer que se esteja preparado para lidar com políticas de opacidade, em face dos dados disponibilizados, porém não tratados, pelos mesmos entes governamentais. Dessa forma, podemos estender o questionamento de Manovich (2008, p. 41) sobre o que é a cultura após ser “*softwarized*” à política e à economia. Como compreender e questionar a política e a economia, em suas diferentes instâncias, sem lançar mão de recursos computacionais capazes de

processar e mediar dados e informações disponíveis, apesar de desconexos e “sujos” em seu estado de *Raw Data*? Afinal, a política e a economia não foram igualmente “*softwarized*”?

Este capítulo trata fundamentalmente dessa questão, ao apontar possibilidades de uso e aplicação dos recursos das HDs no desvelamento de alguns elementos capitais no contexto eleitoral brasileiro, com destaque para relação entre financiamento de campanhas políticas, atuações parlamentares e reverberações (ou silenciamentos) midiáticos e informacionais.

Especificamente no Brasil, na área das Ciências Políticas e da Economia Política da Informação, houve um crescimento considerável no volume de informações disponíveis em função da promulgação da Lei de Acesso à Informação (LAI) em 2012 (BRASIL, 2012). Com a implementação da lei, o Estado brasileiro passou a disponibilizar dados gerados e administrados pelo governo e também pelos poderes legislativos e judiciário. Apesar do número do conjunto de dados (*datasets*) disponibilizados ainda estar longe de ser o ideal, estas novas fontes de informação permitem que pesquisadores, ativistas políticos e os próprios órgãos governamentais de controle, como a Controladoria Geral da União (CGU) e o Ministério Público Federal (MPF), tenham à disposição um volume de informações que antes estavam muradas dentro de cada tribunal, casa legislativa ou órgão público.

Na realidade brasileira, três conjuntos de dados se destacam nas pesquisas quantitativas na área de ciências políticas: a base de resultados eleitorais, a base de financiamento de campanhas políticas e as proposições parlamentares que tramitam no Congresso Nacional. Todas as transações de financiamento de campanha política e os resultados das eleições são disponibilizados em formato aberto pelo Tribunal Superior Eleitoral (TSE), já que este tribunal exerce o controle e a fiscalização da arrecadação e também dos gastos de candidatos e partidos políticos. Já a base de dados de proposições legislativas, que contém todas as propostas de emenda à constituição, projetos de lei e indicações que são encaminhadas na Câmara dos Deputados e no Senado Federal, é disponibilizada por ambas as casas legislativas.

A análise destes conjuntos de dados tem sido utilizada com muita frequência através de métodos tradicionais nas Ciências Sociais, como a estatística descritiva, e os resultados têm sido expressivos (MANCUSO, 2015; PEIXOTO, 2016). Diversos estudos indicam que quanto maior o valor arrecadado nas campanhas, maior as chances de sucesso nas eleições (MANCUSO; SPECK, 2015). A relação

entre o poder econômico e o poder político é o tema central de boa parte das pesquisas acadêmicas que trabalham com o conceito de que a luta pela hegemonia política parte essencialmente do resultado da luta de classes sociais ou frações de classes que disputam o poder (MARX, 2011).

Entretanto, com o surgimento das redes sociais e do complexo fenômeno da desinformação que assola o século XXI, novas variáveis passaram a ter papel de destaque nos processos capazes de influenciar ora direta, ora indiretamente, o resultado das eleições. Por conta dos escândalos de corrupção envolvendo a elite política, os esquemas de montagem do chamado “caixa dois” têm se sofisticado, passando a contar com o uso de robôs nas mídias sociais e na montagem de redes complexas de *WhatsApp* para disseminação de notícias falsas¹.

Assim, buscar técnicas e estratégias capazes de analisar e extrair informações capazes de auxiliar na produção de conhecimento não só relacionado aos conjuntos de dados oficiais, mas também de mídias tradicionais e das redes sociais, é de considerável importância para que se possa entender melhor nossa realidade política. É nesse ponto que, para que se consiga tratar com grandes volumes de dados, ou o chamado *Big Data*, a mineração de dados ou *data mining* reúne modernas metodologias para a aquisição de conhecimento, utilizando diversos algoritmos computacionais próprios para tratamento de grandes volumes de dados (HAN; KAMBER; PEI, 2011).

Dentro desta perspectiva, o grupo de pesquisa Perspectivas Filosóficas da Informação (Perfil-I), do Programa de Pós Graduação em Ciência de Informação do IBICT-UFRJ, formulou o projeto Financiamento, Atuação política e Repercussão Midiática e Informacional (FARMI) com o objetivo de implantar uma plataforma digital de acesso aberto para apresentar cruzamentos, análises e visualizações de informação referente a financiamentos eleitorais, atuações parlamentares e repercussões midiáticas, tanto nas mídias convencionais como nas redes digitais, partindo da materialidade sócio-histórica dos registros. Este projeto acabou por compor à época de sua criação o rol de atividades do Laboratório em Rede de Humanidades Digitais (LARHUD) que atua naquele mesmo programa.

¹ *Documento confirma oferta ilegal de mensagens por WhatsApp na eleição.* Disponível em: <<https://www1.folha.uol.com.br/poder/2018/10/documento-confirma-oferta-ilegal-de-mensagens-por-whatsapp-na-eleicao.shtml>>. Acesso em: 30 out. 2010.

Existe uma série de estudos sociológicos e de natureza política sobre o uso da mineração de dados na sociedade. O *European Journal of Cultural Studies* dedicou uma edição inteira buscando entender como a mineração de dados está influenciando a sociedade e aprofundando ainda mais as relações de poder e controle (ANDREJEVIC; HEARN; KENNEDY, 2015). Apesar destes estudos serem de grande relevância, aqui a nossa preocupação recai no âmbito das HDs no tocante à reflexão, necessária e urgente, sobre como a apropriação das práticas de mineração de dados por parte dos cientistas das humanidades têm se tornado relevante e imperativo para o novo cenário cada vez mais digitizado das práticas sociais, culturais e políticas na contemporaneidade.

O recurso tecnológico que exponencialmente se produz em face das demandas do cenário científico das humanidades traz um desafio duplo: produzir competências, assim como repensar currículos na formação dos cientistas sociais² para os novos desafios que o campo lhes apresenta, e não isentar-se do pensamento crítico, uma vez que os recursos tecnológicos não devem ser vistos como panacéia. A utopia tecno-neoliberal (BEZERRA; SCHNEIDER; SALDANHA, 2013), junto com o libertarianismo vivido no início da internet por seus criadores e idealizadores, não deram suficiente atenção às novas clivagens políticas e socioeconômicas da cultura digital, cada vez mais capturada pelo mercado e espetacularizada (DEBORD, 2015) no cerne de suas práticas, de sua *doxa*, o que nos apresenta um novo horizonte de desafios teóricos, metodológicos e políticos.

Em síntese, pretende-se fazer uma análise das possibilidades de pesquisa que podem ser realizadas utilizando técnicas de mineração de dados à luz das práticas e da inovação metodológica características das HDs (BERRY; FAGERJORD, 2017; PIMENTA, 2016; SPIRO, 2013; RUPPERT; LAW; SAVAGE, 2013), especialmente sobre o cenário eleitoral brasileiro para o qual realizamos uma iniciativa de elaboração de um sítio eletrônico em que pudéssemos agregar informações sobre tal panorama.

² O termo “humanista” refere-se ao conceito de humanismo enquanto “prática crítica” do intelectual, cientistas das humanidades, que busca por uma condição humanista “praticante” (SAID, 2007). Essa perspectiva aponta para o fato de que há um papel público cada vez maior dos cientistas das ciências humanas, sociais e sociais aplicadas em se pensar enquanto agentes que produzem para ciência e, por meio da ciência, para a sociedade civil, já que o conhecimento que produzem atua na compreensão mútua entre sujeitos sociais, movimentos e instituições como amálgama na compreensão de seus papéis e idiossincrasias.

FARMI

As análises desenvolvidas no âmbito do FARMI pretendem adotar uma perspectiva crítica, em diferentes graus de complexidade, sobre as relações de interesse entre agentes políticos e econômicos diversos, incluindo as corporações de mídia. Ao construir uma base de dados aberta, o projeto permite o desenvolvimento de novos cruzamentos, análises e visualizações, inclusive para que outros pesquisadores possam fazer uso dos dados já tabulados para suas pesquisas.

Uma das primeiras bases de dados resultantes deste projeto, denominada Politlab, foi disponibilizada para testes em 2017 e permite análises e visualizações, de modo amigável, dos resultados das eleições de 1998 a 2016, incluindo aí os dados de financiamento de campanha e o histórico de cada candidato ao longo do período em questão. Esta primeira versão, ainda em fase de testes, permite a busca pelo nome do político (com direito a desambiguação, mostrando os políticos que tenham o mesmo nome, o Estado onde ele concorreu e o nome utilizado por ele na eleição). O software utiliza dados do TSE, do projeto CEPESP DATA da Fundação Getulio Vargas e do projeto Tribuna.³ Como objetivo secundário, o projeto FARMI procura mapear as diversas iniciativas construídas pela sociedade civil, acadêmicas ou não, que busquem contribuir para a promoção de transparência dos poderes políticos, competência crítica em informação e cidadania ampliada. Até dezembro de 2018, o mapeamento contava com 45 projetos divididos em três principais categorias: Financiamento de Campanha, Atuação Política e Repercussões Midiáticas. Dentro da categoria de Atuação Política, foram catalogadas iniciativas que investiguem as atuações no Legislativo, no Executivo e no Judiciário. O resultado do mapeamento foi divulgado no site do projeto e algumas conclusões interessantes foram obtidas a partir desse mapeamento (GOMES; SCHNEIDER; BEZERRA, 2018).

A contribuição do projeto para a transparência dos poderes políticos não requer maiores esclarecimentos, com base no que foi dito acima. Sobre a questão da cidadania ampliada, cabe esclarecer que empregamos a expressão na acepção de Boito Jr. (2009), que agrega aos direitos civis, políticos e sociais a reivindicação de direitos econômicos, para além da exploração de classe, que permitam tornar efetivos os três direitos anteriores:

³ Disponível em: <<http://www.farmi.pro.br/politlab/>>. Acesso em: 01 abr. 2018.

Os direitos civis seriam os necessários à liberdade individual (liberdade de ir e vir, liberdade de imprensa, pensamento e fé, direito à propriedade, à livre escolha do trabalho e à justiça); os políticos estariam vinculados à participação no exercício do poder político; os sociais, à assistência social, educação pública e demais elementos que configuram o Estado do bem-estar (MORETZSOHN, 2011, p. 143).

Quanto à competência crítica em informação, entendemos estar contribuindo com o seu fortalecimento ao municiar o cidadão de informação relevante sobre a dinâmica interna de uma dimensão fundamental da vida social que, a despeito da sua permanente exposição midiática, e talvez também por causa da forma típica dessa exposição, espetacular, permanece opaca para os não especialistas. Competência crítica em informação diz respeito primariamente à aquisição de determinados conhecimentos e técnicas referentes à busca e recuperação de informação relevante, mas também à formação de hábitos de pensamento críticos, questionadores, quanto à própria relevância da informação. Desse modo, ao acessar informação que torne transparente o que era opaco, ou seja, a dinâmica subjacente ao jogo político-econômico-informacional do momento, o cidadão, assim espera-se, há de amadurecer seu senso crítico em relação à relevância da informação espetacular inflacionada sobre esse mesmo jogo, e que aliás faz parte do próprio jogo que espetaculariza, sem assumir publicamente que o faz, aparentando transparência, mas produzindo mais opacidade. Isso vale tanto para as formas midiáticas convencionais de informação espetacularizada quanto para as novas, que circulam no ambiente *on-line*. Por isso, um conhecimento sobre as novas formas de circulação, acesso e mediação da informação propagada nas redes sociais e demais plataformas digitais, que atualmente representam uma das principais fontes de informação dos indivíduos com acesso à internet, é igualmente necessário para o fortalecimento da competência crítica em informação e, concomitantemente, da cidadania. Além disso, ao municiar o cidadão não só de dados limpos e informação relevante, mas igualmente de ferramentas amigáveis que permitam a produção de novos cruzamentos e visualizações, o projeto estará estimulando não somente o consumo de informação relevante, capaz de favorecer o amadurecimento do senso crítico do usuário, mas também a produção e circulação de informação relevante por parte do próprio usuário, fomentando assim novos fluxos informacionais,

num círculo virtuoso. Dependendo então do alcance da iniciativa, práticas na vida cotidiana tornam-se tema e problema essenciais para a formulação de soluções e ferramentas digitais capazes de intermediar o processo crítico e analítico, até mesmo preventivo, concernente ao financiamento de campanhas políticas e suas consequências, informacionais ou de outra ordem.

SOBRE O FINANCIAMENTO DE CAMPANHA

Um dos componentes centrais do funcionamento da democracia é o sistema eleitoral. Para falar em eleições, especialmente no Brasil, é imprescindível falar de financiamento das campanhas políticas, já que esse é um dos principais fatores que determinam o sucesso no pleito. Salvo raras exceções, a regra é bem definida: quanto maior o poder econômico do candidato, maiores as chances que ele terá de se eleger (MANCUSO; SPECK, 2015).

Desde 1965, o financiamento de campanhas no Brasil é misto: parte do financiamento é feito pelo Estado e parte é arrecadado de pessoas (físicas ou jurídicas) dispostas a patrocinar um candidato ou partido. O financiamento feito pelo Estado se dá por dois canais: a partir do fundo partidário e pela concessão do horário gratuito eleitoral, além de spots no rádio e na televisão.

Segundo Mancuso (2015), as primeiras pesquisas sobre o cenário brasileiro com um corte empírico sobre financiamento de campanha datam de 1997, tendo o trabalho de Samuels na década de 2000 se convertido na base metodológica para as pesquisas subsequentes. Segundo o pesquisador, que fez uma extensa revisão bibliográfica sobre o tema, até 2014 pode-se categorizar as linhas de pesquisa em três grandes questões: “(i) a relação entre investimentos e resultados eleitorais; (ii) a relação entre investimentos eleitorais e benefícios para os financiadores e (iii) os determinantes do investimento eleitoral”. (p.158)

Em boa parte da literatura, é ponto pacífico que o financiamento empresarial tem sido o maior influenciador do resultado eleitoral e que este vem crescendo acima da inflação desde 1998, eleição após eleição. Samuels (2001) já indicava que, desde 1994, a maior parte do financiamento vinha de fontes empresariais e que a participação da sociedade civil tenderia a diminuir. Somente em 2010, o investimento para todos os cargos em disputa chegou a quase 3 bilhões de reais e 74% deste valor veio de pessoas jurídicas (MANCUSO, 2015).

Após 2014, houve uma mudança significativa, que afetou a realidade do pleito: até as eleições de 2014, tanto cidadãos quanto empresas privadas estavam aptos a contribuir, mas a partir da minirreforma eleitoral (Lei n.º 13.165, de 29 de setembro de 2015) foi vedada a contribuição de pessoas jurídicas. Por conta dessa minirreforma e do agravamento das denúncias e condenações pelo uso de contas paralelas, chamadas de “caixa dois” (dinheiro fornecido por empresas para políticos e partidos, não contabilizado no Tribunal Eleitoral), o volume de dinheiro envolvido nas eleições em 2016 e 2018 foi bem menor que o apurado nos anos anteriores, como podemos ver no Quadro 1:

QUADRO 1 - FINANCIAMENTO ELEITORAL PRIVADO ENTRE 2010 E 2018

Ano	Recursos Privados
2010	2.642.498.331
2014	6.205.357.689
2018	1.138.266.640

Fonte: Fonte: Dados calculados pelo autor com base nas informações divulgadas pelo TSE; Santos (2016).

Entretanto, há indícios de que as empresas continuaram patrocinando legalmente os candidatos, indiretamente, mediante doações feitas pelos sócios majoritários das mesmas empresas que antes financiavam as campanhas.

Um elemento dessa problemática particularmente sensível para a Ciência da Informação é o fato de apesar dos dados de financiamento de campanha estarem disponíveis, eles não são propriamente acessíveis, ou seja, a compreensão e o tratamento destes dados não são triviais, pois os metadados são difíceis de obter e entender, e a quantidade e complexidade das variáveis envolvidas tampouco facilitam o processo de análise. Além disso, os dados brutos carecem de informações suplementares, que os pesquisadores precisam adicionar à base para que possam extrair algum conhecimento relevante.

Tais desafios apontam inequivocamente para uma espécie de *turning point* na academia – em específico relacionada aos campos disciplinares das Humanidades – que precisa rever continuamente suas práticas de investigação, suas metodologias, de modo a lidar com uma tipologia documental multimodal, ubíqua, cifrada, de volume (em termos de memória, bytes) e velocidade (em termos de capacidade de atualização dos dados e de transferência dos mesmos entre canais info-comunicacionais, ambos

digitizados) jamais mensurados. Não obstante, a própria natureza da informação, neste caso ligada à base de dados do TSE, requer uma preparação por parte do agente da pesquisa relacionada às normatividades características desse conjunto de fontes ora trabalhadas, ora disponibilizadas pelo TSE. O desenvolvimento e emprego de um método, neste caso, opera razoavelmente por aquilo que Rogers (2014, p. 78-79, tradução nossa) aponta serem os “métodos digitais”:

[...] prática de readaptar dispositivos, não sendo apenas caixas de ferramentas ou instruções de operação para pacotes de softwares; eles [os métodos digitais] lidam com questões mais amplas sobre como fazer pesquisas on-line. Eles encorajam uma visão ou imaginação sociológica sobre oportunidades de pesquisa existentes na cultura on-line indo atrás do meio [da mediação], em vez de submetê-lo ao preço [às exigências] disciplinar.⁴

Ademais, sem conhecimento prévio do cenário e da estrutura eleitoral brasileira cometer-se-á impertinências quanto ao uso e emprego dos dados e de suas informações provenientes. Por exemplo, a base de financiamento fornecida pelo TSE, de fato, traz o CNPJ da empresa que doou para um determinado partido ou político, mas nada nos diz de que setor econômico essa empresa faz parte, nem se é uma empresa de pequeno ou de grande porte. Outras questões se impõem com a nova determinação de que somente pessoas físicas podem fazer doações: o doador está ligado ou é sócio de alguma empresa? Qual é a sua atividade profissional? Até que ponto é possível investigar esse doador sem que o pesquisador esbarre em questões éticas e de privacidade?

A partir destes questionamentos, pode-se perceber que existe um número muito grande de variáveis que podem ser utilizados na análise do financiamento de campanha e de sua relação com o sucesso eleitoral. E que talvez outros métodos de análise possam ser utilizados para o cruzamento de todas as variáveis envolvidas para que se consiga entender melhor o fenômeno social em pauta. Nesse contexto, o uso de técnicas de mineração de dados pode ser útil.

⁴ [...] the practice of repurposing devices, are not just toolkits or operating instructions for software packages; they deal with broader questions about how to do research online. They encourage a sociological outlook or imagination about research opportunities that exist in online culture by following the medium rather than asking it to do one's disciplinary bidding.

MINERAÇÃO DE DADOS

Nos últimos anos, a mineração de dados (*data mining*) tem atraído muita atenção da pesquisa científica, da indústria da informação e da sociedade como um todo. Se buscarmos na história da ciência da computação, o uso das técnicas utilizadas hoje em mineração de dados vem sendo feito desde 1960, mas somente a partir da década de 80 que o termo passou a ser empregado para consolidar um conjunto de processos e algoritmos que tinham um mesmo objetivo:

“[...] a extração automatizada de padrões que representem algum conhecimento implícito que esteja armazenado em grandes volumes de dados, sejam eles armazenados em banco de dados, *data warehouses*, na web ou qualquer outro tipo de repositório massivo ou fluxo de dados.” (HAN; KAMBER; PEI, 2011, p. 23, tradução nossa).⁵

Para Fayyad, Piatetsky-Shapiro e Smyth (1996), a mineração de dados é parte de um processo maior denominado *Knowledge Discovery on Databases* (KDD), que envolve a coleta, a limpeza dos dados, o processamento em si e a análise e visualização dos resultados encontrados. Entretanto, consideramos a escolha do termo KDD equivocada por diversas razões. Primeiro, porque a mineração de dados é utilizada em repositórios que não estão necessariamente em banco de dados (*databases*), podendo estar em coleção de textos, em imagens e até mesmo em fluxos de dados enviados em tempo real por dispositivos como medidores de temperatura e de posicionamento georreferenciado. Mesmo se o termo *databases* fosse substituído por *data* (dado), alguns autores ressaltam que o processo de pré-processamento e limpeza de dados é intrínseco ao processo de mineração, ou seja, não existe mineração sem modelagem e pré-processamento dos dados (HAND, MANNILA; SMYTH, 2001). E, por fim, do ponto de vista epistemológico, o termo KDD é muito amplo e pode se confundir com a própria definição de Ciência. Afinal, o que é a ciência senão a busca de conhecimento através de dados empíricos coletados?

⁵ “the automated or convenient extraction of patterns representing knowledge implicitly stored or captured in large databases, data warehouses, the Web, other massive information repositories, or data streams.”

Assim, este capítulo segue o entendimento de outros autores (COENEN, 2011; HAN, KAMBER; PEI, 2011; TRYBULA, 1999)⁶ e utiliza a expressão “Mineração de Dados” como mais apropriada que KDD, podendo ser utilizada para identificar todo processo que envolve a descoberta de novos conhecimentos a partir de grandes volumes de dados.

Pode-se categorizar as diversas estratégias de mineração de dados a partir do tipo de conhecimento a ser descoberto, dividindo assim estas estratégias em duas categorias distintas: descoberta supervisionada e descoberta não supervisionada.

A descoberta *supervisionada* requer que parte dos dados sejam pré-classificados a priori, isto é, seleciona-se um subconjunto dos itens a serem estudados e associa-se uma etiqueta única que signifique a classe à qual o item pertence, para que o algoritmo possa ter uma base inicial de conhecimento. A partir daí o algoritmo “constrói” uma função baseado nas entradas e saídas, e quando um novo elemento (do qual não se sabe a qual grupo pertence) for inserido no sistema, o algoritmo irá indicar a qual grupo ele provavelmente irá pertencer.

Para melhor compreensão desse conceito, pode-se imaginar um exemplo no contexto da ciência política, seleciona-se um grupo aleatório de parlamentares e cada um recebe um rótulo indicando se ele é de esquerda, de direita ou de centro. Ou seja, o ser humano indica ao algoritmo que um determinado político (com suas inúmeras características, como sexo, profissão, idade, partido político etc.) faz parte de um determinado grupo. Uma vez que o algoritmo tenha calculado a função que faz este mapeamento de forma precisa, é possível aplicar este mesmo algoritmo para um novo conjunto de elementos em que não se conhece, por exemplo, a orientação política de um determinado candidato.

Já a aprendizagem não supervisionada procura inferir uma função que descreva estruturas ocultas de dados que ainda não foram categorizados. No exemplo dos políticos, digamos que temos apenas as características dos políticos (dados de entrada) e nenhuma categorização aparente (dado de saída), ou seja, não temos um modelo ou categorização prévia para aplicar ao nosso conjunto de dados. Como podemos aprender mais sobre esse conjunto de dados? No nosso exemplo, como

⁶ É interessante notar que a obra de Han e Kamber vai mudando a sua percepção entre os termos KDD e Data Mining ao longo de suas três edições. Em sua primeira edição, os autores seguem a definição de Fayyad. Já na segunda edição já concordam que o Data Mining é considerado pela academia e pelo mercado como o processo integral e, por fim, na terceira edição, citam apenas a perspectiva de Fayyad como uma forma de ver o processo, mas que a grande maioria dos autores já colocam os dois termos como sinônimos, sendo que a preferência é pelo termo Data Mining.

podemos dividir os políticos em categorias que não conseguimos identificar? Será que existe apenas esquerda, direita e centro, ou há outra categorização que permita distinguir os políticos, mas que não esteja aparente? Esse processo de classificação ou categorização é chamado de segmentação (*clustering*) não supervisionada, pois não há uma supervisão humana para determinar quais ou quantas categorias existirão para agrupar elementos de um grupo. Estes agrupamentos são denominados *clusters*, partições ou segmentos. O modelo teórico que fundamenta esse processamento é baseado no princípio de maximizar as similaridades intra-classes e minimizar a similaridade inter-classes (HAN; KAMBER, 2001).

Ainda dentro dos algoritmos não supervisionados, destaca-se a detecção de anomalias (*outliers*) ou, em jargão popular, os pontos fora da curva. A detecção de *outliers* e os processos de segmentação são duas tarefas altamente relacionadas. O agrupamento encontra os padrões de maioria em um conjunto de dados e organiza os dados em função da similaridade de suas características, enquanto a detecção de *outliers* tenta capturar os casos excepcionais que se desviam substancialmente dos padrões da maioria (HAN; KAMBER; PEI, 2011).

É importante também ressaltar um outro recurso que pode ser utilizado tanto nas estratégias supervisionadas quanto nas não supervisionadas: a visualização de dados. A construção de visualizações planas ou espaciais dos dados procura aproveitar as capacidades cognitivas dos seres humanos para favorecer *insights* visuais sobre os dados que muitas vezes os algoritmos de mineração de dados não conseguem fornecer. Entretanto, o processo de montar visualizações em grandes bases não é trivial, pois é necessário encontrar uma visão ou projeção dos dados que reduza a complexidade enquanto captura informações importantes. O objetivo é reduzir a complexidade ao mesmo tempo que se tenta perder a menor quantidade de informação possível (FAYYAD, 2001).

O uso de técnicas de mineração de dados tem sido crescente em artigos na área das humanidades. Desde 2017, existe uma revista científica especializada em Mineração de Dados para Humanidades Digitais – *The Journal of Data Mining & Digital Humanities*⁷ –, que conta com mais de 58 artigos publicados. Em pesquisa bibliográfica na base Web of Science utilizando o termo “Data Mining”, foram encontrados 79 artigos somente nas áreas de Sociologia e Ciência Política. Além do uso de bases de dados tradicionais, pesquisas em humanidades têm consultado outras fontes de dados com resultados bem promissores, a saber:

⁷ Disponível em: <<https://jdmhd.episciences.org/>>. Acesso em: 09 set. 2020.

- Mineração de Textos (*Text Mining*): procura categorizar ou encontrar padrões em grandes coleções de documentos como notícias de jornal, páginas na web, artigos de revistas científicas e obras de um determinado autor ou período. Também é muito utilizado na mineração de opiniões ou análise de sentimentos, quando se procura encontrar padrões em conversações realizadas nas redes sociais ou formulários *on-line* sobre um tema específico.
- Mineração de Imagens: analisa imagens ou vídeos para simplesmente categorizá-los a partir de critérios pré-definidos ou para encontrar padrões específicos em um grande volume de imagens, como, por exemplo, imagens de satélite ou recuperadas por telescópios.
- Mineração de Grafos: é o estudo de grafos gerados a partir do relacionamento entre entidades ou pessoas, como o relacionamento de pessoas nas redes sociais, operações financeiras ou citações bibliográficas.

CONSIDERAÇÕES FINAIS

Quando o projeto FARMÍ foi apresentado no Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), em 2017, no Grupo de Trabalho (GT) de Política e Economia da Informação (GT5), um número considerável de participantes presentes no GT expuseram a sua preocupação com o excesso do uso de algoritmos, do chamado Big Data, e as possibilidades de uso de mineração de dados com os nossos dados pessoais e principalmente a falta de controle da sociedade sobre eles. Como resposta para tais questionamentos, os pesquisadores do FARMÍ trouxeram o seguinte dilema: Qual alternativa nos restava? Aceitar passivamente a influência de tais algoritmos nas nossas vidas? Ou nos apropriar dessas tecnologias, entendê-las, adaptá-las a novos usos cidadãos, buscando novos métodos de pesquisa e uma visão mais crítica deste fenômeno, amparada nestes mesmos volumes de dados?

Efetivamente, desenvolver modelos para entender a vida real raramente é apenas um problema de matemática aplicada ou de computação. Além do domínio da técnica, a construção requer consciência do contexto, das preocupações das partes interessadas, do momento histórico e das repercussões que tal pesquisa trará para a sociedade.

Ações que visem capacitar cientistas sociais para que possam usar técnicas de mineração de dados em suas pesquisas e, ao mesmo tempo, envolver cientistas de dados, sejam eles da computação, da estatística ou da matemática para que entendam o contexto social no qual a pesquisa se insere, abre-se um novo leque de possibilidade para as humanidades.

REFERÊNCIAS

ALVES, D. As Humanidades Digitais como uma comunidade de práticas dentro do formalismo acadêmico: dos exemplos internacionais ao caso português. *Ler História* [Online], v. 69, 2016. Disponível em: <<http://journals.openedition.org/lerhistoria/2496>>. Acesso em: 01 fev. 2018.

ANDREJEVIC, M.; HEARN, A.; KENNEDY, H. Cultural Studies of Data Mining: Introduction. *European Journal of Cultural Studies*, v. 18, n. 4-5, p. 379-394, 2015. <https://doi.org/10.1177/1367549415577395>

BERRY, D. The Computational Turn: Thinking about the Digital Humanities. *Culture Machine*, v. 12, 2011. Disponível em: <<http://sro.sussex.ac.uk/49813/>>. Acesso em: 19 jan. 2019.

BERRY, D.; FAGERJORD, A. *Knowledge and Critique in a Digital Age*. Cambridge, MA: Polity Press, 2017.

BEZERRA, A. C.; SCHNEIDER, M.; SALDANHA, G. Ascensão e queda da utopia tecnoliberal: a dialética da liberdade sociotécnica. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. *Anais...* Florianópolis: UFSC, 2013. Disponível em: <<http://enancib.sites.ufsc.br/index.php/enancib2013/XIVenancib/paper/viewFile/364/358>>. Acesso em: 23 jul. 2016.

BOITO JR., A. Cidadania e classes sociais. In: _____. *Estado, política e classes sociais*. São Paulo: Unesp, 2009. p. 247-261.

BRASIL. *Lei nº 12.527, de 18 de novembro de 2012*. Lei de Acesso à Informação. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/12527.htm>. Acesso em: 02 mar. 2019.

COENEN, F. Data Mining: Past, Present and Future. *The Knowledge Engineering Review*, v. 26, n. 1, p. 25-29, 2011.

CHEN, S. Y.; LIU, X. The Contribution of Data Mining to Information Science. *Journal of Information Science*, v. 30, n. 6, p. 550-558, Dec. 2004.

DEBORD, G. *A sociedade do espetáculo*. Rio de Janeiro: Contraponto, 2015.

FAYYAD, U.; PIATETSKY-SHAPIRO, G. SMITH, P. From Datamining to Knowledge Discovery in Databases. *AI Magazine*, v. 17, n. 3, p. 37-54, 1996.

FAYYAD, U. et al. *Information Visualization in Data Mining and Knowledge Discovery*. San Francisco, USA: Morgan Kaufmann Publishers, 2001.

FEKETE, J-D.; VAN WIJK, J. J.; STASKO, J. T.; *et al.* The Value of Information Visualization. In: KERREN, A.; STASKO, J. T.; FEKETE, J-D. *et al.* (Org.). *Information Visualization: Human-Centered Issues and Perspectives*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. p.1-18. (Lecture Notes in Computer Science). https://doi.org/10.1007/978-3-540-70956-5_1

GOMES, J. C.; SCHNEIDER, M.; BEZERRA, A. C. Aplicativos cívicos: apropriação de dados abertos governamentais pela sociedade. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2018, Londrina. *Anais...* Londrina: ANCIB, 2018.

HAN, J.; KAMBER, M. *Data Mining: Concepts and Technics*. Londres: Academic Press, 2001.

HAN, J.; KAMBER, M.; PEI, J. *Data Mining: Concepts and Techniques*. San Francisco, USA: Morgan Kaufmann Publishers Inc., 2011.

HAND, D. J.; MANNILA, H.; SMYTH, P. *Principles of Data Mining (Adaptive Computation and Machine Learning)*. Cambridge, MA: MIT Press, 2001.

KIRSCHENBAUM, M. G. The Remaking of Reading: Data Mining and the Digital Humanities. *Proceedings of The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation*. Baltimore, 2007.

KLEIN, J. T. *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*. [s.l.]: University of Michigan Press, 2015.

MANCUSO, W. P.; SPECK, B. W. Financiamento empresarial na eleição para deputado federal (2002-2010): determinantes e consequências. *Teoria & Sociedade*, v. 23, n. 2, p. 103-125, 2015.

MANCUSO, W. P. Investimento eleitoral no Brasil: balanço da literatura (2001-2012) e agenda de pesquisa. *Revista de Sociologia e Política*, v. 23, n. 54, p.155-183, 2015.

MANOVICH, L. *Software Takes Command: Extending the Language of New Media*. London: Bloomsbury, 2013.

MARX, K. *0 18 de brumário de Luís Bonaparte*. São Paulo: Boitempo, 2011.

MORETZSOHN, S. A cidadania através do espelho: do estado do bem-estar às políticas de exceção. *Sinais Sociais*, Rio de Janeiro, v. 15, n. 15, p. 138-169, 2011.

PEIXOTO, V. M. *Eleições e financiamento de campanhas no Brasil*. Rio de Janeiro: Editora Garamond, 2016.

PIMENTA, R. M. Os objetos técnicos e seus papéis no horizonte das Humanidades Digitais: um caso para a ciência da informação. *Revista Conhecimento em Ação*, v. 1, n. 2, p.

1-14, 2016. Disponível em: <<http://www.brapci.ufpr.br/brapci/v/a/23535>>. Acesso em: 20 dez. 2018.

ROGERS, R. *Digital Methods*. Cambridge, London: MIT Press, 2014.

ROSA, H. *Accélération: une critique sociale du temps*. Tradução Didier Renault. Paris: La Découverte, 2014.

RUPPERT, E.; LAW, J.; SAVAGE, M. Reassembling Social Science Methods: The Challenge of Digital Devices. *Theory, Culture & Society*, v. 30, n. 4, p. 22-46. 2013. <https://doi.org/10.1177/0263276413484941>

SAID, E. W. *Humanismo e crítica democrática*. São Paulo: Companhia das Letras, 2007.

SAMUELS, D. Money, Elections and Democracy in Brazil. *Latin American Politics and Society*, v. 43, n. 7, p. 27-48. 2001.

SANTOS, B. C. dos. *Interesses econômicos, representação política e produção legislativa no Brasil sob a ótica do financiamento de campanhas eleitorais*. 2016. Tese (Doutorado) – Faculdade de Direito, Universidade Federal de Minas Gerais, Belo Horizonte, 2016.

SPIRO, L. This Is Why We Fight: Defining the Values of the Digital Humanities. In: GOLD, M. K. (Ed.). *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2013.

TRYBULA, W. J. Text Mining. *Annual Review of Information Science and Technology*, v. 34, p. 385-419, 1999.

Análise e mapeamento da organização técnica dos acervos digitais das instituições federais vinculadas ao Ministério da Cultura: fontes de informação para o desenvolvimento das Humanidades Digitais

Dalton Lopes Martins
Joyce Siqueira

INTRODUÇÃO

Atentos às novas necessidades da sociedade e acompanhando a evolução tecnológica, diversos museus, bibliotecas, arquivos e centros culturais do Brasil e do mundo têm se interessado e buscado caminhos para a digitalização de seus acervos. Para Baggio e Flores (2013), tal movimento também corrobora a necessidade de garantir a preservação digital da informação por meio de uma tecnologia flexível para arquivar, preservar e possibilitar o acesso a documentos em formato digital.

Vale salientar que a digitalização por si só não possibilita o pleno acesso aos acervos. Neste tocante, os repositórios digitais são tendência. Shintaku (2017) explica que, em sua concepção inicial, estes foram vinculados ao “repor”, ou seja, tratavam-se de um local para reposição ou relocação de coisas. Entretanto, o contexto mudou, e atualmente são considerados, com frequência, como primeira fonte, como por exemplo, os diversos portais de teses e dissertações.

Um dos pontos fundamentais para a pesquisa científica e o desenvolvimento conceitual e prático das Humanidades Digitais (HDs) é a possibilidade de se utilizar bases de dados de ampla representatividade e interesse cultural como fontes de informação para testes de algoritmos, experimentos com ferramentas de análise, aplicação de metodologias quantitativas e qualitativas, comparação com resultados obtidos internacionalmente, além da possibilidade de se testar hipóteses e avançar no conhecimento empírico.

Desse modo, encontrar fontes de informação que possam ser utilizadas como base para pesquisas em HDs, ou seja, que forneçam informações sobre seus acervos e ainda permitam que essas possam ser reutilizadas, a partir de sua disponibilização em formatos abertos e interoperáveis, torna-se um problema de grande interesse.

Apesar disso, há pouca informação sistematizada no Brasil sobre a existência de fontes de informação e, menos ainda, da qualidade técnica dos dados disponibilizados e se podem ou não serem facilmente reutilizados. Segundo a pesquisa TIC Cultura 2016, do Comitê Gestor da Internet (2017), a disponibilidade de acervos dos arquivos brasileiros na Internet é de 30% das unidades, sendo que menos que 15% dos museus e bibliotecas possuem a mesma prática. Ou seja, o catálogo dos acervos dessas instituições não estão disponíveis em rede e, muitas das vezes, sequer digitalizados. Tal fato, como se pode avaliar pelos números acima, torna muito difícil o avanço de pesquisas nas ciências humanas, que se aproveitem dos recursos das HDs. É questão fundamental a ser resolvida para avanço do campo no país.

É, de fato, pela perspectiva da digitalização e do reuso desses conteúdos que muitos laboratórios e centros de pesquisa têm se interessado, devido à possibilidade de utilização dos acervos das instituições memoriais. Como ressalta Puntoni (2017, p. 129-130):

[...] não podemos deixar de perceber o enorme papel desempenhado neste processo de aproximação da cultura digital e das instituições memoriais pelos centros de pesquisa e pelas universidades, sobretudo pelos profissionais ligados à ciência da informação, aos estudos do patrimônio ou mesmo ao que começava a se definir como humanidades digitais (*digital humanities*).

Como forma de problematizar a questão acima e avançar na análise empírica dessas fontes de informação, optou-se, nesta pesquisa, em focar nas instituições vinculadas ao Ministério da Cultura (MinC), sobretudo por sua relevância e abrangência nacional, sendo elas, de acordo com MinC (BRASIL, 2013c): o Instituto do Patrimônio Histórico e Artístico Nacional (IPHAN), o Instituto Brasileiro de Museus (IBRAM), a Agência Nacional do Cinema (ANCINE), a Fundação Casa de Rui Barbosa (FCRB), a Fundação Cultural Palmares (FCP), a Fundação Nacional das Artes (FUNARTE) e a Biblioteca Nacional (BN).

Das sete instituições supracitadas, cinco possuem acervos digitalizados e disponíveis na internet, em 11 projetos distintos: o IPHAN (1), o IBRAM (4), o FCRB (1), a FUNARTE (3) e a BN (2). Por ter como objetivo mapear a organização técnica destes acervos e discutir sua disponibilidade para a potencialização de pesquisas no campo das HDs, definiu-se importantes questões a serem respondidas: Qual *software* gestor de repositórios digitais é utilizado? A licença do *software* é livre ou proprietária? Qual o tamanho do acervo? Como os itens foram estruturados em coleções? Quais são os formatos dos arquivos armazenados? Qual padrão de metadados foi implantado? Há linguagem documentária para indexação dos conteúdos? Se sim, qual? Permite a coleta de dados via protocolo OAI-PMH?

Além disso, esta pesquisa retrata a experiência de acesso dos pesquisadores a estes acervos, descrevendo, basicamente, se oferecem interfaces amigáveis e maneiras eficientes de se realizar buscas, além de relatar as dificuldades em encontrar os itens avaliados nesta pesquisa.

INSTITUIÇÕES VINCULADAS AO MINISTÉRIO DA CULTURA

O MinC¹ foi criado em 1985, a partir do desmembramento do Ministério da Educação e Cultura, e desenvolve ações para o reconhecimento da importância da cultura para a construção da identidade nacional (BRASIL, 2013b).

Desde então, compete a este Ministério a:

- I. política nacional de cultura;
- II. proteção do patrimônio histórico e cultural;
- III. regulação de direitos autorais;
- IV. assistência e acompanhamento da Casa Civil da Presidência da República e do Instituto Nacional de Colonização e Reforma Agrária - Incra nas ações de regularização fundiária, para garantir a preservação da identidade cultural dos remanescentes das comunidades dos quilombos; e
- V. desenvolvimento e implementação de políticas e ações de acessibilidade cultural (BRASIL, 2013a).

¹ Após o encerramento desta pesquisa, a estrutura da Cultura foi incluída no Ministério da Cidadania e criada a Secretaria Especial da Cultura, Conforme o Decreto nº 9.674, de 2 de janeiro 2019. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/decreto/D9674.htm>. Acesso em: 31 maio 2020.

De acordo com o MinC (BRASIL, 2013d), a sua atual estrutura regimental é regida pelo Decreto nº 9.411/2018, no qual constam seis órgãos de assistência direta e imediata, seis secretarias e sete entidades vinculadas, sendo três autarquias e quatro fundações e escritórios regionais, que abrangem o país territorialmente. Vale dizer que o decreto extingue o ministério e cria a Secretaria Nacional de Cultura, vinculada ao atual Ministério da Cidadania. Destes setores, o foco desta pesquisa são as entidades vinculadas, cujas informações gerais serão apresentadas a seguir.

INSTITUTO DO PATRIMÔNIO HISTÓRICO E ARTÍSTICO NACIONAL

O IPHAN foi criado em janeiro de 1937, por meio da Lei nº 378, para responder pela preservação do Patrimônio Cultural Brasileiro, assegurando sua permanência e usufruto para as gerações presentes e futuras (IPHAN, 2014b). Além disso, é responsável pela conservação, salvaguarda e monitoramento dos bens culturais brasileiros inscritos nas Listas do Patrimônio Mundial e do Patrimônio Cultural Imaterial da Humanidade.

De acordo com o IPHAN (2014a), a iniciativa de digitalizar os documentos resultou, em 2013, na criação da Rede de Arquivos Iphan, projeto selecionado na Chamada Pública de Seleção para Apoio a Projetos de Preservação de Acervos, que integra os conteúdos produzidos em 80 anos de política de preservação do patrimônio cultural.

INSTITUTO BRASILEIRO DE MUSEUS

O IBRAM foi criado em janeiro de 2009, com a assinatura da Lei nº 11.906, para, entre outras finalidades, promover e assegurar a implementação de políticas públicas para o setor museológico, contribuindo para a organização, gestão e desenvolvimento de instituições museológicas e seus acervos (IBRAM, 2018b; BRASIL, 2013c).

De acordo com o MinC (BRASIL, 2013c), o IBRAM se propõe a estimular a participação de instituições nas políticas públicas para o setor museológico e nas ações de preservação, investigação e gestão do patrimônio cultural musealizado. O órgão, também responsável pela Política Nacional de Museus, visa o aumento

de visitação e arrecadação dos museus, o fomento de políticas de aquisição e a preservação de acervos e criação de ações integradas entre os museus brasileiros (IBRAM, 2018a).

O IBRAM é responsável pela administração direta de 30 museus, cuja lista completa está disponível no site do Instituto², no entanto, destes, apenas seis estão tendo ou têm seu acervo digitalizado, são eles: Museu Histórico Nacional, Museu Arqueológico de Itaipu, Museu Villa-Lobos, Museu da República, Museu Imperial e Museu Nacional de Belas Artes. No ano de 2016, o IBRAM optou por aderir a plataforma Tainacan, e vem implementando esse sistema de gestão e a publicação de coleções digitais em seus museus como estratégia para abertura e disponibilização digital dos acervos musealizados. Destaca-se o lançamento do repositório digital do Museu Histórico Nacional (2018) no ano de 2018.

AGÊNCIA NACIONAL DO CINEMA

A ANCINE foi criada em 2001 pela Medida Provisória 2228-1, e está vinculada desde 2003 ao MinC (ANCINE, 2018; BRASIL, 2013c). Trata-se de uma agência reguladora cuja missão é desenvolver e regular o setor audiovisual em benefício da sociedade brasileira, além proporcionar o desenvolvimento de uma indústria forte, competitiva e autossustentada.

FUNDAÇÃO CASA DE RUI BARBOSA

A FCRB foi instituída em 1928. Sua missão é desenvolver a cultura, a pesquisa, o ensino, a divulgação e o culto a obra e vida de Rui Barbosa, para assim contribuir para o conhecimento de diversidade cultural e o fortalecimento da cidadania (FUNDAÇÃO CASA DE RUI BARBOSA, 2018b, 2018c).

Segundo a Fundação Casa de Rui Barbosa (2018a), suas principais atividades são: manter, preservar e difundir o Museu Casa de Rui Barbosa, seu jardim histórico e o acervo bibliográfico e documental; oferecer qualificação e desenvolver pesquisas e publicações nas áreas de sua competência, além de utilizar seu espaço físico para atividades culturais.

² Disponível em: <<http://www.museus.gov.br/os-museus/museus-ibram/>>. Acesso em 18 set 2018.

FUNDAÇÃO CULTURAL PALMARES

A FCP foi criada em agosto de 1988, pela Lei nº 7.668, sendo a primeira instituição pública voltada à promoção e preservação da arte e da cultura afro-brasileira (FUNDAÇÃO CULTURAL PALMARES, 2016; BRASIL, 2013). Dessa forma, preocupa-se com a igualdade racial e a valorização das manifestações de matriz africana, além de implantar políticas públicas que potencializam a participação da população negra brasileira nos processos de desenvolvimento do País.

FUNDAÇÃO NACIONAL DAS ARTES

A FUNARTE foi criada em 1975 pela Lei nº 6.312, e tem como finalidade promover, estimular e desenvolver atividades culturais, incentivando à produção e à capacitação de artistas, o desenvolvimento da pesquisa, a preservação da memória e a formação de público para as artes no Brasil (FUNARTE, 2010; BRASIL, 2013). Assim, é responsável pelo desenvolvimento de políticas públicas de fomento às artes visuais, à música, ao teatro, à dança e ao circo.

Concede bolsas e prêmios, mantém programas de circulação de artistas e bens culturais, promove oficinas, publica livros, recupera e disponibiliza acervos, provê consultoria técnica e apoia eventos culturais em todos os estados brasileiros e no exterior. Além de manter espaços culturais no Rio de Janeiro, São Paulo, Minas Gerais e Distrito Federal (FUNARTE, 2010).

A FUNARTE vem adaptando seu acervo e digitalizando itens que compõem sua vasta coleção – fotos, arquivos sonoros, textos, documentos – para serem disponibilizados ao público, na internet, tornando-o acessível a cada vez mais brasileiros (FUNARTE, 2013, 2018).

Esse processo se iniciou nos anos 2000, mas apenas em 2010 entrou no ar o projeto Brasil Memória das Artes, totalmente dedicado aos acervos digitalizados da FUNARTE. Além deste, a FUNARTE é responsável pelo Acervo Sergio Britto Digital e Acervo Funarte. Vale destacar o projeto Funarte Digital iniciado no ano de 2018 que, em parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), tem como um de seus objetivos dar maior acesso aos acervos da instituição e, para isso, também utilizará a plataforma Tainacan para difusão dos acervos.

BIBLIOTECA NACIONAL

Conforme a Biblioteca Nacional (2018) e o MinC (BRASIL, 2013c), o núcleo original da Biblioteca Nacional do Brasil teve início em 1808, como Real Biblioteca no Brasil, com a chegada de cerca de 60 mil peças, entre livros, manuscritos, mapas, trazidos por D. João VI e sua corte ao Rio de Janeiro. Em 1876, a instituição passa a se chamar definitivamente Biblioteca Nacional. Sua missão consiste na coleta, tratamento e conservação do patrimônio documental brasileiro em língua portuguesa e sobre o Brasil, bem como assegurar seu estudo, divulgação e as condições para sua disseminação.

Em 2006, foi criada a Biblioteca Nacional Digital (BNDigital), que integrou todas as coleções digitalizadas, igualando a BN às maiores bibliotecas do mundo no processo de digitalização de acervos e acesso a obras e serviços via internet (BIBLIOTECA NACIONAL, 2018; BNDigital, 2018; BRASIL, 2013). Em 2014, foi adquirido o *software* SophiA, de automação bibliográfica, e iniciada a migração dos catálogos para a plataforma, utilizada até hoje. Além da BNDigital, há disponível a Hemeroteca Digital Brasileira, um portal de periódicos nacionais que proporciona ampla consulta, pela internet, ao seu acervo de periódicos – jornais, revistas, anuários, boletins etc. – e de publicações seriadas.

METODOLOGIA

Este é um estudo exploratório e aplicado, pois traz pesquisa bibliográfica, citações e exemplos para compreensão do tema, e os seus resultados poderão ser úteis na resolução de problemas práticos. Além disso, é de abordagem quantitativa e qualitativa, visto que apresenta levantamento de dados e discussão.

O primeiro passo para investigação é verificar quais instituições vinculadas ao MinC possuem acervos digitais para, em seguida, identificar os projetos mantidos por elas. Listados os projetos, inicia-se o mapeamento técnico dos acervos digitais, buscando em cada um os seguintes itens: a) *software*; b) licença; c) tamanho do acervo; d) estrutura de organização das coleções; e) formato dos arquivos disponibilizados; f) padrão de metadados; g) linguagem documentária e h) coleta de dados.

Alguns acervos possibilitam o acesso por meio de *login* e senha, no entanto, esta pesquisa realizou a avaliação dos itens sem usuário logado. Além disso, durante a coleta de dados, os pesquisadores retratam a experiência de acesso aos acervos,

constatando, basicamente, se oferecem interfaces amigáveis e maneiras eficientes de se realizar buscas. Relata-se também as dificuldades em encontrar os itens avaliados. O Quadro 1 apresenta os 11 projetos avaliados.

QUADRO 1 - ACERVOS DIGITAIS DOS PROJETOS MANTIDOS POR INSTITUIÇÕES VINCULADAS AO MINC

Instituição	Projetos	Endereço Eletrônico
Iphan	Rede de Arquivos Iphan	http://acervodigital.iphan.gov.br/xmlui/
Ibram	Museu Histórico Nacional	http://mhn.acervos.museus.gov.br/
	Museu Arqueológico de Itaipu	http://museudeitaipu.museus.gov.br/
	Museu Imperial	http://www.museuimperial.gov.br/dami/
	Museu de Belas Artes	http://mnba.gov.br/portal/colecoes/
Fundação Casa de Rui Barbosa	Repositório Rui Barbosa de Informações Culturais	http://rubi.casaruibarbosa.gov.br/
Funarte	Brasil Memória das Artes	http://www.funarte.gov.br/brasilmemoriadasartes/
	Acervo Sergio Britto Digital	http://sbrittod.funarte.gov.br/sophia_acervo/
	Acervo Funarte	http://cedoc.funarte.gov.br/sophia_web/
Biblioteca Nacional	Biblioteca Nacional Digital	https://bndigital.bn.gov.br/
	Hemeroteca Digital Brasileira	https://bndigital.bn.gov.br/hemeroteca-digital/

Fonte: Elaborado pelos autores (2018)³.

Os resultados constituem material essencial para uma investigação mais abrangente, voltada às possibilidades de agregação e interoperabilidade entre os repositórios vinculados ao MinC, pois permite esclarecer e delimitar o entendimento técnico dos repositórios, trazendo como resultado um problema mais esclarecido, passível de investigação mediante procedimentos mais sistematizados. Entende-se que compreendendo a estrutura desses acervos, pode-se avaliar melhor a possibilidade de reuso das informações disponibilizadas para o campo da pesquisa em HDs.

³ O Quadro 1 retrata a extração dos dados realizada em 2018, ano da realização desta pesquisa.

MAPEAMENTO TÉCNICO DOS ACERVOS DIGITAIS

Após análise dos acervos, obteve-se os dados dispostos no Quadro 2 e as considerações a respeito da forma de acesso a esses dados.

QUADRO 2. MAPEAMENTO TÉCNICO DOS ACERVOS DIGITAIS DOS PROJETOS MANTIDOS PELAS INSTITUIÇÕES VINCULADAS AO MINC

Projeto	Software - Licença	Tamanho do acervo	Estrutura das coleções	Formato dos arquivos	Padrão de metadados	Linguagem documentária	Coleta de dados	
							Suporte OAI-PMH?	Em uso?
Rede de Arquivos IPHAN	DSpace - Livre	38.022	Identificação, [...] e Regulamentação sobre preservação.	imagem	Dublin Core	Possui	Sim	Não
Museu Histórico Nacional	Tainacan - Livre	501	Acervo Museológico com subcoleções	imagem	Próprio	Tesouro de Acervo Museológico	Sim	Não
Museu Arqueológico de Itaipu	Tainacan - Livre	836	Acervo Museológico e Pessoas e Memórias	imagem e áudio	Próprio	Não disponível	Sim	Não
Museu Imperial	DSpace - Livre	8.777	Museologia, [...] e Artistas	documento e imagem	Dublin Core	Possui	Sim	Não
Museu Nacional de Belas Artes	Não informado	3.000	Pintura Brasileira, [...] e Novas Linguagens	imagem	Não informado	Não informado	Não informado	-
Repositório Rui Barbosa de Informações Culturais	DSpace - Livre	6.527	Arquivo histórico, [...] e Centro de Referência de Políticas Culturais.	documento e imagem	Dublin Core	CDD	Sim	Sim

continua

continuação

Projeto	Software - Licença	Tamanho do acervo	Estrutura das coleções	Formato dos arquivos	Padrão de metadados	Linguagem documentária	Coleta de dados	
							Suporte OAI-PMH?	Em uso?
Brasil Memória das Artes	Wordpress - Livre	4.874	Atores do Brasil, [...] <sup><sup> e Walter Pinto.	documento, imagem, áudio e vídeo	Próprio	Não disponível	Não	-
Acervo Sergio Britto Digital	SophiA - Proprietária	4.205	Comercial de TV, O Canto da Cotovia e Propaganda	imagem	Dublin Core	CDD	Não	-
Acervo Funarte	SophiA - Proprietária	134.600	Organizado pelos doadores dos fundos e vínculo institucional	arquivos não disponíveis	Marc e Dublin Core	CDD e Tesouro de assuntos	Sim	Não
Biblioteca Nacional Digital	SophiA - Proprietária	1.869.706	Abrão Koogan, [...] <sup><sup> e Wandenkolk	documento e imagem	Dublin Core	CDD	Sim	Não
Hemeroteca Digital Brasileira	SophiA - Proprietária	5.433	Periódicos	imagem	Dublin Core	CDD	Sim	Não

Fonte: Elaborado pelos autores (2018)⁴.

Os dados do Quadro 2 apresentam um cenário que reforça o problema central da pesquisa. Tem-se um total de 2.076.481 itens disponíveis nessas diferentes bases de dados, constituindo, em seu conjunto, o maior acervo de objetos culturais brasileiros digitalizados e sistematizados na internet que se tem conhecimento até o presente momento. Por si só, esse resultado demonstra o potencial que essas fontes possuem para estimular a pesquisa e o interesse no uso de metodologias e técnicas convergentes com as práticas das HDs.

⁴ O Quadro 2 retrata a extração dos dados realizada em 2018, ano da realização desta pesquisa.

Os *softwares* são diversificados, sendo que, dos 11 projetos, quatro utilizam o SophiA, dois o Tainacan, três o Dspace, um o Wordpress, e apenas um não informou, o que demonstra que 54% deles está adotando ao menos um *software* livre para a implementação de suas ações de disponibilização de acervos digitais na internet. Vale destacar que dos projetos que utilizam o Tainacan, apenas dois haviam sido lançados, e que os demais, apesar de acessíveis pela internet, ainda não disponibilizaram a documentação dos acervos.

Em relação ao formato dos arquivos, na maioria dos projetos não há essa informação de forma clara e direta, sendo necessário analisar os itens do repositório, de forma amostral. Assim, de maneira geral, são guardados arquivos nos formatos de imagem, extensões JPG e PNG; áudios em MP3; documentos em PDF e HTML, e vídeos em MP4.

O padrão de metadados mais encontrado foi o Dublin Core e o protocolo de coleta de dados o OAI-PMH. A informação não é transparente nas páginas, sendo necessárias mais pesquisas em outras fontes ou observação dos arquivos. Não foi possível identificar com facilidade os pontos de acesso OAI-PMH, com exceção da RUBI. Desse modo, estudos futuros são indispensáveis para confirmar a possibilidade de coleta e tratamento dos dados das demais iniciativas.

A linguagem documentária utilizada na maioria dos projetos é a Classificação Decimal de Dewey (CDD). Constata-se que o uso desta classificação em outros projetos se dá pelo uso do *software* SophiA, também direcionado a bibliotecas digitais.

Em suma, os projetos vinculados ao MinC apresentam heterogeneidade de soluções propostas para divulgação de acervos. O Quadro 2 demonstra que as instituições tomam as decisões de forma isolada, não havendo padrões bem definidos a serem adotados. A falta de articulação e padronização entre os projetos gera soluções com diferenças importantes do ponto de vista técnico, dificultando projetos e experimentos que tenham por objetivo, por exemplo, coletar os dados dessas instituições e gerar bases de dados unificadas para pesquisa e experimentação com métodos oriundos das Humanidades Digitais.

O ponto mais preocupante é a possibilidade de coleta de dados desses serviços. Apenas quatro dos 11 projetos menciona a possibilidade explícita de coleta. Além disso, não há a coleta de dados para além do uso do protocolo OAI-PMH, funcionalidade que não é tão simples e tecnicamente viável para muitos pesquisadores das ciências humanas. O que se pode concluir é que os projetos atendem aos usuários de internet, no sentido de acesso para visualização digital aos conteúdos, mas não estão bem resolvidos do ponto de vista da disponibilização de

dados abertos para coleta e reuso por pesquisadores que queiram desenvolver outros tipos de pesquisa. Esse aspecto dificulta em grande medida o uso dessa informação qualificada para pesquisas em Humanidades Digitais, o que constitui um desafio a ser enfrentado pela pesquisa e desenvolvimento de soluções para as instituições culturais.

RELATO DE EXPERIÊNCIA NO ACESSO AOS ACERVOS

Nesta seção, apresenta-se as experiências obtidas durante o acesso aos repositórios, destacando as facilidades e dificuldades encontradas. Os repositórios estão agrupados por instituição.

IPHAN

A tela principal da Rede de Arquivos Iphan (Figura 1) apresenta para buscas apenas um campo para inclusão de palavras-chaves, tornando complexa a interação de usuários leigos no conteúdo disponível, afinal, não há facetas ou qualquer recurso que colabore com a pesquisa. Logo abaixo, há um espaço destinado a “Bens diversos”, que apresenta itens contidos no repositório, mas que também não contribuem à formação de *strings* de busca.

FIGURA 1 - TELA PRINCIPAL DA REDE DE ARQUIVOS IPHAN



Fonte: Rede de Arquivos Iphan (2018).⁵

⁵ Disponível em: <<http://acervodigital.iphan.gov.br/xmlui/>>. Acesso em: 02 out. 2018.

Caso o usuário opte pelo botão “Navegar”, a interface é mais completa, com facetas e filtros de ordenação de resultados, além da busca por palavras-chaves, que se mantém. A Figura 2 apresenta esta tela.

FIGURA 2 - TELA DE NAVEGAÇÃO DA REDE DE ARQUIVOS IPHAN



Fonte: Rede de Arquivos Iphan (2018).⁶

Nesta tela, foi possível coletar várias das informações contidas no Quadro 2, tais como: tamanho do acervo; estrutura das coleções, para a qual considerou-se os itens contidos no menu Assunto; formato dos arquivos, que embora classificado por fotografias, textual, imagens e mapas/plantas, são arquivos em formato de imagem; e o padrão de metadados, analisado por meio das informações apresentados em cada item. No entanto, não há informações sobre a linguagem documentária e o protocolo para coleta de dados.

IBRAM

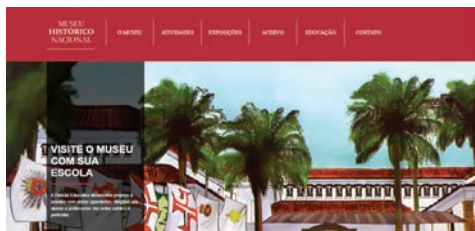
Conforme supracitado, apenas seis, dos 30 museus vinculados ao IBRAM, possuem seu acervo digitalizado e disponibilizado na Web. Para tal, foram escolhidos *softwares* gestores de repositórios digitais diferentes, o Tainacan e o DSpace.

⁶ Disponível em: <<http://acervodigital.iphan.gov.br/xmlui/discover>>. Acesso em: 06 out. 2018.

O Tainacan é fruto da parceria entre o Laboratório de Políticas Públicas Participativas do MediaLab/UFG com o MinC e o Instituto Brasileiro de Museus em torno do desenvolvimento de uma plataforma comum para a produção e organização de acervos digitais em rede. Dessa forma, o *software* Tainacan está em pleno processo de desenvolvimento e vem sendo implantado em alguns museus do IBRAM, o que explica alguns projetos ainda em andamento e funcionalidades técnicas ainda indisponíveis.

Destaca-se que o Tainacan oferece, além do repositório, uma página institucional aos museus. Assim, as Figuras 3 e 4 são as telas principais do Museu Histórico Nacional e do Museu de Arqueologia de Itaipu.

FIGURA 3 - TELA INICIAL DO MHN



Fonte: Museu Histórico Nacional (2018).^{<?>}

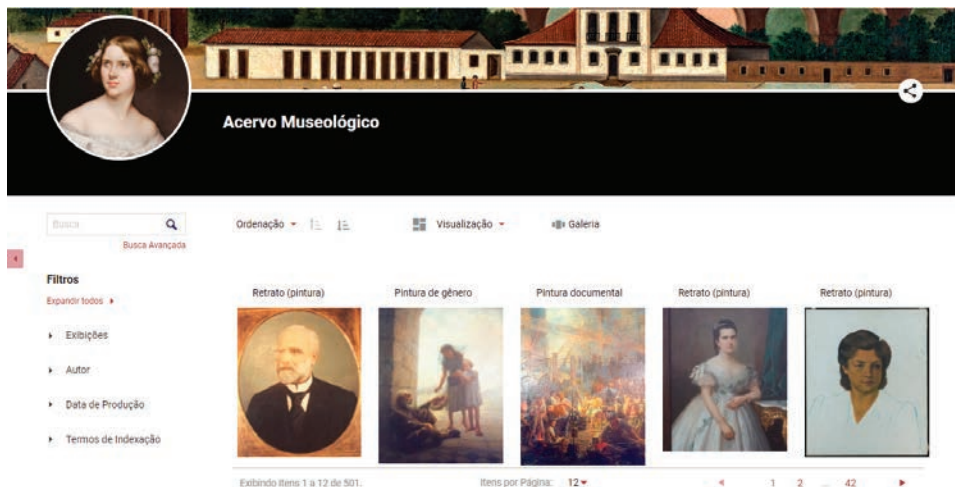
FIGURA 4 - TELA INICIAL DO MUSEU DE ARQUEOLOGIA



Fonte: Museu de Arqueologia de Itaipu (2018).^{<?>}

Por se tratar do mesmo *software*, optou-se por ilustrar apenas a tela do Museu Histórico Nacional (Figura 5).

FIGURA 5 - TELA INICIAL DO ACERVO DO MUSEU HISTÓRICO NACIONAL



Fonte: Museu Histórico Nacional (2018).⁷

A página apresenta busca por palavras-chaves e um *link* que direciona à busca avançada, além de facetras que facilitam o acesso à informação, em uma interface amigável, que permite o compartilhamento de informação via redes sociais.

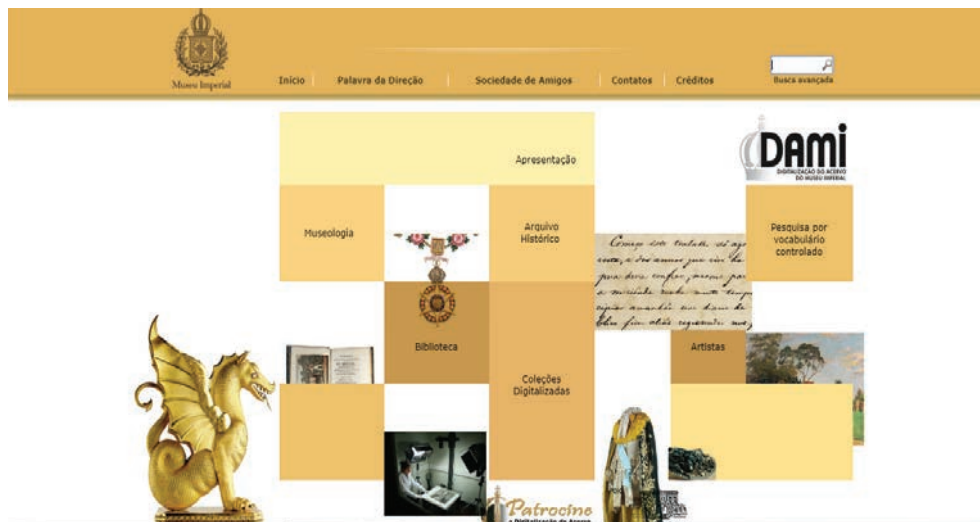
Em relação aos itens referentes ao mapeamento técnico, foi simples encontrar o tamanho do acervo e identificar o padrão de metadados. Informações tais como linguagem documentária e suporte ao protocolo OAI-PMH foram encontradas a partir de pesquisas em outras fontes.

Em processos de digitalização de acervos anteriores a parceria que resultou no Tainacan, há o Museu Imperial, que utiliza o DSpace, e o Museu de Belas Artes, que utiliza site próprio, não usando uma ferramenta específica para repositório digital.

O site do Museu Imperial apresenta em sua página principal um menu com todas as coleções, nesta pesquisa, consideramos apenas “Coleções Digitalizadas”, um campo de busca e um *link* para a busca avançada, além de ter como diferencial, a pesquisa por meio de um vocabulário controlado, conforme mostra a Figura 6.

⁷ Disponível em: <http://mhn.acervos.museus.gov.br/reserva-tecnica/#/?view_mode=masonry&perpage=12&paged=3&order=DESC&orderby=date&fetch_only%5B0%5D=thumbnail&fetch_only%5B1%5D=creation_date&fetch_only%5B2%5D=author_name&fetch_only%5B3%5D=title&fetch_only%5B4%5D=description>. Acesso em: 17 dez. 2018.

FIGURA 6 - TELA INICIAL DO MUSEU IMPERIAL



Fonte: Museu Imperial (2018).⁸

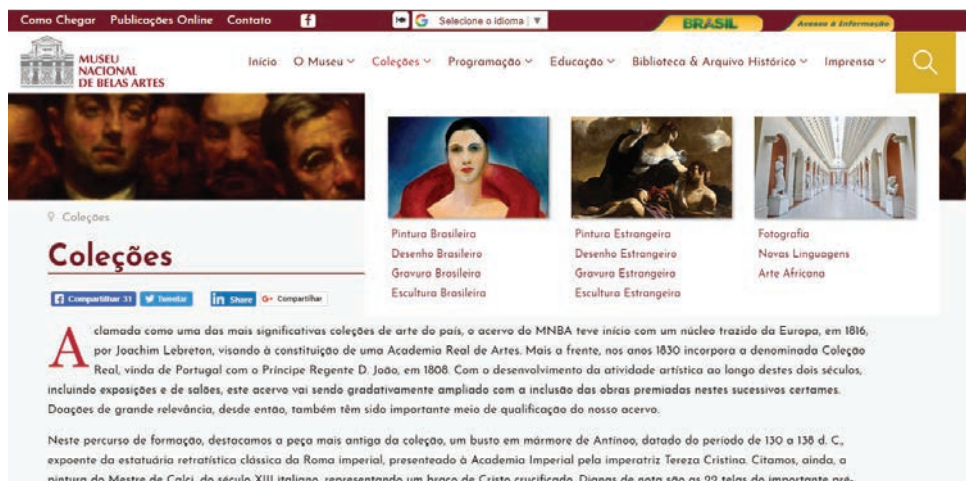
O acesso a busca ou busca avançada seguem o padrão do DSpace e, por isso, permite busca por diferentes filtros e facetas, além de recursos de ordenação. Ao acessar algum arquivo, são apresentados os metadados e o *link* para abrir o arquivo, de fato.

Das informações necessárias ao mapeamento, o tamanho do acervo foi calculado a partir da quantidade de arquivos de cada coleção, o formato dos arquivos foi definido por amostragem e a existência de linguagem documentária constada via página de pesquisa por vocabulário controlado.

A página principal do Museu Nacional de Belas Artes (Figura 7) apresenta, além das coleções do acervo, uma série de informações institucionais. Há um campo para busca por palavra-chave, mas é essencial que o usuário tenha conhecimento do que está buscando.

⁸ Disponível em: <<http://www.museuimperial.gov.br/dami/>>. Acesso em: 08 nov. 2018.

FIGURA 7 - TELA INICIAL DO MUSEU NACIONAL DE BELAS ARTES



Fonte: Museu Nacional de Belas Artes (2018).⁹

Quando uma coleção é acessada, há um breve resumo sobre ela e os itens são apresentados individualmente, sendo a navegação entre eles feita um a um. A cada item são apresentados um conjunto simplificado de metadados. As coleções e itens podem ser compartilhados em redes sociais.

Não é possível conhecer o tamanho do acervo, no entanto, em alguns dos resumos das coleções há um número aproximado de itens digitalizados, com valores como 3.000 e 7.000 itens por coleção. Com esta informação, é possível inferir o quão grande é o acervo disponibilizado, o que torna a interface de navegação entre os itens ainda mais ineficiente.

Não foi localizado o *software* gestor do repositório, nem tampouco o padrão de metadados, a linguagem documentária e a forma para coleta dos dados.

FUNDAÇÃO CASA DE RUI BARBOSA

A tela principal do Repositório Rui Barbosa de Informações Culturais (RUBI), Figura 8, apresenta informações diversas, tal como uma página institucional. Na mesma interface, estão: o campo para busca no repositório, por

⁹ Disponível em: <<http://mnba.gov.br/portal/colecoes/>>. Acesso em: 08 nov. 2018.

meio de palavras-chaves; links para acesso as facetas (autor, assunto, título e data) e links para acesso as coleções (Arquivo Histórico, Arquivo Museu da Literatura Brasileira, Bibliotecas, Museu Casa de Rui Barbosa e Centro de Referência de Políticas Culturais), que quando acessadas, listam suas subcoleções, junto a novas facetas. Esta organização dos dados facilita o acesso a informação para usuários que desconhecem ou têm pouca experiência em relação ao conjunto de dados do acervo.

FIGURA 8 - TELA PRINCIPAL DO RUBI



Fonte: RUBI (2018).¹⁰

Boa parte dos dados pertinentes a esta pesquisa não foram fáceis de localizar, com exceção ao *software* gestor do repositório e a estrutura das coleções, ambos apresentados na página principal. O tamanho do acervo foi obtido por meio do acesso a faceta títulos. O formato dos arquivos foi obtido por amostragem, a partir dos arquivos disponíveis. O padrão de metadados, a linguagem documentária e o protocolo utilizado para coleta dos dados não foram encontrados, no entanto, foram informadas por e-mail (LESSA, 2018), após solicitação. Foi nos relatado que, o padrão de metadados utilizados é o Dublin Core e que utilizam o mesmo padrão do vocabulário controlado utilizado na biblioteca, disponível na base de dados referenciais <acervos.casaruibarbosa.gov.br>. Questionados sobre o uso de dois repositórios, informaram que a principal diferença entre eles, é que o Sophia não fornecemos acesso ao objeto digital, é apenas referencial.

¹⁰ Disponível em: <<http://rubi.casaruibarbosa.gov.br/>>. Acesso em: 03 out. 2018.

De todos os repositórios testados, apenas o RUBI, por meio do DSpace, oferece o acesso aberto aos dados via protocolo OAI-PMH.

FUNARTE

O Brasil Memória das Artes utiliza o WordPress com tema próprio, cuja página principal (Figura 9) disponibiliza pesquisa por palavras-chaves, a estrutura das coleções, e no rodapé da página, de forma bem discreta, algumas facetas – imagens, áudio, vídeos e textos. A busca por palavra-chave oferece o recurso autocompletar formulado à medida que se escreve as palavras, facilitando a busca. Quando opta-se pelo acesso às coleções, os resultados são classificados pelo formato do arquivo, tais como textos, imagens etc.

FIGURA 9 - TELA INICIAL DO BRASIL MEMÓRIA DAS ARTES



Fonte: Brasil Memória das Artes (2018).¹¹

No que diz respeito ao mapeamento técnico, as informações foram difíceis de serem encontradas. A primeira, *software*, a página mostra o uso do WordPress, mas não do tema, descoberto por meio da tecla F12, que apresenta informações

¹¹ Disponível em: <<http://www.funarte.gov.br/brasilmemoriadasartes/>>. Acesso em: 03 out. 2018.

referente a página. O tamanho do acervo foi calculado manualmente a partir do somatório dos itens das coleções. As coleções são ordenadas pelo formato dos arquivos, logo, o item “formato dos arquivos” foi facilmente identificado. O padrão de metadados, a linguagem documentária e o protocolo para coleta dos dados não foram localizados.

O Acervo Sergio Britto não possui interface amigável. A tela principal (Figura 10) possui seis campos para palavras-chaves e dois para tipos de categorização de dados. Um usuário que desconheça o acervo dificilmente conseguirá formular as palavras-chaves necessárias para a localização dos arquivos.

FIGURA 10 - TELA INICIAL DO ACERVO SERGIO BRITTO DIGITAL



Fonte: SophiA Acervo (2018).¹²

Para fins de testes, simulou-se a pesquisa com a palavra-chave “sergio”, mantendo os demais campos vazios, com exceção dos campos selecionáveis, nos quais manteve-se “Todos”. Os resultados são agrupados em *tags* e objetos, em uma

¹² SophiA Acervo. Disponível em: <http://sbrittod.funarte.gov.br/sophia_acervo/>. Acesso em: 03 out 2018.

página intitulada de Resumo, que permite novas pesquisas por palavras-chave . Em ambos os casos, apresenta-se a listagem dos itens, que devem ser selecionados manualmente pelo usuário. Na simulação supracitada, o objeto Hemeroteca, por exemplo, resultou em 1.332 itens. Observou-se então, que, a menos que o usuário esteja munido de palavras-chaves direcionadas, será muito difícil encontrar o objeto de interesse.

Dos dados necessários à pesquisa, apenas o *software* foi localizado. Não há a possibilidade de acessar todo acervo em uma única pesquisa, logo, é inviável conhecer o tamanho do acervo facilmente. Sobre a estrutura das coleções, considerou-se as *tags*, disponibilizadas para pesquisa. O formato dos arquivos podem ser encontrados a partir do acesso aos documentos, mas não há a informação clara. Sobre a linguagem documentária, o SophiA é uma solução direcionada a bibliotecas, desta forma, infere-se que a linguagem documentária seja o CDD. Em contato via e-mail com a empresa Prima, responsável pelo SophiA, nos informaram que: “no caso dos acervos do Sergio Britto, ele trabalha com Sophia Acervo e ele não contempla integração via OAI” (BRAZ, 2018).

O Acervo Funarte também utiliza o SophiA, no entanto, com o produto SophiA Biblioteca e, por isso, conta com um leiaute mais empolgante que o do Acervo Sergio Britto Digital. A página principal (Figura 11) é composta tanto pela busca rápida, por meio de palavras-chaves, quanto pela busca avançada, intitulada combinada, que permite ao usuário definir diferentes *strings* para pesquisa. Ela também oferece de maneira rápida e intuitiva, a seleção de uma ou várias coleções, no entanto, para conclusão da busca é necessário que o usuário digite o mínimo de três caracteres no campo destinado às inserções das palavras-chaves. Assim, caso o usuário desconheça o conteúdo da coleção terá uma pequena dificuldade em acessar o itens do acervo.

FIGURA 11 - TELA PRINCIPAL DO ACERVO FUNARTE



Fonte: Acervo Funarte (2018).¹³

Na parte inferior da página estão disponíveis as últimas aquisições, que provavelmente referem-se as inserções mais recentes do repositório, e somente a partir do *link* “ver mais” é possível ter acesso a alguns itens sem a necessidade de uma pesquisa prévia. Na tela “ver mais” é possível selecionar itens por meio de facetar.

Em relação à pesquisa, foram encontrados com facilidade o *software* e as coleções do acervo. Não é possível identificar o formato do arquivo, pois não há informação nos metadados e nem tampouco direcionamento aos objetos, uma vez que eles estão apenas referenciados no Acervo.

Por utilizar o SophiA, constatou-se que o Acervo Funarte utiliza o padrão de metadados MARC e Dublin Core, disponível para consulta na interface. Sobre a linguagem documentária, o SophiA é uma solução direcionada a bibliotecas, desta forma, infere-se que a linguagem documentária seja a CDD. Mas no acervo, não há essa informação.

¹³ Disponível em: <http://cedoc.funarte.gov.br/sophia_web/>. Acesso em: 03 out. 2018

BIBLIOTECA NACIONAL

A Biblioteca Nacional possui a BNDigital e a Hemeroteca Digital. Diferentemente dos projetos supracitados, que possuem páginas próprias, nesse caso, ambos os projetos coexistem na mesma página. Assim, a página principal da BNDigital (Figura 12) oferece a busca rápida no acervo, por palavras-chaves, e a busca avançada no acervo e na Hemeroteca. Por utilizar o SophiA, a BNDigital apresenta características semelhantes aos acervos que a utilizam, mencionados anteriormente. A página conta com “Últimas aquisições” e com o *link* “ver mais”, que dá acesso a itens do repositório e facetas.

FIGURA 12 - TELA PRINCIPAL DA BIBLIOTECA NACIONAL DIGITAL

Busca rápida no acervo digital

BUSCA AVANÇADA NO ACERVO DIGITAL BUSCA AVANÇADA NA HEMEROTECA

ARTIGOS DOSSIÊS EXPOSIÇÕES **ACERVO DIGITAL** HEMEROTECA DIGITAL [+ SOBRE A BNDIGITAL](#)

Página inicial > ACERVO DIGITAL

ACERVO DIGITAL

Home Pesquisa Minha seleção

Ajuda | Acessibilidade | Alto contraste

Busca rápida Busca combinada

Todos os campos [] E [] Ano edição [] a [] Buscar

Título [] E [] Coleção Qualquer [] Limpar

Autor [] E [] Acervo Qualquer []

Assunto [] Material Qualquer []

Últimas aquisições igual a [] Idioma Qualquer []

Ordenação Título - crescente []

Registros com conteúdo digital

Últimas aquisições

6. 7. 8. 9. 10.

[+ Ver mais](#)

Fonte: BNDigital (2020), adaptada.¹⁴

¹⁴ Disponível em: <<http://bndigital.bn.gov.br/acervodigital/>>. Acesso em: 28 maio 2020.

Os dados da pesquisa foram encontrados com facilidade: o *software*, as coleções do acervo e o tamanho do acervo. O formato dos arquivos foi retirado por amostragem, verificando os arquivos. Constatou-se que a BNDigital utilize o padrão de metadados Dublin Core. De acordo com a empresa responsável pelo SophiA (BRAZ, 2018), o *software* permite a integração via OAI-PMH, no entanto, em módulo separado que precisa ser adquirido. Sobre a linguagem documentária, infere-se, assim como os demais, que a linguagem seja a CDD.

CONSIDERAÇÕES FINAIS

Os acervos criados e mantidos pelas instituições vinculadas ao MinC são extremamente relevantes para sociedade, pois diminuem as distâncias entre o público e o rico material cultural disponibilizado. No entanto, seu crescimento e a nova demanda da sociedade exige que, além de disponibilizar, haja também ferramentas de buscas eficientes, que retornem com facilidade objetos de interesse e, muito além disso, ferramentas que permitam a interoperabilidade desses acervos, tornando a pesquisa transparente ao usuário.

Para isso, uma etapa necessária é a realização do levantamento técnico dos acervos, para conhecer a realidade atual deles. O resultado desta pesquisa deixou nítida a heterogeneidade de soluções, adquiridas a partir da composição de cada projeto, de suas particularidades, sem que houvesse, aparentemente, a necessidade que eles “conversassem”. Também percebe-se que informações pertinentes aos acervos, como sua linguagem documentária e seu padrão de metadados e protocolos para coleta de dados, não são claramente explicitada nas páginas.

Do ponto de vista do uso dessas informações para aplicações em HDs, constata-se que há pouca possibilidade técnica de acessar os dados sistematizados em formatos abertos e de fácil acesso. Vale frisar que esses acervos representam informações culturais altamente qualificadas e curadas ao longo de gerações pela sociedade brasileira, logo representando bases de dados de enorme potencial para aplicações computacionais e seu uso para pesquisas e desenvolvimentos no campo das HDs.

É fundamental que as iniciativas possam repensar seu plano de desenvolvimento na incorporação de funcionalidades técnicas que facilitem o

acesso aos dados para manipulação por outros pesquisadores e usuários. É mais do que necessário que uma política para a produção e manutenção de acervos digitais seja pensada e implementada no país nos próximos anos, visando fornecer condições estruturantes para o melhor reuso dessas informações valiosas da sociedade brasileira em tempos de disseminação ampla de novas técnicas, métodos e processos de pesquisa permeados pela cultura digital.

REFERÊNCIAS

ANCINE. Agência Nacional do Cinema. *Apresentação*. 2018. Disponível em: <<https://www.ancine.gov.br/pt-br/ancine/apresentacao>>. Acesso em: 18 set. 2018.

BAGGIO, C. C.; FLORES, D. Documentos digitais: preservação e estratégias. *Biblos*, Rio Grande, v. 27, n. 1, p. 11-24, 2013. Disponível em: <<https://periodicos.furg.br/biblos/article/view/2654/2395>>. Acesso em: 10 nov. 2018.

BIBLIOTECA NACIONAL. *Histórico*. 2018. Disponível em: <<https://www.bn.gov.br/sobre-bn/historico>>. Acesso em: 18 set. 2018.

BNDigital. Biblioteca Nacional Digital. *Hemeroteca Digital Brasileira*. 2018. Disponível em: <<http://bndigital.bn.gov.br/hemeroteca-digital/>>. Acesso em: 18 set. 2018.

BRASIL. Ministério da Cultura. *Competências*. 2013a. Disponível em: <<http://www.cultura.gov.br/competencias>>. Acesso em: 17 dez. 2018.

BRASIL. Ministério da Cultura. *Histórico*. 2013b. Disponível em: <<http://www.cultura.gov.br/historico>>. Acesso em: 13 nov. 2018.

BRASIL. Ministério da Cultura. *Instituições vinculadas*. Brasil: Governo Federal. 2013c. Disponível em: <<http://www.cultura.gov.br/instituicoes-vinculadas>>. Acesso em: 18 set. 2018.

BRASIL. Ministério da Cultura. *Institucional*. Brasil: Governo Federal. 2013d. Disponível em: <<http://www.cultura.gov.br/institucional>>. Acesso em: 18 set. 2018.

BRAZ, R. Suporte. Destinatário: Joyce Siqueira. [S. l.], 10 dez. 2018. 1 mensagem eletrônica.

COMITÊ GESTOR DA INTERNET NO BRASIL. *Tic cultura 2016 = Survey on the use of information and communication technologies in brazilian cultural facilities: Ict in culture 2016*. São Paulo: Núcleo de Informação e Coordenação do Ponto BR, 2017.

FUNDAÇÃO CASA DE RUI BARBOSA. *Sobre a Fundação*. 2018a. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=1>. Acesso em 19 set 2018.

FUNDAÇÃO CASA DE RUI BARBOSA. *Histórico*. 2018b. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=73>. Acesso em: 19 set. 2018.

FUNDAÇÃO CASA DE RUI BARBOSA. *Missão*. 2018c. Disponível em: <http://www.casaruibarbosa.gov.br/interna.php?ID_S=10>. Acesso em: 19 set. 2018.

FUNDAÇÃO CULTURAL PALMARES. *Apresentação*. 2016. Disponível em: <http://www.palmares.gov.br/?page_id=95>. Acesso em: 18 set. 2018.

FUNARTE. *Sobre a Instituição*. 2010. Disponível em: <<http://www.funarte.gov.br/a-funarte/>>. Acesso em: 18 set. 2018.

FUNARTE. *Brasil Memória das Artes*. 2013. Disponível em: <<http://www.funarte.gov.br/brasilmemoriadasartes/acervo/o-projeto>>. Acesso em: 18 set. 2018.

FUNARTE. *Funarte Digital: novo portal a caminho da implementação*. 2018. Disponível em: <<http://www.funarte.gov.br/artes-integradadas/funarte-digital-novo-portal-a-caminho-da-implementacao/>>. Acesso em: 16 dez. 2018.

IBICT. *Sistema para Construção de Repositórios Institucionais Digitais (DSpace)*. 2012. Disponível em: <<http://www.ibict.br/pesquisa-desenvolvimento-tecnologico-e-inovacao/Sistema-para-Construcao-de-Repositorios-Institucionais-Digitais/apresentacao>>. Acesso em: 18 set. 2018.

IBRAM. *Estrutura*. 2018a. Disponível em: <<http://www.museus.gov.br/acessoainformacao/o-ibram/estrutura/>>. Acesso em: 18 set. 2018.

IBRAM. *O Ibram*. 2018b. Disponível em: <<http://www.museus.gov.br/acessoainformacao/o-ibram/>>. Acesso em: 18 set. 2018.

IPHAN. *Rede de Arquivos Iphan*. 2014a. Disponível em: <<http://portal.iphan.gov.br/pagina/detalhes/1223>>. Acesso em: 20 set. 2018.

IPHAN. *O Iphan*. 2014b. Disponível em: <<http://portal.iphan.gov.br/pagina/detalhes/872>>. Acesso em: 18 set. 2018.

LESSA, L. Informações. Destinatário: Joyce Siqueira. [S. l.], 02 out. 2018. 1 mensagem eletrônica.

MUSEU HISTÓRICO NACIONAL. *Museu Histórico Nacional disponibiliza online 500 obras da sua coleção de pintura*. 2018. Disponível em: <<http://mhn.museus.gov.br/index.php/museu-historico-nacional-disponibiliza-online-500-pinturas-do-acervo/>>. Acesso em: 16 dez. 2018.

PUNTONI, P. Rede Memorial: cultura digital, redes colaborativas e a digitalização dos acervos memoriais no Brasil. In: GOBEL, Barbara, CHICOTE, Gloria (Org.). *Transiciones inciertas: archivos, conocimientos y transformación digital en América Latina*. La Plata: Universidad Nacional de la Plata, Faculdade de Humanidades y Ciencias de la Educación; Berlin: Ibero-Amerikanisches Institut, 2017.

RUBI. *RUBI Repositório Institucional e Temático da FCRB*. 2016. Disponível em: <<http://rubi.casarui Barbosa.gov.br/sobre.html>>. Acesso em: 19 set. 2018.

SHINTAKU, M. Tecnologias para Gestão da Informação. In: VECHIATO, F. *et al.* (Org.). *Repositórios digitais: teoria e prática*. Curitiba: EDUTFPR, 2017. p. 65-89. Disponível em: <<http://repositorio.utfpr.edu.br/jspui/>>. Acesso em: 13 nov. 2018.

O processo de representação de ativos de informação: Web Semântica e Web de Dados no contexto das Humanidades Digitais

Cláudio José Silva Ribeiro

INTRODUÇÃO

De forma global, é possível afirmar que a sociedade contemporânea vem procurando repensar as maneiras de conduzir as atividades laborais, bem como sua relação com a produção de dados e informação geradas como fruto destas atividades. Esta sociedade também vem sendo impulsionada pelo uso das tecnologias disruptivas¹, que causam um deslocamento dos objetos do cotidiano na direção do digital. Por outro lado, a cultura passou a ser influenciada pelo *cultural hacking*², que pressupõe pequenas mudanças apoiadas em artefatos digitais. O que poderia ser entendido como um contrassenso acaba sendo um fator importante para o sucesso de ações para a mudança, na medida em que favorece o entrelaçamento de ações culturais com o ambiente digital.

É nesse sentido que surge a abordagem das Humanidades Digitais (HDs), pois estas procuram explicitar as interseções entre cultura e comportamento, além da própria história, em conjunto com os artefatos computacionais presentes no mundo digital.

Essa abordagem repousa sobre um conjunto de esforços desenvolvidos desde os anos 1950, com a reunião das unidades de informação residentes em bibliotecas e arquivos para sua representação em ambientes eletrônicos. O tratamento do amplo

¹ A noção de Tecnologia disruptiva está ligada às inovações tecnológicas que derivam de uma mudança de paradigma. Esta mudança não é fruto de melhoria contínua ou evolução (WIKIPEDIA, 2016).

² A noção de *cultural hacking* pressupõe a identificação de áreas vulneráveis em sistemas socioculturais e nas organizações, de forma que seja possível promover ações de mudança de cultura organizacional (MOORE, 2018).

volume de informações hospedadas nos centros de documentação, com o uso de processamento usando linguagens textuais, a análise e criação de métricas sobre frequência de termos, dentre outros temas, eram investigações que despontavam no cenário das universidades e centros de pesquisa da época. (BURDICK *et al.*, 2012)

Já no campo da arte, esforços encaminhados na década de 1980, mais precisamente em 1981 na Bienal de SP, podem ser considerados como um dos primeiros exemplos para ir além das formas tradicionais de exposição. A tentativa de construir um movimento de arte postal, reunindo contribuições, colaborações, produção e difusão em uma grande rede pode ser considerado como um passo preparatório na direção das HDs. (BEIGUILMAN; MAGALHÃES, 2014).

Por outro lado, a década de 1980 também promovia outras transformações. O início do movimento da computação pessoal impulsionou ainda mais as investigações, pois foi possível dotar a dinâmica de pesquisa com instrumentos que agilizavam o processamento de múltiplos dados e fontes de informação, liberando os investigadores da dependência dos grandes centros de processamento de dados. Este movimento se tornou mais intenso com a visão proposta por Bernes-Lee nos anos 1990. A chegada do ambiente da Internet e da *World Wide Web* aproximou em definitivo o ambiente computacional do cotidiano da sociedade e da pesquisa aplicada.

Como consequência desses movimentos, verifica-se um deslocamento no conteúdo dos currículos de formação de profissionais nas Ciências Sociais e Humanas das Instituições de Educação Superior (IES), em relação à evolução dos assuntos ligados às temáticas de: preservação; uso; e disseminação de informação; todas abordagens apoiadas na representação digital (LOURENÇO, 2017). Ramalho (2017) reforça este entendimento em um relato no qual é possível averiguar que o processo de representação da informação está imbricado com o uso de aparatos tecnológicos.

Observa-se então que a sociedade está tomando a direção do mundo digital de forma plena, pois estes aparatos estão cada vez mais presentes no seu cotidiano. Esta percepção catalisou os processos de transformação dos objetos de estudo, visto que estes aparatos surgiram como mediadores entre o humano e o digital. Assim, os pesquisadores ligados às Ciências Sociais e Humanas buscaram debater sobre esta mediação e foi no bojo destas discussões que foram formuladas indagações que nos trazem de volta às questões das HDs: com a opção pelo mundo digital, como poderá

ser endereçada a abordagem sobre o contato físico com o objeto? A representação digital dará conta disso? Como esta representação digital poderá interagir com o real? Como tratar o licenciamento e o reuso de dados e da informação? E nos aspectos da preservação, como tratar as diferentes formas e caminhos para auxiliar na preservação de objetos digitais a longo prazo?

O debate sobre a guarda de objetos físicos, materializados em um suporte e de forma tangível, está sendo incrementado para dar lugar às discussões sobre metadados para os documentos digitais e as novas formas de representar os registros. Pode-se deduzir que as afirmações de Medeiros *et al.* (2017) convalidam estas percepções quando observam:

Ao surgir como campo interdisciplinar, as Humanidades Digitais propõem a reflexão sobre as práticas sugeridas pela introdução da tecnologia digital no âmbito das unidades de informação e cultura. Elas vão além da preocupação do uso de ferramentas tecnológicas empregadas às humanidades, geram questões filosóficas próprias relacionadas a apropriação das tecnologias aliadas aos processos de disseminação, acesso, recuperação da informação e criação de novos conhecimentos. (MEDEIROS *et al.*, 2017, *on-line*)

A MOTIVAÇÃO DAS CIÊNCIAS SOCIAIS QUE FOI IMPULSIONADA NA DIREÇÃO DO DIGITAL

Como mencionado anteriormente, pode-se inferir que a sociedade fez a opção de caminhar na direção do digital. Segundo o DH-Manifesto (2010):

[..] 2. Para nós, as *digital humanities* referem-se ao conjunto das Ciências humanas e sociais, às Artes e às Letras. As humanas digitais (sic) não negam o passado, apoiam-se, pelo contrário, no conjunto dos paradigmas, *savoir-faire* e conhecimentos próprios dessas disciplinas, mobilizando simultaneamente os instrumentos e as perspectivas singulares do mundo digital.

3. As *digital humanities* designam uma transdisciplina, portadora dos métodos, dos dispositivos e das perspectivas heurísticas ligadas ao digital no domínio das Ciências humanas e sociais. (DH-MANIFESTO, 2010, *on-line*)

Portanto, cercados por Documentos e Unidades Documentárias, os investigadores em Ciências Sociais precisaram se movimentar e estender o

entendimento sobre objetos físicos e tangíveis na direção do intangível e do virtual. Para além das características do suporte da Unidade Documentária, outras propriedades precisavam ser identificadas para descrever esses novos objetos (RIBEIRO, 2017).

Esta intrincada relação entre as práticas tradicionais de registro do conhecimento e as novas tecnologias marca o movimento das HDs. As HDs incorporam os métodos e as questões legadas pelas Ciências Sociais e Humanas, ao mesmo tempo que mobilizam as ferramentas e perspectivas abertas pela tecnologia digital (DH-MANIFESTO, 2010). Não se trata de uma única área, e nem mesmo de uma nova área do conhecimento. A interdisciplinaridade é um traço característico dos seus projetos, sendo comum a presença de historiadores, linguistas, bibliotecários, matemáticos e cientistas da computação na mesma equipe de trabalho. Por essa razão, as iniciativas podem dar origem a temas bastante distintos, que vão desde o desenvolvimento de novas técnicas, métodos e ferramentas até a entrega de produtos e serviços relacionados a políticas públicas de urbanização; a reflexão sobre ética envolvendo a privacidade e o uso de redes sociais, a manifestação da realidade nas performances midiáticas, na moda, na história e na literatura.

É nessa perspectiva que as Ciências Sociais e Humanas vêm discutindo não apenas a disseminação e a criação de novas obras, mas também as novas formas de expressão e manifestação. O tratamento de múltiplas fontes textuais, o processo de cartografia na Web, o uso de sistemas de informações geográficas, as atividades para *data mining* (seleção e garimpagem de dados) e a sonificação de dados (áudio e sons) convergem para a noção de Humanidades Digitais.

A explosão informacional (Figura 1), consequência das múltiplas fontes e diversos formatos, trouxe mais alguns reflexos no processo de representação do ente digital. A entrada de outros formalismos para representação de coleções (SANTOS NETO *et al.*, 2013; PELLEGRINO *et al.*, 2017), em especial para informações publicadas no ambiente Web, permitiu que novas discussões fossem formuladas. Com isto houve a criação de laços que promoveram a aproximação do mundo digital e das Ciências Sociais e Humanas.

FIGURA 1 - Explosão informacional



Fonte: O autor (2018).

Para dar conta desta explosão surgiram diferentes projetos ligados a temas como: *Big Data*, para tratar redes sociais e bancos de dados, trabalhando com grandes volumes (MALINI; CIARELLI; MEDEIROS, 2017; BARBOSA; KOBASHI, 2017); novos papéis, comportamentos e perfil profissional (RIBEIRO, 2014; ROZSA; DUTRA; NHACUONGUE, 2017); aplicativos para museus e bibliotecas (GODOY VIERA; VARVAKIS; FORESTI, 2018; KREBS; ROCHA; RIBEIRO, 2017; SANTOS PUTON; MANGAN, 2015); além de outras formas de interação entre coleções e a história dentro desta nova realidade, como a solução A Voz da Arte³, trazendo reflexões para os pesquisadores nas Ciências Sociais e Humanas.

Por outro lado, o cotidiano está repleto de informação. A noção de Internet das Coisas (IoT)⁴ incorporou na nossa vida as interações dos objetos do lar, acarretando alterações na forma de lidar com os diferentes aparatos na vida contemporânea. São equipamentos diversos (fornos de micro-ondas, geladeiras, aparelhos de som,

³ O aplicativo A Voz da Arte mostra uma outra forma de interagir das pessoas com as obras. Ele procura incorporar artefatos e objetos tecnológicos, informações correlacionadas e hiperligações. Disponível em: < <https://www.youtube.com/watch?reload=9&v=1rOAgvCnZpw> />. Acesso em: 24 maio 2020.

⁴ Acrônimo em inglês para Internet of Things – Internet das coisas – aparatos tecnológicos dotados de tecnologia embarcada que podem interagir na rede.

TVs etc.) que ligam, desligam, aquecem e preparam alimentos, resfriam bebidas e apresentam listas de compras, tudo por comandos de um aparelho celular. A IoT também promoveu os aplicativos que facilitam o deslocamento de pessoas, chamando táxis, precificando corridas, ajudando no compartilhamento de veículos e racionalizando os meios de transporte, ligando e desligando carros, monitorando, fotografando e registrando muitas ações desenvolvidas por nós em sociedade.

Diante deste cenário, surgem as nossas muitas inquietações para representar dados e informação em diferentes contextos. A questão de partida que motiva este debate passa pela indagação sobre as possibilidades de interpretação automática, ou não, dos signos e significados, nos contextos semântico, sintático e pragmático.

O PROCESSO DE REPRESENTAÇÃO DESTES SIGNOS E SEUS SIGNIFICADOS

A multiplicidade de documentos e unidades documentais no cotidiano (nomeados neste relato de ativos de informação⁵) impulsionaram a necessidade de representar diferentes tipos de objetos. Estes ativos podem ser originados a partir de diferentes coleções, como por exemplo: documentos pessoais, documentos decorrentes da prestação de serviços ao cidadão, objetos originários de projetos nomeados de Galleries, Libraries, Archives and Museums (GLAM) e Linked Open Data in Libraries, Archives and Museums (LODLAM), ou ainda fruto de investigações e projetos de pesquisa. Cabe destacar algumas características que devem ser observadas no âmbito destas coleções:

- Os documentos pessoais e de prestação de serviços podem ser considerados como de simples representação, pois possuem características, atributos e propriedades conhecidos. Em geral, metadados para descrever elementos, tais como autoria, localização e datas, são suficientes para descrever estes ativos de informação.
- Em projetos GLAM e LODLAM, as instituições de memória necessitam trabalhar com obras de arte tanto clássicas quanto contemporâneas. Nestes projetos, a descrição das obras transcende o contexto atual, pois a necessidade de correlacionar textos e informações descritivas das obras com

⁵ Esta noção transmite que o entendimento de uma informação deve ser considerado como elemento-chave de uma instituição, um ativo, tal qual bens, propriedades ou ativos financeiros.

outras fontes de dados na Web permite que o indivíduo interessado possa utilizar e reutilizar estas informações na concepção de obras derivadas. Este processo contribui para a geração de novos conhecimentos e trazem outras experiências para a sociedade, em especial sobre os aspectos de licenciamento e direito de uso. Metadados externos⁶ são esperados, na medida em que é possível entrelaçar documentos e referências a outros ativos de informação relacionados. As obras decorrentes de filmes e vídeos precisam de uma atenção especial, pois eles são gerados agora pelo celular e enviados para sítios de divulgação, ou vazados, intencionalmente ou não, e publicados imediatamente. Portanto, o processo de representação já precisa ser trabalhado na fonte.

- Dados científicos. Os dados de investigações decorrentes da pesquisa em Ciência e Tecnologia possuem características diversas. Podem ser dinâmicos ou estáticos, efêmeros ou duráveis. No âmbito do experimento podem ser observacionais ou derivados de algoritmos computadorizados. Em relação ao processo de levantamento podem ser obtidos por meio de entrevistas ou originários de simulações de processos e fluxos de trabalho (*workflow*). Podem ter passado por algum processo de limpeza com depuração, retirada de duplicatas e identificação de registros com elos e, portanto, sua descrição precisará ser completa, incluindo os critérios utilizados neste processo de limpeza. Podem ser textos, vídeos, *softwares*, gráficos, cobrindo diferentes áreas como astronomia, física, medicina, ecologia, ciências sociais e história. Podem ainda ser classificados em relação à disponibilidade e licenciamento, podendo ser abertos ou fechados, livres ou dependendo de algum tipo de autorização de acesso. Em síntese, é preciso trabalhar com múltiplos dados, metadados e descrições.

O processo de representação traduz uma relação entre o representante e representado. Explicita o elemento que se comporta como *um substituto sintético*. Representar o conhecimento é descrever os conceitos, as propriedades, as características e os significados. O ativo de informação assume o papel em busca da convergência dos atos de representação do objeto e da representação da informação sobre o objeto.

⁶ Metadados externos contribuem para o entendimento de dados não estruturados (textos e outros tipos de manifestação sem estrutura prévia). Eles são criados para endereçar informação correlacionada aos dados primários (INMON; NESAVICH, 2008).

Cabe destacar ainda que o processo de representação destes ativos requer a verificação da validade desta informação. Esta preocupação em relação à capacidade probatória de informação remonta à Idade Média, quando havia a necessidade de aferir a autenticidade dos documentos devido às falsificações que se proliferavam (FRAENKEL, 1995). O uso de instrumentos que permitam garantir a validade e veracidade das informações entendidas e representadas se faz mister especialmente nos tempos atuais.

A junção de todas essas demandas de entendimento, representação, descrição e verificação catalisou a formação de um papel importante para representação e gestão de dados: o curador digital. Cabe ao curador garantir que o ativo de informação represente não apenas a evidência do ato praticado, mas também o propósito para o qual o ativo foi gerado e sua história.

Nessa direção, pode-se inferir que apoiado na experiência da Análise Documentária e nos postulados de Otlet (1934), Briet (1951) e Buckland (1997), o processo de representação conduzido pelo curador deve se apoiar em atributos ligados à materialidade, cobrindo as propriedades constitutivas; à intencionalidade representando o ato que será evidenciado; e, por fim, o pertencimento ao acervo, como uma possibilidade de auferir credibilidade a partir de uma coleção. Mas, como desenvolver estas propriedades no contexto digital?

A resposta a essa indagação revela que o entendimento dos pressupostos da Análise Documentária serve como insumos para implementar o deslocamento na direção de uma nova mediação, agora entre o usuário e a informação, seja ela analógica ou digital (RIBEIRO, 2017).

O que pode ser percebido é que para os ativos de informação existem visões diferenciadas em acervos museológicos, bibliográficos, pesquisa científica e ambientes virtuais na Web. Além disto, surgem cada vez mais normas, estatutos (tantos sociais quanto jurídicos) e acordos prévios garantindo o consenso da comunidade, com o intuito de garantir a validade e capacidade probatória destas unidades documentárias. Os parâmetros de ordem e controle para tratar estes novos ativos presentes no cotidiano da sociedade são outros, principalmente ao observar as novas dimensões de tempo e espaço.

A expressão consenso da comunidade nos traz de volta para a questão das HDs, pois elas são constituídas a partir de comunidades de prática, de

forma solidária, com filosofia aberta e livre acesso. Se comportando de maneira multilíngue, decorrente da ausência de fronteiras, e multidisciplinar, o que remete para uso de múltiplas linguagens de representação, as HDs pressupõem a transformação dos objetos de estudo em artefatos digitais com representação numérica (bits e bytes). Estes artefatos devem ser compostos por metadados, de forma a interoperar com outros ativos e favorecer o fluxo de troca de informação, viabilizando a circulação das informações entre diferentes domínios. (ROSEMBLOOM, 2012).

UM ROL DE TECNOLOGIAS PARA REPRESENTAÇÃO DOS SIGNOS E SIGNIFICADOS

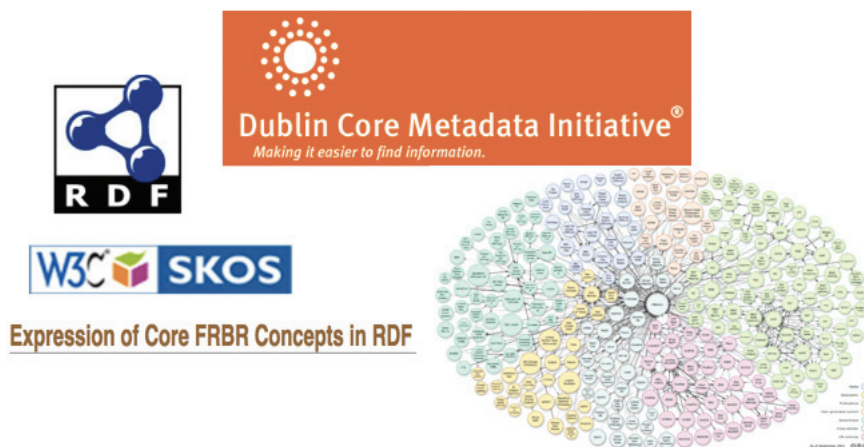
No passado, tanto as bibliotecas quanto as demais instituições de memória atendiam de forma satisfatória aos usuários, graças ao processo de descrição que era efetuado com uma diversidade de instrumentos para representar de forma fidedigna as Unidades Documentárias. Estas eram, predominantemente, registradas em suporte analógico (papel e fichas) e a diversidade de instrumentos descritivos se tornou conhecida como Knowledge Organization Systems (KOS)⁷. Com a evolução das tecnologias para representação e disseminação de informação, novos instrumentos foram criados e passaram a ser apoiados por formalismos e vocabulários para representação. Surgiram então diferentes diretivas tecnológicas para o desenvolvimento de soluções no contexto da Web Semântica, pois esta é mais que somente disponibilizar dados na Web, é sobre fazer ligações (*links*) para que pessoas e máquinas possam explorar esse conjunto de dados. Com o uso da abordagem de Dados Interligados e da noção de *Web* de Dados⁸, quando se tem estas ligações é possível encontrar outros dados relacionados.

É nesse contexto que se destaca o conjunto de tecnologias e formalismos (Figura 3), expressos por padrões, normas e instrumentos apresentados pelo consórcio World Wide Web Consortium (W3C).

⁷ Em português nos referenciaremos a Sistemas de Organização do Conhecimento (SOC).

⁸ Vale lembrar que a Web de Dados faz parte da Web Semântica, uma vez que as ligações (ou conexões) entre dados precisam de semântica por meio de *links* tipados.

FIGURA 2 - PADRÕES E FORMALISMOS DA W3C



Fonte: O autor (2018).

O recorte adotado nesse relato elenca brevemente formalismos e vocabulários mais utilizados para descrever ativos de informação em instituições de memória, segundo a literatura: Dublin Core Metadata Terms (DCMITERMS)⁹, Resource Description Framework (RDF Concepts)¹⁰, Simple Knowledge Organization System (SKOS)¹¹, Expression of Core FRBR Concepts in RDF (FRBRCore)¹² e a tecnologia de *Linked Data*, como forma de incrementar a expressividade das relações entre os ativos de informação.

- **Dublin Core (DC)** – O padrão propõe um *cartão virtual* para a descrição de objetos digitais, que vem a ser composto por uma estrutura de metadados. É um padrão registrado como ANSI, por meio da norma ISO 15836-2009. A correlação entre padrões para armazenar e comunicar informações bibliográficas em formato legível por máquinas não é novidade, pois Alves e Souza (2007) desenvolveram pesquisas para investigar a convergência e integração de representação com uso de DC e MARC21. Em um panorama

⁹ Disponível em: <<http://dublincore.org/documents/dcmi-terms/>>. Acesso em: 01 jun. 2019.

¹⁰ Disponível em: <<https://www.w3.org/TR/rdf-concepts/>>. Acesso em: 01 jun 2019.

¹¹ Disponível em: <<https://www.w3.org/TR/swbp-skos-core-spec/>>. Acesso em: 01 jun 2019.

¹² Disponível em: <<http://vocab.org/frbr/core>>. Acesso em: 01 jun 2019.

sobre o uso do DC no Brasil, Arakaki, Santos e Alves (2015) exploram a descrição de recursos bibliográficos na Web por meio da aplicação de DC e sua correlação com AACR2r.

- **RDF Concepts** – Com diferentes notações para representação (uso de grafos, RDF/XML, Turtle e N-Triples), este formalismo pode ser entendido como uma arquitetura de metadados voltada para a descrição de recursos na Web. Ele possibilita o processamento e a interpretação por máquinas e tem utilização em diferentes áreas de aplicação: descrição de recursos de informação e seus componentes, elaboração de mapas de navegação e estruturação de sítios, comércio eletrônico, bem como no desenvolvimento de agentes de *software* específicos. Em conjunto com SKOS, possibilita a representação da informação na *Web* no âmbito dos Tesouros, além de terminologias e representações específicas para CDD e CDU (CATARINO, 2014).
- **SKOS** – Com o formalismo SKOS é possível representar estruturas de organização do conhecimento de maneira semi-formal no ambiente Web, especialmente tesouros, taxonomias e esquemas de classificação. Em função da forte ligação entre SKOS e RDF, seu uso torna a representação processável por máquinas e facilita a troca de dados e informação entre aplicações de *software* no ambiente Web. (SKOS PRIMER, 2009). O modelo é apoiado em conceitos, propriedades e relacionamentos que permitem a descrição com semântica entre as unidades de informação (RAMALHO, 2017). Ademais, SKOS estabelece um conjunto de padrões na representação de conceitos e esquemas conceituais, além das relações de vizinhança necessárias para compreender melhor um conceito (*broader, narrower e related*). (CASTRO, 2016; SIMIONATO; ARAKAKI; SANTOS, 2017).
- **FRBR com RDF (FRBRCore)** – Em 2005, Ian Davis e Richard Newman desenvolveram o FRBRCore. Esta proposta de formalismo contemplou as entidades e relacionamentos do FRBR¹³, uma vez que é compreendido o conceito de classe no formalismo RDF como uma representação análoga à entidade no modelo FRBR. Sua estrutura principal é composta por classes em RDF que contemplam os Grupos 1, 2 e 3 do modelo FRBR.

¹³ O modelo Functional Requirements for Bibliographic Records (FRBR) apresenta os requisitos mínimos que os registros bibliográficos em formato eletrônico devem possuir. É baseado no modelo Entidade-Relacionamento proposto por Peter Chen (CHEN, 1976), fazendo uso dos conceitos de entidades, relacionamentos e atributos para representar recursos de informação.

Contudo, apesar de não contemplar em sua proposta inicial a totalidade de elementos do modelo FRBR, o formalismo FRBRCore permitiu que a essência dos requisitos para registros bibliográficos pudesse estar presente tanto no contexto da Web Semântica quanto no contexto da Web de Dados, dando início a outra etapa no compartilhamento das informações. Neste sentido, o uso da abordagem de Dados Ligados (*Linked Data*) propiciou uma maior visibilidade do uso do FRBRCore com RDF, pois identificou-se que este recurso está presente em mais de 24 milhões de triplas RDF no contexto da *Web* de Dados (COYLE, 2016). O uso do FRBRCore permite a convergência e convivência entre padrões para o ambiente digital com a possibilidade de reunir catálogos de registros bibliográficos usando Ontologias e RDF, em conjunto com o tratamento e armazenamento em SKOS e com destaque para a convivência dos formalismos estudados com MARC21, AACR2, RDA e CDD (CASTRO, 2016).

- **Uso de Dados Ligados** – Podem ser utilizados para descreverem um conjunto de práticas para publicar, compartilhar e conectar dados estruturados na Web de forma a aumentar o seu valor e uso (ARAÚJO; SOUZA, 2011). A proposta de Dados Ligados nos leva a criar ligações que relacionem dados de diferentes fontes. Estas fontes podem ser bancos de dados, diferentes sistemas informatizados, ou até mesmo outras páginas Web. A correlação obtida em dados ligados deve possuir uma descrição semântica (BIZER; HEATH; BERNERS-LEE, 2009). É possível correlacionar e descrever um sentido (ou propriedade) na ligação, utilizando um predicado para tipificar e ligar os elementos descritos. Fazendo uso de URIs (*Universal/Uniform Resource Identifier*) para identificação e endereçamento de ativos de informação, bem como a descrição em RDF para representá-los, os dados ligados possibilitam que a Web lance mão do verdadeiro conceito de teia de dados. Vale lembrar que nesta teia semântica as descrições precisam de um esquema de nomenclatura global, com dicionários e vocabulários publicados e com algum tipo de controle, permitindo a constante troca de dados e informação (RIBEIRO; ALMEIDA, 2011). A utilização de Dados Ligados permite ampliar a capacidade de encontrar e correlacionar informação por usuários de bibliotecas e museus. A busca por documentos para possivelmente correlacioná-los com obras (biblioteca e museus), bem como suas possíveis expressões e manifestações (vídeos e arquivos de áudio), impulsionará o processo de criação de novos conhecimentos pela sociedade. (SIMINONATO; ARAKAKI; SANTOS, 2017).

Conforme mencionado anteriormente, vale lembrar que o papel do curador digital deve estar em sintonia com as definições apresentadas por esses instrumentos. Será necessária a adoção de representações digitais para os ativos de informação englobando as diferentes dimensões para representação (propriedades, contexto, ligações externas, proveniência, transformações, dentre outras). Descrever diferentes objetos originários de acervos de distintos – sejam bibliográficos, museológicos ou arquivísticos – compõe o cenário desafiador para o curador digital (MARCONDES, 2016).

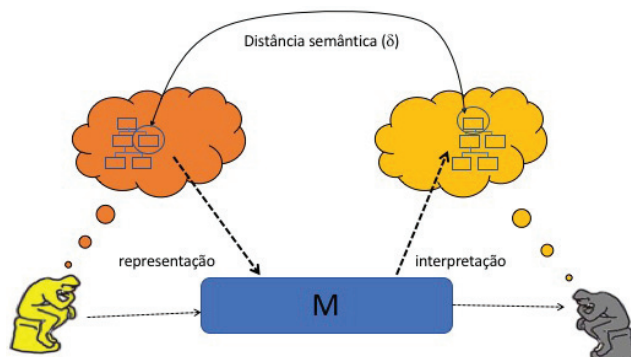
Explorar essas tecnologias para o processo de representação com semântica no ambiente Web é condição importante, pois os ativos de informação transbordam o conceito de Unidades Documentais na direção das Humanidades Digitais. O uso destas tecnologias vem provocando várias inquietações tanto no Brasil quanto no exterior, na medida em que os autores buscam estudar técnicas e práticas para possibilitar a interoperabilidade entre bases de dados. Tudo isto de forma que esses ativos sejam descritos, para relacioná-los com outros inúmeros documentos e/ou dados. (CASTRO, 2016; CLOUGH, 2013; COYLE, 2016; MARCONDES, 2012; 2016; PELLEGRINO et. al., 2017; SANTOS NETO *et al.*, 2013; SIMIONATO; ARAKAKI; SANTOS, 2017).

POSSÍVEIS DISTÂNCIAS SEMÂNTICAS NA REPRESENTAÇÃO DE SIGNIFICADOS

O rol de tecnologias apresentado anteriormente não esgota o tema da representação no ambiente Web, mas demonstra a riqueza de instrumentos tecnológicos que objetivam descrever ativos de informação. Vale lembrar que dependendo do domínio alvo de análise, algumas discrepâncias de entendimento podem ocorrer na compreensão das qualidades e predicados descritos, bem como seus significados para o propósito adequado. As possíveis associações e relações de dependência que estes elementos possuem também podem ser alvo de diferenças. Todas estas diferenças poderão causar algum nível de distância semântica na compreensão da representação obtida (Figura 3).

Nesse sentido, surgem as propostas de modelos ontológicos que possam expressar adequadamente os ativos de informações avaliados, enriquecendo a discussão sobre signos e significados.

FIGURA 3 - SIGNIFICAÇÃO SEMÂNTICA



Fonte: Guizzardi (2011).

A procura de um modelo M que represente e possibilite a interpretação com alto grau de convergência entre conceitos leva este debate para o campo das Ontologias. Uma Ontologia traz uma aproximação do mundo real e remete para o uso de vocabulários e conceitos, além de relações claras e não ambíguas, explicitando o conhecimento no domínio alvo de análise (GRUBER, 1995; GUARINO, 1997).

As Ontologias podem ainda possibilitar o suporte à interoperabilidade, pois permitem que diferentes sistemas e usuários façam o intercâmbio de dados e informação. Contudo, não é objetivo deste relato aprofundar a discussão sobre Ontologias, mas apenas referenciá-las como artefato importante na descrição e compartilhamento de ativos de informação.

Com visões diferenciadas, mas convergentes, nas áreas de Ciência da Computação e Ciência da Informação o seu uso é sustentado por vasta literatura que apresenta desde propostas de categorizações pelo grau de formalismo até as funções para aplicação da Ontologia. (ALMEIDA; BAX, 2003;; BORST, 1997; CAMPOS, 2017; GUIZZARDI, 2005; GUIZZARDI; WAGNER; ALMEIDA; GUIZZARDI, 2015).

O uso de Ontologias no processo de representação de ativos de informação dentro do contexto das Humanidades Digitais pode ser profícuo, em especial no âmbito do compartilhamento de informações sobre recursos de informação cultural digital. Neste sentido, cabe exemplificar que uma das propostas de evolução do modelo FRBR está ligado à compatibilidade e harmonização com

o modelo de referência CIDOC-CRM, proposto pelo Comité International pour la Documentation. Este esforço procurou representar a visão unificada sobre as entidades do patrimônio histórico-cultural: as bibliotecas, os museus e arquivos (LIMA, 2010). A International Federation of Library Associations and Institutions (IFLA) também desenvolveu o processo de evolução para tornar a proposta do FRBR mais próximo de uma Ontologia, o que leva à compatibilidade com formalismos presentes nesta proposta de projeto e que foram definidos como padrão para representação pelo W3C (IFLA, 2015).

CONSIDERAÇÕES FINAIS

A sociedade contemporânea está passando por transformações intensas impulsionadas pela evolução das tecnologias de informação e comunicação. O comportamento dos usuários está cada vez mais voltado para o papel representado pelo neologismo proposto por Alvin Toffler em 1981 e estendido por Bruns em 2008. Segundo a visão apresentada por Toffler (1981), pode-se afirmar que o papel de *prosumer* representa o comportamento de todos nós dentro deste contexto de avalanche de informações na atualidade. Entretanto, Bruns (2008) faz um contraponto e complementa o conceito na direção do papel de *produsage*, pois destaca que a verdadeira motivação só foi viável com a chegada da proposta implementada pela Web 2.0 nos aspectos do compartilhamento e da colaboração.

Produzir, consumir e reutilizar em diferentes contextos, tais como Artes, Ciências Sociais e Humanas, Instituições de Memória, Educação, Engenharias e Ciências da Saúde, exige um processo de entendimento e representação cada vez mais completo. Este processo precisa permitir interpretações convergentes e devem reduzir os riscos de adotarmos visões equivocadas na atividade de reuso de dados e informação.

A garantia desta convergência permitirá o sincronismo entre projetos desenvolvidos no presente, mas que possam reutilizar dados e informação já produzidos no passado para atingirmos resultados significativos no futuro. Como afirma Rufus Pollock¹⁴, fundador da *Open Knowledge Foundation*, o melhor uso que poderá ser feito com os dados obtidos e produzidos pela sua investigação certamente será feito por outros e não por você!

¹⁴ Disponível em: <<https://rufuspollock.com/misc/>>. Acesso em: 20 dez. 2018.

REFERÊNCIAS

ALMEIDA, M; BAX, M. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Revista Ciência da Informação*, v. 3, n. 3, p. 7-20, 2003.

ALVES, M. D. R.; SOUZA, M. I. F. Estudo de correspondência de elemento metadados Dublin Core e Marc21. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 4, n. 2, p. 20-38, jan./jun. 2007.

ARAKAKI, F. A.; SANTOS, P. C. L. V. A. C.; ALVES, R. C. V. Panorama das pesquisas sobre o padrão de metadados dublin core no Brasil. *Revista ACB: Biblioteconomia em Santa Catarina*, v. 20, n. 1, p. 86-97, 2015.

ARAÚJO, L. R.; SOUZA, J. F. Aumentando a transparência do governo por meio da transformação de dados governamentais abertos em dados ligados. *Revista Eletrônica de Sistemas de Informação*, v. 10, n. 1, p. 1-15, 2011.

BARBOSA, E. C.; KOBASHI, N. Y. Extroversão e descoberta: visualização de dados no auxílio a buscas e recuperação de informações. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 13, p. 115-120, set. 2017. Disponível em: <<https://rbbd.febab.org.br/rbbd/article/view/766>>. Acesso em: 20 jun 2019.

BEIGUILMAN, G.; MAGALHÃES, A. G. *Futuros possíveis: arte, museus e arquivos digitais*. São Paulo: Ed.USP, 2014.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. *Linked Data: the Story so Far*. 2009. Disponível em: <<https://goo.gl/fvnecy>>. Acesso em: 19 out. 2018.

BORST, W. *Construction of Engineering Ontologies for Knowledge Sharing and Reuse*. PhD thesis, University of Twente, P.O. Box 217 - 7500 AE Enschede - The Netherlands, 1997.

BRIET, S. *Qu'est-ce que la documentation?* Paris: Éditions Documentaires Industrielles et Techniques, 1951. Disponível em: <<http://martinetl.free.fr/suzannebriet/questcequeladocumentation/>>. Acesso em: 12 nov. 2018.

BUCKLAND, M. What is a Document? *JASIS*, Berkeley, v. 48, n. 9, p. 804-809, Sept. 1997.

BURDICK, A. *et al. A Short Guide to the Digital Humanities*. Cambridge, MA: MIT Press, 2012.

BRUNS, A. *Blogs, Wikipedia, Second Life, and Beyond: From Production to Produsage*. New York: Peter Lang, 2008.

CAMPOS, M. L. de A. Elaboração de modelos de domínio em ontologias: a abordagem onomasiológica e a função da definição. *Ciência da Informação*, v. 46, n. 1, dez. 2017. <https://doi.org/10.18225/ci.inf.v46i1.4016>.

CASTRO, F. F. Functional Requirements for Bibliographic Description in Digital Environments. *TransInformação*, v. 28, n. 2, p. 223-231, 2016.

CATARINO, M. E. Simple Knowledge Organization System: construindo sistemas de organização do conhecimento no contexto da Web Semântica. *Informação & Tecnologia*, v. 1, n.1, p. 4-20, 2014.

CHEN, P. P. S. The Entity-Relationship Model: Toward a Unified View of Data. *ACM Transaction. Database Syst*, v. 1, p. 9-36, 1976. <http://dx.doi.org/10.1145/320434.320440>.

CLOUGH, G. W. *Best of Both Worlds: Museums, Libraries, and Archives in a Digital Age*. Washington, DC: Smithsonian Institution. 2013.

COYLE, K. *FRBR, Before and After: a Look at our Bibliographic Models*. Chicago: ALA Editions, 2016.

DH-MANIFESTO. *Manifeste des Digital Humanities*. 2010. Disponível em: <<http://tcp.hypotheses.org/443>>. Acesso em: 25 nov. 2018.

FRAENKEL, B. A assinatura contra a corrupção do escrito. In: BOTTÉRO, J.; MORRISON, K. *et al. Cultura, pensamento e escrita*. São Paulo: Ática, 1995. p. 81-99.

GODOY VIERA, A. F.; VARVAKIS, G.; FORESTI, F. Perspectivas e desafios dos dispositivos móveis para as bibliotecas universitárias brasileiras. *Revista Interamericana de Bibliotecología*, v. 41, n. 1, p. 19-35, jan. 2018. Disponível em: <<https://aprendeenlinea.udea.edu.co/revistas/index.php/RIB/article/view/330639>>. Acesso em: 12 dez. 2018.

GRUBER, T. R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies*, v. 43, n. 5-6, p. 907-928, 1995.

GUARINO, N. Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In: PAZIENZA, M. T. *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*. Berlim: Springer Verlag, 1997. p. 139-170.

GUIZZARDI, G. *Ontological Foundations for Structural Conceptual Models*. Enschede, The Netherlands, CTIT PhD Thesis Series, No. 05-74 Telematica Instituut Fundamental Research Series, No. 015 (TI/FRS/015). 2005. Disponível em: <http://doc.utwente.nl/50826/1/thesis_Guizzardi.pdf>. Acesso em: 01 dez. 2018.

GUIZZARDI, G. *Infraestrutura Nacional de Dados Abertos*: Curso de Engenharia de Ontologias. Brasília, DF: Ministério do Planejamento, Desenvolvimento e Gestão, 2011.

GUIZZARDI, G; WAGNER, G; ALMEIDA, J. P. A.; GUIZZARDI, R. S. S. Towards Ontological Foundations for Conceptual Modeling: The Unified Foundational Ontology (UFO) Story. *Applied Ontology*, v. 10, n. 3-4, p. 259-271, 2015.

IFLA. Interstitial Federation of Library Associations and Institutions. Definition of FRBROO - A Conceptual Model for Bibliographic Information in Object-Oriented Formalism. Edimburgo, UK: IFLA, 2015.

INMON, W. H.; NESAVICH, A. *Tapping into Unstructured Data: Integrating Unstructured Data and Textual Analytics into Business Intelligence*. Boston: Prentice Hall, 2008.

KREBS, L. M.; ROCHA, R. P.; RIBEIRO, C. Quem leu este também leu...: sistema de recomendação na biblioteca universitária. *Perspectivas em Ciência da Informação*, v. 22, n. 1, p. 151-169, 2017. <https://doi.org/10.1590/1981-5344/2496>.

LIMA, J. A. O. FRBROO: conceitos básicos. In: ROBREDO, J.; BRÄSCHER, M. (Org.). *Passeios no bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: IBICT, 2010.

LOURENÇO, C. de A. Representação da informação: sua abordagem nos cursos de biblioteconomia e nas pesquisas em Ciência da Informação. In: ZAFALON, Z. R.; DAL'EVEDOVE, P. R. *Perspectivas da representação documental: discussão e experiências*. São Carlos: CPOI/UFSCar, 2017. p. 229-240.

MALINI, F.; CIARELLE, P.; MEDEIROS, J. O sentimento político em rede sociais: big data, algoritmos e as emoções nos tweets sobre o impeachment de Dilma Rouseff. *Liinc em Revista*, v. 13, n. 2, p. 323-342, 2017. Disponível em: <<http://revista.ibict.br/liinc/article/view/4089>>. Acesso em: 20 nov. 2018.

MARCONDES, C. H. Linked data: dados interligados e interoperabilidade entre arquivos, bibliotecas e museus na web. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, Florianópolis, v. 17, n. 34, p. 171-192, maio/ago. 2012.

MARCONDES, C. H. Interoperabilidade entre acervos digitais de arquivos, bibliotecas e museus: potencialidades das tecnologias de dados abertos interligados. *Perspectivas em Ciência da Informação*, v. 21, n. 2, p. 61-83, jun. 2016. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/2735>>. Acesso em: 20 dez. 2018

MEDEIROS, A. L. S *et al.* Humanidades Digitais na Fundação Casa de Rui Barbosa: um estudo aplicado de seu conceito. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 18., 2017, Marília. *Anais...* Marília: UNESP, 2017. Disponível em: <<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/viewFile/195/845>>. Acesso em: 01 dez. 2018.

MOORE, S. *Smart With GARTNER: Learn the Art of Cultural Hacking*. 2018. Disponível em: <<https://www.gartner.com/smarterwithgartner/learn-the-art-of-culture-hacking-for-culture-change/>>. Acesso em: 12 nov. 2018.

OTLET, P. *Traité de Documentation: le livre sur le livre: théorie et pratique*. Bruxelles: Mundaneum, 1934. Disponível em: <https://lib.ugent.be/fulltxt/handle/1854/5612/Traite_de_documentation_ocr.pdf>. Acesso em: 10 dez. 2018.

PELLEGRINO, A. L. *et al.* Bibliotecas e instituições de memória na Web, Dados ligados e Web Semântica: Diálogos interdisciplinares. *Memória e Informação*, v. 1, n. 1, p. 53-72, jul./dez. 2017.

RAMALHO, R. A. S. Representação SKOS categoria tecnologia da informação e comunicação do Tesouro Brasileiro de Ciência da Informação: um estudo preliminar. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2017, Marília. *Anais ...* Marília: UNESP, 2017.

RIBEIRO, C. J. S. Big Data: os novos desafios para o profissional da informação. *Informação & Tecnologia*, v. 1, p. 96-105, 2014.

RIBEIRO, C. J. S. Desafios para a representação documental no âmbito da Web Semântica. In: ZAFALON, Z. R.; DAL'EVEDOVE, P. R. (Org.). *Perspectivas da representação documental: discussão e experiências*. São Carlos: CPOI/UFSCar, 2017.

RIBEIRO, C. J. S.; ALMEIDA, R. F. Dados Abertos Governamentais (Open Government Data): instrumento para exercício de cidadania pela sociedade. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. *Anais...* Brasília: ANCIB, 2011. p.2568-2580.

ROZSA, V.; DUTRA, M. L.; NHACUONGUE, J. A. Linked open data no contexto acadêmico: identificação e análise de vocabulários utilizados na academia e na pesquisa científica. *Brazilian Journal of Information Science*, v. 11, n. 3, p. 34-52, 2017. <https://doi.org/10.36311/1981-1640.2017.v11n3.05.p34>

ROSENBLUM, P. S. Towards a Conceptual Framework for the Digital Humanities. *Digital Humanities Quarterly*, v. 6, n. 2, 2012. Disponível em: <<http://digitalhumanities.org:8081/dhq/vol/6/2/000127/000127.html#>>. Acesso em: 15 nov. 2018.

SANTOS NETO, A. L. dos *et al.* Tecnologias de dados abertos para interligar bibliotecas, arquivos e museus: um caso machadiano. *TransInformação*, Campinas, v. 25, n. 1, p. 81-87, jan./abr. 2013.

SANTOS PUTON, B.; MANGAN, P. K. V. BibliotecAR: Realidade Aumentada em um Sistema Android para Auxílio de Localização de Livros em uma Biblioteca. *Revista Eletrônica Argentina-Brasil de Tecnologias da Informação e da Comunicação*, v. 1, n. 2, mar. 2015. Disponível em: <<https://revistas.setrem.com.br/index.php/reabtic/article/view/62>>. Acesso em: 20 dez. 2018.

SIMIONATO, A. C.; ARAKAKI, F. A.; SANTOS, P. L. V. A. da C. Descrição em bibliotecas, arquivos, museus e galerias de arte: linkando recursos e comunidades. *Informação & Informação*, v. 22, n. 2, p. 449-466, 2017.

SKOS PRIMER. *SKOS Simple Knowledge Organization System Primer*. 2009. Disponível em: <<http://www.w3.org/TR/2009/NOTE-skos-primer-20090818>>. Acesso em: 16 nov. 2018.

TOFFLER, A. *The Third Wave*. London: Pan Books Ltd., 1981.

WIKIPEDIA. *Tecnologia Disruptiva*. 2016. Disponível em: <https://pt.wikipedia.org/wiki/Tecnologia_disruptiva>. Acesso em: 10 dez. 2018.

Humanidades Digitais, Memória e Filatelia: uma aplicação prática para a Web

Diego A. Salcedo

INTRODUÇÃO

O projeto de desenvolvimento do Repositório Filatélico Brasileiro (REFIBRA), no qual serão inseridos a Bibliografia Filatélica Brasileira (BIFIBRA) e o Dicionário Filatélico Brasileiro (DIFIBRA), todos eles produtos inéditos no país, está devidamente homologado pelas respectivas instâncias da Universidade Federal de Pernambuco (UFPE), bem como cadastrado no grupo de pesquisa do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esse projeto articula atividades de pesquisa, ensino, extensão e cultura coordenadas pelo Grupo de Pesquisa Imago e Humanidades Digitais (IHD), em parceria com o Departamento de Ciência da Informação (DCI) e o Programa de Pós-Graduação em Ciência da Informação (PPGCI), da UFPE, desde 2014.

Em maior escala, o REFIBRA contribui com um conjunto de princípios que norteiam os programas estratégicos da uma rede nacional de instituições comprometidas com políticas de preservação e digitalização de acervos memoriais brasileiros. O REFIBRA, assim, tem se constituído como o ambiente em que está acontecendo um processo de curadoria digital do patrimônio memorial filatélico brasileiro que, por sua vez, tem sido produzido ininterruptamente desde 1843. Esse patrimônio é constituído por diversos e distintos documentos filatélicos: selos, inteiros, blocos, editais, catálogos, livros, periódicos, boletins, folhetos etc. Vale lembrar que o Brasil, em 1843, foi o terceiro país a emitir e utilizar o selo postal para franquear correspondências, de fato, um pequeno embaixador de papel (ALTMAN, 1991; SALCEDO, 2010).

O projeto corrobora com as ações de preservação da memória digital recomendadas pela Organização das Nações Unidas para a Educação, a Ciência e

a Cultura (UNESCO), na sua Declaração de Vancouver (2012). Logo, insere o Brasil como um dos pioneiros no cenário internacional, particularmente, no campo da Filatelia, da História e da Memória Postal. Na esfera nacional, o projeto tem desenvolvido um ambiente digital que contribuirá para a preservação, conservação e difusão da memória, da cultura e da identidade nacional em consonância, também, com o Plano Nacional de Cultura (PNC), do Ministério da Cultura do Brasil (2010).

Nesse contexto, este capítulo apresenta o desenvolvimento da etapa inicial do REFIBRA, como ocorreu sua concepção e quais as tecnologias computacionais utilizadas. Como metodologia faz a revisão bibliográfica desde alguns elementos propostos na literatura científica especializada da Biblioteconomia, Ciência da Informação, Computação e Filatelia, caracterizando o que pensamos constituir certo desafio interdisciplinar. Ele ilustra uma proposta preliminar de aplicação web semântica para o tratamento informacional de 32 selos postais comemorativos brasileiros e utiliza o padrão de metadados Dublin Core, com necessárias adequações, o sistema de visualização *Linked Data* e o *Resource Description Framework*.

O capítulo está estruturado em cinco seções. Após esta introdução, segue a seção 2, que justifica a importância do selo postal e sua relação com a memória nacional. Em seguida, a seção 3 explora o que deve ser conceituado, entendido e aplicado para que se possa trabalhar com recursos na web semântica, nesse caso, os selos postais. A seção 4 indica como foi desenvolvida a fase inicial do REFIBRA com a aplicação das tecnologias apresentadas na seção anterior. Por fim, a seção 5 discorre sobre as conclusões e as proposições de continuidade do projeto.

MEMÓRIA E FILATELIA: O CASO DO O SELO POSTAL COMEMORATIVO BRASILEIRO

O que dizer de tão rica, lúdica e poderosa interface de informação? Esse pequeno pedaço de papel, indiferente às diversas formas como se apresenta e aos suportes aos quais é agregado, elimina distâncias, preserva com criatividade uma possível história da humanidade. Resgata na forma de documento, pessoas, fatos, eventos, processos e o tempo, de forma geral, funcionando como um elo entre o humano, sua história e o conhecimento político, econômico, social e cultural.

Assim, conforme sugere Salcedo (2010, p. 73),

O selo postal oferece a oportunidade para que possamos, se olharmos atentamente, perceber as transformações pelas quais temos passado, como conduzimos o desenvolvimento tecnológico, como nos distanciamos ou aproximamos do Outro, como lidamos com as diferenças e as semelhanças, como continuamos contando a nossa própria história e a da Natureza, como dizemos ou silenciemos nossos discursos e como os Estados ramificam os seus.

O selo postal, enquanto interface documental tem características marcantes, entre as quais, ser aproveitado tanto para enviar mensagens ao grande público como educar novos leitores de mundo. Isso é possível porque sua estrutura física enquanto objeto material é de fácil manuseio. Aos selos poder-se-ia aplicar a síntese feita por Moles (2004) para o cartaz. Eles têm a função de informar (qual a comemoração, o feito, o objeto da emissão) e de educar (dando a conhecer valores ou temas que escolheram).

Os selos postais servem, por exemplo, para reconhecer e legitimar uma nação, e estão entre os produtos do Estado que possuem maior visibilidade interna e externamente. Enquanto a moeda é produzida conceitualmente para o uso interno de uma unidade política, o selo postal, mais que isso, assume um papel diplomático ao ser utilizado, também, além das fronteiras de sua unidade política emissora. Selos postais são produtos do Estado emissor e difundem universalmente o simbolismo público e as tradições nacionais tanto dentro de suas fronteiras quanto ultramar.

De acordo com as especificações de texto e imagem impressos, os selos postais podem ser categorizados como uma mídia, um meio de comunicação. Além disso, a mensagem impressa no selo é intencional, detentora de mensagens subliminares que, em certa medida, passam despercebidas ao leitor comum, ou ao grande público, que apenas o identifica como taxa devida ao Correio para o envio de missivas postais.

Ora, os selos constituem um excelente meio de propaganda e, assim sendo, justo é que os governos se esforcem de representar fatos, episódios e vultos do nosso país. Os selos postais tornaram-se úteis artefatos ideológicos e culturais para os governos no sentido de promover certa imagem em casa e fora dela.

Interessam, então, as mudanças ocorridas nos períodos imperial e republicano brasileiros, particularmente, a partir de 1843, posto que refletiram diretamente na produção e circulação de selos postais, seus desenhos, suas tarifas e seus tipos.

Ao mesmo tempo que os próprios selos postais, de alguma forma, integravam um acervo documental (coleções) que materializava àquelas mudanças.

Ao ser traçado um panorama das linguagens visuais dos selos postais imperiais brasileiros, é possível perceber um interesse em arrecadar receita por meio do erário elitizado, impondo um limite intransponível tanto em torno do sistema de comunicação quanto no do colecionismo às pessoas de menor poder aquisitivo. A popularização dessa linguagem diz respeito não só ao meio, propriamente dito, ao regime de informação em que se constitui, mas, também, às pessoas que o formam e seus parâmetros socioculturais.

Não é adequado separar esses pequenos textos do quadro social em que foram produzidos e nos quais circularam. No entanto, também não é adequado analisar o selo postal sem considerar a interface documental iconográfica como registro produzido pelo observador; o documento visual como registro ou parte do observável, na sociedade observada; e, finalmente, a interação entre observador e observado.

Por volta de 1900, no Brasil, existia o que pode ser denominado de “dialética do colecionismo filatélico”. Ela se baseia, principalmente, nas conexões entre o colecionismo individual e o coletivo. Parte fundamental desse movimento é o papel da instituição produtora dos objetos colecionáveis, no caso dos selos postais, os Correios. Assim, é possível perceber que, naquele período e que dura até hoje, teve início um movimento de retroalimentação entre os Correios e os colecionadores, incluindo os comerciantes filatélicos. Desde então, os Correios passaram a produzir uma série de produtos para atrair a atenção dos colecionadores e comerciantes filatélicos. É nesse contexto que emerge, dentre variados tipos de documentos filatélicos, o selo postal do tipo comemorativo, objeto específico de que trata este capítulo e que serviu como objeto para o protótipo do REFIBRA, tema tratado nas seções seguintes.

WEB SEMÂNTICA E HUMANIDADES DIGITAIS

As tecnologias de informação e comunicação têm aumentado a facilidade de distribuição de informação através da WEB, tendo o volume de documentos e *links* navegáveis disponíveis, para qualquer pessoa que tenha acesso à Internet, crescido muito mais se comparado ao começo da década. Todo esse volume e diversidade

de informações têm criado dificuldade para que o usuário encontre conteúdos relevantes para suas necessidades informacionais.

Como abordado por Souza e Alvarenga (2004), apesar da Web ter sido projetada para facilitar a troca de informações, seu crescimento se deu de forma rápida e caótica, resultando em um enorme repositório de documentos recuperáveis. Esse referido repositório é conhecido atualmente como Web Sintática, uma vez que trata as informações comparando a estrutura de padrões de texto idênticos para recuperação da informação (MARCONDES, 2012) desconsiderando o contexto e as ligações de inferência dos documentos recuperados.

Nesse cenário, os resultados de pesquisas na WEB feitos por meio dos motores de busca apresentam como resultado uma listagem de documentos que muitas vezes contêm os termos pesquisados pelo usuário, mas que nada tem de relevante com o contexto da sua necessidade informacional, assim as consultas apresentam problemas de atinência com baixa revocação e precisão (SOUZA; ALVARENGA, 2004).

A nova abordagem semântica da WEB proporciona uma outra forma de se pensar a respeito de informação disponível na Internet, como os conhecimentos disponíveis podem ser melhor recuperados e indexados, ou seja, essa nova arquitetura de dados tem influência direta no dia a dia do profissional da Ciência da Informação e de áreas correlatas.

Dessa forma, Souza e Alvarenga (2004) afirmam que a Web Semântica e a Ciência da Informação são campos de pesquisa e desenvolvimento que convergem para um mesmo ponto e se complementam. Os autores observam possíveis melhorias nas atuais atividades do profissional da informação e do tratamento do seu objeto de trabalho: melhoria nos motores de busca, novas construção de interfaces para o usuário, construção automática de tesouros e vocabulários controlados, melhor indexação automática de documentos, melhor gestão do conhecimento organizacional e uma gestão da informação estratégica mais eficaz.

Nesse contexto, a web semântica também influencia diretamente na forma como bibliotecas, museus e arquivos podem trabalhar e disponibilizar seus catálogos, além de promover uma interoperabilidade entre eles, como proposto por Marcondes (2012).

Outros autores, como Serra e Santarém Segundo (2017), sugerem exemplos práticos de como as bibliotecas atuais, físicas e digitais, podem converter seus

catálogos, por exemplo, em MARC (sigla para Machine Readable Cataloging – catalogação legível por computador), para um modelo de dados RDF (sigla para Resource Description Framework – um conjunto de especificações da World Wide Web Consortium – W3C, originalmente planejada como um modelo de dados para metadados). Essa conversão teria como objetivo atender exigências de princípios do *Linked Data* (dados ligados entre si com a finalidade de publicar e estruturar dados na WEB).

Essa diversidade de agentes interconectados em rede (humano-humano, humano-máquina e máquina-máquina) foi observada por Floridi (2014), ao apresentar, num positivo e pertinente debate a respeito da ética da informação contemporânea, o conceito de infosfera, o qual consiste em um ecossistema informacional constituído por entidades informacionais chamadas *Inforgs*, não necessariamente humanas, que tem como principal função colaborar com a dinâmica da informação num ecoambiente computacional.

De fato, há muita pesquisa e desenvolvimento em andamento com relação a projetos da Web Semântica e *Linked Data* existindo um grande número de vocabulários controlados e padrões de dados propostos, como por exemplo o *Bibframe* abordado por Ramalho (2016a; 2016b). No entanto, o escopo do desenvolvimento do REFIBRA tratado neste capítulo é o estudo de um modelo de dados para um repositório do patrimônio filatélico nacional de acesso aberto, baseado em tecnologia web semântica com princípios do *Linked Data*, conectado e interoperável com outras bases de dados abertas (*machine to machine*), de maneira que o conteúdo esteja acessível ao público potencial e que, como pesquisa em andamento, está caracterizada pelo que pensamos constituir certo desafio interdisciplinar para as áreas envolvidas.

Por essa razão, foi escolhido o padrão de metadados Dublin Core para descrever os itens catalogados até então. Tal padrão possui um conjunto de 15 elementos extensíveis para descrever recursos eletrônicos na WEB apresentando características como simplicidade, interoperabilidade semântica, extensibilidade, flexibilidade e certo consenso internacional, uma vez que mais de 20 países aceitam esse modelo como um metadado usável para buscas em meio eletrônico (GRÁCIO, 2002).

Para atualizar a forma de como os documentos são tratados e recuperados na Web, Berners-Lee, Lassila e Hendler (2001) propuseram que tanto as máquinas

quanto as pessoas pudessem entender melhor o contexto em que o documento está inserido, fazendo inferências sobre eles (a web semântica). Ducharme (2013, p. 35, tradução nossa) define a web semântica como “um conjunto de padrões e melhores práticas para compartilhamento de dados e seus elementos semânticos na WEB para uso aplicado”¹, sendo esse “conjunto de padrões” tecnologias que definem uma estrutura necessária para que as informações sejam disponibilizadas de uma maneira interoperável; tais padrões, como o RDF e SPARQL, (linguagem para recuperação/consultas de dados em formato RDF) são mantidos pelo W3C.

Concomitantemente com a ideia de dados inseridos em uma Web que não trabalha mais apenas com a comparação de texto, mas também com a inferência semântica da informação, emergiu a de *Linked Data*, que são princípios e boas práticas que tratam qualquer informação como um recurso representado por *Uniform Resource Identifier* (URI) e potencialmente ligado à outros recursos, formando uma teia de interligações: “Berners-Lee criou o Linked Data como melhor prática para compartilhar dados por meio da infraestrutura Web, oferecendo, assim, excelentes guias para o desenvolvimento de uma infraestrutura de web semântica”² (DUCHARME, 2013, p. 36, tradução nossa).

Atualmente, a proposta da web semântica está em pleno desenvolvimento, tendo grandes expoentes incentivadores como a Dbpedia, um projeto que extrai informação da Wikipédia para ser estruturada de maneira semântica e disponibilizada no modelo necessário para que agentes inteligentes consigam usar essa informação. Porém, o processo de mudança da web sintática para o novo modelo semântico é demorado, uma vez que depende de uma estrutura de colaboração universal de diversos interesses para disponibilização de informação interconectada (SANTOS; CARVALHO, 2007, p. 5). Vale lembrar que hoje é possível apontar três tipos distintos de Internet, todas elas imbricadas entre si, mas internas a uma quarta: a Internet das coisas conectadas.

Nesse sentido, Ferry (2018) sugere que esta nova fisiologia social das máquinas está dominada pela produção e consumo colaborativo e que isso advém

¹ “[...] a set of standards and best practices for sharing data and the semantics of that data over the Web for use by applications”

² “Berners-Lee came up with the idea of Linked Data as a set of best practices for sharing data across the web infrastructure so that [...]. They provide excellent guidelines for the creation of an infrastructure for the semantic web.”

do somatório de três momentos tecnológicos bem demarcados: os adventos da Internet da Comunicação, da Internet da Energia e da Internet da Logística (ou dos Transportes), todas elas comandadas pela Internet das Coisas Conectadas.

Por sua vez, o RDF é um modelo de representação de dados largamente utilizado para troca de informações semanticamente relacionadas por sistemas na web (PEREIRA; HEINRICH; SCHROEDER, 2016). Tal modelo é baseado na ideia de que todo item pode ser descrito e relacionado por meio de um conceito de triplas no qual o *subject* é o recurso em questão, o *predicate* é a propriedade que tem relação com o *object* que, por sua vez, tem o valor da propriedade (LIMA; CARVALHO, 2005). Com a incorporação do valor de URIs tanto para o *subject* como para o *object*, o RDF permite a criação de um mapeamento relacional entre os itens, de modo a permitir uma navegabilidade e relacionamento compreensível por humanos e máquinas (COYLE, 2017).

No entanto, os problemas começam a surgir quando é preciso pensar na forma de armazenamento de documentos em formato RDF, em como persistir os itens mapeados em meio digital. Inicialmente, é possível pensar numa forma de armazenamento feito em um banco de dados relacional, porém ainda assim é preciso refletir sobre a melhor forma de armazenamento de dados e seus relacionamentos, pois eles podem ser muito complexos quando o usuário precisar recuperar itens relacionados com itens de natureza diferentes. Nesse cenário, existe a proposta do banco de dados relacional no modelo *triple-store*, que consiste na criação de uma única tabela contendo apenas três colunas (*subject*, *predicate* e *object*), em que qualquer tipo de item pode ser mapeado de modo simples.

Apesar de existirem críticas relacionadas à degradação de desempenho dessa abordagem, como estudado por Pereira (2016) e Bayer, Nese e Schroeder (2015), o RDF é utilizado em diversos projetos, como no caso *Jena TDB*³, um repositório que faz parte da estrutura do projeto Jena. Esse projeto permite o armazenamento de informações no modelo de dados RDF e também consultas complexas usando a ferramenta Jena Fuseki, a qual fornece uma interface para utilização da tecnologia SPARQL. Por sua vez, esta é uma linguagem de consulta eficiente para recuperação de dados em banco de dados RDF, bem como recomendada pela W3C.

³ Disponível em: <<https://jena.apache.org>>. Acesso em: 08 jan. 2019.

Pelo fato do repositório *Jena TDB* ser um projeto de uso irrestrito de código e de fácil distribuição, optou-se por adotar sua estrutura no desenvolvimento inicial do REFIBRA, uma vez que ela atende as necessidades iniciais do repositório e as suas finalidades, assunto apresentado e debatido na próxima seção.

HUMANIDADES DIGITAIS E O REPOSITÓRIO FILATÉLICO BRASILEIRO

É comum, mas não para todos, que a utilização do computador (hardware e software), articule, conecte, invente, produza, consuma e apague, por vezes de maneira tecnologicamente determinada, um conteúdo denominado como patrimônio cultural. Essas práticas rotineiras e mediadas por computador acontecem nas universidades, nas instituições memoriais (arquivos, museus e bibliotecas), nos centros de documentação, nos institutos de pesquisa, nas organizações públicas, no setor privado e na privacidade dos lares. Esse é o contexto contemporâneo no qual está circunscrito um campo de saber, para alguns de ordem científica, chamado de Humanidades Digitais (HDs).

Entretanto, ao contrário de outras áreas autodenominadas interdisciplinares, é consenso e reconhecido que o seu nascimento foi no ano de 1949 a partir do monumental estudo feito pelo jesuíta italiano Roberto Busa. A sua aventura intelectual tratou de fazer um *índice verborum* de todas as palavras encontradas nas obras de São Tomás de Aquino e autores relacionados, totalizando, assim, cerca de onze milhões de palavras no Latim medieval. Para conseguir trabalhar sobre essa coleção de termos, ele buscou auxílio de Thomas J. Watson, da International Business Machines (IBM), nos Estados Unidos. Estabelecidas as formas de ajuda, o estudo inteiro foi gradualmente transferido para cartões perfurados por computadores (HOCKEY, 2004). Vale lembrar que nesse período não havia sido cunhada a expressão ‘humanidades digitais’, mas já era comum a expressão ‘computação humana’ (em Inglês ‘*humanities computing*’ ou em Espanhol ‘*informática humanística*’).

Certamente, desde então, o desenvolvimento acelerado das tecnologias computacionais tem uma relação direta com as HDs. A sua institucionalização não para de expandir mundialmente, particularmente, em instituições universitárias,

por meio da pesquisa científica e de cursos que levam seu nome, bem como em institutos de pesquisa do setor privado e público.

Do ponto de vista do debate acadêmico, por exemplo, as conferências organizadas pela Aliança das Organizações de Humanidades Digitais são consideradas como marco da visibilidade internacional das HDs. A primeira conferência aconteceu em 1989 na cidade de Toronto, no Canadá, e segue anualmente até hoje. Em 2004, na Inglaterra, Schreibman, Siemens e Unsworth organizam e publicam um livro intitulado *A companion to digital humanitie*, considerado até hoje a obra seminal sobre esse campo.

Por sua vez, numa apresentação em Stanford, nos Estados Unidos, um importante estudo mostrou como eram entendidas e representadas as HDs em 134 currículos universitários distintos (SPIRO, 2011). Em 2015, Portugal organizou o Congresso Internacional de Humanidades Digitais, em que o debate entre o papel social das humanidades não poderia prescindir das transformações influenciadas pelas tecnologias digitais. No Brasil, em 2018, aconteceu o primeiro Congresso Internacional de Humanidades Digitais, no qual ficou evidente sua amplitude de possibilidades e de tendências tanto de pesquisas quanto de aplicações sociais e culturais.

No contexto do trabalho em andamento, as HDs têm amplitude conceitual: para alguns tem o sentido de uma disciplina em construção, para outros parece ser um conjunto de práticas que elenca a inserção das humanidades tanto no contexto da cultura digital quanto no debate científico do paradigma da pós-custódia. Articula-se com a Ciência da Informação num universo amplo e complexo, tanto a partir dos debates teóricos quanto sob uma perspectiva das práticas de organização, seleção, disseminação e curadoria digital. De fato, o debate epistemológico na Ciência da Informação não pode prescindir das práticas das Humanidades Digitais e dos constantes desafios demandados pela contemporânea cultura digital e sua forte característica interdisciplinar.

Nesse contexto, para o desenvolvimento do REFIBRA foi adotado o padrão de metadados Dublin Core, justificado pela sua aceitação internacional e capacidade de extensão. Um selo postal comemorativo brasileiro (Figura 1) ilustra o que foi exposto aqui até o momento.

FIGURA 1 – SELO POSTAL BRASILEIRO: CENTENÁRIO DE NASCIMENTO DE GILBERTO FREYRE (2000)



Fonte: Coleção particular do autor.

A partir da análise desse selo postal comemorativo foi definido um padrão mínimo, não exaustivo, de metadados apresentado na Lista 1:

Lista 1 - Relação dos metadados em Língua Inglesa e seus respectivos atributos em Língua Portuguesa

Title:	Atributo referente ao nome dado ao título do recurso, podendo o ser o nome oficial do selo postal ou batizado no momento do cadastro.
Subject:	Atributo envolve todos os assuntos que o selo postal trata, levando em consideração a imagem impressa no item e todo seu contexto relevante.
Relation:	Atributo referente a qualquer recurso que tenha relação com o selo postal e que não faça parte do assunto principal tratado pelo item.
Description:	Atributo referente a descrição extensiva e mais detalhada do contexto que o selo postal está envolvido, nesse atributo é envolvido a bagagem cultural e cognitiva do catalogado.
Creator:	Atributo referente entidade responsável pela elaboração do conteúdo do recurso. No presente estudo, todos os selos são de responsabilidade da Comissão filatélica nacional.
Publisher:	Atributo referente a instituição responsável pela difusão do selo postal. No presente estudo, todos os selos são de responsabilidade dos Correios.
Contributor:	Ao artista que fez a imagem estampada no selo postal, seja ela um desenhou, uma foto ou qualquer outra arte expressa.

Date:	Atributo referente ao ano de publicação do selo postal.
Identifier:	Atributo referente a identificação não ambígua selo postal.
Language:	Atributo referente ao idioma do conteúdo do selo postal.
Coverage:	Atributo referente a região a qual o selo postal pôde ser utilizado.
Paper Type:	Atributo referente tipo do papel utilizado no selo postal.
Color Type:	Atributo referente tipo de coloração do selo postal.
Filigree:	Atributo referente ao tipo de filigrana utilizado no selo postal.
Print Process:	Atributo referente ao processo de impressão utilizado para emitir o selo postal.
Print Location:	Atributo referente ao local de impressão do selo postal.
Stamp Dimension:	Atributo referente às dimensões do selo postal.
Stamps Issued:	Atributo referente ao número de selos emitidos.
Stamp Series:	Atributo referente ao nome da série que o selo postal faz parte.
Face Value:	Atributo referente ao valor facial estampado no selo postal.
Perforation:	Atributo referente dentição do selo postal

Por sua vez, na Lista 2, a seguir, os respectivos atributos estão indicados, alguns em formatos de *link*, considerada a leitura analítica do selo:

Lista 2 - Metadados em Língua Inglesa e seus atributos em Língua Portuguesa (*Links*)

Title:	Selo Comemorativo do Centenário de Gilberto Freire.
Subject:	< http://dbpedia.org/resource/Gilberto_Freyre >
Relation:	< http://dbpedia.org/resource/Sociology > < http://dbpedia.org/resource/Centenarian > < http://dbpedia.org/resource/Anthropology > < http://dbpedia.org/resource/History > < http://dbpedia.org/resource/Pernambuco >
Description:	Centenário do Nascimento de Gilberto Freyre
Creator:	Comissão Filatélica Nacional do Brasil
Publisher:	< http://dbpedia.org/resource/Correios >
Contributor:	Márcio Rocha
Date:	24.03.2000
Identifier:	S002
Language:	< http://dbpedia.org/resource/Portuguese_language >
Coverage:	< http://dbpedia.org/resource/Brasil >
Paper Type:	< http://dbpedia.org/resource/Coated_paper >
Color Type:	< http://dbpedia.org/resource/Polychrome >

Filigrana:	Não possui.
Print Process:	<http://dbpedia.org/resource/Offset_printing>
Print Location:	http://dbpedia.org/resource/Casa_da_Moeda_do_Brasil
Stamp Dimension:	55 x 22mm
Stamps Issued:	2.160.000
Stamp Series:	Não possui.
Face Value:	R\$ 0,36
Perforation:	11.5 x 12 mm

Determinado o padrão de metadados para esse selo postal, assim como um protótipo para os demais selos postais constituintes do REFIBRA, foi possível catalogá-los numa base de dados RDF. Para tal, como indicado na Figura 2, foi aplicado o *Jena Fueki* para a inserção das triplas RDF utilizando os metadados propostos. Vale ressaltar que estamos ilustrando, unicamente, os atributos *title* (<http://purl.org/dc/elements/1.1/title>) e *print location* (<http://metadadosrefibra/printLocation>). O valor desses *predicates* tem como seus respectivos *objects* (*Title* = Selo Comemorativo do Centenário de Gilberto Freire) e (*print location* = Casa da moeda do Brasil).

FIGURA 2 - INSERINDO OS PRIMEIROS METADADOS

```

1 ▾ INSERT DATA{
2   <http://refibra/resource/S001>
3     <http://purl.org/dc/elements/1.1/title>
4       "Selo comemorativo do Centenário do Nascimento de Gilberto Freyre" .
5   <http://refibra/resource/S001>
6     <http://metadadorefibra/printLocation>
7       <http://dbpedia.org/page/Casa_da_Moeda_do_Brasil>.
8 }

```

Fonte: O autor (2018).

Observe como na Figura 2, entre as linhas de código 1 e 8, estão os comandos em SPARQL para inserir triplas RDF no banco de dados *Jena TDB*. A primeira tripla inserida pode ser identificada entre as linhas 2 e 4. Na primeira linha citada está o valor “<http://refibra/resource/S001>” para representar o *subject* tratado; segue com “<http://purl.org/dc/elements/1.1/title>”, que representa o atributo *title* do Dublin Core e do padrão de metadados proposto e, por fim, o valor “Selo comemorativo do Centenário do Nascimento de Gilberto Freyre” para representar

o *object*. Nas linhas 5 a 8 é inserida outra tripla, na qual fica evidente que o valor do *object* do atributo “<http://metadadorefibra/printLocation>” é outro item semântico disponível em outra base de dados: a Dbpedia, da Wikipédia: (<http://dbpedia.org/page/Casa_da_Moeda_do_Brasil>).

O que está exposto na Figura 2 denota a maneira potencial de utilizar um padrão de metadados existente dentro de uma lógica de triplas RDF, bem como a forma de empregar o Linked Data, já que esse selo postal inserido na base de dados é um *link* semântico que direciona a máquina e/ou o usuário para uma outra base de dados, para um outro local de conexão em rede e compartilhamento de informação. Uma vez executados os comandos apresentados na Figura 2, o *Jena Fuseki* armazenará as triplas RDF em um banco de dados, a partir do qual o usuário poderá recuperar essas informações no futuro, nessa mesma interface, como apresentado na Figura 3 abaixo:

FIGURA 3 - SELEÇÃO DE TRIPLAS INSERIDAS

```

1 select ?subject ?predicate ?object
2 where {
3     ?subject ?predicate ?object
4 }
```

Fonte: O autor (2018).

Com a execução do comando apresentado na Figura 3, o *Jena Fuseki* apresenta o resultado da consulta como um quadro de três colunas contendo o valor do *subject*, *predicate* e *object*, respectivamente, como apresentado na Figura 4, a seguir:

FIGURA 4 - RESULTADO DA CONSULTA

	subject	predicate	object
1	<http://refibra/resource/S001>	<http://purl.org/dc/elements/1.1/title>	"Selo comemorativo do Centenário do Nascimento de Gilberto Freyre"
2	<http://refibra/resource/S001>	<http://metadadorefibra/printLocation>	<http://dbpedia.org/page/Casa_da_Moeda_do_Brasil>

Fonte: O autor (2018).

Como pode ser percebido ao analisar a Figura 4, o retorno da consulta realizada são as triplas RDF que foram cadastradas anteriormente. Isso significa que as tecnologias computacionais desenvolvidas e aplicadas no REFIBRA apresentaram resultados satisfatórios em relação ao objetivo da pesquisa em andamento. O REFIBRA, então, está emergindo como um ciberambiente baseado em tecnologia web semântica a partir do qual dezenas de tipos de documentos filatélicos convergirão, não, apenas entre si, mas com hiperlinks na rede mundial de computadores. Assim, o REFIBRA pode servir como modelo que articula áreas como a CI e as HDs no mesmo sentido em que Koltay (2016) aproxima processos investigativos no campo das Humanidades Digitais.

CONSIDERAÇÕES FINAIS

O projeto REFIBRA contribui para uma nova abordagem de se trabalhar itens informacionais na web pautada basicamente em aplicações de tecnologias computacionais articuladas ao campo interdisciplinar das Humanidades Digitais. Parte do que tem feito o Grupo de Pesquisa IHD da UFPE é, justamente, estimular que estudantes e professores de cursos como Biblioteconomia, Gestão da Informação, Comunicação, Design, Museologia e Filosofia sedimentem novas formas de conhecimento por meio de modelos de produção e uso de informação em ambientes ou espaços digitais. Os passos que estão sendo dados, a partir dos selos postais, não exclui outros documentos, sejam eles de natureza verbal ou verbovisual (SALCEDO, 2010).

O crédito que, como o nome do grupo indica, damos ao trabalho colaborativo e multidisciplinar tem sido salutar e positivo no desenvolvimento de dispositivos que constituem a estrutura desse ambiente digital, bem como na experiência de leituras, debates e convivência das pessoas que ali participam. O REFIBRA, então, sem que pudéssemos planejar, tem se constituído num dispositivo digital que oferecerá acesso irrestrito e aberto, por meio de Linked Data, ao conhecimento na rede mundial de computadores. De fato, uma experiência interdisciplinar que articula saberes diversos da Ciência da Informação e das Humanidades Digitais. Além disso, é salutar registrar que tudo isso demanda olhares éticos e reflexões críticas constantes sobre a nossa contemporaneidade, um complexo sistema que

não, apenas, é caracterizado por interfaces convergentes e ações de prossumidores, mas que também é um lugar de disputas embrenhadas no tecido social, lugar de concretizações individuais e de individuações.

Nesse sentido, temos um papel social de mediadores de informação, apesar de isso estar articulado com ensino, pesquisa e extensão na esfera acadêmico-científica (SALCEDO; CRUZ, 2017), e o REFIBRA uma função de mediação cultural (SALCEDO; PESSOA E SILVA, 2017), na medida em que serve como plataforma digital de patrimônio filatélico brasileiro conectado em rede com diversas e distintas bases de dados no mundo.

Outrossim, tanto o conjunto de dados abertos quanto as informações contidas nos documentos podem servir como amostra para futuras análises que busquem compreender de forma crítica como se dão as inter-relações socioculturais de produção, circulação e disseminação do conhecimento. Nesse contexto contemporâneo, é importante ressaltar que o REFIBRA, em que pese o desafio tecnológico que teremos adiante, articula três dimensões das inter-relações supracitadas, a saber: interatividade, hipertextualidade e transmediação (MIRANDA; SIMEÃO, 2005; COYLE, 2017). O trabalho prosseguirá com o processamento de 32 selos postais comemorativos relacionados a Pernambuco, depois 3.500 selos postais do tipo comemorativo brasileiro e, em seguida, mais de 5.000 selos postais diversos e distintos e, por fim, a documentação filatélica nacional.

Vale apontar para duas dimensões que, por meio do desenvolvimento do REFIBRA, indicam o futuro dos campos científicos supracitados. Uma dimensão diz respeito ao que chamaremos de solução técnica e a outra de solução teórica. Ambas estão articuladas e são mais complementares do que suplementares. As duas dimensões aludem ao futuro aplicado e teórico dos campos supracitados neste trabalho, no sentido de que algumas terminologias estão ganhando força e terreno no cenário internacional e nacional, a saber: curadoria digital, ciências da memória, estudos das interfaces, web semântica e, por fim, aquela terminologia que hoje, em vários lugares, está alcançando status de ciência: humanidades digitais.

Considerando a curadoria digital como um processo de interação e ciclo vital entre dados, informações, documentos, pessoas e instituições, desde a produção, passando pela distribuição até o seu consumo e reutilização, é possível articulá-la com as duas dimensões supracitadas num contemporâneo que designa certo momento histórico idiossincrático e que, em certa medida, traz para um conflituoso

debate alguns paradigmas consolidados nas áreas emergentes nos séculos XIX e XX, criando as condições necessárias para pesquisas futuras.

REFERÊNCIAS

ALTMAN, D. *Paper Ambassadors: the Politics on Stamps*. London, UK: Angus and Robertson, 1991.

BAYER, F. R.; NESE, L.L.; SCHROEDER, Rebeca. *ntSQL: um conversor de documentos RDF para SQL*. In: ENCONTRO REGIONAL DE BANCO DE DADOS, 9., 2015, Joinville. *Anais...* Joinville: Universidade do Estado de Santa Catarina, 2015.

BERNERS-LEE, T., LASSILA, O.; HENDLER, J. The Semantic Web: A New Form of Web Content that is Meaningful to Computers will Unleash a Revolution of New Possibilities. *Scientific America*, May 2001.

DUCHARME, B. *Learning SPARQL: Querying and Updating with SPARQL 1.1*. 2. ed. Sebastopol, CA: O'reilly, 2013.

CONEGLIAN, C. S.; SANTARÉM SEGUNDO, J. E. Europeana no Linked Open Data: conceitos de Web Semântica na dimensão aplicada das Humanidades Digitais. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 22, n. 48, p. 88-99, jan./abr. 2017.

COYLE, K. *Two FRBRs, Many Relationships*. 17 maio 2017. Disponível em: <<http://kcoyle.blogspot.com.br/2017/05>>. Acesso em: 29 maio 2017.

FERRY, L. *A revolução transumanista*. Barueri [SP]: Manole, 2018.

FLORIDI, L. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. United Kingdom: Oxford Press, 2014.

GRÁCIO, J. C. A. *Metadados para a descrição de recursos da Internet: o padrão Dublin Core, aplicações e a questão da interoperabilidade*. 2002. 127 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2002.

HEERY, R. Metadata Futures: Steps toward Semantic Interoperability. In: HILLMANN, D. I.; WESTBROOKS, E. L. (Ed.). *Metadata in Practice*. Chicago: American Library Association, 2004. p. 257-271.

HOCKEY, S. The History of Humanities Computing. In: SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Ed.). *Companion to Digital Humanities*. Oxford: Blackwell Publishing Professional, 2004. p. 3-19.

KOLTAY, T. Library and Information Science and the Digital Humanities: Perceived and Real Strengths and Weaknesses. *Journal of Documentation*, v. 72, n. 4, p. 781-792, 2016. Disponível em: <<https://bit.ly/2QLvZ60>>. Acesso em: 02 ago. 2018.

LIMA, J. C. de; CARVALHO, Cedric L. de. *Resource Description Framework (RDF)*. Goiânia: Instituto de Informática da Universidade Federal de Goiás, 2005.

MARCONDES, C. H. “Linked data” – dados interligados – e interoperabilidade entre arquivos, bibliotecas e museus na web. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, v. 17, n. 34, p.171-192, maio/ago. 2012.

MIRANDA, A.; SIMEÃO, E. (Org.). *Informação e tecnologia: conceitos e recortes*. Brasília: UNB, 2005.

MOLES, A. *O cartaz*. 2. ed. São Paulo: Perspectiva, 2004.

PEREIRA,W.; HEINRICH, T.; SCHROEDER, R. Avaliação de desempenho de sistemas relacionais para armazenamento de dados *RDF*. In: ENCONTRO REGIONAL DE BANCO DE DADOS, 12., 2016, Caxias do Sul. *Anais...* Caxias do Sul: UFRGS, 2016.

RAMALHO, R. A. Sá. Bibframe: modelo de dados interligados para bibliotecas. *Informação & Informação*, Londrina, v. 21, n. 2, p.292-306, 2016a.

RAMALHO, R. A. Sá. Análise dos modelos de dados SKOS e BIBFRAME: novas perspectivas de representação na era dos dados interligados. In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 17, 2016, Salvador. *Anais...* Salvador: Enancib, 2016b.

SALCEDO, D. A. *A ciência nos selos postais comemorativos brasileiros: 1900-2000*. Recife: EDUFPE, 2010.

SALCEDO, D. A.; CRUZ, M. Biblioteconomia, Ciência e Filosofia: um debate necessário sobre teoria e prática no campo acadêmico-científico. *Inf. & Soc.:Est.*, João Pessoa, v. 27, n. 1, p. 47-58, jan./abr. 2017.

SALCEDO, D. A.; PESSOA E SILVA, J. R. Disseminação da informação: o papel do bibliotecário-mediador. *Revista ACB: Biblioteconomia em Santa Catarina*, Florianópolis, v. 22, n. 1, p. 23-30, dez./mar. 2017.

SCHREIBMAN, S.; SIEMENS, R.; UNSWORTH, J. (Ed.). *Companion to Digital Humanities*. Oxford: Blackwell Publishing Professional, 2004.

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a Ciência da Informação. *Ciência da Informação*, Brasília, v. 33, n. 1, p. 132-141, jan./abr. 2004.

SANTOS, F. C.; CARVALHO, C. L. de. *Aplicações de suporte à Web Semântica*. Goiânia: Instituto de Informática da Universidade Federal de Goiás, 2007. 25 p.

SERRA, L.G.; SANTARÉM SEGUNDO, J. E. O catálogo da biblioteca e o linked data. *Em Questão*, Porto Alegre, v. 23, n. 2, p. 167-185, maio/ago. 2017.

SPIRO, L. *Making sense of 134 DH syllabi*. 2011. Disponível em: < <https://digitalscholarship.wordpress.com/2011/06/20/making-sense-of-134-dh-syllabi-dh-2011-presentation/>>. Acesso em: 06 mar. 2020.

W3C SPARQL Query Language for RDF. Disponível em: <<https://www.w3.org/TR/rdf-sparql-query/>>. Acesso em: 08 set. 2017.

A arte espectro neural no limiar das ciências e tecnologias: uma luta pela qualidade da vida

Luiza Helena Guimarães¹

Na luta pela vida, a arte, em conjunção com a ciência e as tecnologias, direciona-se ao universo das sensações humanas e, desse modo, questiona a ética e a liberdade. Ela carrega em si o poder de convocar o desejo, criar imaginários e gerar processos de subjetivação. Mas, é diante dos difíceis desafios, já presentes no final do século do XX, que neste século a arte vem sendo mais fortemente requisitada para defender a sobrevivência da espécie humana e a de nosso planeta.

A luta da arte implica na criação de modos de vida entrelaçados com tecnologias de *hardwares* e de *softwares*. Inclusive, visando melhorar a qualidade de vida, em vários domínios do saber e do poder dos corpos, crescentemente fazemos uso de implantes corporais ligados às tecnologias de comunicação e informação. Os avanços científicos e tecnológicos que estruturam a vida, estruturam a arte, que, por sua vez, entra na disputa com diversos mecanismos sociais pelo domínio de nossa percepção da realidade. Algumas tecnologias que aparecem direta ou indiretamente neste texto, podem servir de exemplo: inteligência artificial (AI), banco de dados, visualização de dados, imagens em 360° e 3D, ressonâncias eletromagnéticas, espectrometria, sonificação de dados, sons em 360° e espacializados.

Entre tantas possibilidades vindas dos campos da ciência e das tecnologias que interferem na realidade, a criação de novas e distintas interfaces para o mundo marca a quase indiscernibilidade da arte com o novo início do cinema,

¹ Artista e Pesquisadora. Pós-doutora em Comunicação e Cultura pela Universidade do Estado do Rio de Janeiro (UERJ). Doutora pelo Núcleo de Subjetividade da Pontifícia Universidade Católica de São Paulo (PUC-SP), com bolsa pela CAPES na Facultad de Educación Visual y Plástica y Comunicación Audiovisual e no Laboratorio de Medios Interactivos, ambos da Universidade de Barcelona, Espanha. Mestra em Comunicação e Cultura pela Escola de Comunicação da Universidade Federal do Rio de Janeiro (ECO/UFRJ). Criou e dirige o Laboratório de Arte Espectro Neural (LArtEN), Brasil.

por exemplo: Virtual Reality (VR), Augmented Reality (AR), Mixed Reality (MR), *fulldome*, *videomapping*, inclusive os Panoramas 360°, historicamente, anteriores ao cinema hollywoodiano, tradicional ou hegemônico. À medida que a arte se instrumentaliza e incorpora conhecimentos de outras áreas, ela ganha possibilidades de se constituir e existir de modos antes possíveis apenas na ficção científica.

Progressivamente nos é dada a possibilidade de realizar rupturas espaço-temporais, não traduzíveis por coordenadas geográficas e nem redutíveis ao tempo cronológico. Os dispositivos artísticos que são pensados e criados com estas características, normalmente propiciam ambientes interativos com simultâneas e múltiplas sequências narrativas, o que já os torna bastante interessantes. No entanto, os ambientes interativos podem ter como fio condutor narrativas sensoriais em ambientes imersivos, o que consideramos ser um desafio a mais para os criadores. De qualquer forma, diferentes modos de construção de narrativas, com maior ou menor grau de interatividade, são capazes de proporcionar ricas experiências imersivas de um único ou de multiusuários, em espaços reais, virtuais ou em ambos a um só tempo. Experimentos artísticos dessa natureza esbarram nos limites perceptivos dos cérebros humanos.

Este capítulo se desenvolve em torno da poética, da estética e de alguns aspectos científicos e tecnológicos que o projeto *Espectros Computacionais 360°* (EC360)² problematiza ou coloca em tensão. EC360 é um projeto transdisciplinar, imersivo e interativo, visual e sonoro, realizado a partir de ressonâncias eletromagnéticas de cérebros humanos. Contempla aspectos da neurociência, das tecnologias eletrônicas e digitais, da ciência de dados e da física

² Criamos um perfil no *Facebook* para o projeto com o intuito de que ele seja um canal aberto de comunicação de eventos, campanhas e, também, de registros dos processos de produção: <<https://www.facebook.com/espectros360/>>. Nele destacamos a matéria que saiu no jornal *O Globo*, disponível em: <https://bit.ly/3j0y8sq>. Acesso em: 20 ago. 2020; dois vídeos no qual apresentamos o projeto *Papo Sonica: Espectros Computacionais* – com palestra de Ricardo Dal Farra – Disponível em: <https://bit.ly/3hqwYpY>. Acesso em: 20 ago. 2020; *Espectros in process com Luiza Helena Guimarães*. Primeira parte: <https://bit.ly/3l8DCDn>, segunda parte: <https://bit.ly/2Yr0qV0>. Acesso em: 20 ago. 2020.

relativa a propagação de ondas eletromagnéticas. Ricardo Dal Farra³ e João Silveira⁴ integram o projeto que foi julgado e selecionado para ser desenvolvido durante a residência artística ArtSonica / LabSonica / Oi Futuro-RJ, 2018, com exposição no Centro Cultural Oi Futuro-RJ, 2019.

O projeto foi a continuidade de trabalhos com ressonâncias eletromagnéticas de cérebros humanos iniciados em 2002 e, na verdade, deve muito a eles em termos de estética e poética. No entanto, com o EC360 nos aproximamos de pesquisas neurocientíficas transdisciplinares, como a descrita por Francisco Teixeira, diretor do projeto Neurobios do Instituto de Neurociências de Lisboa. De acordo com ele, a conjugação de conhecimentos de áreas como neurociência, bioengenharia, AI e robótica permitiu o desenvolvimento de *brain-computer-interfaces* capazes de mostrar a atividade mental, expressa pelas ondas e ritmos cerebrais, em tempo real. Assim, fazendo uso de *neurofeedback* os seres humanos podem modelar sua atividade mental ao controlar estímulos visuais e sonoros. Aqui fica bem caracterizada a relação entre ondas cerebrais, luminosas e sonoras.

Diferentemente dos fins da clínica neurológica, no EC360 não existe a finalidade de obter qualquer espécie de controle mental, embora seja prevista a interferência na atividade mental de quem experimenta. Em paralelo com os conhecimentos científicos, as obras se dirigem ao campo das sensações. Guimarães trabalhou com o objetivo de criar uma instalação artística imersiva e um ambiente de VR. Ela partiu das ressonâncias eletromagnéticas que já haviam lhe inspirado sua primeira exposição com cérebros humanos, mas conforme referido acima, também de associações teóricas relativas aos conhecimentos de *neurofeedback* e da

³ Artista e pesquisador. Doutor em Artes pela UQAM, Canadá. Professor de música e midiarte na Concordia University, Canadá e diretor do Centro de Experimentación e Investigación en Artes Electrónicas (CEIARTE) da Universidad Nacional de Tres de Febrero, Argentina. Foi diretor do Hexagram, Canadá; pesquisador da UNESCO para o Projeto Digi-Arts, França; diretor artístico da bienal mexicana de Artes Eletrônicas: Transitio, em 2015. Membro do conselho editorial do Leonardo (MIT Press, USA) e do Organised Sound (Cambridge Press, UK). Membro do conselho diretivo do ISEA International. Criou e dirige o Simpósio Understanding Visual Music (UVM) e a Conferência Internacional Balance-Unbalance (BunB).

⁴ É consultor educacional, pesquisador, artista e empreendedor. É doutor em Educação, Gestão e Comunicação Científica pela UFRJ. Foi fellow na Universidade de Harvard e atualmente trabalha como especialista em aprendizagem na IFP School - Axens, França.

propagação de ondas eletromagnéticas⁵. Os distintos comprimentos e frequências de ondas eletromagnéticas representam distintas frequências luminosas, sonoras e cerebrais, passíveis de serem traduzidas, transformadas, metamorfoseadas em sons e cores que compõem o espectro luminoso e sonoro.

As obras expostas no Centro Cultural Oi Futuro constam de uma instalação com centenas de ressonâncias cerebrais e um ambiente imersivo e interativo de VR para óculos HTC Vive. Na etapa laboratorial, conforme o projeto, realizamos a espectrometria, a sonificação e o mapeamento de dados provenientes de ressonâncias eletromagnéticas cerebrais. Essa metodologia foi denominada por Guimarães de “espectro-cérebro-sônico-magnético”. Dal Farra escolheu os *softwares* para a realização do mapeamento de ondas, obtendo objetos e estruturas musicais. A parte sonora da exposição, constituída pelas composições eletroacústicas criadas para a instalação e ambiente VR, foi o resultado de uma complexa elaboração que envolveu a poética de trabalhos pré-existentes, o plano de trabalho e a metodologia de pesquisa existente no projeto de EC360, os resultados do mapeamento sonoro das imagens contidas nas ressonâncias eletromagnéticas e a liberdade de criação de Dal Farra, capaz de transformar a ciência em arte sonora.

A criação e concepção dos ambientes imersivos em 360° e 3D seguiu os mesmos princípios advindos da física de processamento de sinais, da neurociência e dos processos artísticos descritos para a obtenção das composições eletroacústicas. Rumando para a descrição do ambiente VR, as imagens em movimento foram desenvolvidas na plataforma da Epic Games em Unreal Engine, pelo engenheiro Randolpho Julião⁶. Optamos por usar essa ferramenta de criação 3D com renderização em tempo real, por ser a mais aberta e avançada do mundo.

⁵ No espectro luminoso cada uma das ondas oferece a sensação de uma cor que são associadas às frequências. As ondas mais longas estão localizadas mais próximas do vermelho e têm a frequência mais baixa. Por sua vez, as ondas mais curtas estão perto do azul e têm a frequência mais alta. O espectro sonoro é o conjunto de todas as ondas que compõem os sons audíveis e não audíveis pelo ser humano. Conseguimos apenas captar vibrações com frequências compreendidas entre os 20 Hz e os 20.000 Hz.

⁶ Analista de Sistemas, Administrador de Sistemas e Engenheiro de Software, especializado em Análise, Projeto e Gerência de Sistemas (APGS/PUC-RJ) e MBA em Gestão de Negócios (UERJ). Atualmente, dedica-se a pesquisa e desenvolvimento nas áreas de Virtual Production, Game Engines, Visão Computacional, Computação Gráfica, Realidades Aumentada, Virtual e Mixada.

Ela permite aos criadores muita liberdade e controle sobre o desenvolvimento de conteúdos tecnológicos de ponta para o cinema, *games* e mundos virtuais imersivos e interativos, por exemplo.

O engenheiro procurou entrar no universo de referência da artista e idealizadora de EC360°/VR e juntos foram dando vida aos seus ambientes abstratos, imersivos e interativos. Seguindo o *storyboard* do projeto para o ambiente VR, quem o experimenta navega por diferentes ambientes, denominados por Guimarães de “zonas sinápticas”. Assim: entrando pela “zona cinzenta” chegaremos a “zona do vazio”, de onde poderemos escolher vivenciar outras cinco diferentes “zonas sinápticas”, cada uma apresentando a equivalência entre frequências de ondas luminosas e sonoras. Por sua vez, elas nos levam para a “zona escura ou do medo” e assim por diante. Em cada zona um ambiente com diferentes sons e imagens gráficas cerebrais em movimento, evocando em nós distintas sensações cerebrais e corporais.

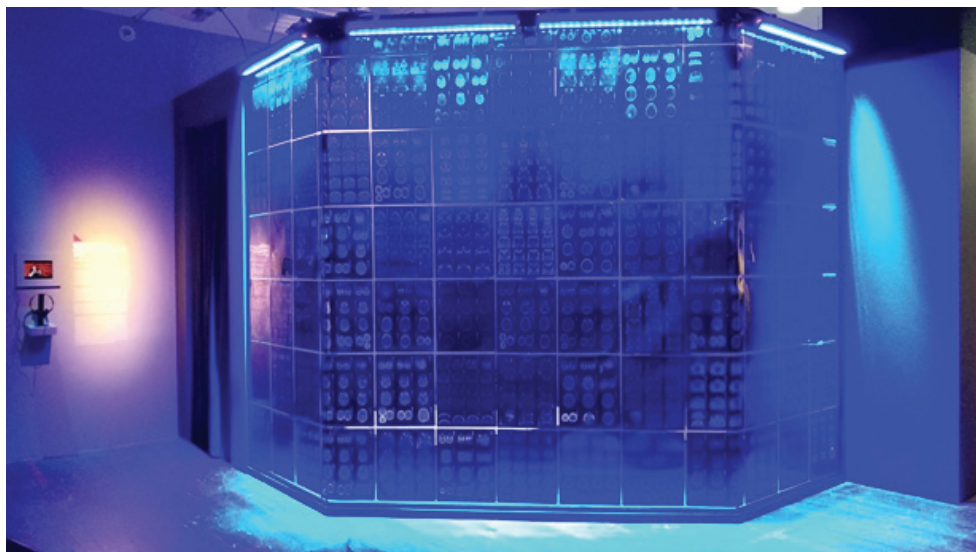
Em nossos cérebros, os estados mentais são alterados diante de imagens e sons dos ambientes em que nos inserimos. Embora os ambientes de EC360 sejam planejados, não buscamos o controle cerebral e sim tornar visível e sensível, por meio dos conhecimentos da arte, as forças que nos afetam cotidianamente e escapam a nossa percepção. Durante o processo de criação dos audiovisuais para VR, juntamente com desenvolvedor, Guimarães foi trabalhando-os no sentido de diferenciar e equiparar as frequências das ondas cerebrais (beta, alfa, teta, delta e gama) conforme as sensações corporais experimentadas. Fez isso em termos de cores, luzes, sombras e texturas; de velocidades, deslocamentos e ritmos entre as imagens; de profundidade de campo e assim por diante. Visando sempre o resultado esperado, buscamos conseguir criar um campo de afetos capazes de tornar sensível a potência de criação em outros corpos. Entrando assim, no campo da produção de subjetividade.

Quanto a parte sonora dos ambientes de VR, Guimarães seguiu o mesmo processo de trabalho e pensamento acima descrito, isto é, buscou a diferenciação e equiparação entre frequências e ondas cerebrais, luminosas e sonoras, relacionando-as aos afetos sentidos, primordialmente, nos corpos dos criadores. A artista tinha em mente as sensações que gostaria de alcançar e provocar em cada zona sináptica. Em consenso de ideias, Dal Farra criou diferentes composições para as diferentes zonas projetadas.

Após ter realizado suas obras sonoras, ele resolveu pedir para Guimarães escutá-las, todavia, sem o prévio conhecimento da zona para qual haviam sido criadas. Foi uma experiência bastante marcante durante a etapa laboratorial. Significava um grande teste e deu certo! Não por coincidência, durante a escuta, ela conseguiu distinguir as sensações provocadas pela música, associando de imediato as peças sonoras apresentadas às zonas sinápticas projetadas. A música de Dal Farra provou ser capaz de se materializar no imaginário, nos sentidos e nas sensações.

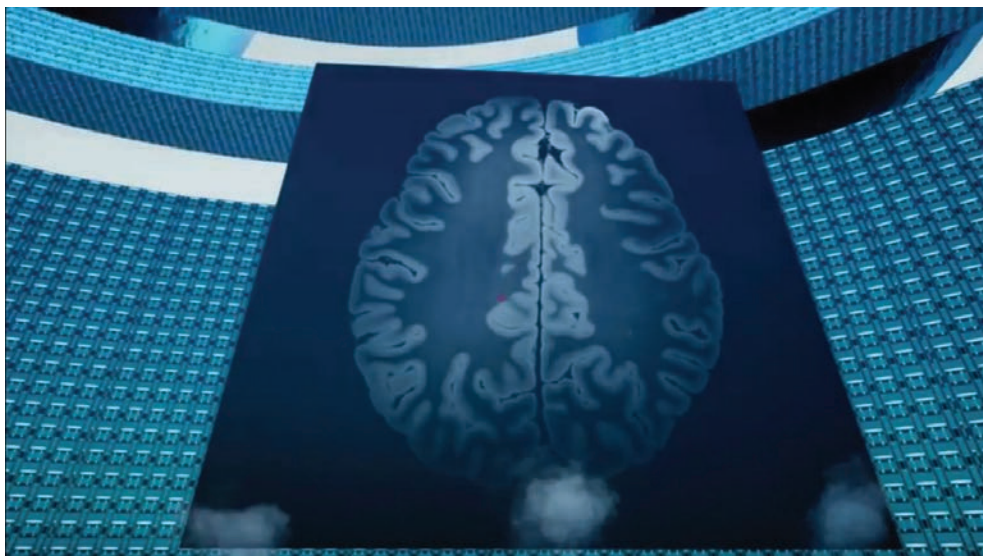
Portanto, uma grande etapa se cumpriu, pois percorremos por conhecimentos científicos, realizamos escolhas entre inúmeras tecnologias, desde os *softwares* utilizados na construção do mapeamento “espectro-cérebro-sônico-magnético” e do ambiente VR, passando pela escolha dos equipamentos visuais e sonoros, chegando, juntamente com uma grande equipe de produção, até a montagem e abertura da exposição. Tudo isso, sem nos afastarmos do campo da arte (Ver imagens 1, 2 e 3).

IMAGEM 1 - “ESPECTROS COMPUTACIONAIS 360º”, AMBIENTE VR, *EXPOSIÇÃO ARTSONICA*



Fonte: Centro Cultural Oi Futuro, Luiza Helena Guimarães, Rio de Janeiro, Brasil, 2019.

IMAGEM 2 - “ESPECTROS COMPUTACIONAIS 360°”, AMBIENTE VR, *EXPOSIÇÃO ARTSONICA*



Fonte: Centro Cultural Oi Futuro, Luiza Helena Guimarães, Rio de Janeiro, Brasil, 2019.

IMAGEM 3 - “ESPECTROS COMPUTACIONAIS 360°”, AMBIENTE VR, IMAGEM DA “ZONA SINÁPTICA” AZUL, *EXPOSIÇÃO ARTSONICA*



Fonte: Centro Cultural Oi Futuro, Luiza Helena Guimarães, Rio de Janeiro, Brasil, 2019.

Sinteticamente, os ambientes se basearam em conhecimentos da física e da neurociência, estabeleceram determinadas relações científicas e tecnológicas entre zonas cerebrais, ritmos e ondas eletromagnéticas, contudo, seguiram primordialmente o imaginário dos artistas. A sinopse que colocamos na exposição foi a mesma escrita por Guimarães em seu projeto:

Você se sentirá imerso em imagens provenientes de cérebros semelhantes ao seu, na música de sinais sonoros por eles emitidos. Trata-se de um ambiente computadorizado em 3D e 360° que emerge do conhecimento das tecnologias e da neurociência, mas também da arte que busca tornar sensível e dizível o universo das sensações e da cognição humana e maquínica.

O conceito poético-científico da exposição se refere, por um lado, aos dados sonoros que se tornam visíveis em cores, luzes e sombras, e, por outro, imagens cerebrais se tornam audíveis. No universo dos dados, da AI, as imagens são submetidas à padrões e convertidas em novas paisagens visuais e sonoras. Nas atualizações, entre o real e o virtual, acontecem incorporações e metamorfoses, propagam-se espectros cerebrais, luminosos, sonoros e eletromagnéticos, irradiados por ondas de sons-imagens-pensamentos.

Assim, surgem as questões: que espectros se irradiam e pairam nos maquinismos entre o humano e o tecnológico? Entre cérebros e mentes, *hardwares e softwares*, como se produz o futuro? Seria pela produção de imaginários como uma potência do falso ou pela verdade científica trazida pelos dados de predição de futuros possíveis? Em debate, está a permeabilidade entre os campos da ciência e da arte, do humano e do tecnológico, modulados pelos mecanismos de AI.

O entendimento da poética e estética de EC360 está em sua origem. Fazendo uso de ressonâncias magnéticas de cérebros, em 2002, Luiza H. Guimarães realizou a exposição *Senha-Teu Limite* na Galeria Mira Schendel – Rio de Janeiro - RJ; em 2004, *Catedral de Cérebros* na Galeria de Arte da Universidade Federal Fluminense (UFF) – Niterói - RJ. O curador da exposição, Pierre Caprez, inspirado pelo ambiente da instalação e com base em conhecimentos filosóficos, escreveu sobre o trabalho no catálogo da mostra *A poética da miscigenação: para uma floração dos signos* (2004). Sobre essas duas instalações, Santaella (2018, p. 40) faz as seguintes observações:

A primeira impressão que disso resulta é a dos vitrais das catedrais que filtram a luz através de seus desenhos. São cérebros-vitrais, a metáfora poética é evidente. Entretanto, para além ou aquém da metáfora, a profusão de imagens cerebrais é perturbadora porque expõe, no nu e cru da luz invasiva, a penetração pela ciência e tecnologia nos recessos mais íntimos do nosso corpo. Nesse caso, o mais íntimo entre os íntimos, o cérebro, a sede pensante e senciente do corpo, paradoxalmente exposta em uma multiplicidade agigantada que lembra (outra metáfora) a mente coletiva e planetária de cérebros interligados graças às redes comunicacionais da internet. As metáforas se desdobram. Basta para isso se colocar dentro do movimento da passagem de uma imagem dos vitrais a outra nas sucessivas telas ou, então, entregar-se à experiência de vivenciar o espaço da instalação, não como mero observador distanciado, mas na simbiose com o arquicérebro de que somos indistinta parte.

As instalações citadas podem ser pensadas como sendo de cinema, pois apesar de serem imagens fixas dispostas lado a lado em vidros, as imagens cerebrais se projetavam no interior das galerias e se movimentavam mediante o deslocamento luminosidade natural e entre o nascer e o pôr-do-sol. Ou mesmo, no caso da galeria da UFF, quando a noite chegava, *spots* instalados do lado de fora projetavam os cérebros nas paredes e no teto, lembrando as pinturas de catedrais, só que com imagens do interior do corpo humano. A instalação de EC360 segue essa mesma lógica acrescida da expansão da linguagem, conforme descrevemos até aqui. Trata-se, deste modo, de arte baseada no tempo, tanto no que diz respeito à imersão quanto na interatividade.

Reunindo cinema e neurociência, Weibel (2004) prevê um possível cinema do futuro, nomeia-o de neurocinema e define-o como o cinema da recepção de estímulos diretamente no cérebro. O autor parte da ideia que o cinema do século XIX havia sido desenvolvido por engenheiros e fisiologistas, enquanto que físicos e neurocientistas geram as tecnologias cinematográficas para o século XXI. Seguindo por essa linha de raciocínio, ou seja, da evolução dos meios tecnológicos e dos conhecimentos científicos, Weibel diz que no século XIX as câmeras tinham a função de capturar e analisar os movimentos, e os projetores de sintetizá-los. Mesmo perante os avanços da fisiologia e da psicologia experimental da época, relativos aos mecanismos da percepção humana, as máquinas ainda se limitaram a simular

visualmente o movimento. Assim, o cinema que começou preocupado com a visão, firma-se no século XX com cinema hollywoodiano e suas máquinas de movimento.

No entanto, a intenção original do cinema de criar máquinas de visão e investigar como se dá a percepção no cérebro é retomada de tempos em tempos. Existiu nas vanguardas dos anos vinte, cinquenta e sessenta, e voltou com o digital. A busca da percepção cerebral já havia produzido algumas das mudanças mais profundas na pintura, primeiro com Cézanne, a fim de obter a “sensação visual” que chamava *petite sensation*, e depois com Picasso, em sua busca de representar a quarta dimensão, o Tempo. A descoberta dos mecanismos cerebrais vem sendo uma constante inquietação da humanidade, o que se reflete nas tecnologias, pois o modo de visualização vai se aproximando das descobertas da neurociência e da ciência cognitiva, natural e artificial.

Essa busca também se encontra na história da cibernética. Em 1947, Wiener (1954) fez um estudo comparativo entre homens e máquinas envolvendo as áreas de comunicação e controle voltadas para sistemas de informação: retroalimentação, estabilidade, regulação, codificação e ruído. Para Wiener, tanto quanto os organismos vivos, as informações mediadas pelas máquinas adquirem uma forma adequada à natureza complexa das comunicações na sociedade. Com o mecanismo de *feedback*, a informação se volta para dominar a entropia, isto através da organização e ajuste intencional de uma sequência temporal entre eventos. Ele queria que as máquinas simulassem o pensamento e a vida, mais especificamente, as formas de comportamento, deste modo, suas investigações recaíram sobre o cérebro.

Com base em seus estudos, em 1953, W. Grey Walter cria a *Machina Speculatrix*⁷ um primeiro exemplar dos sistemas que tentam imitar o processo de pensamentos humanos, mas que já exibia um comportamento complexo. Walter tinha por intuito provar que quanto maior o número de conexões entre os neurônios cerebrais, mais complexo se torna o comportamento. Seus primeiros robôs, os *tortoises*, começavam a desvendar os segredos do funcionamento cerebral. Ele usava circuitos puramente analógicos para simular processos do cérebro, enquanto isso, cientistas como Alan Turing e John Von Neumann estavam voltados para a visão de processos mentais em termos de computação digital.

⁷ Disponível em: <<https://bit.ly/3ghOE5C>>. Acesso em: 20 ago. 2020.

Como Wiener (1954), Leibnitz também se preocupava com a mensagem em termos comunicacionais. Pensava a totalidade do mundo como uma coleção de seres chamados “mônodas” e a interação mecânica entre elas como consequência de uma interação ótica, isto é, da comunicação por imagens visuais. A linguagem da computação, um cálculo lógico ou *Calculus Ratiocinatur*, que antecedeu a lógica matemática, é uma extensão da ideia de linguagem artificial, analisada por Leibnitz a partir da visão da ótica. Tanto ele quanto Wiener, estavam interessados na comunicação através de máquinas, entre autônomos. No entanto, Wiener, iguala homens e máquinas pelo processo da homeostase: ou os seres vivos resistem ao fluxo de desintegração pelo processo de homeostase ou morrem, já que mecanismos homeostáticos do homem obedecem a padrões de *feedback* do mesmo modo dos descritos para autômatos. Corpos, cérebros e máquinas reproduzem padrões, dizia ele:

É divertido, tanto como instrutivo imaginarmos o padrão global do corpo humano, do cérebro humano com suas recordações e suas conexões entrecruzadas, de modo que um hipotético instrumento receptor pudesse reincorporar tais mensagens numa matéria apropriada, capazes de dar continuidade a processos já em curso no corpo e na mente, e de manter a integridade necessária a tal continuidade por via de um processo de homeostase (WIENER, 1954, p.95).

Na imaginação de Wiener, relativas a um padrão global do cérebro humano, ele pensava estar invadindo o terreno da ficção científica, mas compreendeu que a transmissão de comunicações por mensagens (de ideias e de linguagens) é uma extensão do domínio e da percepção humana. Assim, a teoria dos sistemas dinâmicos, cibernética, começando com a ideia de *feedback*, novamente tenta redefinir o conceito de humano através de sua equiparação as máquinas, pois ambos podem ser vistos como processadores de informação.

Donna Haraway (2018) radicaliza esta teoria em direção ao ciborgue, Katherine Hayles (2008) define informação de acordo com Claude Shannon que, por sua vez, baseou-se em Wiener. Sendo equiparados as máquinas, somos programáveis, mas será que somos passíveis de controle e de comando do mesmo modo que elas? Colocados diante da possibilidade das máquinas programarem os homens, mais importante do que se tentar saber o quanto estamos nos tornando inumanos, maquínicos, ciborgues, faz-se necessário entender os devires entre

humanos e máquinas. Então, a questão que se impõe seria, portanto, sobre a digitalização do desejo, imaterial e efêmero.

No relato pelos caminhos para revelar o funcionamento do cérebro, incluímos mais dois acontecimentos que marcaram a ficção e a realidade. De *Neuromancer* (2015), escrito em 1984, no qual o ciberespaço da ficção gibsoniana poderia ser habitado por corpos imateriais, pois os corpos, quando conectados a implantes neurais, desconectariam-se do mundo físico e passavam a habitar o mundo da informação e da simulação, até, por outra via de pensamento, chegarmos a *World Wide Web* (WWW), 1989, desenvolvida pelo cientista de computação Tim Berners-Lee, no Centro Europeu de Energia Nuclear (CERN), na Suíça. A WWW foi criada como uma ferramenta que permitiria compartilhar informações acadêmicas, mas veio a formar a base do funcionamento da rede mundial de computadores na internet, reunindo vários aspectos da vida global. A internet se configurou em um grande, em termos de qualidade e quantidade, banco de dados que guarda a memória da civilização e que pode ser acessado de modo descontínuo, rizomaticamente, semelhante ao modo que o cérebro se movimenta no tempo.

As tecnologias digitais, que se desenvolveram a partir de 1940 com a invenção do computador digital, foram mudando as formas de comunicação e de subjetivação. Com elas se transformaram a economia, a política, as relações sociais, a ecologia, etc. Entre homens e máquinas os desenvolvimentos das tecnologias e das linguagens acontecem sob o viés de uma dupla efetuação. No entanto, a capacidade humana de criar e mudar o rumo dos acontecimentos vem sendo cada vez mais requerida.

Grusin (2015) diz que entre os problemas enfrentamos no século XXI estão aqueles que envolvem os não humanos, tais como: a mudança climática, a seca e a fome; a biotecnologia, a propriedade intelectual e a privacidade; o genocídio, o terrorismo e a guerra. O autor engloba vários fatores para fundamentar a reviravolta não humana. Ela é compreendida de várias maneiras, como animais, afetividade, corpos, tecnologias, sistemas orgânicos e geofísicos, entre outros. Até mesmo o novo paradigma do Antropoceno, que nomeia o humano como influência dominante sobre o clima desde o industrialismo, reconhece que os seres humanos devem ser entendidos como forças climáticas ou geológicas do planeta, pois operam da mesma forma que os não humanos, independentemente da vontade humana, de suas crenças ou desejos.

Grusin (2015), em sua argumentação, chega ao mesmo ponto que venho defendendo em relação “a luta da arte”. Embora nem sempre seja fácil perceber como os afetos são importantes para a reviravolta não humana, nas obras de Deleuze já existia a ideia que os sistemas de afetos, como capacidade de afetar e ser afetado, acontecem nas relações entre os corpos. Tomando o caminho dos sistemas de afetos, eles lançam luz sobre o processo de criação de tecnologias e linguagens em tensão com as forças do mundo sentidas nos corpos. Como indica Grusin (2015), os afetos podem servir para marcar o sentido de reviravoltas e revoluções.

Sendo assim, como convocar a potência criadora quando estamos sendo demandados de tantos modos para replicar padrões comportamentos? Como sair dos automatismos característico do território das representações e criar a partir dos afetos sentidos nos corpos? São questões fundamentais para pensarmos o espaço das tecnologias e das linguagens. Lembrando que o território intensivo da criação é o do presente vivo, tempo do vir a ser, que é o sentido encarnado de um movimento que, em Deleuze (2007), é entendido como tornar-se.

A estética dos dispositivos artísticos ativa a produção de afetos micropolíticos, ou seja, diz respeito a política de subjetivação da arte como ética da construção de si e do mundo. Em *Senha-Teu Limite* ou em *Catedral de Cérebros*, já citadas, se a obra é capaz de ativar campos relacionais ou processuais, instaura territórios potenciais e múltiplos, capturando outros corpos para formar “espaços catedrais de mentes”. Na instalação EC360 criada para o Centro Cultural Oi Futuro, as imagens e sons de cérebros são o limite entre o dentro e o fora, isto é, apresentam-se como uma membrana entre os movimentos contínuos de cérebros biológicos e de imagens e sons cerebrais. Para além de efeitos da luz e sombras ou sonoros, o que se move através do espaço e do tempo são os pensamentos.

No cinema, segundo Deleuze (1986), em entrevista para os *Cahiers du cinéma*, afirma que “o cérebro é a tela”. Ele faz uma analogia entre a biologia do cérebro e o pensamento, pois ambos são moleculares e constituídos por velocidades moleculares, o que estabeleceria um forte contraste com os seres naturalmente lentos que somos. O cinema, ao colocar a imagem em movimento, extrapola-a em direção às associações entre imagens. Ao envolver os circuitos cerebrais, o cinema induz a um automovimento e ganha vida em encadeamentos e reencadeamentos. Produz histórias com a velocidades do pensamento e coloca em curso processos de subjetivação individuais e coletivos.

Se, por um lado, o filósofo Deleuze buscou no cinema a elaboração de conceitos que respondessem suas inquietações, por outro, os artistas e os cientistas seguem confrontando e desafiando, com suas próprias ferramentas, a imaterialidade e a materialidade do mundo em que vivem. São nos encontros e embates com a realidade sensível (invisível e indizível / visível e dizível) que se estabelece, no limiar entre as disciplinas, o campo da transdisciplinaridade.

Deslocando-nos entre os seres humanos, seus cérebros, suas tecnologias e sua ciência, tensionados pela arte, alcançamos um dos caminhos para investigar os modos de aumentar o grau de imersividade nas experiências audiovisuais. É um segmento de pesquisa que configura os principais problemas para o cinema do século XXI. Evidencia que os limites para a criação de experiências em ambientes imersivos deixam de ser tecnológicos e passam a ser da neurociência. O papel preponderante da neurociência para o cinema do futuro já havia sido pontuado por Weibel (2004). Ela se esforça em desvendar o funcionamento cognitivo cerebral e explorar a percepção humana. Em dispositivos VR, por exemplo, a sensação de presença, um termo técnico que define o quão imersiva é uma experiência, tem sido uma das grandes preocupações dos desenvolvedores.

As criações para ambientes interativos e imersivos, como Panoramas 360°, *fulldome*, VR, AR e MR buscam, no conhecimento cognitivo sobre como se processam as percepções e as sensações no cérebro, descobrir maneiras de proporcionar uma experiência de maior qualidade. Por exemplo, usando óculos VR, compreendemos na prática que o tempo de resposta à estímulos da parte lógica e da parte instintiva do cérebro são diferentes, pois a informação visual que chega na retina é percebida pelos instintos milissegundos antes do cérebro poder racionalizá-la. Esse pequeníssimo intervalo de tempo enriquece a experiência em VR por ser capaz de enganar o cérebro fazendo parecer que existimos verdadeiramente em outro espaço e tempo. A qualidade da experiência imersiva e sensação de presença também dependem de o dispositivo conseguir ativar nossas memórias fazendo-as entrar em composição com as imagens, sons e sensações táteis em tempo presente.

As memórias do passado diante da vivência do presente em ambiente de VR se confundem no cérebro. No entanto, com base no que ele já sabe e no que apreende de imediato, logo consegue traçar as diferenças entre o mundo físico e a realidade simulada. O fato da tela dos óculos de VR estar muito próxima aos olhos é outro fator relevante para a diminuição da sensação de presença. Nesse

caso, o cérebro detecta a falta de profundidade de campo e rejeita o cenário como sendo real. Contudo, as experiências imersivas e interativas em VR proporcionam sensações bem mais intensas do que as do mundo físico.

Nas experiências com óculos de VR, ainda podem ocorrer problemas de latência, *delay*, que causa enjoos ao movermos a cabeça ou visualizarmos passagens entre quadros. A explicação fisiológica para isso diz que os sistemas sensorial e motor são acoplados, gerando um descompasso entre os movimentos que sentimos nos ouvidos internos e a imagem que enxergamos.

Os conflitos perceptivos da passagem do mundo físico para o óculos de VR continuam. Existe, também, o conflito vergência-acomodação dos olhos. Ele é causado pela falta de integração, pois independentemente do lugar onde os olhos focarem as imagens, na verdade, toda a tela está a poucos centímetros dos olhos. Em alguns óculos, os objetos renderizados já atingem uma profundidade mais próxima do real, entrando e saindo de foco.

Além disso, mesmo sem alterar muito nosso senso de orientação espacial, a visão periférica e a central seguem caminhos separados em direção ao cérebro. Todavia, está na integração entre distintas áreas a resolução para os problemas do cinema que se inicia em nossos dias.

Assim, demonstramos que os ambientes de alta imersividade precisam ser pensados como um sistema vivo integrado ao cérebro. Os resultados obtidos nas criações para ambientes de VR produzem experiências magníficas para seus jogadores ou performers, termo mais adequado aos trabalhos artísticos. Com esses conhecimentos os desenvolvedores buscam na fase de criação se alicerçar dos melhores *softwares* e *hardwares*, no entanto, é necessário escolher dispositivos VR de qualidade para garantir a experiência imersiva e interativa do trabalho realizado.

A criação do ambiente de VR do projeto EC360 apresenta um importante diferencial em relação a maior parte da produção artística da atualidade, pois elas normalmente seguem a orientação das narrativas sequenciais com base em paisagens reais, como a dos games. No ambiente criado de EC360/VR visamos as sensações “cérebro-corporais” e trabalhamos com imagens de bancos de dados científicos (ressonâncias magnéticas de altíssima definição e neurônios modelados a partir dos existentes no cérebro humano) para criarmos narrativas sensoriais. Logo, os audiovisuais imersivos e interativos se dirigem ao campo de afetos puramente sonoros e visuais.

EC360 despertou mais fortemente a vontade de ampliar as discussões em torno dos temas que ele coloca em pauta, como a AI e a cognição. Também, por meio de ECO360, vivenciamos o fato de estarmos inseridos na construção de grandes mudanças estão em curso na humanidade. Por conseguinte, nasceu o Laboratório de Arte Espectro Neural (LArtEN)⁸. Foi idealizado por Luiza Helena Guimarães e recebeu a mentoria do LABORA/Oi Futuro (2018). Tem entre seus membros mais ativos Ricardo Dal Farra (já citado), Fernando Ferreira (parceiro de projetos anteriores, engenheiro, diretor da empresa Twist de IA), Eric Barbosa (músico, cineasta e produtor), Flávia Werneck (designer e professora) e Luiza Helena Guimarães (autora desse capítulo).

O LArtEN é um laboratório de arte imersiva e interativa, experimental, transdisciplinar e descentralizado. Dedicase a criação, desenvolvimento e produção de “Arte Espectro Neural”, termo criado para marcar a arte e o cinema experimental do século XXI. A Arte Espectro Neural se caracteriza por requerer aptidões inéditas dos cérebros humanos e ser produzida concomitante as descobertas de físicos e neurocientista. Na prática, O LArtEN se dedica às pesquisas, estudos e desenvolvimentos de tecnologias para instalações artísticas e ambientes imersivos e interativos de sons e imagens, como VR, AR, MR, Panoramas 360°, *fulldome*, entre outras.

O termo Arte Espectro Neural revela mais diretamente as preocupações, problemas e projetos do laboratório. Utilizamos a palavra “arte” como referência ao campo de estudos capaz de tornar sensível um modo de conhecimento singular, que não é alcançado por meio da ciência e das tecnologias. Já a instigante e tão controversa palavra “espectro”, aponta para o que no LArtEN acreditamos ser a especificidade do campo da arte diante das outras disciplinas, ou seja, para ser arte deve ter a capacidade de carregar em si a potência, invisível e indizível, de afetar outros corpos. Esse é o poder da arte! Ou seja, o de ocupar o campo dos acontecimentos incorporais, o vir a ser, o devir animal. Coincidentemente, “espectro” também é utilizado na psicanálise, no cinema e na física de processamento de sinais. Em todos os casos citados, “espectro” se mostra sempre como potência de ativação de um campo dos afetos.

⁸ Disponível em: <<https://bit.ly/3aRKNLI>>. Acesso em: 20 ago. 2020.

Disponível em: <<https://www.facebook.com/lartenlab/>>. Acesso em: 20 ago. 2020.

Ainda sobre o nome do laboratório, por fim, a palavra “neural” que interconecta distintos campos de conhecimentos que se dedicam a dar conta do “humano” da contemporaneidade. O projeto EC360 foi a inspiração, a origem e serve para exemplificar projetos transdisciplinares de arte e ciências. Conforme dissemos, em síntese, a “Arte Espectro Neural” se distingue tanto por requerer aptidões inéditas dos cérebros humanos quanto por estar sendo criada e produzida em paralelo com as descobertas de físicos e neurocientista em um fluxo contínuo de trocas e atravessamentos.

Logo ao início afirmamos que a arte vem sendo solicitada para defender a vida, abrangendo a do planeta. Por conseguinte, os projetos do LArtEN se dividem em duas grandes linhas temáticas, ambas com caráter emergencial e intrinsecamente relacionados. A primeira diz respeito as alterações ambientais mundiais e produção de subjetividade e larga escala. Está alicerçada no dispositivo de arte e clima *#EntrancedEarthPanorama*⁹, já realizado, mas que continua sendo desenvolvido. A segunda linha é justamente a da arte e neurociência que estamos discutindo neste capítulo.

Conhecendo os graves problemas políticos, econômicos e sociais mundiais, também para Santaella (2018), existem dois temas centrais que convergem para as preocupações internacionais e artísticas, são eles, justamente, os campos das mudanças ambientais e climáticas e o da neurociência:

Tendo em vista a ecologia e as questões climáticas que, sob o nome de antropoceno (PARIKKA, 2014; YUSSOF, 2017), estão hoje no centro das preocupações internacionais, há artistas que têm se envolvido com trabalhos que exploram uma noção expandida da ecologia em tempos de emergência. Outro campo que tem chamado atenção recentemente é o da neuroestética e neuroarte. A neuro-estética é entendida como “o estudo científico das bases neuronais para a contemplação e a criação de obras de arte” (MACCLURE e SIEGEL, 2015). Os avanços nas neurociências e neuroimagens têm atraído os artistas para os variados aspectos que essas ciências revelam sobre o cérebro (p. 38).

Se assim for, a convergência de ideias vindas desses dois campos se traduz pelo fato de que nossas decisões individuais e coletivas tem o poder de mudar a sorte da

⁹ Cf. GUIMARÃES, 2016, 2018a, 2018b; SANTAELLA, 2018

humanidade: seu modo de existir, sua sobrevivência como espécie e o destino da Terra.

No início de 2019, a IBM lançou para demonstração, na Feira de Produtos Eletrônicos de Consumo (CES) em Las Vegas, o primeiro computador quântico. Ele possuía algumas das tecnologias mais avançadas do mundo dentro de um cubo de vidro, com mais ou menos três metros quadrados, capaz de condensar o imaginário de muitos campos de conhecimentos, as pesquisas de ponta de AI e sobre cognição, os paradoxos da teoria quântica etc. De tal forma, o silício dos transistores dos computadores poderá ser substituído pelos moleculares, o que na filosofia, na arte e na ficção científica já havia sido pensado de distintas maneiras. No entanto, está em curso uma transição, talvez, a maior da história humana. Em contrapartida, está longe de o horizonte científico decifrar e dominar a complexidade da mente humana. Mergulhados em sinapses neuronais de velocidades moleculares, os humanos vão interagindo e criando na emergência do vir a ser.

Em *O Ato de Criação*, Deleuze (1999) fala sobre a relação existente entre o trabalho de arte e o ato de resistência. Deleuze argumenta a partir de um conceito de André Malraux, escritor e diretor francês (1901-1976), para quem um trabalho de arte *“é a única coisa que resiste à morte”*. Ele concorda, mas acrescenta: o trabalho de arte resiste à morte, embora não seja a única coisa. O fato de resistir é o que existe em comum entre o trabalho de arte e o ato de resistência. Embora nem todo o trabalho de arte resista e nem todo o ato seja de resistência, há uma estreita relação que os perpassa. Existe uma dupla face no ato de resistência, ele é ato humano e de arte: ambos resistem à morte e a sua relação é misteriosa, declara Deleuze.

Desta vez, ele cita a afirmação do artista Paul Klee: *“pois bem, falta o povo”*. O que significa dizer falta o povo? Quando um trabalho de arte é criado ainda não existe um povo preparado para entendê-lo, o que falta é uma afinidade primordial entre o trabalho artístico e “o povo”. Mesmo que essa afinidade possa jamais vir a existir, todo trabalho de arte fala para um povo por vir. Sendo assim, o trabalho artístico é um ato humano, um ato de resistência e um apelo a um povo por vir. Existiria aqui a pista para decifrar a misteriosa afinidade entre a arte e atos de resistência?

Os artistas criam imersos nas forças do mundo que o atravessam, criam para responder a necessidade de dar corpo expressivo aos afetos do mundo em seu próprio corpo, eles criam para um povo por vir. Ainda assim, o mistério sobre a

afinidade entre a luta imanente do homem pela vida com os trabalhos de arte ainda está por ser decifrado.

REFERÊNCIAS

- CAPREZ, P. *A poética da miscigenação*: para uma floração dos signos. In: Interculturalidades. Rio de Janeiro: Ed. UFF, 2004.
- DELEUZE G. *A imagem-tempo*. São Paulo: Ed Brasiliense, 1995.
- DELEUZE G. *Francis Bacon*: lógica da sensação. Rio de Janeiro: Zahar, 2007.
- DELEUZE G. *O ato de criação*. Palestra de 1987. Edição brasileira: Folha de São Paulo, 27 jun. 1999.
- DELEUZE G. O cérebro é a tela. Entrevista com Deleuze. *Cahiers du cinéma*, n. 380, 1986. Disponível em: <<https://bit.ly/2EvXpf2>>. Acesso em: 20 ago. 2020.
- GRUSIN, R. Introduction. In: GRUSIN, R. *The Nonhuman Turn*. Minneapolis, MN: Ed. University of Minnesota Press, 2015.
- GIBSON, W. *Neuromancer*. São Paulo: Aleph, 2015.
- GUIMARÃES, L. H. *Cartografia_Vital_Artístico_Acadêmica*. *Medium.com*, 2018a. Disponível em: <<https://bit.ly/3j1CwHE>>. Acesso em: 20 ago. 2020.
- GUIMARÃES, L. H. Exposição Espectros Computacionais 360° (EC360). *Medium.com*, 2019. Disponível em: <<https://bit.ly/34nPcol>>. Acesso em: 20 ago. 2020.
- GUIMARÃES, L. H. Deslocamentos entre latitudes e longitudes. In: ENCONTRO INTERNACIONAL DE ARTE E TECNOLOGIA, 15., 2016, Brasília. Brasília: UnB, 2016. p. 18-325. Disponível em: <<https://art.medialab.ufg.br/p/18056-15-art-2016>>. Acesso em: 20 ago. 2020.
- GUIMARÃES, L. H. LArtEN: Laboratório de Arte Espectro Neural. *Medium.com*, 2020. Disponível em: <<https://bit.ly/3hj1IzH>>. Acesso em: 20 ago. 2020.
- GUIMARÃES, L. H. Transbordamentos: imaginários transdisciplinares entre a mídia e a ciência. *TransObjeto*, 2018b. Disponível em: <<https://transobjeto.wordpress.com/2018/09/24/transbordamentos/>>. Acesso em: 20 ago. 2020.
- HARAWAY, D. *Manifesto cyborg*: donne, technologie e biopolitiche del corpo. Milão: Feltrinelli Editore, 2018.
- HAYLES, K. *Como nos tornamos pós-humanos*: corpos virtuais em cibernética, literatura e informática. Chicago, Ill: University of Chicago Press, 2008.

SANTAELLA, L. Arte, ciência & tecnologia: um campo em expansão. In: GOBIRA, P. (Org.). *Percursos contemporâneos: realidades da arte, ciência e tecnologia*. Belo Horizonte: EdUEMG, 2018.

WEIBEL, P. *La imagen inteligente: neurocinema o cinema cuántico? Arte Algorítmico. De Cezane a la Computadora*. Seminário organizado por UNESCO y MECAD/ESDi, 2004.

WIENER, N. *Cibernética e sociedade: o uso humano de seres humanos*. São Paulo: Ed. Cultrix, 1954.

Sobre os autores

Adriana Silvério é licenciada em Estudos Culturais pela Universidade do Minho, com bolsa de mérito atribuída nos 3 anos do curso (2012-2015). Tem uma Pós-graduação em Comunicação, Arte e Cultura pela Universidade do Minho (2016). Participou no Congresso de Humanidades Digitais em 2015. Foi co-organizadora dos Encontros Minho-Galiza (<https://www.facebook.com/encontrominhogaliza/>) entre 2015 e 2018. Trabalha na administração pública portuguesa, numa autarquia local.

Alejandro Bia é doutor em Engenharia da Computação pela Universidade de Alicante e trabalha como professor na Universidade Miguel Hernández de Elche. Foi Vice-Diretor de P&D da Biblioteca Virtual Miguel de Cervantes, membro do Conselho TEI (2002-2004, 2004-2006 e 2017-2018), membro do Comité Executivo da Association for Literary and Linguistic Computing, agora EADH (2004 -2008 e 2008-2011) e Secretário da Hispanic Digital Humanities Society (2015-2019).

Aline Silva Costa é Professora Efetiva do Instituto Federal de Educação, Ciência e Tecnologia da Bahia (IFBA), graduada em Ciência da Computação e mestre em Linguística pela Universidade Estadual do Sudoeste da Bahia (UESB). Atualmente é doutoranda do Programa de Pós-graduação em Linguística da mesma Instituição, sob a orientação da professora Dra. Cristiane Namiuti. Atua na área de Computação, com ênfase em Linguagem de Programação e Engenharia de Software.

Ana Alcântara é Historiadora, doutorada em História Contemporânea pela Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa mestre em Ciência e Sistema de Informação Geográfica pela NOVA Information Management School e licenciada em História variante Arqueologia pela Faculdade de Ciências Sociais e Humanas da Universidade NOVA de Lisboa. É investigadora do Instituto de História Contemporânea e docente da Escola Superior de Educação do Instituto Politécnico de Setúbal. Tem-se dedicado à História Urbana, História

da Indústria e do Associativismo Operário (séc. XIX) e à aplicação dos SIG na investigação histórica.

Aquiles Alencar Brayner atuou como Curador Digital na Biblioteca Britânica em Londres onde foi responsável pelo desenvolvimento de novos produtos e modelos de gestão para os recursos eletrônicos da instituição. Lecionou em várias universidades britânicas e foi pesquisador em residência da Fundação Biblioteca Nacional. Trabalha atualmente como consultor de projetos e curadoria digital para universidades e instituições de memória.

Cristiane Namiuti Professora Titular da Universidade Estadual do Sudoeste da Bahia, em seu Programa de Pós-Graduação em Linguística. Pesquisadora da Fundação de Amparo à Pesquisa do Estado da Bahia e do Conselho Nacional de Desenvolvimento Científico e Tecnológico. Doutora em Linguística pela Universidade Estadual de Campinas. Fez parte das equipes fundadoras do Corpus Anotado do Português Histórico Tycho Brahe. Atua nas áreas da Linguística Histórica, da Linguística de Corpus e das Humanidades Digitais.

Dalton Lopes Martins Docente na Universidade de Brasília na Faculdade de Ciência da Informação. Doutor em Ciência da Informação pela Escola de Comunicação da Universidade de São Paulo. Mestre em Engenharia da Computação pela Universidade Estadual de Campinas e graduado em Engenharia Elétrica pela mesma universidade. Coordena o projeto de pesquisa Tainacan – software livre para a construção social de repositórios digitais – em parceria com o Instituto Brasileiro de Museus, Fundação Nacional das Artes, Instituto do Patrimônio Histórico e Artístico Nacional e Instituto Brasileiro de Informação em Ciência e Tecnologia.

Daniel Alves é Professor Auxiliar no Departamento de História e investigador no Instituto de História Contemporânea, ambos da NOVA-FCSH, Universidade NOVA de Lisboa. Tem um mestrado em História do Século XIX e um doutoramento em História Económica e Social Contemporânea. Tem publicado sobre História de Portugal, História Económica e Social, SIG aplicado à História e Humanidades Digitais. Coordena desde 2019 o Laboratório de Humanidades Digitais do IHC (<http://dhlab.fcsh.unl.pt/>) e desde 2020 é editor-chefe da versão em português do site *Programming Historian* (<https://programminghistorian.org/>) e editor da revista *IJHAC: A Journal of Digital Humanities*, publicada pela Edinburgh University Press.

Diogo Cunha é licenciado na Universidade do Minho em Línguas Aplicadas com três pós-graduações em Comunicação e Cidadania, em Tecnologia e Arte Digital, e em Marketing Digital. Está neste momento a finalizar o mestrado em Tecnologias Interativas na Escola de Engenharia UMinho onde também trabalha no gabinete de comunicação. É *host* do *podcast* de comunicação de ciência “No Núcleo da Engenharia”. Tem experiência como formador nas áreas das tecnologias para o ensino e design/multimédia.

Fernanda Bonacho é doutorada em Ciências da Comunicação, especialidade Comunicação e Linguagem, pela Universidade Nova de Lisboa. Mestre em Estudos Literários Comparados pela mesma instituição. Desempenha funções docentes e de gestão na Escola Superior de Comunicação Social/Instituto Politécnico de Lisboa desde 1998. Coordena o projeto Academia da Leitura do Mundo, na área da literacia comunicacional e mediática, integrado na rede das Academias Gulbenkian do Conhecimento.

Francisco Abrunhosa é licenciado em Estudos Culturais e mestre em Comunicação, Arte e Cultura pela Universidade do Minho. É cocriador, mentor e coorganizador do trabalho cultural do ‘Encontro Minho - Galiza’, realizado anualmente em parceria com a Universidade do Minho. É também cocriador, mentor e coorganizador do trabalho cultural, de carácter mensal, da tertúlia ‘A Velha Escrita’, formada em 2012 e sediada em Braga.

Giovane Santos Brito é graduado em Letras Vernáculas e mestre em Linguística, ambos pela Universidade Estadual do Sudoeste da Bahia. Especializou-se em Metodologia do Ensino a Distância – EaD pela Uninassau. Atualmente é aluno de doutorado do Programa de Pós-graduação em Língua e Cultura no Instituto de Letras da Universidade Federal da Bahia. Atua como analista de educação/EAD na Faculdade Sudoeste – Unigrad e como docente na Faculdade Einstein.

Isabel Oliveira Martins é Professora Auxiliar na Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa. Lecciona, desde 1983, nas áreas de Literatura Norte-Americana e Tradução Literária. Investigadora do Centro de Estudos Ingleses, de Tradução e Anglo-Portugueses (CETAPS) e do CEAUL (Centro de Estudos Ingleses da Universidade de Lisboa) e membro do *PEnPAL in Trans*, projecto colaborativo interinstitucional de tradução literária aplicada em plataforma digital.

Jesús Javier Rodríguez-Sala é engenheiro em Ciências da Computação pela Universidade de Alicante (Espanha) e doutor em Engenharia da Computação pela Universidade Miguel Hernández de Elche (Espanha). A sua investigação centra-se principalmente na área de mineração de dados e no desenvolvimento de algoritmos para processamento e análise de dados. É também desenvolvedor de aplicações web, tanto na parte visual (Frontend) quanto na parte de servidor e banco de dados (Backend).

João Brígola é licenciado em História, mestre em História Cultural e Política e doutor em História/Museologia, Agregado em Teoria e História da Museologia. Docente na Escola de Ciências Sociais do Departamento de História da Universidade de Évora. Director do Curso de Doutoramento em História. Titular da Cátedra UNESCO/Património Imaterial. Foi Professor Convidado na Universidade Nova de Lisboa e Director Geral do Instituto dos Museus e da Conservação, do Ministério da Cultura de Portugal.

Jorge Viana é Professor Titular da Universidade Estadual do Sudoeste da Bahia. Docente do Programa de Pós-Graduação em Linguística. Doutor em Linguística pela Universidade Estadual de Campinas. Mestre em Comunicação e Semiótica pela Pontifícia Universidade Católica de São Paulo. É pesquisador da Fundação de Amparo à Pesquisa do Estado da Bahia e do Conselho Nacional de Desenvolvimento Científico e Tecnológico. Membro do Laboratório de Pesquisa em Linguística de Corpus e do Laboratório Virtual de Humanidades Digitais da Universidade de São Paulo. Tem experiência em Linguística, Semântica, Semiótica, Linguística de Corpus, Humanidades Digitais, Memória e Fotografia.

Josir Cardoso Gomes é doutorando em Ciência da Informação no Instituto Brasileiro de Informação em Ciência e Tecnologia em convênio com a Universidade Federal do Rio de Janeiro. Possui mestrado em Administração e graduação em Processamento de Dados. Tem mais de 30 anos de experiência na área de Ciência da Computação, com ênfase em Desenvolvimento de Sistemas de Informação, atuando principalmente com banco de dados relacionais, mineração de dados, software livre, arquitetura orientada a serviços e computação em nuvem.

Joyce Siqueira Doutoranda em Ciência da Informação na Universidade de Brasília. Docente na Universidade Católica de Brasília e pesquisadora do Laboratório de Inteligência de Rede da mesma universidade. Mestre em Ciência da Computação

pela Universidade Federal de Goiás. Pós-graduada em Orientação a Objeto em Internet, pela Uni-Anhanguera e Planejamento, Implementação e Gestão de EaD pela Universidade Federal Fluminense. Graduada em Processamento de Dados pela Universidade Estadual de Goiás.

Marco Schneider é doutor em Ciências da Comunicação da Escola de Comunicações e Artes da Universidade de São Paulo. Mestre e bacharel em Comunicação Social Escola de Comunicação da Universidade Federal do Rio de Janeiro. Possui estágio pós-doutoral em Estudos Culturais. Pesquisador titular do Instituto Brasileiro de Informação em Ciência e Tecnologia. Professor associado do departamento de Comunicação da Universidade Federal Fluminense Professor do Programa de Pós-graduação em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia em convênio com a Universidade Federal do Rio de Janeiro. Pesquisador do CNPq PQ 2. Cientista do Nosso Estado da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro.

María Luisa Fernández Rodríguez Doutoranda em Ciência da Informação na Universidade de Brasília. Docente na Universidade Católica de Brasília e pesquisadora do Laboratório de Inteligência de Rede da mesma univesidade. Mestre em Ciência da Computação pela Universidade Federal de Goiás. Pós-graduada em Orientação a Objeto em Internet, pela Uni-Anhanguera e Planejamento, Implementação e Gestão de EaD pela Universidade Federal Fluminense. Graduada em Processamento de Dados pela Universidade Estadual de Goiás.

Mariana Fagundes de Oliveira Lacerda é doutora em Linguística pela Universidade Federal da Bahia, com estágio de doutoramento na Universidade de Lisboa. Professora Titular na Universidade Estadual de Feira de Santana. Coordena o Núcleo de Estudos de Língua Portuguesa e o projeto Corpus Eletrônico de Documentos Históricos do Sertão (FAPESB). Membro do Projeto Nacional para a História do Português Brasileiro e da Comissão Científica Internacional do Projeto Pombalia – Pombal Global.

Nicola Schiavottiello é doutorando em História no centro de investigação do CIDEHUS da Universidade de Évora. A sua investigação é baseada na Interpretação do Património Cultural para o público, com técnicas de visualização 3D em tempo real e com narrativas históricas. Os seus principais interesses

são a realidade virtual, a realidade aumentada, a interpretação do património arqueológico, histórico e cultural para o público através da ligação entre base de dados e ambientes virtuais.

Pedro Sepúlveda é Professor Auxiliar na Faculdade de Ciências Sociais e Humanas e investigador do Instituto de Estudos de Literatura e Tradição, ambos da Universidade Nova de Lisboa. Publicou o ensaio *Os livros de Fernando Pessoa* (Ática, 2013) e o estudo antológico *O planeamento editorial de Fernando Pessoa* (INCM, 2016). Coordena o projeto de investigação *Estranhar Pessoa* (estranharpessoa.com) e, com Ulrike Henny-Krahmer, a *Edição Digital de Fernando Pessoa: Projetos e Publicações* (pessoadigital.pt).

Renata Araujo Professora na Faculdade de Computação e Informática da Universidade Presbiteriana Mackenzie, São Paulo. Pesquisadora do Programa de Pós-Graduação em Sistemas de Informação da Escola de Artes, Ciências e Humanidades da Universidade de São Paulo e do Laboratório Ludologia, Engenharia e Simulação do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro. Graduada em Informática, mestre e doutora em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro. Coordena o Grupo de Pesquisa e Inovação em Ciberdemocracia ([HYPERLINK "http://ciberdemack.com.br/"](http://ciberdemack.com.br/) ciberdemack.com.br). Membro da comunidade científica em Sistemas de Informação junto à Sociedade Brasileira de Computação.

Ricardo M. Pimenta é Pesquisador Titular do Instituto Brasileiro de Informação em Ciência e Tecnologia e professor do Programa de Pós-graduação em Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia em convênio com a Universidade Federal do Rio de Janeiro. Doutor e Mestre em Memória Social pela Universidade Federal do Estado do Rio de Janeiro, com estágio na École des Hautes Études en Sciences Sociales. É bolsista de produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico, Jovem cientista do nosso estado da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro e coordena o Laboratório em Rede de Humanidades Digitais no departamento de ensino e pesquisa do Instituto Brasileiro de Informação em Ciência e Tecnologia e o grupo de pesquisa Memória e Sociedade da Informação. Atualmente é editor associado, responsável pela América Latina e África, do *International Journal of Humanities and Arts Computing* (IJHAC).

Roberto Samartim (Redondela, Galiza, 1971) é doutor em Filologia Galega pela Universidade de Santiago de Compostela e professor no Departamento de Letras da Universidade da Corunha. Integra o Grupo de Estudos Territoriais na UdC e a Rede Galabra (grupos na Universidade de Santiago de Compostela e na Universidade do Minho). Estuda o funcionamento dos campos culturais (com foco no literário) e os processos políticos e ideológicos com eles relacionados.

Silvia Araújo é doutorada em Ciências da Linguagem pela Universidade do Minho / Université Paris 7 – Denis Diderot. É Professora Auxiliar do Departamento de Estudos Românicos da Universidade do Minho. Desenvolve investigação em Linguística de Corpus, tecnologias aplicadas às línguas e Humanidades Digitais, coordenando projetos financiados pela FCT. É diretora do Mestrado em Humanidades Digitais do Instituto de Letras e Ciências Humanas da Universidade do Minho e diretora da H2D - Revista em Humanidades Digitais. Desde 2016, coordena a conferência Internacional techLING.

Susana Valdez é Professora Auxiliar de Estudos de Tradução na Universidade de Leiden (Países Baixos) onde leciona disciplinas teóricas e práticas de tradução. Em investigação dedica-se às normas de tradução e expectativas em tradução biomédica e está principalmente interessada em como os processos de tomada de decisão dos tradutores são influenciados pelas expectativas da comunidade linguística.

Ulrike Henny-Krahmer é doutoranda em Filologia Computacional na Universidade Würzburg, Alemanha. Entre 2015 e 2020 trabalhou no projeto *Computational Literary Genre Stylistics* (CLiGS) na mesma universidade e anteriormente no *Cologne Center for eHumanities* (CCeH) da Universidade de Colónia, onde se estabeleceu a cooperação com Pedro Sepúlveda e o projeto Estranhar Pessoa, com vista à elaboração da *Edição Digital de Fernando Pessoa: Projetos e Publicações* (pessoadigital.pt).

Zenaide de Oliveira Novais Carneiro é doutora em Linguística e pós-doutora em Linguística de *Corpus* pela Universidade Estadual de Campinas. Atualmente é Professora Plena da Universidade Estadual de Feira de Santana, onde coordena o projeto Corpus Eletrônico de Documentos Históricos do Sertão. Integra a equipe de pesquisadores do Projeto Nacional para a História do Português Brasileiro e a Comissão Científica Internacional do Projeto Pombalia – Pombal Global.

