

## Da 1.0 até a 3.0: a jornada da Paleografia no mundo digital

From 1.0 to 3.0: the journey of Paleography in the digital world

DOI: <https://doi.org/10.24206/lh.v7i3.42368>

*Livia Borges Souza Magalhães*

Doutora e mestre em Língua e Cultura pelo Programa de Pós-Graduação em Língua e Cultura do Instituto de Letras da Universidade Federal da Bahia. Licenciada e bacharel em Letras Vernáculas. É líder do Grupo de Pesquisa Modus Scribendi (CNPq-UFBA) e desenvolve pesquisas sobre acervos do século XVI ao XX, analisando o uso de recursos tecnológicos no labor filológico/paleográfico.

E-mail: [maglivia@gmail.com](mailto:maglivia@gmail.com)

ORCID: <https://orcid.org/0000-0003-2134-0883>

## RESUMO

A navegação na *web* tornou-se uma atividade comum na sociedade atual, sendo um reflexo direto do cenário em que a tecnologia computacional avança constantemente, possibilitando a constituição de verdadeiras redes de troca de informação, conhecimento e cultura. O trabalho ora apresentado configura-se como um panorama sobre o impacto dessa tecnologia para os estudos da Paleografia, com vistas a compreender as imbricações da presença dos computadores e da *web* nas práticas paleográficas. Para tanto, são dispostos os estágios evolutivos da *web* e a reverberação deles para os usuários, constituindo aquilo que, conseqüentemente, está reproduzido nas formas de fazer Paleografia, bem como da publicação de trabalhos na área.

**Palavras-chave:** Paleografia. Computadores. Evolução. Web. Paleografia Digital.

## ABSTRACT

*Web* browsing has become a common activity in today's society, being a direct reflex of the scenario in which computer technology is constantly advancing, enabling the establishment of actual networks for the exchange of information, knowledge and culture. The work presented here is an overview of the impact of this technology on the studies of Paleography, with regards to understanding the imbrications of the presence of computers and the *web* within paleographical practices. For this purpose, the evolutionary stages of the *web* and their reverberation for users are arranged, constituting what, consequently, is reproduced in the ways of doing Paleography, as well as publishing works in the field.

**Keywords:** Paleography. Computers. Evolution. Web. Digital Paleography.

## Introdução

Quando surge o primeiro computador da humanidade? Essa pergunta pode ter várias respostas plausíveis, uma vez que existem categorias de computador. Entretanto, é possível dizer que desde o primeiro computador eletrônico e multiuso, o *Electronic Numerical Integrator and Computer* (ENIAC), essas máquinas passaram a ocupar um lugar importante na vida de muitos seres humanos, chegando a constituir uma configuração de cultura, definida por Pierre Lévy, em 1997, como *Cibercultura*. Martino (2014, p. 27) nos apresenta, de maneira objetiva, uma descrição da *Cibercultura*:

[...] a reunião de relações sociais, das produções artísticas, intelectuais e éticas dos seres humanos que se articulam em redes interconectadas de computadores, isto é, no ciberespaço. Trata-se de um fluxo contínuo de idéias, práticas, representações, textos e ações que ocorrem entre pessoas conectadas por um computador — ou algum dispositivo semelhante — a outros computadores.

Podemos dizer que *ciberespaço* é a expressão acadêmica para definir esse novo cenário termo acadêmico. No dia a dia, até por um desconhecimento, nos referimos a esse espaço como *a Internet*. Fazemos pesquisas na *Internet*, usamos nossas redes sociais *na Internet*, pedimos comida *na Internet*. Entretanto, na realidade, Internet é a rede que conecta os computadores. Nós somos usuários dos sistemas que são disponibilizados nessa rede, a *World Wide Web*, ou, *web*. Em outras palavras, quando você abre o *Mozilla Firefox* ou o *Google Chrome* e faz a pesquisa de algum termo, você está usando a *web*, pois nessa atividade há uso de protocolos HTTP/HTTPS, específicos desse sistema; contudo, quando você usa o *Instagram* no seu aplicativo de celular, você não está na *web*, mas usando a comunicação da rede de computadores (a Internet) para fazer a sua navegação.

A nossa convivência, enquanto sociedade civil comum (numa distinção aos cientistas da computação ou programadores), com a rede mundial de computadores ocorre especificamente quando há uma popularização dos sistemas de navegação, os chamados navegadores. O primeiro da história foi criado em 12 de março de 1989, por Tim Berners-Lee, logrando êxito em 25 de dezembro de 1990, quando conseguiu estabelecer uma comunicação bem sucedida usando o protocolo HTTP. Esse movimento inaugurou a Era da *web* nas sociedades.

A primeira era da *web* foi a 1.0 (lê-se um ponto zero) e ela era caracterizada da seguinte forma:

[...]as read only *web* where there are small amount of producer create *web* pages (interlinked) and a large number of customers access those *web* pages through browser via internet. Here user can only read information, user cannot interact with the content of the pages (like comment, answers etc). Technologies used in *web* 1.0 are HTML, HTTP, URI these are core *web* protocols, some newer protocol are also in used [sic] like XML, XHTML and CSS.<sup>1</sup> (NATH; DHAR; BASISHTHA, 2014, p. 86)

Já a segunda era, a *web* 2.0, é marcada por um pouco mais de interatividade. Ela é conhecida como "[...] read-write *web*. It is basically a new way to use existing internet technologies. In *web*2.0 the *web* user cannot only read the content but also write, modify and update the content online, it supports collaboration and help to gather collective intelligence rather *web* 1.0.<sup>2</sup>" (Nath; Dhar; Basishtha, 2014, p. 87).

A fase da *web* 3.0 foi conceituada por Tim Berners-Lee, em 2001. É nela que estamos atualmente e, por isso, sua caracterização efetiva se torna um pouco mais difícil, uma vez que poucos se arriscam no exercício de definir as configurações dessa vivência atual, tão marcada por constantes mudanças e atualizações. Contudo, algumas informações já podem ser firmadas como, por exemplo, o fato de a fase também ser chamada de *web* Semântica, demarcando que, neste momento, os protocolos estão sendo programados para buscar sentido e interpretar os dados apresentados pelos usuários. Esse processo é feito por meio da Inteligência Artificial e tem como resultado a constituição de uma interação cooperativa entre os sistemas e seus usuários.

Exposto esse cenário, apresentamos o objetivo deste artigo: pontuar uma analogia da Paleografia com essa evolução da *web*, observando, inclusive, os impactos dessa evolução no fazer paleográfico.

## Paleografia 1.0

Quando tomamos manuais introdutórios de Paleografia como, por exemplo, o *Noções de Paleografia e de Diplomática*, de Ana Regina Berwanger e João Eurípedes Leal, encontramos, convencionalmente, uma parte introdutória dedicada a apresentar o conceito de Paleografia. Normalmente, essa apresentação começa com a definição etimológica baseada nas raízes gregas:

---

<sup>1</sup> Tradução nossa: [...] como web somente leitura, em que há pequena quantidade de páginas da web criadas por produtor (interligadas) e um grande número de clientes acessa essas páginas da web por meio de navegador via internet. Aqui o usuário pode apenas ler informações, o usuário não pode interagir com o conteúdo das páginas (como comentários, respostas, etc.). As tecnologias usadas na Web 1.0 são HTML, HTTP, URI, esses são os principais protocolos da web, alguns protocolos mais novos também são usados, como XML, XHTML e CSS.

<sup>2</sup> Tradução nossa: [...] ler-escrever na web. É basicamente uma nova maneira de usar as tecnologias de Internet existentes. Na web 2.0 o usuário da web pode não apenas ler o conteúdo, mas também escrever, modificar e atualizar o conteúdo online, ele suporta a colaboração e ajuda a reunir inteligência coletiva em detrimento da web 1.0

παλαιός, significando antigo e γραφή, para escrita. Essa acepção faz entender que o fazer paleográfico se configura somente como o exercício de ler as escritas antigas, o que Núñez Contreras (1994) caracterizou como *Paleografía de Lectura*. O autor nos informa que

Responde al más antiguo concepto que de la Paleografía se tuvo [...] Desde la óptica del pragmatismo en orden a leer escrituras antiguas se publicaron los primeros tratados de Paleografía, sobre todo a partir del siglo XVII, tratados que recogían, completaban y mejoraban los elencos de abreviaturas y las elementales doctrinas establecidas en la Antigüedad y en la Edad Media sobre las escrituras antiguas y sobre su interpretación.<sup>3</sup> (NÚÑEZ CONTRERAS, 1994, p. 19).

Apesar de ser a forma mais antiga, ela não foi abandonada frente à evolução que a Paleografia sofreu ao longo dos anos, coexiste e continua sendo extremamente eficaz para alguns estudos desenvolvidos ainda hoje, no século XXI. Podemos citar, como exemplo, o trabalho da professora Norma Suely da Silva Pereira, publicado no *e-book Paleografia e suas Interfaces*, em que a pesquisadora se vale da Paleografia para fazer a leitura de três documentos: o *Notícia do Gov(ern)o Ecleziástico*, uma parte da obra *Notícia geral de toda esta capitania da Bahia desde o seu descobrimento até o presente anno de 1759*, escrita por José Antônio Caldas; o *Regimento da Sé da Cidade da Bahia*, de 1719; e o *Regimento do Coro da Sé da Bahia*, de 1720. Essa leitura paleográfica foi norteadora para o desenvolvimento de um trabalho de cunho filológico e linguístico, como ela cita no próprio texto:

[...] procurou-se demonstrar a necessidade de adequada preparação do pesquisador para que obtenha bom êxito na leitura de documentos manuscritos, de modo a otimizar a sua utilização. Nesse sentido, para a superação de possíveis obstáculos destaca-se a importância da utilização dos métodos da Filologia textual para o estabelecimento de um texto inteligível, credível e que conserve as características da *scripta*, para que possam ser estudados os elementos que contribuam para a reconstrução de fatos históricos e para que se possam melhor interpretar os fatos linguísticos. (PEREIRA, 2018, p. 229).

Professora da Universidade Federal da Bahia, a autora é uma pessoa que participa da configuração social do século XXI e está inserida na *Cibercultura* e, por isso, é influenciada, direta ou indiretamente, pelos recursos da tecnologia computacional. Contudo, no desenvolvimento do trabalho anteriormente citado, assim como convencionalmente em todo e qualquer trabalho de Paleografia de leitura, o uso de tecnologia computacional foi incipiente e, provavelmente, nem sequer problematizado.

---

<sup>3</sup> Tradução nossa: Responde ao mais antigo conceito de Paleografia que se tinha [...] Do ponto de vista do pragmatismo para a leitura de escritos antigos, foram publicados os primeiros tratados de Paleografia, especialmente a partir do século XVII, tratados que recolhiam, completavam e melhoravam os conjuntos das abreviaturas e das doutrinas elementares estabelecidas na Antiguidade e na Idade Média sobre os escritos antigos e sua interpretação.

O comum é, nesse nível de desenvolvimento de trabalho paleográfico, tomar a tecnologia para dois fins: o primeiro deles é a produção de fac-símiles, gerados por meio do processo de digitalização utilizando câmeras digitais ou *scanners*, tarefa comum, atualmente, em muitas instituições arquivísticas detentoras de acervos. Para o CONARQ (2010, p. 4), fazer esse trabalho

[...] é uma das ferramentas essenciais ao acesso e à difusão dos acervos arquivísticos, além de contribuir para sua preservação, uma vez que restringe o manuseio aos originais, constituindo-se como instrumento capaz de dar acesso simultâneo local ou remoto aos seus representantes digitais como os documentos textuais, cartográficos e iconográficos em suportes convencionais [...]

Já o segundo é, de posse dos documentos, estejam eles apresentados de maneira física ou digital, utilizar processadores de texto, como o *Microsoft Word*, para produzir as transcrições paleográficas. Convencionalmente, o usuário desses processadores não se ocupa de analisar a mediação que existe entre o ato de teclar e a exposição da letra teclada na tela, pois "Esses programas são desenvolvidos para simular as funções de uma máquina de escrever sofisticada, contendo ferramentas de formatação que favoreçam a manipulação do texto de modo rápido e prático, diretamente na tela do computador." (BARREIROS, 2015, p. 178).

## Paleografia 2.0

Assim como na *web 2.0*, a Paleografia 2.0 é demarcada por um elemento particular: a interação. É por meio dela que o paleógrafo institui efetivamente a realização do processo de leitura, mas com um objetivo: constituir "a un riguroso examen relativo a 'todos los problemas de identificación, autenticación y reagrupamiento de las escrituras sea cual fuera el periodo al que pertenecen'"<sup>4</sup> (NÚÑEZ CONTRERAS, 1994, p. 20). Em outras palavras, o fazer paleográfico é marcado pela interação do pesquisador com o texto, estabelecendo uma relação genuinamente hermenêutica, uma vez que cabe ao sujeito interpretar o texto para coletar a informação e mediar a entrega dela para outros estudiosos como, por exemplo, os historiadores.

Núñez Contreras (1994) sinaliza, inclusive, que esse fazer paleográfico interpretativo apresenta uma proximidade com os princípios da Crítica Textual, outra ciência genuinamente interpretativa que, na sua metodologia clássica, busca restituir a forma genuína dos textos, posto que, no processo de

---

<sup>4</sup> Tradução nossa: [...] a um exame rigoroso sobre "todos os problemas de identificação, autenticação e reagrupamento de títulos, independentemente do período a que pertençam".

transmissão, ao longo do tempo, por ato voluntário ou não, terminam sofrendo mudanças (CAMBRAIA, 2005).

Para exemplificar o desenvolvimento de trabalhos que se enquadrem naquilo que estamos chamando de Paleografia 2.0, tomamos a dissertação de mestrado de Ana Cristina Pereira da Silva Ferreira, intitulada *Análise Paleográfica de uma escrita de Chancelaria Régia: A letra Joanina, 1370-1420*. Nessa pesquisa, a autora recorre aos conhecimentos paleográficos para ler o texto manuscrito, constituindo um estudo da Paleografia de leitura, mas, além disso, estuda as características da escrita para identificar a existência de uma letra que configuraria uma chancelaria régia em determinado período da história de Portugal. Nas palavras da autora:

O objectivo principal desta tese de mestrado é demonstrar a existência de uma letra própria da Chancelaria Régia portuguesa que teve início [sic] com o começo do reinado de D. João I e que prolongou-se [sic] praticamente até ao reinado de D. Manuel I. Por esse motivo, decidimos abordar a baliza cronológica 1370-1420, datas que achamos serem suficientes para provar e caracterizar a existência desta letra, que o Professor Doutor Borges Nunes gostava de apelidar de letra joanina. (FERREIRA, 2011, p. 4).

No trabalho, o uso da tecnologia é mais presente do que o encontrado na atuação da Paleografia 1.0. No caso que tomamos como exemplo, a autora constituiu uma base de dados utilizando o *Microsoft Excel* para registrar informações sobre:

Nº de inventário, Tipo de letra, Década, Ano, Mês, Dia, Local, Tipo de documento, Nome do escrivão, Profissão, Desembargadores (quando existem), Nº de ficha (de acordo com o sumário, cada fundo visto começa a sua ficha com o nº1), Fundo, Cota e Descrição da letra. (FERREIRA, 2011, p. 65).

O *Excel* é um editor de planilhas dotado de ferramentas de cálculos e, por isso, facilitador do desenvolvimento de análises de dados especificados, como os anteriormente listados. A própria autora nos apresenta informações sobre esse uso:

O facto de o inventário ser feito em Excel, permite-nos depois, através das células da folha de cálculo seleccionar a informação que pretendemos: por exemplo, se queremos ver apenas os documentos feitos por escrivães, na coluna de “Profissão” seleccionamos escrivães, seguindo a mesma lógica para seleccionarmos um fundo específico, um determinado individuo [sic], etc. (FERREIRA, 2011, p. 66).

Temos, aqui, a pesquisadora, na composição de seu trabalho, fazendo uma delimitação das informações que são essenciais para construção da sua análise, ou seja, interpretando os dados do texto e utilizando a tecnologia de maneira mais consciente, como uma aliada do ato de pesquisa. Entretanto,

ainda na fase 2.0 do fazer paleográfico, encontramos estudos que se apropriam ainda mais da tecnologia, chegando ao ponto de convidar o leitor a interagir com o texto lido. Falamos, aqui, das chamadas edições digitais:

A edição digital, e não edição meramente em formato digital, mostra-se um tipo completamente adequado à Filologia que precisa não somente trabalhar o texto, mas também o paratexto, as informações que contextualizam e dão sentido ao documento editado. Nas edições anteriores tais informações vinham como arredores, mas na edição digital esse arcabouço informacional está totalmente integrado ao texto transcrito, criando assim uma sintonia perfeita entre a transcrição e todas as informações que foram necessárias para que o filólogo adentrasse esse texto, e, conseqüentemente, desempenhasse sua função (de trazer o texto fidedigno) com mais confiança e clareza. O entorno do texto é sempre fundamental para uma boa edição e a edição digital possibilita esse diálogo de forma natural e soberana. A edição digital mostra-se completa, pois o editor pode escolher os critérios de qualquer tipo de transcrição já existente e fazer dialogar isso através de hiperlinks com seu paratexto, além de desdobramento de abreviaturas, movimentos de correção do autor, em caso de texto moderno, entre outras possibilidades. Além disso, tornar o texto digital é possibilitar sua divulgação de forma mais fácil, acessível e abrangente. O que se sente é que a Filologia estava à espera desse tipo de edição para conseguir fazer promover este diálogo entre as suas muitas faces, as suas tantas atividades, de forma democrática e interativa (LOSE; OLIVEIRA; NUNES; MAGALHÃES; ANDRADE, 2011, p. 78).

O paleógrafo, na constituição desses trabalhos, assume uma série de responsabilidades, todas elas mediadas pela interpretação crítica do texto que está estudando. Podemos citar, inicialmente, a ponderação sobre a disponibilização do texto na rede mundial de computadores. Deve considerar, inicialmente, por mais que o texto estudado seja significativo enquanto corpus de pesquisa, se a publicação dela teria um impacto para leitores e outros pesquisadores. Consideramos, para fins de exemplo desse processo, os trabalhos do Grupo de Pesquisa Memória em Papel na constituição da edição da *Coleção de Livros do Tombo do Mosteiro de São Bento da Bahia*. Os pesquisadores desenvolveram uma edição clássica, em um processador de textos, para ser impressa em formato de livro, mas ponderaram que a disponibilização *on-line*, simplesmente acessibilizando para interessados a leitura dos documentos manuscritos, seria uma contribuição para comunidade acadêmica, posto que o manancial apresentado na coleção é de interesse de pessoas do mundo inteiro, fato comprovado pela nomeação, em 2012, como patrimônio da Memória do Mundo pela UNESCO. O resultado dessa percepção foi a publicação do *website* <http://saobento.org/livrosdotombo/>.

Para além desses trabalhos de disponibilização do manuscrito para ser lido na *web*, é importante ressaltarmos, também, aquelas edições que convidam o leitor a escolher os seus caminhos de leitura por meio do acesso ou não aos *hiperlinks* que integram o texto. Essa constituição do *texto linkado* é capaz de redirecionar de uma, digamos assim, base textual para vídeos, imagens, outros textos, músicas e esses elementos também podem remeter para outros, e outros... fazendo com que aquela leitura



inicial do manuscrito se perca completamente na imensidão da *web* (ou não, pois fica a critério do leitor). Aqui, o paleógrafo é o sujeito que atua como desenvolvedor de uma cadeia de dados, inserindo o elo entre o texto que está sendo transcrito e todos os outros dados a mais que achar que são necessários e eficazes para uma interpretação do texto. Contudo, é importante lembrarmos que ele jamais será o mediador dessa leitura, pois “ [...] os canais de distribuição são ágeis, não dependem de cadeias de distribuição ou limitações espaciais ou geográficas, alterando a informação no contexto social em diferentes escalas, de acordo com a natureza constituída na documentação circulante.” (SERRA, 2014, p. 77).

Aqui exemplificamos com o trabalho desenvolvido por Livia Magalhães (2018) ao construir uma edição digital unindo edições de documentos do acervo do Mosteiro de São Bento da Bahia, mais especificamente:

- A edição da já citada *Coleção de Livros do Tombo*, série de documentos onde estão registradas

[...] informações que alcançam um período que vai dos séculos XVI ao XVIII. Os textos constituem-se de traslados do original, autenticados, de acordo com os termos de abertura e de encerramento, por tabelião, afirmando-se que, depois de copiados, foram lidos e achados conforme os originais (ANDRADE, 2010, p. 18).

- A edição do Livro de Aforamentos

[...] um códice cujas páginas, escritas do ano de 1743 até 1874, guardam uma série de contratos de aforamentos de terras que haviam sido doadas ao Mosteiro de São Bento da Bahia, ou seja, trata-se de um documento notarial que remonta a configuração social, geográfica, histórica e linguística da Cidade de Salvador entre os séculos XVIII e XIX (MAGALHÃES, 2010, p. 9).

- A *Edição das Cartas de Profissão dos Monges*, um conjunto de documentos “[...] tidos como símbolo de total entrega à vida religiosa do iniciante à vida monástica e testemunhos do compromisso e da obediência que este prometeu à sua ordem e à sociedade em geral” (JESUS, 2014, p. 23), posto que registram o momento em que se inicia a vida monástica, funcionando, então, como certidão de nascimento dos monges para a vida religiosa.
- A *edição do Livro de Crônicas*, “[...] um manuscrito cunhado de 1914 até 1934 onde estão registradas, como o nome sinaliza, as crônicas do mosteiro baiano” (MAGALHÃES, 2013, p. 21).

- A edição do *Diretório da Congregação*, um documento que apresenta, de maneira objetiva, a biografia dos monges citados no *Livro de Crônicas*, trazendo, assim, dados consideravelmente relevantes para enriquecimento do trabalho proposto.

Feita a seleção, criaram-se elos entre os documentos, partindo da observação de que todos os textos foram construídos por meio do resgate da história de pessoas e lugares que, direta ou indiretamente, estavam ligados à abadia baiana, o que Greimas e Courtés (1983) chamaram de *actantes*. A toponímia e a antroponímia passaram, então, a funcionar como *hiperlinks*, permitindo ao leitor navegar entre os documentos, fazer buscas diretas sobre qualquer parte do texto (por meio da ferramenta de busca dos navegadores), bem como acessar informações paratextuais, imagens, outros *links* da *web* que explicavam alguma informação presente no manuscrito de base. Tal trabalho pode ser acessado por meio do endereço: [www.papeisquenarram.com.br](http://www.papeisquenarram.com.br).

### Paleografia 3.0

Antes de tratarmos da Paleografia 3.0, é importante lembrarmos algumas informações importantes. José D'Assunção Barros (2020) firma a fonte histórica como a base empírica responsável por legitimar as pesquisas na área da História, cabendo ao historiador, frente a sua constituição de pesquisa, considerar qual documento deve ser escolhido, num trabalho de interpretação sobre a adequação da fonte para composição do *corpus* de estudo. O autor informa os elementos que devem ser observados nesse processo de escolha documental – a pertinência, a suficiência, a exaustividade, a representatividade, a homogeneidade e a organização – e, citando Ladurie (1990), ressalta que é importante perceber que de um fragmento documental é possível extrair as mais diversas informações, abrangendo desde a "cultura material [...] modos de pensar e sentir [...]". (BARROS, 2020, p. 34)

Temos, aqui, um olhar para História sob forte influência da Antropologia moderna, que passou a considerar um novo jeito de registrar a história, tendo como principal proposta o abandono da tradição historiográfica, caracterizada pelo viés político, baseado em feitos oficialmente documentados e objetivos – em prol de uma tentativa de buscar, como aponta Peter Burke (1992), construir uma história total, capaz de mostrar os diversos pontos de vista sobre um fato histórico, configurando, assim, uma valorização das múltiplas verdades.

Esse movimento alarga o campo de seleção das fontes e faz com que o trabalho do paleógrafo passe por um outro processo evolutivo, indicado por Núñez Contreras (1994) como o estabelecimento da Paleografia como História da Escrita. Esse direcionamento coloca a ciência como a

responsável por se dedicar ao dado escrito que serve de objeto de estudo nas ciências históricas, como sinaliza Antonio Castillo Gómez e Carlos Sáez (2016):

[...] a Paleografia, mediante o estudo da escritura no marco da sociedade que a produz, o que constitui o campo de estudo da história social da escritura, reorienta seu método de trabalho e suas pretensões científicas; substitui seu velho interesse pela escritura das classes dominantes, produto também da procedência social daqueles que se ocuparam da Paleografia até princípios deste século, para se ocupar de detectar os rastros escritos das classes populares.

Para exemplificar a constituição desse tipo de fazer paleográfico, tomamos algumas pesquisas que têm sido desenvolvidas atualmente pelo Grupo de Pesquisa Memória em Papel. Desde 2019, as pesquisadoras Lucia Furquim Werneck Xavier, Livia Borges Souza Magalhães e Alícia Duhá Lose estão coordenando a constituição de um modelo de leitura automatizada em Língua Portuguesa para o aplicativo *Transkribus*, compondo, mais especificamente, o projeto *Muitas Mãos e Muitas Penas*, que conta, ainda, com a participação dos pesquisadores, em nível de iniciação científica, João Guilherme Veloso e Leonardo Coelho Marques de Jesus; e da mestra em história Cláudia Coimbra do Espírito Santo.

Nesse projeto, os pesquisadores têm ensinado a ferramenta a reconhecer as características da Língua Portuguesa para que, posteriormente, o computador faça a edição automatizada e de maneira rápida do conjunto documental selecionado, mais especificamente, os documentos do período inquisitorial disponibilizados pelo Arquivo Nacional da Torre do Tombo, de Portugal.

Essa escolha justifica-se, primordialmente, pelo fato de os documentos já estarem digitalizados com uma qualidade considerada adequada para o trabalho proposto, um quesito relevante por facilitar o processo de reconhecimento, por parte da *Transkribus*, das regiões de texto que precisam ser selecionadas para leitura. Outro ponto que justifica a escolha é a própria relevância social, cultural e histórica que os processos da Inquisição salvaguardam em si por serem testemunhos da atuação do Santo Ofício no Brasil Colônia [...] Por fim, a extensão do conjunto documental conta como um outro ponto relevante que nos influenciou na escolha pois, convencionalmente, cada processo apresenta um tipo de *scripta*, dotado de determinadas características. Diversificar essa leitura significa diversificar a capacitação da plataforma na *tokenização*, o que faz com que o modelo seja mais exitoso. (MAGALHÃES, XAVIER, 2021, p. 264-265)

Todo trabalho tem sido feito na plataforma *Transkribus*. Ela é “[...] a comprehensive platform for the digitisation, recognition, transcription and searching of historical documents”<sup>5</sup> (TRANSKRIBUS,

---

<sup>5</sup> Tradução nossa: [...] uma plataforma abrangente para a digitalização, reconhecimento, transcrição e pesquisa de documentos históricos.

2021), desenvolvida pela Universidade de Innsbruck, na Áustria, e disponibilizada no site <https://readcoop.eu/transkribus/>, podendo ser usada em uma versão *on-line* ou, também, em estilo aplicativo para *desktop*.

Para digitalização, a plataforma dispõe de uma sugestão: o *DocScan*: um aplicativo para celular, programado tanto para funcionar no sistema *Android* como para *IOs*, que é capaz de digitalizar livros ou documentos com celeridade, pois detecta automaticamente as páginas, dispensando qualquer interação com o celular durante o processo de digitalização. Além disso, o aplicativo assegura produção de imagens de alta resolução.

De posse das imagens a serem trabalhadas, parte-se para a segunda etapa do processo, o reconhecimento do texto. A plataforma está programada dentro daquilo que chamamos de *Handwritten Text Recognition* (HTR). Essa tecnologia usa a base do desenvolvimento do *Optical Character Recognition*, responsável por tomar imagens digitalizadas de impressos e convertê-las em texto editável, num processo que ocorre porque a máquina compara os caracteres com modelos que estão registrados no banco de dados. O HTR é a construção de um banco de dados de reconhecimento de escrita manuscrita que, segundo Raashid Hussain, Ahsen Raza, Imran Siddiqi, Khurram Khurshid e Chawki Djeddi, "[...] offers a number of interesting pattern classification problems including handwriting recognition, writer identification, signature verification, writer demographics classification and script recognition, etc<sup>6</sup>", o que demanda, então, um fazer paleográfico para tratar de todas essas questões e, assim, construir um banco de dados realmente válido.

O trabalho da equipe do Grupo de Pesquisa Memória em Papel tem sido exatamente esse: construir um banco de dados para que o *Transkribus* consiga atuar com documentos em Língua Portuguesa. Para tanto, as imagens a serem usadas devem ser sinalizadas, pelo paleógrafo ou automaticamente, pela máquina, apontando exatamente onde tem texto, o que gera um elemento denominado *região de texto*. Após isso, inicia-se o processo de inclusão de dados: transcrever exatamente as palavras que estão na região indicada, respeitando todos os detalhes e sendo o mais conservador possível. Após a feitura desse processo, deve-se rotular, para máquina, aquilo que é chamado de *verdade fundamental*, ou, em inglês, o *ground truth*. No *Transkribus*, esse processo figura como uma confirmação do usuário para a ferramenta de que aquele fólio está pronto para entrar no banco de dados.

Quanto mais fólios forem transcritos, mais robusto será o quantitativo de *ground truth* e, conseqüentemente, mais consistente será o banco de dados para construção da automação de

---

<sup>6</sup> Tradução nossa: [...] oferece uma série de problemas interessantes de classificação de padrões, incluindo reconhecimento de escrita, identificação de escritor, verificação de assinatura, classificação demográfica do escritor e reconhecimento do script, etc.

transcrições. É, efetivamente, um processo de treinamento. No projeto *Muitas Mãos e Muitas Penas*, tem-se tomado o percentual de 30% de todo conjunto documental inserido no *Transkribus*. Exemplificamos: A equipe inseriu as imagens do processo inquisitorial do Padre Manoel de Moraes, um documento com 268 fólios e, conseqüentemente, 268 imagens e, tirando a amostragem de 30 %, 81 delas foram editadas para construir o *ground truth*, que será somado ao de outros conjuntos documentais e, assim, alimentará a base de dados de maneira mais efetiva.

Feita a base dados, é hora de construir o modelo de leitura automatizada. A plataforma recomenda que esse processo só seja feito se você tiver entre 5000 e 150000 palavras já transcritas como *ground truth*. Entretanto, para realizá-lo, basta acessar a opção de ferramentas da plataforma e clicar na opção "Train". Ela logo iniciará a disponibilização de uma série de formulários para coletar informações sobre o modelo que está por vir e, após o preenchimento deles, iniciará o processo de treinamento. É um processo demorado, mas, quando finalizado, entrega o modelo pronto para uso. A demora em si ocorre em função da articulação das redes neurais, ou seja, os sistemas de computação que usam algoritmos para reconhecer padrões e gerar correlações, agrupando dados, classificando e aprendendo constantemente para se aperfeiçoar. Citamos a descrição desse processo de construção do modelo por Muehlberger *et al'* (2018):

Neural networks can be trained to recognise a particular style of writing by examining and processing digitised images and transcriptions of documents. The result of the training process is what is known as an HTR "model", a computational system tailored to automatically transcribe a set of historical material. HTR technology is language-independent: the neural network training process for any type of alphabet, from any date, is the same. This means there is potential to train up models for any script, from any period (Leifert et al. , 2016). The technology follows a line-oriented approach where the image of a baseline (a horizontal line running underneath a line of text in a digitised image) and the corresponding correctly transcribed text represent the input for the learning algorithms of neural networks (Romero et al. , 2015).<sup>8</sup>

---

<sup>7</sup> Optamos, na construção deste artigo, pelo não uso da expressão *et al.* No entanto, neste caso, fizemos o uso do recurso, pois o texto citado tem a autoria de mais de dez pessoas e a explicitação de todos esses nomes quebraria a dinâmica de leitura do texto. Ressaltamos, entretanto, que todos os autores estão devidamente citados na referência.

<sup>8</sup> **Tradução nossa:** As redes neurais podem ser treinadas para reconhecer um estilo específico de escrita, examinando e processando imagens digitalizadas e transcrições de documentos. O resultado do processo de treinamento é o que é conhecido como HTR "modelo", um sistema computacional sob medida para transcrever automaticamente um conjunto de material histórico. A tecnologia HTR é independente da linguagem: o processo de treinamento da rede neural para qualquer tipo de alfabeto, de qualquer data, é o mesmo. Isso significa que há potencial para treinar modelos para qualquer escrita, de qualquer período. (Leifert et al., 2016). A tecnologia segue uma abordagem orientada por linha, em que a imagem de uma linha de base (uma linha horizontal passando por baixo de uma linha de texto em uma imagem digitalizada) e o texto transcrito corretamente correspondente representam a entrada para os algoritmos de aprendizagem de redes neurais (Romero et al., 2015).

Apresentado o processo de uso do *Transkribus*, é hora de retomarmos as exposições sobre a Paleografia 3.0. Vamos considerar que esse fazer paleográfico, automatizado, firma-se no mesmo momento em que há uma abertura daquilo que pode ser considerado como fonte, ou seja, há um aumento nos mananciais de documentos a serem estudados. Somado a isso, devemos considerar que as instituições arquivísticas estão, paulatinamente, disponibilizando seus acervos em formato digital, usando a *web* como espaço para preservação e divulgação de documentos, o que faz com que o acesso aos textos seja sensivelmente mais facilitado. O resultado desse somatório é a constituição de um cenário em que há mais dados a serem estudados pelo sujeito paleógrafo e poder contar com a tecnologia para transcrições automatizadas é, sem dúvida, um diferencial, visto que o trabalho ocorrerá com celeridade.

É importante sinalizarmos que não estamos aqui propondo que as máquinas substituam o trabalho humano, mas, sim, ajudem. Vamos tomar, por exemplo, o trabalho de constituição da edição da coleção de Livros do Tombo do Mosteiro de São Bento da Bahia anteriormente citada. Na produção daquele trabalho, um grupo significativo de pessoas atuou durante dez anos. De maneira automatizada, supomos que todas as transcrições estariam prontas em, aproximadamente, seis horas e, com elas nas mãos, a equipe partiria para as outras etapas do trabalho, que incluiriam, sem sombra de dúvidas, a revisão, uma vez que a *finesse* da percepção humana é um elemento peculiar que as máquinas ainda não dominam (e acreditamos que nunca dominarão.)

## Paleografia 4.0 ou considerações finais

Uma pesquisa em sites de busca na *web* já nos indica um encaminhamento para a existência da *web* 4.0, mas ainda com a apresentação de discussões sobre o que está por vir. Se tratando de tecnologia e pensando na velocidade com que ela evolui, é difícil firmar exatamente o próximo cenário, mas algo é circunstancial: a tecnologia se firmou e constituiu um mundo conectado, que não irá mais se desconectar e as práticas científicas, como a Paleografia, são constantemente perpassadas por esse processo. É preciso que nós, cientistas dos tempos da *web*, nos apropriemos dessa dinâmica para, com propriedade, fazer jus ao nosso tempo.

## Referências bibliográficas

- ANDRADE, Marla Oliveira. **Uma porta para o passado: edição de documentos dos séculos XVI e XVII do Livro I do Tombo do Mosteiro de São Bento da Bahia** – 2009. 342 f. Dissertação (Mestrado em Letras e Linguística) – Programa de Pós-Graduação em Letras e Linguística do Instituto de Letras, Universidade Federal da Bahia, 2010.
- BARREIROS, Patrício Nunes. **O Pasquineiro da Roça: edição dos panfletos de Eulálio Motta.** Feira de Santana: UEFS Editora, 2015.
- BARROS, José D'Assunção. **A fonte histórica e seu lugar de produção.** Petrópolis: Vozes, 2020.
- BURKE, Peter. A nova história, seu passado e seu futuro. *In*: BURKE, Peter (org.). **A escrita da História: novas perspectivas.** Tradução de Magda Lopes. São Paulo: Editora UNESP, 1992  
Disponível em:  
<[http://etnohistoria.fflch.usp.br/sites/etnohistoria.fflch.usp.br/files/Burke\\_Nova\\_Historia.pdf](http://etnohistoria.fflch.usp.br/sites/etnohistoria.fflch.usp.br/files/Burke_Nova_Historia.pdf)>.  
Acesso em: 02 mar. 2021
- CAMBRAIA, César Nardelli. **Introdução à crítica textual.** São Paulo: Martins Fontes, 2005.
- FERREIRA, Ana Cristina Pereira da Silva. **Análise paleográfica de uma escrita de Chancelaria Régia: a letra joanina, 1370–1420.** Tese de mestrado, História, Universidade de Lisboa, Faculdade de Letras, 2012. Disponível em: <<http://hdl.handle.net/10451/695>>.
- CONARQ. **Recomendações para digitalização de documentos arquivísticos permanentes** Brasília, DF: CONARQ, 2010. Disponível em:  
<<http://www.ufopa.edu.br/media/file/site/proad/documentos/2018/38de0a593ed36b780ed061b6304075f8.pdf>>. Acesso em: 22 fev. 2021.
- GREIMAS, Algirdas Julien; COURTES, Jacques. **Dicionário de semiótica.** Tradução de Alceu Dias Lima et al. São Paulo: Cultrix, 1983.
- GÓMEZ, Antonio Castillo; SÁEZ, Carlos. Paleografia versus Alfabetização: reflexões sobre História Social da Cultura Escrita. **Labor Histórico**, Rio de Janeiro, v. 2, p. 164–187, jan. | jun. 2016.
- HUSSAIN, Raashid; RAZA, Ahsen; SIDDIQI, Imran; KHURSHID, Khurram; DJEDDI, Chawki. A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation. **EURASIP Journal on Image and Video Processing**. v. 46, 2015.  
<<https://doi.org/10.1186/s13640-015-0102-5>>.
- JESUS, Tamires Alice Nascimento de. **Cartas de profissão dos monges do Mosteiro de São Bento da Bahia: “certidão de nascimento” de um ser renovado.** 2014. 215f. Monografia (Bacharelado em Letras) – Instituto de Letras, Universidade Federal da Bahia, 2014.
- LOSE, Alícia Duhá; OLIVEIRA, Jaqueline Martins Carvalho; MAGALHÃES, Livia Borges Souza; NUNES, Marília Andrade; ANDRADE, Marla Oliveira. Edições digitais de manuscritos: do

- século XVI ao século XXI. In: CIRILLO, José; PASSOS, Marie- Hélène Paret (Org.). **Materialidade e virtualidade no processo criativo**. Vinhedo: Horizonte, 2011. p. 77-99.
- MAGALHÃES, Livia Borges Souza. **Fazendo filologia entre tags e dígitos binários**: uma proposta de edição do *Livro de Crônicas* do Mosteiro de São Bento da Bahia. 2013. 406f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras e Linguística, Universidade Federal da Bahia, 2013. DVD.
- MAGALHÃES, Livia Borges Souza. **Pequenas análises do Livro de Aforamentos do Mosteiro de São Bento da Bahia**. 2010. 85f. Trabalho de conclusão de curso (Bacharelado em Letras Vernáculas) – Instituto de Letras, Universidade Federal da Bahia, 2010.
- MAGALHÃES, Livia Borges Souza; Lucia Furquim Werneck Xavier. Can machines think? Por uma paleografia digital para textos em língua portuguesa. In: LOSE, Alcía Duhá;
- MAGALHÃES, Livia Borges Souza; MAZZONI, Vanilda Salignac de Sousa. **Paleografia e suas interfaces**. Salvador: Memória e Arte, 2021.
- MARTINO, Luís. **Teoria das mídias digitais**: linguagens, ambientes, redes. Rio de Janeiro: Vozes, 2014.
- MUEHLBERGER, Guenter; SEAWARD, Louise; TERRAS, Melissa; OLIVEIRA, Sofia Ares; BOSCH, Vicente; BRYAN, Maximilian; COLUTTO, Sebastian; DÉJEAN, Hervé; DIEM, Markus; FIEL, Stefan; GATOS, Basilis; GREINOECKER, Albert; GRÜNING, Tobias; HACKL, Guenter; HAUKKOVAARA, Vili; HEYER, Gerhard; HIRVONEN, Lauri; HODEL, Tobias; JOKINEN, Matti; KAHLE, Philip; KALLIO, Mario; KAPLAN, Frederic; KLEBER, Florian; LABAHN, Roger; LANG, Eva Maria; LAUBE, Sören; LEIFERT, Gundram; LOULOUUDIS, Georgios; MCNICHOLL, Rory; MEUNIER, Jean-Luc; MICHAEL, Johannes; MÜHLBAUER, Elena; Philipp, Nathanael; PRATIKAKIS, Ioannis; PUIGSERVER, Joan Pérez; PUTZ, Hannelore; RETSINAS, George; ROMERO, Verónica; SABLATNIG, Robert; SÁNCHEZ, Joan Andreu; SCHOFIELD, Philip; SFIKAS, Giorgos; SIEBER, Christian; STAMATOPOULOS, Nikolaos; STRAUß, Tobias; TERBUL, Tamara; TOSELLI, Alejandro Héctor; ULREICH, Berthold; VILLEGAS, Mauricio; VIDAL, Enrique; WALCHER, Johanna; WEIDEMANN, Max; WURSTER, Herbert; ZAGORIS, Konstantinos. Transforming Scholarship in the Archives through Handwritten Text Recognition. **Journal of Documentation**, v. 75, n. 5, p. 954-976. Disponível em: <doi:10.1108/JD-07-2018-0114>.
- NATH, Keshab; DHAR, Sourish; BASISHTHA. Subhash. *web 1.0 to web 3.0 – Evolution of the web and its various challenges*. **International Conference on Reliability Optimization and Information Technology (ICROIT)**. Faridabad, 2014. p. 86-89. Doi: <10.1109/ICROIT.2014.6798297>.
- NÚÑEZ CONTRERAS, Luis Nuñez. **Manual de paleografía**: fundamentos e historia de la escritura latina hasta el siglo VIII. Madrid: Cátedra, 1994,



PEREIRA, Norma Suely da Silva. Fontes primárias e o estudo das práticas culturais na Bahia colonial: dificuldades de leitura. In: LOSE, Alícia Duhá; SOUZA, Arivaldo Sacramento. **Paleografia e suas interfaces**. Salvador: Memória e Arte; Edufba, 2018.

SERRA, Liliana Giusti. **Livro digital e bibliotecas**. Rio de Janeiro: FGV, 2014.

TRANSKRIBUS. Disponível em: <<https://readcoop.eu/transkribus/?sc=Transkribus>>. Acesso: 02 mar. 2021.