

HISTÓRIA DIGITAL: DA TRANSCODIFICAÇÃO À RESSIGNIFICAÇÃO²⁹

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

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

1. INTRODUÇÃO

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

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

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

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

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

3. HISTÓRIA E TRANSCODIFICAÇÃO

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

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

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

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

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

4. HISTÓRIA DIGITAL

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

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

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

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

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

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

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

5. TRANSCODIFICAR E RESSIGNIFICAR

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

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

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

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

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

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

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

REFERÊNCIAS

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

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