

# Lendo milhares de páginas em um dia: uma análise digital de textos críticos produzidos entre os séculos XVIII e XIX

Reading thousands of pages in one day: a digital analysis of critiques produced between the 18th and 19th centuries

---

MÁRCIA ABREU<sup>1</sup>

ADIEL MITTMANN<sup>2</sup>

**RESUMO:** Este artigo<sup>3</sup> apresenta uma ferramenta digital denominada *Digital Library Notes 2* (DLNotes2), cujo propósito é produzir anotações semânticas que permitam a realização de análises de grandes volumes de textos. Neste caso, ela foi empregada para o estudo de críticas sobre romances, produzidas entre o final do século XVIII e o início do XIX, em Portugal. Pretende-se mostrar que se pode obter uma nova compreensão sobre as reações de leitura por meio de anotações digitais e análises estatísticas. A ferramenta foi capaz de identificar traços de estilo e personalidade em dois homens de letras, bem como oferecer indícios sobre as maneiras pelas quais se relacionavam com os romances.

**PALAVRAS-CHAVE:** Análise digital de textos; DLNotes2; circulação de impressos.

**ABSTRACT:** This article presents a digital tool called DLNotes2, whose purpose is to produce semantic annotations that allow the analysis of large volumes of text. In this case, it was used for the study of literary criticism about novels, produced between the end of the eighteenth century and the beginning of the nineteenth, in Portugal. Our intention is to show that a new understanding of reading reactions can be obtained through digital annotations

1. Universidade Estadual de Campinas, Campinas, SP, Brasil.

2. Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil.

3. Este trabalho foi realizado no interior do projeto temático “Circulação Transatlântica dos Impressos — a globalização da cultura no século XIX”, que contou com apoio da FAPESP e do CNPq.

and statistical analyses. The tool was able to identify traits of style and personality in two literary critics as well as to provide clues as to the ways in which they related to the novels.  
 KEYWORDS: Digital text analysis; DLNotes2; circulation of printed material.

*These are new methods we are using, and with new methods  
 the process is almost as important as the results.*

Literary Lab pamphlet 1, p. 1

Muitos recursos já foram usados para produzir anotações enquanto se lê. Há quem faça notas nas margens, há quem grife o texto com canetas coloridas, há quem tenha um caderno para copiar trechos associados a comentários. O *Digital Library Notes 2* (DLNotes2)<sup>4</sup> é uma ferramenta computacional que se inscreve nessa linhagem, permitindo que se façam apontamentos em textos digitais. Boa parte da atividade tradicional de leitura e anotação pode ser realizada por meio da ferramenta, mas ela também permite realizar procedimentos impossíveis (ou muito difíceis) quando se trabalha com papel e caneta, por exemplo, comparar aspectos presentes em centenas de textos que tenham algum denominador comum.

Um grupo de pesquisadores reunido no projeto “Circulação Transatlântica dos Impressos”<sup>5</sup> decidiu empregar essa ferramenta digital para a análise de críticas e comentários sobre romances produzidos entre o final do século XVIII e o final do XIX. Para este artigo, em particular, foi concebido um experimento, cujo objetivo foi examinar como dois homens de letras reagiram aos romances que leram, comparando um conjunto de textos produzidos por eles em uma mesma época e sob condições semelhantes. Interessava, também, verificar se traços de estilo e personalidade poderiam ser percebidos por meio da análise de anotações produzidas com a ferramenta.

Para tanto, este artigo divide-se em três partes. Na primeira, apresenta-se o DLNotes2; em seguida, é feita uma descrição do *corpus* de textos e dos procedimentos de anotação especificamente desenvolvidos para o estudo da recepção crítica de romances. Finalmente, faz-se uma análise do material, visando

4. Produzido pelo Núcleo de Pesquisas em Informática, Literatura e Linguística (NUPILL) e pelo Laboratório de Pesquisas em Sistemas Distribuídos (LAPESD), ambos da Universidade Federal de Santa Catarina. (MITTMANN et al., 2013).
5. O projeto desenvolveu-se entre 2010 e 2016 e contou com financiamento da FAPESP, do CNPq, do FAEPEX-UNICAMP, da Université de Versailles e da Universidade Nova de Lisboa. Há informações detalhadas sobre as pesquisas no site <<http://www.circulacaodosimpressos.iel.unicamp.br/>>, onde se encontram também as referências às publicações realizadas até junho de 2017.

responder às questões levantadas acerca das reações de leitura de dois homens de letras coetâneos.

### A FERRAMENTA DIGITAL LIBRARY NOTES 2 (DLNOTES2)

O DLNotes2 é uma ferramenta, utilizável em navegadores *web*, que permite produzir anotações em conjuntos de textos. Para que as informações sejam coletadas e agrupadas de forma sistemática e eficiente, as anotações não são do tipo tradicional, em que o leitor digita um comentário em forma de texto livre. Em vez disso, ele deve escolher entre um conjunto de elementos predefinidos, preenchendo um formulário.<sup>6</sup>

A definição desse conjunto de elementos – designado como *ontologia computacional* no DLNotes2 – é o ponto central do trabalho, pois ela corresponde à definição de todas as informações e dados que serão pertinentes para as anotações que se queira realizar, bem como de todas as correlações que serão produzidas entre eles. Assim, a *ontologia* deve prever todas as possibilidades de exame de um *corpus* de textos. Ela deve conter as categorias de análise – denominadas *classes* – e as informações a elas associadas – designadas como *atributos*. Uma *classe* pode ser mais detalhada por meio do estabelecimento de elementos internos a ela, nomeados *subclasses*.

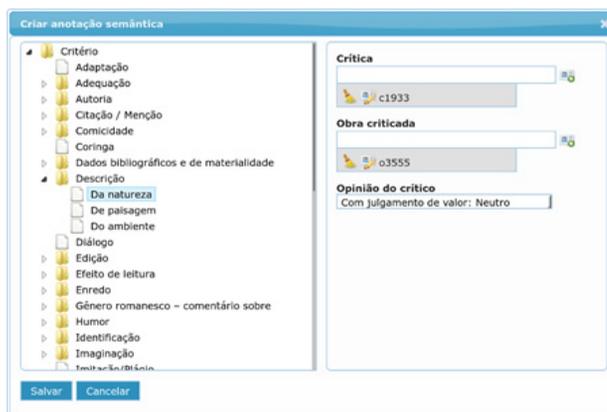


Figura 1. Tela de criação de anotação semântica. – À esquerda estão as classes da ontologia definida para a atividade de anotação, estruturadas em classes e subclasses; à direita estão os atributos, que variam conforme a classe selecionada.

6. Apresentaremos a ferramenta a partir do que é feito no projeto “Circulação Transatlântica de Impressos”. No entanto, ela pode ser configurada para atender diferentes necessidades, inclusive interesses didáticos, como tem sido feito na Universidade Federal de Santa Catarina.

No nosso caso, a *ontologia* contém *classes* e *subclasses* relativas aos elementos empregados pelos críticos para a avaliação dos romances e para a composição de seus textos (que serão apresentados detalhadamente na próxima seção), à nomenclatura empregada para a definição do gênero (que, à época, variava bastante) e ao estabelecimento do resultado geral da avaliação crítica. Contém também *atributos* destinados à identificação da crítica e dos textos criticados (seus títulos, locais e datas de publicação, bem como nome, gênero e nacionalidade de seus autores) e à identificação da opinião do crítico sobre a classe em questão. A Figura 1 exemplifica a estrutura de classes no painel à esquerda e os atributos à direita.

Assim, uma *ontologia* destinada à análise de textos críticos, como esta, pode especificar que há uma *classe* destinada à “Descrição”, que ela tem como *subclasses* “Da natureza”, “De paisagens”, “Do ambiente”, permitindo diferentes graus de minúcia na anotação, ao especificar se o texto aborda genericamente uma descrição ou se trata especificamente de uma descrição de paisagem. Pode especificar, também, que a *classe* (ou a *subclasse*) escolhida receberá *atributos* que identificarão a opinião que o crítico expressou sobre ela.

Os *atributos* variam conforme a *classe*. Se há uma *classe* para indicar que há uma “Citação” em um trecho do texto anotado, faz sentido haver um *atributo* designado para “nome de escritor” e para “título de obra”, para especificar a qual autor ou a qual obra a citação se refere — o que seria totalmente ilógico para a *classe* “Descrição”. Há, também, *atributos* invariáveis, destinados a identificar o texto crítico e a obra criticada.

Estabelecida a *ontologia*, é preciso preparar o *corpus*. O DLNotes2, por funcionar no navegador, precisa que os arquivos do *corpus* estejam no formato HTML (Hypertext Markup Language), que é o padrão da *web*. Arquivos em alguns outros formatos, como é o caso daqueles do Microsoft Word, podem ser convertidos em HTML e, então, enviados ao DLNotes2. No entanto, livros escaneados, mesmo quando em formato PDF, não podem ser utilizados.

Para inserir os textos na ferramenta, o pesquisador cadastra uma *atividade de anotação* e faz o *upload* de um ou vários arquivos. As *atividades de anotação* (isto é, as listas de textos já anotados ou em que serão inseridas anotações) são apresentadas ao pesquisador quando ele entra no sistema (ver Figura 2), podendo conter qualquer número de textos em qualquer tamanho.



Figura 2. Lista de documentos disponíveis na atividade.  
Entre colchetes consta o número de anotações já feitas em cada documento.

É possível que o pesquisador já disponha de textos e de metadados — os dados acerca dos textos — cadastrados em algum sistema antes de iniciar o uso do DLNotes2. Para evitar esforço dobrado, um protocolo de sincronização pode ser estabelecido entre o outro sistema e o DLNotes2, permitindo que eles sejam importados de uma só vez. Os metadados também podem ser sincronizados, o que faz com que eles se tornem disponíveis no momento em que o pesquisador está criando as suas anotações.

No caso da pesquisa desenvolvida no projeto “Circulação Transatlântica de Impressos”, havia um banco de dados denominado CiTrIm, em que haviam sido cadastrados milhares de textos críticos sobre romances, produzidos em diversas partes do mundo. O banco, além de conservar o texto integral da crítica, armazena informações como autoria, local de publicação, título do romance analisado etc. O DLNotes2 e o CiTrIm cooperam de modo que os textos e os metadados residentes primariamente no CiTrIm sejam sincronizados com o DLNotes2, fazendo com que os *atributos* relativos à identificação da crítica e de seu autor sejam convertidos em códigos e automaticamente preenchidos (ver Figura 1, onde os códigos aparecem). Quando as anotações são extraídas do DLNotes2 para serem analisadas de forma quantitativa, os metadados as acompanham, carregando informações sobre, por exemplo, a data e o local em que o texto crítico foi produzido, permitindo observar, assim, como a quantidade de comentários sobre determinada categoria variou ao longo do tempo em determinado lugar.

Com os textos e os metadados inseridos no DLNotes2, é possível começar a produzir as *anotações semânticas*. Trata-se de um processo razoavelmente simples. O pesquisador, enquanto lê um dos textos críticos, seleciona com o *mouse* o trecho ao qual quer associar uma anotação. Nesse momento, o DLNotes2 mostra uma tela como aquela da Figura 1. À esquerda encontram-se as *classes* e as *subclasses* predefinidas e, à direita, estão os atributos disponíveis para a categoria selecionada. Assim, uma *anotação semântica* é a associação de um trecho de um texto com uma *classe* (ou *subclasse*) e seus *atributos*, resultando em uma *instância*, ou seja, um item de análise.

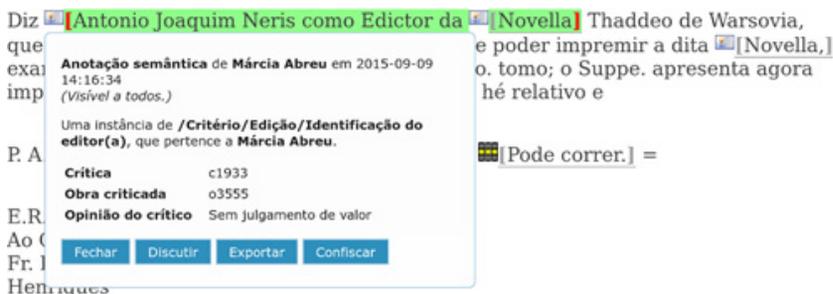


Figura 3. Instância criada e associada ao trecho do texto.

O trecho selecionado fica claramente delimitado por colchetes e sublinhado, como mostra a Figura 3. A marcação é livre, podendo se restringir a uma única palavra ou abranger o texto inteiro. A anotação salva se reduz a um ícone, que precede o trecho anotado e que serve de âncora — ao clicar sobre ele, a *instância* produzida se abre, exibindo os apontamentos feitos. Várias classes podem ser associadas a um mesmo trecho, tornando a anotação mais completa — por exemplo, um trecho da crítica pode afirmar que o enredo do romance é licencioso, o que produzirá duas anotações, uma relativa à classe *enredo* e outra relativa à classe *moral*.

Para que o leitor tenha uma visão geral das anotações feitas em um texto, os ícones são configuráveis: eles podem refletir a *classe* (por exemplo, um ícone em forma de livro associado à *classe* “Citação”) ou um de seus *atributos*, como foi estabelecido no nosso caso, em que os ícones refletem o *atributo* “opinião do crítico”, como se vê na Figura 4. As carinhas verdes são opiniões positivas; as amarelas, neutras; as vermelhas, negativas; e as caixinhas quadradas são informações em que não há juízo de valor.

●[Quanto a dizer o Autor que não escrevera a pag 19 §2 mai mas sim mui cessa a reflexão que fize suposto se pode bem equivocar do modo com que se escrito Similhante a este he aquelle lugar a pag 20V linh 4a e 5a do §2 nas palavras hum encontro que parece ler-se "hum encontra. Quanto á expressão do Autor a pag 13 linh 8a "as mais honestas por entre ellas gostao &a" como o Autor emendou o periodo tem cessado a minha reflexão . Equanto [sic] finalmente á 😊[nota que o Autor julga desnecessaria a pag 12V § ult. referindose áquella sentença quam glorioso he enganar huma formosa Donzella = devia o Autor advertir que eu não a exigi de neccessidade, mas que 🟢[tao somente me pareceo conveniente pôr-se para inspirar mais horror ao vicio.] Todas as mais passagens que notei como capitaes para se formar juizo destes defeitos, as acho emendadas agora pelo Autor, e parece que elle poderia emendar tao bem algumas mais que omitti por não formar maior catalogo, como são logo na pag 1a §5 não se que dizer-te = em lugar de = não sei que te diga = e no paragrapho 6o em que principia huma oração com hum ponto final no meio, sem acabar o sentido. E no §pho 7o e ultimo da mesma pagina em que 🟠[parece estar se ouvindo hum estrangeiro fallando mal a Lingoa Portugueza - Gosto muito de tu ter o pensamento] &c.]

Figura 4. Trecho de texto anotado com o DLNotes2.

Na tela do computador, as carinhas têm cores diferentes para refletir a opinião do crítico.

A ontologia mais todas as instâncias criadas pelos pesquisadores formam uma base de conhecimento, ou seja, o banco de dados como um todo, que pode ser analisado por outras ferramentas computacionais (como planilhas eletrônicas ou pacotes estatísticos) e para o qual se podem fazer as mais diversas questões — o que seria impossível com anotações livres do tipo “aqui o autor critica negativamente uma descrição de ambiente”.

Os dados podem ser exportados em formato de planilha, em que as linhas correspondem às anotações e as colunas, a diferentes propriedades das anotações. Assim, existem colunas que indicam o nome do pesquisador que realizou a anotação, a data em que ela foi feita, o texto no qual foi criada e o trecho que foi demarcado. Os atributos definidos para as classes também podem ser exportados, inclusive aqueles oriundos de uma integração com outros sistemas.

Classe	Crítica/ID	Opinião do crítico	Crítica/Autor/Nome	Crítica/Ano
/Critério/Edição/Identificação do editor(a)	c1933	Sem julgamento de valor	Izidoro Ignacio Henrique	1829
/Critério/Moral	c1933	Com julgamento de valor: Positivo	Izidoro Ignacio Henrique	1829
/Critério/Temas abordados pelo crítico/Política	c1933	Com julgamento de valor: Positivo	Izidoro Ignacio Henrique	1829
/Critério/Temas abordados pelo crítico/Religião	c1933	Com julgamento de valor: Positivo	Izidoro Ignacio Henrique	1829
/Critério/Autoria	c1934	Sem julgamento de valor	Fr. Joze Machado	1824
/Critério/Edição	c1934	Sem julgamento de valor	Fr. Joze Machado	1824
/Critério/Moral	c1934	Com julgamento de valor: Positivo	Fr. Joze Machado	1824
/Critério/Tradução	c1934	Sem julgamento de valor	Fr. Joze Machado	1824
/Critério/Tradução/Qualidade da	c1934	Com julgamento de valor: Positivo	Fr. Joze Machado	1824
/Critério/Autoria	c1935	Sem julgamento de valor	Matheus da Assumpção	1824
/Critério/Enredo/Quantidade de episódios	c1935	Sem julgamento de valor	Matheus da Assumpção	1824
/Critério/Gênero romanesco – comentário sobre	c1933	Com julgamento de valor: Negativo	Izidoro Ignacio Henrique	1829
/Critério/Gênero romanesco – comentário sobre	c1935	Com julgamento de valor: Neutro	Matheus da Assumpção	1824
/Critério/Gênero romanesco – comentário sobre	c1935	Com julgamento de valor: Negativo	Matheus da Assumpção	1824
/Critério/Moral/Relativa ao enredo	c1935	Com julgamento de valor: Negativo	Matheus da Assumpção	1824

Figura 5. Trecho de banco de dados obtido a partir de anotações feitas no DLNotes2.

A Figura 5 mostra um trecho de um banco de dados exportado a partir do DLNotes2. Com base nele, ferramentas como o Microsoft Excel podem ser empregadas para filtrar e explorar os dados. O recurso de tabela dinâmica do Excel é particularmente útil para gerar tabelas e gráficos que contem e agrupem os dados presentes na planilha.

#### UMA ONTOLOGIA COMPUTACIONAL PARA A RECEPÇÃO CRÍTICA DE ROMANCES

Recorremos ao DLNotes2 a fim de analisar grandes volumes de textos críticos produzidos acerca de romances entre meados do século XVIII e meados do XIX. Este período é particularmente interessante, pois nele se desenrolou um processo de canonização do gênero, que passou da desvalorização, tendo em vista sua novidade e popularidade, à aceitação como um gênero literário de prestígio (AUGUSTI, 2011). Diversos agentes tomaram parte neste movimento, que ocorreu em distintas partes do mundo simultaneamente.

Parte deste material, produzido na Inglaterra, na França, em Portugal e no Brasil, foi recolhido pelos pesquisadores do projeto e inserido no DLNotes2, a fim de que fosse possível estabelecer comparações e observar a existência de consonâncias e discrepâncias ao longo do tempo e em diferentes lugares (ASSUMPÇÃO, 2015; GABRIELLI, 2015). Os textos críticos provêm de fontes diversas: publicações em periódicos (generalistas ou especializados em literatura) produzidos nos quatro países; pareceres de censura destinados a aprovar a publicação ou a circulação de romances em Portugal e no Brasil;<sup>7</sup> bibliografias críticas publicadas na França, comentando os romances tidos como mais relevantes ou mais aceitáveis.

Embora as críticas sejam bastante diversas do ponto de vista de sua finalidade, forma de circulação e local de produção, elas se assemelham pelo fato de terem sido produzidas em um mesmo período e acerca de um mesmo tipo de texto — as narrativas em prosa ficcional. Ainda que diversas, elas são notavelmente homogêneas nos elementos empregados para a avaliação dos romances e, até mesmo, na forma de composição dos textos (ABREU, 2016). Assim, tendo lido algumas centenas deles, foi possível definir a *ontologia* com a qual trabalharíamos. Estabelecemos as *classes*, ou seja, os critérios empregados pelos críticos na avaliação de romances e os

7. Os pareceres produzidos pelos censores luso-brasileiros têm muitas características semelhantes aos textos críticos publicados em periódicos, extrapolando, em muito, a avaliação da ortodoxia das obras. Para uma discussão detalhada sobre as semelhanças entre a atividade crítica e a censória, ver Abreu (2008a, 2011).

elementos utilizados na confecção de seus textos.<sup>8</sup> As *classes* contêm um conjunto de elementos de avaliação textual: “adequação” (ou decoro), “descrição”, “diálogo”, “enredo”, “humor”, “ironia”, “linguagem”, “narrador”, “moral”, “personagem” e “verossimilhança”. Ainda como avaliação textual, há *classes* relativas à questão da “nacionalidade”, à qualidade da “tradução”, à “originalidade” da composição, a que se contrapõem o “plágio” e a “imitação”.

Elementos extratextuais também foram definidos nas *classes*: a “autoria”, os “dados bibliográficos e de materialidade”, a “edição” (enquanto atividade do editor), os “efeitos de leitura” provocados pelo texto, a “imaginação” do escritor, os “paratextos”, o “público leitor” e a “recepção” do romance.

Além disso, sabíamos que havia discussões sobre o próprio gênero romanesco, tratando de sua nomenclatura, de seus subgêneros e de seu valor; portanto, uma *classe* foi criada para anotar esses elementos.

Estávamos interessados também em observar procedimentos discursivos empregados pelos críticos ao elaborar seus textos, por isso criamos *classes* relativas ao uso de “citações” e “menções”, ao recurso à “ironia” ao comentar o romance, à elaboração de reflexões “metacríticas” sobre o próprio ato de avaliar obras, à busca de “identificação” de modelos ou, mesmo, da vida do autor à obra. Finalmente, observamos que os críticos também faziam comentários genéricos, às vezes com pouca ou nenhuma conexão com o romance lido, por isso criamos a *classe* “temas abordados pelo crítico”, que tem como *subclasse* “política”, “filosofia”, “história”, “religião”, “sociedade”, “vida pessoal”, entre outras. Praticamente todas as *classes* contêm *subclasses*, que detalham os critérios, mas que não serão apresentadas aqui por razão de brevidade.<sup>9</sup>

Como os textos críticos, em geral, eram judicativos, criamos *atributos* para anotar a opinião expressa pelo autor sobre cada uma das *classes*, informando se havia

8. A definição das *classes* foi feita a partir da leitura de centenas de textos críticos produzidos na primeira metade do século XIX, às quais foram associados critérios comumente empregados pela teoria literária posterior. A definição dos elementos de análise e a produção das anotações semânticas que validaram o sistema foram realizadas por uma equipe de pesquisadores, coordenada por Márcia Abreu: Isabella Spatti Candido (bolsista Iniciação Científica FAPESP), Beatriz Gabrielli (bolsista Iniciação Científica FAPESP e, atualmente, Mestranda com bolsa do CNPq), Larissa de Assumpção (bolsista de Iniciação Científica FAPESP e, atualmente, Mestranda com bolsa da mesma instituição), Priscila Velloni (bolsista de Treinamento Técnico FAPESP), João Lucas Magalhães Moraes (bolsista SAE – UNICAMP), Taís Francison (Mestranda, CNPq), Etienne Sauthier (Pós-doutorado FAPESP).
9. A *ontologia* é explicada com mais detalhes alhures (ABREU; MITTMANN, 2017).

ou não juízo de valor; em caso afirmativo, é possível definir se ele era “positivo”, “negativo”, “neutro” e, em caso contrário, anota-se “sem julgamento de valor”.

Ao final das anotações, há uma *classe* reservada ao “resultado da crítica”, que possui como *atributos* “positivo”, “negativo”, “neutro”, “predominantemente positivo” ou “predominantemente negativo”.

#### UMA ANÁLISE DIGITAL DAS REAÇÕES DE DOIS LEITORES DE ROMANCES

O banco de dados do projeto “Circulação Transatlântica dos Impressos” (CiTrIm) contém, atualmente, 3.258 textos críticos produzidos na França, no Reino Unido, no Brasil e em Portugal entre 1797 e 1850, cadastrados com metadados relativos ao título da crítica e seu autor, local e data de publicação, à língua em que foram escritos e a que romances se referiam. O nome do autor da crítica também está associado a metadados que informam seu gênero, o local e a data de seu nascimento e morte. Da mesma forma, as obras criticadas vêm acompanhadas de informações sobre seu autor, língua em que foi escrita originalmente e gênero textual. Caso a crítica explicita o fato de estar comentando uma edição específica, esses dados também são acoplados à identificação da obra, que passa a conter informações sobre local, data e número da edição, e sobre tradução e adaptação, nos casos em que isso é pertinente. Desse conjunto, 2.333 textos críticos foram incorporados ao DLNotes2 e receberam anotações.

Para este artigo, recortamos uma pequena parte desse material, selecionando textos críticos produzidos por dois censores portugueses,<sup>10</sup> a fim de verificar se o DLNotes2 produziria dados interessantes para análise de características individuais de leitura e avaliação, além de elementos que permitissem pensar sobre estilo e personalidade.

Entre o final da década de 1790 e o início dos anos de 1820, dois homens leram romances incessantemente em Lisboa: João Guilherme Christiano Müller e Francisco Xavier de Oliveira. Eles estavam a serviço da Coroa e trabalhavam como censores, examinando os livros que pretendiam circular em Portugal e seus domínios, em língua original ou em tradução, fossem eles impressos no país ou importados de diversas localidades. Havia muitos outros censores no período, mas João Guilherme Christiano Müller e João Xavier de Oliveira foram os que receberam

10. Para uma discussão sobre a proximidade entre a atividade censória e a atividade crítica, ver Abreu (2008a; 2008b; 2011; 2013).

maior número de romances para avaliar, o que faz deles, possivelmente, os maiores leitores de romance da cidade, uma vez que liam tudo o que seus contemporâneos leriam, mais aquilo que saía de circulação devido à ação censória. Embora as determinações legais exigissem que eles se manifestassem apenas sobre a ortodoxia moral, religiosa e política das obras, seus pareceres excediam em muito essas atribuições, convertendo-se em verdadeiras peças de crítica literária, em que se discutiam problemas de linguagem e verossimilhança, construção de personagens e enredo, qualidade da tradução e da forma material do impresso, entre outras.

Ambos trabalhavam para as mesmas instituições, sob as mesmas determinações, chegando, algumas vezes, a ler os mesmos textos. Mas, nem por isso, chegavam às mesmas conclusões. Uma análise numérica de suas reações aos romances indica que eles estavam em polos opostos. Considerando a *classe* “Resultado da crítica”, Oliveira avaliou 50,0% dos romances que leu como “negativo” ou “predominantemente negativo”, enquanto Müller os considerou como “positivo” ou “predominantemente positivo” em 68,4% dos casos.

Este resultado é consistente com os comentários que cada um deles fez ao longo de seus pareceres. Em números absolutos, Francisco Xavier de Oliveira produziu 100 comentários negativos contra apenas 26 positivos em suas avaliações. Já João Guilherme Christiano Müller fez 232 comentários positivos contra 183 negativos. Ou seja, Müller avaliou positivamente 55,90% dos elementos que analisou nos romances, enquanto Oliveira o fez em apenas 20,63% dos casos. Para perceber quão relevante é o índice de positividade nos comentários de Müller, basta ver que as críticas inglesas publicadas em periódicos no mesmo período apresentam 48,71% de positividade.

Suas diferenças em relação aos romances são perceptíveis de várias formas. Oliveira escreveu, por crítica, mais comentários negativos (8,3 contra 4,8 de Müller) e menos comentários positivos (2,17 contra 6,11 de Müller). E foi mais enfático do que seu colega, ao apresentar os defeitos das obras que leu: Oliveira empregou, em média, 30,39 palavras quando quis manifestar sua reprovação e apenas 9,38 quando tratou de elogiar. Müller parece mais equilibrado, empregando, em média, 30,54 palavras para desaprovar e 22,72 para exaltar as qualidades do livro.

Estes dados mostram não apenas uma atitude mais favorável aos romances por parte de Müller, mas deixam evidente que para reprovar utilizavam-se mais palavras do que para enaltecer, uma vez que, em ambos os casos, os trechos negativos são significativamente mais longos do que os positivos. Um maior detalhamento nos comentários negativos não ocorre, por exemplo, nas críticas inglesas, em que

reprovação e elogio empregavam mais ou menos o mesmo número de palavras (31,89 e 32,60, respectivamente). A exposição minuciosa dos elementos negativos, nos pareceres produzidos no interior da censura lusitana, talvez se deva à necessidade de argumentar, a fim de convencer o tribunal censório — que era quem tomava a decisão, em última instância — sobre a necessidade de proibição da obra. Não havia qualquer necessidade de mostrar as qualidades do texto, mas, mesmo assim, centenas de palavras eram gastas para enaltecer as belezas do romance, mesmo por parte de um censor pouco afeito a eles como Francisco Xavier de Oliveira.

É possível verificar também que eles pensavam sobre os textos de maneiras diferentes. Embora ambos se irmanem, ao considerar o enredo como um elemento central de suas análises (75,0% dos comentários de Oliveira e 71,1% de Müller), Francisco Xavier de Oliveira dá prioridade à observação de elementos textuais, como linguagem (58,3% dos comentários de Oliveira e 36,8% de Müller) e verossimilhança (25,0% dos comentários de Oliveira e 13,2% de Müller), enquanto Müller se preocupa mais com elementos extratextuais, como os efeitos que a leitura poderia provocar no público (presente em 57,9% de seus comentários, contra 25,0% de Oliveira) ou a recepção que a obra obteve no mundo literário (52,6% dos comentários de Müller e 0,0% de Oliveira).

Assim, Oliveira volta-se mais para a qualidade interna da obra, enquanto Müller inquieta-se com as consequências de sua presença no mundo social e, em particular, com as avaliações que seus pares fizeram dos livros que ele deveria julgar. Isso tem impactos na maneira como eles organizam seus textos e seus argumentos. Enquanto Müller refere a recepção crítica da obra como fundamento de suas opiniões (Müller o faz em 52,63% dos casos, enquanto Oliveira em 0,0%), Oliveira recorre a citações de autores, especialmente Horácio (75,00% dos comentários de Oliveira contra 52,63 de Müller).

Parte dessas diferenças pode ser explicada pela inserção profissional desses dois letrados. João Guilherme Christiano Müller era filho de uma família de professores da renomada Universidade de Gottingen, na Alemanha. Nascido em 1752, cursou Humanidades, Línguas orientais e Teologia naquela instituição. Em 1773, mudou-se para Portugal para tornar-se pastor luterano, mas acabou por converter-se ao catolicismo, ocupando diversos postos no governo: foi professor dos príncipes, tradutor do Almirantado, bibliotecário da Rainha D. Maria I, sócio supranumerário da Academia Real das Ciências, diretor da Impressão Régia, além de ter sido censor régio desde

1792 até sua morte, em 1814.<sup>11</sup> Mantinha estreito relacionamento com homens ricos, influentes e cultos residentes em Lisboa (FERRO, 1991; FERRO, 1996, p. 315 e segs.).

Pouco se sabe sobre Francisco Xavier de Oliveira, que, em um de seus pareceres menciona o fato de ter nascido na América. Foi Professor Régio de Retórica e Poética, em Lisboa, a partir de 1771, assumindo, a seguir, o cargo de censor régio, em 1792, função que desempenhou por, pelo menos, 20 anos, emitindo pareceres até 1812.<sup>12</sup> É incerta a data de sua morte, mas um documento de junho de 1819 menciona seu nome como falecido.<sup>13</sup>

Parece fazer sentido que um professor de retórica e poética prestasse muita atenção a elementos textuais, como verossimilhança e linguagem, enquanto um homem do Paço estivesse mais preocupado com o impacto social das obras aprovadas; que o professor de retórica citasse textos clássicos da Antiguidade, enquanto o membro da Academia levasse em conta a reputação da obra no mundo literário.



Figura 6. Palavras mais utilizadas nos comentários negativos de Müller (à esquerda) e Oliveira (à direita).

11. Sua biografia foi apresentada por Francisco Manuel Trigoso de Aragão Morato, seu colega na Academia de Ciências de Lisboa, na Assembléa Pública de 24 de junho de 1815 (MORATO, 1815). Ver também verbete relativo a João Guilherme Cristiano Müller na *Grande Enciclopédia Portuguesa e Brasileira*. (s.d., p. 115-116).
12. Segundo Innocencio da Silva (2001, p. 93), Francisco Xavier de Oliveira foi “Professor Regio de Rhetorica e Poetica em Lisboa pela resolução regia de 10 de Novembro de 1771, com exercicio no Collegio Real de Nobres, e depois no antigo estabelecimento d’estudos do bairro d’Alfama. Impossibilitado a final, por sua idade e molestias chronicas acompanhadas de cegueira, viveu assim alguns annos, morrendo (ao que parece) no de 1823, ou pouco depois”.
13. “Relação dos Censores falecidos, impedidos, e mais ou menos desembaraçados, depois da Real Resolução de 18 de Março de 1807 que substituiu outros falecidos e impedidos que haviaõ sido Nomeados por Decretos de 28 de Agosto de 1795, e 31 de Mayo de 1798” (ARQUIVO NACIONAL DA TORRE DO TOMBO).

Os dados produzidos pelo DLnotes2 permitem fazer ainda outras ilações, pois mostram características que poderiam ser atribuídas à personalidade de cada um desses homens. Uma das marcas mais salientes nos pareceres de Francisco Xavier de Oliveira é o emprego da ironia para se referir aos romances que examinava (58,33%) ou aos seus autores (25,00%). Enquanto isso, uma característica marcante nos documentos produzidos por João Guilherme Christiano Müller é o caráter autorreflexivo de suas observações, presente em 23,68% dos casos. Assim, enquanto Oliveira faz graça com defeitos das obras ou com as intenções de autores e tradutores, Müller se questiona sobre seu papel, expõe dúvidas sobre seus julgamentos e lamenta-se sobre a dificuldade das tarefas que deve desempenhar. A Figura 6<sup>14</sup> reforça as características próprias de cada crítico: nela estão exibidas as 20 palavras mais utilizadas nos trechos negativos. Enquanto termos como “livro”, “pag”, “ibid” e “fol” destacam-se nos comentários negativos de Müller, sugerindo que o autor se preocupa em ser específico quando é negativo, em Oliveira vemos “portuguez”, “portuguesa”, “lingoa” e “traductor”, o que indica que este crítico dedicava atenção especial à língua, sobretudo nas traduções.

Observa-se, assim, uma interessante coincidência entre ironia e negatividade; autorreflexão e positividade. Percebe-se também que, para reprovar, se utilizavam mais palavras do que para enaltecer, já que os trechos negativos de ambos os censores são significativamente mais longos do que os positivos — o que é particularmente digno de atenção no caso de Müller, que revela uma atitude geral mais favorável aos romances.

Finalmente, os dados mostram que os organismos de censura do Antigo Regime não eram entidades monolíticas, abrigando, pelo contrário, opiniões divergentes e ímpetus reprobatórios diversos. Além disso, podiam ser um espaço de elogio às obras lidas, como ocorre em ambos os casos examinados, ainda que com ênfases diferentes.

## CONCLUSÃO

Ao iniciar o trabalho com o DLNotes2, tínhamos um conhecimento amplo sobre a crítica dos séculos XVIII e XIX, suficiente para estabelecer elementos de análise e definir critérios de avaliação, ou seja, para especificar uma *ontologia*. No

14. As nuvens de palavras foram criadas com o auxílio do sítio <<http://www.wordclouds.com/>>.

entanto, o próprio trabalho de estabelecimento de uma *ontologia* força a uma nova e diferente reflexão sobre o objeto, pois não se trata apenas de lembrar quais foram os critérios de análise que, um dia, vimos em algum texto crítico, mas, sim, de organizá-los de maneira lógica, criando hierarquias entre *classes* e *subclasses*, estabelecendo relações entre *classes* e *atributos*, que permitam colocar novas questões para textos críticos produzidos há mais de cem anos.

Extraír dados e identificar padrões a partir de anotações tradicionais é, obviamente, possível, mas é muito difícil fazê-lo, e os resultados não são tão exatos como os obtidos a partir de anotações digitais estruturadas. Some-se a isso o limite que um ser humano tem de leitura, que o obriga, necessariamente, a trabalhar com uma quantidade restrita de dados. No caso do DLNotes2, um grupo de pesquisadores pode colaborar para anotar um *corpus*. Todas as anotações feitas nos textos, independentemente do pesquisador que as criou, são igualmente conservadas e podem ser analisadas de forma conjunta. Assim, diversas pessoas podem anotar um único texto ou podem dividir o trabalho entre si, atribuindo um conjunto de textos a cada pesquisador.

Isso permite trabalhar com uma quantidade muito maior de documentos, mas também cria a necessidade de zelar pela consistência dos dados, o que leva a discussões de interpretação que refinam a compreensão do texto.

O DLNotes2 é uma ferramenta digital interessante, pois, diferentemente de programas empregados em algumas pesquisas em Humanidades Digitais, ela não produz as anotações automaticamente. São os seres humanos que leem, selecionam trechos, criam anotações e interpretam textos, associando-os a elementos de análise predefinidos. Assim, a subjetividade se mantém, e as diferenças de leitura afloram. Se, no grupo de pesquisadores, não costuma haver discussão sobre o fato de o crítico estar comentando a construção de uma *personagem* ou tecendo considerações sobre a *linguagem*, a decisão sobre o fato de seu comentário ser positivo, neutro ou negativo pode levar a acaloradas discussões, já que, nesses casos, a interpretação do pesquisador sobre o texto pesa mais fortemente. Por esse motivo, as anotações são periodicamente revisadas e discutidas pelo grupo.

Assim, esse sistema tem algumas vantagens interessantes. Ele permite o trabalho coletivo, mantendo a subjetividade das leituras, possibilita o compartilhamento dos dados entre vários pesquisadores, alarga enormemente a quantidade de textos considerados e, uma vez criadas as anotações, gera resultados com grande rapidez.

Além dessas vantagens operacionais e metodológicas, o sistema é valioso por permitir enxergar coisas que antes não estavam visíveis ou eram pouco claras.

Uma primeira surpresa foi perceber como eram restritos os elementos empregados para se pensar sobre romances até meados do século XIX. Definimos 14 critérios de análise textual, o que já nos parecia bastante sucinto. Para nossa surpresa, apenas 5 critérios textuais foram realmente produtivos na avaliação de romances — “enredo”, “personagem”, “moral”, “verossimilhança” e “linguagem” (incluindo observações sobre a qualidade da “tradução”) — não apenas nos casos aqui examinados, mas no conjunto de textos críticos produzidos em Portugal, no Brasil, no Reino Unido e na França.

Além disso, nossa percepção sobre o peso desses critérios e sua forma de utilização foi bastante alterada, desde que começamos a obter resultados que permitiram olhar para centenas de textos de uma só vez e com um mesmo padrão. Por exemplo, por muito tempo acreditamos que a avaliação do aspecto moral dos romances era o elemento mais mencionado pelos críticos até a primeira metade do século XIX; entretanto, a quantificação produzida a partir do DLNotes<sup>2</sup> mostrou que o elemento de maior destaque eram os comentários sobre o enredo, aí incluída a apresentação de um resumo da obra. Embora as considerações morais fossem também importantes (elemento destacado por Müller em 73,68% dos casos, contra 50% nos pareceres redigidos por Oliveira), as observações sobre o enredo eram tão ou mais relevantes (Müller menciona o enredo em 71,05% dos casos, e Oliveira o faz em 75% das vezes) – o que é particularmente curioso numa situação em que se analisam textos críticos produzidos no interior de organismos de censura, os quais, por hipótese, deveriam estar mais preocupados com a moral do que com a trama das narrativas.

A leitura humana, ou o “Humanscope”, na sugestiva designação de Sarah Allison et al. (2011), havia dado pouca atenção ao elemento mais presente nos textos críticos — o enredo —, possivelmente por ser ele corriqueiro e pouco rentável em termos de análise. Já a moral chamara a atenção, provavelmente porque a teoria literária do século XX a descartara inteiramente de suas preocupações, tornando esse critério de análise *sui generis* para os olhos de um leitor contemporâneo. Assim, o DLNotes<sup>2</sup> minimizou a interferência de ideias preconcebidas pelos pesquisadores.

Por outro lado, o “Humanscope” havia detectado a evidente ironia que perpassa os textos de Oliveira, mas a relação entre essa característica e o grau de negatividade de suas avaliações tinha passado inteiramente despercebido. Mais inesperada ainda foi a relação entre a reflexão metacrítica em que Müller se enreda e a positividade de

suas avaliações. Essas relações tornaram-se evidentes quando os dados mostraram não apenas as porcentagens de aprovação, mas também a quantidade de palavras empregadas em cada situação.

Assim, o DLNotes2 não substitui a leitura humana, mas a torna mais atenta e aguçada, permitindo a realização de interpretações novas sobre dados há muito conhecidos.

## REFERÊNCIAS

- ABREU, Márcia. Censure et critique – les réactions des premiers lecteurs de romans. *Cahiers du Brésil Contemporain* – Maison des Sciences de l’Homme, Paris, n. 69-70, p. 11-35, 2008a.
- \_\_\_\_\_. O “Mundo Literário” e a “Nacional Literatura”: leitura de romances e censura. In: ABREU, Márcia (Org.). *Trajétórias do romance: circulação, leitura e escrita nos séculos XVIII e XIX*. Campinas-São Paulo: Mercado de Letras; FAPESP, 2008b. p. 275-306.
- \_\_\_\_\_. Nos primórdios da crítica - julgamentos literários produzidos pela censura luso-brasileira. In: FIGUEIREDO, Carmen Lúcia N. de; HOLANDA, Sílvio Augusto de O.; AUGUSTI, Valéria (Org.). *Crítica e literatura*. Rio de Janeiro: De Letras, 2011. p. 197-220.
- \_\_\_\_\_. Escrever sob censura – considerações históricas e literárias. In: ANDRADE, Francisco Eduardo de; GONÇALVES, Andréa Lisly; JESUS, Ronaldo Pereira de. *Itinerários da pesquisa histórica: métodos, fontes e campos temáticos*. Belo Horizonte: Fino Traço, 2013. p. 103-121.
- \_\_\_\_\_. Uma comunidade letrada transnacional: reação aos romances na Europa e no Brasil. In: ABREU, Márcia (Org.). *Romances em movimento: a circulação transatlântica dos impressos (1789-1914)*. Campinas-São Paulo: Editora da Unicamp; FAPESP, 2016. p. 365-394.
- ABREU, Márcia; MITTMANN, Adiel. Ler o passado com ferramentas do futuro: uma análise digital de textos críticos do início do século XIX. *Alea: Estudos Neolatinos*, v. 19, n. 3, p. 661 - 677, 2017.
- ALLISON, Sarah et al. Quantitative formalism: an experiment. *Literary Lab - Pamphlet 1*, p. 8, January 15, 2011. Disponível em: <<https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>>. Acesso em: 15 ago. 2017.
- ASSUMPÇÃO, Larissa de. 2015. *O lugar do romance em bibliotecas oitocentistas* – a presença de obras ficcionais em livros sobre a formação de bibliotecas e nos catálogos da Biblioteca Fluminense e da Biblioteca Imperial. Monografia (Curso em Letras)- Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2015.
- AUGUSTI, Valéria. *Trajétórias de consagração: discursos da crítica sobre o romance no Brasil Oitocentista*. Campinas: Mercado de Letras, 2011.
- FERRO, João Pedro. *Influências germânicas em Portugal na segunda metade do século XVIII*. Separata das Atas do Congresso Internacional Portugal no século XVIII de D. João V à Revolução Francesa. Sociedade Portuguesa de Estudos do Século XVIII. Lisboa: Universitária Editora, 1991.
- \_\_\_\_\_. Intelectuais alemães em Portugal: Johann Wilhelm Christian Müller (1752 – 1814). In: MARQUES, Antonio Henrique de Oliveira; OPITZ, Alfred; CLARA, Fernando. (Coord.).

*Lendo milhares de páginas em um dia: uma análise digital de textos críticos...*

- Portugal – Alemanha – África do imperialismo colonial ao imperialismo político*. Actas do Encontro Luso-Alemão, 4., 2 e 4 de outubro de 1995, Lisboa. Lisboa: Colibri, 1996.
- GABRIELLI, Beatriz. *A presença de romances no jornal francês La Presse (1836 – 1850)*. Monografia (Curso em Letras)- Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2015.
- GRANDE ENCICLOPÉDIA PORTUGUESA E BRASILEIRA. Lisboa e Rio de Janeiro: Editorial Enciclopédia, v. XVIII, p. 115-116, s.d.
- MITTMANN, Adiel et al. DLNotes2: anotações digitais como apoio ao ensino. In: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 24., 25 a 29 de novembro de 2013, Campinas-SP. *Anais...* 2013. p. 527-536.
- MORATO, Francisco Manuel Trigoso de Aragão. Elogio histórico de João Guilherme Christiano Müller. In: ACADEMIA REAL DAS CIÊNCIAS DE LISBOA. *Memórias da Academia Real das Ciências de Lisboa*, lic. Paul Wilhelm Gennrich, 1815. Tomo IV, parte II.
- SILVA, Innocencio Francisco da. *Diccionario Bibliographico Portuguez*. V. III. Ophir – Biblioteca Virtual dos Descobrimentos Portugueses, 9, 2001.

#### SOBRE OS AUTORES

Márcia Abreu é professora do Departamento de Teoria Literária do IEL – UNICAMP. Publicou vários trabalhos, dentre os quais se destacam os livros *The transatlantic circulation of novels between Europe and Brazil, 1789-1914* (Palgrave MacMillan, 2017); *Romances em movimento* (Editora da Unicamp, 2016); *The cultural revolution of the nineteenth century: Theatre, the book-trade and reading in the transatlantic world* (org. com Ana Claudia Suriani da Silva. I. B. Tauris, 2015); *Impresso no Brasil – dois séculos de livros brasileiros* (org. com Aníbal Bragança. Editora da Unesp / Fundação Biblioteca Nacional, 2010), com o qual ganhou o Prêmio Jabuti 2011; *Cultura letrada: literatura e leitura* (UNESP, 2006); *Os caminhos dos livros* (Mercado de Letras; ALB; FAPESP, 2003) e *Histórias de cordéis e folhetos* (Mercado de Letras; ALB, 1999).

*E-mail: marcia.a.abreu@gmail.com.*

Adiel Mittmann possui mestrado e doutorado em Ciências da Computação pela Universidade Federal de Santa Catarina (2016). Atua principalmente nos seguintes temas: Humanidades Digitais, Versificação, Linguística Computacional.

*E-mail: adiel@mittmann.net.br.*

*Recebido em 31 de março de 2017 e aprovado em 26 de abril de 2017.*