

Crítica textual e tecnologia*

Textual scholarship and technology

Luana Batista de Souza 

E-mail: luana.souza@alumni.usp.br

Universidade de São Paulo, São Paulo, SP, Brasil

Resumo

O cotejo de testemunhos é um trabalho árduo e que requer atenção do editor. Com o avanço das Humanidades Digitais, o crítico textual pode ter seu trabalho auxiliado por programas desenvolvidos para o cotejo e edição crítica de textos de diferentes épocas. Esse artigo tem como objetivo apresentar alguns desses *softwares* disponíveis *on line* para o cotejo de documentos e preparo de edição crítica digital, a saber: *CollateX*, *DV-Coll*, *Juxta Commons* e *FairCopy*, e mostrar quais são os aspectos positivos e negativos do seu uso, além dos desafios impostos por essas ferramentas. Como exercício prático, apresentaremos a colação do poema *José*, de Carlos Drummond de Andrade, publicado pela primeira vez em 1942. Serão cotejadas a versão publicada na edição crítica da editora Cosac Naify (testemunho A) e a primeira publicação do poema (testemunho B) nos seguintes programas: *Collatex* e *Juxta Commons*.

Palavras-chave

Cotejo, Testemunhos, *Collatex*, *DV-Coll*, *Juxta Commons*.

Editores-chefes

Marcus Dores
Célia Lopes

Editoras convidadas

Maria Clara Paixão
de Sousa
Vanessa Martins
do Monte

Dossiê

“Humanidades digitais”

Recebido: 08/06/2022

Aceito: 12/04/2023

Como citar:

SOUZA, Luana Batista de. Crítica textual e tecnologia. Revista LaborHistórico, v.9, n.1, e52857, 2023. doi: <https://doi.org/10.24206/lh.v9i1e52857>

* Este artigo é resultado de uma apresentação realizada durante o estágio do Programa de Aperfeiçoamento de Ensino da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. A apresentação se deu em 2016 para os alunos de graduação do curso de Letras, na disciplina Filologia Portuguesa.

Abstract

Collating witnesses is a tough task and requires the scholar's attention. Due to the development of Digital Humanities, the scholar can make use of softwares specially developed to collate and critically edit texts from different times to assist his/her task. This paper aims to present some of those online softwares available for the collating and preparing documents that will be used to create the digital scholarly edition, for example *CollateX*, *DV-Coll*, *Juxta Commons* and *FairCopy*. Also, the positive and negative aspects of using such softwares are presented, as well as the challenges imposed by the tools. As a practical exercise, we present a collation of the poem *José*, by Carlos Drummond de Andrade, first published in 1942. Both published versions will be collated, the one published in the scholarly edition by Cosac Naify (witness A) and the first published version of the poem (witness B), for that purpose, *Collatex* and *Juxta Commons* will be used.

Keywords

Collate, Witness, Collatex, DV-Coll, Juxta Commons.

Introdução

A crítica textual feita ainda hoje no século XXI tem como uma de suas bases o método lachmanniano, que data de meados do século XIX. Tal método, dividido em duas partes: recensão e emenda, representa um marco na história da crítica textual e continua ecoando nos estudos da área dois séculos depois.

Como é sabido, a edição crítica é a etapa final da reconstituição de um texto e o caminho percorrido até ela é longo e árduo. Deve o crítico textual estudar e escolher as fontes selecionando os testemunhos de tradição direta e indireta. Vencida esta etapa, dá-se início à colação com a comparação de diversos testemunhos de um texto a fim de localizar os pontos divergentes, os lugares-críticos¹.

A tecnologia computacional, desde a década de 1960, tem contribuído bastante para agilizar a emenda dos textos através de *softwares* de cotejo e edição (CAMBRAIA, 2005: 182-183). Ao contrário do que se pode imaginar, essas ferramentas não substituem o editor, que segue como elemento fundamental na preparação de edições críticas, visto que seu papel é revisar e analisar as variantes, apresentando resultados e reflexões.

O presente trabalho tem como objetivo apresentar quatro programas de cotejo e edição de texto, a saber: *CollateX*, *Dv-Coll*, *Juxta Commons* e *FairCopy*, bem como as vantagens e desvantagens de usá-los como uma ferramenta no trabalho filológico.

¹ Terminologia utilizada por Cambraia (2005, p. 135) para referir-se a “um ponto do texto em que os testemunhos divergem”.

À título de exemplo, apresentaremos a colação do poema “José”, de Carlos Drummond de Andrade, publicado pela primeira vez em 1942. Não se trata de uma proposta de edição crítica para o poema, e menos ainda, uma revisão da edição crítica feita por Júlio Castañón Guimarães, mas sim um exercício prático para mostrar como esses programas funcionam. A escolha de “José” se dá pela sua importância para a Literatura Brasileira e seus estudos, sua ampla difusão em manuais de Língua Portuguesa e presença em provas de vestibular.

Os programas de colação e edição

Há hoje muitas ferramentas *on-line* disponíveis para o crítico textual utilizar em seu trabalho. Em sua maioria, elas são de fácil acesso e dispõem de uma versão gratuita, muitas vezes dispensando a instalação do programa no computador e podendo ser acessadas a partir de um navegador (*Chrome, Firefox, Safari*). A seguir apresentamos quatro delas.

CollateX

Segundo informações disponíveis em seu *site*², *CollateX* começou a ser desenvolvido em 2010 como um projeto em conjunto dentro da iniciativa *Interedition* da União Europeia. Ele foi planejado com o objetivo de ser o sucessor do *Collate*, programa já utilizado para a comparação de textos e produção de aparato crítico. Embora fosse largamente utilizado em edições críticas, com o passar do tempo, *Collate* tornou-se difícil de ser implementado visto que foi desativado pelo seu fornecedor.

Diferentemente do seu predecessor, *CollateX* é um *software* que pode ser incorporado a outro programa ou, ainda, fazer parte de um sistema de *softwares*. Seu objetivo é o fornecimento e o avanço da pesquisa atual no campo da colação computadorizada envolvendo textos de línguas naturais.

Na aba “*Documentation*” (documentação) do *site*, há todas as informações sobre o modelo seguido para a criação do *software*, sua estrutura central de dados (*variant graphs* – grafos variantes), os algoritmos de alinhamento disponíveis para o usuário, os formatos de texto aceitos (*input*), os formatos gerados (*output*), a interface de linha de comandos, webservice no padrão RESTful, programação baseada em JavaScript API e também os recursos e a bibliografia utilizada.

Uma das características do *CollateX* é a leitura de múltiplas versões (duas ou mais) de um texto dividindo cada versão em partes (*tokens*) a serem comparadas. Para isso, ao utilizar o programa é necessário escolher entre os três algoritmos disponíveis: Dekker, Needleman-Wunsch e MEDITE (ver Figura 1). Cada um deles tem suas

² Disponível em: <<http://collatex.net>>. Acesso em: 10/02/2022

particularidades, mas todos operam em estruturas de grafos variantes e pertencem ao grupo de algoritmos de alinhamento progressivo.

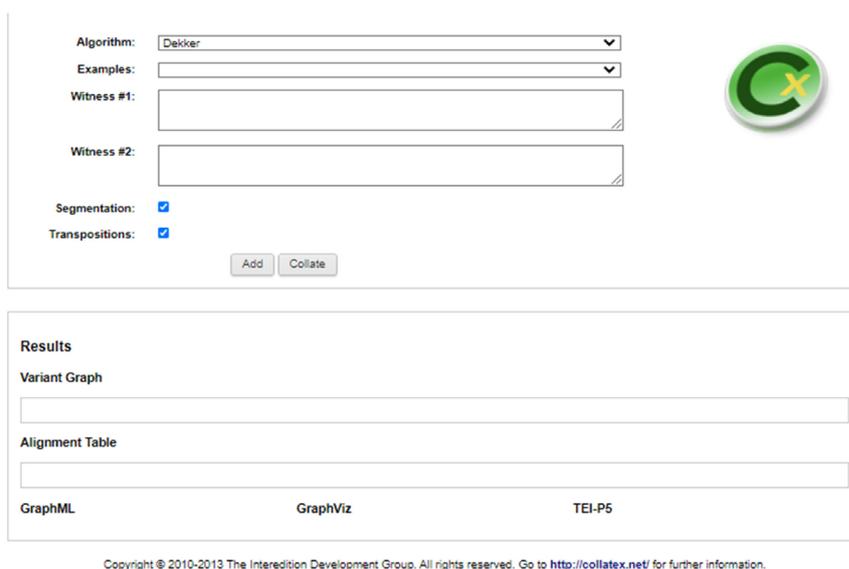


Figura 1. Versão demonstrativa do Collatex.

Fonte: <http://collatex.net>

Desse modo, a comparação entre as versões do texto é feita de maneira progressiva. Primeiro comparam-se duas versões. Desse cotejo, resulta um grafo variante que será comparado com outra versão do texto. Essa nova comparação produz outro grafo variante que é comparado com outra versão do texto até que todas as versões tenham sido mescladas.

CollateX identifica similaridades e diferenças entre edições (incluindo segmentos movidos/transpostos) alinhando *tokens*, e fornece *output* com os resultados em uma variedade de formatos para processamento posterior, tais como *Java Script Object Notation (JSON)*, *TEI P5 (text encoding initiative – linguagem web para a criação de edições digitais)*, *XML*, *GraphML3* e *GraphViz4 DOT*. São aceitos como *input* (os testemunhos a serem inseridos), textos nos seguintes formatos: texto simples (UTF-16), *JSON* e *XML*.

No *site* é possível encontrar uma versão demonstrativa que dispensa *download* e instalação no computador, permitindo seu uso diretamente *on-line*. Nesta versão demonstrativa, é possível visualizar exemplos de cotejo em três línguas diferentes: inglês, alemão e grego. Ao selecionar um dos exemplos, a página é atualizada e os resultados do cotejo são apresentados em diferentes formatos como podemos ver na Figura 2.

³ Formato de arquivo baseado em *XML* para gráficos.

⁴ Software de código aberto, criado pela *AT&T Labs Research* para visualização, gráfica. É utilizado em diversas áreas do conhecimento, tais como Bioinformática, Engenharia da Computação, base de dados, etc.

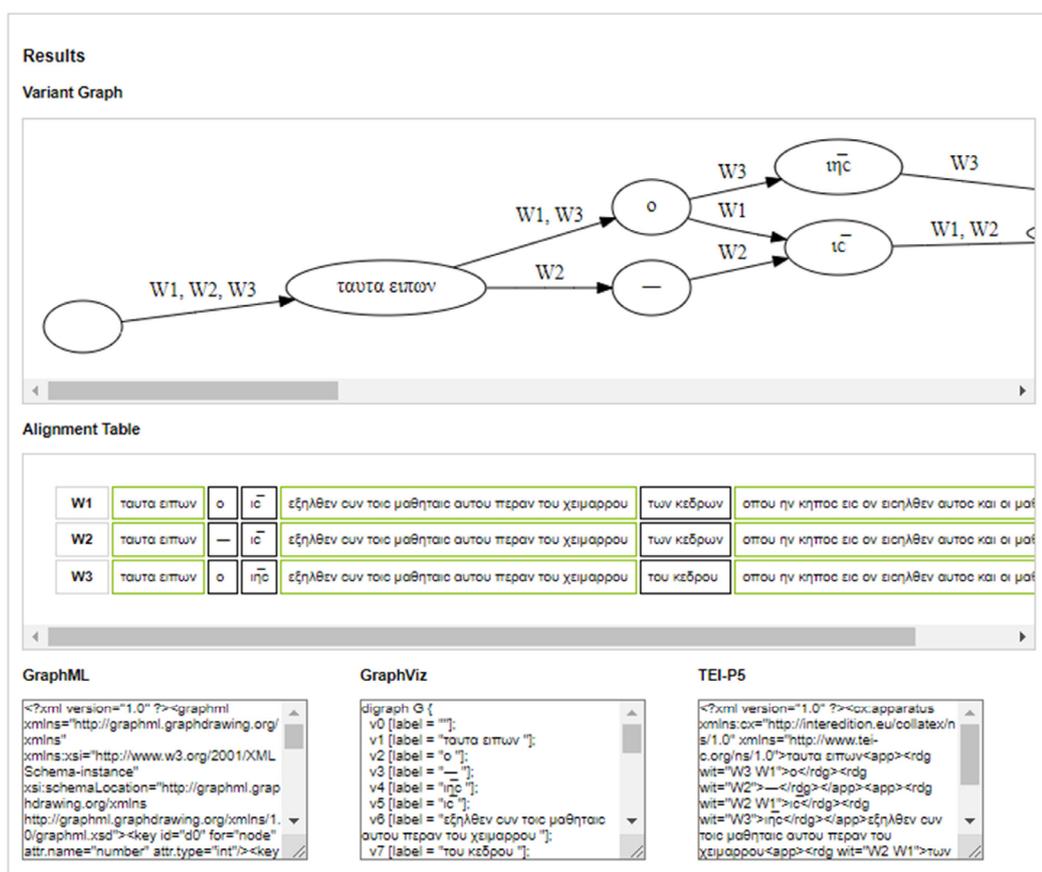


Figura 2. Resultados do cotejo do exemplo em grego

Fonte: <http://collatex.net/>

Na imagem acima, vemos um exemplo de modelo gráfico de variante gerado através do *CollateX*. A letra *W* representa testemunho (*witness*, em inglês) e podemos constatar que foram comparados três. As flechas indicam a sequência das palavras na frase e as variantes são indicadas dentro dos balões. Quando mais de um testemunho apresenta a mesma variante, esses são separados por uma vírgula.

O primeiro passo para usar a versão demonstrativa é selecionar o algoritmo desejado. Em seguida, é necessário digitar no campo correspondente ao testemunho o texto a ser cotejado. No formato padrão, aparecem duas caixas de textos nomeadas como *witness #1* e *witness #2* (testemunho número 1 e testemunho número 2, respectivamente). Para acrescentar mais testemunhos basta clicar no botão “*add*” (adicionar) e uma nova caixa de texto é aberta. Ao clicar no botão “*collate*” (cotejar), todas as versões são comparadas em poucos segundos e os resultados são apresentados (ver Figura 2).

DV-Coll

DV-Coll é a versão mais recente do programa de colação *Donne Variorum* que foi criado nos anos 1980 para a edição dos poemas de John Donne. Trata-se de

uma revisão de *Coll60*, um programa no formato DOS, originalmente escrito em *GW-BASIC* por Bill Vilberg e Bruce Dinoff. A versão Windows foi escrita em *Visual Basic* por Cameron Purvis, modificada por Johan Bjursell e atualizada novamente por Purvis. Funciona em todas as versões do *Windows* de 95 a *Windows* 10.

Criado explicitamente para o trabalho textual no projeto *Donne Variorum*, *DV-COLL*, de acordo com as informações apresentadas no *site*⁵, poderá ser igualmente útil na edição de obras de poetas cujo trabalho existe em múltiplas cópias que precisam ser comparadas e analisadas letra por letra.

Como a etapa da colação requer o preparo dos textos a serem cotejados, *DV-COLL* (ver Figura 3) surge como uma ferramenta útil para o editor, já que providencia um gerador de arquivos no formato *F-GEN*.

Uma das funcionalidades do programa é a capacidade de comparar até 250 transcrições de um poema de uma só vez salvando os resultados da colação em um arquivo de texto simples (*ASCII*) no formato de coluna para facilitar a análise e a manipulação com um processador de palavras. Seu uso requer instalação no computador e todas as informações para *download* e instalação do programa estão disponíveis no *site*.

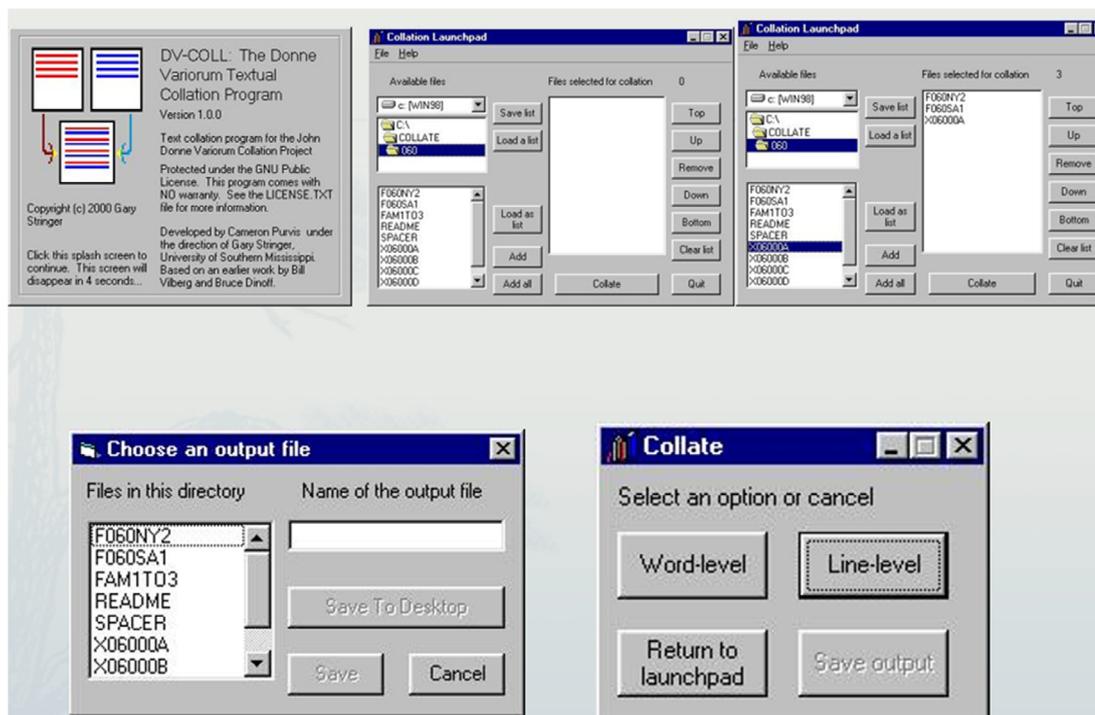


Figura 3. Interface do DV-Coll

Fonte: DV-Coll

⁵ Disponível em: <<http://donnevariorum.dh.tamu.edu/toolsandresources/collation-software/>>. Acesso em: 25/03/2022.

Juxta Commons

O *software Juxta Commons*, criado pela Universidade de Virgínia em 2012 e encerrado em 2020⁶, é uma ferramenta de código aberto para a comparação e confronto de vários testemunhos. Seu objetivo era ajudar os estudiosos e editores a examinar a história de um texto do manuscrito a suas versões impressas.

O acesso era feito on-line através do *site* www.juxtacommons.org. Com uma interface simples e intuitiva (ver Figura 4), era possível comparar e cotejar diferentes versões de um mesmo tempo. Diferente da versão demonstrativa do *CollateX*, que permite apenas como *input* o texto digitado (ou copiado de um editor de texto do tipo Word e colado na caixa de texto do *site*), era possível inserir no *Juxta Commons*, arquivos de texto nos seguintes formatos: *txt*, *xml*, *html*, *doc*, *rtf*, *pdf* e *epub*.

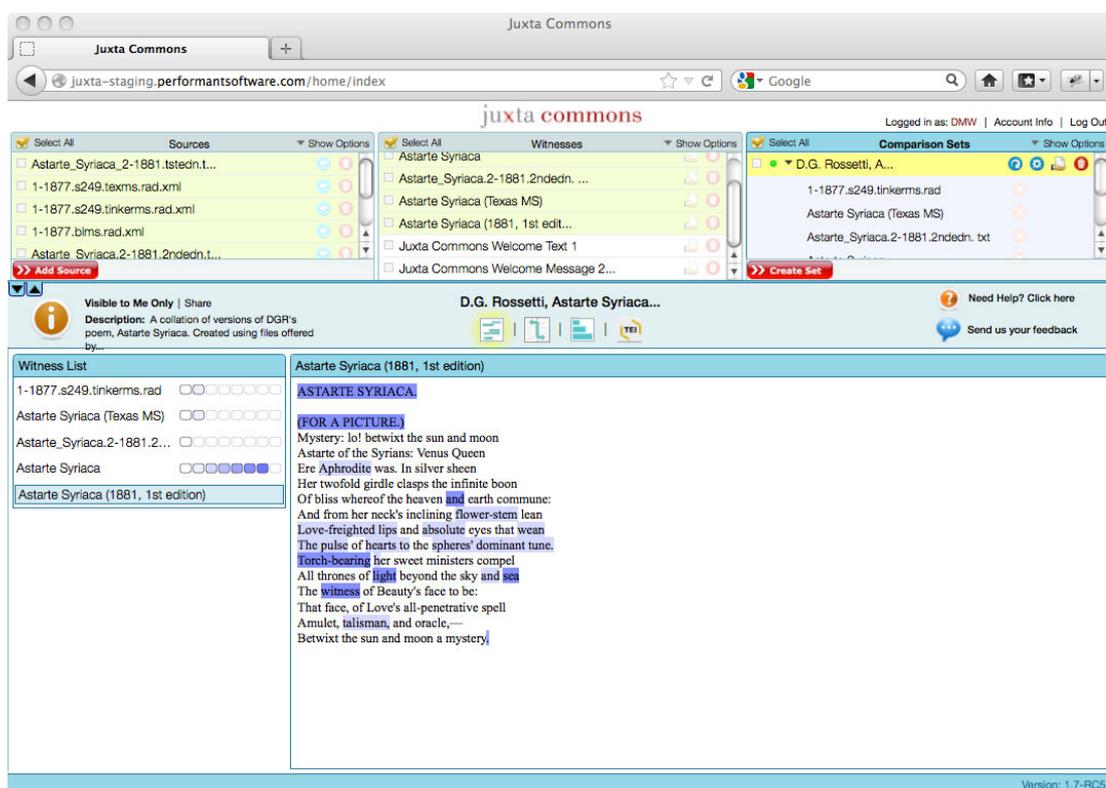


Figura 4. Interface do Juxta Commons

Fonte: Juxta Commons

⁶ *Juxta Commons* foi desenvolvido como um serviço gratuito pela *Performant Software Solutions* e estava disponível gratuitamente *on-line* desde 2014. No dia 25 de agosto de 2020, foi enviado um *e-mail* para os usuários comunicando o encerramento das atividades no dia 8 de setembro do mesmo ano. Na mesma mensagem, foi informado que um novo editor *TEI*, *FairCopy*, mais robusto, estava sendo criado. Para mais informações sobre o *FairCopy*, ler o item 1.4 *FairCopy*.

Outra característica era a criação de uma conta do usuário através da qual era possível acessar remotamente os testemunhos inseridos e os cotejos realizados, não havendo um limite de uso. Além disso, era possível fazer a descrição bibliográfica do testemunho informando o nome pelo qual seria identificado no cotejo, o autor, a data, a editora, uma curta descrição e notas pessoais.

Após a adição dos arquivos à plataforma e listados os testemunhos, era possível gerar o cotejo automático, apresentado em diversos *outputs*, a saber: mapa de calor (texto-base com as variantes realçadas), apresentação justaposta dos testemunhos com indicação da localização das variantes, histograma (distribuição da frequência das variantes), *tei* e *versioning machine* (permite a comparação a partir da edição diplomática dos testemunhos).

Para obter o arquivo com as notas do aparato crítico, era necessário selecionar dentre os testemunhos comparados o texto de base, a sigla utilizada para nomeá-lo e o formato de *output* desejado: *html*, que facilita a criação de códigos de linguagem *web* para a realização de edições críticas digitais, ou *.docx* para ser editado no *Word*.

A comparação do texto era feita a partir do confronto de unidades mínimas e a listagem das variantes no aparato era apresentada a partir da linha em que se encontra (numerada automaticamente pelo programa).

FairCopy

FairCopy foi anunciado, no *email* de encerramento do Juxta Commons, como um programa robusto baseado nas orientações *TEI* e *IIIF*

O *software* está disponível para *download* no *site* <https://faircopyeditor.com/>. O programa oferece dois tipos de licença: individual e em grupo. Essas versões são pagas e antes de adquiri-las é possível se inscrever para instalar e utilizar a versão *trial* (teste) por 14 dias. Ao receber o *e-mail* de confirmação, é fornecida a chave de acesso que permitirá o uso durante o período de teste. Após esse período, para continuar a ferramenta, é necessário comprar a licença.

O programa é descrito como uma ferramenta simples e poderosa para ler, transcrever e codificar o texto⁷. Com ele é possível criar edições críticas digitais através da transcrição, estruturação e anotação das fontes para a publicação *online*. No que tange às imagens, *FairCopy* permite organizar as fontes de imagens e linká-las às transcrições.

Tal qual *CollateX*, há um guia no *site* na aba “*Documentations*” explicando o passo a passo de como usar o programa e criar um projeto. Esse guia é importante porque a interface do *FairCopy* não é intuitiva como a do *Juxta Commons* e sem

⁷ No original: “(...) a simple and powerful tool for reading, transcribing, and encoding text”. (FairCopy, 2021).

as instruções fica difícil utilizá-lo. Um avanço em termos de *input* comparado aos outros *softwares*, é a possibilidade de inserir imagens na edição crítica, tais como os fac-símiles ou figuras que ilustram o texto.

Com a promessa de criar edições críticas digitais, o programa apresenta um refinamento nas marcações disponíveis de *TEI elements* (elementos TEI) que podem ser personalizadas pelo usuário de acordo com o seu objetivo. Com mais autonomia na escolha das marcações, cada edição crítica digital terá a sua própria interface.

E agora, José: um exercício prático

O poema “José”, de Carlos Drummond de Andrade, possui um dos versos mais conhecidos da poesia brasileira: “E agora, José?”. Publicado pela primeira vez no suplemento literário do jornal *A Manhã*, em 11 de janeiro de 1942, o poema tem seu texto estabelecido na edição crítica da poesia de Drummond (1930-1962) publicada em 2012 pela Cosac Naify.

Ao utilizar esse poema como exemplo, não temos a pretensão de rever sua edição crítica preparada por Júlio Castañon Guimarães. O intuito é apenas ilustrativo para que aquele, que não conhece os programas de colação e edição, possa vir a ter o primeiro contato e entenda o seu funcionamento e o seu comportamento com relação às variantes textuais encontradas.

Conforme demonstra Guimarães (2012), nem sempre o famoso verso “E agora, José?” teve essa forma. O poeta oscilou entre escolhas dos nomes Raimundo, Joaquim e José “até chegar na insistência em um único nome, uma única figura, que desse modo encaminha um dos sentidos mais fortes do poema”.

Abaixo apresentamos os dois testemunhos selecionados do poema. O da edição crítica foi chamado de testemunho A e a primeira edição de testemunho B.

TESTEMUNHO A – Cosac Naify (2012)

José

E agora, José?
A festa acabou,
a luz apagou,
o povo sumiu,
a noite esfriou,
e agora, José?
e agora, você?
você que é sem nome,
que zomba dos outros,
você que faz versos,

que ama, protesta?
e agora, José?

Está sem mulher,
está sem discurso,
está sem carinho,
já não pode beber,
já não pode fumar,
cuspir já não pode,
a noite esfriou,
o dia não veio,
o bonde não veio,
o riso não veio,
não veio a utopia
e tudo acabou
e tudo fugiu
e tudo mofou
e agora, José?

E agora, José?
Sua doce palavra,
seu instante de febre,
sua gula e jejum,
sua biblioteca,
sua lavra de ouro,
seu terno de vidro
sua incoerência,
seu ódio - e agora?

Com a chave na mão
quer abrir a porta,
não existe porta;
quer morrer no mar,
mas o mar secou;
quer ir para Minas,
Minas não há mais.
José, e agora?

Se você gritasse,
se você gemesse,
se você tocasse
a valsa vienense,
se você dormisse,

se você cansasse,
se você morresse...
Mas você não morre,
você é duro, José!

Sozinho no escuro
qual bicho do mato,
sem teogonia,
sem parede nua
para se encostar,
sem cavalo preto
que fuja a galope,
você marcha, José!
José, para onde?

TESTEMUNHO B – *A manhã* (1942)

José

E agora, José?
A festa acabou,
a luz apagou,
o povo sumiu,
a noite esfriou,
e agora, Raimundo?
e agora, você?
você que é sem nome,
que zomba dos outros,
você que faz versos,
que ama, protesta?
e agora, José?

Está sem mulher,
está sem carinho,
está sem discurso,
já não pode beber,
já não pode fumar,
cuspir já não pode,
a noite esfriou,
o dia não veio,
o bonde não veio,
o riso não veio,
não veio a utopia

e tudo acabou
e tudo fugiu,
e tudo mofou
e agora, José?

E agora, José?
sua doce palavra,
seu instante de febre,
sua gula e jejum,
sua biblioteca,
sua lavra de ouro,
seu terno de vidro
sua incoerência,
seu ódio - agora?

Com a chave na mão
quer abrir a porta,
não existe porta;
quer morrer no mar,
mas o mar secou;
quer ir para Minas,
Minas não há mais!
José, e agora?

Se você gritasse,
se você gemesse,
se você tocasse
a valsa vienense,
se você dormisse,
se você cansasse,
se você morresse...
Mas você não morre,
você é duro, José!

Sozinho no escuro
qual bicho do mato,
sem teogonia,
sem parede nua
para se encostar
sem cavalo preto
que fuja a galope,
entretanto forte,
você segue, José!
José, para onde?!

Metodologia

Cotejo automatizado

Para a realização do cotejo dos testemunhos A e B, utilizamos os softwares *CollateX* e *Juxta Commons* por termos mais familiaridade com a sua interface. Ambos os textos foram transcritos respeitando sua ortografia e pontuação original e submetidos no formato UTF-8 em arquivos individuais. Uma vez feita a inclusão dos testemunhos e dado o comando para a comparação, o software faz todo o trabalho de colação e ao final gera diferentes *outputs* (Figuras 5 e 6) que podem ser escolhidos pelo crítico textual de acordo com o tipo de edição que ele pretende fazer, como por exemplo uma edição digital a ser publicada *on-line* usando o formato TEI-P5.

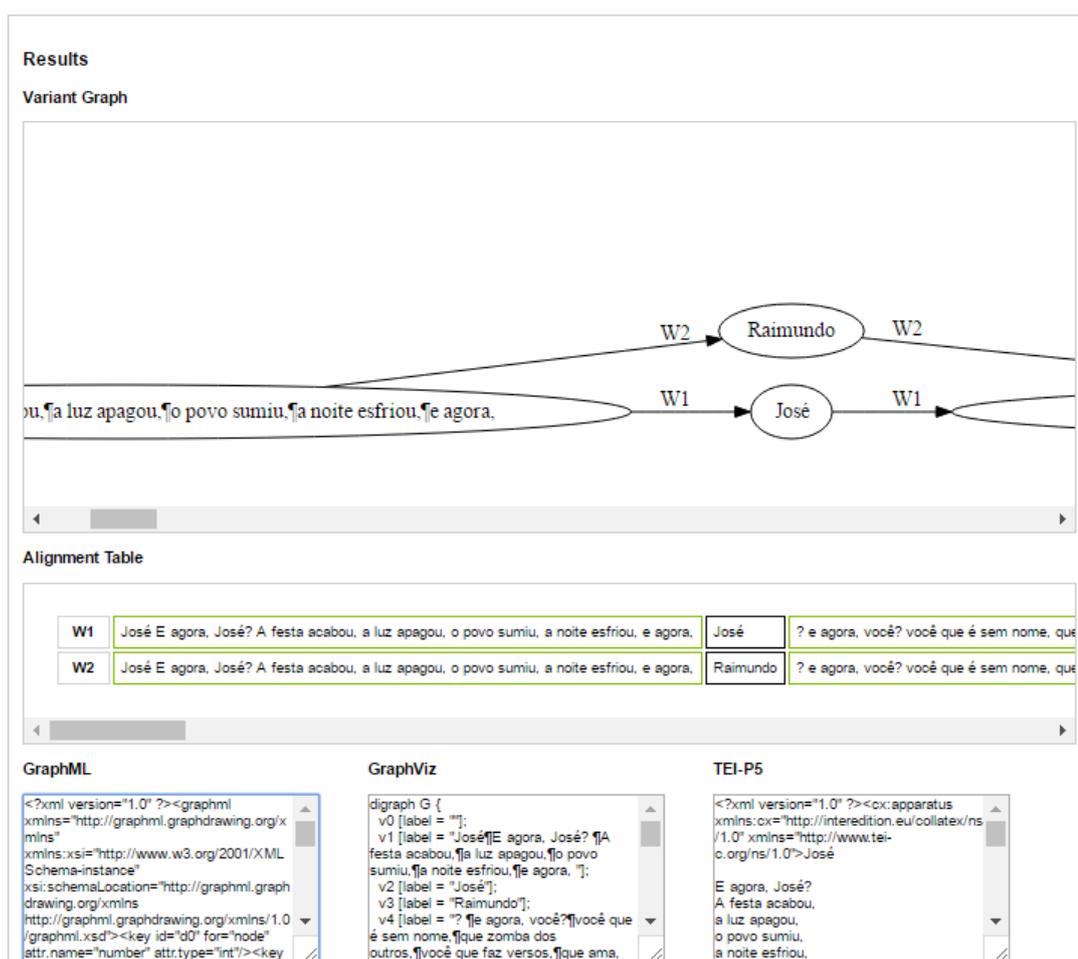


Figura 5. Cotejo CollateX

Fonte: CollateX

Na imagem acima, vemos a página do *CollateX* com os resultados do cotejo. O primeiro é o gráfico de variante, logo abaixo a tabela de alinhamento que mostra,

de forma horizontal, o lugar-crítico nos dois testemunhos comparados. Por fim, os *outputs* em *GraphML*, *GraphViz* e *TEI-P5*.

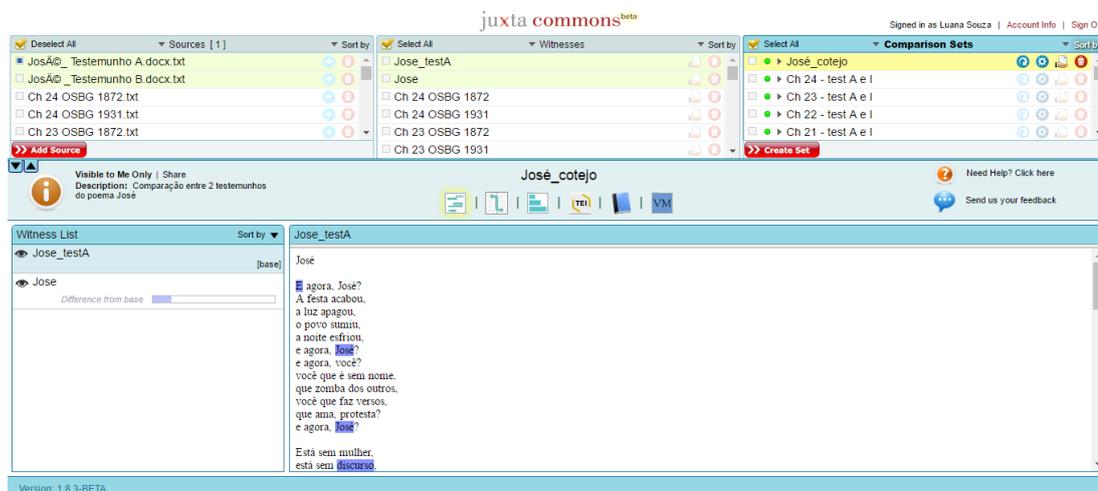


Figura 6. Cotejo realizado no Juxta Commons

Fonte: Juxta Commons

Na Figura 6 vemos a interface do *Juxta Commons* com cotejo já realizado. Na parte superior esquerda da página, estão as fontes (*Sources*), os testemunhos a serem comparados, na tabela ao lado os testemunhos selecionados e com a ficha bibliográfica preenchida (*Witnesses* - testemunhos) e no canto superior direito o arquivo contendo o cotejo realizado (*Comparison Sets* – Conjuntos de comparação).

Conforme mencionado anteriormente, a colação é realizada comparando cada versão do poema salva em arquivos individuais. A partir dessa colação, o software gera o aparato crítico tendo como texto-base o testemunho A e assim o indica logo abaixo de “*Textual Notes*” (notas textuais) com a frase “*This is the best text*” (esse é o melhor texto).

O aparato gerado é uma listagem das variantes que são indicadas a partir da numeração das linhas, facilitando assim a sua localização. Neste aparato gerado automaticamente, primeiro é indicado o número da linha em que está localizada a variante, na linha seguinte a variante seguida por um colchete direito e a sigla do testemunho. As variantes dos dois testemunhos são separadas por um ponto e vírgula como se pode ver no exemplo abaixo:

7
 José] A; Raimundo B
 13
 José] A; Jose B

É importante frisar que tanto as variantes ditas substantivas como as acidentais são contempladas⁸.

Nas próximas linhas vemos o aparato crítico tal qual ele foi gerado pelo *Juxta Commons* (inclusive os elementos em negrito):

Textual Notes

A: Jose_testA - This is the base text.

B: Jose

7

José] **A**; Raimundo **B**

13

José] **A**; Jose **B**

15

discurso] **A**; carinho **B**

16

carinho] **A**; discurso **B**

26-27

fugiu] **A**; tudo fugiu, / e **B**

30

Sua] **A**; sua **B**

37

e] **A**; *not in B*

44

.] **A**; ! **B**

59

,] **A**; *not in B*

61-62

galope,] **A**; galope, / entretanto forte, / você **B**

62

marcha] **A**; segue **B**

63

?] **A**; ?! **B**

As vantagens

A modernização do trabalho de edição de texto através de ferramentas automatizadas é muito importante para a sua dinamização e celeridade, permitindo

⁸ Trata-se de uma terminologia anglo-saxônica adotada por Ivo Castro (1990, p. 52). As variantes substantivas referem-se à estrutura linguística e semântica do texto e as variantes acidentais estão relacionadas à sua forma gráfica e ortográfica. Blecua (1983), apresenta uma categorização das variantes substantivas, denominadas por ele como “erros”. Ele as separa em quatro: adição, omissão, alteração de ordem e substituição, que posteriormente são divididas em subcategorias.

ao estudioso ter ao seu alcance um número elevado de testemunhos cotejados ao mesmo tempo.

Dentre as vantagens apresentadas pelos *softwares* de colação e edição, está sem dúvida, em primeiro lugar, a agilidade no desempenho da tarefa de cotejar os testemunhos, visto que um cotejo óptico, letra por letra, palavra por palavra, exige grande atenção e cuidado. Para além da análise das variantes, que são as diferenças mais evidentes num primeiro momento, os *softwares* também analisam as alterações de ortografia e pontuação.

Em comparação ao método manual e tradicional, focado sobretudo na edição crítica impressa, ter à disposição *outputs* em diferentes formatos permite ao editor que seu trabalho tenha maior alcance porque uma edição crítica digital pode ser publicada em um *site* e assim, já não há mais a dependência de uma editora para publicá-la. Isso confere mais liberdade ao trabalho, que deixa de ser submetido às imposições próprias do mercado editorial, que muitas vezes, ao visar economias nos custos de produção, exige edições mais simplificadas.

O editor ou crítico textual interessado em publicar sua edição crítica *online* pode facilmente fazê-lo a partir do *output* em *TEI* fornecido pelo *CollateX* e *FairCopy*. Esse, por ser mais atual que aquele, e por dispor de uma versão paga, oferece mais funcionalidades e permite a personalização das marcações utilizadas.

Os *outputs* fornecidos pelo *CollateX* são: gráficos de variantes (que permitem analisar as alterações de maneira mais visual através das estruturas em balões), tabelas de alinhamento, *GraphML*, *GraphViz* e *TEI- P5*.

O *software Juxta Commons*, por sua vez, gerava seu *output* em diversas interfaces: mapa de calor (que mostra em destaque no texto base quais são os lugares críticos), comparação justaposta dos testemunhos (ao passar o cursor é possível ver destacado no texto as alterações), histograma (distribuição da frequência das variantes), *tei* (*text encoding initiative*) e *versioning machine* (permite a comparação a partir da edição diplomática dos testemunhos). Além disso, ao cotejar os textos, o software gera o aparato crítico automaticamente, indicando em quais linhas do texto estão os lugares críticos.

As desvantagens

Embora esses *softwares* sejam de grande ajuda, eles também têm suas desvantagens e impõem desafios.

A primeira desvantagem e desafio é a língua em que eles estão disponíveis, já que todos esses programas foram desenvolvidos por universidades ou empresas localizadas no exterior. Para utilizá-los é necessário saber inglês (mesmo que a nível instrumental) e ter um certo domínio dos termos técnicos da Crítica Textual e da informática nessa língua.

A segunda desvantagem é a familiaridade com a linguagem da programação. Ainda que os próprios programas gerem o *output* em *XML*, que permite a criação de um *site* com a edição crítica digital, sem o conhecimento mínimo de *TEI* será difícil fazer modificações e personalizações na *template* (aspecto visual do *site*).

A terceira desvantagem está relacionada ao cotejo. A colação automatizada não é personalizável e reconhece apenas quatro categorias de variantes, a saber: adição, alteração de ordem, omissão e substituição. Os casos de reelaboração e paragrafação não são reconhecidos e é necessário cotejá-los manualmente.

Além disso, os textos a serem cotejados devem ser transcritos e, a depender da quantidade de testemunhos, a transcrição pode ser demorada, principalmente se não forem usados conversores de imagem para texto como o OCR⁹. Os *softwares* de colação e edição ainda não são capazes de reconhecer os textos contidos nos fac-símiles dos testemunhos.

Os arquivos das transcrições devem ser salvos no formato UTF-8 (*Juxta Commons* e *FairCopy*) ou UTF-16 (*CollateX*) para não desconfigurar caracteres gráficos como o til ou a cedilha presentes na língua portuguesa. Como o *Juxta Commons* também aceitava arquivos no formato *docx.*, ao cotejar os textos nesse formato, o til e a cedilha não eram reconhecidos e o texto ficava desconfigurado, causando muitos problemas na edição porque exigia a revisão ortográfica manual. Sendo elaborado para textos em língua inglesa, que não contém esses caracteres, não havia nenhuma menção aos problemas causados em arquivos em outros formatos diferentes de UTF-8 e na colação de textos em línguas estrangeiras.

Com relação à estrutura textual, nem todos os programas são capazes de editar poema e prosa. O *DV-COLL* é um exemplo, servindo apenas para a colação de poemas. Isso restringe um pouco as opções de *softwares* que podem ser utilizados para o cotejo e edição de textos em prosa.

Se por um lado as versões demonstrativas ou de teste permitem o livre acesso ao *software* numa versão simples ou por tempo determinado, por outro elas são um indicativo que há versões pagas. A compra da licença de uso garante o financiamento e aprimoramento do programa, condições imprescindíveis para evitar a obsolescência intrínseca ao campo da tecnologia e da informática. Por requererem a atualização constante de versões e correções de erros, esses *softwares* podem tornar-se obsoletos e ultrapassados com o passar dos anos devido à falta de financiamento. Foi o que aconteceu com o *Juxta Commons*, que após oito anos foi retirado do ar devido à degradação do servidor e questões financeiras gerando um impacto direto nos pesquisadores e estudiosos que dele dependiam para a realização de seus projetos.

⁹ OCR é a sigla de *Optical Character Recognition* (Reconhecimento de Caracteres Óticos).

Por fim, uma limitação imposta pelo *FairCopy* é a não-finalização da ferramenta. Por ser um programa muito recente que ainda está sendo finalizado, *FairCopy* não faz o cotejo dos testemunhos, sendo necessário fazer o cotejo óptico ou utilizar outro programa para a colação. Desse modo, ele apresenta lacunas que podem demorar a ser preenchidas e não dá conta da etapa mais importante que precede à edição crítica. Segundo o designer e desenvolvedor do *FairCopy*, Nicholas Laiacona (2022), esta funcionalidade será incluída em breve.

Considerações finais

Como mostramos nesse artigo, a utilização dos programas de colação e edição apresenta vantagens, desvantagens e muitos desafios.

Uma vantagem incontestável é a otimização do trabalho do filólogo, uma vez que é possível cotejar diversos testemunhos ao mesmo tempo em questão de segundos. A partir do cotejo, são gerados formatos variados de arquivo que atendem à diversas necessidades, dentre elas a edição de texto em formato eletrônico a partir do formato *TEI* ou *XML* e o processamento automático do aparato crítico.

Se a utilização da tecnologia computacional é um ganho para o labor filológico devido a modernização dos processos, ela também é seu calcanhar de Aquiles uma vez que, ao apoiar-se totalmente nos programas como única ferramenta para a colação e reconstituição do texto, o editor crítico torna-se refém das atualizações dos softwares e renovações de licença. Embora atuem como um facilitador eles não substituem o trabalho do filólogo cujo papel é fundamental para a filtragem e análise dos dados.

Além disso, ela está sujeita à obsolescência própria do seu campo. Uma mostra é a desativação do programa *Juxta Commons*, oito anos após a sua disponibilização *on-line*.

Desse modo, um dos grandes desafios é aperfeiçoar cada vez mais os *softwares* e mantê-los atualizados, visando abarcar diferentes estruturas textuais, como poema e prosa, cotejar os testemunhos, fornecer ferramentas para o aparato crítico e para a elaboração de edições críticas digitais.

Referências

ANDRADE, Carlos Drummond. *José In: Autores e Livros*, suplemento literário de *A Manhã*, Rio de Janeiro:11/1/1942.

ANDRADE, Carlos Drummond. *Poesia 1930-62. Edição crítica*. São Paulo: Cosac Naify, 2012.

BLECUA, Alberto. *Manual de Crítica Textual*. Madrid: Castalia, 1983.

CAMBRAIA, César Nardelli. *Introdução à crítica textual*. São Paulo: Martins Fontes, 2005.

CASTRO, Ivo. *Editar Pessoa*. Lisboa: Imprensa Nacional Casa da Moeda, 1990.

GUIMARÃES, Júlio Castañon. Os materiais do poema: isto às vezes é aquilo em CDA. In: ANDRADE, Carlos Drummond. *Carlos Drummond de Andrade – Poesia 1930-62: edição crítica*. São Paulo, Cosac Naify, 2012.

LAIACONA, Nicholas. *Juxta Commons Project [mensagem pessoal]*. Mensagem recebida por <luana.souza.usp@gmail.com> em 05 jun. 2022. Acesso em: 05/06/2022.

Softwares de edição

COLLATEX. Disponível em: <<http://collatex.net/>>. Acesso em 18/01/2022.

DV-COLL. Disponível em: <<http://donnevariorum.tamu.edu/>> Acesso em:10/08/2021.

FAIRCOPY. Disponível em: <<https://faircopyeditor.com/>> Acesso em: 28/05/2022.

VERSIONING MACHINE. Disponível em: <<http://v-machine.org/>> Acesso em: 27/04/2018.