

Edição de documentos e constituição de corpora linguísticos no acervo de Eulálio Motta¹

Edition of documents and linguistics corpora constitution in the holdings of Eulálio Motta

Iago Gusmão Santiago

Universidade Estadual de Feira de Santana (UEFS)

Stephanne da Cruz Santiago

Universidade Estadual de Feira de Santana (UEFS)

Liliane Lemos Santana Barreiros

Universidade Estadual de Feira de Santana (UEFS)

DOI: <https://doi.org/10.5902/2176148542321>

Resumo: Eulálio Motta (1907-1988) arquivou, durante a sua vida, uma grande quantidade de documentos pessoais. Essa documentação é bastante diversificada e demonstra um grande potencial para estudos linguísticos. No projeto *Edição das obras inéditas de Eulálio Motta*, busca-se editar todos esses documentos para preservá-los e transformá-los em dados confiáveis para pesquisa. Assim, recorre-se à Filologia para realizar a transcrição desse material, bem como à Linguística de corpus para a análise automática dos dados. Nesse artigo, apresentam-se os procedimentos utilizados no acervo do escritor Eulálio Motta, tendo em vista o preparo da documentação para o estudo linguístico.

Palavras-chave: Filologia. Linguística de Corpus. Humanidades Digitais. Eulálio Motta.

Abstract: Eulálio Motta (1907-1988) saved, during his life, a great number of personal documents. This documentation is very diverse and shows a great potential for linguistics studies. In the project *Edição das obras inéditas de Eulálio Motta*, it is aimed to edit all these documents in order to preserve and transform them into reliable data for researches. Therefore, textual criticism is used for making this material's transcription, as well as corpus linguistics, that is used for the automatic analysis of data. In this article, are presented the procedures used in the holdings of the writer Eulálio Motta, aiming to prepare the documentation for linguistic studies.

Keywords: Textual criticism. Corpus linguistics. Digital Humanities. Eulálio Motta.

¹ O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

1 Introdução

O escritor mundonovense Eulálio Motta (1907-1988) guardou em seu acervo pessoal uma vasta documentação que contém diários, postais, borradores de cartas, coleções de jornais etc., além de prototextos de obras literárias editadas e inéditas. Essa documentação serve como lugar de memória, preservando não só a história daquele que se arquivou, mas também das práticas da sociedade em que estava inserido, do lugar onde viveu, dos grupos sociais com os quais se envolveu e como fonte para o estudo da história da língua.

As potencialidades que apresentam os documentos do acervo para os estudos linguísticos são inúmeras. A documentação que o constitui é bem diversificada e abrange uma grande variedade de gêneros: cartas, poemas, crônicas, trovas, discursos, resenhas, rascunhos dos mais variados tipos e anotações do cotidiano. Essa documentação também reflete usos linguísticos característicos de alguns grupos sociais aos quais Eulálio Motta pertenceu como a comunidade católica, comunista, integralista, farmacêutica, além dos falares populares representados nos cordéis e nos causos. É importante destacar o fator temporal, pois se trata de uma documentação produzida em vários momentos, dentro de um período de aproximadamente sessenta anos. Tais aspectos presentes no acervo do escritor possibilitam a realização de diversas pesquisas no âmbito da Linguística, nos diversos níveis de estruturação e usos da língua.

Para a realização dessas pesquisas é importante a constituição de *corpora* confiáveis, o que se dá por meio de uma série de procedimentos que vão desde a digitalização dos documentos à adaptação das transcrições filológicas aos *softwares* de análise linguística. Nesse percurso, a Filologia e a Linguística necessitam trabalhar juntas, estabelecendo um diálogo contínuo em todas as etapas. Ao levar em consideração essas questões, o presente artigo discute as relações existentes entre Filologia e Linguística e o progresso alcançado por elas em consequência do advento das Humanidades Digitais. Em seguida, são detalhadas as etapas para a constituição de *corpora* nos projetos *Edição das Obras Inéditas de Eulálio Motta* (UEFS/CONSEPE, Resolução n. 070/2016) e *Estudos Lexicais no Acervo de Eulálio Motta* (UEFS/CONSEPE, Resolução n. 137/2017), com um destaque especial à questão dos rascunhos do caderno *Farmácia São José* e a coletânea dos textos jornalísticos de e sobre o escritor Eulálio Motta que integram o acervo, publicados nos periódicos *Mundo Novo*, *O Lidador*, *Gazeta do Povo*, *O Serrinhense* e *Vanguarda*.

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

102

2 Diálogos entre Filologia e Linguística

Apresentar uma definição sucinta e geral de Filologia é algo difícil de fazer de forma satisfatória. Para Cambraia (2005, p. 18, grifo do autor), a Filologia se encarrega do “*estudo global de um texto*, ou seja, a exploração exaustiva e conjunta dos mais variados aspectos de um texto: linguístico, literário, crítico-textual, sócio-histórico, etc.”. Nessa acepção ampla, a Filologia vem sendo situada, a partir da análise do seu percurso histórico, como uma área de investigação que compreende os estudos históricos, linguísticos, culturais e literários, convergidos com a intenção de possibilitar um estudo completo do objeto texto. Entretanto, com a fragmentação no interior das Humanidades e a consolidação dessas áreas como ciências autônomas, há um refinamento e um aumento da complexidade teórica no interior de cada uma delas. Os filólogos se viram impossibilitados de acompanhar os avanços internos de cada área e a totalidade das novas perspectivas de abordagem do texto, sendo possível apenas desenvolver expertise em alguns destes domínios e tangenciar outros. Assim, diante das limitações impostas, a ambição de esgotar as possibilidades de estudo que um determinado documento pudesse oferecer foi abandonada.

O uso dos termos ‘global’ e ‘exaustivo’ também reflete essa confusão epistemológica existente no campo da Filologia. Já que muitas atividades foram cunhadas como filológicas ao longo dos séculos, tornou-se uma prática recorrente a afirmação da sua especialidade, a Crítica Textual, que se ocupa especificamente da atividade de edição de documentos. Ter-se-ia, portanto, uma Filologia *lato sensu*, uma área mais abrangente que mobiliza os conhecimentos das mais diversas ciências para o estudo dos documentos, e a Filologia *stricto sensu* ou Crítica Textual, que parte do uso desses conhecimentos para a determinação dos métodos a serem adotados nas edições a partir da análise das especificidades de cada documento. No Brasil, atualmente, não tem sido comum a realização de investigações filológicas dissociadas da práxis editorial, sendo, por conseguinte, a relação entre Filologia e Crítica Textual, no contexto brasileiro, uma dicotomia que reflete a constituição teórica e prática de uma mesma ciência, o que também justifica o recorrente uso dos termos como sinônimos na bibliografia especializada.

Outra questão a ser discutida diz respeito à natureza do objeto da Filologia, já que a noção de texto que se desenvolveu na tradição filológica corresponde a de construto linguístico, constituído e materializado

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

104

em uma sequência de caracteres alfanuméricos (BARREIROS, 2015). Tal noção de texto não consegue dar conta da complexidade do objeto de interesse filológico, o qual abrange uma grande variedade de elementos que vão além da sua dimensão linguística. Ressaltar outros aspectos do objeto da Filologia que transcendem essa primeira dimensão é uma atitude necessária para evitar reducionismos. Moreira (2011), ao definir documento e texto, argumenta em favor da indissociabilidade destas duas realidades que constituem o objeto de estudo filológico:

[...] 'documento', em nosso trabalho, significa 'artefato bibliográfico' – constituído não apenas dos materiais, tinta, papel *et cetera* – cujas partes componentes e integradoras – as marcas grafemáticas seriam um dos elementos da integridade documental – não podem ser dissociadas sob pena de mutilação de sua natureza intrínseca e que só pode ser compreendido quando contextualizado historicamente; portanto, não se atribui ao supramencionado vocábulo uma existência autônoma que se contraporá à do 'texto' – conjunto de signos (palavras e pausas) gravados sobre um suporte (MOREIRA, 2011, p. 91).

A noção de documento apresentada pelo autor contempla, assim, duas dimensões: a textual, a língua gravada no documento, e a material, os aspectos físicos que o constituem, inclusive a materialidade da sua parcela alfanumérica. A distinção feita por Moreira (2011) ajuda a especificar melhor o objeto do estudo filológico, o qual consiste não apenas no código linguístico, o texto, mas na materialidade que envolve esse código, também responsável pela construção de sentidos, ou seja, o documento. A análise desses códigos, junto à exploração de informações contextuais em outras fontes, os chamados códigos contextuais, revelam outros aspectos envolvidos nos processos de produção, circulação e apropriação dos documentos, constituindo a abordagem triádica que a Filologia tem adotado contemporaneamente.

A partir destas considerações, pode-se pensar a Filologia, em linhas gerais, como a ciência que se ocupa do estudo de documentos, considerando-os a partir de uma perspectiva triádica, que explore os códigos alfanuméricos, bibliográficos e contextuais, incluindo os processos de produção, circulação e apropriação, ou seja, sua sócio-história. Cumprida a etapa analítica, inicia-se o trabalho editorial, em que os

conhecimentos são empregados em prol da atividade de transposição textual, ou seja, um dado texto verbal é removido do seu suporte e contexto original e inserido em uma nova materialidade, num contexto social, cultural e historicamente distinto. A atividade editorial é, por excelência, o núcleo central da Filologia. No entanto, cabe ressaltar que tal aspecto não a põe em uma posição de ciência auxiliar, pelo contrário, assegura a sua existência e reafirma a sua relevância para o desenvolvimento das ciências humanas.

A Linguística, por outro lado, assume o seu lugar como ciência, no início do século XX, a partir da publicação do *Curso de Linguística Geral*, de Ferdinand de Saussure. Inicialmente, após a distinção apresentada por Saussure (2006 [1916]), o seu objeto de estudo acaba por se restringir ao sistema linguístico, a *langue*, ficando de lado os estudos da fala, a *parole*. Posteriormente, a Linguística passa por uma grande ampliação, durante o século XX, e inclui no escopo dos estudos linguísticos as relações entre língua, cultura e sociedade, bem como a pesquisa histórica, abandonando o enfoque exclusivamente sincrônico e intrasistêmico do seu estágio inicial. Assim, Filologia e Linguística passam a configurar-se como distintas áreas do conhecimento com objetivos e métodos próprios, a Linguística volta-se para questões concernentes à natureza da língua e a Filologia para a preservação, estudo e edição dos documentos escritos.

Apesar disso, a Filologia nunca chegou a dissociar-se por completo dos estudos linguísticos, pois “[...] não é possível o estabelecimento de um texto, nem mesmo a leitura de um manuscrito, se o pesquisador não possuir fortes conhecimentos da língua com que está trabalhando” (TELLES, 2005, p. 90). Esse conhecimento não consiste apenas no domínio do sistema linguístico em si, mas da sua formação histórica, de aspectos variacionais e das teorias linguísticas que servem de base para compreender a língua presente no documento. A Linguística, por outro lado, depende dos estudos filológicos para garantir a procedência e a correta interpretação das fontes de pesquisa utilizadas, considerando seus aspectos materiais e sócio-históricos.

A Filologia passou por um período de menor produtividade no início do século XX em comparação com o grande desenvolvimento que teve durante o século XIX, impulsionado principalmente pela Linguística Histórica oitocentista. Com a ênfase dada às investigações sincrônicas na modernidade e à preocupação teórica em torno de questões

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

106

predominantemente intralinguísticas, os estudos da língua em uma perspectiva histórica terminaram por se distanciar das discussões centrais da Linguística, apesar de ainda serem realizados na forma de Linguística Diacrônica (PAIXÃO DE SOUZA, 2006). O desinteresse dos linguistas pelos documentos de natureza literária também foi um fator que contribuiu para esse distanciamento. No Brasil, segundo Mattos e Silva (2008), a Linguística Moderna, aqui efetivamente estabelecida em 1960, fez com que a Filologia recuasse por um tempo. Vinte anos depois, o problema da mudança voltou a impulsionar as pesquisas em Linguística Histórica no país e a Filologia retomou o seu lugar de relevância na Linguística brasileira, tendo em vista que “não se pode fazer linguística histórica ou diacrônica sem a documentação remanescente do passado” (MATTOS e SILVA, 2008, p. 14).

Assim, a busca pelos conhecimentos da Filologia sobre os métodos de edição e interpretação dos documentos têm se tornado mais frequentes, visto que, além de possibilitarem o acesso a fontes confiáveis, as pesquisas filológicas fornecem elementos para a contextualização dos dados nelas contidos e a avaliação da sua relevância para o estudo da língua que se pretende realizar.

3 O labor filológico nas Humanidades Digitais

Nos últimos anos, com o advento e a democratização da informática, um novo campo do conhecimento surge para transformar o paradigma científico existente nas ciências humanas: as Humanidades Digitais. Seu principal objetivo é integrar a tradição científica dessas ciências à aplicação dos recursos informáticos. Suas origens remetem às pesquisas realizadas no ano de 1946, nos Estados Unidos, pelo padre italiano Roberto Busa. Na ocasião, ele iniciou a construção de um índice eletrônico, *index verborum*, utilizando o computador, que visava lematizar 11 milhões de itens lexicais do latim medieval, a partir da obra de Tomás de Aquino e autores relacionados. A pesquisa monumental é considerada pioneira na aplicação da informática aos estudos na área das Humanidades (HOCKEY, 2004).

As Humanidades Digitais, pautadas no conhecimento já instituído no âmbito das ciências humanas e sociais, das artes e das letras, buscam o aprimoramento das suas técnicas através da inserção das novas tecnologias proporcionadas pelo ambiente digital na sua prática de pesquisa com o intuito de impulsionar o seu desenvolvimento. Na Filologia,

dentre as novas práticas adotadas estão a de utilizar fac-símiles para a realização de edições no computador, possibilitando a preservação do material original e a manipulação das imagens para facilitar a leitura (por meio do aumento do *zoom*, do ajuste de brilho, contraste e cor), e *softwares* que auxiliam na transcrição e cotejamento dos textos. Esses novos recursos incorporados à prática editorial otimizam o labor filológico, tornando os procedimentos mais eficazes e precisos.

Nesse contexto, também surge a Linguística Computacional, definida por Othero (2006) como a área que investiga o tratamento computacional dado à linguagem e às línguas naturais. Segundo o autor, a Linguística Computacional se divide em duas subáreas: o Processamento da Linguagem Natural (PLN) e a Linguística de Corpus. A área de PLN se utiliza dos conhecimentos linguísticos para a construção de *softwares* capazes de processar a linguagem, tais como tradutores e corretores automáticos, reconhecedores de voz, decodificadores de fala etc., ou seja, de ferramentas capazes não só de gerar informações em linguagem natural, mas também de interpretá-las, possibilitando, em alguns casos, verdadeiros atos comunicativos entre homem e computador.

A Linguística de Corpus é o campo da Linguística que se preocupa com a constituição de bancos de dados para o estudo da língua. Segundo Othero (2006, p. 342), a “Linguística de Corpus preocupa-se basicamente com o trabalho a partir de corpora eletrônicos que contenham amostras de linguagem natural”. Com os avanços da Linguística de Corpus, atualmente é possível criar bancos de dados constituídos por amostras da língua em sua modalidade falada e escrita. Os novos recursos computacionais permitem a exploração eficiente desse *corpus* de modo a dinamizar processos que demorariam muito se fossem realizados manualmente, por meio de buscas inteligentes, análise da frequência de um determinado fenômeno e classificações automáticas, por exemplo. O fugaz avanço da Linguística de Corpus tem auxiliado a Linguística em seus mais diversos âmbitos. Barreiros (2017a), ao tratar do impacto das novas tecnologias no âmbito da lexicografia, ressalta que:

Os benefícios alcançados certamente são inúmeros. Além da velocidade na execução das atividades e da ampla capacidade de armazenamento de dados, as novas tecnologias permitem ao lexicógrafo coletar, selecionar, registrar, analisar, aperfeiçoar, recuperar os dados e gerar documentos publicáveis com baixo

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

108

custo. Ressalta-se que essas vantagens não se limitam apenas na execução das etapas de elaboração de grandes dicionários, mas favorecem ao progresso das pesquisas linguísticas de diversas áreas, atrelando produtividade, com qualidade e acessibilidade (BARREIROS, 2017a, p. 218).

Percebe-se que as contribuições dadas pelas Humanidades Digitais para a Linguística e para a Filologia, principalmente com relação ao uso de *corpus* para estudo da língua, tem sido o mais importante elo que une as duas disciplinas na atualidade. Essa tarefa interdisciplinar exige preparo do pesquisador para lidar com a documentação utilizada, que necessita ser preservada, e com o processo de adaptação das transcrições para as ferramentas computacionais, visando superar alguns problemas de incompatibilidade. Para tanto, se faz necessário executar procedimentos básicos que devem ser considerados no momento de elaboração de um *corpus* constituído por documentos escritos, a saber: digitalização, transcrição, plano de revisão e adaptação. Na sequência, exemplificam-se esses procedimentos em documentos que integram o projeto *Edição das obras inéditas de Eulálio Motta*.

3.1 Digitalização

Ao deparar-se com um documento de uma determinada época, o filólogo tem o dever de preservá-lo, não apenas em sua dimensão linguística, mas também em sua dimensão material. Assim, a primeira medida a ser tomada é a observação do seu estado de conservação para avaliar se há a necessidade de restaurá-lo ou não. Em seguida, deve-se iniciar o processo de digitalização do documento, para que não seja mais necessário o manuseio constante do original, o que acentua a sua deterioração. Contudo, nas diretrizes preliminares apresentadas em 2002 para o fornecimento do selo de *Approved Edition* atribuído a edições digitais, o *Committee on Scholarly Editions* (CSE), da *Modern Language Association of America* (MLA), na seção *Editorial methods and procedures*, já ressaltava a obrigatoriedade do filólogo, em algum momento, verificar a exatidão de sua transcrição, comparando-a com os artefatos originais.

A digitalização do material, além de contribuir para a conservação do original, traz inúmeras vantagens para o editor e o leitor do texto. Esse procedimento possibilita uma apresentação mais fiel do documento, respeitando aspectos de sua materialidade que não

podem ser transmitidos apenas com a descrição e se perdem com a transcrição. Também permite que o próprio leitor possa apreciar suas características estéticas e o pesquisador revisar o dado a ser utilizado. No que diz respeito ao editor, facilita o exercício da transcrição, já que, por meio da manipulação da imagem, é possível resolver impasses que dificilmente se conseguiria apenas com a consulta aos originais.

É importante levar em consideração que o processo de digitalização também é prejudicial ao documento, principalmente, por conta do desgaste ocasionado pelo contato com a luz intensa dos escâneres, sendo recomendado que seja feito apenas uma vez ou que sejam usadas câmeras de alta resolução com o flash desativado. Para isso, é necessário contar com o equipamento adequado para garantir fac-símiles com boas resoluções. No caso dos manuscritos, foi possível efetuar uma digitalização de boa qualidade utilizando um escâner convencional para papéis no tamanho A4. A alta resolução da imagem permitiu a visualização ampliada dos códigos alfanuméricos, tornando possível a transcrição de segmentos de difícil decodificação, como os acréscimos nas entrelinhas e as rasuras (Cf. Figura 1).

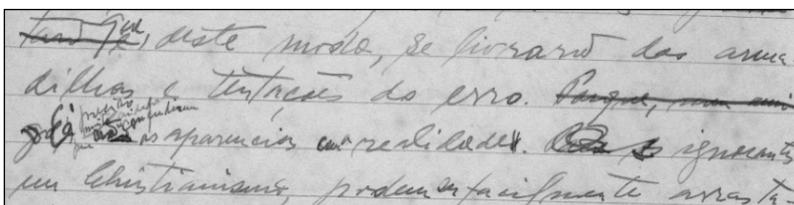


Figura 1 - Rasuras e acréscimos no fólio 27r do caderno *Farmácia São José*

Fonte: Acervo de Eulálio Motta.

Com relação aos periódicos, além da possibilidade de utilizar uma câmera de alta resolução para fotografá-los, foi necessário para digitalizá-los um equipamento de proporções maiores, como o escâner para papéis no tamanho A3, usado no projeto de pesquisa para a digitalização do jornal *O Serrinhense*. A experiência tem revelado que a primeira opção é a que exige mais cuidados para ser realizada, pois o processo de fotografar documentos, se feito de forma descuidada, pode resultar em fac-símiles com caracteres opacos que afetam a sua estética, além de dificultar ou impossibilitar a leitura. Isso pode ser visto no fac-símile de um texto de Eulálio Motta publicado no jornal

Gazeta do Povo (Cf. Figura 2), que apesar de não ser de boa qualidade, não impossibilitou a transcrição, entretanto, tornou imprecisos alguns dados linguísticos como a acentuação, por exemplo.

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

110

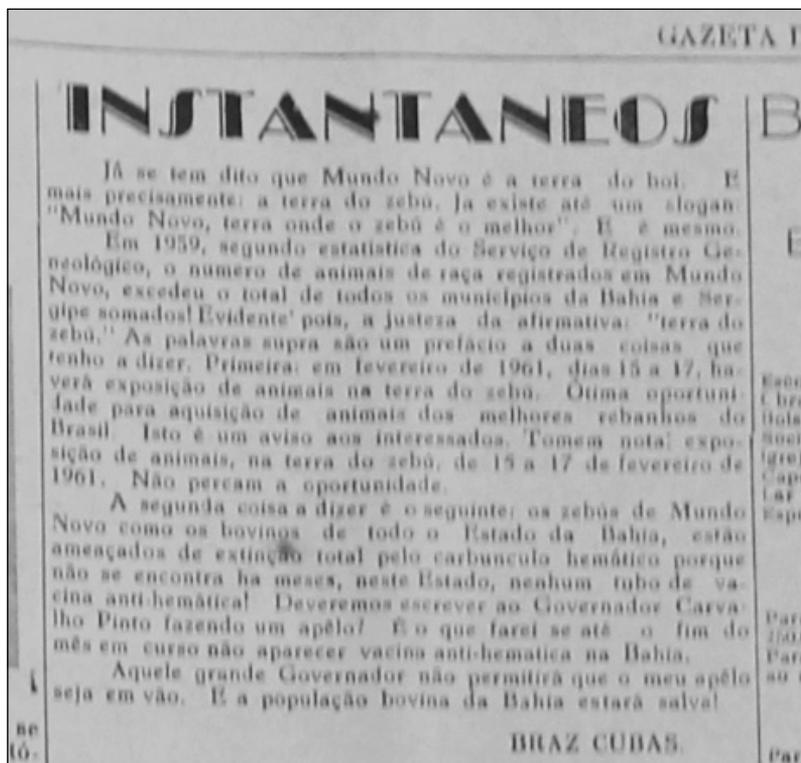


Figura 2 - Fac-símile de texto do jornal *Gazeta do Povo*

Fonte: Acervo de Eulálio Motta.

3.2 Transcrição

O segundo passo é a transcrição dos textos que deve seguir critérios que respeitem os códigos alfanuméricos e bibliográficos dos documentos. O recomendado é que se faça uma transcrição justalinear, ou seja, inserindo o fac-símile ao lado do espaço reservado para a transcrição. No caso dos periódicos, faz-se necessário recortar o trecho do fac-símile que contém o texto a ser transcrito para assegurar uma melhor visualização. O mesmo deve ser feito quando o texto se estender para a página seguinte, deixando sempre o trecho do fac-símile a ser editado ao lado do espaço da transcrição (Cf. figura 3). A transcrição dos documentos é indispensável ao estudo linguístico, visto que as ferramentas computacionais não podem analisar os dados da língua a partir dos fac-símiles.

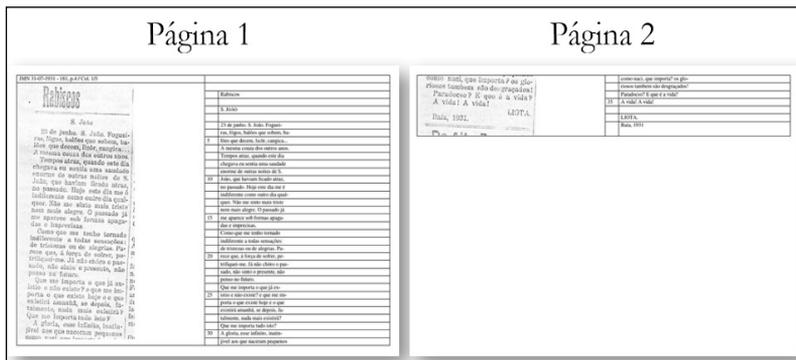


Figura 3 - Arquivo com a transcrição justalinear

Fonte: Elaborada pelos pesquisadores.

Edição de documentos e constituição de corpora linguísticos no acervo de Eulálio Motta

Existem softwares de transcrição automática com a tecnologia *Optical Character Recognition* (OCR), capazes de reconhecer caracteres a partir de um fac-símile e transformá-los em um arquivo de texto editável no computador. Entretanto, eles ainda têm se mostrado limitados, pois não conseguem reconhecer escritas cursivas individuais e apresentam dificuldades na leitura de fac-símiles de impressos com baixa qualidade de digitalização ou provenientes de documentos em mau estado de conservação. Os leitores OCR também cometem alguns erros frequentes que provém de falhas em seus mecanismos de leitura tais como a substituição de caracteres, *m* por *rn* (*r* e *n*), *e* por *c*, ou a supressão de espaços em branco. Por conta disso, os softwares de OCR ainda não provaram precisão suficiente para que possam substituir o trabalho de um editor.

Apesar desses problemas, os leitores OCR são ferramentas que auxiliam a Filologia e a Linguística de Corpus quando a questão é quantidade e tempo, visto que o trabalho de transcrição manual não consegue dar conta de textos extensos em um curto período. Nesses casos, tais ferramentas podem ser adotadas para a transcrição de textos impressos, mas apenas se houver a possibilidade de implementação de um plano de revisão ainda mais rigoroso. No caso do jornal *O Serrinhense*, a transcrição automática pôde ser feita sem gerar problemas graves de leitura dado o bom estado de conservação do documento e a boa qualidade da digitalização, ocorrendo apenas alguns pequenos erros de leitura, dentro dos problemas já previstos, que foram facilmente identificados no plano de revisão.

3.2.1 Transcrição de São João

Para transcrever um texto é necessário adotar critérios que respeitem as especificidades de cada tipo de documento. Assim, o projeto dispõe de critérios tanto para a transcrição de impressos, como para manuscritos. Os textos publicados nos periódicos tratam-se de fenotextos, versões concluídas, que não apresentam modificações feitas pelo escrevente ou por terceiros. Nesse caso, o modelo adotado foi o semidiplomático, considerado mais conservador que as transcrições modernizadas e interpretativas e mais interventivo que a diplomática. Para a transcrição do texto utilizou-se os seguintes critérios, elaborados a partir das propostas de Cambraia (1999; 2005) e Barreiros (2015a):

- 1) Indicar a sigla do jornal (J + iniciais), a data, o número, a página, a coluna e a quantidade de colunas da página. Ex.: JMN, 31-07-1931, n. 181, p. 4, col. 1/3;
- 2) Enumerar as linhas de 5 em 5 à margem esquerda;
- 3) Quando o texto ocupar mais de uma coluna do periódico, sinalizar entre colchetes o número da coluna antes de transcrevê-lo. Ex.: [Coluna 1];
- 4) Transcrever em fonte *Times New Roman*, tamanho 11, recuado à margem esquerda;
- 5) Manter a formatação do original: itálico, sublinhado, negrito, versalete e outros recursos tipográficos utilizados para destacar trechos ou itens lexicais, com exceção da alternância do tipo ou do tamanho das fontes, que serão indicados em notas de pé de página;
- 6) Manter a ortografia, a acentuação, o uso de maiúsculas e a pontuação;
- 7) Manter separação e junção entre caracteres e entre itens lexicais: sinais de pontuação, números, símbolos, itens lexicais em quebra de linha e no corpo do texto;
- 8) Manter caracteres trocados e demais erros tipográficos que venham a ocorrer;
- 9) Preservar as abreviaturas na transcrição e apresentar a abreviatura desdobrada em nota de rodapé, tomando como base, sempre que possível, as formas desenvolvidas no próprio documento;
- 10) Utilizar o símbolo {†} para indicar trechos, sentenças, itens lexicais e letras ilegíveis;
- 11) Indicar as conjecturas por meio do símbolo /*/;

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

112

- 12) Utilizar notas de pé de página para indicar informações complementares tais como: alternância da cor da tinta, de fonte, rasgões, furos, manchas, erros tipográficos, explicações e justificativas complementares do editor.

A transcrição semidiplomática permite pequenas intervenções editoriais que possibilitam a preservação das características relativas ao código alfanumérico, mas podem apagar aspectos materiais que envolvem este código e seus respectivos sentidos. Dentre as intervenções editoriais autorizadas pela transcrição semidiplomática, encontra-se a possibilidade de unir segmentos separados e separar os unidos. No entanto, esse tipo de alteração apaga aspectos da materialidade do texto tipográfico, como o rompimento da estrutura da coluna do jornal e a remoção de recursos utilizados para destacar itens lexicais no texto, como o caso da separação dos grafemas (ex.: “há uma q u e i x a”, *Jornal Gazeta do Povo*, 18-09-1960). Por conta disso, é extremamente importante a correta interpretação do código bibliográfico, pois a língua, ao ser registrada em um documento, assume uma forma material. Apresenta-se o fac-símile e a transcrição de *São João*, texto publicado por Eulálio Motta no jornal *Mundo Novo* em 31 de julho de 1981 (Cf. Figura 4).

Rabiscos

S. JOÃO

23 de junho. S. João. Fogueiras, fôgos, balões que sobem, balões que decem, licôr, cangica: . . . A mesma couza dos outros anos.

Tempos atrás, quando este dia chegava eu sentia uma saudade enorme de outras noites de S. João, que haviam ficado atrás no passado. Hoje este dia me é indiferente como outro dia qualquer. Não me sinto mais triste nem mais alegre. O passado já me aparece sob formas apagadas e imprecisas.

Como que me tenho tornado indiferente a todas sensações: de tristezas ou de alegrias. Parece que, á força de sofrer, petrifiquei-me. Já não choro o passado, não sinto o presente, não penso no futuro.

Que me importa o que já existio e não existe? o que me importa o que existe hoje e o que existirá amanhã, se depois, fatalmente, nada mais existirá? Que me importa tudo isto?

A glória, esse infinito, inafinível nos que naceram pequenos como naçi, que importa? os gloriosos também são desgraçados!

Paradoços? E que é a vida? A vida! A vida!

LHOTÁ.

Baía, 1931.

De Alto Bonito!

Transcrevemos, abaixo, o conteúdo da acta de inspecção, lavrada pelo inspector escolar, por ocasião de sua visita ao collegio particular do Sr Pedro S. Góz, no arraial de Alto Bonito.

«Em chegando nesta localidade onde pernôtei, hoje 23 de Maio, encontrei a escola particular sob a direcção do Sr. Pedro S. Góz funcionando regularmente, suprimido destarte, a falta de uma escola publica.

Julgo de justiça sã que o municipio o auxiliasse com uma pequena subvenção, uma vez que, alem de na epocha actual lhe serem poucos os rendimentos advindo dos seus alumnos, seria mais um incentivo ao Regente, que se dedica com tanto carinho a este punhado de crianças, homens futuros de amanhã, e que á Patria não de servir illuminados pela luz bendicta da instrucção.

Alto Bonito, 24 de Maio de 1931.

(Assignado)—Prof. Alfredo Lino da Costa—Inspector Escolar.

CONEGO JOSÉ DIAS D'AFFONSECA

Vindo da cidade de Cachoeira, onde se achava em gozo de licença chegou a esta cidade a 23 da findante, o nosso digno perecho Conego José Dias de Affonseca, assumindo os destinos de sua Parochia.

O Rev.º veio em uma chic «li-moisine» Ford que adquiriu para seu uso, graças ás estradas de rodagem para todos os pontos do nosso municipio.

Ao conego José Dias mandamos o nosso abraço e votos de boas vindas.

CREANÇA PHENOMENO

De «Folha do Norte», jornal que se edita em Feira de Santana, deste Estado, transcrevemos a seguinte noticia:

«Pessoa merecedora de fé informamos que no lugar denominado Passagem de Umburanas, no municipio de Conceição da Feira, nasceu, ha dias, uma criança do sexo feminino, com duas caras, sendo que uma ao lado da cabeça e está mais perfeita que a anterior, principalmente os olhos de que é provida.

A creança de Passagem de Umburanas, tem, pois, 4 olhos pelos quaes chora e duas bocças, pelas quaes recebe alimentação.

Tem 18 dias de nascida e ainda vivia ao tempo em que nos forneceram esta curiosa noticia que deve interessar á sciencia medica.

Declaração

Domingos Martins Brandão, que outrora usava Domingos Felix Brandão, declara para todos os effeitos que de hoje por diante assigna legitimamente Domingos Martins Brandão.

F. Ladeira do Mel, (Mundo Novo) 25 de Julho de 1931.

Domingos Martins Brandão

VISITAS

Estiveram em nossa redacção dando-nos o prazer de amavel visita, no decorrer desta semana os srs. Manoel Benjamin Moreira zeloso Inspector dos Telegraphos e Arnimiro Oliveira, negociante nesta praça.

Tambem visitou a nossa tenda de trabalhos o Cel. Isalino Queiroz Mattos pessoa conceituada na villa de Baixa Grande, onde reside.

A todos os nossos agradecimentos e votos de boa viagem.

JOSÉ DE NOVAES BARRETO

(ADVOGADO PROVISIONADO)

Accepta o patrocínio de causas civis, commerciaes e criminaes Inventarios—Partilhas—Desquites etc.

MUNDO NOVÔ—BAIA

Politiqueice

Especialmente no recalcitrante sertão bahiano, onde o costume das chefias locais ja é cousa comum e arraizada, tem, por esta mesma razão, a politiqueice, a sua razão de ser.

O «Chefe» é cousa que abaixo não se uae para tão cedo, em que peze ja merecer de muita gente, que começa a enxergar com os olhos de ver, as mais justas repulsas.

Mas a verdade é que, aqui para cima, a politiqueice impera e prepondera!

Trez são as maneiras de praticar a chefia no sertão, o que obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

—A primeira é a do «pinguelo!» E' pelo direito da força: ou obedece sempre á aptidão dos taes chefes:

Figura 4 - Fac-símile do texto São João

Fonte: Acervo do escritor Eulálio Motta.

Rabiscos

S.² João

- 23 de junho. S. João. Foguei-
ras, fógos, balões que sobem, ba-
lões que decem, licôr, cangica...
5 A mesma couza dos outros anos.
Tempos atraz, quando este dia
chegava eu sentia uma saudade
enorme de outras noites de S.
10 João, que haviam ficado atraz,
no passado. Hoje este dia me é
indiferente como outro dia qual-
quer. Não me sinto mais triste
nem mais alegre. O passado já
15 me aparece sob formas apaga-
das e imprecisas.
Como que me tenho tornado
indiferente a todas sensações :
de tristezas ou de alegrias. Pa-
20 rece que, á força de sofrer, pe-
trifiquei-me. Já não chóro o pas-
sado, não sinto o presente, não
penso no futuro.
Que me importa o que já ex-
25 istio e não existe ? e que me im-
porta o que existe hoje e o que
existirá amanhã, se depois, fa-
talmente, nada mais existirà ?
Que me importa tudo isto ?
30 A gloria, esse infinito, inatin-
jível aos que naceram pequenos
como naci, que importa ? os glo-
riosos tambem são desgraçados !
Paradocso ? E que é a vida ?
35 A vida ! A vida !

LIOTA.

Baía, 1931.

*Edição de
documentos e
constituição
de corpora
linguísticos
no acervo de
Eulálio Motta*

115

3.2.2 Transcrição de Dr. Agenor Brandaõ: Saudações

No acervo de Eulálio Motta, encontra-se também uma grande quantidade de rascunhos que revelam as alterações feitas pelo escritor durante a produção dos seus textos. Para a transcrição desses documentos foram adotados os critérios estabelecidos por Barreiros (2013; 2015b), que visam preservar não apenas os códigos alfanuméricos, mas representar por meio de marcadores genéticos, a disposição espacial destes códigos

2 SÃO.

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

116

e de outros que revelam os procedimentos escriturísticos empregados, como as rasuras e borrões, que marcam os cancelamentos, as substituições etc. Trata-se de uma transcrição genética, que se ocupa não só da representação do texto verbal, mas também dos procedimentos de criação que ficaram registrados no documento. Na transcrição genética dos rascunhos emprega-se os seguintes critérios:

- 1) Indica-se a folha;
- 2) Enumera-se as linhas de 5 em 5 à margem esquerda da folha;
- 3) Transcreve-se o texto como se encontra no original, interferindo apenas com marcadores genéticos estabelecidos;
- 4) A transcrição linearizada acomoda as rasuras, substituições, correções e acréscimos na sequência lógica do texto, não obedecendo à topografia do original;
- 5) São mantidas as interpolações, os lapsos do autor, a ortografia, a acentuação, o uso de maiúsculas, a pontuação e registraram-se todas as correções, emendas, rasuras e acréscimos, através da utilização de símbolos;
- 6) A rubrica do autor indica-se entre colchetes.

Quanto aos símbolos utilizados nas edições:

- 1) {} seguimento riscado, cancelado;
- 2) {†} seguimento ilegível;
- 3) {†} / \ segmento ilegível substituído por outro legível na relação {ilegível} /legível\;
- 4) {} / \ substituição por sobreposição, na relação {substituído} / substituto\;
- 5) {} [↑] riscado e substituído por outro na entrelinha superior;
- 6) {} [→] riscado e substituído por outro na margem direita;
- 7) {} [←] riscado e substituído por outro na margem esquerda;
- 8) [] acréscimo no curso da linha;
- 9) [↑] acréscimo na entrelinha superior;
- 10) [↓] acréscimo na entrelinha inferior;
- 11) [→] acréscimo na margem direita;
- 12) [←] acréscimo na margem esquerda;
- 13) [↑{}] acréscimo na entrelinha superior riscado;
- 14) [↑{†}] acréscimo na entrelinha superior ilegível;
- 15) [↑{} / \] acréscimo na entrelinha superior riscado e substituído por outro na sequência;

- 16) [$\uparrow\{\text{f}\}$ / \] acréscimo na entrelinha superior ilegível e substituído por outro na sequência;
- 17) [$\downarrow\{\}$] acréscimo na entrelinha inferior riscado;
- 18) [$\downarrow\{\text{f}\}$] acréscimo na entrelinha inferior ilegível;
- 19) [$\downarrow\{\}$ / \] acréscimo na entrelinha inferior riscado e substituído por outro na sequência;
- 20) [$\downarrow\{\text{f}\}$ / \] acréscimo na entrelinha inferior ilegível e substituído por outro na sequência;
- 21) [$*\uparrow$] parte do texto localizada à margem superior indicada pelo autor através de seta, linha ou números remissivos;
- 22) [$*\downarrow$] parte do texto localizada à margem inferior indicada pelo autor através de seta, linha ou números remissivos;
- 23) [$*\rightarrow$] parte do texto localizada à margem direita indicada pelo autor através de seta, linha ou números remissivos;
- 24) [$*\leftarrow$] parte do texto localizada à margem esquerda indicada pelo autor através de seta, linha ou números remissivos;
- 25) [$*(\text{f. ou p.})$] parte do texto localizada em outro fólio ou página indicada pelo autor a partir de números e letras remissivos ou anotações. Nesses casos, o número do fólio ou da página aparece entre parênteses;
- 26) / * / leitura conjecturada;
- 27) () intervenção do editor (acréscimos e informações).

A seguir, apresenta-se o fac-símile (Cf. Figura 5) seguido da transcrição de um rascunho de carta³ para Agenor Brandão, que faz parte do caderno *Farmácia São José*.

3 No caderno Farmácia São José existem diversos rascunhos de carta. Não se sabe se todos os rascunhos foram passados a limpo e enviados como cartas propriamente ditas. Por meio da análise desses rascunhos, é possível confirmar o envio de algumas delas.

- Dr. Agenor Brandaõ:
Saudações
 Em mãos a sua carta de 4 do corrente, que
 respondo: Você julgou a minha [↑recusa] em trabalhar
 5 no domingo, como falta de interesse de mi-
 nha parte. Devo-lhe dizer o seguinte: o 3º
 Mandamento da Lei de Deus manda: “guarda
 os domingos e festas”; o 1º Mandamento da
 Santa Igreja Católica Apostólica Romana,
 10 diz: – “Ouvir missa inteira nos domingos e festas
 de guarda.” [↑E este é o meu trabalho nos domingos: ouvir missa,] ler os evangelhos, etc.
 Como vê, Sr. Dr., o motivo de minha recu-
 sa é o respeito devido a Deus, á Igreja de
 Deus, á Fé em Deus.
 15 Ha creaturas, entretanto, para as quaes {Deus} [↑a realidade [↓suprema]]
 [↑e motora de tudo] é o interesse., Tudo na vida de taes criaturas
 é relacionado com interesse. O interesse, {no}
 {pior de} /para\ taes] creaturas, é o motivo de
 tudo. Confere.
 20 Atenciosamente,
 [Eulalio de Miranda Motta.]
 5-11-941.
- P.S. Para o seu {†} /trabalho\ de dia domingo, peço[↓-]
 -lhe que aguarde a vinda do socio “que tem
 25 mais interesse.”
 O mesmo.

(A página foi enumerada a lápis pelo autor, contém o número ‘1’ na parte superior direita da folha.

Na linha 23, abaixo do item lexical ‘sócio’, há dois traços feitos com lápis de cor azul celeste.

Na linha 19 há um ‘x’ no grafema ‘C’ do item lexical ‘Confere’.)

3.3 Plano de revisão

Após a digitalização e a transcrição, o terceiro passo é a preparação de um plano de revisão, o qual consiste em um processo de inspeção feito por mais de um filólogo. Nessa etapa, é recomendável que o texto seja revisado primeiramente pelo editor e, posteriormente, por dois ou mais integrantes do grupo de pesquisa, retornando ao final do ciclo para as mãos do editor que irá avaliar as correções sugeridas e estabelecer a versão final. O plano de revisão é de suma importância para realizar uma transcrição confiável, pois ajuda a resolver problemas de digitação e a corrigir erros provenientes de lapsos de leitura.

A obrigatoriedade do plano de revisão pode ser encontrada também nas diretrizes apresentadas pelo *Committee on Scholarly Editions* (CSE). Na seção introdutória, o CSE afirma que é necessário:

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

A proofreading plan that provides for meticulous proofreading at every stage of production so that the accuracy of the text, textual essay, and textual apparatus is not compromised. Automated proof-reading programs ('spell checkers'), word lists, and computerized collation or file comparing programs can be used to alleviate the burden, but they cannot substitute for manual proof-reading nor should they ever be allowed to make unverified changes in the text (MLA, 2002, on-line).⁴

120

Nesse processo, os corretores automáticos se tornaram grandes inimigos do filólogo, pois, ao aplicar correções em um texto, estas ferramentas alteram a grafia, removem e adicionam diacríticos e substituem itens lexicais que não reconhecem, comprometendo a confiabilidade da transcrição. Recomenda-se que, para evitar alterações indesejáveis no texto editado, se desative a função do corretor ou se imprima uma versão para a revisão. É necessário também que no caso dos textos manuscritos tanto o editor, quanto o revisor estudem os padrões caligráficos do escrevente para elaborar um material de consulta que ajude a solucionar problemas de leitura de forma mais objetiva.

A revisão deve sempre ser um processo cuidadoso, em que se busque preservar cada intervenção realizada, a fim de que o editor do texto possa avaliá-las. Assim, recomenda-se que cada revisor utilize canetas ou marquem as alterações feitas na transcrição com cores diferentes, quando a revisão for no arquivo digital, para que o editor possa comparar as observações de cada um dos revisores. Além disso, no caso dos arquivos digitais, os revisores devem sempre buscar preservar as versões anteriores da transcrição, em vez de substituí-las, para facilitar o processo de avaliação final.

4 Tradução nossa: Um plano de revisão que provê a revisão meticolosa em todos os estágios de produção, de modo que a precisão do texto, do ensaio textual e do aparato textual não seja comprometida. Programas de revisão automática ('corretores ortográficos'), listas de palavra, e colação computadorizada ou programas de comparação de arquivos podem ser usados para aliviar o trabalho, mas eles não podem substituir a revisão manual e jamais devem ser autorizados a fazer modificações não verificadas no texto.

3.4 Adaptação

As transcrições conservadoras, diferentes das modernizadas e interpretativas, buscam preservar ao máximo as características linguísticas e, na medida do possível, representar aspectos bibliográficos do documento editado. Para tanto, faz-se necessário o uso de operadores genéticos (Cf. critérios apresentados na seção 3.2) para sinalizar na transcrição outras características do texto que transcendem o código alfanumérico, assim como os sinais que indicam problemas na compreensão de algum segmento e demais interferências do editor, como acontece com as conjecturas.

No momento de transpor os dados para um *software* de análise linguística que não é capaz de processar esses códigos, é necessário remover os operadores para possibilitar a análise dos dados. No entanto, a remoção deliberada desses símbolos pode ocasionar problemas para a decodificação dos dados em um *corpus* de dimensões maiores. Um exemplo disso é o que ocorre com o *AntConc*, *software* desenvolvido para análise lexical, que consegue processar um determinado *corpus* apenas se o arquivo estiver em formato TXT⁵ e se os itens lexicais aparecerem integralmente, não sendo divididos por espaços em branco ou, no caso das edições filológicas, pelos operadores empregados.

O *AntConc* foi um dos *softwares* utilizados para o tratamento lexicográfico de textos do escritor Eulálio Motta por Barreiros (2017b), em sua tese intitulada *Vocabulário de Eulálio Motta* e, posteriormente, adotado nos estudos lexicais desenvolvidos no projeto de pesquisa *Estudos lexicais no acervo de Eulálio Motta*. Dentre as diversas vantagens oferecidas pelo *software* pode-se destacar que:

A interface do *AntConc* é bem simples e, na mesma janela, é possível navegar por diferentes opções de análise, que permitem descobrir como uma palavra ocorre, o quanto ocorre, em que contextos e quais a acompanha, encontrar padrões e variáveis de uso na escrita e fazer levantamento terminológico. Sua praticidade de uso possibilita a extração de listas de palavras (*Word List*), listas de concordâncias (*Concordance*) e de palavras-chaves (*KeyWord*), além de gerar gráficos com os dados analisados. Estas ferramentas são de grande relevância para o linguista, em

5 Atualmente, está em desenvolvimento uma versão capaz de analisar arquivos no formato pdf.

especial, para o lexicógrafo, pois fornece o conjunto das combinações e das colocações que a palavra pode ter em um determinado *corpus* (BARREIROS, 2017b, p. 89).

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

122

Nas edições de rascunhos, por exemplo, é comum encontrar uma grande quantidade de problemas que dificultam a análise linguística automática em diversos níveis. Nesse caso, é imprescindível um processo de adaptação dos arquivos editados para possibilitar a análise dos dados. A adaptação da transcrição deve ser realizada por um filólogo com experiência em análise de fenômenos linguísticos e não pode ser um exercício intuitivo, mas feito de acordo com critérios rígidos como no trabalho desenvolvido anteriormente, a elaboração de uma transcrição confiável do texto verbal.

Como exemplo, observar-se na Figura 6 alguns casos de sobreposição de itens lexicais. A transcrição desses trechos implica em um processo de linearização que deve ser sempre sinalizado pelo editor. No entanto, a linearização de uma escrita não linear gera algumas anomalias linguísticas. Assim, no caso da Figura 6, os itens lexicais concorrentes se apresentam dispostos um ao lado do outro, construindo sentenças que, além de serem, muitas vezes, incoerentes, necessitam ser analisadas e interpretadas individualmente.

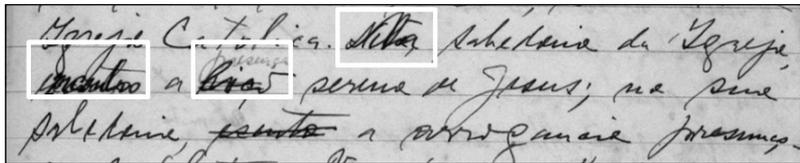


Figura 6 - Segmento substituído na folha 28r do caderno *Farmácia São José*.

Fonte: Acervo de Eulálio Motta.

Diante dessa problemática, poderia haver defensores da remoção das formas alternativas do corpo do texto, estabelecendo uma versão analisável do *corpus*. Entretanto, sabe-se que tal iniciativa implicaria na perda de dados linguísticos. Assim, a questão da linearização do texto consiste ainda em um problema irresolúvel, visto que os programas computacionais desenvolvidos atualmente não foram pensados para compreender a dinâmica do texto em processo, mas que pode ser contornado ao se adotar algumas estratégias, considerando o *software* que se pretende utilizar e o estudo a ser realizado. A sinalização dos

itens linearizados é a única solução para essa questão, pois possibilita a identificação dos segmentos que foram substituídos durante a escrita do texto.

Na Figura 7, apresenta-se um trecho do fac-símile do fólio 30r do caderno *Farmácia São José*, em que o escrevente substituiu o trecho ‘em Cristianismo’ por ‘no assunto’, que na transcrição aparecem de forma linearizada com operadores, ‘é um sábio {em}/no\ {Cristianismo;}/[↑assunto;]’, construindo uma sequência que em nenhum momento foi realizada pelo escrevente, ou na língua portuguesa, ‘é um sábio em no Cristianismo assunto’. Dessa forma, é necessário interpretar o *corpus* por meio da análise dos processos escriturísticos para atribuir coerência ao texto e sinalizar as formas alternativas de modo que possibilite a análise automática.

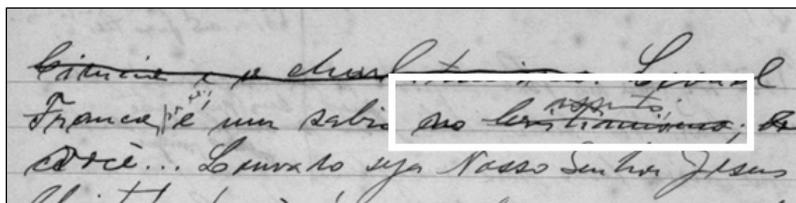


Figura 7 - Exemplo de substituição no fólio 30r do caderno *Farmácia São José*

Fonte: Acervo do escritor.

Outro desafio está na preservação da integridade dos itens lexicais para que o *software* possa reconhecê-los. Em alguns casos, as sobreposições que se encontram no interior dos itens lexicais resultariam em anomalias grafemáticas (ex.: ‘ca{z}/s\’a’, para a troca do s pelo z, ‘caza’), morfemáticas (‘me mand{aste}/ou\’, resultando em ‘mandasteou’) ou outras formas de fragmentação não estruturáveis, como no caso apresentado na Figura 7. De fato, essas formas nunca seriam aceitas como válidas por qualquer filólogo. No entanto, ao mantê-las separadas na transcrição, considerando a forma em que aparecem (‘caza’ e ‘s’, ‘mandaste’ e ‘ou’), tem-se problemas na análise estatística, pois a forma ‘ou’, por exemplo, poderia ser contabilizada como uma conjunção.

No caso dos impressos, é possível mencionar o mesmo exemplo utilizado anteriormente de separação de grafemas (ex.: ‘há uma q u e i x a’, *Jornal Gazeta do Povo*, 18-09-1960). Essa fragmentação não é identificada pelo *software*, o qual irá apresentar uma lista com partes de itens lexicais, dificultando a etapa de levantamento das unidades de análise, conforme

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

124

pode ser visto na Figura 8. É necessário reconstituir as unidades sobrepostas e apresentá-las também de modo linearizado (ex.: 'ca{z}/s\`a' por 'casa {caza}', seguindo a ordem: forma substituta e forma substituída). Deve-se também discutir o que fazer com os casos de reescrita da mesma unidade, que resultam em repetição, alterando os dados na análise estatística, especialmente ao lidar com *corpus* de grande escala.

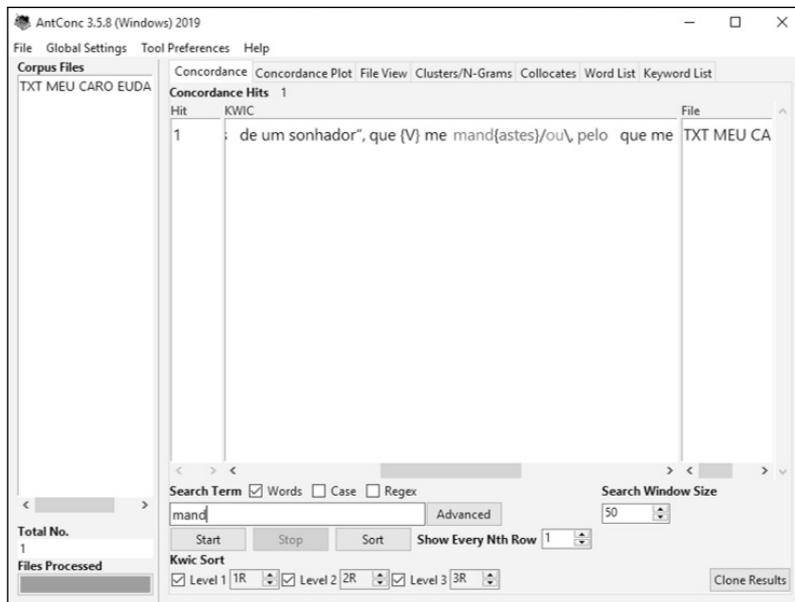


Figura 8 - Interface do AntConc na aba Concordance

Fonte: Elaborada pelos pesquisadores.

No caso das conjecturas, em sua condição de dados duvidosos, devem ser sempre sinalizados no processo de adaptação, pois o dado conjecturado passa a ser considerado pela ferramenta como os demais, ou seja, será incluído junto aos dados seguros, podendo comprometer os resultados das futuras análises. Assim, em casos como em que se conjectura parte de um item lexical, não se deve simplesmente apresentá-lo de forma reconstituída, como nos exemplos mencionados, mas deve-se adotar também um sinal de identificação. O mesmo deve ser pensado com relação ao desdobramento das abreviaturas, que consistem em dados artificiais.

A partir dessa análise prévia, foram pensadas adaptações preliminares: sinalizar a linearização tanto de trechos alterados, 'é um sábio {em}/no\ {Cristianismo;}/[↑assunto;]\` por 'é um sábio no assunto {no Cristianismo}', como de itens lexicais reconstituídos,

‘ca{z}/s\’a’ por ‘casa {caza}’, ou apenas adaptados, como no caso dos impressos, ‘q u e i x a’ por ‘queixa {q u e i x a}’. Os itens conjecturados serão sinalizados entre barras inclinadas e precedidos de asterisco /*/. Vale ressaltar que esses são critérios provisórios e que ampliações e modificações deverão ser realizadas a partir dos testes com uma quantidade maior de documentos para que seja apresentada uma proposta definitiva.

Os procedimentos filológicos mencionados, juntamente com os já empregados pela Linguística de Corpus, como a junção dos itens separados ou separação dos itens unidos, a remoção de símbolos e notas do editor, constituem o processo que se denomina aqui de adaptação, ou seja, uma metodologia de caráter mais complexo do que a limpeza. A adaptação é crucial para o estudo linguístico feito a partir de edições filológicas, já que, muitas vezes, os linguistas interessados em analisar fenômenos em uma transcrição complexa como a genética, por exemplo, nem sempre estão familiarizados com o significado dos símbolos utilizados, o que exigiria uma enfadonha e recorrente consulta às transcrições para a verificação de problemas nos dados do *corpus* em análise.

4 Considerações finais

A Filologia e a Linguística, apesar de apresentarem convergências, são áreas distintas, com objetivos particulares. Enquanto a primeira interpreta e analisa o documento como um fenômeno social, histórico, cultural e linguístico, tendo em vistas a sua reinserção em um novo contexto por meio da preparação de edições, a segunda o utiliza como fonte para interpretação da língua. Assim, conforme aponta Cunha (2004 [1987]), o método da Filologia é o estudo *in praesentia*, um olhar para as particularidades dos documentos, inclusive as particularidades linguísticas, que deve dialogar com o estudo *in abstracto* desenvolvido pela Linguística, enquanto a Linguística parte dos dados *in praesentia* para interpretá-los *in abstracto*.

É necessário, portanto, reconhecer que a construção de um *corpus* para análise linguística, tanto a partir de textos escritos, como falados, não pode ser realizada sem recorrer às discussões filológicas. Do mesmo modo, não se pode editar um documento antigo sem conhecer a história da língua, conhecimento esse que é fornecido pela Linguística. Graças a essa relação simbiótica, a Linguística tem acesso a outras

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

126

modalidades de fontes, como os rascunhos de cartas, por exemplo, que marcam a gênese textual e trazem dados mais próximos do espontâneo, ao passo que a Filologia pode se valer dos estudos linguísticos para compreender a complexidade do processamento da língua na construção dos textos verbais.

Com o avanço das Humanidades Digitais, essas duas ciências se veem instrumentalizadas com novos recursos que ampliam as suas possibilidades, promovendo uma renovação nas esferas teórica e metodológica. Assim, novos *softwares* vêm sendo incorporados à rotina investigativa do filólogo e do linguista, permitindo a realização de pesquisas de forma rápida e segura. No acervo do escritor Eulálio Motta, a aplicação dos novos recursos ao trabalho com o *corpus* do acervo tem possibilitado a preservação e a difusão da obra deixada pelo escritor, além do uso desses dados para o estudo da língua.

Nesse contexto, adotar um método rigoroso como o aqui apresentado possibilita tanto a preservação dos documentos, por meio da digitalização, como também a preparação de transcrições confiáveis, com o uso de critérios adequados e um plano de revisão bem articulado. Ademais, a reflexão sobre a preparação das transcrições para se tornarem *corpora* para o estudo linguístico demonstra o quanto o rigor filológico é relevante para a Linguística de Corpus, já que os problemas aqui apontados são facilmente verificáveis num estudo de pequenas dimensões, mas tendem a se agravar conforme se aumenta a proporção do tamanho dos *corpora*, podendo comprometer a confiabilidade de todo um banco de dados.

Com base no exposto, pode-se afirmar que a pesquisa no acervo do escritor Eulálio Motta é de grande relevância para as Humanidades Digitais e para a preservação do patrimônio cultural da Bahia. A pesquisa e as edições feitas a partir dos documentos do acervo têm permitido descortinar aspectos históricos do sertão baiano, a história preservada pela língua inscrita na documentação e a história contada indiretamente pela materialidade desta documentação que, assim como as fotografias e os objetos, servem como indícios das práticas culturais da época de seu arquivamento. Dessa forma, a pesquisa em acervos torna visível e acessível o que foi invisibilizado pela historiografia tradicional.

REFERÊNCIAS

BARREIROS, Liliane L. S. O uso de ferramentas computacionais na elaboração do Vocabulário de Eulálio Motta: AntConc e FLEx. **A Cor das Letras** (UEFS), Feira de Santana, v. 18, n. 2, 2017a, p. 216-241.

BARREIROS, Liliane L. S. **Vocabulário de Eulálio Motta**. Orientador: Célia Marques Telles. 2017. 360f. Tese (Doutorado em Língua e Cultura) – Instituto de Letras, Universidade Federal da Bahia, UFBA, Salvador, 2017b.

BARREIROS, Liliane L. S.; BARREIROS, Patrício N; SANTIAGO, Stephanie da C. A trajetória do escritor Eulálio Motta no caderno Farmácia São José. In: LOSE, Alícia Duhá; SOUZA, Arivaldo Sacramento; BARREIROS, Patrício Nunes, DUARTE, Rosinês de Jesus. (Orgs.) **Filologia, cultura escrita e estudos culturais**. Feira de Santana: UEFS Editora, 2018. p. 379-403.

BARREIROS, Patrício N. **Projeto Edição das obras inéditas de Eulálio Motta**: Critérios de Edição dos Textos do Jornal Folha do Norte. Feira de Santana, 2015a. Disponível em: <https://eulaliomotta.wordpress.com/criterios-de-edicao-dos-textos-do-jornal-folha-do-norte/>. Acesso em: 19 ago. 2019.

BARREIROS, Patrício N. **O Pasquineiro da Roça**: a hiperedição dos panfletos de Eulálio Motta. Feira de Santana: UEFS Editora, 2015b.

BARREIROS, Patrício N. **O Pasquineiro da roça**: edição dos panfletos de Eulálio Motta. Orientador: Célia Marques Telles. 2013. 386f. Tese (Doutorado em Letras), Programa de Pós-Graduação em Letras e Linguística – Instituto de Letras – Universidade Federal da Bahia, Salvador, 2013.

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

CAMBRAIA, César N. Subsídios para uma proposta de normas de edição de textos antigos para estudos lingüísticos. In: RODRIGUES, Angela C. de S.; ALVES, Ieda M.; GOLDSTEIN, Norma S.; (orgs.) **I Seminário de Filologia e Língua Portuguesa**. São Paulo: Humanitas/FFLCH/USP, 1999, p. 13-23.

Edição de documentos e constituição de corpora lingüísticos no acervo de Eulálio Motta

127

Iago Gusmão
Santiago

Stephanne da
Cruz Santiago

Liliane Lemos
Santana
Barreiros

128

CUNHA, Celso. O ofício de filólogo [1987]. In: CUNHA, Celso. **Sob a pele das palavras**. Rio de Janeiro: Nova Fronteira/Academia Brasileira de Letras, 2004, p. 341-359.

HOCKEY, Susan. The History of Humanities Computing. In: SCHREIBMAN, Susan; SIEMENS, Ray; UNSWORTH, John (Org.). **A Companion to Digital Humanities**. Oxford: Blackwell, 2004. Disponível em: <http://www.digitalhumanities.org/companion/>. Acesso em: 10 jan. 2020.

MATTOS E SILVA, Rosa V. **Caminhos da Linguística histórica: ouvir o inaudível**. São Paulo: Parábola Editorial, 2008.

MODERN LANGUAGE ASSOCIATION. **Sun Site: Preliminary Guidelines for Electronic Scholarly Editions**. Berkeley, jun. 2002. Disponível em: <http://sunsite.berkeley.edu/MLA/guidelines.html>. Acesso em: 19 ago. 2019.

MOREIRA, Marcello. **Crítica Textualis in caelum revocata?: uma proposta de edição e estudo da tradição de Gregório de Matos e Guerra**. São Paulo: Editora da Universidade de São Paulo, 2011.

OTHERO, Gabriel. A Linguística Computacional: uma breve introdução. **Letras de Hoje** (PUCRS), Porto Alegre, v. 41, n. 2, 2006, 341-351.

PAIXÃO DE SOUSA, Maria C. Linguística Histórica. In: NUNES, José Horta; PFEIFFER, Claudia. (Orgs.). **Introdução às Ciências das Linguagens: Linguagem, História e Conhecimento**. Campinas: Pontes, 2006. p. 11-48.

SANTIAGO, Iago G.; BARREIROS, Liliane L. S.; BARREIROS, Patrício N. Os itinerários do jornalista Eulálio Motta no jornal Mundo Novo. In: LOSE, Alícia Duhá; SOUZA, Arivaldo Sacramento; BARREIROS, Patrício Nunes, DUARTE, Rosinês de Jesus. (Orgs.) **Filologia, cultura escrita e estudos culturais**. Feira de Santana: UEFS Editora, 2018. p. 405-427.

SAUSSURE, Ferdinand de. **Curso de Linguística Geral**. São Paulo: Cultrix, 2006 [1916].

TELLES, Célia M. Filologia textual e Linguística românica. **Scripta Philologica** (UEFS), Feira de Santana, v.1, 2005, p. 90-99.

UEFS/CONSEPE. Resolução CONSEPE N° 070/2016. Aprova o Projeto de Pesquisa **Edição das Obras Inéditas de Eulálio de Miranda Motta (IV Etapa)**, sob a coordenação do Prof. Dr. Patrício Nunes Barreiros, do Departamento de Letras e Artes, desta Universidade, financiado pela FAPESB. Feira de Santana-BA: D.O.E., 2 set. 2016.

UEFS/CONSEPE. Resolução CONSEPE N° 137/2017. Aprova o Projeto de Pesquisa **Estudos lexicais no acervo de Eulálio Motta**, sob a coordenação da Profa. Dra. Liliane Lemos Santana Barreiros, do Departamento de Letras e Artes, desta Universidade. Feira de Santana-BA: D.O.E., 12 dez. 2017.

*Edição de
documentos e
constituição
de corpora
linguísticos
no acervo de
Eulálio Motta*

129

