



Carta do século XIX sobre Mesa Cartesiana. Abaixo, transcrição em caracteres para estudo do português da época

001228
 00 Casarão da Silva e Melo: "Carta de liberdade da Cabra de nome Sofia"
 00 Versão Texto original: [D.O. 5864/LJWS]
 00 Arquivo gerado pela ferramenta E-Doctor

Livros: folha 40 verso

Carta de liberdade da Cabra de nome Sofia passada por Antonio Jose de Souza Paes, outrora Senhor daquela

Eu Antonio Jose de Souza Paes abaixo assignado, sou possuidor da Cabrinha Sofia sem embargo algum, e por que he minha vontade, e lhe tenho grande amor, e fi em diante lhe confiro a liberdade, e fi ca fora, como si tal nascesse: podendo seguir o destino, que lhe parecer como arbitra de si mesma, e para seo titulo lhe passo a presente carta por mim escripta, e assignada, que quero tenha validade, como si fosse verba de titulo, pedindo as Justicas do Imperio lhe deem toda a validade que o Direito outorga. São Felipo cinco de abril de mil oito centos e quatro dize mil oito centos e trinta e quatro = Antonio Jose de Souza Paes = Reconheço verdadeiras e dou fé. Caetite

Registro da escrita

Novo método facilita a transformação de documentos manuscritos históricos em arquivos digitais

Marcos de Oliveira

A dificuldade em manusear documentos históricos raros e manuscritos para análise dos textos levou um grupo de pesquisadores da Universidade Estadual do Sudoeste da Bahia (Uesb) a desenvolver um método de fotografia que facilita a transcrição e compreensão de fenômenos linguísticos de uma época. “Existem documentos e livros antigos para os quais o método tradicional de obtenção da imagem por escaneamento pode prejudicar ou até destruir o original porque é preciso, muitas vezes, dobrá-los ou desencaderná-los para uso no escâner”, diz o professor Jorge Viana Santos, do Laboratório de Pesquisa em Linguística de Corpus (Lapelinc) da Uesb. O objeto de estudo dos pesquisadores são livros e documentos cartoriais manuscritos do século XIX que já tiveram grande manuseio e cujo estado é bem frágil. “Diferentemente da fotografia, no escaneamento o documento é que se adapta ao aparelho e não o contrário”, diz. Para a digitalização de documentos impressos, já existem *softwares* bem difundidos que levam o nome de reconhecimento óptico de caractere (OCR na sigla em inglês) e podem ler o documento a partir de escâneres e transformá-lo em digital. Em documentos manuscritos não existe essa possibilidade.

O método criado pelo professor Santos em colaboração com a professora Cristiane Namiuti Tempon, também da Uesb, começa com a captura da imagem em uma câmera fotográfica. Para isso, o documento é assentado em uma espécie

de placa plana de plástico de cor cinza e quadriculada milimetricamente, característica que serve para informar no computador a exata medida do papel. Denominada pelo grupo de Mesa Cartesiana, sobre ela também são colocadas escalas de tom de cores, informações catalográficas, paginação e sequência. A página do documento pode tanto ser apresentada no computador com todas essas informações como também de forma recortada, apenas a parte manuscrita.

DETALHES NA TELA

A transposição do documento do mundo físico, intermediado pela fotografia, para a formatação digital, é feita por um *software* desenvolvido também no Lapelinc. Ele permite interpretar esses dados e recuperar numa tela de computador os tons e cores originais de um documento. Assim, o método faz a transposição de documentos manuscritos históricos para a formação de conjuntos de textos eletrônicos com aspecto próprio para pesquisa científica.

As vantagens do Método Lapelinc se expandem também na facilidade de aumentar o texto original na tela do computador para verificar detalhes ou tirar dúvidas em relação à escrita. Com o documento digital é possível fazer várias consultas sem deteriorar o material histórico. Segundo Santos, o novo método contribui para a análise dos paleógrafos, especialistas que leem o texto para estudos de linguagem e fazem a transcrição e adaptação ao português atual se for o caso. A linguística de *corpus* (texto para

análise) necessita do original em caracteres para a compilação de *corpora* (conjunto de *corpus*) para análise linguística automática. “Nosso método permite montar o *corpus* eletrônico que forma um banco de dados no qual é possível identificar cada palavra e etiquetá-la, facilitando o trabalho do linguista na busca pelo seu objeto de estudo; pode-se, assim, etiquetar substantivos e verbos, por exemplo”, diz Santos. “O historiador pode ler na linguagem de hoje, mas o linguista quer saber como o texto foi concebido naquela época para determinar o padrão e a evolução da linguagem.”

O trabalho de estrutura do Método Lapelinc começou em 2008 e ainda não terminou, faltando a finalização do *software* para fazer a transcrição e a edição do texto. Todo o sistema criado na Uesb também pode ser útil em outras instituições acadêmicas e até em empresas. “Fazemos pesquisa e um apoio externo ou comercial não muda nosso trabalho, mas o protótipo pode levar a um produto, porque o método é passível de uma patente. No momento estamos finalizando seu desenvolvimento”, explica Santos. O trabalho teve financiamento da Fundação de Amparo à Pesquisa do Estado da Bahia (Fapesb), do Conselho Nacional de Pesquisa Científica e Tecnológica (CNPq) e da própria universidade. ■

Artigo

Santos, J. V. e Brito, G. S. Fotografia técnica de documentos para formação de *corpora* digitais eletrônicos: o método desenvolvido no Lapelinc. *Letras & Letras*, v. 30, n. 2, p. 421-30, jul./dez. 2014.