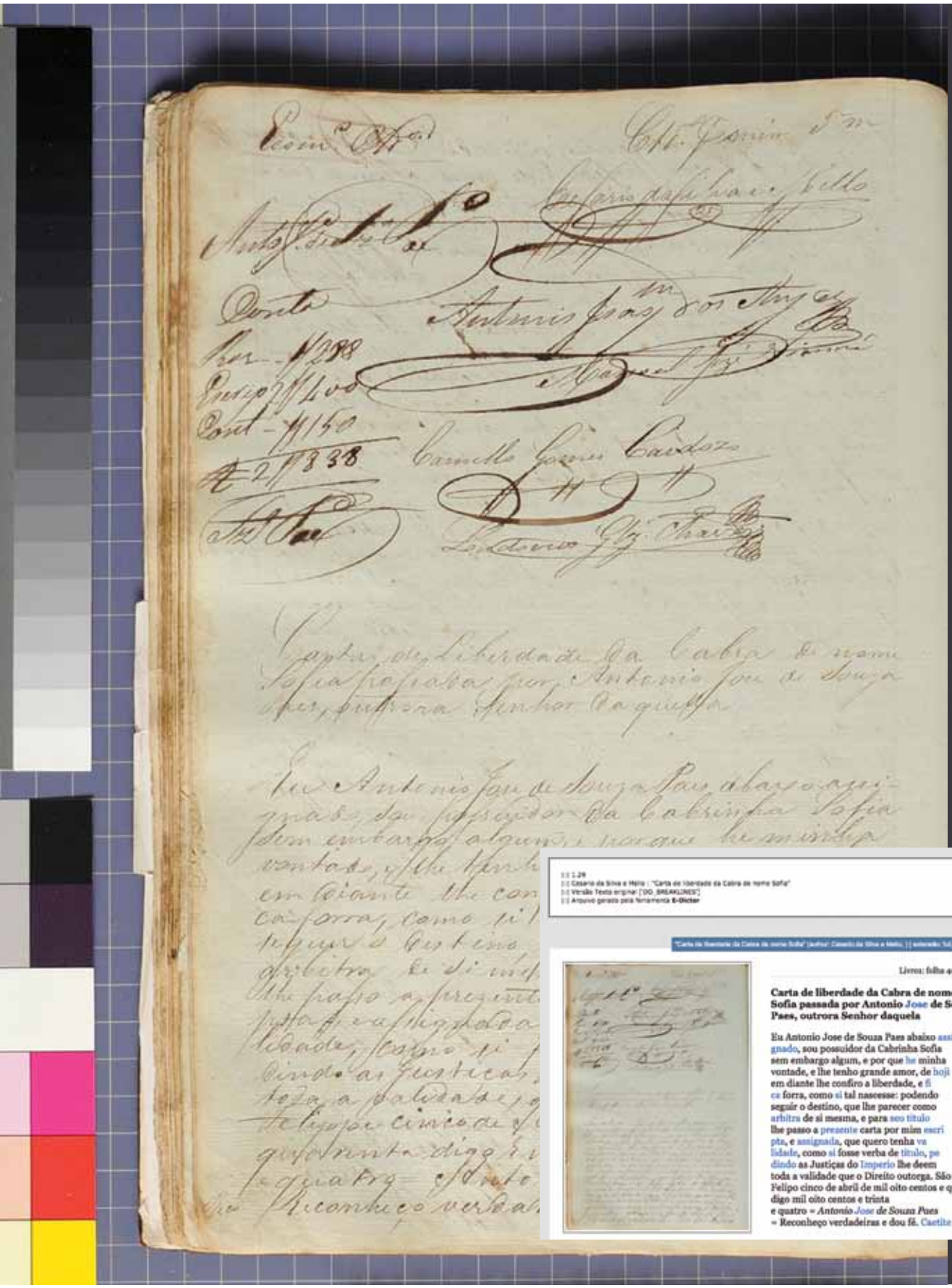

1.29
Cesario da Silva e Mello | : "Carta de liberdade da Cabra de nome Sofia"
Versão Texto original [DO_BREAKLINES]
Arquivo gerado pela ferramenta E-Dictor

"Carta de liberdade da Cabra de nome Sofia" (author: Cesario da Silva e Mello,) [extensão: 540, 440 palavras]

Livro: 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, de hoji em diante lhe confiro a liberdade, e fi ca forra, como si tal nascesse: podendo seguir o destino, que lhe parecer como arbitra de si mesma, e para seo titulo lhe passo a prezente carta por mim escri pta, e assignada, que quero tenha va lidade, como si fosse verba de titulo, pe dindo as Justiças do Imperio lhe deem toda a validade que o Direito outorga. São Felipo cinco de abril de mil oito centos e quatro digo mil oito centos e trinta e quatro = *Antonio Jose de Souza Paes* = Reconheço verdadeiras e dou fé. Caetite

A 19th century letter on the Cartesian table. Below, text transcription for the study of period Portuguese

Recording writing

A new method facilitates the transformation of historical handwritten documents into digital files

PUBLISHED IN APRIL 2015

The challenges involved with handling rare historical documents and manuscripts led a group of researchers from the State University of Southwestern Bahia (UESB) to develop a photographic method that facilitates the transcription and analysis of texts from previous eras. "There are old documents and books for which the traditional methods of obtaining an image through scanning can damage or even destroy the original, because they often require folding it or removing it from its bindings in order to place it on a scanner," says Professor Jorge Viana Santos of the UESB Corpus Linguistics Research Laboratory (Lapelinc). The researchers study 19th century official registry books and documents, texts that have been handled often over the years and are very fragile. "Unlike with photography, in scanning the document must adapt to the device, and not the opposite," he says. There is software currently available that is able to convert typed or printed text into a text file using a method called optical character recognition (OCR), which takes the scanned image of a document as its input. This cannot be done with documents written by hand.

The new method, created by Professor Santos in collaboration with fellow UESB Professor Cristiane Namiuti Tempon, begins with a photograph of the text. Before taking the photo, the document is placed on a flat sheet of gray plastic with millimeter markings that allow the computer to identify the exact measurements of the document. Color tone scales, cataloging, pagination and sequence information are also placed on this Cartesian table. The document page can be displayed on the computer with all of this information or with the handwritten part only.

DETAILS ON THE SCREEN

Through photography, software (also developed at Lapelinc) transposes the document from the physical world to the digital one. It interprets this data, recovering the colors and tones of the original document and recreating them on the computer screen. Thus, the method translates historical handwritten documents into electronic texts suitable for scientific research.

One advantage of the Lapelinc Method is the ease with which the original text can be magnified on the computer screen, allowing researchers to check details or answer questions about the writing. These digital documents can be consulted several times without damaging the historical materials. According to Santos, the new method contributes to analyses performed by paleographers or specialists who study the language of a text, transcribe it, and translate it to modern Portuguese, if necessary. Corpus linguistics (the study of electronic texts) requires that the documents studied be in text format, allowing researchers to compile corpora (the plural of corpus) for automated linguistic analysis. "Our method allows compilation of an electronic corpus, forming a database in which each word can be identified and labeled, facilitating the linguist's work when searching for the object of study; for example, nouns and verbs can be tagged," says Santos. "The historian can read the text in modern Portuguese, but the linguist wants to know how the text was written in the original language, to analyze the patterns and evolution of the language."

Development of the Lapelinc method began in 2008 and is ongoing. Text transcription and editing functions must still be incorporated into the software. The system created at UESB could also be useful to other academic institutions and even to businesses. "We do research, and external or commercial support would not change our work, but a prototype could lead to a product, since the method could result in a patent. We are currently wrapping up development," explains Santos. The project was financed by the Bahia Research Foundation (FAPESB), the National Council for Scientific and Technological Development (CNPq) and the university itself. ■ **Marcos de Oliveira**

PHOTOS LAPELINC

Article

Santos, J. V. and Brito, G. S. Fotografia técnica de documentos para formação de corpora digitais eletrônicos: o método desenvolvido no Lapelinc. **Letras & Letras**. V. 30, No. 2, p. 421-30. July/Dec. 2014.