

# Fotografia técnica de documentos para formação de *corpora* digitais eletrônicos: o método desenvolvido no Lapelinc

Technical Photography of documents for construction of digital corpora: the method developed in Lapelinc

Jorge Viana Santos\*  
Giovane Santos Brito\*\*

---

**RESUMO:** O presente artigo apresenta descritivamente etapas do método fotográfico que vem sendo desenvolvido e utilizado, desde 2008, no Lapelinc (*Laboratório de Pesquisa em Linguística de Corpus* - UESB). Nos limites deste trabalho, por recorte metodológico, destaca-se um dos instrumentos desenvolvidos especificamente para uso neste método, no processo de transposição de documentos manuscritos históricos do tipo jurídico para formação de *corpora* linguísticos eletrônicos: a *Mesa Cartesiana*.

**ABSTRACT:** This paper shows steps of Lapelinc Photographic Method, developed since 2008, at the Lapelinc (Corpus Linguistics Research Laboratory). Methodologically, we emphasize the Cartesian Table (*Mesa Cartesiana*), a instrument developed by us for material transposition of historical documents to construction of digital linguistic corpora.

**PALAVRAS-CHAVE:** Linguística de *Corpus*. Fotografia. *Corpora* Eletrônicos. Documentos Históricos.

**KEYWORDS:** Corpus Linguistics. Photography. Electronic Corpora. Historical Documents

---

## 1. Introdução

Como explicam Brito e Santos (2013), ao depararmos com documentos antigos, podemos encontrar respostas para fenômenos linguísticos complexos relacionados a períodos remotos e presentes das línguas, a exemplo do que foi feito por Santos (2008). Investigações desse tipo podem ajudar pesquisadores na busca da compreensão, descrição e explicação de fenômenos concernentes a uma língua, tanto em seus aspectos históricos e filológicos quanto no que se refere à própria gramática.

---

\* Doutor em Linguística pela Universidade de Campinas (Unicamp). Professor do Departamento de Estudos Linguísticos e Literários da Universidade Estadual do Sudoeste da Bahia (DELL/UESB). Professor do quadro permanente do Programa da Pós-Graduação em Linguística (PPGLIN/UESB). Professor colaborador do Programa de Pós-Graduação em Memória, Linguagem e Sociedade (PPGMLS/UESB). Pesquisador da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). Membro do Grupo de Pesquisa em Estudos da Linguagem (UESB/Cnpq), do Grupo de Pesquisa Lugares de Enunciação e Processos de Subjetivação (Unicamp/Cnpq), do Grupo de Pesquisa Mulheres em Discurso (Unicamp/Cnpq) e do Grupo de Pesquisa PROHPOR (UFBA/UESB/Cnpq).

\*\* Mestrando em Linguística pelo Programa da Pós-Graduação em Linguística da Universidade Estadual do Sudoeste da Bahia (PPGLIN/UESB). Pesquisador bolsista da Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB). Membro do Grupo de Pesquisa em Estudos da Linguagem (UESB/Cnpq).

Tal realidade, porém, encontra algumas dificuldades: acesso e disponibilização de documentos históricos para fins de pesquisa; e, no caso de reprodução e/ou cópia de documentos desse tipo, a garantia de fidedignidade necessária para a pesquisa científica.

Neste sentido, o presente artigo objetiva apresentar descritivamente etapas do método fotográfico quem vem sendo desenvolvido e utilizado no Lapelinc (*Laboratório de Pesquisa em Linguística de Corpus* - UESB), desde 2008, destacando especificamente, a *Mesa Cartesiana*, um dos instrumentos desenvolvidos para uso no processo de transposição, para o formato digital, de documentos manuscritos históricos do tipo jurídico para formação de *corpora* linguísticos eletrônicos.

Considerando a questão “Qual a viabilidade do uso da fotografia para a captação fidedigna de documentos para compor *corpora* digitais, visando estudos linguísticos e científicos?”, no Lapelinc postulamos por hipótese que, desde que metodicamente controlada em suas fases de captura, catalogação, edição, armazenamento, e leitura, a Fotografia apresenta-se como forma altamente viável e produtora de digitalização, permitindo à Linguística, ou outra ciência, acessar imagetivamente, de modo confiável, o documento não disponível no local da pesquisa.

Para tanto, nos limites deste trabalho, consideraremos, para comprovação, dois itens: em primeiro, lugar apresentamos as etapas do *Método Lapelinc*; em segundo, descrevemos exemplificadamente a *Mesa Cartesiana*, instrumento de fotografia, desenvolvido para tal método.

## **2. Fotografia técnica de documentos: o método desenvolvido no Lapelinc**

O *Método Lapelinc* é um método de fotografia cientificamente controlada que, desenvolvemos e temos aplicado e aperfeiçoado desde 2008 no processo de transposição<sup>1</sup> de documentos manuscritos originais em papel para o formato digital, conforme descrito nos trabalhos de Santos (2008, 2010a, 2010b, 2013a, 2013b), Namiuti, Santos e Leite (2011), Namiuti-Temponi, Santos, Costa e Farias (2013), Brito, Santos e Namiuti-Temponi (2013),

---

<sup>1</sup> Sobre o conceito de transposição, ver Santos e Namiuti (2013).

com vistas a integrar *corpora* eletrônicos anotados, a exemplo do *Corpus Dovic<sup>2</sup> Beta*. Possui três características:

- a) Pressupõe domínio da Fotografia (*Photography*) enquanto linguagem e enquanto técnica.
- b) Necessita de equipamento e aparato técnico auxiliar específicos, a exemplo da *Mesa Cartesiana* (cf. item 2).
- c) Visa à construção, sobretudo, de *corpora* manuscritos para uso científico: Linguística, História, Direito, Memória, dentre outras.

Em conjunto, com estas três características, como postulado em Namiuti-Temponi, Santos, Costa e Farias (2013, p. 12), objetivamos algo muito importante: transformar a Fotografia não num simples meio de reprodução de um documento, uma fotografia pragmática, que serve apenas a uma pesquisa e não tem compromisso de futuro, mas sim praticá-la com método científico de reprodução digital, como se vê no gráfico 1:

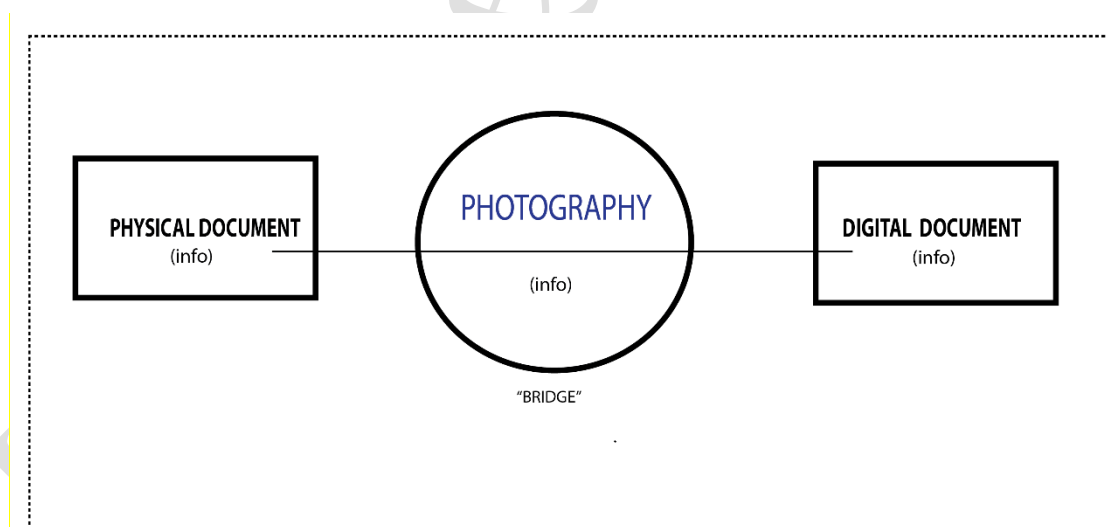


Gráfico 1: Fotografia praticada com método científico de reprodução digital: a *ponte* entre Documento Físico (DF) e Documento digital (DD)<sup>3</sup>

Fonte: *Material transposition of Documents for a digital corpus implementation* (SANTOS; NAMIUTI, 2013b)

<sup>2</sup> Dovic (Documentos Oitocentistas de Vitória da Conquista e região) (SANTOS; NAMIUTI, 2014) é um corpus eletrônico anotado, desenvolvido no âmbito do Projeto Corpora Digitais Para a História do Português Brasileiro – região Sudoeste da Bahia: Aliança PHPB – Tycho Brahe (FAPESB: 6171/2010) (SANTOS; NAMIUTI, 2010).

<sup>3</sup> Gráfico originalmente apresentado em inglês em Santos e Namiuti (2013b).

No gráfico [...], a Fotografia funciona como uma espécie de *ponte* entre DF (documento físico) e o DD (documento digital). Mas para isso, defendemos que é preciso que ela registre, na própria imagem, dados/informações que façam com que a imagem gerada não perca o vínculo com o documento que lhe deu origem (NAMIUTI-TEMPONI; SANTOS; COSTA; FARIAS, 2013, p. 15)

Tal método, em sua aplicação, apresenta cinco etapas principais:

- 1) Controle: etapa da captura de informações da fonte (por exemplo, catalogação de dados de um livro a ser fotografado);
- 2) Captura fotográfica da imagem do original: fotografia sequenciada dos documentos utilizando equipamentos adequados, inserindo na imagem a quantidade necessária de dados que garanta a sua relação com o objeto que a originou. Ou seja: fotografa-se o DF para se formar o DD;
- 3) Catalogação no *Database Dovic* das folhas-imagens componentes do documento;
- 4) Edição;
- 5) Criação de imagens de uso co-indexadas à imagem-original.

Dentre essas etapas, destaque-se que para a uso na 2, desenvolvemos no Lapelinc um equipamento com vistas a viabilizar de modo controlado e cientificamente padronizado a Fotografia: a *Mesa Cartesiana*. Vejamos.

### **3. Um instrumento do método Lapelinc: a *mesa cartesiana***

#### **3.1 O projeto**

A *Mesa Cartesiana* foi idealizada e concebida para ser um equipamento que possibilite, junto com o registro do documento, o registro visual cientificamente controlado de suas características físicas, tais como: tom, cor, tamanho, acervo/arquivo do original físico, tipologia, paginação e sequenciação.

### 3.2 O layout

No tocante ao *layout*, a fim de alcançar tal objetivo, a *Mesa Cartesiana* apresenta-se como na figura 1, em que se observam: escala de tom (1), escala de cores (2), instrumentos de medição calibrados (3, 4, 5), informações catalográficas (6), paginação (7) e sequenciação (8).



FIGURA 1: *Mesa Cartesiana* (lay out)

Fonte: *Corpus Dovic Beta*

### 3.3 O funcionamento

Assim configurada, a *Mesa Cartesiana* possui as condições técnicas necessárias para contribuir para a garantia de fidedignidade necessária à pesquisa científica. Isto porque, durante a captura fotográfica digital (cf. item 2), os elementos componentes funcionam do seguinte modo:

- a) Escala de tom (1) e escala de cores (2): sendo escalas científicas elaboradas para o controle fotográfico, possui amostras de tons e cores com parâmetros que, podem ser interpretados por programas e *softwares* de edição e leitura de imagem, capazes por isso de, por exemplo, recuperar numa tela de computador as tom/cores originais de um documento, independente da leitura que o olho humano faça.
- b) Instrumentos de medição (3, 4, 5): sendo escalas científicas elaboradas para controle milimétrico, do modo como estão dispostas,

formam um perfeito plano cartesiano, capaz de matematicamente permitir o cálculo preciso das medidas de quaisquer documentos (livros, folhas...), independente da sua posição.

c) Informações catalográficas (6), paginação (7)<sup>4</sup>, sequenciação (8)<sup>5</sup>: garantem um vínculo permanente entre o DF e o DD.

Os formatos das imagens capturadas atendem basicamente à orientação do CONARQ (Conselho Nacional de Arquivos) (CONARQ, 2010, p. 14-15), que sugere a captura de uma matriz no formato *Raw*<sup>6</sup> que, em nosso caso, tem a dimensão de 4256 x 2832 *pixels*, com profundidade de 14 bits por canal (RGB), gerando um arquivo não comprimido (*uncompressed*) de 12.1 *megapixels*. Juntamente com esse arquivo, a câmara gera um arquivo em JPEG<sup>7</sup>, com resolução mínima de 300 *dpi*. O mesmo arquivo em *Raw* possibilita, ainda, a criação de arquivos de alta resolução para armazenamento no formato TIFF<sup>8</sup>, outra recomendação do CONARQ.

Destaque-se que, por se tratar de um método que se desenvolve em função da realidade do objeto para qual se destina, e considerando que os livros notariais antigos normalmente eram numerados em apenas um lado da folha (no anverso), ficando o verso sem o respectivo número, equivalendo, pois, ao verso de mesmo número do anverso – como se vê nas figuras 2 e 3 abaixo –, após o uso de um numerador manual para indicar a paginação com codificação de cores: vermelho, para folhas frente, e verde, para folhas verso, combinado com uso de um sequenciador (em azul) para indicar a ordem da imagem no original, já está em desenvolvimento no setor de informática do Lapelinc um numerador e um sequenciador eletrônicos, para serem utilizados nas coletas a partir de 2015, com vistas a eliminar as possibilidades de erro humano.

---

<sup>4</sup> Numerador com números sequenciais que seguem a ordem de numeração das folhas (frente e verso) conforme constam no livro físico. Folhas não numeradas no original, por regra, são indicadas como “000”, caso, por exemplo, de uma capa.

<sup>5</sup> Sequenciador com números sequenciais, de 1 (capa) a “n” (verso da última capa ou folha do documento), que indicam a ordem das folhas-imagens, isto é, de cada imagem capturadas para fins de edição.

<sup>6</sup> Raw, arquivo digital com dados não processados, que é gerado pela câmara, e não pode ser diretamente manipulado, apenas possibilitando a geração de outro arquivo em formato diverso, a exemplo de JPEG ou TIFF (cf. LONG, 2004, p. 26)

<sup>7</sup> JPEG (Joint Photographic Experts Group), arquivo digital comprimido (cf. LONG, 2004, p. 27)

<sup>8</sup> TIFF, arquivo digital não comprimido, o que, em comparação, por exemplo, ao JPEG, o torna superior em qualidade de imagem (cf. LONG, 2004, p. 26).



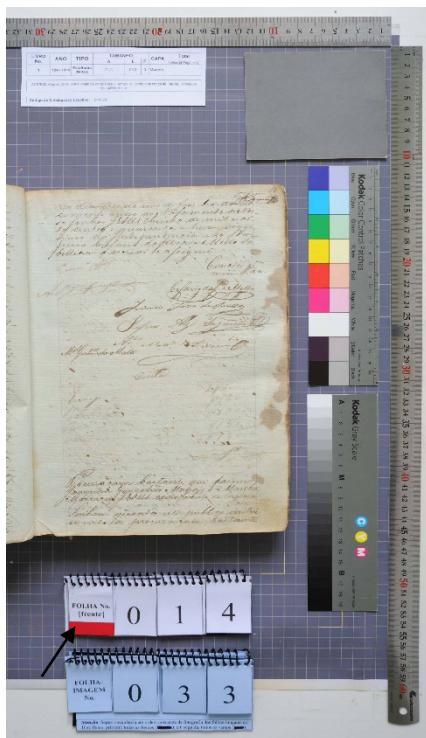


FIGURA 2: folha frente.  
Fonte: *Corpus Dovic Beta*.

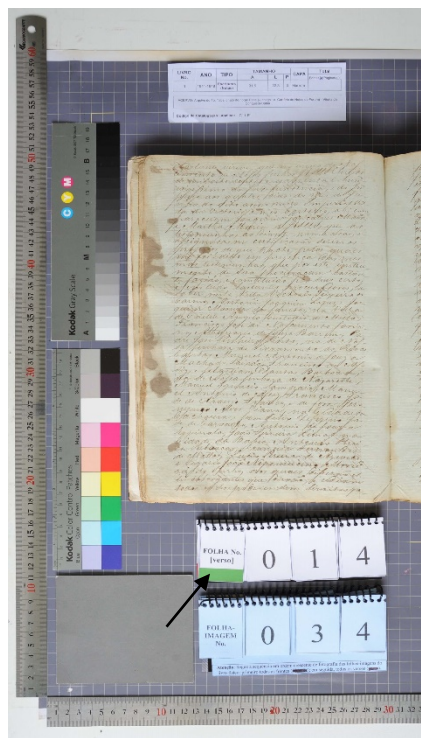


FIGURA 3: folha verso.  
Fonte: *Corpus Dovic Beta*.

Tais informações são de suma importância para o controle da edição, pois, como se vê nas figuras 2 e 3, as imagens das folhas são capturadas uma a uma, com excesso de borda suficiente para serem editadas posteriormente, quando se recortam todas as informações inseridas na *Mesa Cartesiana*, para obtermos a imagem de uso, contendo apenas o documento, como exemplificam as figuras 4 e 5:

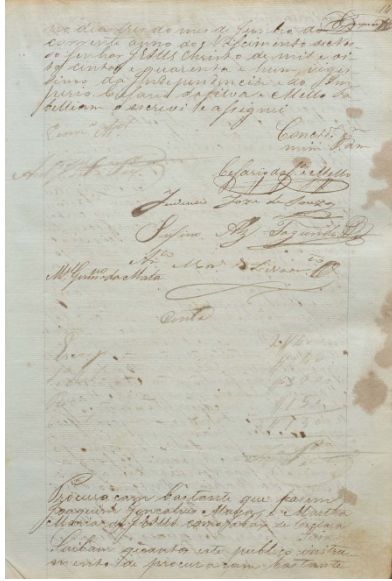


FIGURA 4: folha frente editada.  
Fonte: *Corpus Dovic Beta*.

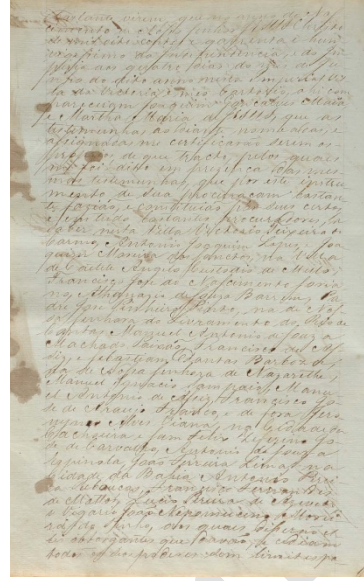


FIGURA 5: folha verso editada.  
Fonte: *Corpus Dovic Beta*.

Cabe ressaltar que o *Método Lapelinc* se propõe como método que visa à digitalização por Fotografia e não por escaneamento. Isto devido à natureza do objeto que compõe o acervo com que lidamos: livros notariais manuscritos que, dada a sua idade/datação (Século XIX) e grande tempo de manuseio (muitos ainda estão em uso), apresentam-se hoje em estado de extrema fragilidade, sendo impraticável a sua digitalização por *scanner*. O processo de escaneamento por mais moderno que seja é fixo e exige que o objeto/documento se adeque a ele em termos de tamanho, iluminação, etc. Diferentemente, a Fotografia, enquanto linguagem e enquanto técnica, dada à sua flexibilidade, devida a quase infinita possibilidade de variação de parâmetros, se adequa à realidade do objeto e não o inverso.

#### 4. Considerações finais

Pelo exposto, verifica-se que o *Método Lapelinc*, no processo de transposição de documentos manuscritos históricos para formação de *corpora* linguísticos eletrônicos, configura-se de fato como passível de transpor um documento do formato papel, um DF, para um formato digital, um DD, com fidedignidade requerida pela Ciência.

Isto porque, ao pô-lo ao adotá-lo, torna-se necessário, não só postular, mas também praticar a Fotografia de modo cientificamente controlado, utilizando equipamentos especificamente criados para esse fim, tal como a *Mesa Cartesiana*.



É, portanto, a Fotografia viabilizando a ciência, preservando a história, preservando a memória e – muito importante – contribuindo para o desenvolvimento da Linguística de *Corpus*.

## Referências

BRITO, G. S.; SANTOS, J. V. (2013). **A designação dos sobrenomes de escravos como estigma da escravidão em fontes judiciais do Sudoeste baiano**. Projeto de pesquisa. Fapesb. 2013.

BRITO, G. S.; SANTOS, J. V.; NAMIUTI-TEMPONI, C. Photograph(y) as a means for transposition of paper to digital text. Poster Session. **Workshop on construction and use of large annotated corpora**. Unicamp. Campinas: 2013.

CONARQ – Conselho Nacional de Arquivos. **Recomendações para digitalização de documentos arquivísticos permanentes**. Rio de Janeiro: Arquivo Nacional, 2010.

LONG, B. **Complete digital photography**. Hingan: Charles River Media, 2004.

NAMIUTI, C.; SANTOS, J. V.; LEITE, C. M. B. Propostas e Desafios dos Novos Meios das Antigas Fontes: A Preservação da Memória pela Linguística de Corpus. In: X Colóquio Nacional e II Colóquio Internacional do Museu Pedagógico UESB, 2011, Vitória da Conquista. **Anais do X Colóquio Nacional e II Colóquio Internacional do Museu Pedagógico UESB**. Vitória da Conquista: UESB, 2011. v. 1. p. 1-11.

NAMIUTI-TEMPONI, C.; SANTOS, J. V.; COSTA, A.; FARIAS, I. S. Computação e Linguística: importante diálogo para pesquisas e preservação da memória nos novos meios das antigas fontes. **RBBA**, v. 2, n. 1, p. 9-34, jul./2013.

SANTOS, J. V. (2008). **Liberdade na escravidão: uma abordagem semântica do conceito de liberdade em cartas de alforria**. Tese (Doutorado em Linguística) – Instituto de Estudos da Linguagem da UNICAMP, Campinas, 2008.

SANTOS, J. V. **Apresentação de meios para o transporte: digitalização de documentos manuscritos e impressos**. Conferência ministrada na I Oficina de Linguística de Corpus da Bahia (UEFS, UESB, UFBA). Feira de Santana: UEFS, 2010a.

SANTOS, J. V. **Técnicas de transporte do texto manuscrito para o meio digital**. Conferência ministrada na I Oficina de Linguística de Corpus da Bahia (UEFS, UESB, UFBA). Feira de Santana: UEFS, 2010b.

SANTOS, J. V.; NAMIUTI, C. **Corpora Digitais Para a História do Português Brasileiro - região Sudoeste da Bahia**: Aliança PHPB - Tycho Brahe. Projeto de pesquisa. UESB, Vitória da Conquista, 2010. (FAPESB: 6171/2010)

SANTOS, J. V.; NAMIUTI, C. **DOVIC (Documentos Oitocentistas de Vitória da Conquista e região)**. Corpora Eletrônico. UESB. Vitória da Conquista, 2014.

SANTOS, J. V.; NAMIUTI, C. Material transposition of Documents for a digital corpus implementation. Lecture. **Workshop on construction and use of large annotated corpora**. Unicamp. Campinas: 2013b.

SANTOS, J. V. **Um método de Fotografia técnica documental para formação de corpora digitais de documentos históricos manuscritos**. 2013. Uesb. Vitória da Conquista. 2013a. (Curso).

Artigo recebido em: 15.10.2014

Artigo aprovado em: 15.12.2014