

O Corpus M.A.P.: Desenvolvimento de um sistema de edições filológicas digitais para a documentação sobre as Mulheres na América Portuguesa

**Maria Clara Paixão de Sousa
Universidade de São Paulo**

Resumo

O objetivo desta proposta é dar início à formação do **Corpus M.A.P.**, uma reunião inédita de manuscritos de mulheres e sobre mulheres da América Portuguesa editados filologicamente com tecnologias digitais de última geração. As fontes do Corpus integram o **Catálogo M.A.P.**, repositório digital criado no âmbito do **Projeto M.A.P. (Mulheres na América Portuguesa)**. O projeto, ativo desde 2017, tem como objetivo central sistematizar e tornar visível para pesquisas futuras um conjunto de fontes documentais imensamente importantes para os estudos filológicos e para os estudos da história da língua, da história social, da história da escrita e da leitura, e da história das mulheres no Brasil, por meio da construção de um catálogo eletrônico de documentos escritos por mulheres na América Portuguesa entre 1500 e 1822. A relevância do Projeto reside fundamentalmente na possibilidade de organização inédita dessa documentação a um tempo escassa e fundamental para a compreensão da história da formação do Brasil. Presentemente, o Catálogo comporta 150 entradas detalhadamente descritas e um texto-piloto com edição filológica digital. A sistematização integral dessa documentação na forma do Corpus aqui proposto permitirá a ampliação do alcance do público leitor dos textos e a disponibilização desse material para fins de pesquisa nos campos da linguística e da ciência da computação. A proposta se filia aos trabalhos do Grupo de Pesquisas Humanidades Digitais (formado por pesquisadoras do M.A.P) junto ao Centro de Inteligência Artificial (C4AI) da Universidade de São Paulo, atualmente responsáveis pela formação do Corpus Carolina (com lançamento previsto para março de 2022).

The M.A.P. Corpus: Developing a digital edition system for documents about Women in Portuguese America

Maria Clara Paixão de Sousa
University of São Paulo

Abstract

The aim of this proposal is to initiate the formation of the **M.A.P. Corpus**, composed of manuscripts written by women and about women in Portuguese America, philologically edited with the latest digital technologies. The Corpus fonts are part of the **M.A.P. Catalog**, a digital repository created within the scope of the **M.A.P. (Women in Portuguese America) Project**. The project, active since 2017, aims to systematize and make visible for future research a set of documentary sources that are immensely important for philological studies and for studies of the history of language, social history, the history of writing and reading, and the history of women in Brazil, through the construction of an electronic catalog of documents written by women in Portuguese America between 1500 and 1822. The relevance of the Project lies fundamentally in the possibility of organizing this unprecedented documentation at a scarce and essential time for the understanding of the history of formation in Brazil. At present, the Catalog comprises 150 detailed entries and a pilot text with digital philological editing. The integral systematization of this documentation in the form of the Corpus proposed here will allow the expansion of the readership of the texts and the availability of this material for research purposes in the fields of linguistics and computer science. The proposal is part of the work of the Digital Humanities Research Group (formed by MAP researchers) at the Artificial Intelligence Center (C4AI) of the University of São Paulo, currently responsible for the formation of Corpus Carolina (scheduled for launch in March 2022).

1 Enunciado do problema

1.1 Motivações

“No teatro da memória,
as mulheres são sombras tênuas”

Michelle Perrot,
Práticas da memória feminina, 1989

Em algum dia do ano de 1592, Catarina Garcia de Cabreira escreve de Arraiolos a seu marido Antonio do Vale de Vasconcelos, em Salvador, pedindo notícias e mandando saudades, pois seus olhos “já não veem de tanto chorar” (Cabreira, 1592); em um outro dia daquele mesmo ano, na Madeira, Inês Fernandes escreve a João Gonçalves, seu primo e pai de seu filho, pedindo ajuda para ser embarcada com o menino para junto dele no Brasil, por conta das necessidades que passam, suplicando “que se alembre de uma órfã tão desamparada e de seu filho que passa muitos bocados de fome” (Fernandes, 1592). Em 24 de março de 1591, uma outra mulher escrevia a seu marido, de Oeiras a Pernambuco, pedindo provimentos para o filho e contando do “muito trabalho que tenho levado por amor de vós” – e assina: “Desta que não devera ser, Vicência Jorge” (Jorge, 1591). Em São Paulo, nos idos de 1730, Maria Clara da Anunciação escreve a seu namorado: “Sr. Antônio José, vossa mercê não me quer bem... eu quero a sua pessoa bem... peço a vossa mercê por quem é, não faça cousa que se diga cousa de menino” (Anunciação, 1730). Em 16 de março de 1765, Ana Maria Cardosa escreve ao alferes de Atibaia, Domingos Leme do Prado, pedindo que ele prenda seu pai e seu irmão, que abusam sexualmente dela e das irmãs, e que agora, ela revela, “...andam me jurando a pele” (Cardosa, 1765)¹.

Essas palavras registradas em raros exemplares de escritos feitos por mulheres ao longo dos primeiros séculos da formação da América Portuguesa chegaram até nós por diferentes acidentes históricos: as cartas de Catarina, Inês e Vicência foram preservadas como provas em processos da Inquisição de Lisboa (pois os destinatários das três missivas foram acusados e processados como bígamos); a carta de Maria Clara, como prova no processo movido contra o namorado Antônio por quebra de promessa de casamento; a de Ana Maria Cardosa, por ter chegado a uma instância importante da organização administrativa-militar da época e pela sorte de ter sido enviada a um alferes cioso de seus papéis, que legou vasta documentação preservada até hoje.

Para além da condição fortuita de terminarem inseridos nos processos do Santo Ofício ou nos maços frios da correspondência administrativa colonial, foi muito rara a preservação de documentos escritos por mulheres no reino de Portugal e na América Portuguesa ao longo do período colonial – tanto por, na maioria dos casos,

¹ Para visualizar entradas detalhadas das cinco cartas citadas, cf. <http://map.prp.usp.br/MAP=Fichas.html>.

terem feito parte das esferas não letradas e de baixa condição social, quanto por, mesmo quando letradas, terem sido impedidas de participar das relações de poder, e portanto, do espaço mais amplo da circulação da escrita.

Assim, os apelos, as súplicas, os protestos de amor e de vingança de Vicêncio, Inês, Catarina, Maria Clara e Ana Maria chegam até nós como réstias de luz que irromperam, por pequenos rasgos, o manto espesso que cobria a vida e o cotidiano das mulheres no contexto da América Portuguesa – luzes tênues lançadas sobre as sombras das mulheres *“no teatro da memória”*, a lembrarmos Michelle Perrot (Perrot, 1989).

De fato, tendo em conta o que se sabe sobre as condições de vida das mulheres no contexto colonial, e sobre seu acesso ao letramento e às instâncias públicas de expressão (como mostrado, entre outros, por Priore 1990, 1994; e Algranti, 1992, 1998), a surpresa não recai sobre a escassez de registros escritos por elas na época, mas sim sobre o fato de chegarmos a nos deparar com algum testemunho deles, séculos depois. E de fato: se a historiografia em tantos momentos se calou, os documentos, de seu lado, encerraram vozes cristalinas, ainda que enclausuradas em uma documentação opaca.

À raridade e escassez desse conjunto documental soma-se a dificuldade de sua reunião, explicada talvez pela natureza díspar que motivou o registro escrito acerca das mulheres e (mais raramente) dos documentos escritos pelos próprios punhos femininos, talvez pelo diminuto grau de interesse sobre o tema do cotidiano feminino de parte da historiografia mais tradicional. A historiografia que se debruçou sobre a história das mulheres na América Portuguesa a partir da década de 1980 bebeu em fontes primárias majoritariamente inéditas e cuja principal característica é a dispersão custodial. Organizar essa documentação para o leitor erudito e especialista é uma tarefa importante; mais importante, porém, será tornar mais transparentes as vozes ali encerradas para a leitora leiga.

É esse nosso intuito com o Projeto **M.A.P. – Mulheres na América Portuguesa**, iniciativa em curso desde 2017, e que forma o horizonte retrospectivo e prospectivo da presente proposta de trabalho, como se descreve abaixo.

1.2 Antecedentes e redes de pesquisa

1.2.1 O Projeto M.A.P

O Projeto M.A.P. vem reunindo virtualmente essa documentação dispersa em único ponto de acesso, o Catálogo eletrônico online **Mulheres na América Portuguesa**, possibilitando que as vozes relatadas presentes nas fontes primárias tornem-se vozes autorais, narradoras de suas próprias histórias.

O Catálogo contém informações arquivísticas, filológicas e temáticas sobre cada documento encontrado e um índice onomástico das mulheres escreventes e das mulheres com discurso relatado nos documentos.

A ideia de reunir documentos de mulheres e sobre mulheres forma-se por força da contingência da raridade da documentação autoral, que já comentamos; para complementá-la, buscamos e catalogamos também textos coetâneos escritos sobre mulheres. Mais especificamente, que incluem ‘falas’ de mulheres na forma de discursos relatados (tipicamente, na forma de confissões, denúncias, petições e

outros elementos componentes de processos inquisitoriais ou instrumentos administrativos), um material que, embora não traga a voz imediata das mulheres, como no caso do primeiro grupo documental, ainda assim traz elementos importantes para a compreensão e a contextualização daquele.

O Catálogo M.A.P. pretende assim compor um mapa polifônico de vozes de mulheres que escreveram no período colonial, somadas ao registro do discurso relatado de mulheres cujo comportamento, por diferentes razões, mereceu a atenção da sociedade da época – em geral, da parte das instâncias disciplinadoras da Igreja e da administração colonial. Nesse mapa importa, centralmente, a literalidade da expressão e a literalidade do relato da expressão, sendo esta uma investigação originária do campo da filologia e da linguística histórica. Assim, colocamos a fidedignidade documental como pedra de toque do trabalho, para compor um conjunto que atenda aos interesses de diferentes linhas de pesquisa, notadamente a história do cotidiano e a história das mulheres no Brasil.

O Projeto M.A.P. é pensado para um horizonte de longuíssimo prazo. Reunir e sistematizar esses documentos se configura, de fato, como a tarefa de uma vida – ou de muitas vidas. Desde seu momento inicial, em 2017, essa ideia ampla passou a se materializar em etapas sistematizadas que transformaram o impulso inicial em ações concretas. Assim, o projeto maior cresceu pela execução sucessiva de 11 projetos menores e pontuais².

Graças a essa forma de trabalho, o Catálogo M.A.P. apresenta hoje 150 entradas, sistematizadas a partir de um trabalho minucioso de recolha documental, decifração paleográfica e ordenação informática. A partir da codificação de base, diferentes visualizações dos dados são possíveis – por exemplo, em mapa georreferenciado (Figura 1) ou em entradas detalhadas (Figura 2).

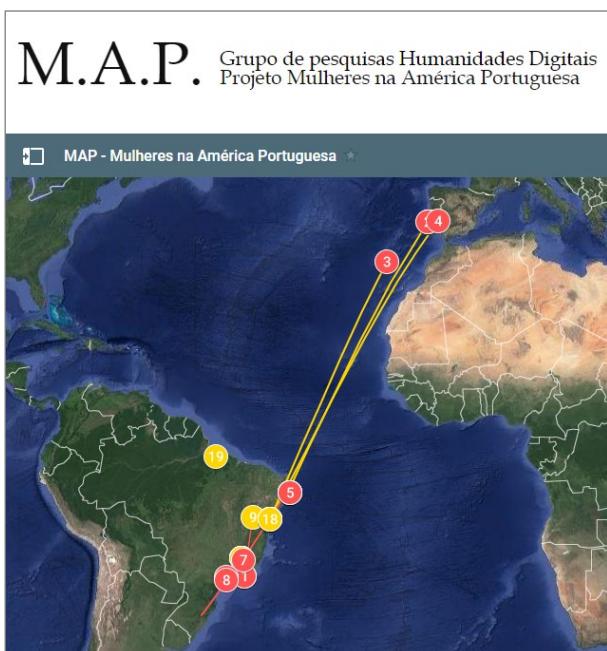


Figura 1:
Visualização do Catálogo M.A.P.:
inserção georreferenciada no Google Maps

² A lista completa dos 11 projetos, com ligações para a versão completa de cada um, está no Anexo I – Versões do Projeto M.A.P., e também em http://map.prp.usp.br/MAP_Historico.html.

Anna Maria Cardosa

agora andam me iurando a pelia por querer me matar e andam dizendo que não ei de escapar mesmo e assim quero que prendam esta gente que não tom[bem] mais para esta terra [...] são uma gente maus e assim que quero que façam o que quiserem dele que eu não posso viver mais receosa deles

De uossa merce Serua e Criada Anna [M]aria Cardosa

(1765)



Imagen: Biblioteca Nacional do Rio de Janeiro (BNJ)

Autora de carta enviada para o alferes de Atibaia, 1765

Documento acessível, digitalizado e
transcrito
Não há menção na bibliografia conhecida

Figura 2:
Visualização do Catálogo M.A.P.:
detalhe da entrada de Anna Maria Cardosa

No mesmo período, a equipe do M.A.P. cresceu da formação original, com as duas coordenadoras e quatro bolsistas de Iniciação Científica, para seu tamanho atual, comportando 24 pesquisadoras da graduação e da pós-graduação (algumas delas, bolsistas; em sua maioria, pesquisadoras voluntárias). As pesquisadoras atualmente ativas no Projeto se organizam em três '*frentes de trabalho*': a Frente de Filologia, que se ocupa mais concentradamente da prospecção documental e do trabalho de decifração paleográfica dos manuscritos; a Frente de Extroversão, que tem se dedicado aos espaços de extroversão dos resultados do projeto para um público mais amplo através de mídias digitais; e a Frente Computacional, que se dedica ao desenvolvimento das tecnologias digitais utilizadas na estruturação do Catálogo. Em 3.3 apresentamos com detalhe a equipe e seu perfil de formação.

A equipe da Frente Computacional tem realizado, mais recentemente, outras atividades de pesquisa com forte impacto no desenho da presente proposta, tal como se expõe a seguir.

1.2.2 Rede de pesquisas e Inserção institucional

O Projeto M.A.P. e a presente proposta de trabalho se filiam ao **Grupo de Pesquisas Humanidades Digitais** (humanidadesdigitais.org) e ao **Laboratório Virtual de Humanidades Digitais** (LaViHD, lavihd.fflch.usp.br), abrigados no **Núcleo de Apoio à Pesquisa em Etimologia e História da Língua Portuguesa** (NEHiLP, nehilp.prp.usp.br).

O Grupo Humanidades Digitais foi criado em 2009, inicialmente funcionando junto à Biblioteca Brasiliiana Guita e José Mindlin da Universidade de São Paulo, e desde 2016 junto ao NEHiLP. O LaViHD foi fundado mais recentemente, em 2019, em conjunto com pesquisadores da **Universidade Estadual do Sudoeste da Bahia** (UESB), e representa a consolidação das parcerias entre o grupo da USP e o **Laboratório de Pesquisas em Linguística de Corpus** (LaPeLinC) da UESB em torno da concepção e desenvolvimento de recursos digitais focalizados no estudo histórico da língua e na

curadoria digital de acervos memoriais. A reunião desses pesquisadores foi motivada por seus históricos comuns de pesquisa, descritos em 3.1.1.

O objetivo do LaViHD é desenvolver ambientes de apoio e ferramentas para acervos digitais e corpora eletrônicos – em particular, ferramentas para captura imagética de documentos, edição filológica e anotação linguística. No longo prazo, o LaViHD pretende funcionar como um polo incubador e difusor de tecnologias de processamento de texto que, a um tempo, refletem o estado da arte das tecnologias em âmbito internacional e sigam um compromisso rigoroso com as melhores práticas filológicas, como mostramos em Paixão de Sousa et al. (2021). O trabalho inicial no LaViHD teve por foco constituir um ambiente colaborativo para o aprimoramento dessas ferramentas e para a concepção e o desenvolvimento de novos instrumentos e metodologias voltados à pesquisa linguística e ao preparo filológico de corpora. Nos primeiros meses, a colaboração se deu, sobretudo, graças ao desenvolvimento do trabalho de tese de Aline Silva Costa (Costa, em curso), orientado por Cristiane Namiuiti com coorientação de Maria Clara Paixão de Sousa (um trabalho de grande impacto para esta proposta, como se descreve em 3.2.2).

Os objetivos do LaViHD receberam um impulso de grande magnitude, de fato, em meados de 2020, a partir do convite de pesquisadores do **Instituto de Matemática e Estatística da USP (IME)** para nossa colaboração em seu projeto junto ao **Centro de Inteligência Artificial da Universidade de São Paulo (C4AI)**. O trabalho no C4AI é o contexto maior que impulsiona a presente proposta, como se descreve abaixo.

1.2.3 O Corpus Carolina

O **Centro de Inteligência Artificial da Universidade de São Paulo (C4AI)**³ tem como missão produzir pesquisa avançada em Inteligência Artificial no Brasil, disseminando e debatendo os principais resultados, treinando estudantes e profissionais, e transferindo a tecnologia para a sociedade. O **Projeto de Processamento de Linguagem Natural (NLP2, <http://c4ai.inova.usp.br/pt/nlp2-pt>)** é um dos desafios do C4AI, e tem como objetivo geral desenvolver sistemas que avancem o estado da arte do Processamento de Linguagem Natural para o português brasileiro, atingindo um novo patamar em qualidade de geração e desempenho em relação ao que existe hoje. O NLP2 está atualmente construindo vários corpora, entre eles o Carolina, o CORAA, Corpus de Áudios Anotados de Português Falado e o Portinari, Corpus Anotado do Português.

A equipe do LaViHD vem trabalhando desde 2020 junto ao C4AI e é atualmente responsável pela formação do **Corpus Carolina** (sites.usp.br/corpuscarolina), um corpus do português contemporâneo para amplo uso – que inclusive servirá como uma ‘nave-mãe’ com relação aos demais corpora produzidos no C4AI-USP. O Carolina é um Corpus Aberto para Linguística e Inteligência Artificial com um volume robusto de textos de tipologia variada em português brasileiro contemporâneo (1970–2021). O corpus possui um importante diferencial: é concebido com uma metodologia original que denominamos WaC-wiPT (‘Web as Corpus com informações de Proveniência e Tipologia’). Consideramos a

³ Projeto especial da USP, IBM e FAPESP (em seu Programa de Centros de Pesquisa em Engenharia, Processo 2019/07665-4).

proveniência um aspecto crucial a se aspirar em corpora baseados na web, combinada à tipologia e ao gerenciamento de equilíbrio. Além de facilitar o cumprimento dos direitos autorais e a rotulagem tipológica, ela permite responder a perguntas sobre a origem dos textos e aumenta o escopo de uso do corpus. O trabalho inicial da equipe do Carolina buscou abordar a tipologia textual em um sentido amplo, livre de um compromisso teórico estrito, como uma ferramenta metodológica crucial no desenvolvimento de um acervo de textos de tamanho tão significativo – permitindo a organização das buscas, da seleção e do balanceamento dos textos, como mostrou preliminarmente Paixão de Sousa (2020).

A versão inicial do corpus (**Carolina Beta**), a ser lançada em março de 2022, incluirá documentos do âmbito judiciário, textos jornalísticos e wikis, entre outros, e conterá cerca 700 Giga bites de texto, o correspondente a 40% do volume de arquivos já prospectados e ainda em tratamento.

Os intensos trabalhos junto ao C4AI para a construção do Carolina tiveram consequências importantíssimas sobre a equipe computacional do M.A.P., pela oportunidade ímpar de formação e capacitação das pesquisadoras. Essa contingência deu o norte e o impulso à presente proposta, como se detalha abaixo.

1.3 Proposta e perspectivas

Graças ao ambiente colaborativo proporcionado pelos trabalhos do LaViHD no C4AI-USP, o Projeto M.A.P. chegou, em 2021, a um novo patamar em sua trajetória. Em particular, a intensificação da capacitação da equipe computacional nos últimos dois anos nos faz constatar que o momento é maduro para darmos início ao plano que sempre esteve no horizonte do Projeto: a construção do **Corpus M.A.P.**, a partir do material reunido e extensivamente trabalhado no Catálogo M.A.P.

Para concretizar essa ideia, planejamos em primeiro lugar encerrar um ciclo importante dos trabalhos: o Carolina Beta será publicado em março de 2022, e os ajustes para o crescimento do corpus devem estar estabelecidos até o fim desse primeiro semestre do ano. A partir da segunda metade de 2022, a equipe se focará na execução dos trabalhos previstos na presente Proposta.

A opção por começar o Corpus M.A.P. depois do lançamento da primeira versão do Corpus Carolina se justifica, entre outros, pela possibilidade de aproveitar importantes conhecimentos tecnológicos desenvolvidos na construção do Carolina para a construção do Corpus M.A.P. (é o caso, sobretudo, da codificação dos metadados, como se descreve em **3.2.1** à frente). Ao lado desse aproveitamento de tecnologias, há outros aspectos do trabalho cujos antecedentes vão além da parceria com o C4AI – em particular, o trabalho no Carolina tem como foco textos do português contemporâneo, enquanto o M.A.P. é composto manuscritos dos séculos XVI a XIX. Nesse sentido, lançaremos mão da ampla experiência de membros da equipe nas áreas da filologia e da construção de corpora históricos, como se descreve em **3.1.2** à frente.

Assim, a presente proposta pretende articular novas soluções metodológicas a partir dos antecedentes de pesquisa aqui descritos, buscando aproveitar as novas expertises desenvolvidas pelas pesquisadoras junto ao C4AI e a experiência da

equipe no âmbito do trabalho em corpora históricos em benefício dos objetivos de longo prazo do M.A.P.

2 Resultados esperados

A meta do projeto aqui proposto é a de publicar o **Corpus M.A.P. Beta** – versão inicial de um corpus inédito formado por manuscritos de mulheres e sobre mulheres da América Portuguesa, editados filologicamente com tecnologias digitais de última geração, e com interesse para pesquisas nos âmbitos da filologia, da história da língua, da história social, da história da escrita e da leitura, e da história das mulheres no Brasil.

Essa versão do Corpus demonstrará a aplicação das tecnologias desenvolvidas, permitirá pesquisas preliminares do próprio grupo e de outros pesquisadores, e formará a base para a expansão do Corpus nos anos seguintes, posteriores à vigência do auxílio. O Corpus M.A.P. Beta será composto por **40** documentos manuscritos editados eletronicamente, organizados em três conjuntos de acordo com suas distintas tipologias documentais de origem:

- A. Documentos do Fundo do Tribunal do Santo Ofício (TSO): **10**
- B. Documentos do Fundo do Governo da Capitania de São Paulo (GCSP): **20**
- C. Documentos do Arquivo do Instituto de Estudos Brasileiros da USP (IEB): **10**

A lista dos documentos está no **Anexo II – Documentos a serem editados**. Os manuscritos serão editados com o software para edição filológica e anotação linguística automática **eDictor**, compreendendo pelo menos duas camadas de edição (diplomática e modernizada), como se mostra em **3.1**. Entretanto, cada um desses conjuntos documentais apresenta desafios metodológicos próprios, tanto do ponto de vista computacional como filológico; assim, diferentes estratégias serão desenvolvidas para cada conjunto, como se descreve em **3.2** a seguir.

O Corpus M.A.P. Beta será o produto mais visível do projeto aqui proposto; além disso, consideramos que a metodologia a ser desenvolvida na construção desse Corpus e as oportunidades de formação e capacitação da equipe serão também resultados importantes – em particular, por virem a constituir as bases para o prosseguimento dos trabalhos do grupo no crescimento do Corpus e pelo seu potencial de transferência para a comunidade de pesquisas. Assim, pontuamos abaixo os resultados esperados nessas três vertentes:

Quadro-resumo dos resultados esperados

- | |
|--|
| <ul style="list-style-type: none">1. O Corpus M.A.P. Beta, com 40 manuscritos editados eletronicamente;2. Metodologias e ferramentas computacionais:<ul style="list-style-type: none">2.1 Sistemas de metadados para o Catálogo e Corpus M.A.P.;2.2 O eDictor 2.0;2.3 Modelos treinados para leitura automática de manuscritos (HTR);3. Formação e capacitação pessoal para a continuidade do Projeto. |
|--|

Os desafios esperados e os métodos previstos para chegarmos a esses resultados serão descritos na seção **3** a seguir.

3 Desafios, meios e métodos

3.1 Planejamento geral para o Corpus M.A.P. Beta

3.1.1 Experiência anterior da equipe e trabalhos preliminares

O tratamento dos textos no Corpus M.A.P. será desenvolvido como recursos digitais que buscam unir tecnologias atuais aos mais rigorosos princípios da prática filológica, nos moldes de Paixão de Sousa (2013) e Monte e Paixão de Sousa (2017).

As pesquisadoras doutoras e as doutorandas envolvidas na presente proposta tem ampla experiência no campo da construção de tecnologias de corpus. Entre as ferramentas e metodologias já desenvolvidas pelas pesquisadoras estão o **eDictor**, software para edição filológica e anotação linguística automática desenvolvido pela pesquisadora responsável por esta Proposta em conjunto com pesquisadores da área da computação (Paixão de Sousa, Kepler e Faria, 2010), e o **WebSync**, sistema web para corpora anotados em XML desenvolvido por Aline Silva Costa sob a orientação de Cristiane Namiuti (Namiuti, Santos e Costa, 2015; Costa, 2015). As ferramentas têm uma origem em comum: entre 1998 e 2007, a pesquisadora responsável e pesquisadora associada Cristiane Namiuti fizeram parte da equipe fundadora do **Corpus Tycho Brahe**, o primeiro corpus sintaticamente anotado do Português (cf. <http://www.tycho.iel.unicamp.br/corpus>), em um grupo de pesquisas ativo por mais de vinte anos, pioneiro no uso de ferramentas computacionais em linguística e precursor das Humanidades Digitais no Brasil (cf. Namiuti, 2004; Paixão de Sousa, 2014). A colaboração entre as docentes prosseguiu, mais recentemente, pela participação da pesquisadora responsável como colaboradora nos projetos de pesquisa do LaPeLinC (Santos e Namiuti, 2010; Namiuti, 2013; Namiuti, 2016). Por fim, a doutoranda Priscila Tuy Batista foi formada, na iniciação científica e no mestrado, junto ao **Corpus Eletrônico de Documentos Históricos do Sertão (CEDOHS)**, da Universidade Estadual de Feira de Santana (<http://www5.uefs.br/cedohs>) – o primeiro ambiente a adotar o eDictor depois do Corpus Tycho Brahe – e tem imensa experiência não só como usuária do software, mas, sobretudo, na condução de oficinas de capacitação para outros usuários.

No que toca o conhecimento e a experiência com o trabalho filológicos, a pesquisadora responsável orienta trabalhos na linha de Filologia do Português junto ao Programa de Pós-graduação em Filologia e Língua Portuguesa da Universidade de São Paulo e é uma das responsáveis pela disciplina de Filologia Portuguesa na graduação em Letras da Universidade desde 2008. A pesquisadora associada Vanessa Martins do Monte, além de orientar trabalhos na mesma linha e ser também responsável pela disciplina de Filologia Portuguesa, tem ainda ministrado diversas oficinas de paleografia na USP e a convite de outras Universidades, e apresenta publicações que a colocam entre os principais especialistas em paleografia portuguesa no Brasil (entre outras, Megale et al. 2007, Monte 2015, Monte 2021). A parceria entre as prof^{as}. Vanessa e Maria Clara na coordenação do M.A.P. tem se fundando na combinação de suas experiências no trabalho computacional e no trabalho filológico, que desejamos reproduzir na presente Proposta.

3.1.2 Edição-piloto

O planejamento para a construção do Corpus M.A.P. tem como antecedente mais concreto a publicação, em 2020, da edição do manuscrito da *'Denúncia contra Francisca Luís'*, composto no contexto da 1ª visitação do Tribunal do Santo Ofício à Bahia⁴, e um dos primeiros documentos do Catálogo M.A.P.

Essa 'edição-piloto', disponível em <http://map.prp.usp.br/Corpus/FL/FL.html>, foi realizada com o software eDictor, e envolveu algumas modificações na produção de sua visualização que também planejamos aplicar no Corpus M.A.P. Para uma descrição detalhada do sistema de edição no eDictor, remetemos a Paixão de Sousa (2014). Aqui, mostramos algumas imagens que podem ilustrar os resultados da anotação. O documento se apresenta para leitura em formato de edição diplomática (cf. Figura 3) e edição modernizada (cf. Figura 4), mantendo sempre garantido o acesso às formas originais, graças ao sistema em camadas do eDictor.



Figura 3:
Versão edição
diplomática

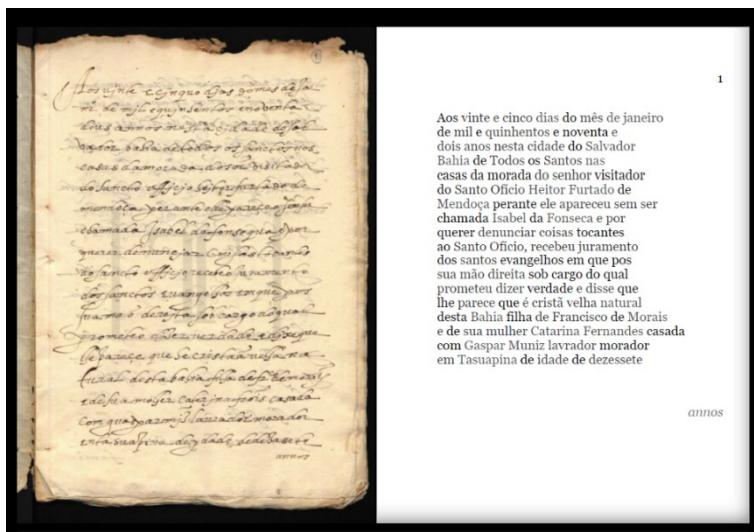


Figura 4:
Versão edição
modernizada

⁴ Tribunal do Santo Ofício (TSO). Denúncia contra Francisca Luís. Processo inquisitorial. Salvador, 1592. ANTT - TSO-IL, 13787. Disponível em: <http://digitarq.arquivos.pt/details?id=4510000>.

As duas versões foram formatadas com uma visualização melhorada com relação à proposta tradicional do eDictor (cf. Figura 5); além disso, a edição incluiu anotações de natureza lexical, não previstas no formato original da ferramenta (no caso, estão anotados todos os topônimos e antropônimos do documento).



Figura 5:
Detalhe de tela:
visualização permitida pela
edição em camadas

Como todo documento trabalhado no eDictor, a base fundadora da anotação é um documento XML no qual estão codificadas (e numeradas) todas as palavras do documento original – e que contém, além disso, a anotação da classe morfológica de cada palavra (Figura 6). A anotação da edição é semiautomática; a anotação das classes de palavras é inteiramente automatizada na ferramenta.

```

▼<p id="p_5">
  ▼<s id="s_5">
    ▼<w id="11">
      <o>Aos</o>
      <m v="P+D-P"/>
    </w>
    ▼<w id="12">
      <o>ujnte</o>
      <e t="gra">vinte</e>
      <m v="NUM"/>
    </w>
    ▼<w id="13">
      <o>&lt;/o>
      <e t="gra">e</e>
      <m v="CONJ"/>
    </w>
    ▼<w id="14">
      <o>cinqno</o>
      <e t="mod">cinco</e>
      <e t="gra">cinquo</e>
      <m v="NUM"/>
    </w>
    ▼<w id="15" t="conjectura">
      <o>dj[a]s</o>
      <e t="gra">dias</e>
      <e t="conj">djas</e>
      <e t="dip">dj[a]s</e>
      <m v="N-P"/>
      <ref refnum="000" pos="N-P" posnum="000" id="ref_0"/>
    </w>
    ▼<w id="16">
      <o>do</o>
      <m v="P+D"/>
    </w>
    ▼<w id="17">
      <o>mes</o>
      <e t="mod">mês</e>
      <m v="N"/>
      <ref refnum="001" pos="N" posnum="000" id="ref_0"/>
    </w>

```

Figura 6:
XML de base,
parágrafo 5, palavras 11 a 17 –
'Aos vinte e cinco dias do mês'.

Legenda:

Marcação das palavras:
categoria 'w'

Indicação dos termos originais:
categoria 'o'

Anotações de edição:
categoria 'e'

Anotações de classe de palavra:
categoria 'm'

A edição-piloto da *Denúncia de Francisca Luís* apresenta já diversas características importantes para os objetivos do Projeto M.A.P. até este momento – entretanto, sabemos, por nossa experiência na área, que a edição dos manuscritos no âmbito geral do Catálogo M.A.P. pode ser muito mais interessante e enriquecedora dos pontos de vista filológico, linguístico e computacional.

Em particular quanto ao aspecto computacional, salientamos que, como já mencionado em 1.2.3, a parceria com o C4AI-USP na construção do Corpus Carolina proporcionou excepcionais oportunidades de formação às pesquisadoras, graças à proximidade com esse ambiente científico e tecnológico de altíssimo nível. Os progressos alcançados pelas estudantes foram notáveis nesse último ano, e elas se encontram hoje plenamente capacitadas para trabalhar com autonomia e independência com tecnologias de ponta. Em combinação com a parceria com a equipe do LaPeLInC (que atualmente está desenvolvendo o eDector 2.0, como se descreve em 3.2.2 a seguir), esse novo patamar de capacitação nos permite prever que o sistema de edição dos textos no novo Corpus será tecnologicamente mais avançado que a edição-piloto.

Assim, a edição-piloto configura, para essa proposta, uma inspiração importante no sentido de um desenvolvimento mais aprofundado, graças à melhoria dos métodos e ferramentas utilizadas, como se expõe abaixo.

3.2 Desenvolvimento de métodos e ferramentas computacionais

3.2.1 Sistemas de metadados para o Catálogo e Corpus M.A.P.

A primeira e mais importante tarefa na construção de um Corpus eletrônico é a formulação de um sistema de metadados robusto, flexível e padronizado. A experiência de construção do Corpus Carolina mostrou essa importância com muita clareza para a equipe: a etapa de discussão, formulação e implementação dos metadados (na forma de um cabeçalho padronizado de acordo com as normas TEI., cf TEI Consortium, 2021) ocupou parte importante do trabalho em 2020 e 2021. Nesse processo, desenvolvemos ideias e aprendemos técnicas fundamentais que desejamos transferir para a formação do Corpus M.A.P.

A estrutura atual do Catálogo M.A.P. já representa efetivamente os metadados de um futuro Corpus. Esse sempre foi o espírito por trás da formação de um catálogo com o grau de detalhamento do M.A.P. – presentemente, cada uma das 150 entradas apresenta 39 categorias descritivas, pertinentes a informações arquivísticas, filológicas, temáticas e textuais de cada documento. Como o Catálogo M.A.P. foi construído em XML, a ideia inicial era a de que cada entrada fosse aproveitada como o cabeçalho do documento do Corpus, que contivesse cada manuscrito editado. Entretanto, nossa experiência com o Carolina mostra que será desejável que, nesse aproveitamento, as categorias M.A.P. sejam adaptadas para o vocabulário geral do padrão TEI.

A adaptação das categorias do Catálogo atual para um formato de metadados TEI formará o modelo para os cabeçalhos do Corpus – e, dialeticamente, nos obrigará a reescrever o próprio Catálogo. Essa contingência é bem-vinda, já que, desde os primeiros meses de 2021, a Frente Computacional do Projeto M.A.P. vem desenvolvendo as sementes de um novo formato para a codificação do Catálogo, com novas formas de inserção de dados e categorias descritivas aprofundadas, de um ponto de vista temático, e mais robustas, de um ponto de vista computacional. A essa nova versão, já preliminarmente desenhada, mas ainda não implementada, denominamos o ‘Catálogo M.A.P. 2.0’.

Prevemos, assim, iniciar a construção do Corpus pelo desenvolvimento desse sistema de metadados avançado. Como a equipe prevista para a tarefa é composta (entre outras) pelas quatro pesquisadoras que desenvolveram o cabeçalho Carolina, acreditamos que esse trabalho será muito produtivo. A meta, nesse caso, é que o 'Catálogo M.A.P. 2.0' (ou seja, o novo formato do Catálogo e, consequentemente, o modelo de cabeçalhos do Corpus) esteja pronto dentro do primeiro semestre da vigência da presente proposta. A partir de então, este modelo será usado nos textos que virão a compor o Corpus.

3.2.2 Desenvolvimento e implementação do eDictor 2.0

No que toca à edição filológica dos textos, seguirá o modelo rigoroso da crítica textual, e ainda incluirá as anotações diplomática e linguística. Esse trabalho será feito no software eDictor, com diversas melhorias técnicas planejadas.

O eDictor foi o primeiro software para edições filológicas eletrônicas inteiramente desenvolvido no Brasil (já a partir de Paixão de Sousa e Kepler, 2007), e é atualmente utilizado por diversos projetos de pesquisa no Brasil e em Portugal (cf. <https://edictor.net>, e Paixão de Sousa 2013, 2019). O software apresenta inúmeras funcionalidades interessantes – entretanto, tomando-se em conta sua idade, ele precisa ser atualizado e reformulado em diversos pontos; algumas limitações conhecidas da versão 1.0 têm sido apontadas desde Paixão de Sousa (2015), e remetem, sobretudo, ao módulo de transcrição do software (Figura 7).

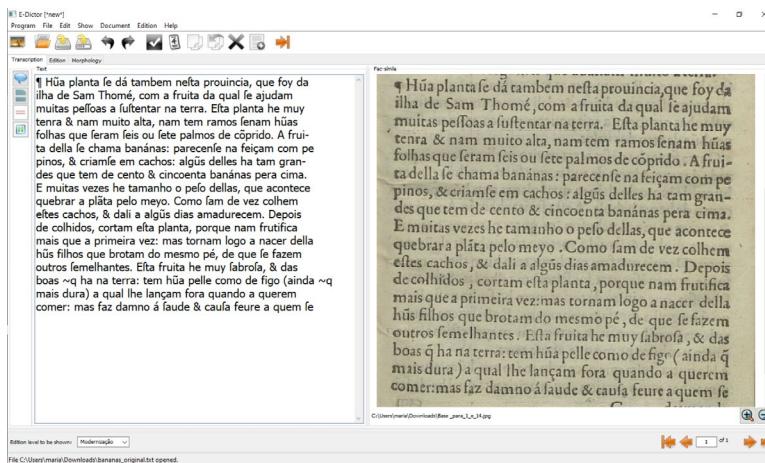


Figura 7 - eDictor 1.0 - Módulo de transcrição filológica

O necessário desenvolvimento do eDictor 2.0 foi iniciado pelo trabalho de tese de Aline Silva Costa (Costa, 2019; Costa, em curso), orientado por Cristiane Namiuti com coorientação da pesquisadora responsável (e autora da primeira versão do software). O eDictor 2.0 será inteiramente baseado na Web, e conseguirá, a um tempo, sanar as limitações conhecidas da versão atual – por exemplo, ao prever um módulo de transcrição consideravelmente mais avançado – e preservar seus pontos fortes, como a flexibilidade na anotação de edição e a precisão da anotação morfológica automática.

No espaço da presente Proposta, pretendemos intensificar a consultoria e o apoio a Aline Costa no desenvolvimento do eDictor 2.0. Além disso, com o progresso de Costa (em curso), passaremos a nos dedicar a etapas de implementação e concretização dos resultados da nova ferramenta nos moldes desejados pelo Corpus M.A.P. Nesse sentido, importa notar que uma característica central do

eDictor é seu potencial notável de customização pelos usuários finais. Assim, a partir do desenvolvimento da nova versão, a equipe do Corpus deverá dedicar tempo à implementação de mudanças que se adequem aos nossos objetivos.

Importa também, por fim, destacar que o eDictor, nas duas versões, é um software livre de código aberto. Dessa forma, assim como a versão 1.0, a versão 2.0 ficará inteiramente disponível para a comunidade de pesquisas.

3.2.3 Treinamento experimental de softwares para leitura automática de manuscritos (HTR)

Uma importe vertente do trabalho de construção do Corpus M.A.P. será o desenvolvimento de modelos para leitura automática dos manuscritos, por meio do treinamento de softwares com a tecnologia **Handwritten Text Recognition (HTR)**.

Ao contrário do trabalho com o eDictor, esta vertente representa, para a equipe do Projeto, uma seara inédita de pesquisas. Apenas em 2021, depois de muitos anos de experiência na área da edição filológica eletrônica, passamos a ter um contato mais sistematizado com ferramentas de HTR – e, a partir disso, decidimos lançar mão delas, ainda que de modo experimental, para a construção do Corpus M.A.P. Selecionamos preliminarmente duas ferramentas para treinamento:

i. Transkribus, software desenvolvido pelo consórcio ReadCoop, sediado na Universidade de Innsbruck, Áustria (<https://readcoop.eu/transkribus>). Em novembro de 2021, a equipe do Projeto M.A.P. realizou uma Oficina de aprendizagem com o Transkribus, oferecida pela Dr.^a Lúcia Xavier, membro do Consórcio ReadCoop, e pela Prof.^a Dr.^a Lívia Magalhães, da Universidade Federal da Bahia. Com a oficina, ficaram evidentes as potencialidades do software para a leitura ‘automática’ de alguns dos manuscritos do Catálogo M.A.P. – em particular, os documentos inquisitoriais, que têm em comum o fato de terem sido produzidos por um mesmo punho (o do notário Manoel Francisco). No treinamento preliminar realizado na Oficina, a análise feita pela ferramenta frente a um manuscrito experimental foi bastante razoável. Isso levou a nossa equipe a dar início a uma parceria com a Dra. Lúcia Xavier (que, depois da oficina, foi convidada a fazer parte de nosso Grupo de Pesquisas) a partir de 2022, no sentido de construir um ‘modelo’ para o Transkribus, inteiramente treinado com documentos da 1^a visitação do Santo Ofício à Bahia.

ii. LaPeLinC Transcriptor, software para a transcrição paleográfica de documentos históricos desenvolvido por Bruno Silvério Costa (Costa, Costa, Santos e Namiuti, 2021). O software desenvolvido no contexto do LaPeLinC-UESB funciona a partir de uma interface que permite a visualização e manipulação dos fac-símiles, (com a ampliação/redução de áreas selecionadas, o fatiamento da imagem em linhas de transcrição, o alinhamento entre texto transscrito e imagem, entre outros), oferece um abecedário que permite a consulta da forma de escrita e a representação das ocorrências no original, de forma fidedigna, relacionando imagem e transcrição para texto em subscrito, sobrescrito, símbolos especiais, abreviaturas e nomes. O software inclui ainda uma interface de busca a dicionário de abreviaturas de Borges Nunes, permitindo buscar, visualizar e escolher formas de grafia e transcrição para as abreviaturas catalogadas. Em dezembro de 2021, a equipe do Projeto M.A.P. reuniu-se com os desenvolvedores do LaPeLinC Transcriptor, e avaliou que o software seria também uma opção extremamente interessante

para o uso no Corpus. A reunião levou à ideia de desenvolvermos um treinamento nessa ferramenta para a transcrição automática do nosso conjunto de documentos do Fundo do Governo da Capitania de São Paulo referentes ao século XIX (que são bastante próximos aos documentos do projeto em que se origina o Transcriptor).

Tendo em vista essas investigações preliminares, o plano da equipe é treinar o Transkribus para o conjunto documental **A-TSO** e o LaPeLinC Transcriptor para o conjunto documental **B-SGCSP**. O conjunto **C-IEB** será transscrito sem auxílio de ferramentas de HTR. Teremos, dessa forma, três grupos trabalhados com métodos diferentes, e resultados que podem ser comparados para pesquisas futuras.

3.3 Gestão de projeto e formação continuada da equipe

3.3.1 Histórico do modo de trabalho no Projeto M.A.P.

Uma marca importante do Projeto M.A.P. tem sido a qualidade da formação acadêmica e da capacitação profissional proporcionada às mais de 25 estudantes que estiveram ligadas ao Projeto entre 2017 e 2021. Algumas dessas estudantes vêm percorrendo trajetórias de formação completas dentro do Projeto, da iniciação científica ao mestrado, com planos de prosseguimento para o doutorado. Essa capacitação de longo prazo, sistemática e continuada contempla tanto as estudantes dedicadas centralmente a uma formação filológica rigorosa como aquelas que passaram a se especializar na área computacional. Esse último horizonte de formação, menos usual na área de Letras, tem inclusive proporcionado oportunidades profissionais importantes para algumas das estudantes graduadas (duas delas, neste momento, estão empregadas no setor de tecnologia da informação em boas empresas do mercado).

A qualidade da formação das estudantes vem sendo favorecida por um modo de trabalho essencialmente coletivo, fundado em sessões de trabalho de dois tipos: reuniões semanais plenárias, dedicadas alternadamente à discussão da bibliografia e ao andamento dos trabalhos gerais, com a presença de todas as pesquisadoras; e sessões semanais das chamadas 'Frentes' (pequenos subgrupos com afinidades temáticas), conduzidas como momentos de trabalho coletivo na construção do Catálogo.

As sessões dos dois tipos, de cerca de quatro horas por semana cada uma, aconteceram ininterruptamente, entre os meses de fevereiro e novembro, desde 2017 (em forma remota durante os anos de 2020 e 2021). A partir de setembro de 2020, oito das estudantes da Frente computacional passaram a trabalhar também na construção do Corpus Carolina, como descrito em **1.2.3**. Nesses casos, as alunas participam de reuniões em conjunto com a equipe do LaViHD-C4AI (semanalmente, por duas horas), além de dedicarem cerca de oito horas de trabalho semanal ao Carolina. Essa carga de trabalho aumentada foi compensada pela excelente oportunidade de formação técnica das estudantes, que passaram a dominar a linguagem Python (graças a um excelente curso intensivo oferecido pelos pesquisadores do IME); três dessas alunas foram ainda contempladas com bolsas TT-II ligadas ao C4AI.

Para a execução da presente proposta, planejamos prosseguir e intensificar esse modo de trabalho, que julgamos extremamente produtivo e bem-sucedido.

3.3.2 Descrição resumida da equipe

A equipe prevista para os trabalhos diretamente ligados a esta proposta será composta por pesquisadoras do Projeto M.A.P. e do LaViHD.

Entre as pesquisadoras do M.A.P., além das duas coordenadoras do Projeto (uma delas a responsável pela presente proposta), prevemos a participação de uma pesquisadora doutora bolsista e de dezessete estudantes com tarefas sistemáticas e bem-definidas, sendo quatro bolsistas e treze não-bolsistas (estudantes voluntárias ou que recebem bolsas de outras origens). O M.A.P conta ainda com outras seis estudantes que terão participação voluntária nos trabalhos sem horas definidas. O desenho da equipe foi realizado com base na experiência de trabalho das coordenadoras do M.A.P. com as estudantes, e tomando em conta o seu perfil atual no que toca à capacitação técnica e o modo de trabalho no projeto: as estudantes pré-selecionadas para o trabalho sistemático com horas dedicadas a esta proposta participam ativamente das Frentes de filologia e computacional do M.A.P., cujo foco, agora, passará a ser o trabalho no Corpus. Além dessas pesquisadoras da equipe principal, o Projeto M.A.P. tem contado com a colaboração voluntária de um pesquisador com formação em computação, cujos trabalhos serão fundamentais para a implementação do Corpus conforme planejado nesta proposta. Por seu perfil de formação, planeja-se que esse pesquisador seja um dos bolsistas de Treinamento Técnico a participar do Projeto.

Por fim, duas pesquisadoras da UESB ligadas ao LaViHD, com ampla experiência no desenvolvimento de ferramentas para corpora, atuarão como consultoras.

A intensidade da dedicação das participantes será portanto variada, de acordo com seus perfis e os objetivos da pesquisa, conforme se detalha no documento **“Descrição das atividades desenvolvidas pela equipe”**. Uma lista simples da equipe está também no **Anexo III. Visão geral da equipe**. Abaixo, apresenta-se um plano resumido do trabalho no projeto.

3.3.3 Plano de trabalho, gestão e formação continuada

1. Gestão de projeto. Ao longo da vigência da proposta, a proponente e as pesquisadoras doutoras se reunirão regularmente para avaliar o andamento dos trabalhos, de duas formas:

- (1.1) reuniões semanais das coordenadoras do M.A.P., com cerca de duas horas;
- (1.2) reuniões semestrais de todas as pesquisadoras doutoras (incluindo as colaboradoras da UESB), em seminários intensivos de cinco horas.

2. Formação e capacitação. Ao longo da vigência da proposta, prosseguiremos fomentando o ambiente coletivo de reflexão e execução de tarefas que é parte essencial do modo de trabalho no M.A.P. As atividades nesse âmbito incluem:

- (2.1) reuniões semanais de acompanhamento;
- (2.2) estudo bibliográfico;
- (2.3) redação coletiva de dois artigos (sendo a redação dos artigos e o estudo da bibliografia acompanhados nas reuniões regulares);
- (2.4) três oficinas de paleografia;

- (2.5) três oficinas de tecnologias de texto (sendo as oficinas momentos de trabalho mais intensivos e dedicados a problemas específicos a serem selecionados pelas coordenadoras);
- (2.6) quatro seminários internos, nos quais as estudantes de iniciação, mestrado e doutorado apresentam suas pesquisas em andamento.

3. Trabalho no catálogo M.A.P. Ao longo da vigência da proposta, o Catálogo M.A.P. precisará ser mantido e alimentado. Esse trabalho não afetará diretamente a construção do Corpus M.A.P. Beta (cujos documentos já estão catalogados), mas seus resultados serão essenciais para o crescimento do Corpus.

4. Pesquisa documental. A exemplo do trabalho de manutenção do Catálogo, a prospecção arquivística é fundamental para o crescimento futuro do Corpus. A prospecção intensa realizada até 2019 no Projeto foi impactada negativamente pelo fechamento dos arquivos físicos durante a pandemia da Covid-19; assim, planejamos intensificar a prospecção em duas vertentes:

- (4.1) visitas periódicas ao Arquivo Público do Estado de São Paulo.
- (4.2) três visitas a arquivos nos estados do Rio de Janeiro, Minas Gerais e Bahia.

5. Trabalho filológico. O trabalho da Frente de filologia do M.A.P. será dedicado à edição dos manuscritos para o novo Corpus ao longo da vigência da proposta. A frente se reunirá em sessões semanais de quatro horas para o trabalho coletivo de planejamento e resolução de problemas. As tarefas incluem:

- (5.1) a transcrição de manuscritos com edição semiautomática com o eDictor;
- (5.2) o treinamento dos modelos de HTR.

6. Trabalho de desenvolvimento tecnológico. O trabalho da Frente computacional do M.A.P. será dedicado à construção e implementação de metodologias e ferramentas para o novo Corpus ao longo da vigência da proposta. A Frente se reunirá em sessões semanais de quatro horas para o trabalho coletivo de planejamento e resolução de problemas, e em sessões quinzenais com as pesquisadoras da UESB, especificamente para discutir a evolução do eDictor 2.0. As tarefas incluem:

- (6.1) desenvolvimento do 'Catálogo M.A.P. 2.0' (i.e., cabeçalhos para o Corpus);
- (6.2) desenvolvimento e implementação do eDictor 2.0;
- (6.3) preparo computacional dos arquivos de base;
- (6.4) preparo do funcionamento em rede;
- (6.5) construção da Interface final do Corpus.

O Cronograma de execução de cada atividade pontuada entre 1 e 6 acima, bem como a previsão da entrega de seus produtos, é apresentado em **4** a seguir.

4 Cronograma

2022					2023												2024
agosto	setembro	outubro	novembro	dezembro	janeiro	fevereiro	março	abril	maio	junho	julho	agosto	setembro	outubro	novembro	dezembro	janeiro

1 Gestão de projeto

- 1.1 Reuniões semanais da coordenação
- 1.2 Reuniões semestrais das associadas

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

2 Formação e capacitação

- 2.1 Reuniões semanais de
- 2.2 Estudo bibliográfico
- 2.3 Redação coletiva de artigos
- 2.4 Oficinas de paleografia
- 2.5 Oficinas de tecnologias de texto
- 2.6 Seminários internos

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
	✓	✓	✓	✓													
		✓	✓	✓													
			✓	✓													
				✓													

3 Trabalho no catálogo M.A.P.

- 3.1 Alimentação regular do catálogo

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

4 Pesquisa documental

- 4.1 Pesquisa digital e local
- 4.2 Pesquisa em arquivos fora do estado

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

5 Trabalho filológico

- 5.1 Leitura e edição de manuscritos
- 5.2 Treinamento de modelos de HTR

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

6 Desenvolvimento tecnológico

- 6.1 Desenvolvimento do Catálogo M.A.P.
- 6.2 Desenvolvimento do e-Dictor 2.0
- 6.3 Preparo dos arquivos de base
- 6.4 Preparo do funcionamento em rede
- 6.5 Preparo da interface final

✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

7 Finalização de produtos

- 7.1 Lançamento do Catálogo M.A.P. 2.0
- 7.2 Lançamento público do e-Dictor 2.0
- 7.3 Lançamento público do Corpus M.A.P.
- 7.5 Submissão de artigos a periódico
- 7.6 Relatório final

5 Disseminação e avaliação

A pesquisadora responsável por esta proposta construiu e mantém os websites do NEHiLP, do Grupo de Pesquisas Humanidades Digitais, do LaViHD e do Projeto M.A.P., com o objetivo de publicar, divulgar e disseminar os resultados das pesquisas. A equipe do Projeto M.A.P., além disso, mantém canais de divulgação nas mídias sociais e um espaço de blogagem acadêmica na plataforma francesa Hypotheses:

Sites

NEHiLP: nehilp.prp.usp.br

Grupo de Pesquisas Humanidades Digitais: humanidadesdigitais.org

LaViHD: lavihd.fflch.usp.br

Projeto M.A.P.: map.prp.usp.br

Projeto M.A.P. – inglês: mapusp.wordpress.com

M.A.P. Blog: mapusp.hypotheses.org

eDictor: edictor.net

Redes sociais do M.A.P. e do NEHiLP

www.instagram.com/map_usp (1.920 seguidores)

www.instagram.com/nehilp_usp (674 seguidores)

www.facebook.com/projetomap (1.035 seguidores)

twitter.com/map_usp (375 seguidores)

Para a disseminação dos resultados da pesquisa da presente proposta, o mesmo espírito de atenção à divulgação científica no meio digital será mantido – inclusive com as redes sociais ligadas ao M.A.P. e ao NEHiLP fazendo a divulgação dos trabalhos para o público amplo.

Ressalte-se, ainda, que, nos últimos meses, tem havido um interesse considerável da mídia em torno do Projeto M.A.P., com reportagens e entrevistas veiculadas em importantes jornais e canais de televisão (as matérias recentes estão listadas no endereço de clipping do Projeto, map.prp.usp.br/midia). É provável que a formação do corpus intensifique esse interesse da mídia, já que os documentos estarão integralmente disponíveis para livre acesso com linguagem atualizada; a equipe seguirá seu compromisso em atender o grande público, o que é compatível com nossos objetivos de disseminação do conhecimento científico.

Na esfera acadêmica, a equipe pretende submeter ao menos dois artigos científicos a periódicos de alto renome na área de linguística de corpus, humanidades digitais ou filologia: um primeiro artigo, no início do projeto, relatando o desenvolvimento e instauração do Catálogo M.A.P. 2.0; e um segundo artigo, ao final do Projeto, relatando o desenvolvimento do Corpus M.A.P.

Finalmente, a principal forma de disseminação da pesquisa será a publicação do próprio Corpus M.A.P., em julho de 2024, com acesso aberto e gratuito. O Corpus será hospedado no site do Projeto M.A.P., cujo domínio pertence à Universidade de São Paulo, e para cuja manutenção o Projeto conta com uma máquina virtual dedicada do sistema Internuvens da USP, com 1 TB de espaço.

6 Outros apoios

As sucessivas versões do Projeto M.A.P. desde 2017 (detalhadas no **Anexo I – Versões do Projeto M.A.P.**) foram financiadas sobretudo com bolsas de Iniciação Científica, em quatro ciclos de fomento do **Programa Unificado de Bolsas da Universidade de São Paulo (o PUB)**: entre setembro de 2017 e agosto de 2018, com 4 bolsas; entre setembro de 2018 e agosto de 2019, com 3 bolsas; entre setembro de 2019 e agosto de 2020 com 2 bolsas; entre setembro de 2019 e agosto de 2022, com 4 bolsas. Em 2018, foi outorgado o Projeto M.A.P. (Mulheres na América Portuguesa): Mapeamento

de escritos de mulheres e sobre mulheres no espaço atlântico português a partir de métodos das Humanidades Digitais, na **Chamada Universal MCTIC/CNPq** n.º **28/2018**. O valor outorgado foi R\$ 13.600,00, para bolsas de IC. Até o momento, nenhuma bolsa foi implementada. Presentemente, a equipe do M.A.P. conta com cinco bolsas de treinamento técnico, mestrado e doutorado, da FUSP, do CNPq (cota do Programa de Pós-graduação em Filologia e Língua Portuguesa) e da Fapesp:

1. Bolsa de doutorado CNPq Priscila Tuy Batista (até jun/2022)
2. Bolsa de mestrado CAPES-DS, Adriana dos Santos Silva (até out/2023)
3. Bolsa de mestrado Fapesp, Beatriz de Freitas Cardenete (até out/2022)
4. Bolsa de mestrado FUSP, Mariana Lourenço Sturzeneker (até dez/2022)
5. Bolsa TT2 FUSP Maria Clara Ramos Morales Crespo (até dez/2022)
6. Bolsa TT2 FUSP Maria Lina de Souza Jeannine Rocha (até dez/2022)

Nesta Proposta, solicitamos 7 bolsas de Treinamento Técnico, conforme se detalha no documento **“Descrição das atividades desenvolvidas pela equipe”**. Durante a vigência do fomento, pretendemos buscar outros auxílios na forma de bolsas individuais que viabilizem a continuidade de algumas das participantes do Projeto que não apresentam perfil de bolsistas de Treinamento Técnico.

7 Bibliografia

- Algranti, Leila Mezan. Famílias e vida doméstica. In: LM e Souza, org. História da vida privada no Brasil, v. 1, Cotidiano e vida privada na América Portuguesa, p. 83-154. São Paulo: Cia. das Letras; 1998.
- Algranti, Leila Mezan. Honradas e devotas: mulheres da Colônia: estudos sobre a condição feminina através dos conventos e recolhimentos do sudeste, 1750-1822. Tese de doutoramento, Universidade de São Paulo; 1992.
- Anunciação, Maria Clara da. Carta pessoal. São Paulo, 1730. ACM-SP/PS - Arquivo da Cúria Metropolitana PGA-100/ Projeto P.S., PSCRI1741. [AA|007|MCA].
- Cabreira, Catarina Garcia de. Carta pessoal . Vila Viçosa, 1592 . ANTT/PS - TSO-IL, 1476 / Projeto P.S., PSCRI143. [AA|006|CGC].
- Cardosa, Anna Maria. Carta pessoal . Atibaia, 1765 . Biblioteca Nacional do Rio de Janeiro (BNJ) - Biblioteca Nacional do Rio de Janeiro, Coleção Morgado de Mateus, Documentos Avulsos. Cota: I 30, 21, 25. [AA|008|AMC].
- Costa, AS. Um sistema de anotação de múltiplas camadas para o corpus digital DOViC [tese]. Tese de doutorado, Departamento de Estudos Linguísticos e Literários, Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista; em curso.
- Costa, AS. WebSinC: Uma Ferramenta Web para buscas sintáticas e morfossintáticas em corpora anotados - Estudo de Caso do Corpus DOViC – Bahia. Dissertação de Mestrado, Departamento de Estudos Linguísticos e Literários, Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista; 2015.
- Costa, B; Costa, A; Santos, JV; Namiuti, C. Lapelinc Transcriptor: Um software para a transcrição paleográfica de documentos históricos. Comunicação ao V Congresso Internacional de Linguística Histórica – CILH. Unicamp, julho de 2021.
- Fernandes, Inês. Carta pessoal . Madeira, 1592 . ANTT/PS - TSO-IL, 2555 / Projeto P.S., PS2517. [AA|005|IF].
- Jorge, Vicência. Carta pessoal . Oeiras, 1591 . ANTT/PS - TSO-IL, 10755 / Projeto P.S., CARDS2253 . [AA|003|VJ].
- Megale, H; Toledo Neto, SA; Monte, VM do; Fachin, PRM; Costa, RF. Crítica textual: análise grafemática e pesquisa lingüística. VEREDAS. REVISTA DA ASSOCIAÇÃO INTERNACIONAL DE LUSITANISTAS, v. 8, p. 127-146, 2007.

Monte, VM do. Correspondências paulistas: as formas de tratamento em cartas de circulação pública (1765-1775). 1. ed. São Paulo: Humanitas – FAPESP, 2015. v. 1. 520p

Monte, VM do. História da Paleografia: movimento entre a ciência e a arte. In: Alícia Duhá Lose; Lívia Borges Souza Magalhães; Vanilda Salignac Mazzoni. (Org.). Paleografia e suas interfaces – volume 2. Salvador: Memória & Arte, 2021, v. 2, p. 36-62.

Monte, VM do; Paixão de Sousa, MC. Por uma filologia virtual: O caso das atas da câmara de São Paulo (1562-1596). REVISTA DA ABRALIN, v. 16, p. 239-264, 2017. <http://revistas.ufpr.br/abralin/article/view/51938>

Namiuti, C. Do português pré-clássico às variantes modernas: contribuições para o estudo da sintaxe e suas interfaces. Projeto de pesquisa. Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB 007/2016) / Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista; 2016.

Namiuti, C. Memória conquistense: implementação de um corpus digital. Projeto de pesquisa. Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 485098/2013-0) / Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista; 2013.

Paixão de Sousa, MC. A Filologia Digital em Língua Portuguesa: Alguns caminhos. In: Ana Paula Banza & Maria Filomena Gonçalves (coord.), Património textual e humanidades digitais: da antiga à nova Filologia. Évora: CIDEHUS/FCT; 2013. <http://dspace.uevora.pt/rdpc/bitstream/10174/10468/1/e-book.pdf>

Paixão de Sousa, MC. eDictor: A chronology. Apresentação na mesa redonda “eDictor: advances and perspectives”. Workshop Construction and use of large annotated corpora. Campinas; 9 de setembro de 2013: <http://manualedictor.wordpress.com/historico>

Paixão de Sousa, MC. e-Dictor: Histórico e perspectivas. Comunicação ao Gallaecia: III Congresso Internacional de Linguística Histórica. Santiago de Compostela; 28 de julho 2015. <http://www.slideshare.net/mariaclaraps/edictor-histrico-e-perspectivas-2015>

Paixão de Sousa, MC. O Corpus Tycho Brahe: contribuições para as humanidades digitais no Brasil. Filologia e Linguística Portuguesa, v. 16, p. 53-93, 2014. <http://www.revistas.usp.br/flp/article/view/88404>

Paixão de Sousa, MC; Kepler, FN. E-Dictor: Uma ferramenta integrada para a anotação de edição e classe de palavras. VI Encontro de Linguística de Corpus, São Paulo; 2007.

Paixão de Sousa, MC; Kepler, FN; Faria, PPF de. e-Dictor. Versão 1.0 beta 10, 2013. Programa de Computador. <https://edictor.net/download>

Paixão de Sousa, MC; Kepler, FN; Faria, PPF de. E-Dictor: Novas perspectivas na codificação e edição de corpora de textos históricos. In: Shepherd T; Sardinha T; Pinto MV., organizadores. Caminhos da linguística de corpus. Campinas: Mercado de Letras; 2010.

Paixão de Sousa, MC; Monte, VM do; Namiuti, C. Carolina: Context, development and current state of na Open Corpus for Linguistics and Artificial Intelligence. Slovenia-Brazil Webinar In Artificial Intelligence. IRCAI – International Research Centre on Artificial Intelligence, United Nations Educational, Scientific and Cultural Organization; 8 de junho de 2021.

Paixão de Sousa, MC; Monte, VM do; Namiuti, C; Santos, JV. LaViHD: Fundamentos para um laboratório de pesquisa dedicado à interoperabilidade e reuso de tecnologias para curadoria de acervos memoriais e edições filológicas digitais. II Congresso Internacional em Humanidades Digitais – HDRio. Rio de Janeiro, 14 de abril de 2021.

Perrot, Michelle. Práticas da memória feminina. Revista Brasileira de História, v. 9, n. 18, p. 9-18; 1989.

Priore, Mary del. A Mulher na história do Brasil. São Paulo: Contexto; 1994.

Priore, Mary del. Ao sul do corpo: condição feminina, maternidades e mentalidades no Brasil Colônia. Tese de doutoramento, Universidade de São Paulo; 1990.

Santos, JV; Namiuti, C. Corpora digitais para a história do Português Brasileiro – Documentos históricos da região sudoeste da Bahia: Aliança PHPP-Tycho Brahe. Projeto de pesquisa. Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB 6171/2010) / Universidade Estadual do Sudoeste da Bahia. Vitória da Conquista; 2010.

TEI Consortium, eds. TEI P5: Guidelines for Electronic Text Encoding and Interchange. [Version 4.3.0.]. [Last updated 31st August 2021]. TEI Consortium. <http://www.tei-c.org/Guidelines/P5>.

Anexos

I. Versões do Projeto M.A.P.

2017

1. Paixão de Sousa, Maria Clara; Monte, Vanessa Martins do. 'Agora andam me jurando a pele': Escritos de mulheres e escritos sobre mulheres na América Portuguesa. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2017. Disponível em: http://map.prp.usp.br/MAP_Projeto_2017.html

2018

2. Monte, Vanessa Martins do. M.A.P. (Mulheres na América Portuguesa): Mapeamento digital de escritos de mulheres e sobre mulheres no espaço atlântico português – Vertente filológica. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2018. Disponível em: http://map.prp.usp.br/Projs/MAP_2018_2019_PUB_VMM.pdf
3. Paixão de Sousa, Maria Clara. M.A.P. (Mulheres na América Portuguesa): Mapeamento digital de escritos de mulheres e sobre mulheres no espaço atlântico português. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2018. Disponível em: http://map.prp.usp.br/Projs/MAP_2018_2019_PUB_MCPS.pdf
4. Paixão de Sousa, MC. M.A.P. (Mulheres na América Portuguesa): Mapeamento de escritos de mulheres e sobre mulheres no espaço atlântico português a partir de métodos das Humanidades Digitais. Projeto de pesquisa, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Edital Universal 2018) / Universidade de São Paulo. São Paulo; 2018. Disponível em: http://map.prp.usp.br/Projs/MAP_2018_EditalUniversalCnpq.pdf

2019

5. Monte, Vanessa Martins do. M.A.P. (Mulheres na América Portuguesa): o garimpar de escritos de mulheres e sobre mulheres no espaço atlântico português. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2019. Disponível em: http://map.prp.usp.br/Projs/MAP_2019_2020_PUB_VMM.pdf
6. Paixão de Sousa, Maria Clara. M.A.P. (Mulheres na América Portuguesa): Mapeamento de escritos de mulheres no espaço atlântico português a partir das Humanidades Digitais. Projeto de pesquisa, Programa Unificado de Bolsas

(PUB) / Universidade de São Paulo. São Paulo; 2019. Disponível em:
http://map.prp.usp.br/Projs/MAP_2019_2020_PUB_MCPS.pdf

2020

7. Monte, Vanessa Martins do. A filologia como janela para o ensino: um experimento paleográfico a partir do acervo do Projeto M.A.P. (Mulheres na América Portuguesa). Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2020. Disponível em:
http://map.prp.usp.br/Projs/MAP_2020_2021_PUB_ENSINO_VMM.pdf
8. Monte, Vanessa Martins do. Da academia à sociedade: o exitoso uso das mídias digitais como ferramentas de extroversão. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2020. Disponível em:
http://map.prp.usp.br/Projs/MAP_2020_2021_PUB_CULTEXT_VMM.pdf
9. Monte, Vanessa Martins do. Filologia no M.A.P. (Mulheres na América Portuguesa): o garimpar de escritos de mulheres e sobre mulheres no espaço atlântico português, Fase II. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2020. Disponível em:
http://map.prp.usp.br/Projs/MAP_2020_2021_PUB_PESQUISA_VMM.pdf
10. Paixão de Sousa, Maria Clara. A Filologia como janela para o ensino: um experimento digital a partir do acervo do Projeto M.A.P. (Mulheres na América Portuguesa). Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2020. Disponível em:
http://map.prp.usp.br/Projs/MAP_2020_2021_PUB_ENSINO_MCPS.pdf
11. Paixão de Sousa, Maria Clara. M.A.P. (Mulheres na América Portuguesa): Mapeamento de escritos de mulheres no espaço atlântico português a partir das Humanidades Digitais, Fase II. Projeto de pesquisa, Programa Unificado de Bolsas (PUB) / Universidade de São Paulo. São Paulo; 2020. Disponível em:
http://map.prp.usp.br/Projs/MAP_2020_2021_PUB_PESQUISA_MCPS.pdf

II. Documentos a serem editados

A. Documentos do Fundo do Tribunal do Santo Ofício

1. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1591. ANTT - PT/TT/TSO-IL/028/01267. Disponível em: <http://digitarq.arquivos.pt/details?id=2301154>. Nomeada no documento: Filipa de Sousa, [0029].
2. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1592. ANTT - TSO-IL, 13787. Disponível em: <http://digitarq.arquivos.pt/details?id=4510000>. Nomeada no documento: Francisca Luís, [0031]. (edição piloto já publicada)
3. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1592. ANTT - PT/TT/TSO-IL/028/01275. Disponível em: <http://digitarq.arquivos.pt/details?id=2301163>. Nomeada no documento: Guiomar Piçarra, [0035].
4. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1591. ANTT - PT/TT/TSO-IL/028/10748. Disponível em: <https://digitarq.arquivos.pt/details?id=2310925>. Nomeada no documento: Maria Gonçalves Cajada, [0056].
5. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1593. ANTT - PT/TT/TSO-IL/028/10754. Disponível em: <http://digitarq.arquivos.pt/details?id=2310931>. Nomeada no documento: Maria Álvares, [0062].
6. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1591. ANTT - PT/TT/TSO-IL/028/03306. Disponível em: <http://digitarq.arquivos.pt/details?id=2303255>. Nomeada no documento: Paula de Sequeira, [0066].
7. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1592. ANTT - PT/TT/TSO-IL/028/10753. Disponível em: <http://digitarq.arquivos.pt/details?id=2310930>. Nomeada no documento: Maria Pinheira, [0057].
8. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1591. ANTT - PT/TT/TSO-IL/028/01273. Disponível em: <https://digitarq.arquivos.pt/details?id=2301160>. Nomeada no documento: Guiomar Lopes, [0034].
9. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1591. ANTT - PT/TT/TSO-IL/028/01279. Disponível em: <https://digitarq.arquivos.pt/details?id=2301167>. Nomeada no documento: Antônia de Barros, [0015].
10. Tribunal do Santo Ofício (TSO). Processo inquisitorial. Salvador, 1593. ANTT - PT/TT/TSO-IL/028/01289. Disponível em: <http://digitarq.arquivos.pt/details?id=2301177>. Nomeada no documento: Catarina Quaresma, [0019].

B. Documentos do Fundo do Governo da Capitania de São Paulo

1. Secretaria de Governo da Capitania (SGO). Requerimento. Vila Rica, Minas Gerais, Brasil, 1734. APM - SG-CX.26-DOC.34. Disponível em: <http://www.siaapm.cultura.mg.gov.br/modules/brtacervo/brtacervo.php?cid=1830>. Autora indireta catalogada: Anastácia da Conceição, [0077].
2. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1819. APESP - 1.1.613/93-3-74. Não disponível em repositórios digitais. Autora indireta catalogada: Ângela Maria de Jesus, [0078].
3. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1819. APESP - 1.1.0613/93-3-73. Não disponível em repositórios digitais. Autora indireta catalogada: Ana Borges, [0079].
4. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. Distrito de Parnaíba, 1816. APESP - 1.1.0612/93-3-15. Não disponível em repositórios digitais. Autora indireta catalogada: Ana Caetana de Jesus, [0080].
5. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1819. APESP - 1.1.0613/93-3-80. Não disponível em repositórios digitais. Autora indireta catalogada: Ana Francisca, [0081].
6. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1822. APESP - 1.1.624/95-3-23. Não disponível em repositórios digitais. Autora indireta catalogada: Ana Maria da Candelária, [0082].
7. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1819. APESP - 1.1.0613/93-3-69. Não disponível em repositórios digitais. Autora indireta catalogada: Ana de Godoy Lima, [0083].
8. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1819. APESP - 1.1.0613/93-3-67. Não disponível em repositórios digitais. Autora indireta catalogada: Antônia de Almeida, [0084].
9. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1819. APESP - 1.1.619/94-2-13. Não disponível em repositórios digitais. Autora indireta catalogada: Antônia Maria de Araújo, [0085].
10. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1817. APESP - 1.1.0613/93-3-44. Não disponível em repositórios digitais. Autora indireta catalogada: Bárbara Maria, [0087].
11. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, sem data. APESP - 1.1.609/93-1-32. Não disponível em repositórios digitais. Autora indireta catalogada: Caterina Dias de Sene, [0088].
12. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1819. APESP - 1.1.0614/93-4-25. Não disponível em repositórios digitais. Autora indireta catalogada: Custódia Maria de Anunciação, [0089].
13. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. Lages, 1801. APESP - 1.1.609/93-1-37. Não disponível em repositórios digitais. Autora indireta catalogada: Custódia Maria Rodrigues, [0090].
14. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1787. APESP - 1.1.601/92-1-25. Não disponível em repositórios digitais. Autora indireta catalogada: Escolástica Francisca, [0091].

15. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. Sem indicação de local do documento. Sem data. APESP - 1.1.697/15. Não disponível em repositórios digitais. Autora indireta catalogada: Escolástica Maria Leonor de Souza França, [0092].
16. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1805. APESP - 1.1.608/93-1-9. Não disponível em repositórios digitais. Autora indireta catalogada: Feliciana Coelha, [0093].
17. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1819. APESP - 1.1.0614/93-4-7. Não disponível em repositórios digitais. Autora indireta catalogada: Floriana Maria Jacinta, [0095].
18. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1810. APESP - 1.1.0611/93-2-74. Não disponível em repositórios digitais. Autora indireta catalogada: Francisca das Chagas, [0096].
19. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, Brasil, 1819. APESP - 1.1.614/93-4-6. Não disponível em repositórios digitais. Autora indireta catalogada: Gertrudes Luiza, [0097].
20. Secretaria de Governo da Capitania de São Paulo (SeGovSP). Requerimento. São Paulo, 1819. APESP - 1.1.0614/93-4-5. Não disponível em repositórios digitais. Autoria indireta catalogada: Gertrudes Maria da Conceição, [0098].

C. Documentos do Instituto de Estudos Brasileiros da Universidade de São Paulo

1. Carta de Ana Teresa para João Antônio de Mendonça, 1794. IEB/USP. Coleção Alberto Lamego, Cota AL-099-033.
2. Carta de Ana Teresa para João Antônio de Mendonça, 1797. IEB/USP. Coleção Alberto Lamego, Cota AL-099-032.
3. Carta de Ana Teresa para João Antônio de Mendonça, 1798. IEB/USP. Coleção Alberto Lamego, Cota AL-099-034.
4. Carta de Ana Teresa para João Antônio de Mendonça, 1798. IEB/USP. Coleção Alberto Lamego, Cota AL-099-007.
5. Carta de Ana Teresa para João Antônio de Mendonça, 1798. IEB/USP. Coleção Alberto Lamego, Cota AL-099-008.
6. Carta de Ana Teresa para João Antônio de Mendonça, s/d. IEB/USP. Coleção Alberto Lamego, Cota AL-099-031.
7. Carta de Ana Vicêncio de Almeida para Antônio Prado, 1815. IEB/USP. Coleção Veridiana Prado, Cota VP-40.
8. Carta de Violante Clara para Excelentíssimo Senhor, 1737. IEB/USP. Coleção Alberto Lamego, Cota AL-050-008.
9. Carta de Violante Clara para Excelentíssimo Senhor, 1737. IEB/USP. Coleção Alberto Lamego, Cota AL-050-007.

Obs.: Um décimo documento do Arquivo do IEB a ser incluído no Corpus será selecionado pela pesquisadora Beatriz Cardenete, cujo trabalho de mestrado inclui a prospecção desse arquivo, entre outros.

II. Visão geral da equipe

Pesquisadora Responsável

1. Prof^a. Dr^a. Maria Clara Paixão de Sousa (USP)

Pesquisadoras Associadas

2. Prof^a. Dr^a. Vanessa Martins do Monte (USP)
3. Prof^a. Dr^a. Cristiane Namiuti (UESB), consultora

Pesquisadora doutora bolsista

4. Priscila Starline Estrela Tuy Batista (defesa prevista para agosto de 2022)

Pesquisadora doutoranda consultora

5. Aline Silva Costa (defesa prevista para 2023)

Pesquisadores bolsistas de Treinamento Técnico - 4

6. Igor Leal (mestrando, defesa prevista para fevereiro de 2022)
7. *Bolsista a definir posteriormente a partir de edital público*

Estudantes bolsistas de Treinamento Técnico - 3

8. Elisa Hardt Leitão Motta (mestranda)
9. Andrea Cristina Natanael da Silva (graduada)

Estudante bolsista de Treinamento Técnico - 2

10. Natalia Zacchi (graduada)

Estudante bolsista de Treinamento Técnico - 1

11. Maria Eduarda Martins Mendes Cordeiro (graduada)

Estudantes com até 20 horas de dedicação

- Bolsistas com fomentos de outras origens:

12. Adriana dos Santos Silva (Bolsa de mestrado CAPES-DS, até out/2023)
13. Beatriz de Freitas Cardenete (Bolsa de mestrado Fapesp, até out/2022)
14. Mariana Lourenço Sturzeneker (Bolsa de mestrado FUSP, até dez/2022)
15. Maria Clara Ramos Morales Crespo (Bolsa TT2 FUSP, até dez/2022)
16. Maria Lina de Souza Jeannine Rocha (Bolsa TT2 FUSP, até dez/2022)

- Estudantes sem bolsas:

17. Raquel de Paula Guets (mestranda)
18. Renata Morais Mesquita (mestranda)
19. Sofia Tonoli Maniezzo Zani (mestranda)
20. Mayara Feliciano Palma (graduada)
21. Giovanna Poloni (graduada, sem bolsa)
22. Mariana Marques da Silva (graduada)
23. Patrícia Brasil Silva (graduada, sem bolsa)

Estudantes sem horas semanais determinadas

24. Gabriele Franco (doutoranda)
25. Ana Carolina Estremadoiro Prudente do Amaral (mestre)
26. Nicóllia de Lima Garcia (mestranda)
27. Carla Angelino Di Lorenzo Midões de Mello (graduada)
28. Laryssa Albino Bezerra Rogério de Oliveira (graduada)
29. Mariana Rodrigues de Vita (graduada)