

Para um hiperdicionário de Epidemiologia Histórica Luso-Brasileira: do século 18 aos dias de hoje

For a hyperdictionary of Luso-Brazilian Historical Epidemiology: from
the 18th century to today

Maria José Bocorny Finatto* 

RESUMO: O artigo relata partes do estudo e da sistematização do vocabulário de um *corpus* digital composto por obras médicas de valor histórico, impressas em português no século 18, associadas ao tema das doenças e seus tratamentos. Esse *corpus* pode interessar à história da Medicina em português e contém dados aproveitáveis para uma base de conhecimento e um hiperdicionário de epidemiologia histórica luso-brasileira. Com técnicas e métodos de Lexicometria, os dados extraídos do *corpus* são cotejados com diferentes registros especializados da atualidade. Ao reunirem-se diferentes terminologias, usos linguísticos e conceitos históricos do passado com o presente, torna-se possível desenhar um futuro hiperdicionário *on-line*, pleno de *hyperlinks*. Esse hiperdicionário, descrito e exemplificado no artigo, oferecerá uma linha de tempo de saberes e conhecimentos antigos associados a uma série de registros da atualidade.

PALAVRAS-CHAVE: Terminologia Histórica. Vocabulário da Medicina. Epidemiologia Histórica. Português. Hiperdicionário.

ABSTRACT: The article reports on parts of the study and systematization of the vocabulary of a digital corpus composed of medical works of historical value, printed in Portuguese in the 18th century, associated with the theme of diseases and their treatments. This corpus may be of interest to the history of Medicine in Portuguese and contains data suitable for a knowledge base and a hyperdictionary of Luso-Brazilian historical epidemiology. Using techniques and methods of Lexicometry, the data extracted from the corpus are linked to different specialized current records. By bringing together different terminologies, linguistic usages, and historical concepts from the past with the present, it becomes possible to design a future online hyperdictionary, full of hyperlinks. This hyperdictionary, described and exemplified in the article, will offer a timeline of ancient knowledge and insights associated with a series of contemporary records.

KEYWORDS: Historical Terminology. Medical Vocabulary. Historical Epidemiology. 18th Century. Hyperdictionary.

* Doutora em Letras, Universidade Federal do Rio Grande do Sul (PPG-LETRAS-UFRGS).
maria.finatto@gmail.com

1 Introdução

Hoje em dia, com a digitalização dos acervos de várias bibliotecas físicas, acessa-se uma série de obras antigas, manuscritas ou impressas, em diferentes tipos de formatos de arquivo, em plataformas e bases de dados. Assim, ficou muito mais fácil consultar obras e documentos que antes somente estavam disponíveis em seções de acervos raros, devidamente protegidos. Nesse cenário, dinamizado pela “relativa modernidade” de uma área conhecida como *Humanidades Digitais* (Finatto, 2023), temos encontrado, sob a ótica da Linguística Aplicada, da Lexicologia, Lexicografia e da Terminologia, estudos dedicados a textos de temática médica, sejam impressos ou manuscritos, disponíveis em português, produzidos em diferentes séculos.

Entre alguns trabalhos pioneiros com obras de temática médica, no âmbito dos estudos lexicais, podemos citar o de Murakawa (2002). Essa autora dedicou-se ao estudo do vocabulário das plantas e das drogas no século 16, conforme registrado por Garcia D’Orta (1501-1568), médico e botânico português. Mais tarde, o trabalho de Barros (2011) já nos trazia um estudo sobre o provável primeiro dicionário de termos médicos escrito em português, produzido pelo padre Pero de Castilho da Companhia de Jesus, em 1613. E, entre os mais recentes, o estudo de Domladovac Silva e Murakawa (2021) realiza um levantamento do léxico temático no domínio da cura das enfermidades que atingiam a população no Brasil colonial, com referência ao século 18. Essas autoras cotejaram informações do Banco de Dados do DHPB - Dicionário Histórico do Português do Brasil (<https://dicionarios.fclar.unesp.br/dhpb/>) - séculos 16, 17 e 18 - com diferentes registros lexicográficos.

Desse modo, em diferentes perspectivas, a história da linguagem médica em português, conforme registrada por seus práticos e agentes formadores, foi atraindo o interesse de diferentes estudos linguísticos de enfoque lexicológico no Brasil e em Portugal. Esses trabalhos se ocuparam de variados aspectos, períodos temporais e utilizaram diferentes referenciais teóricos e metodológicos. Felizmente, ainda testemunhamos a produtividade desse rico filão de estudos, por exemplo, nos

seguintes trabalhos: de Murakawa e Nadin (2020), tratando dos séculos 16, 17 e 18; no artigo de Gonçalves (2022), dedicado a uma obra pioneira da Enfermagem lusófona do século 18; e no trabalho de Finatto, Gonçalves e Lazzari (2023), que buscou contrastar, com recursos de Lexicometria, o vocabulário de uma obra de Medicina do século 18 e esse registro inaugural da Enfermagem em português. Além disso, vale salientar as importantes reflexões teórico- metodológicas de Curti-Contessoto (2024), autora que se debruça sobre o lugar e os rumos da pesquisa diacrônica ou sócio-histórica em Terminologia.

Todavia, ainda que exista a digitalização prévia de obras médicas antigas e que tal digitalização seja um bem e necessidade inquestionável, cumpre-se todo um longo e difícil percurso até chegar-se à organização e compartilhamento de um *corpus* em formato de arquivos de texto editável. Há uma longa jornada até que se obtenha algo com que possa trabalhar e explorar em suas "palavras" (ou *lexias*, significados e conceitos, o que não será discutido aqui). Isto é, desde o acesso inicial a um manual médico em um catálogo, acervo, arquivo ou base bibliográfica, em formato físico ou digital, enfrentam-se várias dificuldades. São dificuldades que podem começar já pela identificação e categorização eficiente dos nomes de temas e dos assuntos tratados em catálogos *on-line* e que podem prosseguir pelo reconhecimento do modo de indexação e das tipologias aplicadas a diferentes naturezas de documentos antigos. Fora isso, após finalmente chegar-se a uma dada obra ou obras de interesse, a devida interpretação dos seus conteúdos requererá a consulta a uma série de materiais de apoio, edições filológicas e obras de referência, como dicionários antigos, *corpora* históricos e bases de dados diversas. Nessa trajetória, o encontro de uma buscada obra médica antiga pode ser comparado à descoberta de uma ilha em meio a um vasto arquipélago, cujo desenho o pesquisador terá que distinguir e situar, em meio a um oceano de informações, correlações e condições sócio-históricas circundantes ao documento.

Vencida a etapa inicial, a pessoa interessada em reconstituir ou depreender a informação contida nos documentos, precisará conseguir contextualizar conceitual e linguisticamente a informação que conseguir ler. Será vital também recorrer à história dos conhecimentos médicos em português, a partir de um dado período. Assim, com essas fontes em mãos (ou à sua tela), o árduo trabalho do pesquisador tende a seguir exitoso. Portanto, a informação que o investigador acessa via texto escrito - seja na imagem de um fac-símile, em um arquivo PDF ou em um simples arquivo de texto sem formatação - precisará ser enriquecida com uma série de insumos de pesquisa. Felizmente, uma série de edições filológicas têm cumprido o importante papel de oferecer esse inestimável apoio em forma de insumos e de referências para facilitar a compreensão da leitura de textos antigos transcritos, com a sua devida contextualização. Essas edições são fruto de um trabalho hercúleo, hoje cada vez mais reconhecido e valorizado, associado à Paleografia, Ecdótica e à Filologia Digital (Lose; Magalhães, 2017).

Nessa linha, vale mencionar um trabalho dos mais recentes, dedicado a um manuscrito que, justamente, arrola o acervo das teses históricas da Faculdade de Medicina da Bahia, reunidas em um volume que cobre o período de 1836 até 1888. Esse trabalho, além de resgatar o próprio livro físico – muito deteriorado – e apresentar todo o conteúdo do volume em edição semidiplomática, traz uma série de imagens, descrições e comentários (Mazzoni; Lose *et al.*, 2023). São informações sobre as dissertações e teses e respectivas sessões de defesas – aprovadas e reprovadas - dos trabalhos daqueles homens e daquelas poucas mulheres que buscaram o título de médico ou cirurgião naquele espaço-tempo. A edição e os devidos cuidados com esse volume sobre as teses médicas da Bahia nos ajudam a resgatar as pessoas e as ideias que fizeram a história da Medicina do Brasil¹.

¹ Esse acervo de teses médicas da Universidade da Bahia, produzidas no séc. 19, conforme se verá mais adiante, será uma das vinculações do nosso futuro hiperdicionário de epidemiologia histórica.

Feita esta introdução, na sequência, este artigo, primeiro, relata o estado atual de uma experiência em curso, que inventaria o vocabulário sobre doenças e seus tratamentos conforme empregado em manuais médicos de valor histórico, impressos em português (Lazzari; Finatto, 2023). Não são apresentados hipóteses, experimentos ou testagens, visto que se trata apenas de um relato de pesquisa em andamento e de uma ideia base para um recurso dicionarístico. Como está em (Lazzari; Finatto, 2023), a transcrição desses textos é um recurso facilitador, oferecido em arquivos de texto de diferentes formatos, para seu processamento com ferramentas informatizadas. Esses arquivos, em unidades de texto sem formatação, abastecem um *corpus on-line* (<https://sites.google.com/view/projeto38597>) que também aponta o acesso para os respectivos fac-símiles, em formato PDF, informa as bibliotecas digitais, acervos e plataformas de origem, como também indica alguns estudos sobre as obras em foco.

Em segundo lugar, este artigo, após situar o *corpus* em construção, apresenta, de um modo genérico, a proposta de um *hiperdicionário* a ser dele derivado. Um hiperdicionário, conforme Termignoni (2015, p. 20), é “um dicionário *on-line* projetado para usufruir das vantagens do hipertexto – um conjunto de informações ligadas entre si e editadas no computador”. Nesse tipo de produto, as informações lexicográficas – isto é, as informações sobre os conteúdos, usos e significados contextuais de palavras e expressões, tornam-se *hiperlinks* ou são por eles ilustradas, demonstradas e ampliadas. Nesse tipo de produto, as ontologias (Gritz *et al.* 2021), que são recursos equivalentes a mapas de noções ou representações dinâmicas de conteúdo, tornaram-se recursos informáticos usuais para dar conta de redes semânticas. A percepção dessas redes, especialmente em trabalhos de Processamento de Linguagem Natural (PLN), tem sido bastante útil para a organização de repertórios de vocabulários vinculados a áreas especializadas do conhecimento. Por último, o artigo discute algumas alternativas para o encaminhamento desse hiperdicionário e os desafios para a sua organização.

2 Uma experiência em curso

Conforme já mencionado, vencida a barreira do acesso aos arquivos e acervos, há todo um trabalho para colocar as obras selecionadas em formato *on-line*, na forma de arquivos-texto. Assim, transforma-se o texto escrito das fontes impressas em dados e úteis para pesquisas dedicadas ao léxico, especialmente as que usam com alguma ferramenta informatizada, conforme tem sido feito hoje no âmbito do PLN e das Humanidades Digitais (Vieira *et al.*, 2024). É sobre esses desafios que tratamos a seguir, tomando por base a experiência com a reunião de um *corpus* de manuais médicos antigos, iniciada em finais de 2018 (Finatto, 2020), com apoio de diferentes órgãos de fomento (CNPq, CAPES e FAPERGS). Além do oferecimento público de um *corpus* amostral delimitado, em diferentes formatos de arquivos, ajustados para diferentes usos em pesquisa linguística, esse empreendimento, que integra uma pesquisa acadêmica na área das Ciências do Léxico e da Terminologia, almeja colocar bases para a construção de um glossário histórico médico e de um futuro hiperdicionário de epidemiologia histórica luso-brasileira.

No âmbito dos projetos “Corpus Histórico da Linguagem da Medicina em Português (Século XVIII): Terminologia Diacrônica e Humanidades Digitais” e da pesquisa “Proposta de um hiperdicionário de epidemiologia histórica do século 18 para conectar bases de dados em português” (<https://sites.google.com/view/projeto38597/>), são analisados e disponibilizados, publicamente, uma série de textos médicos reunidos em um *corpus* de arquivos de transcrições. São oferecidos dados de manuais impressos, publicados em Portugal no século 18 por médicos, enfermeiros e cirurgiões, havendo já boas perspectivas para a inclusão de obras brasileiras do século 19 (Lazzari; Finatto, 2023).

No momento, os destaques do *corpus* ficam para algumas das muitas e volumosas obras do médico J. Curvo Semedo (1635-1719), para um primeiro manual de Enfermagem publicado em português, de 1741, e para uma obra da autoria de dois cirurgiões, dedicada ao atendimento de marujos e pessoas embarcadas, de 1794.

A seleção das obras para o *corpus*, além da temática, época e língua, no segmento do século 18, privilegiou apenas aquelas cuja proposta envolvesse, de modo explícito no texto, apresentar a linguagem com algum tipo de facilitação ou didatismo. Ademais, consideram-se apenas obras em formato de livro dedicadas à formação de práticos e futuros profissionais, as quais buscam instruir pessoas que assistiam doentes e que poderiam não ter tido uma formação erudita. Desse modo, o *corpus*, nesse segmento, é relativamente restrito e específico, tendo um caráter apenas amostral frente a um todo que se busca representar.

O caráter bastante restrito e até singelo do *corpus* pode ser confirmado quando se considera todo um universo de obras médicas impressas em Portugal no século 18. Afinal, conforme o recente levantamento e revisão do pesquisador Baudry (2023), essa produção bibliográfica corresponde a um universo de 576 itens, produzidos por pelo menos 87 editores e 250 autores. Baudry, nessa pesquisa de valor inestimável, catalogou e indexou três categorias de tipos de impressos médicos produzidos entre 1700 e 1899: os efêmeros (ou aqueles compostos por folhas soltas de até quatro páginas); os folhetos (com até 50 páginas); e os livros.

Uma das tantas obras apontadas por Baudry (2023), que também se encaixaria nos nossos critérios, especialmente quanto a um foco na formação de práticos e com publicação no séc. 18, é a seguinte – sendo que dela localizamos apenas a primeira edição de 1667, na base digital da Biblioteca Nacional de Portugal:

LEITÃO, Manuel (?-?) *Pratica de barbeiros. Em quatro tratados. Em os quaes se trata de como se ha de sangrar, e as cousas necessarias para a sangria; e juntamente em que parte do corpo humano se hão de lançar as ventosas assim secas, como sarjadas; e em que parte compitão sanguixugas, e o modo de as applicarem; como outras muitas curiosidades pertencentes a tal officio. Acrescentada com a Guia de Sangradores pelo Doutor Henrique do Quintal*. Lisboa: Na officina de Antonio Duarte Pimenta, 1744; 103 p.; in-8º.

Contextualizada a abrangência e a natureza do *corpus*, a seguir, reproduzimos alguns trechos de três obras nele incluídas, cujos autores tinham formações distintas: um médico douto, um padre-enfermeiro e dois cirurgiões de armada. O trecho reproduzido está conforme a sua apresentação original, mas sofreu ajustes para facilitar o processamento. Está em um formato de texto simples, que, frisamos, não é uma *edição filológica*, ainda que parta de transcrições feitas por nós. Nesse sentido, vale destacar que, na maioria dos arquivos do *corpus*, conforme disponível em 2024, não estão respeitadas as separações de sílabas e quebras de linhas do texto, nem a paginação ou o *mise en page* (ou *layout gráfico*) do original digitalizado. Há também, entre os diferentes arquivos oferecidos, divergências em como foram feitas as transcrições com relação ao uso de *f*, *ſ* (*s* curto e *s* longo) e *S* e dos espaços originais antes dos sinais de pontuação. Usou-se apenas *S* e mantiveram-se os espaços antes dos sinais de pontuação, sem desenvolver as abreviaturas. Com os trechos a seguir, buscase oferecer apenas uma ideia dos conteúdos textuais e tipos de transcrições/arquivos do *corpus*:

a) trecho da obra de Semedo (1707), composta por 101 relatos de caso, com as suas observações e ensinamentos. O volume/edição utilizado tem 635 páginas. O trecho é a narração um atendimento feito pelo médico, em 1688. O paciente era um homem identificado em pormenores, com quadro de *tosse veemente com fluxo de sangue*.

Depois que o illustrissimo doente me deo conta do mal que padecia , e descansou hum pouco da sua pratica , lhe fallei do modo seguinte : Deveis saber senhor , que todas as tosses procedem do bofe offendido , e irritado , ou por idiopathia , quando o mesmo bofe não aparta , nem deita fóra de si os excrementos , que resultaõ do alimento com que se sustenta , e conserva na sua individual nutrição , e entaõ sentindose o bofe aggravado , e estimulado do mesmo excremento , pertende deitallo fóra por meyo da tosse , se o póde conseguir , se tira logo a doença ; mas se a desgraça he tal , que não põde separar os ditos excrementos , se reprezão , e amontoão no mesmo bofe , e ahi se engrossaõ cada dia mais , e andando os tempos , se

convertem em pedra , como já vi ; e daqui procedem as tosses continuas , e desesperadas , ou fymphaticamente se offende o bofe , quando os humores serosos se destillaõ da cabeça , e cahem dentro nelle , e vellicaõ , ou mordem as suas membranas nervosas ; ou quando os taes soros apartandose da massa sanguinaria , se transfundem pelos vasos pulmonares , ou por huma resudação entraõ na membrana interior da traca arteria , que he muito mimosa , e sensitiva ; e daqui nasce que no principio he secca ; porque como a materia he entaõ muito delgada , não obedece aos impulsos da faculdade expulsiva , porque não tem esta em que fazera sua força .

Fonte: Semedo, 1707. Observação 2. Disponível em: <https://sites.google.com/view/projeto38597/observa%C3%A7oens-medicas-1707> Acesso em: 18 mar. 2024.

b) trecho da obra do padre-enfermeiro Santiago (1741), com 300 páginas, do capítulo 7, cujo tema é o atendimento, pelo noviço, também um caso que envolve *fluxo de sangue* (o que possivelmente corresponde a uma *hemorragia* ou *sangramento*, conforme o vocabulário atual). O texto ensinava o novato a acudir o paciente na ausência de um cirurgião.

Se o enfermo for fraco, não dormir, e estiver evacuado, lhe fareis alguns remedios depois de ter lançado quasi huma tigela de sangue, sendo pelo nariz.

Primeiramente lhe atirareis ao rosto com hum pucaro de agua , sem que o enfermo o veja lançar, e com a violência , que puder ser, de fórmula , que lhe apanhe todo o rosto, e testa, cujo remedio se póde repetir por trez , ou quatro vezes , com interpolação de tempo entre huma , e outra vez ; e não bastando este remedio, lhe mandareis meter os testiculos em agoa bem fria, ou panos molhados nelles varias vezes repetidos.

Também são boas as ligaduras em os braços, e pernas; e não querendo parar, vereis de que venta corre o sangue: se for pela direita , lhe lançareis huma ventosa grande em sima do figado, o qual está hum dedo por baixo das costellas mendozas da parte direita; e se sahir pela venta esquerda, lhe lançareis a ventosa em sima do baço , o qual está da parte esquerda, hum dedo por baixo das mesmas costellas.

Se o enfermo lançar o sangue por ambas as ventas copiosamente, lhe lançareis duas ventosas, huma no figado, e outra no baço, e as repetireis até que pare o fluxo de sangue. Também são boas nas barrigas das pernas, e

nos musculos pela parte de dentro bastante tempo. Também he bom remedio disparar huma espingarda junto do mesmo enfermo , sem se lhe advertir antes.

Fonte: Santiago, 1741. Capítulo 7. Disponível em: endereço <https://sites.google.com/view/projeto38597/arte-de-enfermeiros-1741> . Acesso em: 18 mar. 2024.

c) trecho da obra *Aviso à gente do mar sobre a sua Saúde* (Mauran; Carvalho, 1794), com 470 páginas. Esse manual é por nós considerado da autoria de dois cirurgiões de Marinha, embora apareça somente o primeiro devidamente identificado. É uma obra para orientar pessoas que atendiam doentes e feridos em navios em que não houvesse médicos ou cirurgiões. O segmento abaixo integra o Capítulo 13, cujo título é *Das Febres Malignas*:

QUando hum Marinheiro da Equipagem de hum Navio Mercantil he atacado de febre maligna , he raro que os outros o não sejaõ successivamente : eu tenho visto mais de huma vez navios desarmados pela quantidade de doentes , que foraõ atacados dellas ; este mal he tanto mais perigoso , quanto os symptomas , que o acompanhaõ no seu principio parecem muitas vezes de pouca consequencia ; o que faz que ellas se desprezem , e que os doentes cheguem muitas vezes a estar nos ultimos extremos antes que se tenha percebido , que estão doentes .

Ha com tudo febres malignas , que saõ annunciadas muitos dias antes por hum numero de symptomas , que saõ correios , que vem diante dellas , taes como hum pulso lento , opprimido , huma lassidaõ acompanhada de tristeza , de fastio , hum pezo na bocca do estomago , a bocca grossa , a lingua , as gengivas , e os dentes cobertos de huma materia amarellada , glutinosa , e verdoenga ; assim se vaõ arrastando muitos dias os Marinheiros , tendo fastio a todo o alimento , até que chegaõ em fim a succumbir . Poder-se-hiaõ pois prevenir estas doenças , obrigando-os a observar a dieta , a tomar ár sobre o convez , dando-lhes o vomitorio do N. 21. , fazendo-lhes tomar depois duas , ou tres porções por dia dos pós do N.º 17. , limonada , ou agoa com vinagre ; porque se elles ficaõ debaixo dos convezes , aonde estão amontoados confusamente , principalmente quando a Equipagem he hum pouco numerosa , eis-ahi começaõ as emanações mefíticas , de que o ár , que elles respiraõ nestes lugares , está carregado , a

inficionar toda a membrana pituitaria do nariz , dos seios da face , e dos bronquios ; dahi ellas se transportaõ para as tunicas do estomago ; o que causa em pouco tempo febres malignas , que se propagaõ , e pegaõ como por huma verdadeira innoculaçaõ , ou enxertia .

Fonte: Mauran & Carvalho, 1794. Capítulo 13. Disponível em: endereço <https://sites.google.com/view/projeto38597/aviso-a-gente-do-mar-1794>. Acesso em: 18 mar. 2024.

Como se pode perceber pelos trechos exemplificados, há todo um *entorno de significação* (Macedo, 2019) específico, linguístico e sócio-histórico, desenhado por esses manuais médicos do séc. 18. Esse *entorno* manifesta-se tanto no uso de palavras e/ou de terminologias quanto nas concepções e narrativas daquela época sobre os corpos, seus funcionamento e as doenças. No manual de Enfermagem, a indicação de se aplicar, por exemplo, um enorme susto ao doente, a título de um tratamento ou remédio contra *sangramentos* ou *fluxos de sangue*, situa bem o quadro dos conhecimentos, saberes e as condições do ambiente físico em que se davam os atendimentos: “*Também he bom remedio disparar huma espingarda junto do mesmo enfermo , sem se lhe advertir antes*”.

3 Do corpus ao projeto do hiperdicionário

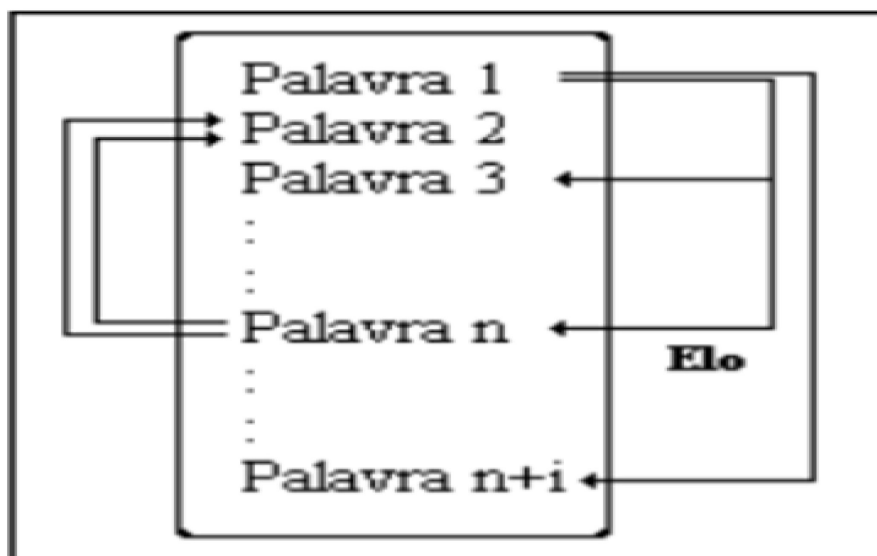
Conforme já mencionado, um dos objetivos da pesquisa em curso é descrever e sistematizar as terminologias, conceitos e modos de dizer relacionados a doenças e seus tratamentos, ao longo das obras do séc. 18 reunidas no *corpus*. Esse tipo de trabalho, conforme apontam Santos, Olival e Sequeira (2020), tende a ser útil para a criação de uma *Base de Conhecimentos* e deve, sempre que possível, ensejar conexões internas e externas, evitando-se “ilhas de dados”.

Com os resultados já alcançados, entendemos que já temos subsídios para o desenho do futuro *hiperdicionário* (Wives, 1998). Esse será um recurso de referência, justamente, pleno de conexões, concretizado em uma estrutura computacional. O

usuário que se tem em mente para tal produto é a pessoa interessada na história luso-brasileira das doenças e de seus tratamentos, que acessa o *corpus* do projeto e que precisa de diferentes informações para apoiar seu entendimento sobre o vocabulário, modos de dizer e as terminologias desses textos. Também é uma pessoa que busca alguma correspondência dos conteúdos antigos com dados da atualidade e que, com eles, tenta ensaiar uma *linha de tempo* até a atualidade.

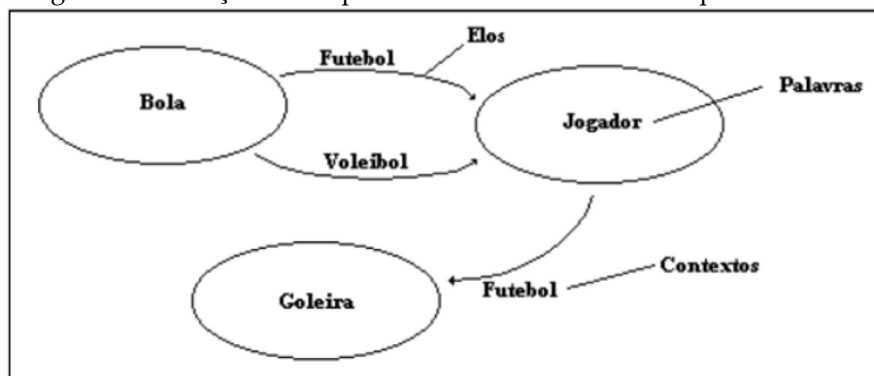
Conforme já foi ilustrado em Wives (1998), um hiperdicionário, sendo ele de natureza e suporte informatizados, é uma estrutura composta por palavras ou por nódulos lexicais interligados por *Elos* conforme vê-se nas figuras 1 e 2 seguir, ambas de sua autoria. O *Contexto* é um conjunto de *Elos* que relacionam as palavras/expressões que pertencem a esse contexto. Cada *Elo* determina um contexto diferente, podendo haver mais de um *Elo* entre duas palavras/expressões, se houver contextos diferentes relacionando-as ou nenhum (caso não exista um contexto que as relacione). As palavras inter-relacionadas pelo mesmo *Elo* constituem um *Contexto*. Nesse caso, os *Contextos* correspondem a um dado âmbito temático que se associa aos itens, exemplificado, adiante, na figura 2, pela área dos esportes coletivos com bola vinculada às palavras Bola, Goleira e Jogador.

Figura 1 — Estrutura abstrata de um hiperdicionário computacional.



Fonte: Wives (1998).

Figura 2 — Relações entre palavras e contextos em um hiperdicionário.



Fonte: Wives (1998).

Para chegar ao desenho desse universo, que pode ser, *grosso modo*, comparado a um *frame semântico* (Santos; Souza; Chishman, 2023) ou a uma ontologia, com esses nódulos de palavras/itens e elos (ou *links*), no contexto das obras médicas impressas do século 18, é preciso identificar e extrair seus relacionamentos e/ou conexões, conforme o *corpus* os espelhe. E esses dados são depreendidos, justamente, pelo exame do vocabulário empregado. Esse exame, para além de um trabalho de *extração computacional* (resumido em Vieira *et al.*, 2024), inclui a leitura detida dos textos originais nos fac-símiles, a leitura atenta dos dados em arquivos de texto, a exploração de ocorrências e de contextos de palavras e de expressões, o que pode ser feito com apoio de ferramentas diversas.

Assim, parte do exame linguístico-conceitual dos textos transcritos que temos feito também conta com a ajuda de ferramentas e de técnicas básicas da Linguística de *Corpus* (Berber Sardinha, 2004), especialmente o *software* AntConc (Anthony, 2021). Com esse *software*, são produzidas listagens e estatísticas diversas a partir das palavras contidas nos arquivos dos textos transcritos, às quais agregam-se procedimentos de Lexicometria (Sanromán; Docío, 2022), guiados pelos preceitos da Terminologia Diacrônica (Finatto, 2020; Dury; Picton, 2009; Curti-Contessoto, 2024). Ademais, o trabalho adota uma perspectiva textual, discursiva e descritiva (Finatto, 2015, p. 225) dos estudos de Terminologia (Krieger, 2001; Finatto, 2015; Ferreira; Marques, 2022), na qual entendemos que as unidades de conhecimento especializado ou os termos

adquirem seu significado em um dado contexto. Isto é, especificamente dentro de um quadro em que o seu papel em um processo, atividade ou evento é destacado e relacionado a outros conceitos, nesse mesmo quadro (Faber, 2018).

Essas referências teórico-práticas dão suporte ao trabalho de identificação e seleção estatística de itens candidatos a encabeçar os verbetes de um glossário histórico e também situam os itens que corresponderão aos núcleos do futuro hiperdicionário. Mais detalhes sobre os termos e palavras mais frequentes em algumas obras desse *corpus* do séc. 18 encontram-se no artigo Finatto, Lazzari (2023) e no capítulo de livro Finatto, Gonçalves, Lazzari, 2023. Por questões de espaço, não reproduzimos aqui o quadro geral desse levantamento lexicométrico.

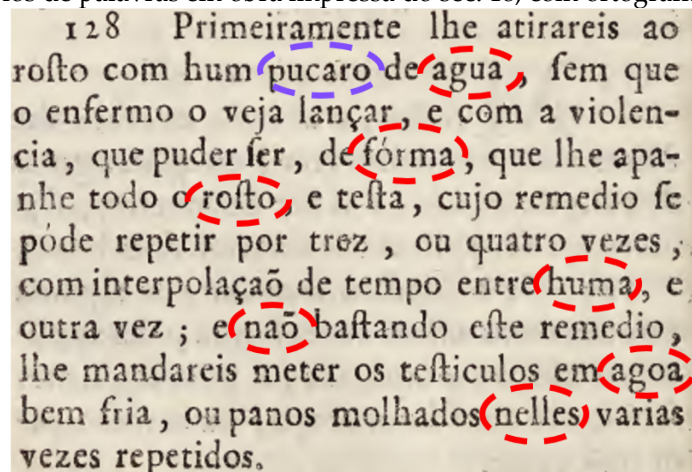
Com esses dados, nesse futuro hiperdicionário, a ideia é associar, em uma rede semântica, nomes de doenças e males (por ex.: FEBRE MALIGNA) e sintomas (por ex.: FLUXO DE SANGUE, TOSSE, CÂMARAS) com designações de partes do corpo (por ex.: FÍGADO, BAÇO, BOFES, BARRIGAS DAS PERNAS, COSTELA MENDOSA, MEMBRANA PITUITÁRIA DO NARIZ), nomes de remédios, preparos, plantas e substâncias (por ex.: ÁGUA DE BELDROEGAS, BEZOÁRTICO, POEJOS, VOMITÓRIO), nomes de tratamentos e procedimentos terapêuticos, nomes de instrumentos (por ex.: PURGA, SANGRIA, VENTOSAS, LANCETAS) e unidades de medida (como GRÃO, ONÇA, PÚCURA(O)).

A perspectiva é organizar esses termos e expressões para além de um ordenamento alfabético tradicional, de um modo que se possa acessar um quadro de informações relacionadas a partir de diferentes pontos de partida, sejam palavras, assuntos ou categorias. Assim por ex., ao iniciar-se a consulta por um dos itens lexicais que correspondem a um dado processo terapêutico, como SANGRIA, seria possível encontrar dados sistematizados sobre seus usos no *corpus* de obras médicas do séc. 18, chegando a seus correspondentes terapêuticos atuais e doenças relacionadas, do passado e do presente. Entre esses dados do *corpus*, são contemplados também os perfis de pessoas/doentes, nomes de pessoas/pacientes, com respectivos dados

geográficos e datas, implicados em um dado tratamento e/ou doença. Esse tipo de informações corresponde, conforme a terminologia do Processamento de Linguagem Natural (PLN), à identificação das *Entidades Nomeadas* ao longo dos textos do *corpus* (Zilio; Finatto; Vieira, 2022).

Naturalmente, há todo um trabalho de geração e tratamento desse *corpus* histórico, que é feito também com auxílio de recursos e técnicas de PLN. Isso porque, no período da impressão dessas obras, verificava-se toda a sorte de variação de escrita e de tipografia, de modo que em uma mesma página, de um mesmo autor, encontrasse, por exemplo a palavra ÁGUA escrita como AGUA ou AGOA, além de haver itens de vocabulário que é preciso situar, como o termo anatômico MADRE por ÚTERO ou as antigas denominações de substâncias químicas, como ROSALGAR, equivalente hoje a um tipo de arsênico (Furtado, 2002, p. 799, Glossário) que seguiu sendo mencionado em outras obras médicas do século 18, como no *Erário Mineral* (Ferreira, 2002). Desse modo, o usuário deverá poder pesquisar o hiperdicionário tanto pela(s) forma(s) antiga (s) das palavras quanto por suas correspondências atuais, a partir de um léxico de variantes, tal como aquele já desenhado por Cândido Junior (2008). A figura 3 a seguir exemplifica um pouco do quadro dessas variações no nosso *corpus*, quanto às diferentes formas de escrita e de tipografia.

Figura 3 — Exemplos de palavras em obra impressa do séc. 18, com ortografia diferente da atual.



Fonte: adaptado de Santiago (1741), cap. VII, p.84. Disponível em:

<https://archive.org/details/b30507340/page/n3/mode/2up>

Por conta disso, o projeto já tenta aproximar, com apoio computacional, a escrita antiga com a sua forma atual ou normalizada. Investe-se, assim, no bom potencial da normalização automática dos textos (Zilio; Lazzari; Finatto, 2024), que serve de apoio para a normalização feita pelas pessoas, sem a substituir. Nessa direção, o *site* da pesquisa já oferece alguns arquivos com a escrita normalizada, o que foi feito manualmente, tendo como referência o português brasileiro, justamente para embasar um processo de Aprendizagem de Máquina. Buscamos esse apoio computacional para a normalização conforme outros grupos de pesquisa já têm feito e reconhecem ser necessário (Cameron; Olival; Vieira, 2023) e importante para melhorar o processamento e a recuperação de informações em fontes que trazem um português antigo e precisam conectar-se a bases com o português atual.

Em que pesem os problemas com a qualidade das informações automaticamente extraídas em dados brutos de transcrições e as variações nas formas das palavras conforme escritas na época, para além do uso do AntConc (Anthony, 2021), algumas ferramentas genéricas de PLN já se mostram capazes de identificar informações relevantes (Quaresma; Finatto, 2020), com bom potencial para ajudar quem lidar com esses textos na criação de bases de conhecimento histórico. Enfim, o processamento desse *corpus*, ainda que siga técnicas muito básicas, como a geração de listagens de palavras gráficas, identificadas como *tokens* e *types*, já ajuda a reconhecer e a refinar a informação escrita e transcrita nele contida. E isso, com todas as limitações inerentes, pode ser muito relevante para várias aplicações, produzindo-se verdadeiros “mapas de navegação” sobre a organização e conteúdos dos documentos reunidos. Esses mapas, em PLN, são denominados de *ontologias*, constituindo esquemas semelhantes às nossas já conhecidas *árvores de domínio* de Terminologia (Fromm; Yamamoto, 2021). Tais representações visam ilustrar a organização e a distribuição das palavras e expressões conforme empregadas, servindo também para oferecer uma imagem do conteúdo de um texto ou de muitos textos ao mesmo tempo. Isso é feito pela identificação de palavras, termos e expressões mencionados, as quais são listadas,

organizadas em hierarquias detalhadas ou mesmo arranjadas em diagramas mais simples, como nuvens de palavras.

A ligação entre as ontologias ou “mapas de conteúdo” é, por fim, um outro dos objetivos futuros da pesquisa. Assim, antigas terminologias de Saúde poderão ser encontradas e mapeadas em seus equivalentes em ontologias da atualidade. Esse mapeamento será mais uma forma de enriquecer os recursos derivados do *corpus* e estabelecer quais as áreas da Medicina, da Anatomia e que tipo de doenças e medicamentos eram conhecidos daquela época. Esses “mapas”, conforme veremos ilustrado adiante, conectam-se à ideia de base do hiperdicionário.

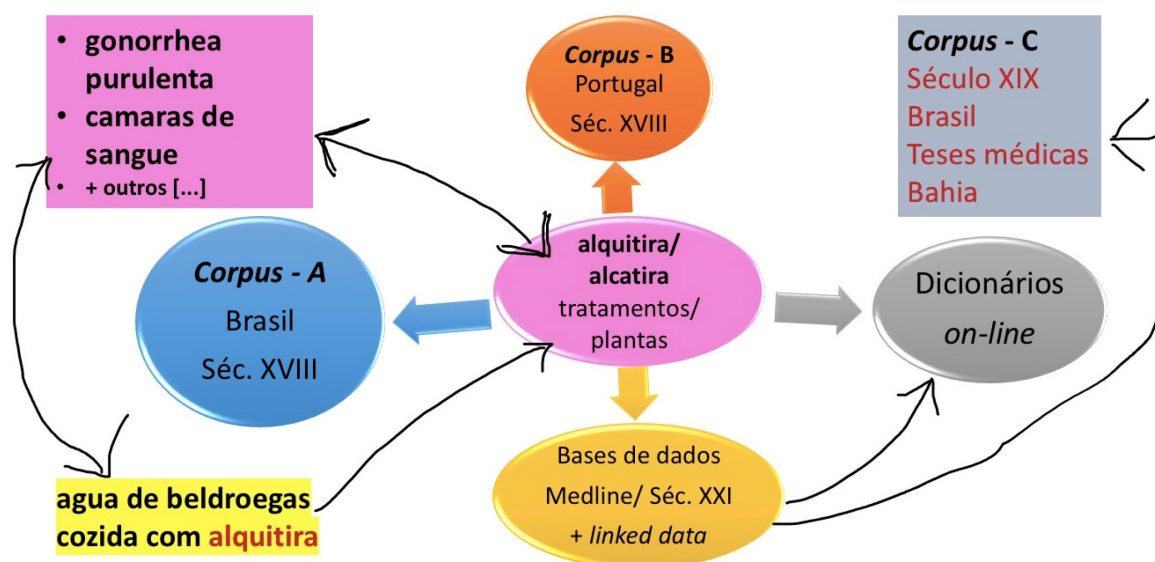
4. Exemplo de nóculo no hiperdicionário: alquitira, uma planta para remédios

Várias terminologias e expressões relacionadas, conforme usadas nos manuais de Curvo Semedo, já se encontram contrastadas com o vocabulário de Enfermagem empregado por Santiago em 1741 (Lazzari; Finatto, 2023). Esse contraste, entre outros, já te ajudado a identificar diferentes vocabulários empregados. A diferença e as semelhanças podem apontar perspectivas, crenças e métodos de cura distintos, além de traços de gêneros textuais emergentes no período (Finatto; Gonçalves; Lazzari, 2023). Assim, percebem-se também elementos da distribuição de conceitos e de saberes dos diferentes práticos e profissionais de saúde em ação e formação no século 18.

A Figura 4 ilustra uma das bases de organização do almejado hiperdicionário. O nome da planta ALQUITIRA/ALCATIRA (var.) ocupa uma posição central. Na figura vemos, como um nóculo e ponto de partida, o nome ALQUITIRA/ALCATIRA. Conforme contextos do *corpus*, essa planta integrava um preparo/remédio adjuvante contra diversos males, como *gonorrhoea purulenta* e *camaras de sangue*, entre outros. Repare que o nosso *corpus* (A-Brasil), pelo nóculo central, associa-se a um outros *corpus* B, de Portugal. Esse acervo português corresponde, entre outros possíveis, aos conteúdos do CIDEHUS-DIGITAL, com destaque para seus *corpora* do séc. 18, como

as *Gazetas Manuscritas da Biblioteca Pública de Évora* (<https://books.openedition.org/cidehus/3083>) e o *corpus* das *Memórias Paroquiais* (<https://www.cidehusdigital.uevora.pt/portugal1758/memorias>). Essa base de textos pode ser percorrida, por palavras, em <http://www.cidehusdigital.uevora.pt/pesquisa>.

Figura 4 — Exemplo das relações do nóculo ALQUITIRA/ALCATIRA no hiperdicionário. Hiperdicionário de Epidemiologia Histórica Luso-Brasileira para vincular bases de dados em português



Fonte: autoria própria.

É das *Memórias* que reproduzimos o seguinte trecho, justamente, sobre *plantas medicinais*, no qual vemos o interesse comercial associado às diferentes espécies fitoterápicas mencionadas em 1758:

E em todos os seus pumares, e hortas, que são muntos, há e se crião inenitas ervas medicináes, como são, bardána, eufrazia, solda, celidonia, malvas violetas, salva, linho em rama, gilbarbeira, pionia, arruda, sabugo, legação, erva moura, tanchagem, rosmaninho, tramagaeira, avenca, barbasco, ourégão, urtigas, funcho, alfavaca de cobra, memendro, héra, erva molarinha, norça, betonica, jarro, salça, ajpo, erva de sam joão, agriogens, e infinitas outras de que os ervedarios da corte vem aqui em todos os annos buscar, para venderem, grande copia; e algumas que não achão pellos, pumáres, vão buscallas dentro do Castello desta mesma Villa, onde tambem são muntas.

Fonte: Memórias Paroquiais, Marvão, Santa Maria. In: Portugal 1758, ANTT, Memórias Paroquiais, Vol. 22, nº 74, pp. 471 a 498. Disponível em: <http://www.cidehusdigital.uevora.pt/s/portugal1758/memorias/marvao-santa-maria/>

Voltando à Figura 4, tem-se, embaixo, o segmento *agua de beldroegas com alquitira*, que é um dos contextos de uso do item destacado no nosso *corpus*. Nesse caso, vale salientar que há o nome de uma outra espécie vegetal, BELDROEGA(S)/BALDROEGA(S), var. – presente em *água de beldroega(s)*. Quanto ao item ALQUITIRA/ALCATIRA, ele também ocorre na expressão *goma de alquitira*. Com essa planta, produzia-se uma espécie de espessante natural, o que facilitava a deglutição dos preparados líquidos por doentes muito debilitados. Esse pode vir a ser um saber e um insumo natural relevante nos dias de hoje.

A partir desses nódulos, pelo hiperdicionário, podem ser recuperados diferentes contextos e situações de doenças e pessoas atendidas que mostram usos terapêuticos antigos desses vegetais. Um exemplo temos nos relatos de Semedo (1707), na sua Observação VI, aqui parcialmente reproduzida – veja a frase final do segmento sobre *agua de beldroegas*:

[...] em 23. de Julho de 1689 . Hum nobre mancebo , cujo nome quero paſſar em ſilencio , contrahio , por cauſa de hum ajuntamento impuro , huma gonorrhœa purulenta , & hum bubaõ na verilha direita ; & como por vergonha propria , ou por reſpeito de ſeus pays não quizeſſe manifellar o ſeu peccado , lhe applicou , por conſelho de hum homem ignorante , certo remedio , com que ſe refolveo a inchaço & ſe ſupprimio o eſquentamento com tanta brevidade , como deſgraça , & deſde aquella hora por diante lhe ſobrivieraõ camaras de ſangue acompanhadas com acerrimas dores de barriga : para remediar a eſtes ſymptomas , ſe lhe deitaraõ innmeraveis ajudas lavativas , abſterſivas , anodinas , vulverarias , & adſtringentes , mas todas ſem proveito [...] & como as dores dos inteſtinos eraõ inſoportaveis , foy neceſſario fazerlhe algumas ſangrias na vea d’Arca do braço direito, aſſim pararevellir os humores que faziaõ as camaras , como para ttemperar a acrimonia delles ; tudo porèm baldadamente . Appellaraõ entaõ os Medicos para o uſo da agua de beldroegas cozida com alquitira [...],

Fonte: Semedo, 1707. Observação 6. Disponível em: <https://sites.google.com/view/projeto38597/observa%C3%A7oes-medicas-1707>
Acesso em: 18 mar. 2024.

Quanto ao vínculos ou elos com dados atuais sobre os nódulos do hiperdicionário, ainda no caso dos itens que correspondem a nomes de plantas, como as já citadas ALQUITIRA e BELDROEGA, a ideia é levar o usuário, via hiperdicionário, a diferentes informações científicas do presente sobre as espécies mencionadas, mencionadas nos tratamentos e nos preparos.

A nossa opção, para esse fim, considerando um cenário português dessas práticas e conhecimentos e plantas, foi de trazer dados da base digital denominada *Flora-on* (Flora-on, 2024). Esse é um portal que sistematiza informação fotográfica, geográfica, morfológica e ecológica de todas as espécies de plantas vasculares autóctones ou naturalizadas listadas para a flora de Portugal (Continente, Açores e Madeira). Nessa base encontram-se, por exemplo, diferentes informações e fotos sobre uma planta denominada *alquitira-do-algarve*, cujo nome científico é *Astragalus tragacantha*, que poderia ser um provável correspondente daquela *alquitira*.

Conforme a plataforma *Flora-on*, aquela *alquitira* é a um pequeno arbusto de distribuição mediterrânica, que em Portugal continental ocorre apenas numa área restrita do litoral sudoeste do Algarve. Infelizmente, tal planta hoje está ameaçada de extinção. Além desse *link* com imagem e dados da planta na *Flora-on*, o hiperdicionário também deverá apontar para outras informações, como as farmacológicas, sobre a planta/espécie referida, como aquelas que há na *Wikipedia*. Também deverá considerar fontes históricas de referência como a que vemos a abaixo, que apontam para um outro nome latino associado à planta:

alquitira – alcatira

Descrição

simultaneamente planta e goma medicinal, à qual os boticários chamavam de *Dragantum gummi*. Goma branca que se extrai de astrágalos.

Fontes: *BLUTEAU*, 1712-1728, vol. I, p. 284; *MORAES*, 1980, vol. I, p. 133.
LR <https://www.museuvirtualdalusofonia.com/glossario/alquitira-alcatira/>

Todavia, como se pode perceber pela informação a seguir, no que se refere aos nomes de plantas e espécies vegetais, desdobram-se verdadeiros leques e até enigmas. Afinal, há nomes populares e inúmeras variações, do passado ao presente. É o caso de TRAGACANTO, algo próximo ao nome científico, também usado para designar a nossa ALCATIRA/ALQUITIRA.

tragacanto

nome masculino

Arbusto do género *Astragalus*, que ressuma uma goma utilizada como emoliente e na preparação de pastilhas, pílulas, etc.

sinónimo

alcatira; alquitira; goma adraganta

Fonte: Porto Editora – tragacanto no Dicionário infopédia de Termos Médicos [em linha]. Porto: Porto Editora. [consult. 2024-02-07 12:32:43]. Disponível em: <https://www.infopedia.pt/dicionarios/termos-medicos/tragacanto> Acesso em: em 07 fev.2024

Por conta disso e do protagonismo que assumem os nomes de plantas nesse cenário discursivo e textual, tornou-se necessário contar com a validação de especialistas em Botânica. Afinal, será importante informar a sua correta caracterização e identificação, o que precisa corresponder ao nome botânico científico normatizado. Ademais, vários nomes de vegetais, usos, preparos e aplicações terapêuticas dessas espécies, conforme citadas nos textos do século 18, resistem hoje apenas na memória da cultura popular portuguesa e também na cultura brasileira. Afinal, várias espécies e usos foram levados ao Brasil colonial, como a(s) BELDROEGA(S) e o(s) POEJO(S). No Brasil, algumas dessas plantas até receberam novos nomes, que podem variar do Norte ao Sul em um país de dimensões continentais e flora autóctone riquíssima. Assim, a proposta do hiperdicionário, no

segmento das plantas mencionadas nesses manuais médicos portugueses, também apontou o imperativo de recorrermos aos estudos atuais de Etnobotânica (Carvalho, 2019).

Os investigadores de Etnobotânica buscam salvaguardar e sistematizar os saberes culturais e populares sobre as plantas, associando-os aos conhecimentos científicos da atualidade. Uma parceria recentemente conquistada para a validação científica dessas denominações do séc. 18 é o apoio do coordenador da *Cátedra UNESCO em Etnobotânica e Ciência Cidadã* do Instituto Politécnico de Beja, em Portugal. Essa Cátedra foi fundada em 2023 e o seu *site* provisório encontra-se em: <https://museu358.wixsite.com/meusite>.

Desse modo, no segmento de *links* do hiperdicionário que partem dos nomes das plantas, prevemos oferecer conexões com dados atuais de estudos culturais e de experimentos farmacológicos associados às plantas citadas no *corpus*. Para finalizar, mencionamos, ainda, o item CARDO, uma outra planta tradicional do Alentejo, em Portugal. Conforme uma notícia de 2017, do Jornal Público, disponível em <https://www.publico.pt/2017/06/26/ciencia/noticia/cardo-uma-planta-mediterranea-que-a-ciencia-portuguesa-esta-a-explorar-1776741>, essa é uma planta mediterrânica que “a ciência portuguesa está explorando”. Conforme a notícia, há registros do uso do CARDO no Alentejo que datam de 1835, mas o nosso *corpus* e suas conexões revelam que os registros sobre seus usos medicinais podem recuar, pelo menos, até 1707.

5 Questões e perspectivas

Relatamos aqui partes uma pesquisa interdisciplinar, em regime de cooperação internacional, que envolve projetar e produzir um hiperdicionário *on-line* de Epidemiologia Histórica, especialmente dedicado ao século 18 e associado a diferentes fontes digitais em português. Esse produto servirá como um ponto de conexão entre

diferentes bases de dados, acervos históricos e trabalhos atuais de grupos de pesquisa que já atuam em parceria.

O hiperdicionário será baseado na sistematização de termos e expressões extraídos de um *corpus* de obras médicas, indexadores de assunto e em dados relacionados de diferentes acervos documentais. Será um produto baseado em técnicas computacionais de mineração de texto, reconhecimento de terminologias e de termos relacionados mobilizados para geração de ontologias e de representações de conteúdo. Essas são técnicas já bastante exploradas e descritas no estado-da-arte das pesquisas de PLN e de Linguística de *Corpus* (Lopes *et al.*, 2009; Loh, 2000). Todavia, demandam vários ajustes quando se lida com fontes históricas, pois essas tecnologias não foram pensadas para lidar com o português escrito de épocas passadas. Nessa via, há todo um conjunto de questões associadas à infraestrutura computacional disponível e aquela necessária para concretizar e fazer funcionar esse tipo ou modelo de hiperdicionário. Afinal, ele precisará conectar-se a um *corpus* de referência de obras médicas selecionadas, já alojado em um dado servidor, e a outros vários, *corpora* e plataformas, que podem utilizar tecnologias distintas ou pouco compatíveis com as que já se tem. Nessa direção, um “sonho de consumo” é o aproveitamento e a nossa conexão com os dados da iniciativa denominada *The Linked Open Data Cloud* <https://lod-cloud.net/> que tem nódulos como a WordNet.

As obras selecionadas para o *corpus* de partida, além do elo comum da língua portuguesa, expressavam o propósito de apresentar seus conteúdos de um modo que pessoas de menor erudição e os práticos em formação daquela época conseguissem entender com mais facilidade as informações veiculadas. Assim, considerando ainda que a maioria das obras médicas mais clássicas publicavam-se apenas em latim, os textos desse *corpus* empregam, lado a lado, termos como o popular “bofes” e o erudito “pulmão” enquanto descrevem processos, substâncias, doenças e sintomas de um modo que, em tese, os diferentes leitores daquela época poderiam acompanhar e situar. Desse modo, a variação terminológica, para além da variação da grafia e

tipografia, seja a variação histórico-temporal ou a variação diacrônica, são os grandes desafios desse projeto. Também o trabalho computacional para apoiar a normalização da escrita dos textos antigos tem ocupado nossos esforços mais recentes (Zilio; Lazzari; Finatto, 2024). Essa normalização – compreendida apenas como um produto computacional - será muito importante para o sistema de remissões “encontrar” as diferentes formas de escrita dos textos antigos e as associar às formas atuais.

Além de um desejado produto *on-line* específico e diferenciado, cuja concretização mostra-se bastante complexa e desafiadora, essa abordagem pode oferecer novos *insights* para os estudos históricos – conceituais, históricos e linguísticos - sobre a assistência à Saúde em determinados recortes cronológicos. De nossa parte, como linguistas, podemos dizer que estudo das terminologias no ambiente comunicativo dessas informações textuais abre a possibilidade de acompanhar a evolução do significado de um termo ou conceito. Ao fornecer informações sobre polissemia, variabilidades denominativas e sinônimas, revelam-se, a partir do léxico e do vocabulário repertoriado, dados sobre a funcionalidade desses fenômenos na história da conceituação, da nomeação e da compreensão de saberes e conhecimentos. Por fim, cabe mencionar que, nesta fase do trabalho com o *corpus*, ficou muito claro o protagonismo das plantas e das espécies vegetais em meio ao cenário sócio-histórico, documental e linguístico das doenças e seus tratamentos no século 18.

Agradecimentos

Este trabalho somente foi possível graças aos apoios do CNPq, Editais 06/2019 - PQ – Proc. 308926/2019-6, APQ Proc. 401770/2022-2 e Pós-Doutorado no Exterior – PDE, Proc. 200051/2023-7. Fica também o nosso muito obrigada a toda equipe do CIDEHUS na Universidade de Évora, Portugal.

Referências

ANTHONY, L. **AntConc**. Versão 4.0.0. Tóquio, Japão: Waseda University, 2021. Disponível em: <https://www.laurenceanthony.net/software.html>. Acesso em: 25 mar. 2022.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, SP: Manole, 2004.

BARROS, L. A. O primeiro dicionário médico do Brasil. **Cadernos de Estudos Linguísticos**, Campinas, SP, v. 46, n. 1, p. 21–42, 2011. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.20396/cel.v46i1.8637156>

BAUDRY, H. **Bibliografia Médica Lusa**: Obras impressas em Portugal no século XVIII. Jul 2023, Lisboa: CHAM / Universidade dos Açores. 191 p. (CHAM eBooks // Estudos; no. 4). Disponível em: https://research.unl.pt/ws/portalfiles/portal/66757186/Bibliografia_Medica_Lusa.pdf Acesso em: 19 mar. 2024.

CAMERON, H. F.; OLIVAL, F.; VIEIRA, R. **Planear a normalização automática**: tipologia de variação gráfica do corpus das Memórias Paroquiais (1758), LaborHistórico, Rio de Janeiro, vol.9, n.1 (2023). DOI <https://doi.org/10.24206/lh.v9i1.52234>

CÂNDIDO JUNIOR, A. **Criação de um ambiente para o processamento de corpús de Português Histórico**. 2008. 143 f. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – Universidade de São Paulo (USP), São Carlos, 2008.

CARVALHO, L. M. de C. **As plantas e os portugueses**. Património, tradição e cultura. Lisboa: Fundação Francisco Manuel dos Santos, 2019.

CURTI-CONTESSOTO, B. O(s) lugar(es) da diacronia na Terminologia: de onde partir para realizar um estudo terminológico-diacrônico hoje? **Acta Scientiarum**. Language and Culture, v. 45, n. 2, p. e67723, 23 fev. 2024. Acesso em: 18 mar. 2024. DOI <https://doi.org/10.4025/actascilangcult.v45i2.67723>

DOMLADOVAC SILVA, C.; MURAKAWA, C. de A. A. Levantamento de unidades lexicais especializadas relativas à cura, em tratados medicinais do Brasil colonial. **Domínios de Lingu@gem**, Uberlândia, v. 15, n. 3, p. 814–834, 2021. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.14393/DL47-v15n3a2021-7>

DURY, P.; PICTON, A. Terminologie et diachronie: vers une réconciliation théorique et méthodologique? **Revue française de linguistique appliquée**, vol. 14, p. 31-41, 2009/2. DOI <https://doi.org/10.3917/rfla.142.0031>

FABER, P. Frames and Specialized Knowledge Representation. Terminologie & Ontologie: Théories et Applications. (Anais) **Actes de la conférence TOTh 2018 Le**

Bourget du Lac – 7 & 8 juin 2018. p. 13-15. Disponível em: http://ontologia.fr/TOTh/Conference/TOTh2018/TOTh_2018.pdf Acesso em: 19 mar. 2024.

FERREIRA, L. G.; FURTADO, J. F. (org.). **Erário mineral**. Belo Horizonte: Fundação João Pinheiro, Centro de Estudos Históricos e Culturais (Coleção Mineiriana); Rio de Janeiro: Fundação Oswaldo Cruz, 2002. Acesso em: 19 mar. 2024. DOI

FERREIRA, R.V.; MARQUES, E. A. Estudos terminológicos: teorias e perspectivas de análise. **ReDILLeT**, Revista Digital Internacional de Lexicología, Lexicografía y Terminología, N.º 5 (2022). Disponível em: <https://revistas.unc.edu.ar/index.php/ReDILLeT/article/view/39957> Acesso em: 19 mar. 2024.

FINATTO, M. J. B. Medicina em português no século XVIII: desafios da Terminologia Diacrônica no cenário das Humanidades Digitais. **Revista Panacea@** – Vol. XXI, N.º 52, p.20-36, segundo semestre de 2020. Disponível em: https://www.tremedica.org/wp-content/uploads/panacea20-52_07_Tribuna_BocornyFinatto.pdf Acesso em: 19 mar. 2024.

FINATTO, M. J. B. Humanidades digitais e estudos históricos do léxico. **Domínios de Lingu@gem**, Uberlândia, v. 17, p. e1769, 2023. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.14393/DLv17a2023-69>

FINATTO, M. J. B.; GONÇALVES, M. F. G.; LAZZARI, R. R. Léxico e terminologia em um novo gênero textual do século XVIII: o manual para enfermeiros. *In*: BRUMME, J.; VINAGRE, N. T. **Emergencia de nuevos géneros textuales y terminología em la historia de los lenguajes de especialidad**. Linguistica philologica 2. Berlin: Peter Lang, 2023. p. 199-232.

FINATTO, M. J. B. Estudos de terminologia no Brasil: diálogos com Portugal. **Oslo Studies in Language** 7(1), p. 223–234, 2015. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.5617/osla.1454>

FLORA-ON: Flora de Portugal Interactiva. (2024). Sociedade Portuguesa de Botânica. Disponível em: www.flora-on.pt. Acesso em : 19 mar. 2024.

FROMM, G.; YAMAMOTO, M. Compilação, reciclagem e padronização de um Corpus Colaborativo de Linguística: percursos metodológicos. **Revista de Estudos da Linguagem**. DOI <http://dx.doi.org/10.17851/2237-2083.29.3.2041-2078>

GONÇALVES, M. F. A Arte de Enfermeiros (1741): aspetos do léxico relativo a doenças e remédios no século XVIII. **Panace@**, vol. XXI, nº 52, 2º sem., p. 68-85, 2020. Disponível em: <http://hdl.handle.net/10174/29311>. Acesso em: 19 mar. 2024.

GRITZ, R. A. J. *et al.* An Ontology Based Natural Language Processing Pipeline for Brazilian COVID-19 EMR. In: BRAZILIAN E-SCIENCE WORKSHOP (BRESKI), 15. , 2021, Evento Online. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 97-104. DOI <https://doi.org/10.5753/breski.2021.15794>

KRIEGER, M. G. Terminologia revisitada. In: KRIEGER, M. G.; MACIEL, A. M. B. (org.). **Temas de terminologia**. Porto Alegre & São Paulo: Ed. Universidade UFRGS & Humanitas/USP.

LAZZARI, R. R.; FINATTO, M. J. B. Exame do vocabulário médico no Português no século XVIII: contribuições da lexicometria para o desenho de um dicionário histórico. **Mandinga**, v. 7 n. 1, p. 102-123, 2023. Disponível em: <https://revistas.unilab.edu.br/index.php/mandinga/article/view/1259/1031> Acesso em: 19 mar. 2024.

LOH, S ; WIVES, L. K. ; PALAZZO M. de O., J . Concept-based knowledge discovery in texts extracted from the WEB. **SIGKDD Explorations** , v. 2, n.1, p. 29-39, 2000. DOI <https://doi.org/10.1145/360402.360414>

LOPES, L.; VIEIRA, R.; FINATTO, M.; MARTINS, D.; ZANETTE, A.; JUNIOR, L. Extração automática de termos compostos para construção de ontologias: um experimento na área da saúde. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, 3(1). DOI <https://doi.org/10.3395/reciis.v3i1.244pt>

LOSE, A. D.; MAGALHÃES, L. B. S. Reflexões sobre edições digitais: fazendo filologia no século XXI. **A Cor das Letras**, [S. l.], v. 17, n. 1, p. 115–126, 2017. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.13102/cl.v17i1.1442>

MACEDO, C. C. Q. A arte da tradução: um breve exercício de terminologia diacrônica. **Cadernos do IL**, [S. l.], v. 1, n. 59, p. 255–270, 2019. Acesso em: 19 mar. 2024. DOI <https://doi.org/10.22456/2236-6385.92537>

MAURAN, G.; CARVALHO, B. J. **Aviso a' Gente do Mar sobre a sua Saude**. R. Typ. de João Antonio da Silva, Lisboa, Portugal, 1794. Trad. adapt. da ed. original francesa, com notas de Bernardo José de Carvalho. Original digitalizado da Coleção Manizola da Biblioteca Pública de Évora (BPE), Portugal.

MAZZONI, V. S. de S. ; LOSE, A. D ; PINHEIRO, L.B.; AZEVEDO, F. C. de; SILVEIRA, É. da. **Livro de theses da Faculdade de Medicina da Bahia** [livro eletrônico]: volume 1, de 1836 a 1888: ed. semidiplomática, descrição e comentários/ Salvador, BA: Memória e Arte, 2023 . Disponível em:

[https://1f11a6e7-5dbd-49ca-a343-](https://1f11a6e7-5dbd-49ca-a343-4afae8a65778.filesusr.com/ugd/d9b288_51f4c1ddeae04bda8ff1eb6d68615894.pdf)

[4afae8a65778.filesusr.com/ugd/d9b288_51f4c1ddeae04bda8ff1eb6d68615894.pdf](https://1f11a6e7-5dbd-49ca-a343-4afae8a65778.filesusr.com/ugd/d9b288_51f4c1ddeae04bda8ff1eb6d68615894.pdf)

Acesso em: 19 mar. 2024

MURAKAWA, C. de A. A. O vocabulário das plantas e drogas na Índia do século XVI. *In: Actas do XVII Encontro nacional da Associação Portuguesa de Linguística*. Lisboa: APL, 2002. p. 329-334. Disponível em: <https://apl.pt/wp-content/uploads/2017/12/2001-27.pdf> Acesso em: 19 mar. 2024.

MURAKAWA, C. de A. A.; NADIN, O. L. De Males e Dores: a variação terminológica na denominação de doenças no Português do Brasil colonial. **Papéis**, Campo Grande, v. 24, Nº Especial, p. 289-320, 2020. Disponível em: <https://periodicos.ufms.br/index.php/papeis/article/view/12405>. Acesso em: 19 mar. 2024.

MURAKAWA, C. de A. A. Vocabulário das enfermidades em documentos do Brasil Colonial: o relato de Prodígiosa Lagoa (1749). *In: MURAKAWA, C. de A. A.; NADIN, O. L. (org.). Terminologia: uma ciência interdisciplinar*. Série Trilhas Linguísticas, n. 22. São Paulo: Cultura Acadêmica, 2013. p. 83-101.

QUARESMA, P.; FINATTO, M. J. B. Information Extraction from Historical Texts: a Case Study. DHandNLP@ PROPOR. **Anais...**2020. Disponível em: <https://ceur-ws.org/Vol-2607/short2.pdf> Acesso em: 19 mar.2024.

SANROMÁN, Á. I.; DOCÍO, S. S. Análise lexicométrica: algumas técnicas aplicadas a entrevistas a visitantes de Santiago de Compostela. *In: FEIJÓ, E. J. T.; PRADO, F. R.; SANROMÁN, A. I. (ed.) Contar o caminho de Santiago: Literatura, discurso(s) e efeitos sociais na comunidade local*. Lisboa: Colibri, 2022. p. 233-260.

SANTIAGO, Fr. D. **Postilla religiosa, e arte de enfermeiros, Guarnecida com eruditos conceitos de diversos Authores. facundos, Moraes, e Escrivarios**. Lisboa Occidental: Na officina de Miguel Manescal da Costa, Impressor do Santo Officio, 1741. Disponível em: <https://archive.org/details/b30507340> . Acesso em: 19 mar. 2024.

SANTOS, A. N. dos ; SOUZA, D. S. de; CHISHMAN, R. L. de O. Conceptualizações de atleta paraolímpico no contexto brasileiro: um estudo diacrônico baseado em frames.

Domínios de Lingu@gem, Uberlândia, v. 17, p. e1768, 2023. Acesso em: 19 mar. 2024.
DOI <https://doi.org/10.14393/DLv17a2023-68>

SANTOS, I.; OLIVAL, F.; SEQUEIRA, O. Excavating the data pit: the Portuguese Parish Memories (1758) as a gold standard. *In: [Anais] DHandNLP 2020: Digital Humanities and Natural Language Processing: Proceedings of the Workshop on Digital Humanities and Natural Language Processing (DHandNLP 2020) co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2020)*. Ed. by M. J. Finatto; R. Vieira; S. Pollak; S. Luz, Évora, 2020, Vol. 2607. Disponível em: <http://ceur-ws.org/Vol-2607/>. Acesso em: 19 mar. 2024.

SEMEDO, J. C. **Observações medicas doutrinaes de cem casos gravissimos**. Lisboa: Na officina de Antonio Pedrozo Galram, 1707. Disponível em: <https://books.google.pt/books?id=qCVH54Hs2i0C&dq=Observac%CC%A7oens+medicas+doutrinaes>. Acesso em: 19 mar. 2024.

TERMIGNONI, S. **Bases teórico-metodológicas para um hiperdicionário semibilíngue de expressões idiomáticas italiano-português em meio a um ambiente virtual de aprendizagem**. Tese de doutorado. Universidade Federal do Rio Grande do Sul – UFRGS, 2015. Disponível em: <http://hdl.handle.net/10183/130768>. Acesso em: 19 mar. 2024.

VIEIRA, R. *et al.* 27 PLN e Humanidades Digitais. *In: CASELI, H. ; NUNES, M. da G. V. Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português – 2a. Edição*. São Carlos : BPLN, 2024. Disponível em: <https://brasileiraspln.com/livro-pln/2a-edicao>. Acesso em: 19mar. 2024.

WIVES, L. K. **Técnicas de Recuperação de Informações com Ênfase em Informações Textuais**. Trabalho Individual / Technical Report, 1998. Disponível em: https://www.academia.edu/2957818/T%C3%A9cnicas_de_Recupera%C3%A7%C3%A3o_de_Informa%C3%A7%C3%B5es_Com_%C3%8Anfase_em_Informa%C3%A7%C3%B5es_Textuais. Acesso em: 19 mar. 2024.

ZILIO, L.; FINATTO, M. J. B.; VIEIRA, R. Named entity recognition applied to Portuguese texts from the XVIII century. *In: TROJAHN, C.; FINATTO, M. J. B.; de PAIVA, V.; VIEIRA, R. (ed.) Proceedings of the Second Workshop on Digital Humanities and Natural Language Processing (2nd DHandNLP 2022) co-located with International Conference on the Computational Processing of Portuguese (PROPOR 2022)*. Virtual Event, Fortaleza, Brazil, 21st March, 2022. Disponível em: <http://ceur-ws.org/Vol-3128/paper10.pdf> Acesso em: 19 mar. 2024.

ZILIO, L.; LAZZARI, R. R.; FINATTO, M. J. B. NLP for historical Portuguese: Analysing 18th-century medical texts. (**Anais**) Proceedings of the 16th International Conference on Computational Processing of Portuguese, 2024. Association for Computational Linguistics (ACL): Santiago de Compostela, Galicia/Spain. 2024. Disponível em: <https://aclanthology.org/2024.propor-1.8.pdf>. Acesso em: 19 mar. 2024.

Artigo recebido em: 08.02.2024

Artigo aprovado em: 22.03.2024