

# Medicina em português no século XVIII: desafios da Terminologia Diacrônica no cenário das Humanidades Digitais

Maria José Bocorny Finatto \*

Medicina en portugués en el siglo XVIII: retos de la terminología diacrónica en el escenario de las humanidades digitales

**Resumen:** El objetivo de este artículo es informar sobre aspectos de una investigación lingüístico-terminológica diacrónica, aún en su fase inicial, alineada con el campo de las humanidades digitales, con la que estamos organizando un corpus histórico del lenguaje de la medicina en portugués. Se tratará de una colección en línea de textos impresos en el siglo XVIII sobre el macrotema «enfermedades y sus tratamientos».

**Palabras clave:** historia de la ciencia, humanidades digitales, lenguaje de la medicina, lingüística aplicada, terminología, terminología diacrónica.

**18th century medicine in Portuguese: challenges relating to diachronic terminology in a Digital Humanities context**

**Abstract:** This article reports on aspects of a diachronic linguistic and terminological study in the Digital Humanities, which is still in its initial phase. Part of the study is to compile a historical corpus of medical language in Portuguese. The online collection comprises printed texts from the 18th century on the broad theme of diseases and their treatment.

**Key words:** applied linguistics, diachronic terminology, digital humanities, history of science, language of medicine, terminology.

*Panace@* 2020; XXI (52): 20-36

Recibido: 15.IX.2020. Aceptado: 5.XI.2020.

## 1. Introdução e objetivos

O objetivo deste artigo é relatar aspectos de uma pesquisa linguístico-terminológica diacrônica, ainda em sua fase inicial, alinhada ao campo das Humanidades Digitais. Nela estamos organizando um *corpus* histórico da linguagem especializada da Medicina escrita em português. Esse *corpus* será uma coleção disponível *on-line*, de acesso gratuito, composto por textos impressos do século XVIII sobre o macro tema *Doenças e seus tratamentos*. Vale salientar que a pesquisa iniciada é uma expansão do que já apresentamos em Finatto (2018). Essa iniciativa é justificada e originada a partir de um quadro de condições e conexões epistemológicas, sobre o que tratamos a seguir.

Em primeiro lugar, vale destacar que, pela mutabilidade de um conjunto de termos científicos e de suas conceituações, ao longo do tempo, desvelam-se diferentes trajetórias do conhecimento. Assim, o exame de denominações e de enunciados, em meio a diferentes textos e discursos, conforme se sucedam, pode oferecer mais uma perspectiva para avaliarmos os processos de evolução dos saberes e das linguagens especializadas.

A propósito, vale lembrar, com Benveniste (1989: 252), que o estabelecimento de uma terminologia própria assinala, em toda ciência, uma conceptualização nova e um momento decisivo de sua história. Para esse autor, a história particular de uma ciência poderia ser resumida à história de seus termos específicos. E, nesse cenário, os termos, a definição terminológica (DT), os enunciados definitórios, em seus diferentes formatos e instâncias textuais, tornam-se elementos-chave para quem se interessar pela história das ciências, das ideias e das sociedades. É sobre isso que estamos tratando, em uma mirada histórica, à medida que compilamos o nosso *corpus*.

Como pano de fundo para a mirada, trazemos os estudos de Terminologia, entendidos como um apartado dos Estudos da Linguagem, em Linguística Aplicada, que se dedicam à descrição e análise dos diferentes fenômenos da comunicação técnico-científica. Em meio a esses estudos, reconhecemos, com diversos autores, que é vital dar conta da variabilidade de terminologias, DTs e conceitos, em diferentes situações comunicativas, seguindo enfoques diafásicos, diatópicos, diastrásticos e, também, diacrônicos.

No cenário brasileiro, o ponto de vista dos estudos e trabalhos de Terminologia tem sido mais descritivo e analítico do que normativo. Além disso, a Terminologia brasileira tem mostrado uma tendência a ir além dos repertórios de léxicos

\* Universidade Federal do Rio Grande do Sul – Programa de Pós-Graduação em Letras, Programa PQ do CNPq, Brasil. Dirección para correspondencia: [maria.finatto@gmail.com](mailto:maria.finatto@gmail.com).

temáticos, buscando alcançar também a «especificidade dos contextos discursivos nos quais as terminologias se manifestam» (Krieger, 2000: 228). Nessa direção, temos travado diálogos recorrentes com os estudos do texto e do discurso, com os estudos da Tradução e da Ciência da Informação. Disciplinas ou programas de estudos de Terminologia, quanto a teorias e práticas, figuram na maioria dos currículos da formação profissional universitária de tradutores do Brasil. Nesses programas de formação, a habilitação para tradução de textos de temáticas técnico-científicas tem um destaque especial.

Para além das abordagens linguísticas dos estudos de Terminologia, também a História vem tentando entender, a seu modo, os léxicos temáticos e sua evolução. Por isso, uma maior aproximação entre os Estudos da Linguagem e os de História, no cenário brasileiro, conforme entendemos, ainda merece ser mais incentivada. Afinal, tal diálogo tenderia a ser benéfico para todos os envolvidos, ainda que qualquer empreitada interdisciplinar seja sempre um desafio.

Em História, escolas como, por exemplo, a de Koselleck (a chamada História dos Conceitos), buscam esclarecer que as terminologias, cunhadas e adotadas em determinados cenários de conhecimento, têm sua especificidade a partir do sujeito que as enunciam e dos contextos sociais, temporais, regionais, entre outros, em que se colocaram. Nesse sentido, alguma proximidade com a Linguística já tem sido considerada positiva também por alguns historiadores (cf. Jasmim, 2005; referido também por Macedo, 2019).

Para Robin (1973: 41), por exemplo, uma Lexicologia, na História, teria como finalidade o estudo do vocabulário das sociedades passadas. E, nesse âmbito, poderíamos considerar a História como uma ciência (Rüsen, 2010) que, naturalmente, tenderia a dialogar com outras disciplinas conhecidas como Ciências Humanas. Desse modo, conforme assinala Witter (2005: 14), o «empenho dos historiadores no diálogo com outras ciências sociais, como a antropologia, a sociologia, a filosofia e a psicologia, fez surgir, no âmbito das pesquisas históricas, novas áreas de interesse, em termos de objetos e abordagens». Nessa dinâmica, a narrativa histórica pode ser vista como fruto da capacidade de o homem interpretar a si e o mundo que o cerca. As diferentes narrativas históricas, ao abrigarem terminologias, vocabulários, visões de mundo e conceituações, favorecem apreciações, podendo ser ampliadas e enriquecidas.

Por outro lado, o enfoque histórico, em meio a estudos de Terminologia *stricto sensu*, conforme nossa percepção, não tem sido muito desenvolvido no Brasil. Entretanto, importa ressaltar as iniciativas de colegas que já lidaram com a presença de termos técnico-científicos em Lexicografia ou com o discurso terminográfico (Krieger e Maciel, 2001; Barros, 2005) e mesmo com a verificação linguístico-conceitual em diferentes fontes históricas (como nos trabalhos de Gonçalves e Murakawa, 2009 ou de Marengo, 2016). Também a Filologia Digital tem avançado entre nós, reunindo pesquisadores brasileiros, linguistas e cientistas da Computação, que têm labutado em torno de textos antigos, especialmente manuscritos. Esses colegas têm procedido ao seu tratamento e edições (Paixão de Sousa e Trippel, 2006) para fins de oferecimento, público e gratuito, de

acervos *on-line*. Entre esses *corpora* e bases de dados, é possível o estudioso de Terminologia encontrar diferentes materiais de seu interesse.

No cenário internacional, verificamos uma produção mais intensa sobre o tema da diacronia, o que se confirma por sucessivos eventos, como uma série de encontros sobre a História das Linguagens Ibero-Românicas de Especialidade. Em 1997, em Barcelona, por exemplo, já houve o colóquio *La història dels llenguatges iberoamericànics d'especialitat (segles XVII-XIX): solucions per al present*, organizado pelo IULA (Instituto Universitário de Linguística Aplicada), evento que se repetiu várias vezes. E, na mesma linha, em 2019, tivemos o *VIII Colóquio Internacional sobre a História das Linguagens Ibero-românicas de Especialidade (CIHLIE)*<sup>1</sup>, em Pisa, Itália.

Ao longo dos últimos vinte anos, o CIHLIE tornou-se uma referência obrigatória para qualquer investigador interessado nas linguagens de especialidade nos idiomas ibero-românicos. Nesses eventos, têm sido tratados, desde enfoques muito variados, de estudos de Terminologia conexos à divulgação das ciências, da tradução de textos científicos. Também têm sido temas a padronização, a investigação diacrônica e a comparação textual, as tradições culturais e discursivas e as perspectivas dialógica intertextual e social das linguagens especializadas.

Considerando uma interface Terminologia-História, Bernt Møller (1998: 426) já havia assinalado a necessidade de impulsionarmos um «estudo da evolução dos termos e das terminologias». Assim, propôs o que denominou *terminocronia* para enfrentar o que percebia como um *déficit diacrônico* e até mesmo uma certa condição de «primo pobre» desse tipo de estudo (conforme relatam Dury e Picton, 2009: 31).

Afinal, se a variação terminológica é tão necessária e óbvia quanto a variação lexical ou linguística em qualquer linguagem ao longo dos espaços e das sociedades, também as variações diacrônicas (cf. Boulanger *apud* Faulstich, 2001: 19) perfazem a essência das linguagens que veiculam conhecimentos. E, a despeito de um relativo desprestígio, diferentes estudiosos de Terminologia, fora do Brasil, em suas diferentes orientações teóricas, como Gaudin (2007) e Temmermann (2000, 2001 e 2004) já chamaram atenção para o fato de que o estudo diacrônico dos termos — e das linguagens técnico-científicas — é, sim, relevante. Essa relevância é justificada à medida que repercute, no seu aspecto linguístico, não só uma história de designações e de conceitos, mas elementos de interesse da História das Ideias.

De nossa parte, reconhecemos, há bastante tempo, a importância desse tipo de estudo (Finatto, 2001). Mas, mesmo sendo legítima a interface, é importante apontar, frente ao que consegue produzir em termos de pesquisa, as diferentes dificuldades e impedimentos que ainda hoje enfrentamos para lidar com a história das linguagens especializadas. Somente assim, reconhecendo entraves, poderemos equacioná-los e ultrapassá-los. Apesar dos muitos problemas apontados por Dury e Picton (2009), muito bem situados no artigo de Macedo (2019), há ainda muito o que fazer para que haja mais estudos diacrônicos em Terminologia.

Em função disso, sem pretender colocar nenhum modelo de investigação inovador, temos tratado do tema das terminologias

e das linguagens especializadas em uma perspectiva diacrônica, em algumas experiências de pesquisa bem recentes. Nosso interesse tem se voltado para o léxico da Medicina conforme foi escrita e publicada em português no século XVIII. Buscamos reunir fontes, contrastar obras e descrever a configuração da informação, sob um ponto de vista linguístico-terminológico e textual, contando com apoio computacional (Quaresma e Finatto, 2020), filológico e lexicográfico.

Este artigo é feito dessas experiências mais recentes. Assim, nesta oportunidade, nosso objetivo é relatar aspectos de nossos, ainda em fase inicial, alinhados ao campo das Humanidades Digitais. Para a empreitada, contamos com o indispensável suporte de historiadores, de especialistas em Medicina e de cientistas da Computação, que já lidaram com o processamento de *corpora* antigos, manuscritos e impressos.

Nosso *corpus* é dedicado às diferentes obras do médico alentejano João Curvo Semedo (1635-1719). Uma obra sua, de 1707, inaugura nosso *corpus*-amostra inicial, já oferecido em um *site* de acesso aberto. Tal *site* foi planejado para dar apoio a atividades remotas de ensino em disciplinas de Terminologia, Lexicografia e de Linguística Histórica junto aos cursos de Letras da Universidade Federal do Rio Grande do Sul (UFRGS). Mais detalhes desse repositório podem ser conferidos em TEXTECC<sup>2</sup>, em Finatto, Gonçalves e Quaresma (2018) e em Finatto (2018).



FIGURA 1. Capa da obra de Semedo (1707)

Tendo conseguido conquistar apoio institucional para essa primeira iniciativa, que vincula atividades de ensino e pesquisa, agora o objetivo é ampliá-la. Essa ampliação estará sob forma de um *corpus* temático digital, com todas as obras de Semedo, ladeado com materiais diversos, como representações de conteúdo com suporte computacional. Com ele, esperamos, além de divulgar esse tipo de estudo, suprir necessidades de diferentes pesquisadores, professores e acadêmicos, haja vista que *corpora* históricos temáticos do Português ainda são relativamente raros no panorama nacional e internacional.

O *corpus* ampliado deverá ser integrado por seis obras completas de Semedo, publicadas entre 1704 e 1783, mais adiante listadas. As figuras 1 e 2 ilustram algumas características de uma das obras de Semedo, da qual retiramos alguns exemplos de análise para este artigo.

A escolha pelo autor e seleção de suas obras baseia-se no fato de ele ter sido uma referência fundadora – e polêmica – do conhecimento médico e farmacológico constituído e divulgado em língua portuguesa. Seus livros foram consumidos – traduzidos e citados – na Europa, no Brasil e em outras colônias portuguesas, por quem atendia doentes e necessitava de algum guia de procedimentos. O leitor-foco preferencial de seus livros, na época em que foram publicados, era a pessoa de menor erudição, pouco afeita a materiais semelhantes disponíveis apenas em Latim. Daí serem objeto, ao longo da nossa pesquisa, de considerações sobre o seu potencial de Acessibilidade Textual e Terminológica (ATT). Isso tem a ver com uma provável facilitação da linguagem empregada, de modo que pudesse ser compreensível pelo leitor tipificado, à época, como uma pessoa dedicada a atender doentes, mediante pagamento, caso, por exemplo, dos boticários ou dos cirurgiões-barbeiros.

As seis obras reunidas correspondem a um acervo de mais de 2 mil páginas impressas na sua versão original, as quais serão tratadas e oferecidas, aos poucos, sob a forma de transcrições e imagens de diferentes formatos. Buscamos produzir e oferecer edições em ortografia original e, se possível, atualizada, acompanhadas de descrições comentadas de suas terminologias e de representações do conteúdo de cada uma, geradas com apoio de ferramentas computacionais. A sistemática do nosso trabalho é próxima à de Marengo (2016), que propõe a edição semidiplomática de documentos para elaboração de glossários terminológicos seletivos. No nosso caso, salientamos, são obras impressas, a que temos acesso já previamente digitalizadas, e os glossários são uma, entre outras modalidades de (re)apresentação do conteúdo dos textos do *corpus*. Os impressos do século XVIII, como veremos mais adiante, também trazem desafios importantes, principalmente em função dos recursos tipográficos da época e da variação ortográfica. Ainda assim, são dificuldades incomparáveis ao trabalho direto com manuscritos.

Na sequência deste artigo, apresentamos nossas bases teóricas de Terminologia e situamos a Terminologia Diacrônica e o campo das Humanidades Digitais. Depois, caracterizamos as fontes textuais sob estudo e trazemos pequena amostra de estudos preliminares, centrados apenas em uma das obras que integra o *corpus*.



FIGURA 2. Amostra do conteúdo da obra (Semedo, 1707), com retrato do autor

## 2. Estudos de Terminologia, Terminologia Diacrônica, e Humanidades e Filologia Digitais

A nossa pesquisa, para reunião, edição e tratamento do *corpus*, toma como base as referências teórico-metodológicas da Teoria Sociocognitiva da Terminologia (Temmermann, 2000, 2001 e 2004). A essa teoria aliamos perspectivas textuais e comunicativas da Terminologia, como também da Linguística Histórica, Linguística de *Corpus* e Processamento de Linguagem Natural (PLN).

Dessa síntese, imaginamos, poder-se-ia chegar a uma Terminologia Diacrônica baseada em *corpus*. Entretanto, como um cenário maior e que precede a esse tipo de estudo, temos o trabalho filológico fundamental, a crítica e edição do texto conforme há muito já se tem feito em Filologia em Linguística Histórica.

Nesse sentido, vale destacar que a Filologia (Paula e Amo-

rin, 2020), tradicionalmente, está associada a três tarefas: a substituição do texto à sua forma genuína, para sua publicação; a uma função adjetiva ou qualificadora, na qual se busca tornar evidentes informações subjacentes ou implícitas; e a uma função denominada *transcendente*, que trata de explicitar relações históricas que teriam motivado o texto. No nosso caso, acreditamos que também essas funções nos movem, ainda que a primeira possa ser ajustada, pois as obras já são acessadas previamente digitalizadas.

Nos dias de hoje, considerando toda uma informatização e o trabalho com documentos digitalizados, já se realiza uma *filologia virtual* (Monte e Paixão de Sousa, 2017). Nesse âmbito, temos uma edição feita em *camadas* de dados, de modo que diferentes dimensões de processamento e de representação de um documento possam servir a diferentes propósitos de estudos e às funções antes referidas. A seguir, apresentamos as principais ideias envolvidas nesse cenário multidisciplinar de orientações.

## 2.1. Terminologia

Uma definição mais restrita de Terminologia é a que já temos em Sager (1993: 21), apresentada como «el estudio y el campo de actividad relacionado con la recopilación, la descripción y la presentación de términos». Mas desde essa concepção, os estudos de Terminologia têm ido desde o ponto focal mínimo de um léxico mais específico ou marcado, no âmbito da *palavra* ou do *termo*, até o ponto máximo do todo do texto (algumas vezes entendido como um *discurso*). Nesse âmbito máximo, além do texto escrito, vemos hoje também a inclusão do texto oral (Serra, 2019) em investigações de Terminologia em que se abarcam também os saberes populares associados a determinadas ocupações e atividades profissionais, como a Agricultura, a construção civil e o processamento de matérias-primas.

Assim, ao lado do plano-texto sob exame, colocam-se concepções de um *discurso especializado*, o qual já foi definido, por pensadores da Terminologia, como um conjunto de interações negociadas, identificado por meio de marcas lexicais ou morfológicas, elementos gráficos ou pelo modo pelo qual a informação é apresentada (Cabré, 2005). A materialidade do *discurso especializado* associa-se, nesse percurso, a um dado uso de língua e a (sub)linguagens<sup>3</sup> que se sucedem e se combinam em um todo de relações. Esse todo é o *macroplano da língua-em-uso*.

No âmbito dos estudos de Terminologia, as linguagens técnico-científicas, em suas diversidades, matizes e apropriações, tendem a ser identificadas pelas denominações *linguagens especializadas*, *línguas de especialidade* ou *linguagens de especialidade*. Em que pese toda uma gama de discussões sobre cada uma dessas denominações e sobre diversos pontos de vista implicados a partir de cada uma, ainda é usual assistirmos, em trabalhos e pesquisas atuais, retomadas de uma distinção inaugural, entre a chamada *linguagem geral* ou *linguagem cotidiana*, não marcada, e a *linguagem especializada*, marcada.

Em paralelo à conservação da discussão sobre essas oposições, divulgam-se hoje, no Brasil, por exemplo, pesquisas sobre vocabulários considerados afins aos terminológicos em cenários de investigação que seriam inusitados, se colocados há 10 anos. Como exemplo desse «caráter inusitado», vemos hoje estudos terminológicos a partir de textos ficcionais (*cf.* Fromm, 2013; Esperandio, 2015; Carneiro, 2016), tratando desde a ficção científica televisiva —em obras como *Star Trek/Jornada nas Estrelas*—, até o vocabulário terminológico empregado em romances escritos ou de narrativas filmicas.

Nessa via, a distinção geral/especializado, em termos absolutos, parece cada vez menos importante, visto que, atualmente, laços de semelhança e de diferenças entre *palavras* e *terminologias* tendem a ser, conforme entendemos, apenas resgatados ou equacionados. Não parece que a oposição ainda precise, simplesmente, ser resolvida. Entre recentes discussões sobre teorias e perspectivas de Terminologia, nas quais geralmente se encontra uma oposição entre conceitos (associados a termos) e significados (associados a palavras *comuns*), tratar-se-ia mais de «mirar lo que asemeja una palabra y un término que no lo que los diferencia» (Krieger, Santiago e Cabré, 2013: 331).

Feitas essas considerações, no cenário dos livros de Medicina escritos em português no século XVIII com que lidamos,

cabem essas perguntas: que tipo de textos e linguagens especializados seriam esses? Seriam obras de divulgação científica? Tal item ou expressão *x*, conforme empregado em uma dada obra, será uma terminologia ou se trata de um vocábulo *comum*, percebido terminologizado pela análise e pelos *enriquecimentos* filológicos feitos no presente?

Naturalmente, a nossa mirada sobre a linguagem empregada nas obras de J. C. Semedo deverá considerar o valor e a significação em meio a um passado determinado e em meio à consolidação de um (pré-)conhecimento que, comparado ao de hoje, seria irrisório. Isso sem contar os saberes peculiarmente construídos pelo próprio autor, muitos à base de tentativas e erros, repletos das crenças da sua época e muitas vezes expressos de modo que servissem de propaganda para os fármacos e misturas que ele mesmo criava e comercializava, fazendo-se também de boticário.

Enquanto linguistas, convocamos, para nos auxiliar em um tal equacionamento ou graduação de oposições, paradoxalmente renovados no passado em foco, a noção de gênero textual e discursivo. Afinal, com o exame das obras de Semedo, podemos observar subsistemas linguísticos e deparamo-nos, na prática, com *socioletos profissionais* e *sublinguagens*. Ao refletir sobre a natureza dessa modalidade de uso da língua, Aubert (1996) já havia postulado —e antecipado— que as *linguagens de especialidade* poderiam ser concebidas como um «conjunto de marcas lexicais, sintáticas, estilísticas e discursivas que tipificam o uso de um código linguístico qualquer em ambiente de interação social, centrado em uma determinada atividade humana» (p. 27, grifo nosso).

Em resumo, pode-se afirmar que o tratamento de uma linguagem especializada, compreendida como um todo multifacetado, em Terminologia, tem ido muito além da descrição de terminologias ou de um vocabulário mais ou menos marcado. Entre vários diálogos instaurados e renovados pela Terminologia atual, será necessário o deslocamento para a (re)valorização dos aspectos históricos e/ou evolutivos das linguagens especializadas conforme eram realizadas no século XVIII em português. Reconhecimento do entorno de significação sócio-histórico, em torno das terminologias, das obras e do domínio em foco, como um todo, deverá ser uma etapa do estudo com esse tipo de material.

## 2.2. Para uma Terminologia Diacrônica

Para Krieger (2006: 162), «o léxico de uma língua, longe de ser um bloco monolítico, constitui-se como um conjunto heterogêneo em vários ângulos de sua composição». O tempo, em sua sucessão, o espaço e os registros humanos determinam e espelham sua constituição. A heterogeneidade, assim, é constitutiva da língua e do léxico, planos que se combinam em um dinamismo natural. É «esse dinamismo que torna o léxico o pulmão das línguas, e, simultaneamente, um objeto multifacetado e em constante mobilidade» (Krieger, 2006: 163).

Desse modo, torna-se lícito considerar que as atualidades terminológicas de uma dada linguagem especializada, seja ela de um âmbito técnico, científico ou tecnológico, no seu aqui e agora do século XXI, sejam tributárias e herdeiras de todo um

processo evolutivo, de uma sucessão de concepções e conhecimentos, desde o estabelecimento do domínio em que se inserem. A história de uma especialidade ou de um domínio de conhecimento, portanto, repercute sobre seus atuais usos linguísticos e convenções de discurso. A Terminologia Diacrônica trata justamente de recuperar e sistematizar essas transformações entre o presente e o passado. Assim, em tese, o estudo de um texto do passado deve auxiliar a melhor entender as motivações para a linguagem dos textos do presente.

A despeito do valor desse tipo de estudo, Dury e Picton (2009: 31) situam e explicam alguns obstáculos para a (in)existência ou para a pouca produção de pesquisas diacrônicas em Terminologia. São obstáculos, conforme as autoras, de natureza: (a) teórica e histórica; (b) técnica; (c) pragmática; e, (d) psicológica. Esses obstáculos encontram-se muito bem situados e explicados no trabalho de Macedo (2019).

Conforme vemos, o principal entrave para uma mirada histórica segue sendo a percepção de que esses estudos seriam algo supérfluo frente às necessidades mais imediatas do trabalho com a linguagem científica do presente, plena de transformações e desafios para o analista.

Apesar desses entraves, Dury e Picton propõem uma «reconciliação» com a diacronia, sugerindo algumas alternativas para o seu enfrentamento. Essa reconciliação, em Terminologia, que entendemos ser muito necessária, dar-se-ia, por uma via teórica, mas também prática, através do acolhimento de propostas como as de Rita Temmerman (Temmerman, 2000, 2001 e 2004), em sua Teoria Sociocognitiva da Terminologia (TSCT).

Afinal, ao defender a importância de descrever o caráter evolutivo das unidades de compreensão —que na TSCT tomam o lugar dos *conceitos*—, para melhor apreendê-las, a TSCT, justamente, acolhe e estimula os estudos diacrônicos. Mas, para que se chegue a concretizar uma Terminologia Diacrônica, pois ainda se trata de apenas uma ideia em meio a um programa teórico, acreditamos ser prudente alguma ponderação sobre os obstáculos acima apontados. A partir deles, poderemos identificar vias alternativas de trabalho, conforme as condições de que dispomos. Destacaremos neste artigo apenas os problemas de ordem técnica, apontados por Dury e Picton (2009: 31) buscando algumas prováveis soluções.

No bloco de «problemas técnicos», está a falta de recursos computacionais, textuais e financeiros para estudos históricos frente a necessidades muito mais prementes do presente. Nesse contexto, compilar *corpora* que cobrissem períodos passados, como o século XVIII, parece empreitada onerosa e pouco viável frente a uma utilidade mais imediata desse tipo de estudos. Esse aspecto, acreditamos, em tempos de uma pandemia mundial de COVID-19, nesse ano de 2020, tornou-se bastante relativo. Afinal, nunca se voltou tanto ao tema de doenças do passado tentando-se dar conta de fenômenos do presente, como é o caso de textos sobre a Gripe Espanhola, sobre a Peste, a Lepre e sobre antigas epidemias e seus tratamentos, incluindo os desenvolvimentos mais primitivos de vacinas.

Outro problema importante, reconhecido em Dury e Picton (2009: 33), é a limitação das ferramentas computacionais

de análise de *corpus* com que contamos hoje. Afinal, são, na maioria, desenvolvidas visando registros da atualidade. Elas, em geral, não conseguem lidar com textos antigos, ortografias e caracteres antigos, manuscritos, transcrições de manuscritos, entre outros elementos. Além disso, há a necessidade de haver *corpora* históricos para contrastar, por exemplo, o vocabulário empregado em um conjunto específico de textos de uma dada época e os padrões gerais, gramaticais e lexicais, que seriam espelhados em um amplo universo de textos dessa mesma época.

Realmente, as dificuldades, em termos técnicos, são inúmeras. Mas, nem por isso, podem nos paralisar. É importante seguir em frente, pois o reconhecimento da importância desse tipo de trabalho, em diferentes cenários de pesquisa, especialmente para além do campo dos Estudos da Linguagem, tende a impulsionar a criação ou mesmo adaptação de recursos técnicos já disponíveis.

Independentemente de obstáculos e do tipo de suporte mais ou menos «moderno» com que se conte, no âmbito da constituição das ciências, acompanhar como os fenômenos se apresentavam e foram interpretados pode ser a chave para os novos enfrentamentos do presente. Se, como a TSCT de Temmermann aponta, as unidades de interpretação estão em constante evolução, como um resultado da busca por um entendimento maior e mais amplo, lidar com os textos antigos nos permite observar uma trajetória de buscas corretas e equivocadas. Nesse empreendimento, acompanhar a variação conceitual (Pereira e Nadin, 2019), que mostra a construção de conhecimentos, é um elemento-chave para a Terminologia Diacrônica.

De nossa parte, acreditamos que muitos dos recursos técnicos disponíveis no presente podem ser adaptados e até recuperados para lidarmos com materiais textuais antigos. Além disso, já temos, sim, importantes iniciativas associadas a estudos históricos, com textos antigos, também no âmbito do Processamento da Linguagem Natural, um ramo da Ciência da Computação. Exemplos disso vemos, recentemente, em Cameron, Gonçalves, Quaresma (2020) e em Santos, Olival e Sequeira (2020). Produzir *corpora* temáticos, como um acervo dedicado a obras impressas sobre doenças e seus tratamentos no século XVIII, pode ajudar a expor essa necessidade. Na próxima seção, após tratar, muito brevemente, do campo das Humanidades Digitais, tentamos demonstrar, justamente, a viabilidade dessas ideias.

### 2.3. Humanidades Digitais e Filologia Digital

Como o âmbito das Humanidades Digitais é ainda relativamente pouco conhecido, cabe apresentá-lo, brevemente. Enquanto expressão agregadora de práticas, de teorias e de métodos que se desenvolvem desde a segunda metade do século XX, as Humanidades Digitais são consideradas, pelos muitos laboratórios e centros de pesquisa surgidos desde 2010 no mundo acadêmico nacional e internacional, como uma *transdisciplina*.

Esse novo ponto de encontro de conhecimentos, que é inter e transdisciplinar, incorpora os métodos, os dispositivos e as perspectivas heurísticas das Ciências Humanas e Sociais (como as áreas de Letras e Linguística, Filologia, Estudos Literários, História, Museologia, Arquivologia, entre outras), ao mesmo

tempo em que mobiliza as ferramentas e abordagens singulares abertas pela tecnologia digital, com contribuições de diferentes ramos das Ciências da Computação, além de outras Ciências Exatas e da Terra.

Como estrelas do cenário internacional da pesquisa das Humanidades Digitais, vemos acervos de valor histórico reunidos e explorados com apoio informatizado, especialmente acervos de textos escritos, mas há também *corpora* de dados de diferentes modalidades e naturezas. Exemplos disso são reconstituições históricas de objetos físicos, (re)construções digitais tridimensionais de paisagens físicas e de cidades antigas, atlas histórico-digitais com mapas de regiões antigas conforme eram no passado, reproduções digitais de monumentos e de artefatos diversos, como fósseis, sofisticadas digitalizações de obras de arte e até reconstruções informatizadas integradas a modelos de animação de representações de organismos vivos.

Algumas instituições universitárias, no exterior e no Brasil, já oferecem formação e informação em Humanidades Digitais. Alguns exemplos a citar são:

- a) Universidade do Minho, em Portugal, que tem um pioneiro programa de Mestrado em Humanidades Digitais<sup>4</sup>;
- b) Universidade Federal Rural do Rio de Janeiro — junto ao PPGIHD — Programa de Pós-Graduação Interdisciplinar, que também abriga um Mestrado em Humanidades Digitais<sup>5</sup>, como também suas atividades de pesquisa em Mineração de Dados Digitais<sup>6</sup>;
- c) Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que abriga o Laboratório em Rede de Humanidades Digitais, iniciativa do Grupo de Pesquisa Informação, Memória e Sociedade (IMeS). Seu objetivo é incentivar a produção científica e tecnológica além de desenvolver ferramentas e metodologias mediadas no vasto campo das Humanidades intermediada pelas plataformas, ferramentas e registros digitais<sup>7</sup>.

Nesse moderno contexto de resgates do passado, um outro ótimo exemplo de várias possibilidades de materiais reunidos está no Centro Interdisciplinar de História, Culturas e Sociedades, o CIDEHUS, da Universidade de Évora, em Portugal. Lá vemos estudiosos de diferentes áreas e também pesquisadores linguistas envolvidos em diversos empreendimentos de estudo. A partir desse centro, acessa-se o CIDEHUS DIGITAL <<http://www.cidehusdigital.uevora.pt/#/>>.

O CIDEHUS DIGITAL é um repositório de dados, de acesso aberto, no qual se oferece um crescente volume de materiais de valor histórico, desde textos — especialmente manuscritos — até reconstruções digitais de mapas de cidades medievais, incentivando a sua reutilização em atividades de pesquisa e de ensino. Sua proposta é tornar o passado mais acessível por meio das tecnologias do presente e contribuir para o constante (re) fazer da ciência.

A despeito de maiores ou menores recursos e funcionalidades oferecidas a partir de acervos que hoje podemos acessar *on-line*, o movimento em direção à digitalização, que move e define as Humanidades Digitais, vem ocorrendo de um modo

bastante intenso, em diferentes bibliotecas de grandes centros europeus e norte-americanos. Poucas instituições têm, entretanto, condições de oferecer uma versão ou edição filológica junto com as imagens digitalizadas dos seus materiais que tenham o formato de textos escritos.

A partir do processamento de diferentes camadas da informação do texto original, transposto para o formato de uma imagem ou digitalização, necessitamos dessa informação acessível em diferentes formas. No caso do estudo linguístico, é necessário poder distinguir unidades de palavras e haver condições para que caracteres impressos sejam opticamente reconhecíveis. Nesse ponto, vemos, novamente, uma Filologia tradicional (conforme Banza e Gonçalves, 2013) avançando para o formato digital.

Mas, até que se tenha um texto antigo digitalizado disponível em um arquivo *on-line*, na sua grafia original e, se possível, também em sua versão em ortografia atualizada, toda uma série de etapas de trabalho humano árduo é necessária. Essas etapas envolvem as *camadas* antes citadas conforme descrevem Monte e Paixão de Sousa (2017). Isto é, frisamos, não há uma *mágica computacional* que acontece partindo-se do arquivo de imagem de um livro antigo. Apenas a digitalização acessível de um livro antigo, por si só, não atenderá todas as necessidades de estudos e pesquisas, ainda que seja um capital hoje indispensável para a conservação do patrimônio histórico documental.

Nesse sentido, partindo-se da digitalização de um texto antigo, uma das etapas mais importantes do trabalho é conduzida pelo linguista-filólogo – e terminólogo. Afinal, ele cuida da edição ou transcrição do texto para seu oferecimento em diferentes formatos, tornando-os enriquecidos também para historiadores. Com esse trabalho, pelo viés da Linguística Histórica, buscamos conservar a informação original e situar a leitura, no presente, de um documento antigo.

Dado o alto custo de se gerar a versão modernizada de um texto antigo, em geral, oferece-se, em muitos casos, apenas o texto conforme a imagem da grafia original, conservando-se caracteres e tipos gráficos. Felizmente, a maioria dos materiais digitalizados de obras impressas oferece a opção de consulta ao texto com o reconhecimento óptico de caracteres.

Por outro lado, importa registrar que há, sobre os diferentes tipos de processos de modernização de escrita de textos antigos, especialmente no âmbito da Linguística e da Filologia, toda uma série de discussões e detalhamentos metodológicos que não serão aqui reproduzidos. Para alguns, a versão modernizada do texto é uma deturpação aplicada à informação.

A despeito de polêmicas sobre como apresentar o texto antigo transcrito, o trabalho de edição de textos, no cenário *on-line* da atualidade, busca suporte nas indicações de Kabatek, autor que é uma referência internacional para a Filologia Digital e para o tratamento de *corpora* históricos em diferentes idiomas (Kabatek, 2005 e outros anos). Para o nosso caso, considerando o português, contamos também com orientações de pesquisadores brasileiros, linguistas e cientistas da Computação, que já têm trabalhado com esse tipo de fontes e procedido ao seu tratamento computacional (Paixão de Sousa e Trippel, 2006) para fins de seu oferecimento *on-line* em diferentes apresentações.

Ainda assim, é importante reconhecer o monumental esforço de conservação de patrimônios históricos, na parte da digitalização mais qualificada de acervos bibliográficos e de documentos antigos. Isso tem ocorrido em arquivos públicos e em diferentes repositórios institucionais do Brasil, cuidando-se de manuscritos e de textos em geral, tais como cartas pessoais, jornais, documentos administrativos, entre outros. Uma iniciativa bem próxima de nós é o Arquivo Público do Estado do Rio Grande do Sul (APERS), que já congrega profissionais, historiadores, linguistas, documentalistas e arquivistas, em torno do tema das Humanidades Digitais<sup>8</sup>.

Conforme já citado, nossa pesquisa centra-se na expansão de um *corpus* de textos impressos sobre Medicina do século XVIII, os quais já estão sendo transformados em digitalizações com maior detalhamento e com mais camadas de dados, começando pela obra de Semedo de 1707.

Nessas novas *camadas de dados*, explicadas por Monte e Paixão de Sousa (2017), trata-se de *manipular* e enriquecer o texto intensamente, sem em nada alterar o objeto original mostrado no texto digitalizado. Isso porque a imagem digitalizada do texto original fica sempre à disposição do leitor.

Este é, a nosso ver, o impacto mais claramente positivo da edição ou filologia virtual. Afinal, torna possível ampliar os conhecimentos ópticos de caracteres e de expressões e busca, já embutidos nos arquivos disponíveis em plataformas como *Google Books*, havendo a inserção, em novas versões do documento, de uma série de meta-dados e de conexões. Por exemplo, a partir de um arquivo de transcrições, é possível buscar apenas pelos verbos, os nomes de pessoas ou obter listas das terminologias mais empregadas ao longo do texto. Quanto aos diferentes recursos que podem ser acoplados a um *corpus*, vale conferir, por exemplo, as experiências do trabalho Cucatto (2012) com a extração semiautomática de elementos deverbiais em *corpora* antigos do Português.

Esse tipo de processamento do texto antigo, conforme vemos nas propostas das Humanidades Digitais, torna viável toda uma série de recursos de busca, tais como estatísticas de distribuição de palavras, obtenção automática de resumos por nuvens de palavras e mapeamento de temas mais recorrentes, comparações com outros textos por perfil vocabular, representações dinâmicas de conteúdo, entre outros. Tudo isso, num *corpus* digital, passaria a funcionar ao longo de todos os textos de uma base de dados, além de permitir conexões com outras bases (ou *corpora*), de arquitetura semelhante. Isso seria um passo rumo a uma base de conhecimento, conforme defendem Santos, Olival e Sequeira (2020) para estudos em torno de documentos históricos em *corpora* on-line.

### 3. Obras do corpus, materiais de apoio e ensaios com Semedo (1707)

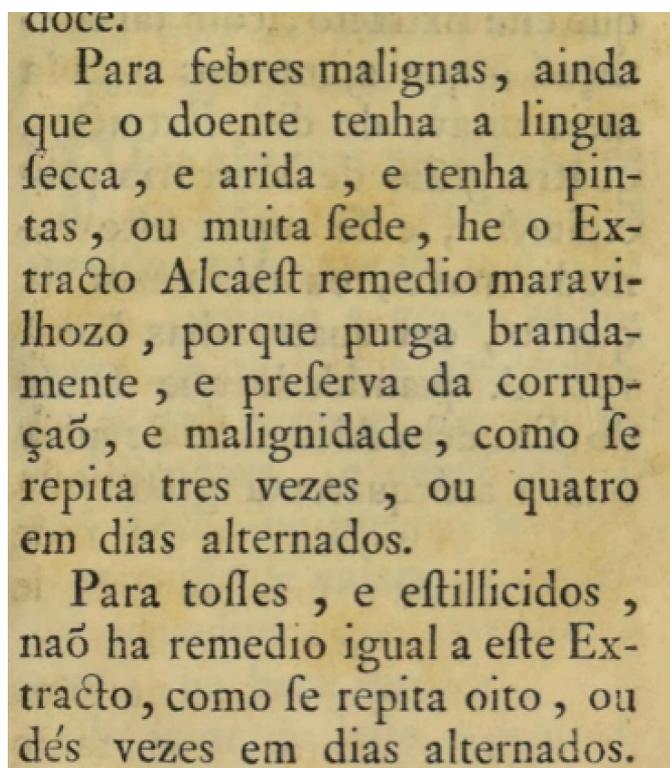
O ponto de partida para a organização do nosso *corpus* foi a relevância do autor e das suas obras para os estudos históricos sobre doenças e seus tratamentos em língua portuguesa. Conforme Lourenço (2016), esse conjunto de obras fornece

um «*frame de significação*» para a feição dos conhecimentos científicos da época e pode contribuir para o entendimento do contexto histórico em que se inserem. Assim, nosso *corpus* contempla as seguintes obras de J. C. Semedo, impressas durante o século XVIII:

1. Semedo, João Curvo. *Observações medicas doutrinaes de cem casos gravíssimos, que em serviço da pátria, & das nações estranhas escreve em língua portugueza, & latina*. Lisboa, 1707. 616 p. Disponível em Google Books
2. ———. *Atalaya da vida contra as hostilidades da morte; fortificada e guarnecida com tantos defensores, quanto são os remédios, que no decurso de cincoenta e oytto anos experimentou [...]*. Lisboa Occidental: na Officina Ferreyrenciana, 1720. 696 p. Disponível em Google Books e na Biblioteca Pública de Évora - Portugal
3. ———. *Polyanthea Medicinal. Noticias galenicis e chymicas*. 2.<sup>a</sup> Impressão. Lisboa: Officina de Antonio Pedroso Galram, 1704. 990 p. Disponível em Google Books  
———. *Polyanthea Medicinal. Noticias galenicis e chymicas*. 3.<sup>a</sup> Impressão. Lisboa: Officina de Antonio Pedroso Galram, 1716. 879p.  
———. *Polyanthea Medicinal. Noticias galenicis e chymicas*. 4.<sup>a</sup> Impressão. Lisboa: Officina de Antonio Predrozo Galram, 1727. 879 p.
4. ———. *Manifesto que o Doutor [...], médico morador em Lisboa faz aos amantes da saúde, e attentos às suas consciências*. Lisboa: Valentim da Costa Deslandes, 1706. 11 p. Disponível em Google Books
5. ———. *Proposta que o Doutor Joam Curvo Semmedo, Medico, morador em Lisboa, faz aos amantes da saúde, & consciências*. 8p. s/l. s/d. Disponível em: Biblioteca Nacional de Portugal, versão digital. <<http://purl.pt/15026>>.
6. ———. *Memorial de vários simpleses que da India Oriental, da América e de outras partes do mundo vem ao nosso Reyno para remédio de muitas doenças, no qual se acharão as virtudes de cada um, e o modo com que se devem usar*. 32p. s/l. s/d. Disponível em Biblioteca Digital Hispânica. Versão digital. <<http://bdh.bne.es>>.

As seis obras citadas, em suas diferentes edições, felizmente, encontram-se digitalizadas e disponíveis gratuitamente para *download* através da plataforma *Google Books* ou a partir de outras plataformas digitais universitárias. Como material complementar, visto que se trata de uma retomada e comentários às obras selecionadas, pretendemos também incluir a seguinte obra, como uma fonte de apoio e de estudos contrastivos. Essa obra foi produzida pelo herdeiro do autor:

Semedo, Manuel José Curvo. *Compendio dos segredos medicinaes, ou remedios curvianos que inventou, e compôs o Doutor Joaõ Curvo Semmedo [...] mandado imprimir per Manoel Jozè Curvo Semmedo actual manipulador, e administrador dos ditos remedios*. Lisboa: na Officina de Jozè de Aquino Bulhoens, 1783. 132p.



**FIGURA 3.** Página 69 do livro *Compendio dos segredos medicinaes, ou remédios curvianos (...)* de 1783

A figura 3 mostra um trecho da página 69 desse livro, cujo arquivo digital se encontra disponível em <<https://wellcome-collection.org/works/sm9uexmx>>.

A versão desse trecho reproduzido na figura 3, transformada na versão .TXT, somente texto, na qual vemos o reconhecimento de caracteres ativo é a que está na figura 4:

A partir do *download* das obras do *corpus*, em formato .PDF, é feita uma leitura inicial e panorâmica, de cada uma, para uma sistematização dos seus principais pontos e características. Nesta primeira etapa, tentamos reproduzir experiências de transformação semi-automatizada de arquivos de obras antigas impressas, obtidos originalmente em formato .PDF, que passarão a ser armazenados em diferentes formatos (cf. Hirohashi, 2004; Candido Jr. e Aluísio, 2009; Candido Jr., 2008). É também o momento da retomada necessária de experiências sobre a detecção de variantes da escrita de uma mesma palavra ao longo dos textos do *corpus* (cf. Giusti *et al.*, 2007). Afinal, conforme já se reconheceu com Candido Jr. (2008), em documentos antigos como os nossos, um mesmo redator pode escrever, num mesmo texto, uma mesma palavra com diferentes formas.

Para a parte de uma *leitura interpretativa*, para situar as obras do *corpus*, contamos com o auxílio de trabalhos que tenham tratado da história da Medicina ao longo do século XVIII e, especificamente, do legado de Smedo (como Lourenço, 2016). Recorremos também a livros de temática semelhante, em Português, publicados no mesmo período. Essas são as nossas fontes de apoio, apresentadas mais adiante.

A partir dessa sistematização inicial, o *corpus* de estudo, con-

Para Afma , he o Extra&o  
Alcaeíl maravilhozo remedio,  
com tal condição , que fe to-  
me dés , ou doze vezes em  
dias alternados, dando depois  
diíto todôs os dias em jejum  
quatro onças de agua de boita  
de boi diítillada no Mez de  
Maio, fobre duas colheres de  
xarope de hyíTopo.

**FIGURA 4.** Versão somente texto do livro *Compendio dos segredos medicinaes, ou remédios curvianos (...)* de 1783. Arquivo TXT disponível em <<http://cesimadigital.pucsp.br/handle/bcd/1660>>

forme fixado, é contrastado com diferentes materiais, que não sejam específicos sobre o tema *Doenças e seus tratamentos*, mas que tenham sido igualmente produzidos em período aproximado no século XVIII. Esses materiais são denominados, em Linguística de Corpus (cf. Berber Sardinha, 2000 e 2004), de *corpus* de referência e servem para uma detecção de especificidades do texto e do vocabulário em foco. Para essa fase do trabalho, o conteúdo de cada um dos arquivos das obras é transformado em arquivo no formato somente texto (.TXT) o qual é gerado a partir do texto copiado bruto do arquivo PDF original. Esse processo já foi por nós testado, com bom rendimento, no trabalho de Finatto, Quaresma e Gonçalves (2018).

A próxima etapa do trabalho envolve a geração de transcrições filológicas de cada obra com a respectiva versão atualizada. Essa é uma etapa de longa duração. Para ela, já contamos com o auxílio de estudantes de Letras da UFRGS, alunos das disciplinas Linguística Histórica e Terminologia.

A etapa final será a revisão, a reunião e a disponibilização pública *on-line* do *corpus* propriamente dito e do material a ele conexo. Nesse ponto, são geradas representações de conteúdo produzidas com ferramentas computacionais. Mais adiante, trazemos um exemplo de um tipo dessas representações.

### 3.1. Fontes de apoio para as edições do texto

Conforme apontam Santos, Olival e Sequeira (2020), o contexto histórico é essencial na área das Humanidades, sendo um aspecto fundamental nesse tipo de trabalho. Assim, como fontes de apoio, para uma boa compreensão do universo de significação dos textos de Smedo, temos de recorrer a diferentes fontes, que acabam relacionadas ao *corpus* à medida que produzimos suas representações de conteúdo. Nesse sentido, é preciso contextualizar as versões de edição do *corpus*, de modo que se consiga compreender razões para o que observamos nos textos, especialmente quanto ao emprego de terminologias, de

conceituações, de expressões e de associações de palavras mais e menos recorrentes.

Exemplo de estudo sobre a história das doenças/tratamentos médicos vemos no trabalho de Grmek (1995), intitulado *Déclin et émergence des maladies*. Nele vemos referência, por exemplo, de que o século XVIII teria sido o século da Nosologia, enquanto o XIX seria o século da Patologia. As referências desse autor, ainda que se apliquem a textos de pesquisadores eruditos e médicos, registram as temáticas recorrentes sobre doenças que hoje conhecemos como varíola, peste bubônica, tuberculose, sífilis, entre outras, temas presentes ao longo das diferentes obras de Semedo.

A preocupação de Grmek (1995) foi a de historiar o conhecimento científico-médico, não sendo sua perspectiva filológica ou linguística, de modo que muitos dos nomes antigos para essas enfermidades não são resgatados. Esse seria o caso, por exemplo, do termo **bexigas/(s)**, que corresponde, pelo que depreendemos de textos dos séculos XVIII e de dados de dicionários antigos, *grosso modo*, à doença conhecida modernamente pelo nome de **varíola**.

Serão também apoio importante as fontes de pesquisa histórica, tais como o artigo *O homem português perante a doença no século XIII: atitudes e receituário*, de Eugénio dos Santos (Santos, 1984), publicado em 1984 na *Revista da Faculdade de Letras*. Esse artigo nos indica fontes documentais do que se denomina de *Medicina Popular*, cujas práticas tendem a ser criticadas, apoiadas ou apenas referidas por Semedo.

Para uma visão de um repertório referencial de nomes de doenças e de seus tratamentos, Santos (1984) cita, justamente, as obras *Polyanthea medicinal, noticias galenicis e chymicas*, Lisboa Occidentia, 1741, e a obra *Observações medicas doutrinaes de cem casos gravíssimos...*, Lisboa, 1727, de João Curvo Semedo e, ainda, a obra *Pharmacopea tubalense chimico-galenica*, de Manuel Rodrigues Coelho. Recorremos a esses materiais na fase de revisão de literatura da nossa pesquisa e também para a construção do já citado «*frame* de significação».

Outra fonte importante citada por Santos é a obra do frei-boticário João de Jesus Maria, da Congregação de São Bento, responsável pela botica do Mosteiro de Santo Tirso, que editou, em 1772, a *Farmacopéia dogmática, médico-química e teórico-prática*, publicada na cidade do Porto, na Oficina de Antonio Álvares Ribeiro Guimarães, tendo sido esta a quinta farmacopeia publicada em Portugal.

Entre outros materiais de apoio, cabe destacar os estudos da obra de Garcia de Orta (*Colóquios dos Simples e Drogas da Índia*, 1563) realizados por Murakawa (2005). Murakawa trabalhou bastante esta fonte, a qual se tornou uma referência pioneira para obras do gênero produzidas mais tarde e também para dicionaristas que se ocuparam de seu vocabulário.

Também a obra de Luís Gomes Ferreira (*Erário Mineral*, 1735), que cita obras e orientações de Semedo, é um texto importante a pesquisar, considerando-se especialmente os trabalhos de Júnia Furtado, da UFMG, que editou a obra e a oferece em ortografia atualizada. Essa pesquisadora da UFMG tem toda uma série de trabalhos em que retoma obras relacionadas ao tema dos tratamentos de Saúde no século XVIII. Esse material

também foi objeto de estudo de Gonçalves (2013), justamente nos aspectos lexicais do português que mais nos interessam.

Além do apoio direto de fontes diversas do período e de pesquisas histórico-conceituais relacionadas ao tema em foco, é importante registrar toda uma produção de teses e de dissertações do Brasil, promovida pela instituição Casa de Oswaldo Cruz – FIOCRUZ e do seu Programa de Pós-Graduação em História das Ciências e da Saúde<sup>9</sup>. Temos, a partir desse PPG brasileiro, importante apoio para uma melhor contextualização histórica das obras de Semedo.

Além desses recursos históricos sobre Doenças/Enfermidades de pesquisas da *Casa de Oswaldo Cruz* e de diferentes relatos sobre conhecimentos médicos do século XVIII e do uso de dicionários e obras de referência, utilizamos *corpora* históricos do português, com preferência para os disponíveis *on-line* com acesso gratuito conforme nosso levantamento (Finatto, 2018). Entre os acervos localizados, um recurso valioso a destacar, mais uma vez, pois permite buscas de palavras com ferramentas computacionais, é o *Corpus Histórico do Português Tycho Brahe (Tycho Brahe Parsed Corpus of Historical Portuguese)* (Galves e Faria, 2010). Esse corpus conta com 68 textos completos disponibilizados para consulta, num total de 2.873.916 palavras, o que é uma dimensão considerável de dados. Seus textos são oferecidos, inclusive, anotados semântica e sintaticamente (Paixão de Sousa e Trippel, 2006). Para o período do século XVIII, traz 13 referências documentais (08 fontes, de 1702 a 1733, e 05 fontes, de 1750 a 1799, com concentração de 03 fontes em 1799), o que será útil para contraste com as obras de Semedo.

Por fim, como fontes de apoio, importa ainda dizer algo sobre os dicionários históricos têm nos ajudado a contemplar o léxico e cultura do século XVIII. Especialmente úteis são os dicionários de Morais Silva (Silva, diferentes anos, a partir de 1789, ver abaixo), Bluteau (1712-1728) e Silva Pinto (1832) e o próprio *Dicionário Histórico do Português do Brasil – DHPB* (Murakawa, 2014), assim que disponível para consulta. Entre vários dicionários, a obra de Bluteau (Bluteau, Raphael. *Vocabulário portuguez e latino*. Coimbra: Collegio das Artes da Companhia de Jesus, 1712-1728. 10 v.) é uma fonte indispensável.

Assim, obras importantes são:

Silva, Antonio. de Morais. *Diccionario da língua portugueza*. Lisboa: Typographia Lacerdina, 1813. 2 v.

Silva, Antonio. de Morais. *Diccionario da língua portugueza*. Lisboa: Officina de Simão Thaddeo Ferreira, 1789. 2 v. Silva, Antonio. de Morais. *Grande dicionário da língua portuguesa*. 10.<sup>a</sup> ed. Orgs. Augusto Moreno, Cardoso Junior e José Pedro Machado. Lisboa: Editorial Confluencia, 1949-1959. 12 v. Silva, Antonio. de. Morais. *Diccionario da língua portugueza*. 6.<sup>a</sup> ed. Lisboa: Typographia de Antonio José da Rocha, 1858. 2 v.

Silva, Antonio. de Morais. *Diccionario da língua portugueza*. 9.<sup>a</sup> ed. Lisboa: Empreza Litteraria Fluminense de Santos, Vieira & Commandita, [18-]. 2 v.

Silva Pinto, Luiz Maria da. *Diccionario da Lingua Brasileira por Luiz Maria da Silva Pinto, natural da Provincia de Goyaz*. Na Typographia de Silva, 1832.

Vieira, Frei Domingos. *Grande dicionario portuguez ou thesouro da lingua portugueza*. Porto: Editores Ernesto Chardron e Bartholomeu H. de Moraes, 1871-1874. 5 v.

A obra de Machado Filho (2019), [Machado Filho, Américo Venâncio Lopes (2019): *Novo Dicionário do Português Arcaico ou Medieval*. Amazon Livros, 697p.] também oferece importante suporte, sobretudo quando detectamos unidades não reperitórias nos léxicos antes citados.

### 3.2. Representações de conteúdo com apoio computacional

Para a representação do conteúdo e do perfil terminológico do *corpus*, além dos resultados gerados com sistemas e princípios da Linguística de Corpus (cf. Biber, Conrad e Reppen, 1998), via panorama de distribuições de palavras, lançamos mão de diferentes ferramentas computacionais. Essas ferramentas, em sua maioria, trabalham com um *input* de texto, geralmente no formato .TXT. Isso significa que é preciso um preparo inicial do arquivo, evitando-se que sejam considerados trechos com problemas de reconhecimento de caracteres. Esses problemas são bastante recorrentes.

Além do arquivo em formato .TXT, que seria uma camada inicial da edição, temos os arquivos que correspondem às transcrições no formato semidiplomático. Essas transcrições são armazenadas em diferentes formatos de editores de texto mais tradicionais, como Word – formatos .DOC ou .DOCX. Uma ferramenta que lida com textos nesse formato é o sistema SOBEK.

SOBEK é uma ferramenta mineradora de textos (Reategui *et al.*, 2018) que produz uma representação de conteúdo do texto em termos de grafos com as palavras nele empregadas. Está programado para lidar com textos em português e inglês, mas não foi desenvolvida para lidar com ortografia e tipografia antiga. Desenvolvida por nossos colegas da Ciência da Computação na UFRGS, assim, funciona a partir de dados da distribuição de palavras. Sendo um recurso comparável a um «gerador de ontologias», é acessado gratuitamente em: <<http://gtech.nuvm.ufrgs.br/sobek/index.html>>.

Tal como resgatam Minghelli e Chishman (2012), citando Guarino (1998), utiliza-se o termo *ontologia*, com inicial maiúscula, para referir um ramo da Filosofia, no qual se estudam tipos de estruturas dos objetos, propriedades, eventos, processos e relacionamentos em todas as áreas da realidade. Nesse âmbito, *ontologia* corresponde a um sistema de categorias que representaria certa visão do mundo. Por outro lado, na Computação, desde a década de 90, o termo *ontologia* tende a referir um artefato de representação de conteúdo, formado por um vocabulário específico, o qual descreve uma realidade a partir de um conjunto de frases, presentes em dado *corpus* textual ou conjunto de documentos.

Conforme já apontaram Ferreira, Lopes e Vieira (2012), a produção de ontologias de domínio, isto é, daquelas que se ocupam de conjuntos de documentos que espelham a produção de determinadas áreas de conhecimento, mormente as científicas ou técnicas, tornou-se um ramo relevante da área de Processamento de Linguagem Natural (PLN). De modo semelhante, re-

conhece-se essa relevância também para a área dos Estudos da Linguagem e/ou da Linguística (Minghelli e Chishman, 2012), para representação de conteúdo, para a descrição linguística em diferentes níveis do funcionamento do léxico e da configuração conceitual e semântica.

Nesse cenário, o sistema SOBEK analisa as palavras recorrentes em um texto, sendo capaz de identificar termos relevantes e representá-los graficamente. A relevância, vale frisar, é de natureza estatística, estabelecida a partir de diferentes cruzamentos entre o texto de entrada, frequência das diferentes palavras que perfazem o texto e suas conexões, e uma base comparativa de padrões de distribuição de palavras de outros textos em português – que são também textos da atualidade.

A título de exemplo, adiante, na figura 5, vemos um texto-fonte do nosso *corpus* sendo submetido à ferramenta SOBEK na sua versão que se acessa e utiliza *on-line*. A figura 6 traz a saída do sistema e a ontologia gerada.

O texto sendo processado é parte da obra intitulada *Observações medicas doutrinaes de cem casos gravíssimos* [...], impressa em Lisboa, em 1707. A obra conta com 635 páginas e é constituída, no miolo temático, por 101 segmentos de texto. Cada segmento corresponde a uma *observação* de caso/paciente com comentários diversos do autor-médico sobre os tratamentos por ele desenvolvidos. Cada *observação*, assim, é um relato de procedimentos adotados, indicação de medicamentos, conselhos, críticas de procedimentos, como também traz remissões, nas laterais das páginas, a obras latinas e a conceitos de autores importantes na Medicina da época.

O segmento submetido ao sistema é um trecho, na sua ortografia original, que trata de um caso com uma mulher que recém havia dado à luz, a *Observaçam XCII*. O trecho completo desta *Observação* 92 tem 1.382 palavras conforme a ferramenta indica. Para uma ideia do tipo de texto sendo processado, segue abaixo um excerto da *Observação* 92:

#### OBSERVAÇAM XCII

De huma Senhora, a quem estando sobre parto faltou a evacuação do puerperio, & logo lhe deu febre acompanhada com muitos cursos, & de tudo a livre dando lhe oito sangrias nos pés.

I. Huma das cousas que faz grande embaraço aos Medicos principiantes, he como se não de aver quando saó chamados para alguma mulher parida, a quem achão com febre, camaras, & faltas de purgação lochial; porque como tem lido em varios lugares de Galeno, (I.) que avendo febre juntamente com camaras, não sangrem, porque se sangrarem, & as camaras continuarem, cahirão os doentes em tal fraqueza, que morrerão: à qual sentença definitiva se ajunta, que a natureza não poderá suportar duas evacuações juntas sem perigo da vida, isto os acovarda de maneira, que não ousão a sangrar; por outra parte estão vendo muitas vezes, que a febre he grande, & que para as grandes febres diz o mesmo Galeno, (2.) & a experiencia o mostra, que não ha remedio mais presenteaneo, & efficaz que as sangrias; outra parte estaó vendo a falta da purgação do puerperio, & que esta por nenhum

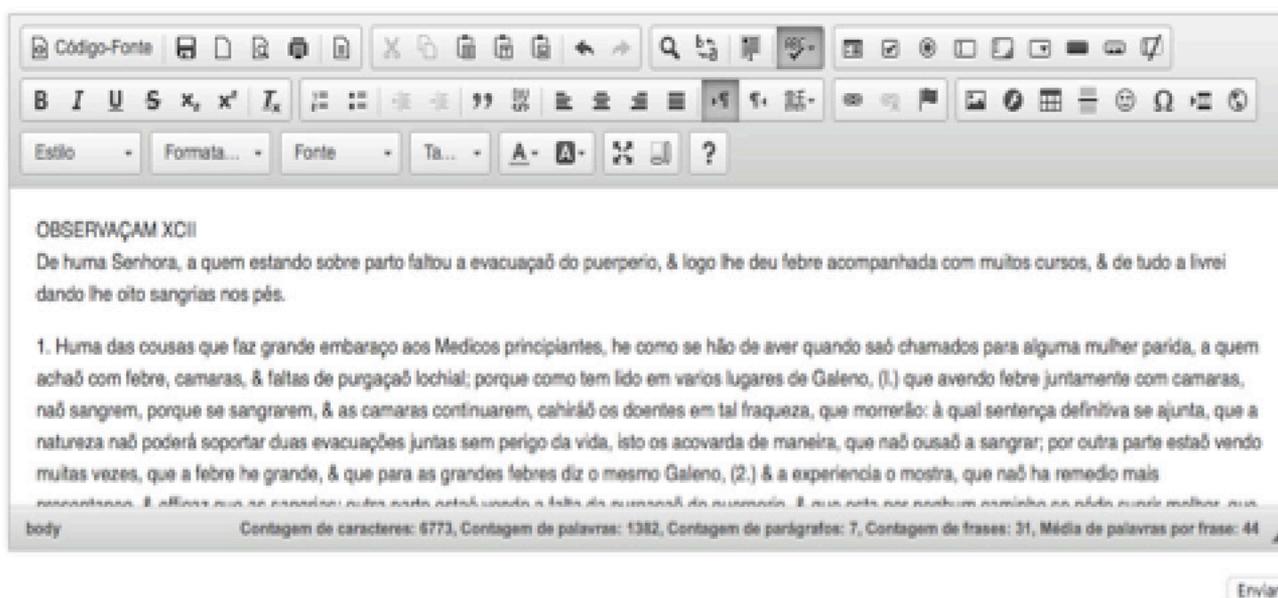


FIGURA 5. Trecho do livro de Semedo (1707) submetido ao sistema SOBEK, na grafia original

caminho se póde suprir melhor, que com as sangrias dos pés: daqui vem o embaraço, & perplexidade no que haõ de obrar; porque para naõ fazer remedio algum, sobre arguir ignoramcia, he impiedade, porque he deixar a natureza de deseparada, & entregue nas mãos do inimigo; para applicar purga, he impropria sobre parto, aonde ha muitas camaras acompanhadas com febre; para sangrar, ha grande temor de perder forças sem alivio da doença, & com risco da vida: para que pois o Medico se naõ veja embaraçado, & confuso, direi nesta Observaçao o que tenho feito em casos semelhantes com prospera fortuna.

Na representação gerada, que pode ser visualizada na figura 6, os grandes nodos representam os termos mais frequentes e as conexões representam as relações entre eles. A partir de experimentos como esse, efetuamos contrastes com os resultados obtidos por outras ferramentas de Linguística de Corpus, como o AntConc (Anthony, 2014) e o TermoStat (Drouin, 2003), que nos apontaram itens mais frequentes e potenciais terminologias de domínio. Nessas comparações, importa considerar que ambas ferramentas foram construídas para dar conta de textos da atualidade. A despeito disso, fica claro que se pode trabalhar

com o material na ortografia antiga, ainda que a origem de seus sistemas não tenha considerado esse tipo de *input*.

Na figura 7, a seguir, temos o mesmo sistema SOBEK, versão para *download*, processando não uma, mas todo um conjunto de Observações em que ocorre a palavra FEBRE. Vê-se que o sistema está *rodando*, relativamente bem, mesmo com o texto com a grafia original. Um segmento transcrito dos textos, conforme inseridos, em bloco, no SOBEK, está reproduzido na figura. Os números entre colchetes indicam as quebras de linhas de cada página.

#### Observaçam XLII

[1] De huma febre cansada de enchimento do estomago ; para [2] remedio da qual se deraõ vinte sangrias, & por isso [3] degenerou a febre em maligna de tam venenosa qualidade , [4] que o doente chegou a estar unguido , & certamente [5] morreria se eu lhe naõ acudira dandolhe o meu [6] Bezoartico ; com cuja virtude solutiva , & cardiaca [7] foi lentamente purgando os humores damonofos , & rebatendo [8] a qualidade venenosa , & por este caminho o [9] livre da morte estando quaifi metido na sepultura.

[10] Naõ averá Medico tam falto de ciência, [11] que ig-

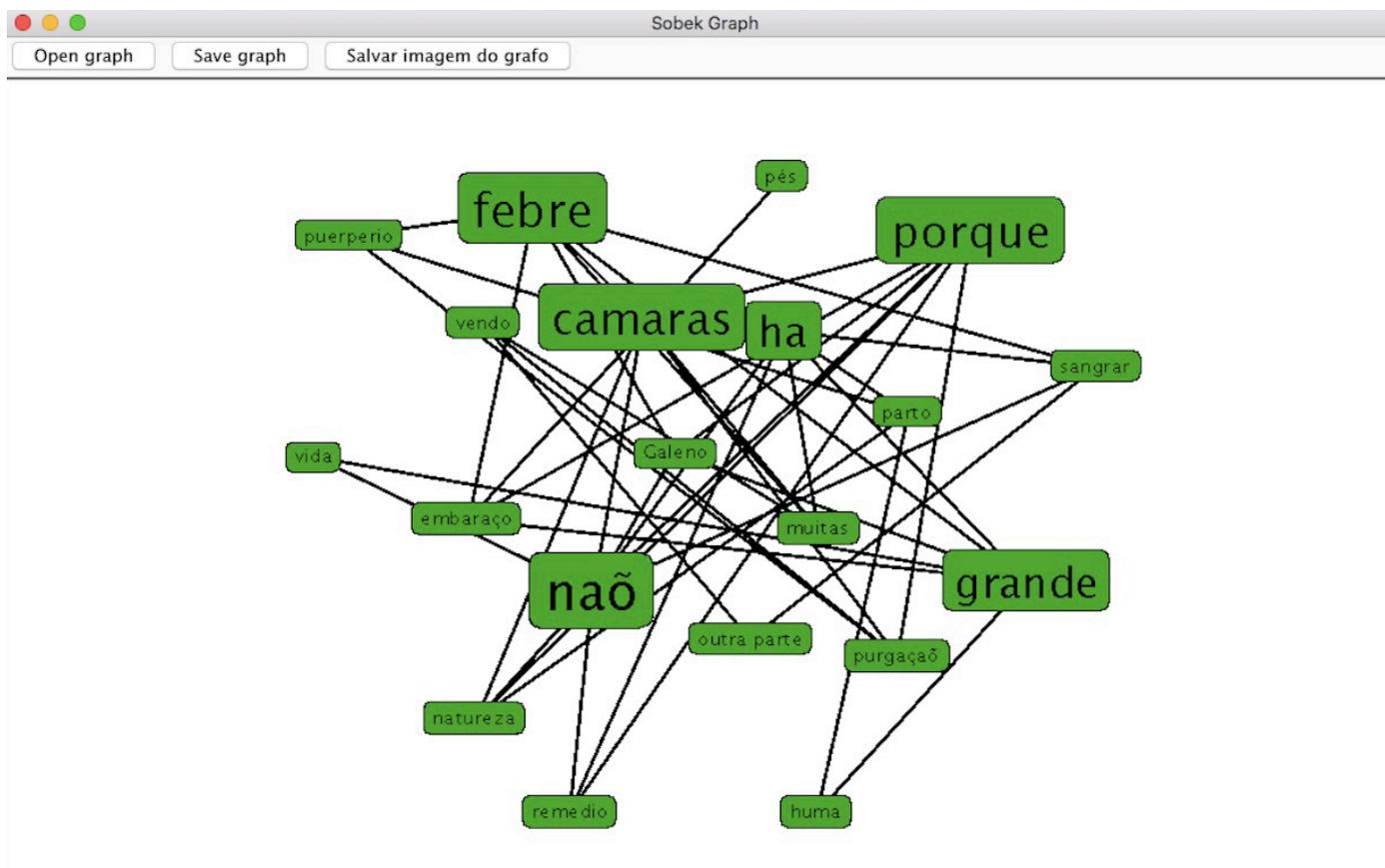


FIGURA 6. Uma representação de conteúdo da Observação 92 de Semedo (1707) pelo sistema SOBEK versão para download

nore o muito que Galeno, & [12] todos os Doutores reprovão as sangrias nas febres & [13] doenças, que procedem de cruexas, ou enchimento [14] do estomago: comtudo não fei com que cegueira, [15] ou fado infelice desprezão alguns homens os preceitos [16] dos Oraculos da Medicina, mandando sangrar em [17] todas as febres, sem diferença, nem distinção das [18] causas dellas, nem dos temperamentos, ou forças dos[19] doentes;

#### 4. Considerações finais e perspectivas

No vasto e novo território internacional de conhecimentos e de compartilhamentos de experiências, insere-se a investigação que estamos levando adiante. Com ela, buscamos inscrever a Terminologia Diacrônica baseada em *corpus* em meio aos estudos de Terminologia e de Filologia. O nosso é um estudo linguístico-filológico e histórico, mas também computacional, sobre um conjunto de textos antigos sobre temas de Medicina, produzidos no século XVIII, em português, e transpostos para o formato digital.

Considerando que o nosso *corpus* está sendo reunido em função da sua autoria, cobertura temporal e temática, em que pesem diferentes critérios para o seu desenho (cf. Atkins, Clear e Ostler, 1992) um ponto de partida importante já é o reconhecimento prévio, em estudos anteriores e em obras de referência, de um vocabulário de enfermidades e de expressões associadas

ao tema do século XVIII, cuja presença/utilização estamos examinando. Para tanto, tomamos como orientação os trabalhos de Temmermann (2000, 2001 e 2004), que nos mostram exemplos de acompanhamento histórico de unidades de compreensão e de terminologias conforme empregadas em textos de uma mesma especialidade, ao longo do tempo.

Um exemplo de levantamento prévio desse vocabulário e do modo de apresentação dessas unidades ou construções associadas a doenças e seus tratamentos tem-se em Murakawa (2013), que encontrou cerca de 112 unidades diferentes em apenas um documento, o relato *Prodigiosa Lagoa*, de J. C. Miranda, de 1749, documento com apenas 27 páginas. Desse estudo, além de uma lista de itens, é possível aproveitar toda uma rede ou *frame* semântico associado ao campo de enfermidades e seus tratamentos no cenário lusitano do século XVIII. Também serão úteis os trabalhos de Barros (2005) sobre dicionários e obras antigas de Medicina.

Conforme Berber Sardinha (2004: 40), «de um modo geral, a padronização é a regularidade expressa na recorrência sistemática de unidades co-ocorrentes de várias ordens (lexical, gramatical, sintática, etc.)». Para que sejam definidos os padrões de uma palavra ou expressão, faz-se necessário averiguar as outras palavras e estruturas frequentemente associadas a ela que, de alguma forma, refletem o seu significado.

Isso representa uma espécie de *diga-me com quem anda e te direi quem és* léxico-gramatical. Ou seja, é por isso que John Firth postulou que *se conhece uma palavra pelas companhias*

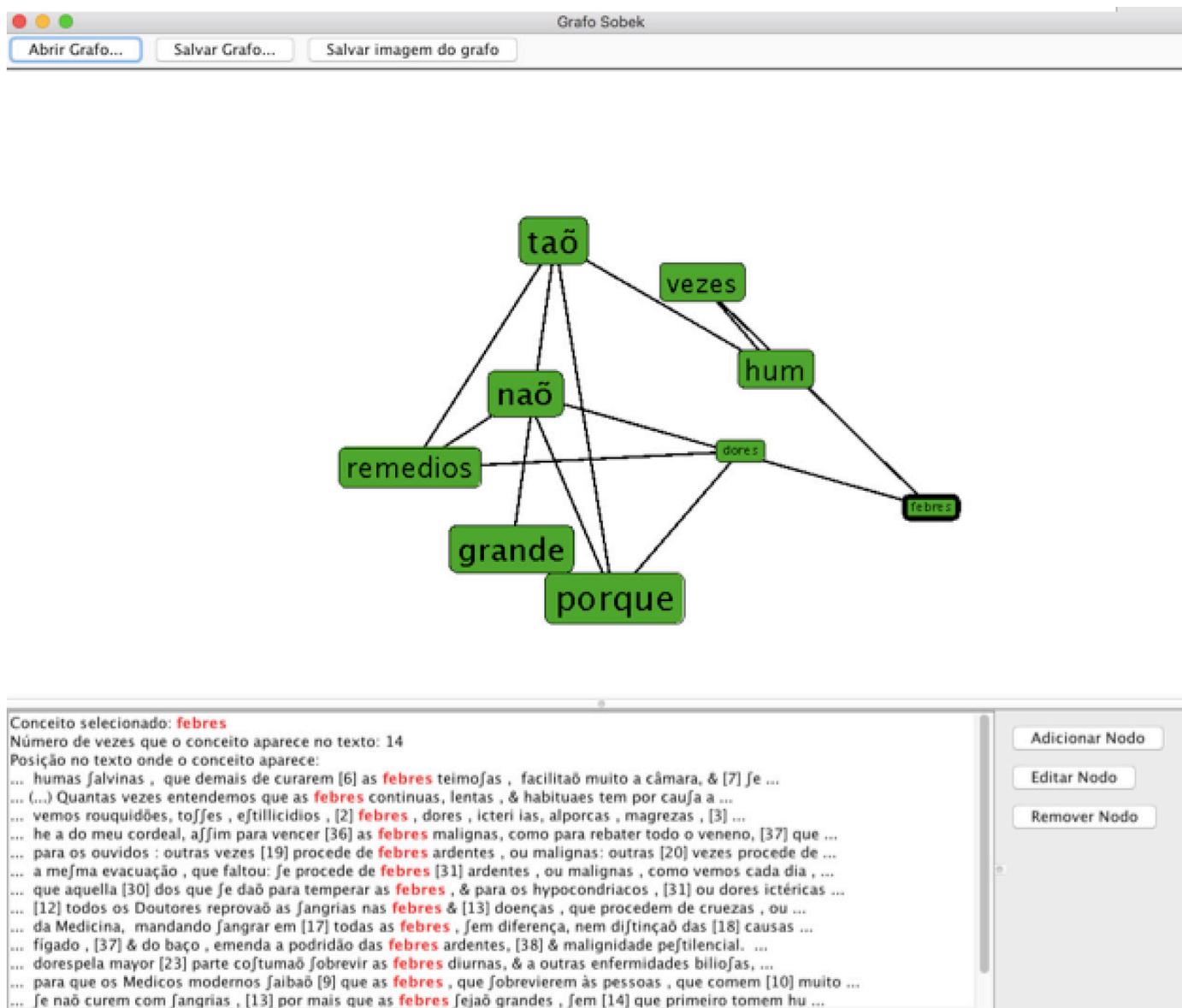


FIGURA 7. Processamento de diferentes observações de Smedo (1707), com o sistema SOBEK, na versão para download

com as quais ela anda (Firth, 1957: 179). Isso é que também pretendemos descrever no nosso *corpus* de documentos sobre doenças e seus tratamentos do século XVIII.

Além desse aspecto terminológico do vocabulário, pretende-se também, com auxílio de ferramentas de PLN, oferecer uma descrição sobre os verbos desse domínio, conforme empregados nas obras de Smedo que integram o *corpus*. Isso será realizado aproveitando-se, entre outras experiências, a proposta de Alves, Chishman e Quaresma (2007) para construção de ontologias. Com elas, temos produtos semelhantes a representações gráficas, geradas automaticamente, que ilustram um mapeamento de conexões entre os conceitos (representados por palavras) em um texto.

Nosso estudo, frisamos, pela natureza do livro digitalizado, é feito com apoio computacional, conforme se reconhece e discute em âmbito internacional e também já entre nós, no Brasil, no trabalho de Freitas (2017). Essa autora alerta para não esquecer que o dado digital nunca será o dado original, no seu

contexto, sendo dele mais uma representação do que qualquer outra coisa.

Nesse sentido, vale já frisar que ter uma obra já digitalizada, «fotografada no todo», é apenas uma primeira etapa de uma série de outras tantas que são necessárias para que o seu conteúdo possa ser devidamente explorado e usufruído em um ambiente *on-line*. De nossa parte, esperamos poder compartilhar um material de base, devidamente enriquecido, para quem se interesse pelo tema da linguagem médica em português conforme foi sendo constituída através dos tempos.

### Notas

1. Mais detalhes do evento em: <<https://8cihilie.sp.unipi.it>>. Livro de resumos em: <<https://8cihilie.sp.unipi.it/wp-content/uploads/2019/08/libro-degli-abstract-2.pdf>>.
2. Projeto TEXTECC: <<http://www.ufrgs.br/textecc/terminologia/index.php>>.

3. Quanto ao termo *sublinguagem*, vale situar as concepções de Hoffmann (2015<sup>a</sup> e 2015<sup>b</sup>). Para esse autor (Hoffmann, 2015<sup>a</sup>: 89-92), as variedades de uso de uma dada língua, como os socioletos, correspondem às linguagens especializadas. Trata-se de um subsistema ou variedade, cujo grupo de falantes pode ser identificado com determinadas camadas sociais estudadas pela Sociologia. As especificidades das linguagens especializadas devem, então, ser avaliadas no sentido de se verificar, principalmente, até que ponto elas permitem reconhecer seus usuários como representantes de uma determinada especialidade e, ao mesmo tempo, como integrantes de um determinado grupo social.
  4. Universidade do Minho, Portugal. Programa de Mestrado em Humanidades Digitais. Disponível em <<https://www.ilch.uminho.pt/pt/Ensino/Paginas/Mestrado-em-Humanidades-Digitais.aspx>>.
  5. Universidade Federal Rural do Rio de Janeiro. Programa de Mestrado em Humanidades Digitais. Disponível em <<https://www.dcc.ufrj.br/ppgihd/>>.
  6. Universidade Federal Rural do Rio de Janeiro. Pesquisa em Mineração de Dados. Disponível em <<https://www.dcc.ufrj.br/ppgihd/index.php/linhas-de-pesquisa/>>.
  7. Laboratório em Rede de Humanidades Digitais, iniciativa do Grupo de Pesquisa Informação, Memória e Sociedade (IMES). Disponível em <<http://www.larhud.ibict.br/sobre/>>.
  8. Arquivo Público do Estado do Rio Grande do Sul. Mais detalhes em: <<https://www.apers.rs.gov.br/inicial>>.
  9. Programa de Pós-Graduação em História das Ciências e da Saúde. Disponível em <<http://www.ppghcs.coc.fiocruz.br/index.php/br/>>.
- Barros, Lidia Almeida (2005): «O discurso terminográfico na obra De Medicina Brasiliensi (1648), de Guilherme Piso», *Tradterm*, 11: 255-307. <<https://doi.org/10.11606/issn.2317-9511.tradterm.2005.49690>> [consulta: 06.VIII.2020].
- Benveniste, Émile (1989): *Problemas de Linguística Geral 11*. Campinas: Pontes.
- Berber Sardinha, Antonio Paulo (2000): «Linguística de Corpus: Histórico e Problemática», *DELTA*, 16(2): 323-367. <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44502000000200005](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005)>.
- Berber Sardinha, Antonio Paulo (2004): *Linguística de Corpus*. São Paulo: Manole.
- Biber, Douglas; Susan Conrad e Randi Reppen (1998): *Corpus linguistics – Investigating language structure and use*. Cambridge: Cambridge University Press.
- Cabré, Maria Teresa (2005): «A Terminologia, uma disciplina em evolução: passado, presente e alguns elementos de futuro», *Debate Terminológico*, 1, s. p. <<https://seer.ufrgs.br/ri-term/article/view/21286>>.
- Cameron, Helena Freire; Maria Filomena Gonçalves e Paulo Quaresma (2020): «Linguistic and orthographical classic Portuguese variants Challenges for NLP», em *PROPOR Workshp: Digital Humanities and Natural Language Processing*, Évora, 1: 43-48. <<https://dblp.org/db/conf/dhandnlp-ws/dhandnlp-ws2020.html>>.
- Candel, Danielle e François Gaudin (eds.) (2006): *Aspects diachroniques du vocabulaire*. Mont-Saint-Aignan: Publications des Universités de Rouen et du Havre.
- Candido Junior, Arnaldo (2008): *Criação de um ambiente para o processamento de corpus de Português Histórico*. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) – São Carlos: Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.
- Candido Junior, Arnaldo e Sandra Maria Aluísio (2009): «Building a Corpus-based Historical Portuguese Dictionary: Challenges and Opportunities», *Revue TAL*, 50: 73-102.
- Carneiro, Raphael Marco Oliveira (2016): *Discurso literário de fantasia infantojuvenil: proposta de descrição terminológica direcionada por corpus*. Dissertação (Mestrado em Estudos Linguísticos). Uberlândia: Universidade Federal de Uberlândia.
- Cucatto, Livia Aluisi (2012): «XTRADEV: um sistema de extração semiautomático de deverbais em corpus do português histórico e contemporâneo», em *Anais do X Encontro do CELSUL – Círculo de Estudos Linguísticos do Sul UNIOESTE*, 1-10. <[http://www.celsul.org.br/Encontros/10/completos/xcelsul\\_artigo%20\(123\).pdf](http://www.celsul.org.br/Encontros/10/completos/xcelsul_artigo%20(123).pdf)>.
- Drouin, Patrick (2003): «Term extraction using non-technical corpora as a point of leverage», *Terminology*, 1 (9): 99-115.
- Dury, Pascaline (2013): «Que montre l'étude de la variation d'une terminologie dans le temps: quelques pistes de réflexion appliquées au domaine médical», *Debate Terminológico*, 9: 2-10.
- Dury, Pascaline e Aurélie Picton (2009): «Terminologie et

### Referências bibliográficas

Alves, Isa Mara da Rosa; Rove Luiza de Oliveira Chishman e Paulo Quaresma (2007): «Verbos do domínio jurídico: uma proposta de organização ontológica com vistas ao PLN», *Veredas*, 6: 55-65.

Anthony, Lawrence (2014): *AntConc (Version 3.4.3) [Computer Software]*. Tokyo: Waseda University. <<http://www.laurenceanthony.net/>>.

Atkins, Sue; Jeremy Clear e Nicholas Ostler (1992): «Corpus design criteria», *Literary and Linguistic Computing*, 7: 1-16.

Aubert, Francis (1996): *Introdução à metodologia da pesquisa terminológica bilíngue*. São Paulo: Humanitas.

Banza, Ana Paula e Maria Filomena Gonçalves (2018): *Roteiro de História da Língua Portuguesa*. Évora: UNESCO Chair in Intangible Heritage and Traditional Know-How: Linking Heritage. <[https://dspace.uevora.pt/rdpc/bitstream/10174/22196/1/Roteiro\\_de\\_História\\_da\\_L%C3%A-Dngua\\_Portuguesa.pdf](https://dspace.uevora.pt/rdpc/bitstream/10174/22196/1/Roteiro_de_História_da_L%C3%A-Dngua_Portuguesa.pdf)>.

Barros, Lidia Almeida (2004): «O primeiro dicionário médico do Brasil», *Cadernos de Estudos Linguísticos*, 1 (46): 21-42.

Barros, Lidia Almeida (2005): «Elementos de historia de la terminografía médica en Brasil», *Revista de Lexicografía*, 11: 7-37.

- diachronie: vers une réconciliation théorique et méthodologique?», *Revue française de linguistique appliquée*, XIV: 31-41.
- Esperandio, Isabela Beraldi (2015): *Legendas de seriadados de tema sobrenatural: uma abordagem terminológica para tradutores*. Dissertação (Mestrado em Estudos da Linguagem) – Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS). Porto Alegre: UFRGS.
- Faulstich, Enilde (2001): «Aspectos de Terminologia geral e Terminologia variacionista», *TradTerm*, 7: 11-40.
- Finatto, Maria José Bocorny (2001): *Definição terminológica: fundamentos teórico-metodológicos para sua descrição e explicação*. Tese (Doutorado em Estudos da Linguagem, Teorias do Texto e do Discurso) – Instituto de Letras. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Finatto, Maria José Bocorny (2003): «A definição de termos técnico-científicos no âmbito dos estudos de terminologia», *Revista de Estudos da Linguagem*, 1 (11): 197-222.
- Finatto, Maria José Bocorny (2018): «Corpus-amostra português do século XVIII: textos antigos de medicina em atividades de ensino e pesquisa», *Domínios da Linguagem*, 1 (12): 435-464.
- Finatto, Maria José Bocorny; Maria Filomena Gonçalves e Paulo Quaresma (2018): «Portuguese Corpora of the 18th century: old Medicine texts for teaching and research activities», em *Proceedings of the Conference on Language Technologies & Digital Humanities*, 1: 114-120. <[http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018-Jose-et-al\\_Portuguese-Corpora-of-the-18th-century-old-Medicine-texts-for-teaching-and-research-activities.pdf](http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTDH-2018-Jose-et-al_Portuguese-Corpora-of-the-18th-century-old-Medicine-texts-for-teaching-and-research-activities.pdf)>.
- Firth, John Rupert (1957): *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Francis, Winthrop Nelson (1992): «Language Corpora BC», em Jan Svartvik (ed.): *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm*, pp. 17-32.
- Freitas, Cláudia (2017): «Estudos linguísticos e Humanidades digitais: corpus e descorporificação», *Gragoatá*, 44 (22): 1207-1227. <<https://periodicos.uff.br/gragoata/article/view/33556>> [consulta: 03.VIII.20].
- Fromm, Guilherme e Mila Bang (2013): «Terminologia em série: House M. D», *EntreLetras*, (4): 114-133.
- Galves, Charlotte e Helena Britto (1999): «A construção do corpus anotado do português histórico Tycho Brahe», em *Anais do IV Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada (PROPOR)*. Évora: Universidade de Évora, pp. 81-92.
- Galves, Charlotte e Pablo Faria (2010): *Tycho Brahe Parsed Corpus of Historical Portuguese*. <<http://www.tycho.iel.unicamp.br/~tycho/corpus/en/index.html>>.
- Gaudin, François (2007): «Quelques mots sur la socioterminologie», *Cahiers du RIFAL*, 26: 26-35.
- Giusti, Rafael; Arnaldo Candido Junior, Marcelo Muniz, Lívia Cucatto e Sanda Aluísio (2007): «Automatic detection of spelling variation in historical corpus: An application to build a Brazilian Portuguese spelling variants dictionary», em *Proceedings of the Corpus Linguistics*.
- Gonçalves, Maria Filomena (2013): «Aspectos do léxico português e brasileiro no século XVIII: Pesos e medidas no Erário Mineral (1735), de Luís Gomes Ferreira», *Confluência - Revista do Instituto de Língua Portuguesa (Rio de Janeiro)*, 43: 47-67.
- Gonçalves, Maria Filomena e Clotilde de Almeida Azevedo Murakawa (2009): «Lexicografia implícita em textos del Padre Jesuíta Fernão Cardin», em Otto Zwartjes, Ramon Arzapalo Marin e Thomas Smith-Stark (eds.): *Missionary Linguistics IV - linguística misionera IV - Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 233-248.
- Grmek, Mirko Dražen (1995): «Déclin et émergence des maladies», *Hist. cienc. saude-Manguinhos [online]*, 2 (2): 09-31.
- Guarino, Nicola (1998): «Formal ontology and information systems», em Nicola Guarino (ed.): *Formal ontology in information systems: Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. Amsterdam: IOS Press.
- Hirohashi, Alexandre Sussumu (2004): *Aprendizado de regras de substituição para normatização de textos históricos*. Dissertação (Mestrado – Instituto de Matemática e Estatística). Universidade de São Paulo. São Paulo: USP.
- Hoffmann, Lothar (2015<sup>a</sup>): «Conceitos Básicos da Linguística das Linguagens Especializadas», em Maria José Bocorny Finatto e Leonardo Zilio (eds.): *Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas*. 1. ed. Porto Alegre: Pallotti, pp. 39-50.
- Hoffmann, Lothar (2015<sup>b</sup>): «Linguagens especializadas como sublinguagens», em Maria José Bocorny Finatto e Leonardo Zilio (eds.): *Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas*. 1.ª ed. Porto Alegre: Pallotti, pp. 89-101.
- Jasmin, Marcelo Gantus (2005): «História dos conceitos e teoria política e social: referências preliminares», *Revista Brasileira de Ciências Sociais*, 57 (20): 27-38.
- Kabatek, Johannes (2005): «Sobre a historicidade de textos», traduzido por: José da Silva Simões, *Linha d'água*, 17: 157-170.
- Kabatek, Johannes [*portugiesische Originalversion von 78*] (2012): «Tradição discursiva e gênero», em Tânia Lobo, Zenaide Carneiro, Juniana Soledade, Ariadne Almeida e Silvana Ribeiro (eds.): *Rosae. Linguística histórica, história das línguas e outras histórias*. Salvador: EDUFBA, pp. 579-588.
- Kabatek, Johannes; Claus Pusch e Wolfgang Raible (eds.) (2005): *Romanistische Korpuslinguistik II: Korpora und diachrone Sprachwissenschaft. Romance Corpus Linguistics II: Corpora and Diachronic Linguistics*. Tübingen: Narr (ScriptOra 130).
- Koselleck, Reinhart; Christian Meier, Horst Günther e Odilo Engels (2013): *O conceito de História*. Traduzido por René E. Gertz. Belo Horizonte: Autêntica.
- Krieger, Maria da Graça (2000): «Terminologia re-

- visitada», DELTA, 2 (16): 209-228. <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44502000000200001&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200001&lng=en&nrm=iso)> [consulta: 29.VI.2020].
- Krieger, Maria da Graça (2006): «Lexicografia: o léxico no dicionário», em Maria Cândida Trindade Costa de Seabra (ed.): *O léxico em estudo*. Belo Horizonte: Faculdade de Letras da UFMG.
- Krieger, Maria da Graça e Anna Maria Becker Maciel (eds.) (2001): *Temas de terminologia*. São Paulo/Porto Alegre: Ed. Universidade/UFRGS/Humanitas/USP.
- Krieger, Maria da Graça e Maria José Bocorny Finatto (2004): *Introdução à Terminologia: teoria e prática*. São Paulo: Contexto.
- Krieger, Maria da Graça; Márcio Sales Santiago e Maria Teresa Cabré (2013): «Terminologia em foco: uma entrevista comentada com Maria Teresa Cabré», *Calidoscópio*, 3 (11): 328-332.
- Lourenço, Tânia Souza (2016): *O médico entre a tradição e a inovação: João Curvo Semedo*. Dissertação (Mestrado em Programa de Pós-graduação em História) – Universidade Federal Fluminense. Niterói: UFF.
- Macedo, Cristian (2019): «A arte da tradução: um breve exercício de Terminologia Diacrônica», *Cadernos do IL*, 59: 255-270. <<https://seer.ufrgs.br/cadernosdoil/article/view/92537/54508>>.
- Marengo, Sandro Márcio Drumond Alves (2016): «Crítica Textual e Terminografia Diacrônica: bases para preparação da socioterminologia histórica», *LaborHistórico*, 2 (2): 86-112.
- Minghelli, Thaís Domênica e Rove Chishman (2012): «Ontologia jurídica e a relação de meronímia», *Veredas (UFF OnLine)*, 16: 85-103.
- Møller, Bernt (1998): «À la recherche d'une terminochronie», *Meta*, 43 (3): 426-438.
- Monte, Vanessa Martins do e Maria Clara Paixão de Sousa (2017): «Por uma filologia virtual: o caso das atas da Câmara de São Paulo (1562-1596)», *Revista da ABRALIN*, [S.l.], 1 (16). <<https://revistas.ufpr.br/abralin/article/view/51938>> [03.08.2020].
- Murakawa, Clotilde de Almeida Azevedo (2005): «Plantas e Drogas da Índia na obra Coloquios de Garcia d'Orta: um estudo do vocabulário», em *Atas do VII Encontro da Associação Internacional de Lusitanistas*, Providence – USA.
- Murakawa, Clotilde de Almeida Azevedo (2013): «O vocabulário das enfermidades em documento do Brasil Colonial: o relato de Prodigiosa Lagoa (1749)», em Clotilde de Almeida Azevedo Murakawa e Odair Luiz Nadin da Silva (eds.): *Terminologia: uma ciência interdisciplinar*. Araraquara: Editora Cultura Acadêmica.
- Paixão de Sousa, Maria Clara e Thorsten Trippel (2006): «Building a historical corpus for Classical Portuguese: some technological aspects», em *Proceedings of the V International Conference on Language Resources and Evaluation (LREC)*. Genova: ELRA, pp. 1831-1836.
- Paula, Maria Helena de e Amanda Moreira de Amorim (2020): «Em vida e na hora da morte também: o que dizem registros de óbito oitocentistas da freguesia de Nossa Senhora da Penha de Corumbá (1847-1855)», *Diacrítica*, 32 (3): 249-270.
- Pereira, Amanda Henrique e Odair Luiz Nadin (2019): «Análise da variação terminológica denominativa em textos jurídicos: o caso do termo petição inicial», *TradTerm*, 34: 121-142.
- Quaresma, Paulo e Maria José Bocorny Finatto (2020): «Information Extraction from historical texts: a Case Study», em *PROPOR Workshop: Digital Humanities and Natural Language Processing*, Évora, 01: 49-56. <<https://dblp.org/db/conf/dhandnlp-ws/dhandnlp-ws2020.html>>.
- Reategui, Eliseo; Ana Paula Metz Costa, Daniel Epstein e Michel Carniato (2018): «Learning Scientific Concepts with Text Mining Support», em *International Conference in Methodologies and Intelligent Systems for Technology Enhanced Learning*. Berlin: Springer.
- Robin, Régine (1973): *História e Linguística*. Editora Cultrix: São Paulo.
- Rüsen, Jörn (2010): *Razão histórica: fundamentos da ciência histórica*. Brasília: Editora Universidade de Brasília.
- Santos, Ivo; Fernanda Olival e Ofélia Sequeira (2020): «Excavating the Data Pit: the Portuguese Parish Memories (1758) as a Gold Standard», em *PROPOR Workshop: Digital Humanities and Natural Language Processing*, Évora, 01: 69-75. <<https://dblp.org/db/conf/dhandnlp-ws/dhandnlp-ws2020.html>>.
- Semedo, João Curvo (1707): *Observações medicas doutrinaes de cem casos gravíssimos, que em serviço da pátria, & das nações estranhas escreve em língua portugueza, & latina*. Lisboa: Oficina de Antonio Pedrozo Galram.
- Serra, Luís Henrique (2019): *A variação denominativa no discurso especializado da cana-de-açúcar no Brasil: uma pesquisa sobre a variação funcional*. Tese (Doutorado em Programa de Pós-graduação em Filologia e Língua Portuguesa) – Universidade de São Paulo. São Paulo: USP.
- Temmerman, Rita (2000): «Une théorie réaliste de la terminologie: le sociocognitivisme», *Terminologies nouvelles*, 21: 58-64.
- Temmerman, Rita (2004): «Teoria Sociocognitiva da Terminologia». Traduzido por Natacha Enzweiler e Luzia Araújo; revisão de Talia Bugel. *Cadernos de Tradução*, 17: 31-50.
- Witter, Nikelen Acosta (2005): «Curar como arte e ofício: contribuições para um debate historiográfico sobre saúde, doença e cura», *Tempo*, 19 (10): 13-25. <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-77042005000200002&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-77042005000200002&lng=en&nrm=iso)> [consulta: 29.VII.2020].