

DOI: <http://dx.doi.org/10.18226/19844921.v12.n25.02>

DicTrans: dicionário on-line multilíngue sobre o Transtorno do Humor Bipolar*

DicTrans: the online multilingual dictionary about the Bipolar Disorder

Heloísa Orsi Koch Delgado**

Carmen Lucas Verneti***

Carlos Alberto dos Santos****

Resumo

Este texto apresenta o DicTrans, dicionário on-line multilíngue sobre o Transtorno do Humor Bipolar, para profissionais e estudantes do âmbito da Saúde. Em termos teóricos, este trabalho fundamenta-se na área da Terminologia e sua relação com os Estudos do Texto, e em termos aplicados, encontra subsídios metodológicos na abordagem baseada em corpus e em estatísticas do Processamento da Linguagem Natural. O corpus especializado, composto de artigos escolhidos de uma revista internacionalmente reconhecida sobre o tema, foi analisado como um todo representativo de propriedades linguísticas, comunicativas e cognitivas específicas. Essa visão textual sobre o texto especializado ofereceu respaldo teórico-aplicado para a metodologia de construção do dicionário, cuja descrição encontra-se detalhada ao longo do artigo.

Palavras-chave

Dicionário on-line. Transtorno do Humor Bipolar. Linguística de Corpus. Processamento da Linguagem Natural. Linguagem Especializada.

Abstract

The article introduces DicTrans, an online multilingual dictionary about Bipolar Disorder (BD) aimed at professionals, undergraduate and graduate students of the Health area. The theoretical framework describes the terminological approaches that show the connection between the field of Terminology and Text Linguistics. As for methodology, we employed the corpus-based approach and Natural Language Processing tools. The specialized corpus,

* Além dos autores titulares, que assinam este artigo, também são autores Débora Montenegro Pasin, Janaína Baladão de Aguiar, Larissa Bauermann Ramos, Lucas Zambrano Rollsing, Lúcia Henriques Maia, Luiza Rico Bini, Maria da Graça Montenegro, Maria José Bocorny Finatto, Rebeca Américo, Vanessa Silva Fischer.

** Doutora em Letras pela Universidade Federal do Rio Grande do Sul (UFRGS).

*** Instituto de Prevenção e Pesquisa em Álcool e Outras Dependências (IPPAD).

**** Mestre em Ciência da Computação pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

composed of articles taken from an internationally acknowledged journal about BD, was analyzed considering the communicative, linguistic, and cognitive features commonly encountered in domain-specific language. This particular viewpoint was the basis for the methodology used, whose description is described throughout the study.

Keywords

Online dictionary. Bipolar Disorder. Corpus Linguistics. Natural Language Processing. Specialized Language.

Introdução

Diante da importância do acesso ao conhecimento sobre os avanços da ciência para qualificar os profissionais das Ciências Médicas, o presente dicionário objetiva descrever o DicTrans, o dicionário on-line multilíngue sobre o Transtorno do Humor Bipolar (doravante THB). Em linhas gerais, a obra apresenta a linguagem médica científica sobre o THB através de verbetes com definições expandidas e áudios explicativos, que buscam facilitar a compreensão sobre uma das doenças mais complexas da atualidade. Em outras palavras, a criação do Dictrans - elaborado por uma equipe multidisciplinar das áreas da Linguística, Tradução, Psiquiatria, Ciências da Computação e do Ensino de Línguas - visa proporcionar uma conjugação entre familiarização e aprofundamento sobre o THB, com enfoque no auxílio diagnóstico e no apoio conceitual de casos clínicos psiquiátricos. Em que pese a importância de acesso ao conhecimento, bem sabemos que há uma grande lacuna de produtos dessa natureza que contemplem o português do Brasil (entre outros idiomas) no âmbito da Medicina e áreas afins. Via de regra, os produtos estão disponíveis, em sua grande maioria, apenas em língua inglesa.

O DicTrans¹, iniciado em 2012 e lançado em 2019, contém a terminologia utilizada em artigos atuais sobre o THB em quatro línguas (inglês – português – italiano – espanhol), sendo o inglês a língua de partida. Esse dicionário tem como público leitor principal os profissionais da saúde e os estudantes de Medicina e áreas afins. Poderá ser igualmente útil a jornalistas, tradutores e revisores linguísticos, que tratem de temas relacionados à saúde humana.

¹ Contamos com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Chamada Universal – MCT/CNPq N° 14/2013, no período de 01/10/2013 a 30/09/2016.

O principal corpus de estudo escolhido para este trabalho constituiu-se de textos em língua inglesa, oriundos da revista *Bipolar Disorder*², internacionalmente reconhecida e voltada para os últimos avanços da doença. A escolha do inglês como língua de partida se deve, principalmente, por ser a *língua franca* das áreas técnico-científicas, enquanto o português do Brasil, o italiano e o espanhol foram escolhidos para serem as línguas de chegada. Vale lembrar que a língua portuguesa brasileira se torna essencialmente importante pela necessidade de maior divulgação sobre o THB, indo ao encontro da preocupação das instituições, associações e centros de pesquisa sobre essa doença no Brasil. Justificam-se as escolhas da língua italiana porque a Itália é um país de referência nos estudos sobre transtornos psiquiátricos, e a língua espanhola, pela extensa literatura existente sobre o THB e por ser uma das línguas mais faladas do mundo³.

No que tange à escolha por esse domínio especializado, baseamo-nos nas pesquisas científicas de importantes pesquisadores brasileiros que destacam a grave condição do THB, doença que provoca episódios maníacos e depressivos em seus pacientes e afeta de 0,5% a 1% da população em sua forma estrita e próximo de 6%, se considerarmos suas formas mais leves, o chamado espectro bipolar (KAPCZINSKI et al., 2009).

A maioria dos pacientes com esse transtorno apresenta os primeiros sintomas antes dos 25 anos de idade. Contudo, o tempo médio entre o início dos sintomas e o diagnóstico correto é, com frequência, de dez anos. Assim, um dos desafios da psiquiatria hoje em dia é diminuir o tempo entre as primeiras manifestações clínicas e o tratamento adequado, pois quanto maior o número de episódios não tratados, maior o comprometimento cognitivo do paciente, que progride de forma inexorável na maioria dos casos.

Muitas vezes, a confusão diagnóstica se deve à própria história natural do THB, pois a depressão é a sua primeira manifestação em mais da metade dos casos; porém, para que o diagnóstico de bipolaridade seja estabelecido, é necessário que haja episódios maníacos ou hipomaníacos na história clínica do paciente. Além disso, a doença se estabelece de forma gradual, e não repentina e está entre as dez

² <https://www.nimh.nih.gov/health/topics/bipolar-disorder/index.shtml>.

³ De acordo com a edição de 2018 do jornal *Clarín*, o espanhol é a segunda língua materna mais falada do mundo e a terceira da Internet. Disponível em: https://www.clarin.com/clarin-em-portugues/espanhol-ja-segunda-lingua-mais-falada-do-mundo_0_En_jOwIbS.html. Acesso em: 01 Mar, 2020.

condições mais debilitantes do mundo, tornando-a muito dispendiosa para a família e a sociedade (KAPCZINSKI et. al. 2009).

Ademais, a depressão talvez seja a fase da doença onde se concentra a maior carga de incapacidade, sendo a sexta causa mais comum de inaptidão entre adultos jovens. Seus portadores não só perdem mais dias de trabalho como produzem menos quando presentes. Mesmo aqueles pacientes que conseguem uma remissão clínica enfrentam dificuldades para recuperar o nível de funcionamento anterior à doença. Trata-se de uma doença com alto grau de mortalidade, cujo risco de suicídio é trinta vezes maior do que outros transtornos (GAZELLE et al., 2005).

Há de se reconhecer que, nos últimos cinquenta anos, a pesquisa psiquiátrica evoluiu significativamente, trazendo uma melhor compreensão sobre seus fatores causais e sobre o desenvolvimento de tratamentos mais eficazes. No entanto, o THB, apesar de ter recebido importantes contribuições da ciência, continua sendo uma doença potencialmente devastadora e, por isso, foco de interesse de psiquiatras do mundo inteiro (KAPCZINSKI et al., 2009).

Para que avanços psiquiátricos nessa importante área do conhecimento estejam disponíveis para especialistas e praticantes da Saúde, faz-se necessário disponibilizar recursos de fácil acesso e conteúdo informativo confiável, que contribuam para sua divulgação na sociedade. Dessa forma, acreditamos que o DicTrans possa se tornar uma ferramenta de busca por informações fidedignas e atualizadas sobre o domínio do THB, pois se configura como produto continuado de investigação em que prevemos a atualização anual de conteúdo, para que se justifique como ferramenta de uso on-line.

Apresentadas as características gerais da obra, descrevemos, a seguir, o enfoque teórico-metodológico escolhido para sua elaboração.

Enfoque teórico-metodológico da obra

Tendo em vista que se trata de um dicionário de linguagem especializada, o presente estudo encontrou apoio na Terminologia, dentro de uma perspectiva textual, e nos recursos e técnicas de uma abordagem em corpus e do Processamento da Linguagem Natural (PLN). As linhas que seguem discorrem sobre esses enfoques teórico-metodológicos, além de descreverem as etapas realizadas para a construção do DicTrans.

Estudos da Terminologia sob a ótica textual

Entendemos Terminologia como uma área de estudos ou disciplina a qual se ocupa, de modo geral, de diferentes fenômenos das linguagens técnico-científicas. Atualmente, os estudos dessa área estão cada vez mais associados às diferentes teorias dos Estudos do Texto e do Discurso, as quais destacam que o texto é parte essencial dessa associação. É importante salientar, no entanto, que as terminologias continuam tendo um papel relevante entre as teorias terminológicas. Numa perspectiva histórica, a partir dos anos 90, a Terminologia sofreu uma mudança paradigmática importante tanto no campo teórico quanto no aplicado. No primeiro, esse campo de estudos passou a se constituir a partir das áreas da Linguística, Comunicação e Cognição, enquanto no segundo, passou a adotar novas metodologias no fazer terminográfico. Diferentemente da Teoria Geral da Terminologia (TGT), que analisava os termos *in vitro*, o novo olhar à forma como os termos deveriam ser analisados deu lugar à sua identificação e descrição *in vivo*, ou seja, dentro de seu contexto de uso, os textos especializados. Dessa forma, e de acordo com a Teoria Comunicativa da Terminologia (TCT), apresentada por Maria Teresa Cabré (1999), os termos, considerados e analisados em toda a sua riqueza e complexidade, passaram a ser entendidos como unidades dinâmicas e passíveis de mudança e evolução (CABRÉ, 1999). Em relação à prática, a inovação foi o uso de corpora textuais e ferramentas computacionais de extração de informação linguística. Nesse sentido, destaca Bevilacqua (2013, p. 11) que “tal inovação implicou uma mudança de metodologia, uma vez que se passou do método onomasiológico⁴, utilizado na perspectiva prescritivista, para o método semasiológico⁵, nas perspectivas textuais e comunicativas⁶”. Essas perspectivas, que fazem parte de propostas descritivistas, preveem que “se identifiquem, analisem e descrevam os termos, considerando as diferentes situações comunicativas em que são utilizados, e que, conseqüentemente, se considerem os aspectos linguísticos, comunicativos e cognitivos” (Bevilacqua, 2013, p.12).

⁴ Parte-se do conceito (significado) ao termo (significante).

⁵ Parte-se do termo ao conceito.

⁶ Passa-se da perspectiva prescritivista de Wüster (1950-60 até início dos anos 90) para a perspectiva descritivista (CABRÉ, 1999, entre outros).

Visto que o termo é um elemento comunicativo dentro do texto especializado, fazemos uso das palavras de Krieger (2008, p.6), que dá destaque à “integração de componentes de textualidade e discursividade no aparato teórico-metodológico da Terminologia, cujo objeto primeiro é termo técnico-científico”. Dessa forma, a aproximação dessa área - vista sob uma perspectiva descritiva - com os Estudos do Texto e do Discurso permitiu que fizéssemos uma análise mais aprofundada da linguagem especializada do THB. Buscamos compreender, por exemplo, os elementos envolvidos em sua formulação discursiva (fatores linguísticos, comunicativos, cognitivos e pragmáticos), o que contribuiu para o estabelecimento da metodologia de elaboração do DicTrans,

Nessa direção, Ciapuscio (2003) salienta que “[...] a ideia básica é a de que o estudo terminológico do texto pode trazer contribuições não apenas para a Terminologia, mas também para o esclarecimento de problemáticas relevantes para os estudos textuais, como, por exemplo, avanços em direção a uma tipologia dos textos especializados e ao estabelecimento de graus de especialização dos textos tomando como base o léxico” (p. 71).⁷ A compreensão, portanto, de que a comunicação especializada se materializa no texto especializado, possui propriedades estruturais específicas - assim como ocorre com outras tipologias textuais - e fundamenta-se nos mecanismos da linguagem em funcionamento, é de essencial importância.

Com relação ao grau de especialização que as comunicações especializadas refletem, Krieger e Finatto (2004, p. 116) salientam que este (o grau) fornece o ponto de partida para uma tipologia. Afirmam que “mais do que o tema, o grau de densidade informativa junto à forma de comunicar, com maior ou menor utilização da terminologia da área em questão, funciona como mecanismo determinante dos graus de especialização de um texto”. As autoras acrescentam que “os propósitos diferenciados de cada tipo de comunicação e os diferentes destinatários previstos explicam não apenas a variação da densidade informativa, como o uso maior ou menor de terminologias”.

⁷ Versão original em espanhol: “[...] intento brindar argumentos adicionales “a la idea básica de que el estudio terminológico del texto puede no solo aportar a la terminología sino que también puede contribuir a esclarecer problemáticas relevantes dentro de los estudios textuales, como por ejemplo, avances hacia una tipología de los textos especiales y al establecimiento de grados de especialización de los textos basado en el léxico”.

No caso do DicTrans, a densidade informativa é alta, pois é uma obra direcionada a profissionais e estudantes das áreas da Saúde. No campo “definição”, por exemplo, embora tenhamos optado pelo uso de uma linguagem objetiva, evitando um estilo de escrita rebuscada, há uma alta densidade terminológica, ou seja, uma concentração de conteúdo específico da área, o que torna a leitura de difícil compreensão para leigos. Abaixo, um exemplo ilustrativo da definição de *abstract thinking* (pensamento abstrato) da obra.

Abstract thinking: O pensamento abstrato engloba três funções executivas nucleares: memória operacional, controle inibitório, e flexibilidade cognitiva, que atuam como base para o desempenho de funções executivas mais complexas como a solução de problemas, planejamento, e raciocínio abstrato. (CAMARGO et al., 2014, p.117).

No que diz respeito às propriedades estruturais específicas da comunicação especializada, podemos dizer que há, de fato, características tradicionalmente conformadoras de textos especializados, marcadamente os de natureza científica. A comunicação tradicional da ciência, por exemplo, tende ao emprego dos verbos em terceira pessoa, produzindo, assim, uma imagem de impessoalidade, de isenção de pontos de vista. Conforme Krieger e Finatto (2004, p. 117), “este mecanismo contribui para o efeito de que o conhecimento relatado está isento de condicionamentos e de pontos de vista particulares, como se fosse a ciência a falar por ela mesma”. Trata-se, aqui, de um recurso linguístico que, junto com o uso de nominalizações e de estruturas passivas, favorece os efeitos de indeterminação e apagamento da subjetividade. Explica-se, dessa forma, a razão pela qual o traço da impessoalidade está associado ao estilo do texto científico. Exemplos dessa impessoalidade podem ser vistos através dos segmentos, representativos da voz passiva⁸ retirados do nosso corpus de estudo⁹ sobre o THB (Revista *Bipolar Disorder*). Além disso, grande quantidade dos termos técnico-científicos se constrói a partir de prefixos e radicais gregos ou latinos, o que beneficia a comunidade internacional, acima das fronteiras linguísticas nacionais ou

⁸ Exemplo: Alterations in genetic coding for the NMDA (N-methyl D-aspartate) receptor subunit 1 and subunit 2B genes *have both been shown* to confer risk for BD, although the effect of these genetic abnormalities on receptor density or function is unclear.

⁹ Esta pesquisa fez uso apenas do corpus de estudo e não de referência. Não sentimos necessidade de compará-los para extrairmos os termos relevantes. Isso se deu por conta da participação da médica psiquiatra, especialista em THB, que fez a seleção dos termos de todo o corpus de forma manual. Essas palavras-chave foram comparadas aos termos relevantes obtidos pela ferramenta ExATOLI de Lopes (2015), descritas mais adiante, e revisadas pela coordenadora do projeto e pela médica. Vale lembrar que um corpus de referência obtém as palavras-chave de um corpus de características genéricas semelhantes ao corpus de estudo e tende a eliminar os elementos genéricos (relativos a um mesmo gênero) em comum, resultando em uma lista de palavras-chave que não inclui esses elementos (SARDINHA, 2005).

regionais (*maladjustment, premorbid*). Outro procedimento de formação de termos científicos, não exclusivo, mas muito típico, está relacionado com as reduções (Siglas: THB – Transtorno do Humor Bipolar) e com a hiponímia (Tipo 1, Tipo 2 – tipos de THB).

Há outras peculiaridades que, juntamente com o uso de estruturas linguísticas monossêmicas, explicam qualificações prototípicas dos textos especializados, tais como a objetividade e a concisão. Esses efeitos estilísticos agregam-se a determinados mecanismos de natureza textual, como a clareza, a coesão e a coerência, funcionando a favor da lógica e da pertinência da comunicação especializada e, assim, de um determinado formato padrão da comunicação da ciência (KRIEGER e FINATTO, 2004).

Em consonância com Krieger e Finatto (2004), Rodilla (1998) salienta que a concisão é uma característica da mensagem científica em que se usa o menor número de palavras necessárias para expressar a ideia que se deseja comunicar. Logicamente, esta propriedade não pode ir contra uma transmissão clara e eficaz do conteúdo. Trata-se, por um lado, de uma economia semântica relacionada com o número de palavras empregadas, cujo expoente máximo é a substituição de frases inteiras por uma única palavra através da criação de um novo termo (ex.: hipomaníaco: surto maníaco de menor intensidade no transtorno de humor bipolar). A economia linguística ressalta as condições implícitas particulares da comunicação técnica, pois sua aplicação supõe uma série de premissas como, por exemplo, a de que os interlocutores tenham os mesmos conhecimentos do assunto. No entanto, as frases da linguagem científica podem ser mais longas do que em outros tipos de linguagem devido à complexidade dos conceitos científicos, exigindo, dessa forma, um número maior de palavras para que possam ser explicados. Tomamos como exemplo o conceito de depressão: é uma doença psiquiátrica crônica e recorrente que produz alteração do humor caracterizada por tristeza profunda e forte sentimento de desesperança¹⁰.

Há de se considerar também a precisão dos termos empregados para a elaboração da mensagem: quanto maior for a sinonímia, a polissemia e a homonímia

¹⁰ Conceito do Dr. Dráuzio Varela. Disponível em: <https://drauziovarela.uol.com.br/doencas-e-sintomas/depressao/>. Acesso em: mar. de 2020.

contidas nesses termos, mais difícil a compreensão (RODILLA, 1998)¹¹. Como exemplo de precisão, o *CID 10*, ou *Classificação Internacional de Doenças*, valida os diagnósticos no mundo inteiro, ou seja, a mesma doença é diagnosticada com os mesmos critérios nos mais diferentes países, ou seja, os critérios são universais. Dessa forma, os sintomas de uma doença como a esquizofrenia devem ser reconhecidos em outras partes do mundo¹², pois são considerados o ponto de partida para o diagnóstico: no caso dos esquizofrênicos, um dos sintomas cardinais é a rigidez afetiva. No entanto, há de se levar em consideração os aspectos culturais locais tais como a nacionalidade do paciente: ao compararmos italianos e alemães, reconhecemos que, de modo geral, no primeiro grupo a afetividade é mais aparente, enquanto no segundo, a demonstração de afeto tende a ser mais oculta.

Explicitamos, de forma breve, a complexidade da comunicação especializada, com ênfase no discurso científico, e a importância de examiná-la à luz da textualidade. Nessa perspectiva, seus componentes aparecem entrelaçados de modo a constituir uma rede: há uma inter-relação, portanto, entre o conteúdo semântico, o léxico empregado, a formulação discursiva, a estruturação textual e os fatores pragmáticos.

Acreditamos que as perspectivas aqui traçadas, longe de esgotarem o assunto, contribuíram para uma melhor representação do THB na obra DicTrans.

A abordagem da terminologia mediada por corpus e pelo Processamento da Linguagem Natural

Por ser uma pesquisa baseada na compilação e análise de termos que compõem a nominata do dicionário, nossa metodologia encontra apoio e inspiração na abordagem baseada em corpus¹³. Sánchez (1995) explica o conceito de corpus que continua atual:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da

¹¹ Texto original, tradução livre com foco na ideia geral do parágrafo: “La homonimia existente em el dominio de los tecnicismos puede dificultar la precisión del discurso científico”.

¹² Vale acrescentar que o reconhecimento se torna mais fácil se as línguas forem escritas com o alfabeto romano (ou latino). Isso vale, portanto, para situações em que outros alfabetos não sejam compreendidos (o coreano, o grego e o armênio, por exemplo).

¹³ Um dos embasamentos para esta pesquisa são os resultados da tese de doutorado de Delgado (2012), na qual a autora propôs um enfoque didático para a tradução de artigos científicos relacionados ao tema dos transtornos psiquiátricos. O diferencial do trabalho, tal como se vê em Teixeira (2010), foi o de uma abordagem da terminologia mediada por corpus.

totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise' (Sanchez, 1995, pp. 8-9)

Nesse sentido, o nosso objetivo foi o de compilar termos e expressões de textos científicos especializados escritos sobre o THB, partindo de documentos em língua inglesa veiculados, em meio eletrônico, entre 2010 e 2016. Para esses documentos em inglês, buscamos textos de igual temática e datação, nas línguas italiana e espanhola escritos originalmente nessas línguas. Houve, então, uma busca pelos equivalentes¹⁴ nesses corpora, em sua grande maioria digitais, dos verbetes em inglês. Nessa fase não fizemos uso de ferramenta computacional; utilizamos tal recurso apenas na etapa de extração e tratamento dos corpora em língua portuguesa (ExATOLP, LOPES, 2012) e inglesa (ExATOLI, LOPES, 2012), esse último especialmente adaptado pela pesquisadora para desenvolver esta pesquisa. Vale dizer que essa ferramenta não foi desenvolvida para outros idiomas.

O corpus em inglês foi constituído dos artigos da referida revista *Bipolar Disorder*; já o corpus em português constituiu-se da *Revista Brasileira de Psiquiatria*. Para validarmos os equivalentes dos termos nessa língua, utilizamos a *Classificação Internacional de Doenças (CID-10)*¹⁵, livros sobre o tema, tais como *Neuropsicologia: Teoria e Prática* (CAMARGO et al., 2014), e o *Manual Diagnóstico e Estatístico de Transtornos Mentais* (2014), entre outros. Para a língua espanhola, fizemos uso de dicionários digitais consagrados na literatura como, por exemplo, o *RAE - Dicionário da Real Academia Espanhola*¹⁶, o *CLAVE - Dicionário de uso do Espanhol Atual*¹⁷ e o *Larousse Dicionário Avançado de Espanhol*, este último, impresso. As equivalências dos termos em espanhol, obtidas através dessas obras, foram validadas pelos seus contextos de uso em revistas on-line atuais sobre o THB nessa língua. Quanto às equivalências em língua italiana, utilizamos artigos digitais publicados sobre a

¹⁴ Utilizamos “equivalência” e suas flexões para indicar o “grau” de equivalência em que uma palavra, uma frase, ou mesmo um texto da cultura de partida pode ser considerado na língua e na cultura receptora. A equivalência tem traços do discurso ou da palavra e depende da tradução.

¹⁵ Disponível em: <https://www.bulas.med.br/cid-10/>. Acesso em: abr. 2020.

¹⁶ <https://www.rae.es/>.

¹⁷ Ediciones SM. Dicionario Clave. Disponível em: <http://clave.smdiccionarios.com/app.php>, atualmente fora do ar, provavelmente para melhorias. Acesso em: mar. 2020.

doença¹⁸ e livros impressos tais como “*La perdita della tristezza*” de 2015¹⁹ e “*Anima*” de 2011²⁰, entre outros. Mais do que procurarmos as equivalências para os termos das entradas do dicionário, procuramos buscar fontes de diversos repertórios e gêneros textuais, para que houvesse consistência do léxico especializado.

Para fins de cumprimento dessa etapa, valemo-nos também de ferramentas da área de Processamento de Linguagem Natural (PLN), que nos auxiliou na etapa de extração automática de termos e visualização de resultados.

O Processamento da Linguagem Natural (PLN) é uma área da Ciência da Computação que desenvolve e utiliza recursos informatizados para o tratamento de fenômenos e de ocorrências das línguas naturais, geralmente lidando com seguintes aspectos: fonologia, morfologia, sintaxe, semântica e pragmática. O PLN produz ferramentas para desempenhar tarefas determinadas que partem da linguagem humana como insumo ou dado, tais como a tradução automática, a indexação automática e a busca/recuperação de informação em grandes acervos textuais. Em resumo, o PLN “denota especificamente o objeto da pesquisa de desenvolvimento de sistemas computacionais capazes de processar objetos de natureza linguística” (DIAS DA SILVA, 1996). Manning e Schütze (1999) também afirmam que o PLN incorpora inúmeras técnicas para interpretação da linguagem com base em métodos estatísticos e de aprendizado de máquina – conjunto de algoritmos que induz um computador a “aprender” um determinado número de regras de funcionamento de uma língua - por meio de análises de corpora de exemplos típicos do mundo real²¹.

De fato, há várias tarefas que podem ser conduzidas pelo PLN tais como a análise sintática (*parsing*), a contagem de *types* (palavras diferentes) e de *tokens* (total de palavras de um texto), lematização, marcação de componentes de um discurso, detecção automática de idiomas, identificação de relações semânticas, enumeração de classes de palavras, análise morfológica, reconhecimento de fala, padrões de tradução de máquina, entre outras. A tarefa de extração de termos com maior

¹⁸ Exemplo: *Disturbo bipolare, un braccialetto con wifi e app per monitorare la malattia* (2015). Disponível em: https://www.ilmessaggero.it/salute/medicina/disturbo_bipolare_braccialetto_wifi_app_pisa-1052353.html. Acesso em: out. 2019.

¹⁹ Original: HORWITZ, A.; WAKEFIELD, J. *The Loss of Sadness: How Psychiatry transformed normal sorrow into Depressive Disorder*. Oxford-New York: Oxford University Press, 2007.

²⁰ HILLMAN, J. *Anima*. New Jersey: Princeton University Press e London: Routledge & Kegan Paul Ltd., 2011.

²¹ Não conseguimos acesso ao livro digital e nem impresso e, portanto, não encontramos a versão original da citação.

relevância em um corpus de domínio é um processo típico da área de PLN, pois parte da anotação linguística de cada texto, extração e contagem dos termos portadores de informação e tratamento estatístico dos resultados encontrados.

No que se refere ao Dictrans, dado o volume de dados textuais tratados (89 textos em inglês e 43 em português²²) e devido à necessidade de imparcialidade na busca por informações, foi necessário fazer a extração proposta por meio de ferramentas automáticas, como o *parser* PALAVRAS e os extratores de sintagmas nominais ExATOLp e ExATOLi (LOPES, 2009; 2012), que auxiliam na extração de termos computacionalmente relevantes²³ para textos em língua portuguesa e língua inglesa, respectivamente.

O ExATOLp e o ExATOLi são ferramentas que recebem um corpus anotado por um *parser* e extraem automaticamente todos os sintagmas nominais que possuem um conjunto de heurísticas com base linguística para refinar o processo de extração, e realizam todo o processo de extração e ordenação dos termos por frequência, bem como a identificação de termos com maior carga semântica para o domínio. As ferramentas também disponibilizam recursos linguísticos, como listas de termos, concordanciador, *tag clouds* e hierarquias de conceitos. O resultado prático dessa tarefa foi a geração de conjuntos de termos representativos do THB e a enumeração adequada das principais informações do corpus.

A geração automática de recursos visuais, tais como as *tag clouds*, capazes de sintetizar informações complexas é um desafio que engloba diversas subáreas da Ciência da Computação. Inicialmente, é preciso conhecer as particularidades da informação a ser constituída; em seguida, deve-se decidir de que forma cada particularidade deve ser representada e, finalmente, deve-se buscar uma forma graficamente clara e agradável para apresentar as informações.

No caso de informações associadas a recursos textuais, optamos pelas *tag clouds*²⁴, que permitem representar os conteúdos associados às palavras e aos termos. Nesse tipo de representação, são escolhidas características como tamanho

²² O número menor de artigos disponíveis em português se dá pelo fato de que os pesquisadores brasileiros escrevem, com frequência, em língua inglesa, a língua atual da ciência.

²³ Lopes, em sua tese de doutorado, provou que o método computacional de extração proposto por ela era capaz de extrair termos relevantes (com maior carga semântica) e, para isso, comparou o resultado extraído com listas de referência, tendo conseguido bom desempenho de extração. Para saber mais sobre a pesquisa de Lopes, acesse <http://meriva.pucrs.br/dspace/bitstream/10923/1651/1/000439085-Texto%2bCompleto-0.pdf>

²⁴ Nuvens de etiquetas. Árvores hiperbólicas e hipertextos.

da fonte, cor, tipo da fonte e localização para representar os conteúdos associadas a cada termo. Na sua versão mais simples, os termos são distribuídos aleatoriamente, com a mesma fonte e cores, porém o tamanho da fonte de cada termo é proporcional à sua relevância descrita por um valor numérico associado. É importante retomar aqui o sentido de *relevância*: característica em que os termos, mesmo não sendo os mais frequentes, possuem uma carga semântica maior devido à sua especificidade ao texto. Podemos remeter o sentido de relevância ao conceito de *chavidade*, ou seja, a relação estatística entre a ocorrência de dada palavra, em um corpus de estudo, e a importância que assume para o léxico de uma área de especialidade.

Com o intuito de facilitar a busca pelo termo no dicionário, *tag clouds* aparecerão automaticamente na primeira página do site da obra, mostrando as expressões que possuem maior frequência de visita²⁵, eliminando a necessidade de o consulente digitar o termo de busca no site. As nuvens serão clicáveis, ou seja, o leitor, ao clicar nos termos dispostos dentro delas, será direcionado às informações sobre os mesmos disponíveis no ambiente.

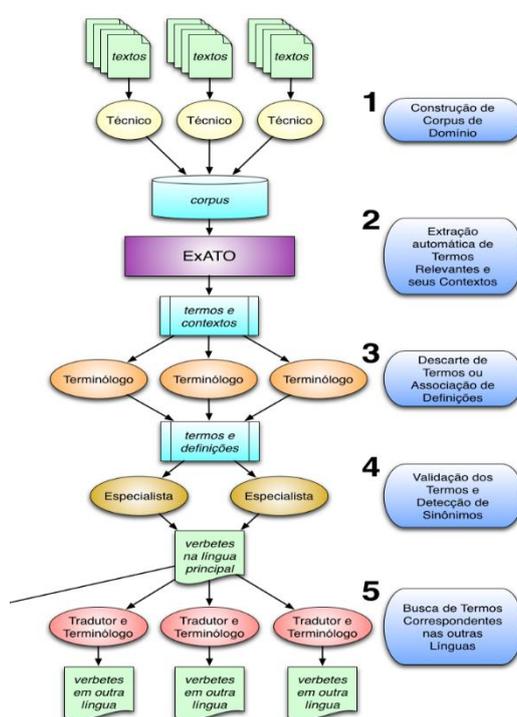
Isso posto, passamos, a seguir, para a descrição da metodologia geral da construção do DicTrans.

Metodologia geral da construção do DicTrans

A figura 1 descreve genericamente o processo de construção do dicionário multilíngue em questão. Nela, pode-se perceber as cinco etapas (em azul à direita) e seus atores (no centro). São também indicados os dados de entrada e saída de cada uma dessas etapas. Um ponto importante para esse processo é a escolha de uma língua principal como ponto de partida para a construção da obra que, neste caso, foi a língua inglesa. As quatro etapas iniciais atuaram somente sobre a língua principal, sendo as demais línguas tratadas somente no contexto da quinta e última etapas do processo.

²⁵ Com base nos dados de armazenamento de acessos, em breve disponíveis ao consulente.

Figura 1 – Processo geral de construção do dicionário proposta pela equipe



Fonte: Equipe DicTrans

Etapa 1 - Construção do corpus de domínio

Nesta etapa, a linguista e a especialista da área do domínio foram as responsáveis por escolherem o corpus de estudo principal do dicionário. No caso de domínios científicos, é natural que a escolha dos textos seja feita com base em revistas científicas especializadas, ou outro tipo de bibliografia definidora (livros e textos reconhecidos como abrangentes e representativos do domínio). Como mencionado anteriormente, utilizamos os artigos da revista *Bipolar Disorder* publicados entre 2010 e 2016.

As tarefas de construção do corpus de domínio foram automatizáveis através de robôs de captura de textos nos diversos formatos disponíveis (*html, pdf, etc.*) e sua

conversão para texto puro (*txt*), tendo sido possível, dessa forma, automatizar a busca e a conversão da informação disponível. Esse corpus, obtido em formato *pdf*, foi limpo quanto a informações não textuais (figuras, gráficos, fórmulas) e à irrelevância para o domínio (endereços, propagandas, códigos) e convertido automaticamente para formato texto puro. A remoção realizada pela máquina foi validada por uma cuidadosa supervisão das pesquisadoras para validar a qualidade da próxima etapa do processo.

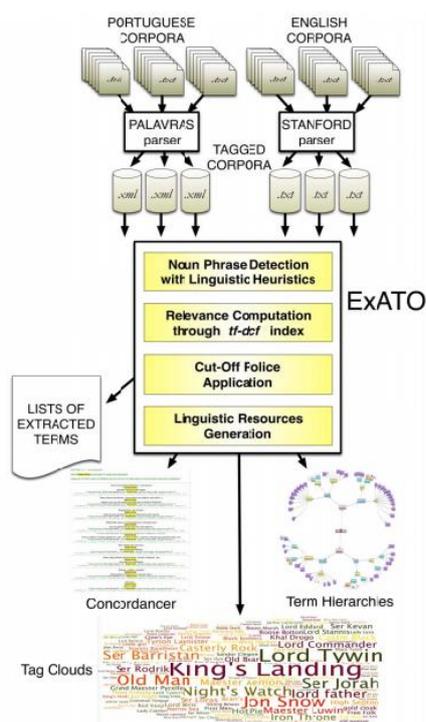
Seguindo os passos definidos na Figura 1, foi construído um corpus em inglês composto por 89 artigos científicos, com um total de 490.460 palavras distribuídas em 18.281 frases e 8.794 n-gramas (15% do total). O número total de sintagmas terminológicos, extraídos computacionalmente pela ferramenta ExATOLI, foi de 147.458 (com repetição) e, dentre estes, 58.629 termos/sintagmas nominais (sem repetições) foram encontrados. Nesse conjunto, o mais frequente foi *patient* encontrado 1.927 vezes. No entanto, o termo *Bipolar Disorder* que foi encontrado 1.191 vezes, foi considerado o mais relevante do corpus no contexto da análise manual (olhar do especialista). A aplicação dos pontos de corte (duas ocorrências, ou seja, todos os termos que tivessem frequência igual ou menor a 2 eram descartados) para os termos detectados resultou na escolha dos 8.794 termos mais relevantes. No entanto, dado o número de pessoas (13) envolvidas no trabalho e seus compromissos profissionais principais, buscou-se a elaboração das entradas (com definições, exemplos, pronúncias e áudios explicativos nas quatro línguas, seleção do tópico do verbete), além da etapa da revisão linguística, de apenas 388 termos. Temos como objetivo ampliar o repertório lexical no que diz respeito ao número de entradas no dicionário; almejamos iniciar essa etapa em médio e longo prazo.

Etapa 2 – Extração automática de termos relevantes e seus contextos

Partimos do corpus construído anteriormente e buscamos extrair os termos mais relevantes do domínio, bem como os contextos em que foram empregados. Essa tarefa, como dissemos, foi automatizada pelo ExATOLI, capaz de extrair os termos mais relevantes em *corpora* escritos em inglês e português, bem como identificar todas as frases onde o termo foi empregado. Lembramos que não fizemos uso dessa ferramenta para as línguas espanhola e italiana, optando, então, pela utilização de material digitalizado e, quando necessário impresso, para a coleta dos equivalentes dos termos escolhidos para compor o dicionário.

A tecnologia por trás do ExATO consiste no processamento automático que leva em consideração diversas técnicas de base linguística como anotação sintática (*PoS-tagging*), identificação de sintagmas nominais e heurísticas de tratamento dos sintagmas detectados. ExATO também possui etapas de processamento de forte base estatística, como o índice *tf-dcf*, que estima a relevância de cada termo para o domínio, e a escolha de limiares, que determina quando um termo é suficientemente relevante para que possa ser considerado um conceito específico do domínio. Além disso, o ExATO provê diversos formatos de saída (*csv*, *html*, *tag clouds*, hierarquias de conceitos), permitindo que termos e contextos extraídos possam ser exportados para outras ferramentas computacionais. No caso do corpus sobre o THB, as saídas do ExATO foram adaptadas como entrada da ferramenta de auxílio aos linguistas, utilizada na terceira etapa. Abaixo, uma imagem da estrutura da ferramenta utilizada.

Figura 2 - Estrutura da ferramenta EXATO



Fonte: Lopes, 2015

O resultado dessa etapa foi a extração automática de termos com qualidade terminológica na forma de listas de n-gramas relevantes, que foram classificados, segundo o número de palavras, em unigramas, bigramas, trigramas, etc. Cada termo detectado foi contabilizado e o processo de estimativa de relevância e aplicação de pontos de corte (duas ocorrências) foi feito sobre o total de termos únicos encontrados.

Para ilustrar numérica e linguisticamente os resultados do nosso corpus, datado de 2015, apresentamos um resumo dos n-gramas, e uma amostra dos sintagmas nominais (trigramas), muitos dos quais incluídos na nominata do DicTrans:

Figura 3 – Resultados e amostra dos n-gramas

			Trigramas	
Textos:	89		Bipolar I Disorder	
Sentenças:	18,281		Bipolar disorder patients	
Tokens:	490,460		Patients with BD	
Selecionados:	8,794	15%	Bipolar II disorder	
(para nuvens)			Borderline personality disorder	
			Major depressive disorders	
			bipolar disorder group	
			parents with BP	
			Structured Clinical Interview	
			Written informed consent	
			paliperidone ER group	
			Substance use disorders	
			age of onset	
			Subjects with BP	
			Children and adolescents	
			course of illness	
			mania and depression	
			Bipolar spectrum disorder	
			Trait rejection sensitivity	
			Individuals with BD	
			Age at onset	
			disorder and schizophrenia	
			maximum monthly increase	
			State rejection sensitivity	
			History of psychosis	
			index PP episode	
			community living skills	
			current mood episode	

	únicos	c/repetição
Unigramas	4,822	62,684
Bigramas	16,285	38,798
Trigramas	11,382	16,749
Quadrigramas	8,906	10,864
Pentigramas	5,623	6,255
Hexagramas	3,559	3,784
Heptagramas	2,331	2,407
Octagramas	1,607	1,641
Eneagramas	1,190	1,220
Multigramas	2,924	3,056
Total	58,629	147,458

Fonte: Lopes (2015)

As etapas seguintes ao processamento da ferramenta EXATO foram: a) leitura manual cuidadosa de cada artigo; b) inclusão de novos sintagmas, tais como *abstract thinking*, *acute treatment phase* e *binge eating disorder*, selecionados manualmente e considerados relevantes por serem palavras-chave para a compreensão sobre o THB; c) exclusão de unigramas e bigramas com pouca relevância (*women*, *participants*, *present study*).

Vale dizer que o processo automatizado foi muito útil como referência inicial da extração de termos, visto que, a partir da verificação e validação manual desse corpus (etapa 4), houve um aproveitamento de 40% do corpus processado automaticamente.

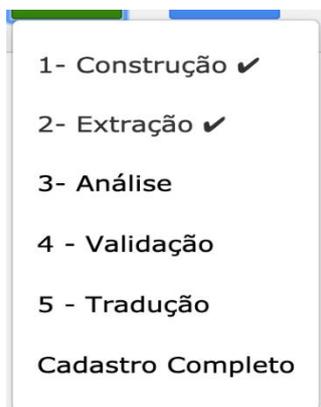
Etapas 3 – Descarte dos termos e associação de definições

Nessa etapa, os termos extraídos e seus contextos alimentaram uma base de dados com acesso via interface dedicada, utilizada por linguistas para analisar, de forma manual, cada termo individualmente. O resultado dessa análise levou à aceitação ou ao descarte do termo, caso ele fosse adequado ou não do ponto de vista

linguístico. Pela natureza dessa atividade, dividimos os termos aceitos entre os membros da equipe de forma a reduzir o tempo total da etapa. Note-se que a atuação em paralelo dos profissionais não representa problema para essa terceira etapa, pois a busca de definição para um determinado termo é, em princípio, independente dos demais termos.

A associação de definições foi realizada pela especialista do domínio, fase em que utilizou livros impressos e artigos digitais. Sua expertise no tema foi essencial para que as definições fossem fidedignas, contribuindo substancialmente para almejarmos a confiabilidade do leitor ao acessar a obra. Após a busca e organização das definições, a especialista e a linguista coordenadora do projeto realizaram a primeira revisão do conteúdo, levando em consideração a pertinência da informação para o perfil do público almejado, às características comunicativas e sociodiscursivas e os aspectos léxico-gramaticais. Tal revisão foi feita dentro do site, utilizando a ficha terminológica já disponibilizada para esse fim; a descrição sobre o tipo de ficha que utilizamos será descrita na seção sobre as etapas metodológicas. Abaixo, uma imagem que mostra a visão que a equipe tinha do processo de organização e revisão das entradas do dicionário.

Figura 4 - Processo de organização e revisão das entradas



Fonte: Equipe DicTrans (2015)

Etapa 4 – Validação dos verbetes e detecção de inconsistências

Na etapa 4, trabalhamos com a mesma base de dados utilizada na etapa anterior: todos os termos e exemplos retirados do corpus bem como as definições extraídas e adaptadas de referências impressas e on-line foram visualizados e validados pela especialista do domínio e linguista coordenadora do trabalho. A equipe

iniciou a revisão de todos os verbetes já inseridos no site e à medida que finalizava a segunda revisão, as autoras titulares²⁶ iniciavam a terceira revisão do conteúdo, sendo responsáveis por informar à equipe quando havia inconsistência e falta de padronização dos itens²⁷. Por exemplo, algumas das observações que fizemos foram relativas à sinonímia e repetição de entradas, e à falta de padronização dos nomes químicos dos medicamentos (nome do composto indicado como substância ativa responsável pela ação farmacológica presente no medicamento), opção escolhida para uso nas entradas. Vale esclarecer que não empregamos os nomes comerciais para não favorecer uma dada indústria de medicamento em detrimento de outra, visto que esse tipo de nome pode ser comercializado sob muitos nomes fantasia. Nossa intenção foi a de informar conteúdo claro e objetivo ao nosso leitor e minimizar a chance da dúvida e incerteza ao utilizar a obra.

Etapa 5 – Busca dos termos correspondentes nas demais línguas

Essa etapa foi na busca por termos e definições nas demais línguas do dicionário (português, espanhol e italiano). A elaboração de todo o conteúdo em língua portuguesa foi auxiliada pela ferramenta ExATOlp (extração automática dos termos frequentes e relevantes do corpus) e por revistas e livros digitais e impressos. A ferramenta contribuiu para que encontrássemos as equivalências dos termos no par de línguas inglês-português, e as referências bibliográficas nos ofereceram um apoio substancial para encontrarmos as definições de tais termos. Para as demais línguas, buscamos, em textos de conteúdo e datação similares, as equivalências desses termos, assunto detalhado na seção sobre a abordagem mediada por corpus e o Processamento da Linguagem Natural.

Vale dizer que não fizemos uso de corpora paralelos anotados sintaticamente através de ferramentas computacionais para esse fim. A razão pela qual esses tipos de corpora não foram utilizados é porque, de acordo com Aubert (1996), nem sempre é vantajoso fazer uso de corpora paralelos na Terminologia, tendo em vista que a busca por termos já traduzidos nesses tipos de corpora não garante a confiabilidade do uso, uma vez que o tradutor nem sempre é especialista na área e talvez não disponha de capacidade linguística (e salientaríamos, terminológica) para, se necessário, criar os termos na língua de chegada. Como existem, na literatura, um

²⁶ Heloísa O.K. Delgado e Carmen Verneti.

²⁷ As decisões sobre todos os detalhes da obra eram tomadas em reuniões realizadas bimensalmente.

número considerável de textos sobre o THB em espanhol e italiano (por razões já expressas na Introdução), não houve a necessidade de se procurar por textos traduzidos nessas línguas. Ademais, a participação ativa da psiquiatra foi fundamental para a validação dos termos nas diferentes línguas pois, como antes mencionado, os termos/sintomas de uma determinada doença, em uma determinada língua, devem ser reconhecidos em qualquer parte do mundo, pois são considerados o ponto de partida para o diagnóstico, mesmo que seu conteúdo seja veiculado por línguas que não utilizam o alfabeto latino ou que, simplesmente, usem sistema de escritas diferentes como é o caso do mandarim (logográfica) e do japonês (silábica). No caso dessa atividade, esse aspecto foi considerado factível por conta da escolha de línguas no dicionário que utilizam o alfabeto latino.

Isso posto, apresentamos, como exemplificação, algumas das etapas específicas de trabalho e salientamos que não seguem uma ordem necessariamente cronológica e que, por vezes, foram realizadas concomitantemente.

Etapas metodológicas específicas com comentários adicionais

- a) Elaboração de um questionário de sondagem sobre a importância do dicionário para um grupo de 13 estudantes de Medicina de uma universidade brasileira e para 5 estudantes de uma universidade italiana, com a qual temos contatos profissionais. Os resultados foram satisfatórios e mostraram que 53% desses jovens expressaram que esse tipo de obra é “importante” ou “muito importante” e 47% acharam que é “interessante, mas não essencial”.
- b) Criação de um domínio para o site do DicTrans. Contratação de um especialista para construir o *site* (www.dictrans.org).
- c) Confecção de um modelo de ficha terminológica para organização das entradas (ex.: verbete, definição, exemplo, equivalências em outras línguas, campo semântico, entre outros). Para isso, realizou-se uma revisão dos modelos de fichas terminológicas existentes na literatura.
- d) Elaboração de uma ficha terminológica que atendesse às necessidades da obra, levando-se em consideração o perfil do consultante.

Figura 5 – Ficha terminológica do DicTrans

The image shows a web form for adding terminology entries. At the top left is the 'DICTRANS' logo. The main heading is 'Incluir Verbetes'. Below it are several text input fields, each with a label to its left: 'Entry in English', 'Concept', 'Source', 'Examples', 'Entrada em português', 'Espressione in italiano', 'Expresión en español', and 'Comments/Observations'. Below the 'Comments/Observations' field is a 'Topic' label and a dropdown menu. The dropdown menu is open, showing a list of categories: Anatomia, Clínica, Diagnóstico, Exames complementares, Funções cognitivas, Funções fisiológicas, Método de validação, and Medicina. To the right of the dropdown is a large empty text area. At the bottom of the form are two buttons: 'Salvar' (Save) and 'Limpar' (Clear).

Fonte: Delgado et al (2014)

As fichas terminológicas elaboradas para cada verbete serviram de base para a construção do DicTrans, oferecido gratuitamente através do sítio **dictrans.org**²⁸. O papel das fichas terminológicas foi o de auxiliar na organização das informações escolhidas, atuando como “pano de fundo” dos verbetes incluídos na obra. Nesse sentido, a ficha apresentada na Figura 5 está disponível apenas para os autores, os quais podem editar as informações contidas na obra. Vale salientar que as fichas variam de modelo para diferentes produtos, pois obedecem aos objetivos e critérios estabelecidos pela equipe de trabalho, levando em consideração os usuários pretendidos. Nesse sentido, as Figuras 6 e 7 mostram o *layout* da entrada de

²⁸ Elaboramos 338 fichas terminológicas, que correspondem ao número de entradas disponíveis para consulta no site.

abstinence from tobacco, e da consulta por letra, que serão visualizados pelo consulente.

Figura 6 – Exemplo de entrada

The screenshot shows the DicTrans interface with a search bar containing 'Ir' and navigation buttons for 'Meu Profile', 'Verbetes', 'Usuários', 'Semantic Fields', and 'Sair'. The main content area is titled 'Visualizar Verbetes' and displays the following information:

- Entry in English:** Abstinence from tobacco
- Concept:** 1) O uso diário de tabaco durante um período mínimo de várias semanas; 2) Cessação abrupta (ou redução do uso de tabaco), seguida, no prazo de 24h, por quatro (ou mais) dos seguintes sinais ou sintomas: irritabilidade, frustração ou raiva; ansiedade; dificuldade de concentração; aumento do apetite; inquietação; humor deprimido; insônia. 3) Os sinais ou sintomas do critério (2) causam sofrimento clinicamente significativo ou prejuízo no funcionamento social, profissional ou em outras áreas importantes da vida do indivíduo. 4) Os sinais ou sintomas não são atribuíveis a outra condição médica nem são mais bem explicados por outro transtorno mental, incluindo intoxicação por ou abstinência de outra substância.
- Source:** American Psychiatric Association. (2014). Manual Diagnóstico e Estatístico de Transtornos Mentais (5. ed) (M. I. C. Nascimento, Trad.). Porto Alegre, RS: Artmed.
- Examples:** 1) In order to address the possibility that any observed worsening of mood might represent the natural course of BD rather than the effects of smoking abstinence, change in mood from pre-abstinence to post-abstinence among abstainers was compared with mood changes observed during a random 13-week series of symptom ratings in continuing smokers. 2) We also found that, on average, smoking abstinence was not associated with an increase in symptoms of depression or mania during the two months following the initiation of abstinence. 3) Another gap in the literature regarding tobacco smoking among individuals with BD is how initiation of smoking abstinence impacts symptoms of the disorder.
- Entrada em português:** Abstinência de tabaco
- Espressione in Italiano:** Astinenza da tabacco
- Expresión en español:** Abstinencia de tabaco
- Comments/Observations:**
- Semantic field:** Diagnóstico

Additional notes on the right side of the interface state: 'Todas as letras minúsculas e no singular' and 'Definições e contextos não podem exceder a 10 linhas textuais'. A button 'Imprimir Verbetes' is located at the bottom left.

Fonte: Equipe DicTrans (2018)

Figura 7 - Exemplo da consulta por letra

The screenshot shows the DicTrans interface with a search bar containing 'a' and a button 'Ir'. Below the search bar, it indicates 'Foram localizados 338 registros'. The results are displayed as follows:

- abstract thinking**
português: pensamento abstrato
italiano: pensiero astratto
espanhol: pensamiento abstracto
- actigraphy**
português: actigrafia
italiano: actigrafia
espanhol: actigrafia
- acute treatment phase**
português: fase aguda do tratamento
italiano: fase acuta del trattamento
espanhol: fase aguda del tratamiento

Fonte: Equipe DicTrans (2018)

e) Inserção dos termos selecionados e revisados em fichas terminológicas por meio do preenchimento das informações estabelecidas. Nessa etapa, incluímos os equivalentes dos termos nas outras línguas, encontrados em corpora especializados. Preenchimento e revisão das fichas terminológicas.

- f) Acréscimo das pronúncias dos termos nos quatro idiomas e de informações sobre a prática clínica da médica psiquiatra através de áudios na língua portuguesa e tradução desses excertos em língua espanhola, inglesa e italiana.
- g) Criação de um logo para o DicTrans e identificação da marca (imagem abaixo). As parábolas representam os momentos críticos de mania e de depressão da doença, como se fossem um gráfico que registra estes pontos altos e baixos. Além disto, eles simbolizam um sorriso e uma boca para baixo, demonstrando a instabilidade emocional. O círculo ao centro é o elemento que passa a sensação de equilíbrio, simbolizando o momento de estabilidade da doença. Já as cores utilizadas possuem grande associação com área da saúde e passam uma sensação de seriedade. Segundo Farina, Perez e Bastos (2011), a cor azul é a preferida por mais da metade da população ocidental e a mais lembrada quando os ocidentais querem se referir à simpatia, à harmonia, à amizade e à confiança. Já a cor cinza, é neutra e possui associação afetiva com seriedade, maturidade e sabedoria.

Figura 8 – Logotipo do DicTrans



Fonte: Furlan (2015)

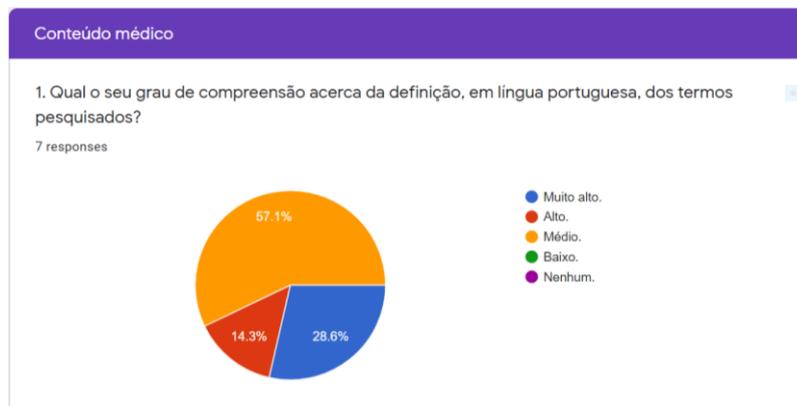
- h) Publicação do DicTrans (gratuito no *site* www.dictrans.org). Divulgação do dicionário para profissionais e estudantes da área da Saúde, e jornalistas, tradutores e revisores linguísticos, que tratem de temas relacionados à saúde humana.
- i) Criação de um endereço eletrônico para que os leitores entrem em contato com as autoras titulares em caso de sugestões e dúvidas (chat@dictrans.org).
- j) Pedido para registro de marca no Instituto Nacional de Propriedade Intelectual (INPI), já concedido sob o número 917698843.
- k) Pedido para o registro de software, já concedido sob o número BR512019001658-0.

l) Solicitação de Direito Autoral, que está sendo analisado pela Biblioteca Nacional.

m) Validação do dicionário: essa etapa, recentemente planejada e posta em prática, será fundamental para futuras melhorias da obra. Elaboramos um questionário, através do *Google Forms*, com 18 (dezoito) perguntas de múltipla escolha (com exceção da última, opcional, para sugestões e críticas) sobre os aspectos que consideramos basilares para avaliarmos o dicionário: conteúdo médico, recursos linguísticos e usabilidade. Esses aspectos representam a multidisciplinaridade inerente ao processo teórico e metodológico de elaboração do DicTrans. Disponibilizamos o questionário a profissionais da saúde (psiquiatras, incluindo médicos de especialidades diversas, fisioterapeutas, biólogos, entre outros).²⁹ Até o momento, sete respondentes participaram da pesquisa de opinião.

Abaixo, amostras de uma pergunta de cada um dos três aspectos, as respostas em forma de gráficos e comentários dos respondentes, e reflexões adicionais de nossa autoria.

Figura 9 – Conteúdo médico



Fonte: Equipe DicTrans (2020)

O gráfico de pizza mostra que o grau de compreensão da definição foi “Muito alto” para 28.6% dos respondentes, ou seja, 2 respondentes assimilaram muito bem o conteúdo, enquanto que 14.3% (1 respondente) teve um grau de compreensão alto e 57.1% (4 respondentes), tiveram um grau médio de assimilação. Não houve respostas para as categorias “Baixo” e “Nenhum”.

²⁹ Convidaremos, em momento oportuno, os estudantes das áreas da Saúde a colaborarem conosco. Suas respostas serão igualmente benéficas para o contínuo aperfeiçoamento do DicTrans.

Os comentários dos participantes foram os seguintes: “Estou familiarizada com os termos”; “Tenho alguém na família portadora do Transtorno Bipolar”; “Entendi tudo” e “Não sou da área da Psiquiatria”. Salientamos que nem todos justificaram suas respostas. Como as justificativas eram necessárias para responderem a pergunta posterior, alguns utilizaram de subterfúgios como números, letras ou símbolos no lugar dos comentários para continuarem a atividade.

Dos respondentes, seis são médicos (quatro psiquiatras e dois de outras especialidades) e um é fisioterapeuta. Dessa forma, acreditamos que os resultados mostram uma consonância com o objetivo da obra: embora tenha sido idealizada para profissionais da Saúde, é compreensível que o grupo de médicos, principalmente, psiquiatras, sintam-se mais confiantes ao ler conteúdo sobre o tema.

Figura10 – Conteúdo multilíngue



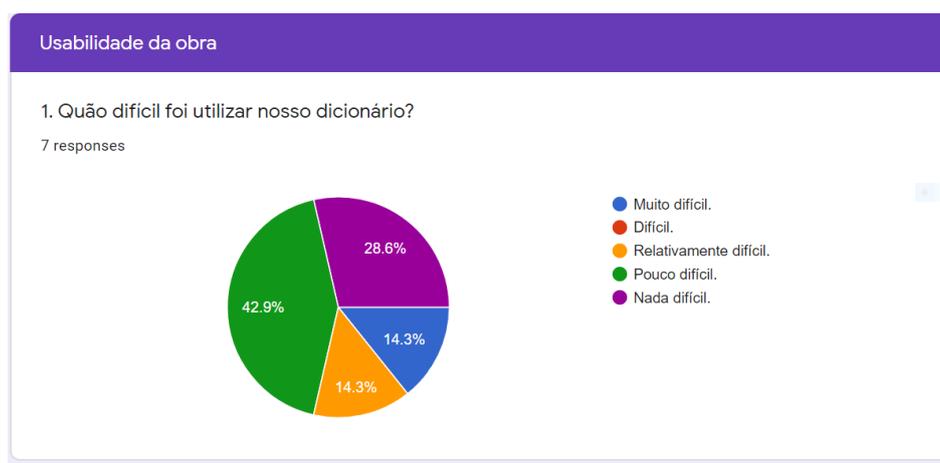
Fonte: Equipe DicTrans (2020)

Com relação ao grau de dificuldade linguística dos exemplos, escritos em língua inglesa, os resultados foram os seguintes: “Muito alto” para 14.3% dos respondentes, ou seja, 1 respondente teve dificuldade em compreender o conteúdo dos exemplos. Ambas as categorias “Médio” e “Baixo” obtiveram o valor de 28.6% (2), indicando que, na primeira, os respondentes tiveram certa dificuldade em compreender o conteúdo dos exemplos, provavelmente apenas assimilando sua ideia principal. Já na segunda categoria, outros dois respondentes acharam que o grau de dificuldade foi baixo, possivelmente pelo fato de compreender melhor o tema abordado e ter um nível de proficiência leitora especializada mais alto. Para a categoria “Nenhum”, houve o mesmo valor de 28.6% (2 respondentes), indicando que leem com muita facilidade o conteúdo sobre o THB em língua inglesa. Não houve respostas para a categoria “Alto”.

Os comentários dos respondentes foram estes: “Dificuldade já existente”, “Precisaria ouvir a pronúncia” (2 respondentes); “Não tive dificuldade”. Da mesma forma que a pergunta acima, alguns respondentes utilizaram números, letras ou símbolos no lugar dos comentários para continuarem a atividade.

A obra oferece áudios das pronúncias dos 388 termos disponíveis, assim como áudios explicativos (ambos nas quatro línguas) com o intuito de ampliar o conteúdo de cada verbete. A expansão do número de áudios explicativos para todos os termos está prevista para o ano de 2020, indo ao encontro de uma das perguntas do questionário, que intenta saber se são recursos adicionais importantes (a resposta foi positiva nesse quesito). Parece-nos que a inclusão de áudios dos exemplos na língua inglesa, como sugerida por um dos respondentes, seja uma opção plausível a longo prazo, visto que iniciamos as buscas por equivalentes, nos quatro idiomas, de todas as definições do dicionário, para o justificarmos como multilíngue.

Figura 11 – Usabilidade da obra



Fonte: Equipe DicTrans (2020)

Com relação à usabilidade da obra, temos os seguintes resultados: 14.3%, ou seja, 1 respondente, apareceu em duas categorias: “Muito difícil” e “Relativamente difícil”. A categoria “Nada difícil” foi escolhida por 2 respondentes (28.6%) enquanto que a categoria “Pouco difícil” obteve o número de 42.9%, que corresponde ao número de 3 respondentes.

Os profissionais reportaram os seguintes comentários: “Não achei fácil”; “É um pouco intuitivo”; “É bem simples de transitar”; “Fácil acesso” e “Sem dificuldade”.

Embora apenas 2 respondentes acharam o dicionário muito ou relativamente difícil, ficaremos atentos ao contínuo aperfeiçoamento do quesito “Usabilidade”. De fato, a usabilidade do site é um ponto sob a atenção dos titulares da obra. Deseja-se que ele seja de fácil acesso, agradável navegação e que atenda diretamente às necessidades do público a qual se destina. Para tanto, pesquisas referentes ao seu uso vêm sendo desenvolvidas, a fim de que se possa determinar a melhor engenharia da informação para o produto. A premissa deste trabalho consiste em definir adequadamente o perfil do leitor, e em seguida testar estratégias de navegação e organização de conteúdo direcionadas a ele. O resultado esperado é manter o foco do navegante no conteúdo literal da obra, e não exigir dele esforço de entendimento da usabilidade.

Descritas as principais etapas metodológicas da obra, passamos, agora, às considerações finais.

Considerações finais

Diante da importância do acesso ao conhecimento dos avanços da ciência para a capacitação do profissional de Ciências Médicas, acreditamos que o DicTrans tenha potencial para cumprir o seu principal objetivo: explicar, de forma contextualizada, os principais conceitos sobre o THB, para facilitar a compreensão, entre profissionais da área da Saúde, sobre uma das mais complexas e devastadoras doenças da atualidade.³⁰

Portanto, acreditamos que o dicionário possa se tornar uma ferramenta de busca por informações fidedignas e atualizadas sobre o domínio do THB, visto que como produto continuado da presente investigação, prevemos sua atualização anual, para que se justifique como ferramenta de uso on-line.

Sem esgotar os objetivos pensados para a obra, esperamos que ela também possa proporcionar aos pesquisadores linguístico-terminológicos sugestões metodológicas baseadas em abordagens quantitativas e qualitativas de análise, que possam ser replicadas em outros domínios de especialidade, assim como fomentar o diálogo entre os especialistas sobre tais metodologias.

Estão previstos, ainda, como produtos continuados da pesquisa, publicação de artigos nas quatro línguas, maior divulgação em eventos e redes sociais, criação de

³⁰ No período de outubro de 2019 e abril de 2020, o número de acessos ao dicionário foi de 978 (visitantes únicos) e de 1.171 (número de visitas).

um blog com conteúdo inteligente sobre o THB, atualização sistemática do software e do corpus de estudo, este último para inclusão de novos termos na obra (há cerca de mais 600 termos disponíveis para análise, extraídos pela ferramenta computacional utilizada).

Há outros dois tópicos que também serão explorados como trabalhos futuros. O primeiro, de cunho computacional, visa explorar outras técnicas de extração automatizada de termos. Apesar desta ter sido apenas uma das etapas de todo o processo de construção, a extração computacional proporciona agilidade na escolha inicial dos termos frequentes e relevantes. Buscar outros softwares ou técnicas de extração mais precisos auxiliarão o trabalho posterior de refinamento realizado pela especialista do domínio, e também servir de apoio aos linguistas durante o processo de busca de equivalentes dos verbetes.

O segundo tópico, concentrado no campo da linguística, almeja reforçar a internacionalização científica, incluindo novos idiomas ao dicionário. Esse trabalho ampliará o uso da ferramenta, bem como atrairá mais pesquisadores que possam colaborar com a construção da obra, sejam linguistas, especialistas de domínio ou cientistas da computação. Nesse sentido, a língua francesa está sendo discutida com uma possível candidata a compor a obra, devido ao grande volume de países francófonos e a uma provável parceria com uma universidade canadense.

Por fim, almejamos compreender possíveis fenômenos linguísticos e terminológicos que potencializam a complexidade da linguagem especializada sobre o THB, no âmbito da Acessibilidade Textual e Terminológica. Dessa forma, e a partir da análise da complexidade dos corpora, nas línguas apresentadas no dicionário, textos acessíveis serão produzidos destinados ao adulto leigo com baixa escolaridade ou limitado letramento científico. Esses textos simplificados serão incorporados no site e no blog e estarão disponíveis através de *banners* destinados a esse fim, como forma de atender outro perfil de leitor. Entendemos que o DicTrans, dentro de um leque de possibilidades de atualização e expansão, tenha potencial para impulsionar a internacionalização científica entre leitores com diferentes níveis de conhecimento e letramento, ser uma opção de compreensão sobre o intrincado labirinto de uma doença como o THB e proporcionar alívio emocional aos seus portadores e familiares.

Referências

.

AUBERT, F. H. *Introdução à metodologia de pesquisa terminológica bilíngue*. São Paulo: CITRAT/FFLCH/USP, 1996.

BEVILACQUA, C. R. Por que e para que a Linguística de Corpus na Terminologia. In: TAGNIN, S.; BEVILACQUA, C. R. *Corpora na Terminologia*. São Paulo: Hub Editorial, 2013.

CABRÉ, M. T. *Terminología: Representación y Comunicación*. Elementos para una Teoría de Base Comunicativa y otros Artículos. Barcelona: Universitat Pompeu Fabra, 1999.

CIAPUSCIO, G. *Textos especializados y terminología*. Barcelona: IULA, 2003.

DELGADO, H. O. K. Proposta de um didática de tradução de linguagens especializadas para licenciandos em Letras. *Tese de doutorado*. Porto Alegre: UFRGS, 2012.

DELGADO et al. *A Ficha Terminológica no Projeto DicTrans: aspectos definitórios e estruturais*. Palestra apresentada no VIII EnELLE. Pontifícia Universidade Católica do Rio Grande do Sul, 2014.

DIAS DA SILVA, B. C. A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais. *Tese de doutorado*. Araraquara: Universidade Estadual Paulista, 1996.

FARINA, M.; PEREZ, C.; BASTOS, D. *Psicodinâmica das cores em comunicação*. São Paulo: Editora Blucher, 6ª Ed. 2011.

FINATTO, M.J.B. *Complexidade Textual em artigos científicos: contribuições para o estudo do texto científico em português*. Organon (UFRGS), v. 50, p. 30-45, 2011.

CAMARGO, C.H.P.; DIAZ FUENTES, D.; DINIZ, L. F.; *Neuropsicologia - Teoria e Prática*. São Paulo: Artmed, 2014.

GAZELLE, F. ANDREAZZA, A.C. SANT'ANNA, M.K. *Diagnóstico precoce do Transtorno Bipolar*. Revista Brasileira de Psiquiatria. v.27, no.1. São Paulo: ABP, março de 2005.

KAPCZINSKI, F.; QUEVEDO, J. et. al. *Transtorno Bipolar: teoria e clínica*. Porto Alegre: Artmed Editora S.A., 2009.

KAPZINSKI, F.; VIETA, E.; MAGALHÃES, P.; BERK, M. (Org.). *Neuroprogressão e Estadiamento no Transtorno Bipolar*. Porto Alegre, RS: Artmed, 2017.

Manual diagnóstico e estatístico de transtornos mentais [recurso eletrônico]: DSM-5 [American Psychiatric Association]. Porto Alegre: Artmed, 2014.

KRIEGER, M. G. Porque Lexicografia e Terminologia: relações textuais. *Anais do CELSUL*. São Leopoldo: UNISINOS, 2008.

KRIEGER, M. G., FINATTO, M. J. B. *Introdução à Terminologia: teoria & prática*. 1a ed. São Paulo: Contexto, 2004.

LOPES, L.; FERNANDES, P.; VIEIRA, R.; FEDRIZZI, G. *ExATOlp: An automatic tool for term extraction from Portuguese language corpora*. LTC'09, Poznam, Poland, 2009.

LOPES, L.; VIEIRA, R.; FERNANDES, P.; COUTO, G. *ExATOlp: extraction of language resources from Portuguese corpora*". In: *International Conference on Computational Processing of the Portuguese Language - PROPOR*, 2012.

LOPES L. *Extração automática de conceitos a partir de textos em língua portuguesa. Tese de doutorado*. Porto Alegre, RS. Pontifícia Universidade do Rio Grande do Sul (PUCRS), 2012.

MANNING, C. D.; SCHÜTZE, H. *Foundations of Statistical Natural Language Processing*. Cambridge: The MIT Press, 1999.

RODILLA, B.M.G. *La Ciencia empieza em la palabra*. Barcelona: Ediciones Península, 1998.

SARDINHA, T. B. *A influência do tamanho do corpus de referência na obtenção de palavras-chave usando o programa computacional Wordsmith tools*. The ESpecialist. São Paulo: PUCSP, vol. 26, nº 2, 2005.

SANCHEZ, A. ET AL (Orgs.) *Corpus Linguístico del Español Contemporáneo – Fundamentos, Metodología, y Aplicaciones*. Madrid: SGEL, 1995.

TEIXEIRA, R. B. S. *Termos de (Onco) Mastologia: uma abordagem mediada por corpus. Dissertação de Mestrado* (Programa de Estudos Pós-Graduados em Linguística Aplicada). Pontifícia Universidade Católica de São Paulo, 2010.

Recebido em: 16 mar. 2020

Aceito em: 18 abr. 2020