

A PROPOSTA DE UM GUIA DE REFERÊNCIA SOBRE TERMINOLOGIA PARA ESTUDANTES DE CURSOS DE LICENCIATURA EM LETRAS ATRAVÉS DO PROCESSAMENTO DE *CORPUS VIA WORDSMITH TOOLS*

HELOÍSA ORSI KOCH DELGADO*

DÉBORA MONTENEGRO PASIN**

LUCAS ZAMBRANO ROLLSING***

RESUMO

Relatamos a etapa inicial da construção de um guia de referência sobre Terminologia para estudantes de cursos de licenciatura em Letras. Focamos na elaboração do *corpus* de estudo e das fichas terminológicas, as quais antecedem à obra *per se*. Utilizamos o *WordSmith Tools* para extrair as palavras-chave, a fim de, posteriormente, tornarem-se verbetes terminológicos. As palavras-chave foram obtidas de acordo com seu valor de chavicidade (razão estatística que caracteriza as palavras-chave oriundas da comparação de uma lista de palavras de um *corpus* de estudo com a lista de palavras de um *corpus* de referência). Para fins de amostra, trouxemos dois exemplos de fichas e seus respectivos verbetes.

PALAVRAS-CHAVE: Guia de referência sobre Terminologia, termos, fichas terminológicas.

1. INTRODUÇÃO

Alguns campos científicos são mais consagrados ou mais populares, pela comunidade acadêmica em geral, em detrimento de

* Doutora em Letras pela Universidade Federal do Rio Grande do Sul (UFRGS), Professora Adjunta do Curso de Letras e Assessora-Chefe da Assessoria para Assuntos Internacionais e Interinstitucionais da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Rio Grande do Sul, Brasil. E-mail: heloisa.delgado@pucrs.br

** Doutoranda do Programa de Pós-Graduação em Medicina e Ciências da Saúde e Especialista em Estudos da Tradução pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Rio Grande do Sul, Brasil. E-mail: debora.pasin@acad.pucrs.br

*** Mestre e Licenciado em Letras pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre, Rio Grande do Sul, Brasil. E-mail: lucas.rollsing@acad.pucrs.br

outros. Isso se dá por inúmeras razões: i) pelo impacto social das diferentes pesquisas; ii) por suas aplicabilidades mais ou menos imediatas; iii) por suscitarem maior interesse ou, mesmo, iv) por terem mais tradição histórica. Nesse âmbito de diferentes repercussões, de forma mais restrita, Krieger (2010, p. 163), por exemplo, afirma que, as áreas humanas do conhecimento:

[...] não são facilmente visíveis pelo cidadão e tendem a ser menos valorizadas. Em nossa situação particular, a “invisibilidade” é comum, sobretudo porque a “palavra” é um bem coletivo, natural ao homem. Consequentemente, o profissional que com ela “lida” não recebe a devida valorização.

Assim, poderíamos dizer que a Linguística, como um todo ou em suas subáreas, ainda sofre da “invisibilidade” comentada por Krieger. Na Linguística, se tomarmos apenas o campo da Terminologia e o entendermos como um ramo da Linguística Aplicada, vemos um exemplo resultante do crescimento da pesquisa das áreas Humanas e Sociais, embora os estudos terminológicos, em nível de graduação, ainda sejam limitados no Brasil. Mesmo com o grande desenvolvimento atual da pesquisa informatizada do português (SARDINHA, 2004), que tem favorecido contatos e reflexões sobre o funcionamento e perfis de vocabulário das linguagens técnicas e científicas, disciplinas de graduação sobre Terminologia tendem a se concentrar em cursos de Letras/ Tradução em alguns cursos de Bacharelado em Linguística e/ou Letras e também em cursos de Biblioteconomia. Nos cursos de Licenciatura em Letras, no cenário da formação de professores de línguas, sua presença ainda hoje segue sendo rara.

A esse respeito, há mais de uma década, Krieger e Finatto (2004, p. 7) já reconheciam que poderiam: “[...] contribuir para diminuir a lacuna de produção bibliográfica sobre Terminologia em língua portuguesa”. Nesse contexto, vale lembrar que a Terminologia, como uma disciplina com contornos delimitados, é relativamente recente em todo o panorama internacional. A lacuna visualizada pelas pesquisadoras há dez anos, ainda pode ser percebida, mesmo que com menos intensidade. Essa realidade acentua-se quando consideramos a diversidade teórico-aplicada da área da Linguística e suas subdivisões no que concerne a sua aplicabilidade, léxicos e termos.

No entanto, sabemos que a Terminologia além de contribuir para o desenvolvimento dos estudos referenciais em geral, faz-se pertinente nas sociedades modernas dada a nossa crescente exposição aos conhecimentos científicos e tecnológicos. No campo das Ciências da Saúde, vemos exemplos diários, em diferentes meios, em que conceitos científicos complexos e suas terminologias tornam-se usos corriqueiros de linguagem, como é o caso atual do termo *microcefalia*. Em um mundo cada vez mais avançado científica e tecnologicamente, temos a rápida expansão das práticas semiolinguísticas que definem esses campos disseminando-se sociedade adentro. Isso, por si só, já mostra a necessidade de descrição e análise de suas metalinguagens.

Em meio a esse quadro, os recursos sobre Terminologia, e sobre os Estudos do Léxico em geral, ainda parecem relativamente densos e escassos, especialmente em se tratando de obras de referência como dicionários, glossários, manuais, entre outros, para alunos dos cursos de Letras em geral. Sobre isso, tendo em mente a questão da lacuna do ensino sobre o léxico, Barbosa (2014, p. 311) afirma que:

[...] o exame das práticas semiolinguísticas dos enunciadores e dos enunciatários do discurso pedagógico tem permitido observar claramente que a questão do ensino do léxico não é considerada importante, é, até mesmo, frequentemente esquecida ou desconhecida, [...], de que resulta qualitativamente e quantitativamente, um baixo rendimento [...].

A bem de se promover uma maior aproximação entre o futuro professor de Letras e a Terminologia – considerando-se sua inserção como uma das mais jovens ciências do léxico –, acreditamos que obras de referência consigam dirimir a lacuna de um campo científico, cuja complexa terminologia pode vir a separar seu leitor principiante de uma experiência mais produtiva ao ingressar em um dado ramo do conhecimento. Nesse sentido, salientamos que nosso texto está longe de esgotar os meandros dessa complexa e multifacetada área. Não abordaremos, por exemplo, vertentes oriundas da Socioterminologia (GAUDIN, 1993), da Teoria Sociocognitiva da Terminologia ou Socioontologia (LAWSON, 2004), da Teoria Realista da Terminologia (TEMMERMMAN, 2001), da Terminologia Cultural (DIKI-KIDIRI,

2000), ou, ainda, da Terminologia de Perspectiva Textual (CIAPUSCIO, 1998). Nossa iniciativa, de base didática, é dirigida a estudantes iniciantes de cursos de licenciatura em Letras, que se deparam com disciplinas sobre Terminologia¹, para as quais, acreditamos, conceitos técnico-científicos possam ser apresentados em uma linguagem mais acessível numa primeira experiência.

2. TERMINOLOGIA: ORIGENS E EVOLUÇÃO²

A Terminologia, enquanto prática que possui como objeto de estudo os termos especializados, não é recente. Recordemos os trabalhos feitos por Lavoisier e Berthold no campo da química e Linné na botânica e zoologia, no século XVIII, para destacar o interesse dos especialistas em denominar os conceitos científicos de seus domínios do conhecimento. No entanto, é no século XIX que botânicos (1867), zoólogos (1889) e químicos (1892) expressam, em colóquios internacionais, a necessidade de estabelecerem, de forma sistemática, regras de formação dos termos para cada área. Vale lembrar que a mola propulsora dessa necessidade está calcada no advento da internacionalização progressiva da ciência.

Enquanto o século XIX é marcado pela preocupação dos cientistas pela sistematização de terminologias, é somente no século XX que pesquisadores das áreas técnicas começam a se inquietar com essa questão. Surge, no campo da engenharia, o austríaco Eugene Wüster (1898-1977), considerado o fundador da terminologia moderna e o principal representante da chamada Escola de Viena e o russo D. S. Lotte (1889-1950), teórico e fundador da Escola Soviética de Terminologia. Em 1904, é fundada a primeira associação internacional de normalização, a CEI (Comissão Eletrotécnica Internacional).

Em 1931, Wüster apresenta sua tese de doutorado *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik* (Padronização Internacional de Linguagem Técnica, especialmente em Engenharia Elétrica) na Universidade de Viena, cuja relevância reside no ponto de partida para a normalização terminológica. Picht (1984) sustenta que a difusão da versão russa de sua obra originou a criação do Comitê Técnico (TC37) da Associação Internacional para a Padronização (ISA – *International Standardization Association*), cujo

objetivo era o de unificar os métodos de trabalho e a apresentação de terminologias especializadas.

As preocupações de Wüster se concentravam nas questões metodológicas e normativas, visto que considerava a terminologia um instrumento de trabalho que deveria auxiliar, de forma eficaz, na desambiguação da comunicação técnica e científica. Seu interesse pela teoria da terminologia surgiria mais tarde. Na abertura do simpósio da Infoterm (*International Information Centre for Terminology* ou Centro Internacional de Informação sobre Terminologia), em 1975, em Viena, o engenheiro austríaco atribui a paternidade intelectual da teoria da terminologia a quatro cientistas, a saber: 1) O alemão A. Scholman, autor de vocabulários técnicos em seis línguas, publicados entre 1906 e 1928 e o primeiro pesquisador a considerar o caráter sistemático dos termos de especialidade; 2) O linguista suíço F. de Saussure (1916), o primeiro a destacar a sistematicidade das línguas; 3) O russo E. Dressen, pioneiro em enfatizar a importância da normalização e o propulsor da organização ISA (anterior a 1991), precedente da ISO (*International Organization for Standardization* ou Organização Internacional para a Padronização), atual organismo de normalização internacional; e 4) O inglês J. E. Holmstrom que, desde a UNESCO, em 1971, impulsionou a difusão internacional das terminologias e o primeiro a reivindicar um organismo que se ocupasse desse assunto (surgiu, então, o InfoTerm).

Wüster (1979), criador da Teoria Geral da Terminologia (TGT), via os termos como dissociados da gramática, do contexto e do discurso, ou seja, como unidades com vida independente; não eram, portanto, polissêmicos, não tinham sinônimos, nem homônimos. Seu pressuposto era de que o conhecimento especializado é uniforme seja qual fosse a língua e cultura que o cercasse, não havendo variação terminológica de acordo com as situações contextuais que se apresentassem em seu entorno. Com isso, o autor afirma seu interesse pela precisão de conceitos, textos escritos, formas internacionais de designação e controle consciente da evolução (planejamento, unificação e padronização), e uso exclusivo da abordagem onomasiológica (parte do conceito para a designação). O que vai de encontro a uma proposta eminentemente pragmática de língua, pois não postula os termos como elementos das línguas naturais e que devem ser estudados dentro de seus contextos de

uso. Ainda assim, a teoria wüsteriana justifica seu papel de referência porque auxiliou a Terminologia a estabelecer-se como campo de conhecimento com fundamentos epistemológicos e objeto próprio de investigação (KRIEGER; FINATTO, 2004).

Voltamos, agora, nossa atenção para os linguistas e especialistas das ciências humanas. Seu interesse pela Terminologia ocorre apenas, e de forma irregular, na primeira metade do século XX, por volta da década de 1950, pois havia pouco entusiasmo desse grupo de estudiosos pelo estudo dos termos, mesmo sendo a língua/linguagem o centro de interesse científico desse coletivo. Apenas nos últimos anos, evidenciou-se o esforço em construir uma teoria que contemplasse os princípios que regem as línguas reais e possíveis concebidas como instrumentos de comunicação.

Com relação ao aspecto comunicativo, reconhecido nas últimas décadas, houve uma evolução subsequente das teorias terminológicas que, à época, não havia sido observada pelos teóricos da área: a da função comunicativa da linguagem especializada. Entra em cena a Teoria Comunicativa da Terminologia (TCT), proposta por Cabré (1999), com o objetivo de valorizar os aspectos pragmáticos das linguagens especializadas. Atestando que as unidades terminológicas formam, sim, parte da linguagem natural e são passíveis de variação assim como qualquer elemento linguístico o é. Essa nova visão trouxe para os estudos terminológicos uma redefinição do conceito de termo, assim como da sua natureza linguístico-comunicativa, o que a autora veio a chamar de constituição poliédrica do termo (CABRÉ, 1999).

Essa nova corrente dos estudos terminológicos define, então, termos como itens lexicais, ou seja, estão inseridos em um contexto linguístico e de uso, os quais contribuem para a articulação do estatuto terminológico da unidade lexical e reconhecem a presença de sinonímias e de variações nos repertórios terminológicos como funcionais. Isso significa relacioná-lo a uma série de componentes de natureza semiótica, pragmática e ideológica que permeiam qualquer processo comunicativo.

Nos últimos decênios, a Terminologia tem sido foco de sistemático desenvolvimento e de profunda reflexão sobre seus princípios e métodos, alcançando reconhecimento de sua importância social e política em nível nacional e internacional. Parece que podemos

avistar um processo, para o futuro da Terminologia, que conduza a uma racionalização progressiva do trabalho e uma funcionalidade mais adaptada a cada tipo de circunstância. Se a Terminologia nasce com a finalidade de suprir a necessidade de criação de novos conceitos em uma sociedade em constante avanço técnico e científico e, assim, garantir a comunicação entre especialistas, seu desenvolvimento tem que conferir um caráter cada vez mais polivalente e seletivo, que deva servir para melhor resolver as necessidades humanas suscitadas por uma nova cultura e aproximar a tecnologia das pessoas, melhorando sua qualidade de vida e suas relações interpessoais e intergrupais, cada vez mais amplas e complexas. A partir desses pressupostos, acreditamos que eles podem ser transpostos para uma experiência de sala de aula de graduação que seja mais profícua e conduza à formação de docentes e pesquisadores com estofamento em sua área de formação.

Na seção seguinte, apresentamos brevemente o conceito e os objetivos das “fichas terminológicas”, uma etapa oriunda da face aplicada da Terminologia, a Terminografia. Com elas, damos prossecução a qualquer produto terminográfico, seja glossários especializados, dicionários terminológicos, entre tantas outras produções.

3. CONCEITO DE FICHA TERMINOLÓGICA

As fichas terminológicas são essenciais para qualquer trabalho terminográfico. É uma fase inicial desenhada conforme as necessidades do projeto em vista. Sua elaboração é um passo metodológico preliminar à construção do produto propriamente dito. A ficha arrola todas as informações sobre uma unidade terminológica. De acordo com Krieger e Finatto (2004, p. 136):

[n]ela, constam informações indispensáveis, tais como a fonte textual de coleta de um termo, segmentos de texto onde esse termo ocorre, seus contextos de uso, informações sobre variantes denominativas, sinônimos, construções recorrentes que o acompanham.

Esse registro congrega tanto dados terminológicos quanto terminográficos, ambos necessários para o andamento do trabalho. Dados terminológicos são informações referentes à natureza linguística

e às relações contextuais de cada unidade terminológica uma com a outra. Segundo Barros (2004, p. 212), dessa ordem terminológica, destacamos: a) a unidade linguística em si (aspectos morfológicos, sintáticos e pragmáticos do possível verbete); b) a descrição da unidade terminológica (exemplos, figuras, contextos de uso etc.); e c) as relações intersignificativas de ordem semântica (sinonímia, parassinonímia, hiperonímia-hiponímia). Dados terminográficos são informações que se complementam com as referidas acima, inserindo uma outra gama de aspectos concernentes à organização estrutural da obra. De acordo com Barros (2004, p. 212), os dados mais utilizados são:

a) domínio de aplicação do termo; b) indicativo de país, de língua e de autoridade; c) notas; d) símbolo de classificação; e) fontes; f) data do primeiro registro e da última atualização da ficha; g) número de série; h) frequência de ocorrência; i) descritores; j) língua de partida e língua de chegada.

Contudo, esses dados podem ser inclusos ou não nas fichas, e nem sempre serão incorporados no produto final, pois muitos deles têm por função orientar o trabalho do terminográfico, a fim de torná-lo mais eficaz. Esses dados, sejam terminológicos, sejam terminográficos, são incorporados ao trabalho sempre que necessário de acordo com os objetivos estabelecidos previamente pelos pesquisadores.

O conjunto de fichas terminológicas com as informações acima descritas constitui um fichário, o qual pode ser de dois tipos: a) sobre os dados de referência do termo; e b) sobre os dados terminológicos em si.

O primeiro remete à “localização” do verbete quanto à bibliografia, contendo informações referentes ao autor que cunhou ou usa o termo, ao título da obra no qual ele pode ser encontrado, à cidade, à editora que detém os direitos autorais, ao ano de publicação, ao número de páginas etc.; itens que caracterizam o verbete não linguisticamente. O segundo se aproxima mais da presente proposta, pois se trata dos verbetes em si, desde sua recolha, registro e análise. Nesse fichário, apresentamos três modelos de fichas terminológicas: 1) ficha de trabalho, 2) ficha de síntese e 3) ficha de remissiva.

A primeira registra os dados encontrados na pesquisa bibliográfica, a segunda se centra no conteúdo conceptual da unidade a fim de

adequá-la à realidade do seu contexto e a terceira remete às possíveis designações do verbete em questão, donde temos a decomposição e análise do conceito, criando campos semânticos e ligando variantes (BARROS, 2004). A ficha aqui trazida tem dados terminológicos, terminográficos e congrega elementos da ficha de trabalho e de síntese. No Quadro 1, trazemos a ficha que foi elaborada para o presente estudo e, em cada campo, uma breve descrição da sua função:

QUADRO 1 - FICHA TERMINOLÓGICA

Entrada: unidade de tratamento sobre a qual a ficha é referente. De acordo com Barros (2004), não deve ser marcada por inflexões verbais (caso seja um verbo), nem vir no feminino (salvo se ela comportar traços relevantes), nem ser pluralizada (salvo se há plurais lexicalizados ou variações semânticas).	
Nº de ocorrências: dado estatístico referente às ocorrências dos termos no <i>corpus</i> de estudo e de referência.	
Área: situa a unidade de tratamento dentro da própria Terminologia nas possíveis linhas de pesquisa, por exemplo, dentro da Socioterminologia, da Teoria Sociocognitiva, da Terminologia Textual e etc.	Inf. gramatical: relativa à composição fonética, morfológica, sintática e semântica da entrada.
Definição: descreve o conteúdo semântico-conceitual da unidade terminológica.	
Contexto: um exemplo de ambiente linguístico natural em que o termo se encontra a fim de elucidar o conceito ao falante quando ele confrontar a definição e o seu uso no texto.	
Sinonímia: possíveis termos concorrentes, tolerados, equivalentes, ou ainda relações de significância que a entrada possa apresentar dependendo do uso.	Antonímia: termos que se colocam em oposição semântica ao verbete.
Ver também: nesse campo, poderão ser incluídas sugestões de leitura para esclarecer mais sobre o termo, informações autorais, sites de pesquisa, dicionários especializados, possíveis relações semânticas que a mesma unidade possa ter em outras áreas e etc.	

Fonte: Elaborado pelos autores (2015).

4. METODOLOGIA DE COLETA E ANÁLISE DOS DADOS

A linguística de *corpus* (LC) investiga línguas naturais através de *corpora* de textos falados ou escritos, a fim de extrair e descrever fatos linguísticos intrínsecos, por meio de *softwares* especializados no tratamento de dados textuais. Novas perspectivas sobre a linguagem em uso são trazidas à tona na inspeção dos resultados obtidos da análise de *corpora*. Essa abordagem baseada em *corpus* possibilita a produção de obras dicionarísticas, o que é muito recorrente em dicionários de língua inglesa, mais especificamente. Sardinha, em sua obra *Linguística de Corpus*, defende que:

[a] influência mais visível no mundo contemporâneo está na preparação de dicionários. Hoje, todos os grandes dicionários da língua inglesa (de Oxford até Cambridge, Collins, Longman) são feitos com base em Linguística de *Corpus* (SARDINHA, 2004, p. XVII-XVIII).

Isso deixa claro como a análise de *corpora* fomenta a prática lexicográfica e terminográfica. No entanto, a LC está intimamente ligada às inúmeras áreas não somente dentro da Linguística, como também em qualquer área que queira analisar a língua em uso. A título de exemplo, empresas privadas financiam projetos para criarem *corpora* de textos de negócios a fim de descreverem aquela modalidade da língua, universidades desenvolvem testes de proficiência em gêneros acadêmicos, e assim por diante.

O nosso *corpus* de estudo, denominado TermCorp, é composto por 12 artigos acadêmicos sobre Terminologia, os quais foram disponibilizados nos repositórios LUME e Termisul, ambos na Universidade Federal do Rio Grande do Sul, e no CITRAT, na Universidade de São Paulo. Já para o *corpus* de referência, que é: “[...] geralmente uma coletânea maior de textos de uma abrangência de gêneros mais ampla e/ou de recursos” (BAKER; HARDIE; MCENERY, 2006, p. 137, tradução nossa)³, foi selecionado o projeto Lácio-Web (LW). Esse projeto foi iniciado em janeiro de 2002, com a parceria entre o Núcleo Interinstitucional de Linguística Computacional (NILC), localizado no Instituto de Matemática e Estatística (ICMC-

USP, IME) e a Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH). O público-alvo do LW é heterogêneo: de um lado linguistas, cientistas da computação, lexicógrafos, e de outro, não especialistas em geral. Formado por seis *corpora* – Lácio-Ref, Lácio-Dev, Par-C, Comp-C, Mac-Morpho e Lácio-Sint (porção do Lácio-Ref etiquetada automaticamente) –, selecionamos o material referente ao Lácio-Ref. Mais especificamente, o Lácio-Ref é composto por 6.241 textos, divididos em oito *subcorpora* de acordo com a temática (textos sobre ciências agrárias, biológicas, exatas, generalidades, humanas, religião e pensamento, saúde e sociais).

Utilizamos o programa *WordSmith Tools versão 6* (SCOTT, 2015), doravante WST, para análise dos *corpora*, o qual disponibiliza três funcionalidades, *WordList*, *Keywords* e *Concord*. A primeira gera listas de palavras de um *corpus* inserido nele, a segunda compara listas de palavras de um *corpus* de estudo com um *corpus* de referência, a fim de extrair as palavras-chave, que são aquelas cuja frequência mais destoa em relação as suas frequências em cada *corpus* separadamente, e a terceira funcionalidade produz as concordâncias ou listagens de ocorrências de um item específico em seu contexto. De ambos os *corpora*, devemos extrair a lista de frequência de palavras⁴.

Ao contrastar diferentes ocorrências, uma oriunda do *corpus* de estudo (nesse caso, artigos científicos sobre Terminologia) e a outra do *corpus* de referência (textos de língua geral extraídos do projeto LW), temos a possibilidade de inferir se a palavra em questão é uma candidata a termo genuíno ou não, ou seja, se cada um dos vários sentidos que as palavras ou frases apresentam de acordo com cada contexto passa a conter uma carga semântica particular em dado campo do saber. A técnica utilizada foi a de *keyword* e de *chavicidade* com unigramas (itens de palavra única).

Os textos originalmente coletados em PDF foram convertidos para TXT, pois o WST trabalha, inicialmente, com arquivos nessa extensão, para então editá-los em suas próprias extensões. Após esta etapa, retiramos qualquer elemento que não fosse linguístico e constituinte da cadeia textual, ou seja, foram retirados os títulos, subtítulos, cabeçalhos e rodapés, tabelas, figuras, paginação, informações sobre o autor e instituições, resumos e palavras em língua

estrangeira. Esta fase de limpeza retira os denominados “ruídos”, os quais podem interferir no processamento e nos resultados. Tendo os textos limpos em mãos, o WST realiza o processamento dos termos, a fim de extrair a lista de palavras mais frequentes do TermCorp e do Lácio-Ref.

Após as listas de palavras terem sido produzidas, as comparamos uma com a outra na funcionalidade *Keywords*, a qual nos deu as quinhentas palavras-chave em ordem decrescente de chavicidade (razão estatística que caracteriza as palavras-chave oriundas da comparação de uma lista de palavras de um *corpus* de estudo com uma lista de palavras de um *corpus* de referência, ou seja, o valor que revela aquelas palavras cujas frequências são mais dissonantes entre o TermCorp e o Lácio-Ref). As palavras cujas frequências são estatisticamente maiores no *corpus* de estudo do que no *corpus* de referência são chamadas de palavras-chave positivas, aquelas cujas frequências são estatisticamente menores no *corpus* de estudo do que no de referência são chamadas de palavras-chave negativas.

A partir disso, selecionamos as cem palavras-chave com o maior grau de chavicidade; depois, somente os substantivos não marcados. A decisão por expressões nominais é recorrente na produção de glossários e dicionários, contudo: “[...] é preciso compreender que não só de substantivos é feito o todo de uma terminologia.” (KRIEGER; FINATTO, 2004, p. 129). Foi gerada, então, uma nova lista com 48 palavras, da qual retiramos todas aquelas com frequência menor ou igual a 20 *tokens*, resultando em 37 palavras⁵, a partir das quais, retiramos os dez termos mais representativos da área. Sendo assim, dessa avaliação qualitativa, trazemos os dez termos abaixo, dispostos por grau decrescente de chavicidade:

TABELA 1 - LISTA FINAL DAS 10 PALAVRAS-CHAVE SELECIONADAS

Rank	Key Word	Freq.	%	RC. Freq.	RC. %	Keyness
4	Terminologia	127	0.37	173		959,28
13	Termo	118	0.35	1.010	0,01	522,96
18	Indexação	42	0.12	0		451,81

22	Glossário	36	0.11	0		387,26
27	Dicionário	32	0.09	0		344,22
33	Temática	31	0.09	4		308,63
39	Vocabulário	27	0.08	0		290,43
47	Léxico	28	0.08	9		260,22
65	Tesouro	21	0.06	2		212,32
88	Unidade ⁶	63	0.18	1.393	0,02	171,70

Fonte: Elaborada pelos autores (2015).

Na tabela acima, há três palavras-chave negativas: “unidade”, “termo” e “Terminologia”, as quais são mais frequentes no *corpus* de referência do que no TermCorp (frequência de 173, 1.010 e 1.393, respectivamente). É interessante analisar que as ocorrências de palavras-chaves negativas, nesse caso, podem ser relacionadas com o seu grau vocabular, ou seja, estão mais propensas a serem vocábulos da língua geral que termos de uma linguagem técnica-especializada. Entretanto, isso não as desclassifica como candidatas a termo, pois na coluna *RC.%*, a qual se refere à porcentagem representativa da palavra no *corpus* de referência, essas palavras-chave negativas representam valores menores do que os presentes na coluna *%*, a qual se refere à representatividade percentual da palavra no *corpus* de estudo. Isso significa que essas palavras são mais representativas no TermCorp, porque para os valores menores do que 0,01⁷ (células em branco), 0,01 e 0,02 no Lácio-Ref, sempre temos um valor maior no *corpus* de estudo (0.37/ 0.35/ 0.12/ 0.11/ 0.09/ 0.09/ 0.08/ 0.08/ 0.06/ 0.18).

Após essas etapas, seguiu-se com o preenchimento e revisão das fichas terminológicas. O *Concordanciador* nos possibilitou extrair todas as informações necessárias para preencher a ficha, conforme vemos a seguir nos dois exemplos trazidos:

QUADRO 2 - FICHA TERMINOLÓGICA

Entrada: UNIDADE TERMINOLÓGICA	
Nº de ocorrências: 63 hits / 0,18% do TermCorp 1.393 hits / 0,02% do Lácio-Ref	
Área: Terminologia	Inf. gramatical: sub. fem.; pl. (s).
<p>Definição: “A unidade terminológica é, portanto, unidade linguística, unidade cognitiva e unidade sociocultural, e seu estudo pode limitar-se a um aspecto, ou pode ser um estudo integrador das várias faces do termo (CABRÉ, 1999). Como as linguagens de especialidade fazem parte da língua geral, as unidades terminológicas adquirem seu valor especializado de acordo com o uso em determinada situação de comunicação. No modelo teórico da TCT, não há, a priori, uma divisão entre palavras e termos. Essa diferenciação só se dá no plano comunicacional, no qual os termos são ativados em uma situação de discurso especializado como forma de veicular conhecimento de uma determinada área de especialidade. É evidente que somente uma abordagem que leve em conta o caráter discursivo pode explicar os fenômenos relativos às unidades de significação especializada que ocorrem na linguagem em uso, como a polissemia, as metáforas e a variação de unidades terminológicas. Sendo assim, é essencial estudar as unidades terminológicas em uso no texto e no discurso.” Fonte: KILIAN; BOCORNY; WILKENS; VILLAVICENCIO, p. 279.</p>	
<p>Contexto: “A unidade terminológica, considerada em seu aspecto poliédrico, a saber, unidade linguística, cognitiva e sociocultural, deve ser descrita tanto em relação ao sistema da língua, como em relação ao âmbito de especialidade, pois se tratam de unidades lexicais da língua geral, cujas especificidades significativas são ativadas nas situações de comunicação especializada.” Fonte: SZABO, p. 56.</p>	
Sinonímia: termo; unidade lexical especializada.	Antonímia: unidade lexical; vocábulo; palavra.
Ver também: UTC (Unidade Terminológica Conceitual); unidade linguística, cognitiva e sociocultural (poliedricidade do termo).	

Fonte: Elaborado pelos autores (2015).

QUADRO 3 - FICHA TERMINOLÓGICA

Entrada: INDEXAÇÃO	
Nº de ocorrências: 42 hits / 0,12% do TermCorp - <i>Nenhuma ocorrência no Lácio-Ref</i> -	
Área: Terminologia.	Inf. gramatical: sub. fem.; pl. (ões); geralmente utilizada como sintagma adjetival: [...] <i>de indexação</i> .
<p>Definição: “Em indexação trabalhamos com os conceitos expressos nos documentos pelos autores. Apesar de indexarmos palavras, estas têm que serem entendidas como rótulos que podem ser expressos de várias formas. O indexador está trabalhando com uma rede conceitual de uma área específica do conhecimento. A compreensão do que é conceito e de suas relações irá facilitar o trabalho de indexação, que não mais será feito por palavras, mas sim por termos representativos daqueles conceitos.”</p> <p style="text-align: right; font-size: small;">Fonte: FERREIRA; LAAN, p. 2.</p> <p>“[...] unidades indexadoras podem ser definidas como signos linguísticos que representam a informação para viabilizar a comunicação entre o usuário e o conteúdo dos documentos.”</p> <p style="text-align: right; font-size: small;">Fonte: EZEQUIEL; LAAN, p. 2.</p>	
<p>Contexto: “Quanto ao processo de indexação, nesses casos de existência de variação concorrente, se repetem os problemas já mencionados. A percepção dessas variações pelo agente indexador, o uso de um vocabulário controlado e a estruturação de uma bem elaborada rede de remissivas pode minimizar o ruído e o silêncio no momento da recuperação da informação.”</p> <p>Fonte: EZEQUIEL; LAAN, p. 13.</p>	
Sinonímia: processos de indexação, representação da informação e representação da informação documentária.	Antonímia: não foi encontrado, em princípio, nenhum termo que se oponha conceitualmente ao processo de “indexação”.
Ver também: instrumentos de indexação; unidades de indexação.	

Fonte: Elaborado pelos autores (2015).

A partir das fichas acima, selecionamos aquelas informações que constituíram o verbete, tendo em vista o guia de referência sobre Terminologia para estudantes iniciantes de cursos de licenciatura em Letras proposto anteriormente. A seguir, respectivamente às fichas, encontram-se os verbetes:

QUADRO 4 - VERBETE

U.ni.da.de(s) terminológica(s),s.f. “A unidade terminológica é, portanto, unidade linguística, unidade cognitiva e unidade sociocultural, e seu estudo pode limitar-se a um aspecto, ou pode ser um estudo integrador das várias faces do termo (CABRÉ, 1999). Como as linguagens de especialidade fazem parte da língua geral, as unidades terminológicas adquirem seu valor especializado de acordo com o uso em determinada situação de comunicação.” **Ex:** “A unidade terminológica, considerada em seu aspecto poliédrico, a saber, unidade linguística, cognitiva e sociocultural, deve ser descrita tanto em relação ao sistema da língua, como em relação ao âmbito de especialidade [...]” **Sin.:** termo; unidade lexical especializada. **Ant.:** unidade lexical; vocábulo; palavra. **Ver:** UTC (Unidade Terminológica Conceitual); unidade linguística, cognitiva e sociocultural (poliedricidade do termo).

Fonte: Elaborado pelos autores (2015).

QUADRO 5 - VERBETE

In.de.xa.ção(ões), s.f. “Em indexação trabalhamos com os conceitos expressos nos documentos pelos autores. Apesar de indexarmos palavras, estas têm que serem entendidas como rótulos que podem ser expressos de várias formas.” **Ex:** “Quanto ao processo de indexação, nesses casos de existência de variação concorrente, se repetem os problemas já mencionados. A percepção dessas variações pelo agente indexador, o uso de um vocabulário controlado e a estruturação de uma bem elaborada rede de remissivas pode minimizar o ruído e o silêncio no momento da recuperação da informação.” **Sin.:** processos de indexação, representação da informação e representação da informação documentária. **Ver:** instrumentos de indexação; unidades de indexação.

Fonte: Elaborado pelos autores (2015).

5. CONSIDERAÇÕES FINAIS

A construção do TermCorp nos forneceu condições de realizar a etapa inicial de uma obra terminográfica, assim como, vislumbrar um guia de referência sobre Terminologia com foco em estudantes de graduação em Letras. Acreditamos que materiais de referência, como manuais, dicionários, glossários, léxicos etc., podem em muito contribuir para dirimir a complexidade lexical dos termos de uma dada área do saber que muitas vezes provocam estranhamento e dificuldades de entendimento em leitores/pesquisadores iniciantes. Barbosa (2017, p. 313) corrobora a proposta atual, quando atesta que: “[a]prender uma ciência básica, uma ciência aplicada, ou uma tecnologia corresponde a aprender (adquirir competência e desempenho) a linguagem de especialidade respectivamente constituída”.

Vimos que um *corpus* é um recurso valioso para extrairmos dados que fundamentam obras terminográficas. Naturalmente, devemos aprimorar o *corpus* quanto à sua extensão e relevância no contexto de graduação com o intuito de ampliarmos o que foi feito até o presente momento. E, da mesma forma, devemos elaborar mais fichas de outros termos oriundos do processamento que representem a área da Terminologia. Além disso, ainda prospectando a continuação do presente trabalho, faremos uso de fórmulas que avaliam a complexidade textual (Índice *Flesch*, por exemplo) e de métricas e índices de coesão e coerência de textos num amplo espectro de medidas (*Coh-Matrix*). Esta última baseia-se em pesquisas de Linguística Computacional e Psicolinguística, ligadas ao Projeto PorSimples – NILC, USP, UFScar, UNESP, 2009 –, em sua versão *on-line*, que oferece 60 métricas para uso livre.

Assim, vemos o fenômeno da acessibilidade de informação – técnica, científica, literária, entre outras – para diferentes níveis escolares e culturais em plena expansão. Inúmeros trabalhos, tais como o de Todirascu et al. (2013), Finatto (2011) e Pasqualini (2012) sustentam essa afirmação. Ao considerarmos os alunos de Letras como usuários com acesso a conteúdos científicos sobre Terminologia, cuja função e natureza são, fundamentalmente, metalinguísticas, devemos incluir, também, a terminologia da Linguística. Palavras de cunho especializado, tais como *lexia*, *termo*, *palavra*, *significante*, *significado*,

sintagma, lexema, língua, linguagem e assim por diante, precisarão ser abordadas em sala de aula à medida que o conhecimento dos alunos progredir. Portanto, promover essa acessibilidade é ampliar o saber-fazer do aluno, possibilitando a circulação e a difusão do conhecimento especializado de uma área do saber que vem galgando seu espaço.

THE PROPOSAL OF A TERMINOLOGY REFERENCE GUIDE FOR
UNDERGRADUATE STUDENTS OF LANGUAGES THROUGH *CORPUS*
PROCESSING ON *WORDSMITH TOOLS*

ABSTRACT

We report the initial step for the construction of a Terminology reference guide for undergraduate students of languages, focusing on the compilation of the study *corpus* and the terminological records, which precede the product *per se*. We used WordSmith Tools to extract the keywords in order to, later, become terminological entries. The keywords were selected according to its keyness value (statistical rationale which characterizes the keywords from the comparison of a study *corpus* and a reference *corpus* wordlists). For sampling purposes, we show two terminological records and their respective entries.

KEYWORDS: Terminology reference guide, terms, terminological records.

LA PROPUESTA DE UN GUÍA DE REFERENCIA SOBRE TERMINOLOGÍA PARA
ESTUDIANTES DE CURSOS DE LICENCIATURA EN LETRAS POR MEDIO DEL
PROCESAMIENTO DE *CORPUS* A TRAVÉS DEL *WORDSMITH TOOLS*

RESUMEN

Relatamos la fase inicial de la construcción de una guía de referencia sobre Terminología para estudiantes de los cursos de la Licenciatura de Letras. Nos centramos en la elaboración del *corpus* de estudio y de las fichas terminológicas, las cuales preceden a la obra. Utilizamos el *WordSmith Tools* para extraer las palabras clave, con el objetivo de, posteriormente, convertirlas en entradas terminológicas. Las palabras clave se seleccionaron de acuerdo con su valor de clavicidad ([keyness] razón estadística que caracteriza las palabras clave provenientes de la comparación de una lista de palabras de un *corpus* de estudio con la lista de palabras de un *corpus* de referencia). Para fines de muestreo, presentamos dos ejemplos de fichas y sus respectivas entradas.

NOTAS

1. Salientamos que nosso *corpus* de estudo é composto somente por textos de Terminologia. Entretanto, estamos cientes de que, para outras áreas de interesse, uma ampliação do *corpus* de estudo se faz pertinente a fim de abarcar outros fenômenos que não os terminológicos.
2. O texto dessa seção (1.1) foi parcialmente extraído, traduzido e adaptado da obra: CABRÉ, Maria Teresa. *La terminología: Teoría, metodología, aplicaciones*. Barcelona: Editora Antártida/Empúries, 1993.
3. Texto original em inglês : “[...] often a larger set of texts drawn from a wider range of genres and/or sources”.
4. Uma lista de todas as palavras que aparecem em um texto ou *corpus*, geralmente usadas na criação de um dicionário. Listas de palavras geralmente dão as frequências de cada palavra (ou *token*) no *corpus*. As palavras são na maioria das vezes ordenadas alfabeticamente, ou em valores de frequência, tanto com a frequência bruta ou com a porcentagem que a palavra tem no texto de forma geral. (BAKER; HARDIE; MCENERY, 2006, p. 169, tradução nossa). Texto original em inglês: A list of all of the words that appear in a text or **corpus**, often useful for dictionary creation. Word lists often give the frequencies of each word (or **token**) in the corpus. Words are most usually ordered alphabetically, or in terms of **frequency**, either with a raw frequency count and/or the percentage that the word contributes towards the whole text.
5. No primeiro momento (elaboração do guia de referência), ocupamo-nos daqueles termos cujo valor de chavidade os tornavam relevantes para o TermCorp, como uma espécie de filtro para a inclusão de termos e conceitos relacionados e relevantes à área. Salientamos que, à medida que houver uma progressão do trabalho, alguns outros termos que não tenham aparecido nos resultados atuais podem também ser representativos à área da Terminologia. Logo, haverá a necessidade de ampliarmos o volume de textos do TermCorp para justamente dar conta da inclusão de termos até então não arrolados como relevantes.

6. O WST apresenta somente “unidade” como palavra-chave para o *corpus* de estudo, visto que foi solicitado na extração somente unigramas, contudo, ao verificarmos as concordâncias para “unidade”, notamos que se trata na verdade de uma “unidade terminológica”, enquanto que no *corpus* de referência, trata-se de uma unidade que não a terminológica. Optamos, então, pela colocação “unidade terminológica”, para constar como termo, devido a sua carga semântica especializada e relevante para a área de estudo do trabalho.
7. O WST, por padrão, não exhibe valores de representatividade menores do que 0,01, razão pela qual as células que têm esses valores se encontram em branco na tabela.

REFERÊNCIAS

BAKER, Paul; HARDIE, Andrew; MCENERY, Tony. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press, 2006.

BARBOSA, Maria Aparecida. *Relações conceptus/ designationes vocabulares e terminológicas no percurso gerativo de enunciação de codificação*. Universidade de São Paulo. Disponível em: <<http://www.filologia.org.br/ixcnlf/17/21.htm>>. Acesso em: ago. 2014.

_____. *Transposições vocabulares e terminológicas em campos lexicais: ensino da metalinguagem técnico-científica*. Disponível em: <<http://www.filologia.org.br/vicnlf/anais/caderno07-15.html>>. Acesso em: jul. 2017.

BARROS, Lidia Almeida. *Curso básico de Terminologia*. São Paulo: Editora da Universidade de São Paulo, 2004.

CABRÉ, Maria Teresa. *La terminologia*. Representación y comunicación. Barcelona: IULA/UPF, 1999.

CIAPUSCIO, Guiomar. La terminología desde el punto de vista textual: selección, tratamiento y variación. *Organon*, Porto Alegre, v. 12, n. 26, p. 43-65, 1998.

DIKI-KIDIRI, Marcel. *Une approche culturelle de la terminologie, Terminologies Nouvelles*. Editorial “Le Rifal”, 2000. p. 27-30. [Trad. para o espanhol de Rodrigue Bigoundou, Un enfoque cultural de la terminología].

FINATTO, Maria José Borcony. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. *Organon*, v. 25, n. 50, 2011.

GAUDIN, François. *Pour un socioterminologie. Des problèmes semantiques aux pratiques institutionnelles*. Rouen: Publications de l'Université de Rouen, 1993. p. 231.

KRIEGER, Maria da Graça. Lexicologia, Lexicografia e Terminologia: Impactos Necessários. In: FINATTO, M. J. B. ISQUERDO, A. N. (Org.). *As ciências do léxico: lexicologia, lexicografia, terminologia*. Campo Grande: UFMS; Porto Alegre: UFRGS, 2010. 624 p. v. IV.

KRIEGER, Maria da Graça; FINATTO, Maria José Borcony. *Introdução à Terminologia: teoria e prática*. São Paulo: Contexto, 2004.

LAWSON, Tony. *A Conception of Ontology*. Cambridge: Faculty of Economics, 2004.

PASQUALINI, Bianca Franco. *Leitura, tradução e medidas de complexidade textual em contos de literatura para leitores com letramento básico*. 2012. 155 f. Dissertação (Mestrado em Teorias Linguísticas do Léxico) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

ROLLSING, Lucas Zambrano. *A banalização de conceitos oriundos da Terminologia vista como um recurso de aprendizagem*. 2014. 65 f. Monografia (Conclusão de curso Letras) – Faculdade de Letras, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2014.

SARDINHA, Tony Berber. *Linguística de corpus*. São Paulo: Manole, 2004.

SCOTT, Mike. *WordSmith Tools Version 6*. Liverpool: Lexical Analysis Software, 2015.

TEMMERMMAN, Rita. Sociocognitive terminology theory. In: *Simpósio Internacional de Verano de Terminología: Terminología y cognición*. Barcelona: IULA-UPF, 2001. p. 75-92.

TODIRASCU, Amalia et al. *Coherence and cohesion for the assessment of text readability*. Estrasburgo: Université de Strasbourg, 2013.

WÜSTER, Eugen. *Introduction to the General Theory of Terminology and Terminological Lexicography*. Vienna: Springer, 1979. (Edição Canadense de 1985). [Trad. do alemão para o inglês].

Submetido em 13 de julho de 2017.

Aceito em 26 de janeiro de 2018.

Publicado em 23 de abril de 2018.
