

ESTUDO-PILOTO SOBRE TERMINOLOGIAS DA CIÊNCIA DA COMPUTAÇÃO

Fabiana Hennies Brigidi

Submetido em 24 de maio de 2019.

Aceito para publicação em 21 de agosto de 2019.

Cadernos do IL, Porto Alegre, n.º 59, outubro. p. 89-110.

POLÍTICA DE DIREITO AUTORAL

Autores que publicam nesta revista concordam com os seguintes termos:

- (a) Os autores mantêm os direitos autorais e concedem à revista o direito de primeira publicação, com o trabalho simultaneamente licenciado sob a Creative Commons Attribution License, permitindo o compartilhamento do trabalho com reconhecimento da autoria do trabalho e publicação inicial nesta revista.
 - (b) Os autores têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada nesta revista (ex.: publicar em repositório institucional ou como capítulo de livro), com reconhecimento de autoria e publicação inicial nesta revista.
 - (c) Os autores têm permissão e são estimulados a publicar e distribuir seu trabalho online (ex.: em repositórios institucionais ou na sua página pessoal) a qualquer ponto antes ou durante o processo editorial, já que isso pode gerar alterações produtivas, bem como aumentar o impacto e a citação do trabalho publicado.
 - (d) Os autores estão conscientes de que a revista não se responsabiliza pela solicitação ou pelo pagamento de direitos autorais referentes às imagens incorporadas ao artigo. A obtenção de autorização para a publicação de imagens, de autoria do próprio autor do artigo ou de terceiros, é de responsabilidade do autor. Por esta razão, para todos os artigos que contenham imagens, o autor deve ter uma autorização do uso da imagem, sem qualquer ônus financeiro para os Cadernos do IL.
-

POLÍTICA DE ACESSO LIVRE

Esta revista oferece acesso livre imediato ao seu conteúdo, seguindo o princípio de que disponibilizar gratuitamente o conhecimento científico ao público proporciona sua democratização.

<http://seer.ufrgs.br/cadernosdoil/index>

Terça-feira, 29 de outubro de 2019.

ESTUDO-PILOTO SOBRE TERMINOLOGIAS DA CIÊNCIA DA COMPUTAÇÃO

PILOT STUDY ABOUT COMPUTER SCIENCE TERMINOLOGIES

Fabiana Hennies Brigidi*

RESUMO: Este estudo-piloto está ambientado nas áreas da Ciência da Computação, Biblioteconomia e Terminologia. Tem como objetivo a identificação das terminologias utilizadas por pesquisadores da Ciência da Computação por meio dos resumos das teses e dissertações defendidas na UFRGS de 2014 a 2018. Justifica-se este recorte em função da verificação dos candidatos a termos para, posteriormente, subsidiar a elaboração de um vocabulário controlado. Fundamenta-se, teoricamente, na Teoria Comunicativa da Terminologia e Teoria Sociocognitiva da Terminologia. Utiliza a Linguística de Corpus como metodologia. A partir dos resultados e discussões, destaca-se a representatividade de termos em língua inglesa e de siglas, concluindo-se que um corpus em maior escala poderá proporcionar resultados mais concretos.

PALAVRAS-CHAVE: Terminologia; indexação; ciência da computação; recuperação da informação.

ABSTRACT: This paper is an interdisciplinary pilot study in the fields of Computer Science, Librarianship, and Terminology. Its primary goal is to identify the terminologies used by researchers in Computer Science. I analyze the summaries of theses and dissertations defended between 2014 and 2018 at the UFRGS. Based on the established criteria, I used the terms that further support the development of a controlled vocabulary. My theoretical framework is informed by Communicative Theory of Terminology and Sociocognitive Theory of Terminology. Methodically speaking, is framed on the Corpus Linguistics. Based on the results and related discussions, I conclude that due to the representativeness of terms in English and acronyms, indeed a larger scale corpus would provide more concrete results.

KEYWORDS: Terminology, indexing, computer science, information retrieval.

1 Introdução

As terminologias estão presentes em todas as áreas do conhecimento tendo como principal função a condução do conhecimento especializado. Para Maria Teresa Cabré, as línguas de especialidade configuram-se instrumentos de comunicação entre os especialistas (CABRÉ, 1993). Segundo essa autora, as terminologias são o principal elemento capaz de diferenciar as linguagens de especialidade da língua comum, além de fazer distinções entre as próprias linguagens especializadas (CABRÉ, 1993).

Contudo, é necessário identificar quando um termo e/ou uma unidade terminológica (UT) assume essa posição ao invés de se comportar como uma palavra. Nas áreas de especialidade, essa identificação é recorrente. Sendo assim, os aspectos comunicativos do discurso especializado, essenciais na concepção da TCT, são determinantes na identificação das UTs, ou seja, uma palavra pode assumir a posição de

* Mestre em Gestão de Unidades de Informação pela Universidade do Estado de Santa Catarina (UDESC). Bibliotecária-Documentalista na Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: fabiana.brigidi@ufrgs.br.

termo no contexto, pois, segundo a TCT, “[...] *a priori* não há termos, nem palavras, mas somente unidades lexicais, tendo em vista que estas adquirem estatuto terminológico no âmbito das comunicações especializadas.” (KRIEGER; FINATTO, 2004, p. 35).

A Biblioteconomia é uma das áreas responsáveis pelo tratamento temático de recursos informacionais¹. É fundamental que os bibliotecários, assim como os terminólogos, tenham conhecimento dos demais universos terminológicos para subsidiar uma indexação de qualidade que satisfaça as necessidades de seus usuários por meio da recuperação da informação (RI).

A indexação de assuntos implica na preparação de uma representação do conteúdo temático de documentos, sintetiza Lancaster (2004). O indexador descreve o conteúdo do documento “[...] ao empregar um ou vários termos de indexação, comumente selecionados de algum tipo de vocabulário controlado.” (LANCASTER, 2004, p. 6). Essa atividade costuma ser realizada por profissionais da informação fortemente influenciados por suas experiências profissionais, conhecimento da área a ser tratada, consulta a ferramentas de trabalho como os vocabulários controlados² e a internet em consonância com a própria intuição. A subjetividade é outra característica da indexação, em função do ponto de vista do indexador, assim como do acervo a que o recurso pertence, isto é, bibliotecas especializadas costumam priorizar termos relacionados à área de conhecimento em que estão inseridas.

Tendo em vista o contexto descrito, o estudo aqui proposto diz respeito ao tratamento temático referente às terminologias da área da Ciência da Computação, desempenhado pelo bibliotecário no ambiente universitário, configurando-se um recorte de pesquisa.

Este trabalho tem a finalidade de identificar parte da Terminologia da área da Ciência da Computação presente nos resumos em língua portuguesa das teses e dissertações produzidas pelo Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande do Sul (PPGC/UFRGS) no período de 2014 a 2018. Optou-se pelos parâmetros citados a fim de se realizar um diagnóstico inicial da terminologia utilizada por pesquisadores da área da Ciência da Computação da UFRGS, configurando-se, assim, a justificativa deste recorte. Destaca-se ainda que, por se tratar de um estudo-piloto, alguns parâmetros poderão ser alterados futuramente, como, por exemplo, expandir a pesquisa para a análise de termos na língua inglesa, visto que a área em questão possui muitos termos nesse idioma.

Em outras palavras, o presente recorte tem o objetivo de identificar, inicialmente, as UTs acerca da área da Ciência da Computação a partir da produção intelectual (PI) representada pelos resumos das teses e dissertações provenientes do PPGC/UFRGS para, posteriormente, criar um protótipo de vocabulário controlado da área para fins de utilização pelo Sistema de Automação de Bibliotecas da UFRGS

¹ Nesta pesquisa, os recursos informacionais englobam qualquer tipologia documental com conteúdo informacional em qualquer suporte disponibilizados por Unidades de Informação.

² Na área da Biblioteconomia, os vocabulários controlados são formados por linguagens documentárias que, por sua vez, “[...] são sistemas artificiais de signos normalizados que permitem representação mais fácil e efetiva do conteúdo documental, com o objetivo de recuperar manual ou automaticamente a informação que o usuário solicita. Entende-se que as linguagens documentárias é que farão a comunicação entre a linguagem natural dos usuários e a unidade de informação, elas são utilizadas para representar o conteúdo dos documentos, por isso alguns autores as definem como sistemas simbólicos instituídos, que visam a facilitar a comunicação.” (TRISTÃO; FACHIN; ALARCON, 2004, p. 162).

(Sabi)³. Optou-se por esse domínio em função dos avanços tecnológicos que são diretamente refletidos nas terminologias a partir do surgimento constante de novas UTs.

No contexto apresentado, as bibliotecas universitárias desempenham um papel fundamental para a recuperação dos recursos informacionais disponibilizados em seus acervos representado pelo tratamento temático, mais especificamente, o processo de indexação. Em síntese, a indexação consiste na atribuição de assuntos por meio de termos pertinentes ao recurso indexado, conforme comentado anteriormente. Contudo, na Biblioteconomia, faz-se uso de controle vocabular para garantir consistência ao catálogo. Esse controle é realizado mediante ferramentas como as listas de cabeçalhos de assuntos e os tesauros, por exemplo. Sendo assim, observa-se a fundamental inter-relação da representação temática com os recursos informacionais ofertados pelas bibliotecas que trabalham para uma eficiente RI.

Com relação à fundamentação teórica, este estudo se baseia nas concepções de Maria Teresa Cabré acerca da TCT que, de modo geral, prioriza o contexto especializado em que as UTs são identificadas, juntamente com a Teoria Sociocognitiva da Terminologia (TST), postulada por Rita Temmerman, que integra aspectos sociais, histórico-culturais e cognitivos para descrever e explicar processos de cognição. Essa última teoria tem o objetivo de contribuir mais especificamente com a identificação dos processos cognitivos envolvidos nas PIs aliado aos aspectos diacrônicos das UTs deste estudo-piloto. Além dessas teorias, a Linguística de *Corpus* (LC), também, faz-se presente devido ao auxílio computacional na análise do *corpus*-amostra, compondo a fundamentação metodológica deste estudo.

Sendo assim, este artigo está estruturado da seguinte forma: discorre-se inicialmente sobre termo e Terminologia, indexação, RI e teorias da Terminologia supracitadas. Os procedimentos metodológicos são explicados na sequência, destacando-se a compilação e análise do *corpus*-amostra fundamentados na LC, seguidos dos resultados da pesquisa, discussão e considerações finais.

2 Elementos teóricos

Com a finalidade de fundamentar o presente estudo, os elementos teóricos aqui apresentados abrangem aspectos referentes à Terminologia propriamente dita, com enfoque para a comunicação especializada; a indexação, considerada um dos processos referentes à representação temática de recursos informacionais; os vocabulários controlados, responsáveis pelo controle terminológico de catálogos de unidades de informação (UIs); a RI, fundamental para atender às necessidades informacionais de usuários de UIs e as teorias da Terminologia, TCT e TST, propostas por Maria Teresa Cabré e Rita Temmerman, respectivamente.

2.1 A Terminologia e o termo

A Terminologia tem como objeto central o termo e sua complexidade, pois, conforme Zilio (2011), defini-lo é uma tarefa árdua. Essa constatação pode representar um problema ao terminólogo no momento da seleção de termos presentes em um

³ O catálogo *on-line* das bibliotecas da UFRGS é comumente chamado de Sabi. Disponível em: <http://www.sabi.ufrgs.br>. Acesso em: 22 jan. 2019.

determinado texto. Embora Barros (2004, p. 40) simplifique a definição de termo ao afirmar que se trata de “[...] uma unidade lexical com um conteúdo específico dentro de um domínio específico.” é comum que o terminólogo se pergunte: “Como distinguir o que é termo daquilo que não é?” (ZILIO, 2011, p. 119). Esse problema também é vivenciado por outros profissionais da informação, como os bibliotecários, por exemplo, durante a análise temática, para posterior indexação de recursos informacionais.

Para explicar essa dificuldade de identificação terminológica, Zilio (2011, p. 119) salienta que

As teorias de Terminologia existentes tentam dar um esclarecimento sobre o que vem a ser um termo, mas, frequentemente, seus apontamentos levam para distinções altamente subjetivas por parte do profissional, que acaba arbitrando a questão baseado principalmente em seus instintos.

Nesse sentido, grande parte dos profissionais da informação alia sua intuição à experiência profissional durante a seleção de termos de um texto, nem sempre relacionando-a, conscientemente, a qualquer teoria existente.

Com exceção da teoria denominada Linguística do Texto Especializado, que tem como foco o próprio texto da área de especialidade e a linguagem empregada, sem preocupar-se em diferenciar os termos das palavras, as demais teorias estão voltadas para o termo em si (ZILIO, 2011). Nesse contexto, ao abordarem o surgimento das teorias e escolas de Terminologia, Krieger e Finatto (2004, p. 30) salientam que “[...] alguns estudiosos passaram a desenvolver reflexões sobre os termos [...]”, pois havia certa preocupação quanto à pragmática terminológica.

Com o objetivo de contextualizar a discussão sobre o que é ou não termo, destaca-se a divisão diacrônica acerca dos estudos da Terminologia. A primeira fase, com um enfoque mais cognitivo e prescritivo no qual o conceito vem antes do termo; e a segunda fase, com enfoque mais linguístico e descritivo, no qual o termo pode ser considerado uma unidade com significante e significado. Complementando essa ideia, Krieger e Finatto (2004, p. 30) explicam que na primeira fase, “[...] prevalece uma perspectiva normativa sobre as terminologias em contraponto às linhas de fundamento descritivo sobre o léxico especializado, que ganham impulso com o desenvolvimento da Linguística.”

Como uma primeira referência às reflexões sobre o termo propriamente dito, destaca-se a Teoria Geral da Terminologia (TGT) proposta pelo engenheiro austríaco Eugen Wüster em meados da década de 1930, considerada um marco na história da área (KRIEGER; FINATTO, 2004). Segundo a TGT, o termo é caracterizado pela univocidade, pois se refere a somente um conceito e vice-versa. Zilio (2011) destaca a divisão explícita entre a língua comum e as línguas de especialidade vislumbrada por essa teoria. “Enquanto as palavras pertencem à língua comum, o termo pertence à língua de especialidade e deve ser controlado de forma a não existir uma mesma denominação para dois conceitos.” (ZILIO, 2011, p. 120).

A inflexibilidade caracterizada pela TGT, entre outras características, proporcionou o surgimento de novas teorias com diferentes posicionamentos teóricos. Surgiram correntes que passaram a entender o termo como parte de um sistema linguístico ao invés “[...] de um sistema de denominações vinculado a um sistema de conceitos [...]” (ZILIO, 2011, p. 121) como preconiza a TGT. Esse é o caso da TCT que entende o termo como um signo linguístico, assim como as palavras, porém caracterizado pela poliedricidade, isto é, formado por uma unidade linguística,

comunicativa e cognitiva. A diferenciação entre termo e palavra ocorre no contexto das linguagens especializadas, por meio da situação comunicativa em que são utilizados, considerando-se a intenção dos usuários envolvidos, a temática veiculada e o tipo de discurso em que se inserem (CABRÉ, 1993). Desse modo, os contextos linguístico e pragmático são determinantes na identificação de um termo do ponto de vista da TCT que, por ser uma teoria diretamente relacionada a esta pesquisa, é comentada novamente na seção referente às teorias da Terminologia.

Outras definições sobre o termo propriamente dito são formuladas por diversos autores, como, por exemplo, Barros (2004, p. 40), ao afirmar que se trata de “[...] uma unidade lexical com um conteúdo específico dentro de um domínio específico.” Nessa perspectiva, o termo somente irá existir nas linguagens de especialidades, pois as palavras pertencem ao léxico comum. Sob o ponto de vista de Laipelt (2015, p. 58), “[...] um termo representa conceitos de uma área de especialização, transmite conhecimento especializado e possui caráter linguístico em função de sua inserção no discurso.” Independentemente da definição, muitos autores da área da Terminologia concordam que o status de termo somente é concebido em meio a um contexto de especialidade.

Embora a Terminologia forneça diferentes concepções acerca dos termos presentes nos textos especializados, para a Biblioteconomia a identificação dos mesmos é indispensável para as atividades voltadas à representação temática de um recurso informacional, independentemente de sua definição. Essa atividade está diretamente relacionada à indexação comentada na próxima seção.

2.2 *Indexação*

A indexação está inserida no contexto do Tratamento Temático da Informação (TTI), mais especificamente na representação temática, pois se configura numa técnica de análise de conteúdo que tem por finalidade a condensação de informações significativas por meio da atribuição de termos referentes aos recursos indexados “[...] criando uma linguagem intermediária entre o usuário e o documento.” (VIEIRA, 1988, p. 43).

O diagrama do TTI apresentado na Figura 1 ilustra o contexto em que a indexação está incluída, isto é, resume-se numa atividade de representação temática que se relaciona diretamente à análise temática com subordinação ao tratamento temático no universo da Ciência da Informação em que a Biblioteconomia se insere.

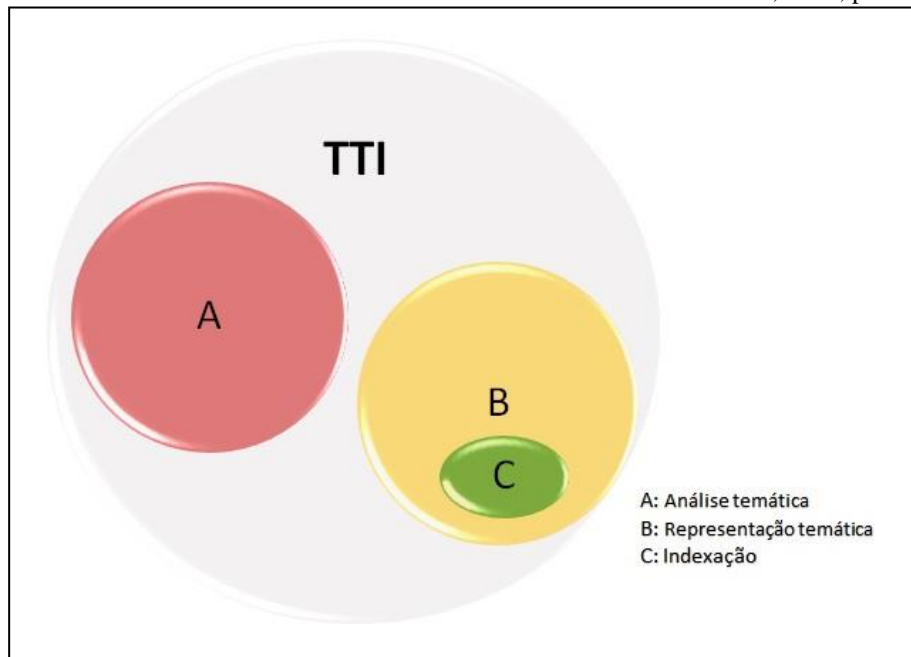


Figura 1 – Diagrama do TTI

De modo geral, o TTI se configura um processo complexo voltado para o acesso do recurso informacional que, “Por sua vez, centra-se basicamente em analisar, descrever e representar o conteúdo informacional dos documentos, com fins de armazenamento e recuperação da informação em sistemas de informação.” (DAL’EVEDOVE, 2010, p. 15). Sendo assim, a indexação pode ser considerada uma atividade essencial e indispensável em UIs com a finalidade de proporcionar meios de recuperar as informações presentes nos acervos, satisfazendo seus usuários, aproximando-se inclusive da função de apoio à pesquisa, “[...] cujo objetivo consiste em adquirir, registrar, controlar, elaborar e transmitir informação relativamente às demandas dos usuários [...]” (CAFFO, 1988, p. 11 *apud* GUIMARÃES, 2009, p. 107).

Centrando-se na indexação propriamente dita, Cintra et al. (2001) salientam que, para a caracterização do assunto de um texto, utiliza-se um código denominado Linguagem Documentária (LD). “A informação é, neste caso, expressa através dos elementos de um código exterior ao texto submetido à análise, supondo, portanto, um procedimento de tradução.” (CINTRA et al., 2001, p. 18). Desse modo, o processo de análise temática se inicia na identificação dos assuntos pertinentes ao recurso informacional para, então, representá-los por meio de LDs pertinentes, subentendendo-se uma etapa intermediária denominada tradução.

As LDs caracterizam-se por um conjunto de termos que tem por objetivo definir as formas de entrada e pesquisa utilizadas por indexadores e usuários num sistema documentário. Assim sendo, as LDs integram “[...] elementos resultantes de escolhas feitas em um universo lexical amplo. Suas unidades têm origem tanto em Linguagens de Especialidade, como na linguagem de uso corrente e nas Terminologias de área.” (CINTRA et al., 2001, p. 18). Observa-se, portanto, a estreita relação da indexação e do trabalho documentário no todo, com a Terminologia que, para Cintra et al. (2001), mantém um diálogo constante.

O termo (unidade terminológica), ao associar denominação e noção, constitui uma unidade referencial. Na Terminologia, uma palavra designa um determinado objeto porque opera com propriedades e características, remetendo a determinados universos de valores consubstanciados nos discursos de especialidade. (CINTRA et al., 2001, p. 20).

Portanto, o contexto comunicativo de especialidade proporciona a identificação de conceitos e UTs, refletindo diretamente nas teorias de Terminologia identificadas para este estudo, pois a TCT privilegia a função comunicativa textual, entre outras características, enquanto a TST se volta para os aspectos cognitivos das unidades de compreensão no discurso comunicativo, bem como nos aspectos diacrônicos observados ao longo de publicações sobre um mesmo tema em diferentes períodos, embora com menor ênfase.

No âmbito da Biblioteconomia, a representação desses assuntos, por meio de LDs, ocorre com instrumentos específicos para subsidiar a etapa da tradução no processo de indexação. Esses instrumentos denominam-se vocabulários controlados, pois têm a função de manter a consistência dos catálogos de assuntos de UIs. Destaca-se que o controle vocabular utilizado por profissionais da Biblioteconomia também se estende a autoridades como nomes pessoais, entidades e eventos, por exemplo. Contudo, esta pesquisa está diretamente relacionada à temática de assuntos, portanto, não faz referência aos demais tipos de controle vocabular.

Vocabulários controlados são comumente representados pelos tesouros, pelas listas de cabeçalhos de assunto e pelas taxonomias. Silva, Souza e Almeida (2008) destacam que, assim como na criação de ontologias, a elaboração de vocabulários controlados demanda uma organização conceitual por processos que inclui a categorização, a classificação e as relações dos conceitos identificados, bem como o tratamento da terminologia empregada nos conceitos e nas relações da estrutura. Essa perspectiva supõe uma ligação direta entre conceito e terminologias que, neste estudo, estão presentes no discurso comunicativo das publicações de uma área de especialidade no decorrer de um determinado período, logo, relaciona-se diretamente à TCT e à TST.

A elaboração das ferramentas de controle vocabular não é o foco deste recorte de pesquisa e, por isso, não foi analisada em profundidade. Destacam-se, apenas, as questões conceituais e terminológicas envolvidas, bem como seu principal objetivo: controlar as Uts, de modo a permitir associações sinonímicas e homonímicas, por exemplo, garantindo consistência e eficiência dos sistemas de recuperação de informações, neste caso, os catálogos de bibliotecas.

Mesmo com todo o ferramental apresentado para auxiliar a atividade de indexação, isto é, a utilização do controle vocabular por meio das diferentes LDs, observa-se uma subjetividade bastante representativa por parte dos profissionais da informação. O ponto de vista, a experiência e até mesmo a intuição são características apresentadas pelos indexadores durante o processo que engloba todo o TTI. Fonseca (2017) alerta para o entendimento individualizado de um mesmo texto por diferentes indexadores perpassando sua familiaridade e nível de especialização na área. A autora complementa ao afirmar que, para algumas pessoas, somente a leitura das palavras-chave pode ser suficiente, enquanto que, para outras, é necessário realizar uma leitura mais apurada do sumário, resumo e outras partes do texto, conforme cada tipo documental (FONSECA, 2017).

Sobre esse assunto, Lancaster (2004, p. 24) alerta que “Usualmente, recomenda-se um misto de ler e ‘passar os olhos’ pelo texto.” Nesse sentido, a leitura técnica do

recurso informacional indexado deve ser realizada mediante a verificação de itens fundamentais, como título, resumo e considerações finais, aliada a outras partes que se fizerem necessárias. O conhecimento prévio do profissional sobre a temática indexada irá definir o nível de aprofundamento necessário nessa etapa. Aconselha-se que o indexador se habitue a considerar o documento completo, incluindo partes lidas em profundidade e partes lidas de relance para que os termos atribuídos reflitam o todo e não somente parte do conteúdo temático (LANCASTER, 2004).

Após a identificação dos principais assuntos, recorre-se à etapa da tradução, fundamental por transformar a subjetividade da indexação em uma linguagem universal representada pelas LDs. “O objetivo desse procedimento [tradução] é permitir a intermediação entre usuário e documento no momento da busca e recuperação da informação em um sistema de informação.” (PIOVEZAN; FUJITA, 2015, p. 112).

Além disso, o indexador também deve estar atento aos fatores de exaustividade e especificidade (LANCASTER, 2004), conforme a política de indexação da UI em que atua. A exaustividade se refere à quantidade, enquanto a especificidade ao conteúdo propriamente dito dos termos escolhidos para representar o recurso. Segundo Piovezan e Fujita (2015), a indexação é uma das principais atividades do profissional da informação, pois o sucesso da missão das UIs depende de seu bom desempenho, entre outras atividades. No caso da indexação propriamente dita, sua repercussão se dá efetivamente com a recuperação da informação, tema comentado a seguir.

2.3 Recuperação da informação

Em sentido literal, é possível entender a RI como o resultado satisfatório de um processo de busca informacional que geralmente ocorre em sistemas informatizados, como os catálogos *on-line*. Seu sucesso depende de estratégias eficientes de pesquisa, bem como do trabalho do indexador que é responsável pela atribuição de termos pertinentes ao recurso informacional indexado.

Embora atualmente se relacione a RI aos ambientes digitais, sua origem se deu em meados da década de 1940, em função do expressivo volume de publicações ocorridas após a Segunda Guerra Mundial, fenômeno conhecido com explosão da informação⁴. Essa terminologia é apontada por Vannevar Bush, respeitado cientista do MIT (*Massachusetts Institute of Technology*), na publicação de 1945 intitulada *As we may think* (Como podemos pensar) abordando o referido fenômeno como um problema a ser solucionado (RUSSO, 2010). No período pós-guerra já se apontava a RI como solução para esse problema informacional.

Originalmente denominada *Information Retrieval* por Calvin Mooers em 1951, a RI abrange os aspectos intelectuais de descrição de informações aliados às especificidades de pesquisa, além de quaisquer sistemas, técnicas ou máquinas utilizados para o desempenho da operação (MOOERS, 1951). Desse modo, observa-se a relevância do trabalho do indexador, bem como do responsável pela busca, pois ambos estão diretamente relacionados às terminologias utilizadas durante a inclusão e busca da informação, respectivamente.

Nessa linha de pensamento, Corrêa (2008, p. 43) explica que

⁴ Russo (2010) aponta dois sinônimos para essa expressão: explosão informacional e explosão bibliográfica.

A RI consiste basicamente em uma interação usuário, profissional da informação e o universo de documentos (entendendo-se aqui documento como a informação registrada e armazenada em qualquer suporte tanto de texto, imagem ou som). A partir de uma necessidade de informação verificada e comunicada, procede-se a busca, aquisição e consulta de documentos relevantes para a solução do problema levantado. Este processo inclui uma negociação intelectual e cognitiva que pode necessitar de ajustes/revisões na comunicação com a finalidade de buscar a correspondência mais clara possível entre o problema de informação e o documento a ser pesquisado.

Sendo assim, além das figuras humanas (usuário e indexador) envolvidas na RI, a informação propriamente dita compõe a tríade que necessita estar em sintonia para que se obtenha um resultado satisfatório. Inserem-se nesse universo duas teorias da Terminologia diretamente relacionadas a esta pesquisa em função de suas características: TCT e TST.

2.4 Teorias da Terminologia

Entende-se que as teorias da Terminologia escolhidas para o embasamento teórico desta pesquisa, TCT e TST, são capazes de proporcionar um melhor entendimento do universo comunicativo da produção científica aqui estudada, visto que ambas, interligadas, relacionam-se diretamente ao tema proposto.

A valorização dos aspectos comunicativos das linguagens de especialidade e o entendimento de que as UTs formam parte de uma linguagem natural e da gramática da língua são características fundamentais da TCT (KRIEGER; FINATTO, 2004). Por pertencerem ao universo da Ciência da Computação, os termos identificados na produção intelectual deste estudo relacionam-se diretamente com os preceitos da TCT, pois pertencem a uma área específica e têm na comunicação entre os pares a exposição de suas terminologias. Nesse sentido, o contexto comunicativo irá definir o status de termo de uma unidade lexical.

Desse modo, a discussão inicial apresentada neste artigo acerca do que é ou não termo, depende das variáveis aqui citadas. Na concepção de Costa (2014), quando na posição de termo, configuram-se unidades léxicas poliédricas que podem assumir características linguísticas, cognitivas ou sociais (comunicativas), simultaneamente em função das áreas e situação comunicativa em que são utilizadas. “Consequentemente, o conteúdo de um termo não é fixo, mas relativo, variando conforme o cenário comunicativo em que se inscreve.” (KRIEGER; FINATTO, 2004, p. 35). Assim, o termo se apresenta como “[...] um elemento linguístico e, ao mesmo tempo, como elemento da produção do saber, uma vez que [representa e] transmite conhecimento especializado de uma determinada área.” (COSTA, 2014, p. 79).

A TCT deve incluir a competência e a atualização dos falantes, contemplando sua heterogeneidade cognitiva e comunicativa; considerando os fenômenos da linguagem natural, descrevendo especificidades cognitivas, linguísticas (gramaticais, pragmáticas, textuais e discursivas) e comunicativas das UTs; e explicando como o falante especialista adquire e utiliza essas especificidades (CABRÉ, 1993).

Do ponto de vista da TST, também identificada nesta pesquisa, “[...] os termos são unidades de compreensão e de representação, funcionando como modelos cognitivos e culturais.” (KRIEGER; FINATTO, 2004). Por estarem em constante

evolução, essas unidades são mais representativas que os conceitos para a TST que também defende uma “[...] perspectiva textualista, pois vê o termo como um constructo que se instaura no contexto da comunicação especializada.” (COSTA, 2014, p. 87-88). Para essa teoria, em princípio, o termo não existe, porém se constitui em um processo de conceitualização e categorização sociocultural que “[...] não acontece fora da língua e é mediado por modelos cognitivos, muitas vezes metafóricos, que facilitam a compreensão da realidade.” (COSTA, 2014, p. 88). Sendo assim, a TST integra aspectos sociais, histórico-culturais e de cognição para descrever e explicar processos cognitivos, inclusive por meio de metáforas.

A seleção das unidades de compreensão é baseada no público-alvo, em *corpus* de discursos reais e na proposta de organizar as terminologias das áreas de especialidades fundamentadas em estruturas prototípicas e modelos cognitivos (COSTA, 2014). Observa-se uma visão linguística do termo e a valorização dos aspectos cognitivos das linguagens especializadas, propostas pela TST (COSTA, 2014).

Outro aspecto a ser considerado se refere às variações terminológicas de polissemia e sinonímia, decorrentes da constante evolução das UTs observados pela precursora da TST, Rita Temmerman (KRIEGER; FINATTO, 2004). Essas variações são essências para o entendimento das unidades de compreensão diretamente relacionadas aos aspectos diacrônicos das linguagens especializadas. “Os períodos históricos no decorrer de sua evolução podem ser mais ou menos essenciais para a compreensão de uma unidade.” (TEMMERMAN, 2004, p. 35). Modelos cognitivos como os metafóricos, por exemplo, “[...] têm seu papel no desenvolvimento de novas idéias, o que significa que os termos são motivados.” (TEMMERMAN, 2004, p. 35).

A questão temporal reflete diretamente na área da Ciência da Computação, analisada neste estudo. O referido domínio está em constante evolução terminológica em função do acelerado avanço tecnológico que Aranalde (2005, p. 340) descreve como uma realidade “[...] onde o novo é instável e está em iminência de ser superado pelo novíssimo, sempre em contínua elaboração [...]”. Embora a citação desse autor tenha mais de uma década, ela permanece atual num mundo em que a tecnologia está em constante desenvolvimento e com ela, as terminologias.

Observa-se uma intersecção entre as teorias aqui citadas, representada pela perspectiva linguística referente ao termo/unidade de compreensão, pelos aspectos cognitivos e pelo contexto comunicativo que ambas consideram, essencialmente.

3 Procedimentos metodológicos

Os procedimentos metodológicos adotados neste estudo reúnem as concepções teóricas da TCT e da TST aliadas aos preceitos da LC, esta última em função da utilização de *corpora* para análise dos dados. Segundo Sanchez e Cantos (1996, p. 8-9 *apud* Berber Sardinha, 2004, p. 18), *corpus* se refere a

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

De acordo com essa definição, na LC, para que a reunião de determinado material seja considerado um *corpus*, critérios de origem, propósito, composição, formatação, representatividade e extensão precisam ser atendidos (BERBER SARDINHA, 2004). A origem se refere à autenticidade dos dados; o propósito requer que os dados sejam objeto de estudo linguístico; a composição pressupõe que os dados foram escolhidos criteriosamente; a formatação sugere que os dados sejam legíveis por computador; a representatividade indica destaque de uma língua ou de uma variedade linguística e a extensão se refere a um vasto material (BERBER SARDINHA, 2004).

Entende-se que esta pesquisa atende aos critérios citados e está de acordo com a concepção de Beilke (2018, p. 378) quando este afirma que “[...] o *corpus* deve atender minimamente a necessidade para o qual foi criado [...]”. Ou seja, embora se trate de um *corpus*-amostra⁵ e por essa razão seja formado pela parte de um todo maior, este estudo se sustenta para o propósito a que se dispõe que é obter um diagnóstico inicial das terminologias da área Ciência da Computação no ambiente universitário.

A LC “[...] ocupa-se da coleta e da exploração de *corpora*, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística.” (BERBER SARDINHA, 2004, p. 3). A exploração ocorre mediante programas de computador criados para essa finalidade relacionando-se diretamente ao critério de formatação mencionado anteriormente. Marian (2015, p. 477) afirma que a LC “[...] percebe a linguagem sob perspectiva probabilística, ou seja, as ocorrências não surgem de forma aleatória, sendo possível evidenciar e quantificar padrões por meio de ferramentas estatísticas.” Neste estudo foi possível coletar os candidatos a termos segundo resultados estatísticos mencionados por Marian (2015), além de outros critérios pré-estabelecidos (Quadro 1 e Figura 2).

Em função das possibilidades de programas de computador que a LC dispõe, optou-se pelo uso do software livre *AntConc*, criado por Laurence Anthony, que mantém o programa atualizado mediante inclusão de “[...] novas e variadas ferramentas, só ou em colaboração com outros pesquisadores, que enseja análises complexas e multifacetadas.” (TAGNIN, 2018, p. 14). O referido software foi indispensável durante as etapas referentes à compilação, análise e processamento do *corpus*-amostra.

Em síntese, os procedimentos metodológicos realizados neste estudo englobam as seguintes etapas: a) definição da tipologia documental a ser analisada; b) pesquisa no catálogo Sabi segundo critérios pré-estabelecidos; c) compilação dos registros selecionados para formação e análise do *corpus*; d) processamento e exploração do *corpus*; e) levantamento dos candidatos a termo. Essas etapas são descritas em maior profundidade nas seções seguintes.

3.1 Dados do *corpus*-amostra

Considerando-se o recorte da pesquisa em desenvolvimento, o *corpus*-amostra analisado foi composto pelos resumos⁶ das teses e dissertações provenientes do

⁵ Tipologia baseada no critério de seleção elaborado por Berber Sardinha (2004, p. 20) acerca da seleção textual “De amostragem: composto por porções de textos ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo.”

⁶ Por ser considerada uma amostra inicial, optou-se pela não utilização das palavras-chave neste primeiro momento.

PPGC/UFRGS dos últimos cinco anos (jan./2014 a dez./2018) por se tratar de uma amostra inicial. Além disso, os demais critérios pré-definidos se limitam a resultados em língua portuguesa e disponíveis *on-line*.

A pesquisa propriamente dita foi realizada em 22 de janeiro de 2019 no catálogo Sabi. A busca foi efetivada por meio da “Pesquisa CCL” que permite a utilização de uma linguagem de comandos específicos. Optou-se por esse tipo de pesquisa com a finalidade de incluir os critérios previamente selecionados, especificados no quadro a seguir:

Quadro 1 – Critérios de pesquisa e códigos de comando

CRITÉRIOS	CÓDIGO DO COMANDO ⁷
Somente documentos dos Programas de Pós-Graduação referentes à Produção Intelectual da UFRGS	WPG
Cursos da Ciência da Computação ⁸ e Computação	CMPP
Código do idioma	WLN
Somente resultados em português	POR
Conteúdo eletrônico: somente com texto completo ⁹	WLI
Formato	WTF
Somente teses e dissertações	TD
Tipo de documento	WTD
Tese, dissertação e trabalho de conclusão de mestrado profissional	M

Fonte: Elaborado pela autora, 2019.

A partir dos critérios e códigos de comando descritos no Quadro 1, a estratégia de pesquisa conta com o operador booleano *and*, conforme ilustrado no Quadro 2:

Quadro 2 – Estratégia de pesquisa

WPG=CMPP AND WLN=POR AND WLI=TEXTO COMPLETO AND WTF=TD AND WTD=M

Fonte: Elaborado pela autora, 2019.

Além da estratégia de pesquisa apresentada no Quadro 2, incluiu-se o recorte temporal citado (2014-2018), de acordo com a representação a seguir (Figura 2).

⁷ Disponível em: <http://www.ufrgs.br/documenta/manuais-sabi/registro-bibliografico/anexos/anexo-campos-subcampos-de-recuperacao-da-informacao>. Acesso em: 22 jan. 2019.

⁸ Até 1998 o programa de pós-graduação se chamava Ciência da Computação, sendo alterado para Computação após essa data.

⁹ Esse critério permite a recuperação de registros disponíveis *on-line*.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

SABI CATALOGO ON-LINE

Catálogo atual: Catálogo UFRGS
Usuário: FABIANA HENNIES BRIGIDI

Encerrar sessão
Usuário
Empréstimos/Renovações
Débitos
Minha pasta
Meus perfis DSI
Ajuda

Nova pesquisa Última pesquisa Pesquisas anteriores Catálogos Agendamentos Fale conosco

Pesquisa básica | Pesquisa avançada | Pesquisa multicampo | Pesquisa CCL | Percorrer índice

Pesquisa CCL - Linguagem de Comandos

Digitar frase em linguagem de comandos
WPG=CMPP AND WLN=POR AND WLI=TEXTO COMPLETO AND WFT=T

Palavras adjacentes? ☒ Não ☐ Sim

Buscar Limpar

Filtros
Idioma: Português
Ano de: 2014 até 2018
AAAA (Use 7 para truncar o ano quando não usar intervalo de anos)
Ex: 2007 recupera o período de 2000 a 2009

Figura 2 – Pesquisa no Sabi

Conforme os Quadros 2 e 3 e a Figura 2, a estratégia de pesquisa teve o objetivo de recuperar somente as teses e dissertações provenientes do PPGC/UFRGS em língua portuguesa com texto disponível *on-line* dos últimos cinco anos. A referida pesquisa recuperou 161 registros.

De posse dos documentos para constituição do *corpus*-amostra, passou-se para a etapa de formatação em arquivo txt, composto de conteúdo textual apenas, conforme preconiza o critério de mesma nomenclatura (formatação) na LC (BERBER SARDINHA, 2004). Todos os registros foram copiados para o formato txt e nomeados com algarismos arábicos para seguir uma numeração sequencial referente à pesquisa no Sabi. Essa formatação em txt é necessária por ser requisito de utilização do software *AntConc*. Além disso, para tornar a análise mais eficiente, optou-se pelo uso de uma *stoplist*¹⁰, formada por conteúdo gramatical dispensável como artigos e preposições que devem ser ignorados pelo programa.

Os 161 arquivos já formatados foram incluídos no *AntConc* para iniciar a análise propriamente dita. Inicialmente, optou-se pela geração de lista de palavras individuais denominada *Word List*¹¹ (Figura 3).

¹⁰ Essa lista foi retirada do blog do Stanley Loh. Disponível em: <http://miningtext.blogspot.com/2008/11/listas-de-stopwords-stoplist-portugues.html>. Acesso em: 22 jan. 2019.

¹¹ Berber Sardinha (2004) explica que a *Word List* contém uma lista de palavras elencadas em conjunto com suas frequências absolutas e percentuais e, neste caso, ordenada por frequência.

Rank	Freq	Word
1	172	dados
2	155	trabalho
3	136	resultados
4	128	abordagem
5	128	redes
6	125	rede
7	103	diferentes
8	102	forma
9	98	método
10	97	sistemas
11	91	informações
12	87	processo
13	87	usuários
14	86	além
15	84	modelo
16	83	aplicações
17	82	desempenho
18	81	podem
19	80	sistema

Figura 3 – Word List

A *Word List* permite uma visão geral das palavras presentes no *corpus*-amostra por ordem de frequência, proporcionando uma especulação inicial de candidatos a termos. Nessa etapa, foram obtidas 25235 palavras (*Word Tokens*), sendo 5679 palavras diferentes (*Word Types*). A partir da *Word List*, também, foi possível verificar em que contexto as palavras estavam inseridas por meio da aba *Concordance*. Na aba *File View* visualizou-se o contexto de um termo em um texto específico com todas as suas ocorrências.

Utilizou-se como critério de frequência até três ocorrências em função da dimensão do *corpus*-amostra, passando-se para a análise de trigramas, conforme ilustra a Figura 4.

Total No. of N-Gram Types		Total No. of N-Gram Tokens	
Rank	Freq	Range	N-gram
1	27	24	estado da arte
2	27	17	um conjunto de
3	25	22	o desenvolvimento de
4	24	19	a partir de
5	21	19	a fim de
6	21	17	o uso de
7	21	17	os resultados obtidos
8	20	6	processos de negócio
9	19	16	de acordo com
10	18	14	a criação de
11	18	12	o número de
12	18	16	o objetivo de
13	18	15	é capaz de
14	17	15	a necessidade de
15	16	12	o método proposto
16	16	15	o processo de
17	16	13	por meio de
18	15	12	deste trabalho é
19	15	13	um modelo de

Figura 4 – Trigramas

A lista de trigramas forneceu 44601 palavras (*Total No. of N-Gram Tokens*), sendo 39329 palavras diferentes (*Total No. of N-Gram Types*). Essa listagem foi comparada com a *Word List* para identificação dos candidatos a termos compostos.

Após a realização das etapas descritas e das análises realizadas, incluindo a verificação dos contextos em que as possíveis UTs se encontravam, chegou-se aos resultados da pesquisa apresentados na próxima seção.

4 Resultados e discussão

Considerando-se os preceitos teórico-metodológicos que fundamentam a presente pesquisa e que incluem os aspectos comunicativos, cognitivos e sincrônicos¹² do *corpus*-amostra segundo a TCT e a TST aliados à LC, foi possível obter resultados preliminares, visto que se trata de um estudo-piloto.

¹² Embora este estudo contenha especificações de diacronia, nesta etapa da pesquisa a medida de tempo segue os critérios estabelecidos por Berber Sardinha (2004) acerca do tempo, neste caso, sincrônico, por compreender um período de tempo delimitado (2014-2018).

A partir da *Word List* elaborou-se uma lista preliminar de candidatos a termo listados em ordem decrescente de ocorrências, conforme ilustra o Quadro 3. Foram recuperadas 1212 palavras com até cinco ocorrências, dessas, 49 possíveis candidatas a termo.

Quadro 3 – Candidatos a termo da *Word List*

Ocorrências	Candidatos a termo	Ocorrências	Candidatos a termo	Ocorrências	Candidatos a termo
128	Redes	8	<i>Finfet</i>	6	ACL
125	Rede	8	interoperabilidade	6	Android
51	software	8	processadores	6	Automação
36	Web	8	Roteamento	6	EDP
15	hardware	8	Transistores	6	EMPS
14	<i>Threads</i>	7	BPMN	6	Flops
12	ontologia	7	<i>Dynamics</i>	6	Online
11	metadados	7	Escalonador	6	TVF
11	polaridade	7	Links	5	Bit
9	<i>Cloud</i>	7	<i>Openflow</i>	5	<i>Codes</i>
9	<i>Flip</i>	7	Sensoriamento	5	Grafo
9	<i>middleware</i>	7	<i>Ubicomp</i>	5	HCS
9	<i>monitoring</i>	7	Ubíqua	5	OLPP
8	<i>benchmark</i>	7	Ubíquas	5	POR
8	<i>Cache</i>	7	VLSI	5	Softwares
8	codificação	6	Access		

Fonte: Elaborado pela autora, 2019.

Os candidatos a termo foram identificados conforme os critérios comentados neste trabalho, como, por exemplo, frequência, representatividade da área, contexto em que aparecem e intuição da autora. Destacam-se as seguintes observações:

- Três palavras aparecem no singular e no plural (rede/redes e software/software, ubíqua/ubíquas): no caso de serem consideradas UTs, será necessário optar por uma das versões apresentadas, pois possuem o mesmo significado;
- Embora a pesquisa possuísse o filtro para língua portuguesa, muitos termos constam em inglês: esse é um indício de que a área da Ciência da Computação possui muitos termos em língua inglesa, conforme já comentado neste texto;
- Foram identificadas várias siglas: entende-se que autores da área pressupõem que o leitor compreenda o significado dessas siglas, no entanto, é responsabilidade do bibliotecário/terminólogo esclarecê-las, bem como definir se são representativas para atingir o status de UT.

A pesquisa realizada a partir dos trigramas resultou em 244 registros com até cinco ocorrências, sendo que 30 foram selecionados como candidatos a termos compostos, segundo ilustrado no quadro a seguir.

Quadro 4 – Candidatos a termo composto dos trigramas

Ocorrências	Candidatos a termo	Ocorrências	Candidatos a termo	Ocorrências	Candidatos a termo
14	aprendizagem de máquina	7	conjunto de dados	5	algoritmos de aprendizagem
14	objetos de aprendizagem	7	cross language links	5	banco de dados
13	modelagem de processos	7	recuperação de informação	5	circuitos integrados
11	redes virtuais	7	técnicas de visualização	5	<i>cross language</i>
10	linguagem natural	6	bancos de dados	5	dispositivos de encaminhamento
10	modelos de processo	6	bases de dados	5	internet das coisas
9	conjuntos de dados	6	<i>dynamic c ran</i>	5	<i>internet of things</i>
9	funções de rede	6	representação de conhecimento	5	processamento de linguagem
8	base de dados	6	sistemas de recomendação	5	redes de computadores
8	injeção de falhas	6	<i>software defined networking</i>	5	requisitos não funcionais

Fonte: Elaborado pela autora, 2019.

Assim como os candidatos a termo da *Word List*, o levantamento dos candidatos a termos compostos provenientes da lista de trigramas seguiu os mesmos critérios. Observam-se algumas similaridades do Quadro 4 com o Quadro 3. São elas:

- Três candidatos a termos apresentam-se tanto no plural como no singular (conjunto/conjuntos de dados, banco/bancos de dados, base/bases de dados): nesse caso também é necessário optar por uma das formas;
- Quatro ocorrências de palavras em língua inglesa pelos mesmos motivos já mencionados;
- Um dos candidatos a termo consta em língua portuguesa e inglesa simultaneamente (internet das coisas/*internet of things*): embora esse caso não tenha ocorrido na *Word List*, este é mais um indício da influência da língua inglesa na área da Ciência da Computação.

A partir do cruzamento de informações dos Quadros 3 e 4, é possível observar que, num primeiro momento, alguns candidatos a termo não pertencem necessariamente à área da Ciência da Computação, como, por exemplo, ‘polaridade’, ‘transistores’, ‘modelos de processo’ e ‘técnicas de visualização’. Essa constatação indica que diferentes áreas de conhecimento podem se inter-relacionar, manifestando, assim, uma ideia de interdisciplinaridade.

Contudo, para a seleção de UTs de uma determinada área, é preciso ter parâmetros que definam o que é ou não condizente com a elaboração de um produto

terminográfico, visto que esse é um dos objetivos futuros desta pesquisa. Considerando que este estudo-piloto permite um levantamento preliminar de candidatos a termo da área da Ciência da Computação para a elaboração de um vocabulário controlado, os exemplos acima mencionados não seriam representativos da área e, portanto, provavelmente não seriam incluídos no referido produto.

Outro aspecto a ser comentado se refere à ausência de relação entre os candidatos a termos da *Word List* com a lista de trigramas. Embora provenientes do mesmo *corpus*-amostra, ambas as listas trouxeram palavras diferentes. Em função disso, entende-se que para a elaboração de um vocabulário controlado será necessário fazer uso de ambas as ferramentas para a identificação dos termos simples e compostos para inclusão.

É preciso destacar algumas características das teorias da Terminologia que fundamentam este estudo, correlacionando-as aos candidatos a termos, como, por exemplo, o contexto comunicativo especializado em que ocorrem incluindo os usuários envolvidos (docentes, discentes, pesquisadores), a situação em que são utilizados, a temática veiculada e o tipo de discurso, nesse caso, a comunicação científica. As questões referentes às variações terminológicas em função da diacronia necessitam de um recorte mais amplo que forneça subsídios para sua identificação. Por fim, entende-se a necessidade de estudos aprofundados de cada candidato a termo, considerando-se a consulta a especialistas, para posterior inclusão em um vocabulário controlado da área da Ciência da Computação.

5 Considerações finais

O recorte de pesquisa aqui apresentado propiciou a identificação de prováveis UTs no contexto da produção intelectual da área da Ciência da Computação referente ao PGCC/UFRGS dos últimos cinco anos. Entende-se que o tipo de documento escolhido para esse estudo se destaca pela atualidade e confiabilidade dos temas abordados, relacionando-se aos conceitos da TCT e TST e, portanto, configurando-se num mecanismo eficiente no uso de terminologias.

Em meio a essa realidade, o *corpus*-amostra deste estudo reflete as últimas pesquisas de mestrado e doutorado da Universidade no domínio em questão, trazendo termos utilizados por pesquisadores. Os resultados obtidos refletem aspectos pragmático-discursivos da área por meio das terminologias identificadas.

Foi possível constatar que alguns candidatos a termo identificados neste estudo poderiam correlacionar-se com outras áreas de conhecimento, não se limitando apenas à Ciência da Computação. Entretanto, o contexto discursivo é determinante na área de especialidade, por isso, no *corpus*-amostra, destacam-se outros candidatos a termos específicos da área.

Importante destacar a relevância da elaboração de diretrizes específicas para inclusão de UTs em produtos terminográficos como os vocabulários controlados. No caso das bibliotecas, recomenda-se a criação de uma política de indexação que abarque todas as possibilidades de tratamento temático, como, por exemplo, a inclusão de remissivas, a definição dos idiomas que serão tratados, a definição dos termos escolhidos, o período de atualização da ferramenta, entre outros. No caso específico da Ciência da Computação, a questão diacrônica é essencial em função dos constantes avanços tecnológicos que propiciam o surgimento e a atualização de UTs.

Carneiro (1985) simplifica e destaca que os elementos a serem considerados numa política de indexação envolvem a cobertura de assuntos; a seleção e aquisição de documentos-fonte; o processo de indexação incluindo níveis de exaustividade e especificidade, escolha da linguagem, capacidade de revocação e precisão do sistema; a estratégia de busca; o tempo de resposta do sistema; a forma de saída e a avaliação do sistema. Todos esses elementos podem ser adaptados às necessidades da UIs bem como à fundamentação teórica terminológica escolhida, no caso deste estudo, a TCT e a TST.

Por fim, este estudo-piloto teve o propósito de apresentar um breve panorama terminológico acerca da área da Ciência da Computação no contexto discursivo-comunicativo do PGCC/UFRGS com a finalidade de fundamentar um futuro projeto de doutorado. Conclui-se que as terminologias utilizadas por pesquisadores da área englobam termos em língua inglesa bem como siglas que necessitam de esclarecimentos. Observa-se a necessidade de um *corpus* de estudo em larga escala para uma melhor identificação das terminologias e suas atualizações. Portanto, entende-se que esta pesquisa necessita aprofundamento para fornecer resultados mais concretos a fim de contribuir com a literatura da área, auxiliando profissionais da informação essencialmente.

REFERÊNCIAS

ANTHONY, Laurence. *AntConc* [Software]. Okubo: Waseda University, 2017. Disponível em: <http://www.laurenceanthony.net/software/antconc/>. Acesso em: 28 jul. 2018.

ARANALDE, Michel Maya. A questão ética na atuação do profissional bibliotecário. *Em Questão*, Porto Alegre, v. 11, n. 2, p. 337-368, jul./dez. 2005.

BARROS, Lidia Almeida. *Curso básico de terminologia*. São Paulo: Edusp, 2004.

BERBER SARDINHA, Tony. *Linguística de corpus*. Barueri: Manole, 2004.

BEILKE, Neubiana Silva Veloso. Pommersche korpora: um conjunto de corpora dialetais da variedade brasileira do pomerano. In: FINATTO, Maria José Bocorny Finatto et al. (Orgs.). *Linguística de corpus: perspectivas* [recurso eletrônico]. Porto Alegre: Instituto de Letras - UFRGS, 2018. p. 365-398.

BRIGIDI, Fabiana Hennies. *Indexação híbrida: vocabulário controlado e folksonomia*. 199 p. Dissertação (Mestrado) - Universidade do Estado de Santa Catarina, Centro de Ciências Humanas e da Educação, Mestrado em Gestão de Unidades de Informação, Florianópolis, 2016.

CABRÉ, Maria Teresa. *La terminología: representación y comunicación*. Barcelona: Universitat Pompeu Fabra, 1993.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 14, n. 2, p. 221-241, set.

1985. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/reb/>. Acesso em: 07 mar. 2019.

CINTRA, Anna Maria Marques et al. Linguagens documentárias e terminologia. In: ALVES, Ieda Maria (Org.). *A constituição da normalização terminológica no Brasil*. 2. ed. São Paulo: FFLCTH/CITRAT, 2001.

CORREIA, Elisa C. D. *O uso do computador e a definição do campo da Ciência da Informação em relação à Biblioteconomia no Brasil: uma análise sociotécnica*. 2008. 246 f. Tese (Doutorado em Sociologia Política) - Universidade Federal de Santa Catarina, Florianópolis, 2008. Disponível em: <http://repositorio.ufsc.br/xmlui/handle/123456789/91932>. Acesso em: 15 out. 2014

COSTA, Maria Izabel Plath da. *Terminologia jurídico-policia: proposta de elaboração de um glossário eletrônico*. 2014. 287 f. Tese (doutorado) - Universidade Federal do Rio Grande do Sul. Instituto de Letras. Programa de Pós-Graduação em Letras, Porto Alegre, RS, 2014. Disponível em: <http://www.bibliotecadigital.ufrgs.br/da.php?nrb=000931729&loc=2014&l=cbda55f32728c2af>. Acesso em: 17 jul. 2018.

DAL'EVEDOVE, Paula Regina. *A perspectiva sóciocognitiva no tratamento temático da informação em bibliotecas universitárias: aspectos inerentes à percepção profissional*. 2010. 301 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2010. Disponível em: https://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/dalevedove_pr_me_mar.pdf. Acesso em: 30 jan. 2019.

FONSECA, Cíntia Cíbele Ramos. *Descrição terminológica com vistas à indexação na área da fisioterapia na UFRGS: um estudo piloto* [Artigo final apresentado na disciplina de Fundamentos de Terminologia da UFRGS]. Porto Alegre, 2017.

GUIMARÃES, José Augusto Chaves. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. *Ibersid*, Zaragoza, v. 3, p. 105-117, 2009.

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

LAIPALT, Rita do Carmo Ferreira. *Metodologia para seleção de termos equivalentes e descritores de tesouros: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário*. 2015. 213 f. Tese (Doutorado em Linguística Aplicada) - Universidade do Vale do Rio dos Sinos, São Leopoldo, 2015. Disponível em: <http://hdl.handle.net/10183/126617>. Acesso em: 17 ago. 2019.

LANCASTER, F. W. *Indexação e resumos: teoria e prática*. 2. Ed. Brasília, DF: Briquet de Lemos, 2004.

MARIAN, Jane. Terminologia especializada: um estudo baseado na linguística. Curitiba, *Caderno PAIC*, v. 16, n. 1, 2015. Disponível em: <https://cadernopaic.fae.edu/cadernopaic/article/view/109>. Acesso em: 21 fev. 2019.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. *American Documentation*, Cleveland, v. 2, 1951, p. 20-32.

PIOVEZAN, Luciana Beatriz; FUJITA, Mariângela Spotti Lopes. Análise de cocitação de autores: uma aplicação em estudos de indexação. *Em Questão*, Porto Alegre, v. 21, n. 1, jan./abr. 2015, p. 110-129.

RUSSO, Mariza. *Fundamentos de biblioteconomia e ciência da informação*. Rio de Janeiro: E-papers Serviços Editoriais, 2010.

SABI: catálogo online [UFRGS]. Disponível em: <https://sabi.ufrgs.br>. Acesso em: 22 jan. 2019.

SILVA, Daniela Lucas da; SOUZA, Renato Rocha; ALMEIDA, Maurício Barcellos. Ontologias e vocabulários controlados: comparação de metodologias para construção. *Ciência da Informação*, Brasília, DF, v. 37, n. 3, p. 60-75, set./dez. 2008.

TAGNIN, Stella. E a Linguística de Corpus vai desbravando novos horizontes. In: FINATTO, Maria José Bocorny Finatto et al. (Orgs.). *Linguística de corpus: perspectivas* [recurso eletrônico]. Porto Alegre: Instituto de Letras - UFRGS, 2018. p. 11-18.

TEMMERMAN, Rita. Teoria sociocognitiva da terminologia. *Cadernos de Tradução*, Porto Alegre, n. 17, out./dez. 2004, p. 31-50.

TRISTÃO, Ana Maria D.; FACHIN, Gleisy Regina B.; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesauros: instrumentos para organização do conhecimento. *Ciência da Informação*, Brasília, DF, v. 33, n. 2, p. 161-171, maio/ago. 2004.

VIEIRA, Simone Bastos. Indexação automática e manual: revisão de literatura. *Ciência da Informação*, Brasília, DF, v. 17, n. 1, jan./jun. 1988, p. 43-57.

ZILIO, Leonardo. Termo e valor linguístico: uma abordagem ensaística. *Cadernos do IL*, Porto Alegre, n. 42, jun. 2011. p. 119-128. Disponível em: <https://seer.ufrgs.br/cadernosdoil/article/view/26016>. Acesso em: 20 jan. 2019.