

A linguagem documentária em catálogos on-line para a política indexação

Vera Regina Casari Boccato

Como citar: BOCCATO, Vera Regina Casari. A linguagem documentária em catálogos on-line para a política de indexação. *In:* LEIVA, Isidoro Gil; FUJITA, Mariângela Spotti Lopes (org.). **Política de indexação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2012. p. 139-152.
DOI: <https://doi.org/10.36311/2012.978-85-7983-199-7.p139-152>



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Marília



**CULTURA
ACADÊMICA**
Editora



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).

CAPÍTULO 5

A LINGUAGEM DOCUMENTÁRIA EM CATÁLOGOS ON-LINE PARA POLÍTICA DE INDEXAÇÃO

Vera Regina Casari Boccato

INTRODUÇÃO

É fato que o uso das tecnologias de informação e comunicação nas atividades de tratamento e recuperação da informação propiciou novas perspectivas para os centros de documentação, serviços de bibliotecas, entre outros espaços e ambientes informacionais.

Sobre isso, Fujita (2005) ressalta a importância das mudanças de paradigmas da informação determinadas por três fatores inovadores e com as quais o bibliotecário deve lidar. São eles:

1. forma: a diversidade de formatos exige tratamento temático e descritivo compatíveis com consequente modificação de normas, diretrizes, manuais e metodologias; a coexistência do formato impresso e do formato eletrônico;
2. acesso: a evolução tecnológica da comunicação de dados facilitou o acesso simultâneo de todos a todos os registros;
3. valor: a informação registrada, tratada e disseminada tem um valor mais alto.

Nesse contexto, a informação é tratada como um bem, adquirindo valor inestimável independentemente do suporte em que é apresentada, exigindo novas abordagens, novos métodos e novas tecnologias no seu ciclo de vida, isto é, na sua coleta, seleção, tratamento e disseminação para uso imediato e de acordo com a necessidade.

No que se refere ao tratamento temático da informação, com destaque para o processo de indexação, Boccato (2009b) aponta para a necessidade de construção de linguagens documentárias pela perspectiva sociocognitiva do

bibliotecário indexador e do usuário, pois segundo Frohmann (1990) a indexação consiste em dois distintos níveis de operações: a primeira envolve a representação implícita ou explícita do documento por termos de indexação, e o segundo a transição dos termos da indexação para o léxico de um vocabulário controlado possuidor de relações semânticas e sintáticas provenientes da linguagem de indexação.

Na abordagem de Hjørland (1997, 2002), a indexação é um processamento intelectual que depende da cognição; e o domínio dos contextos físico, psicológico e sociocognitivo é necessário para que o bibliotecário indexador realize a identificação e a seleção de conceitos na concepção orientada para o conteúdo e para a demanda.

O contexto realiza a intermediação entre a situação real dentro da biblioteca - norteadas pelas concepções de análise de assunto utilizadas pelo bibliotecário indexador – e o sistema linguístico – linguagem documentária adotada pelo sistema, caracterizado pela representação dos documentos indexados.

A abordagem sociocognitiva preconiza a interação existente entre os ambientes organizacionais e informacionais e as relações mentais e culturais dos bibliotecários indexadores e usuários refletidos no desempenho de suas atividades.

Sobre a recuperação da informação verificamos estudos, tais como os de Anderson (1998), Morisson (1999), Miller (2004), Novotny (2004), Guha e Saraf (2005) e Hearn (2009) que demonstram preocupação no desenvolvimento de sistemas automatizados capazes de atender ao máximo as necessidades de diferentes usuários de diversos ambientes informacionais, visando a rapidez no acesso, a qualidade da informação buscada e recuperada, a possibilidade de interação e de interconexão entre usuários e sistemas de recuperação, tais como os catálogos on-line.

Este capítulo, portanto, trata da temática linguagem documentária numa perspectiva sociocognitiva, visando a escolha da linguagem mais adequada de catálogos on-line de bibliotecas para uma política de indexação.

OS CATÁLOGOS ON-LINE DE BIBLIOTECAS CENTRADOS NO CONTEXTO SOCIOCOGNITIVO DO USUÁRIO

Os catálogos on-line, também conhecidos como OPACs¹, são sistemas automatizados de recuperação da informação que se encontram disponíveis de maneira mais geral e os primeiros a que se recorre qualquer usuário quando

¹ OPAC: *Online Public Access Catalog*. Em português, Catálogo Público de Acesso On-line.

tem uma necessidade de informação (FERNÁNDEZ MOLINA; MOYA ANEGÓN, 1998),

Na concepção de García López (2007), os OPACs permitem a consulta e a visualização dos registros que compõem o fundo de uma coleção, de um acervo de uma biblioteca (ou de várias delas) e atualmente são modelados para que os usuários possam interagir diretamente, sem a necessidade de intermediários.

Rubi (2008, p. 12) relata que “[...] os catálogos, antes locais e restritos a determinada comunidade, agora se tornaram disponíveis através da Internet, atravessando fronteiras geográficas, e funcionando como verdadeiras vitrines das bibliotecas”.

Para Fujita (2009, p. 12)

[...] os catálogos são instrumentos plurifuncionais com possibilidades de acesso múltiplo cujas formas de representação documentária estão organizadas em metadados, Estão [...] disponíveis na web para que qualquer usuário, a qualquer tempo e em qualquer lugar, possa acessar.

Os catálogos on-line permitem aos usuários localizar os documentos existentes nos acervos, verificando o seu estado de circulação, disponibilidade e permitindo a realização de empréstimos e reservas desses materiais. Nesse sentido, o OPAC integra funções básicas dos serviços de bibliotecas, além de possibilitar a interconectividade a outros acervos, integrados por sistemas de redes de bibliotecas.

Boccatto (2009a) apresenta, na atualidade, a concepção dos OPACs marcada pelas Interfaces Gráficas de Usuário (GUI), caracterizando-se pelo uso da metodologia de hipertexto, isto é, com a utilização de várias interfaces multimídias, possibilitando a importação e exportação integradas de registros, a partir de um único comando (“*click*”) com a preocupação de desenvolver a interação de novas ferramentas de buscas e fontes de informação, acentuando cada vez mais a interoperabilidade entre sistemas de recuperação da informação.

O nosso ponto de vista encontra, também, suporte nos fundamentos teóricos de Hjørland (1997, 2002) e Jacob e Shaw (1998) sobre a modelagem de catálogos on-line vistos pela perspectiva de um sistema de informação integrado, objetivando o acesso, a recuperação, a localização e a obtenção rápida e fácil das informações e dos documentos impressos e eletrônicos representativos da necessidade real de informação do usuário e do seu contexto social.

Essa iniciativa retrata a necessidade da modelagem e da geração de catálogos on-line centrados no usuário e em suas conexões sociais que envolvem

o processamento da informação, realizada por equipes multidisciplinares e integradas - bibliotecários, usuários, autores, *designers* de catálogos, analistas de sistemas, entre outros - valorizadas pela presença dos bibliotecários como contribuintes no delineamento de pontos de acesso, exemplificados, também, pelos acessos analíticos de assunto, autor e título e por outros recursos aplicáveis à representação e recuperação da informação.

As visões sociocognitivas estão interessadas na cognição individual, porém abordam isso a partir do contexto social. O processo de conhecimento individual está associado ao contexto histórico, social e cultural das unidades e sistemas automatizados de informação responsáveis pelo tratamento e recuperação da informação (HJØRLAND, 2002)

Nessa perspectiva, García López (2007) considera dois componentes na modelagem dos OPACs, a saber: 1) Humanos, sinalizado pela participação dos usuários, dos bibliotecários e o contexto de trabalho (biblioteca ou centro de documentação); 2) Sistema automatizado, caracterizado pelo *hardware*, *software* e os elementos que constituem a estrutura do próprio OPAC – a interface do usuário, a interface do sistema de gestão da base de dados, o sistema de gestão da base de dados e a própria base de dados. (Grifo nosso).

Complementando a proposta de García López (2007), agregamos ao componente Sistema automatizado, particularmente no elemento estrutura do próprio OPAC – a interface do usuário, a linguagem documentária, componente fundamental dos sistemas automatizados de informação para o tratamento e a disseminação da informação de áreas que exigem um controle da terminologia.

Nesse contexto, identificamos as bibliotecas universitárias, responsável pela gestão das informações e dos documentos que integram seus acervos, desenvolvendo e mantendo, por meio das tecnologias de organização e representação do conhecimento, instrumentos, técnicas e sistemas de recuperação da informação que facilitam o acesso ao conhecimento produzido pela Universidade, num processo contínuo de geração e socialização do saber em contextos de áreas científicas especializadas.

Boccatto (2009b, p. 121) relata que os catálogos on-line de bibliotecas universitárias,

[...] inseridos em um contexto de áreas científicas especializadas, necessitam de instrumentos de organização e recuperação da informação compatíveis com suas características e da sua comunidade usuária.

Diante do exposto, vimos os estudos de usuários e de avaliação como estratégias sociocognitivas importantes para a gestão, organização e disseminação

da informação, no delineamento das necessidades informacionais e na observação das ações e dos procedimentos de busca e recuperação da informação em catálogos on-line com o uso de linguagens documentárias.

AS LINGUAGENS DOCUMENTÁRIAS EM CATÁLOGOS ON-LINE DE BIBLIOTECAS PARA POLÍTICA DE INDEXAÇÃO

Hoje é realidade o uso de buscadores como o Google, em que privilegia a linguagem natural como veículo de comunicação entre o usuário e o sistema, na localização de sites, portais, artigos eletrônicos, redes sociais, entre outras fontes de informação disponíveis na Rede Mundial de Dados – Internet.

Em vista disso, a partir de catálogos on-line - exemplificados pelo catálogo coletivo Athena² da Rede de Bibliotecas da Universidade Estadual Paulista – Unesp - podemos nos conectar com esse buscador, por meio dos resultados de buscas recuperados, conduzindo à complementação de informações, a demais assuntos e às pesquisas presentes na Rede associados a estes resultados alcançados.

Consideramos adequada a linguagem natural, isto é, a linguagem expressa pelo usuário, na busca e recuperação da informação em ambientes colaborativos em que ele é participante ativo, integrante desse processo. As folksonomias³ são um exemplo disso, pois são manifestações a partir da linguagem dos usuários na categorização de conteúdos na determinação de palavras-chave (*tags*). Segundo Wall (2006) as *folksonomias* são os resultados das marcações (atribuições) livres e pessoais de etiquetas (*tags*) a informações ou objetos (qualquer coisa com *URL- Uniform Resource Locator*), tendo em vista a sua recuperação.

Todavia, quando focalizamos unidades de apoio ao desenvolvimento do ensino e da pesquisa, caracterizadas pelas bibliotecas universitárias, estas possuem acervos especializados que requerem um tratamento temático para a recuperação da informação com grande precisão na especificidade dos assuntos.

As linguagens controladas possibilitam o acesso e a recuperação de informações pertinentes ao desejo de busca dos usuários a partir do controle do vocabulário que as compõem, decorrente da linguagem natural ou da linguagem de especialidade ou de ambas.

² Banco de dados Bibliográficos da UNESP – ATHENA: <http://portal.biblioteca.unesp.br/portal/athena/>

³ Folksonomia: “[...] é a tradução do termo *folksonomy* que é um neologismo criado em 2004 por Thomas Vander Wall, a partir da junção de *folk* (povo, pessoas) com *taxonomy*”. (CATARINO, M. E; BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. *DataGramaZero - Revista de Ciência da Informação*, Rio de Janeiro, v. 8 n. 3, jun. 2007. Disponível em: <http://dgz.org.br/jun07/Art_04.htm>. Acesso em: 06 jan. 2011).

As listas de cabeçalhos de assuntos são exemplos de linguagens controladas, formadas por cabeçalhos e subcabeçalhos advindos de repertórios da linguagem natural. Para Gil Urdiciain (2004), elas são linguagens pré-coordenadas, de estruturas associativas ou combinatórias que consistem em listas alfabéticas de palavras ou expressões da linguagem natural capazes de representar os assuntos de um documento.

A utilização das listas de cabeçalhos de assunto para a representação temática e recuperação da informação pode ser indicada, por exemplo, em catálogos on-line de bibliotecas públicas e comunitárias, correspondendo aos objetivos organizacionais e aos propósitos de busca dos usuários. O tratamento temático requer uma representação de assunto mais abrangente, priorizando as áreas mais gerais e, preferencialmente, disponibilizado em livros.

Em colaboração, a pesquisa de Boccato (2009b) demonstrou a inadequação no uso da Lista de Cabeçalho de Assunto da Rede Bibliodata (LCARB) para a indexação e recuperação da informação de áreas científicas especializadas em catálogos on-line de bibliotecas universitárias. Dentre os resultados obtidos, destacamos a falta de vocabulário especializado no repertório terminológico da LCARB, com a recomendação sobre a incorporação de termos específicos, tendo em vista a especificidade exigida do tratamento de conteúdos documentários para a recuperação precisa da informação de usuários especialistas.

Nessa perspectiva, a linguagem documentária, vista como uma linguagem controlada, possibilitará a representação de assuntos de áreas científicas que reverterá na qualidade da pesquisa realizada e na credibilidade do catálogo on-line quanto ao seu desempenho na recuperação da informação e na satisfação do usuário.

Para Boccato (2009b, p. 119) as linguagens documentárias são

linguagens estruturadas e controladas, construídas a partir de princípios e de significados advindos de termos constituintes da linguagem de especialidade e da linguagem natural (linguagem do discurso comum), com a proposta de representar para recuperar a informação documentária.

Isso posto, entendemos que o usuário de bibliotecas universitárias - discente de graduação, pós-graduação e docente pesquisador - fazendo uso de uma linguagem documentária representativa de sua área científica e de sua cultura terminológica na busca bibliográfica em catálogos on-line terá mais condições de obter resultados úteis e pertinentes à sua atividade investigativa os quais possibilitarão assisti-lo nas tomadas de decisões, nas resoluções de problemas e na geração de novos conhecimentos.

Os tesauros são exemplos de linguagens documentárias com alto rigor de construção terminológica e de relações conceituais, possibilitando a flexibilização na representação e recuperação da informação, pois são,

linguagens de estruturas combinatórias e pós-coordenadas, constituídas de termos - unidades linguísticas provenientes da linguagem de especialidade e da linguagem natural -, denominados de descritores, providos de relações sintático-semânticas, referentes a domínios científicos especializados, possibilitando a representação temática do conteúdo de um documento, bem como a recuperação da informação. (BOCCATO; RAMALHO; FUJITA, 2008, p. 201).

Os tesauros são formados por uma base léxica (descritores e não descritores) estruturada em relações conceituais dos tipos hierárquicas (termos genéricos e específicos), não hierárquicas (associativas – termos relacionados) e de equivalência (não-descritores – sinônimos ou quase-sinônimos) (BOCCATO, 2008).

A hierarquização dos termos permite realizar uma representação de assunto condizente com as ideias do autor. Por sua vez, a elaboração de uma estratégia de busca a partir de termos genéricos ou específicos propicia a obtenção de respostas satisfatórias em relação às necessidades de pesquisas dos usuários.

Os termos associados proporcionam a ampliação das estratégias de busca, no sentido da linguagem disponibilizar termos que estão relacionados com outros, isto é, um termo passa a lembrar outro, favorecendo a recuperação da informação. As relações de equivalência possibilitam o controle dos termos polissêmicos e sinonímicos, garantindo a recuperação dos assuntos.

Sobre isso, consideramos ser fundamental o incremento das relações associativas e de equivalência nas linguagens documentárias, o que propiciam a aproximação da linguagem do usuário com a linguagem adotada pelo catálogo on-line.

As linguagens documentárias devem ser construídas com base na terminologia de uma área, com categorias e subcategorias delimitadas, a partir de relações sintático-semânticas de termos precisos, explícitos e bem definidos (BOCCATO, 2011). A Terminologia, a Teoria da Classificação Facetada de Ranganathan, a Teoria do Conceito proposta por Hjørland (2009) e as normas internacionais possuem grande importância na construção de linguagens documentárias, com destaque para os tesauros conceituais, fornecendo subsídios teóricos, metodológicos e diretivos na compreensão, definição e compilação de termos, bem como no estabelecimento de relacionamentos conceituais entre eles.

A Terminologia estuda, teoricamente, os termos e seus respectivos conceitos, os sistemas de conceitos e sua representação. Entendemos, pois que a Terminologia é uma disciplina de caráter sistêmico que prepara *corpus* especializados com o fim de elaborar produtos técnico-científicos como dicionários e glossários de uma área do conhecimento em particular. “As vertentes teóricas da Terminologia estão relacionadas a propósitos pragmáticos, visando à comunicação universal do conhecimento e de suas ciências relacionadas” (BOCCATO; FUJITA, 2010).

A Teoria Sociocognitiva da Terminologia (TST), vista como vertente teórica da Terminologia, revela que os termos não podem ser entendidos fora de seu “[...] ambiente natural – os textos – “e de que estes consistem em um conjunto dinâmico de elementos linguísticos, pragmáticos, discursivos e comunicativos [...]” (BARROS, 2006, p. 25). A TST “[...] procura demonstrar que o conhecimento científico corresponde a um padrão sociocognitivamente modelado” e que no “[...] trabalho terminográfico, ela coloca como prioridade a adequação da obra ao perfil cognitivo e ao propósito pragmático do usuário”. (MACIEL, 2007).

Ressaltamos que a Teoria Comunicativa da Terminologia (TCT) e a Socioterminologia são auxiliares, também, na elaboração de linguagens documentárias socioculturais, pragmáticas e comunicativas.

Os preceitos da Teoria da Classificação Facetada de Ranganathan são colaborativos na construção de tesouros conceituais na medida em que eles norteiam o agrupamento de termos em domínios homogêneos e de suas divisões em categorias ou facetas, a hierarquização dos termos componentes de cada faceta e a flexibilidade nas combinações entre termos dessas facetas.

Numa perspectiva pragmática, a Teoria do Conceito de Hjørland (2009) preconiza que um conceito pode possuir mais de uma significação, sendo que pode estar representado em mais de uma categoria ou dentro de uma mesma categoria, com diferentes significados. Entendemos, pois que o conceito deve ser entendido como “significado socialmente construído”, identificado não só pelos discursos científicos, mas também pelos discursos dos usuários.

Os princípios das garantias literária, de uso, cultural e organizacional são contribuintes, também, na construção linguagens documentárias consistentes, a partir de termos consagrados pela literatura científica, do contexto sociocognitivo e cultural do usuário, das áreas científicas e do contexto organizacional das bibliotecas universitárias.

Segundo a norma internacional ANSI/NISO Z39:19 (2005) a garantia literária refere-se a frequente ocorrência do termo na literatura científica. A garantia de uso diz respeito aos termos livres ou controlados utilizados

frequentemente pelo usuário e que são armazenados e, conseqüentemente, coletados pelos sistemas automatizados de busca e recuperação da informação. A garantia organizacional trata dos termos que caracterizam e que representam o contexto da organização.

Para Beghtol (2002, p. 511) a garantia cultural pressupõe que qualquer sistema de organização e/ou representação do conhecimento pode ser apropriado e útil para os indivíduos em alguma cultura, somente se ele for baseado nas suposições, valores e preocupações dessa mesma cultura.

Dessa forma e considerando-se a universalização, no nível global e a particularização, no nível local que a linguagem documentária deve possuir, torna-se fundamental a interação entre os subsídios teóricos e metodológicos da Terminologia, das Teorias apresentadas, das diretrizes estabelecidas pelas normas internacionais e dos princípios de garantia literária, garantia de uso, garantia organizacional e garantia cultural na construção de linguagens consistentes para a representação de assuntos e recuperação da informação científica especializada em catálogos on-line de bibliotecas universitárias.

Esse cenário conduz, também, à uma mudança na postura de atuação do bibliotecário compromissado com a construção de linguagens documentárias compatíveis, não só com a sua comunidade usuária local (interna ou externa), mas também com uma comunidade usuária potencial remota, na busca da qualidade na informação recuperada, disponibilizada pelos catálogos on-line.

Tais profissionais devem ser valorizados em suas funções, considerando-se os conhecimentos prévios, as visões de mundo e os conceitos e linguagens adquiridos durante seu processo de interrelação com o meio social, como fatores contribuintes para a formação e desempenho satisfatório de um sistema de recuperação da informação mediante o uso de linguagens documentárias, a exemplo dos tesouros.

Baseado no estudo de Lopes (2002), analisamos as vantagens e desvantagens na adoção das linguagens natural ou documentária no processo de representação e recuperação da informação. São elas:

Quadro 21 – Vantagens e desvantagens das linguagens natural e documentária

LINGUAGEM NATURAL		LINGUAGEM DOCUMENTÁRIA	
VANTAGENS	DESVANTAGENS	VANTAGENS	DESVANTAGENS
Permite a imediata representação do assunto no catálogo on-line sem a necessidade de consulta a uma linguagem documentária e a interlocução entre o bibliotecário indexador e usuário pelo fato de utilizarem a mesma linguagem.	No processo de busca da informação, o usuário fará um esforço intelectual maior na identificação de termos polissêmicos, sinonímicos e homógrafos.	Controle total do vocabulário utilizado na representação da informação no processo de indexação, amenizando os problemas de comunicação entre bibliotecários indexadores e usuários.	Necessidade de disponibilização da linguagem documentária na interface de recuperação da informação para o usuário final.
Processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem documentária.	Alta revocação e baixa precisão na recuperação da informação no catálogo on-line.	A partir das notas de escopo do tesauros, os bibliotecários indexadores podem escolher mais adequadamente os descritores que retratam os assuntos dos documentos	Necessidade de treinamento no uso do tesauros, tanto para usuários profissionais (bibliotecários), quanto para usuários finais.
Termos de indexação são identificados e selecionados e representados diretamente dos documentos que vão constituir o catálogo on-line. Em ambientes colaborativos (<i>blogs</i> , <i>twitters</i> , e demais redes sociais), a indexação social é realizada pelo usuário na prática da identificação, seleção e representação dos termos a partir dos próprios recursos informacionais.	A estratégia de busca deverá ser exaustiva, arrolando todos os conceitos e seus respectivos sinônimos acerca da temática investigativa do usuário.	Uma linguagem documentária bem elaborada e consistente pode ocasionar alta precisão e relevância na recuperação da informação, transmitindo confiança ao usuário perante um possível resultado negativo e, conseqüentemente, a credibilidade no catálogo on-line.	Desatualização da linguagem documentária poderá conduzir a representação de assuntos inadequados e resultados insatisfatórios e negativos ao usuário.

Temas específicos citados nos documentos podem ser encontrados.	Perda de confiança do usuário no catálogo on-line perante um possível resultado negativo.	As relações hierárquicas, de equivalência e não-hierárquicas dos tesouros auxiliam tanto o bibliotecário indexador, quanto o usuário na identificação de conceitos relacionados e referência de termos associados.	Custo na construção e na gestão, além da necessidade constante de atualização e manutenção da linguagem documentária
---	---	--	--

FONTE: Adaptado de Lopes (2002, p. 48).

Dentre as diversas vantagens e desvantagens elencadas sobre a linguagem documentária, destacamos o fato de que ela traz maiores benefícios no processo de indexação com a produção de representações de conceitos por termos mais estruturados e padronizados, principalmente no contexto de áreas científicas especializadas. A falta de controle do vocabulário para o processo de indexação ocasiona a ambiguidade, a incidência de termos sinônimos e homógrafos e isto são ocorrências importantes a serem analisadas, pois refletem diretamente nos resultados encontrados nas buscas por assunto. Sobre a linguagem natural, esta viabiliza a interlocução entre o bibliotecário indexador e o usuário, pois ambos utilizarão a mesma linguagem no sistema de recuperação da informação.

Com a adoção da linguagem natural por um sistema de informação, a recuperação apresentar-se-á com um índice de revocação mais alto e com um índice de precisão mais baixo. Um dos motivos para essa ocorrência é que quanto mais pontos de acessos forem providos e disponíveis para o usuário na recuperação da informação por assuntos, maior será a possibilidade de recuperar documentos não condizentes com a sua necessidade de busca (LANCASTER, 2004).

Em contraposição, o controle do vocabulário é usado para melhorar a eficácia do armazenamento de informação e dos sistemas de recuperação, de navegação eletrônica e/ou em outros ambientes, aos que procuram identificar e encontrar o assunto desejado por meio da descrição de assunto, usando uma determinada língua. A finalidade preliminar do controle do vocabulário é conseguir a consistência na descrição de assuntos e facilitar a sua recuperação. A necessidade do controle do vocabulário faz-se presente para a resolução de características básicas da linguagem natural, a saber: duas ou mais palavras ou termos podem ser usados para representar um único conceito? Duas ou mais

palavras que têm a mesma ortografia podem representar conceitos diferentes? (ANSI/NISO Z39.19, 2005).

O controle do vocabulário traz benefícios para o bibliotecário indexador ao indexar com qualidade os conteúdos dos documentos para a recuperação, a partir do uso de uma linguagem capaz de representar o seu contexto, a cultura de sua área temática e do catálogo on-line que a utiliza.

Para isso, é fundamental o estabelecimento de uma política de indexação, com o intuito de garantir a realização de uma representação da informação documentária mais adequada aos propósitos de busca do usuário, conferindo credibilidade ao catálogo on-line.

A política de indexação, conforme Rubi (2008, p. 50), “[...] não deve ser vista como uma lista de procedimentos a serem seguidos, e sim uma filosofia que reflita os interesses e objetivos da biblioteca”.

Para Carneiro (1985, p. 221) uma política de indexação

[...] deve servir como um guia para tomada de decisões, deve levar em conta os seguintes fatores: características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

A autora (Carneiro, 1985, p. 229-239) também elenca sete elementos de importância para a definição de uma política de indexação em que, dentre eles, destacamos o item Escolha da linguagem.

A linguagem permitirá ao bibliotecário indexador representar o conteúdo documentário com precisão que conduzirá os resultados das estratégias de buscas elaboradas pelo usuário. A escolha da linguagem influencia o desempenho de um sistema de recuperação da informação, a exemplo do catálogo on-line, e, conseqüentemente, na recuperação da informação relevante ou não.

CONSIDERAÇÕES FINAIS

As diferentes unidades e ambientes informacionais utilizam em sistemas de recuperação da informação diferentes linguagens, sejam elas natural, controlada ou ambas.

Isso faz com que distintos aspectos e abordagens sejam considerados e realizadas para a escolha da linguagem mais adequada, visando o sistema de recuperação da informação em uso.

Em vista disso, os tipos de bibliotecas, o nível de profundidade do tratamento temático da informação empregado - a partir da escolha dos termos correspondentes a especificidade e exaustividade que a linguagem possui e, conseqüentemente, a especificidade e exaustividade do sistema⁴, as áreas de abrangência, os tipos de suportes disponibilizados pelo catálogo on-line e as categorias de usuários são elementos norteadores a serem analisados na escolha da linguagem para indexação e recuperação da informação do sistema automatizado.

Sob essa perspectiva, vê-se a linguagem documentária como um veículo de comunicação social, “imbuída de valores, em que os conceitos representados por termos devem refletir a cultura do indivíduo, do ambiente em que ele está inserido e da área de conhecimento a que ela corresponde” (BOCCATO, 2011).

O uso de uma linguagem documentária pelo bibliotecário indexador e pelo usuário do catálogo on-line destina-se, de forma pragmática, à escolha do termo correto visando diminuir a diversidade e a ambigüidade de vocabulário. Ela realiza uma mediação entre a linguagem do documento e a linguagem de busca do usuário e estabelece uma representação precisa por meio de termos que correspondem aos assuntos tratados pelo autor do documento.

No contexto das áreas científicas especializadas de catálogos on-line de bibliotecas universitárias a linguagem documentária mais apropriada para a indexação e recuperação da informação deve ser a mais específica possível (especificidade de termos), possuindo pouca coordenação entre os termos e que contemple os quesitos de: interação com o usuário, separação hierárquica dos termos, regras explícitas de combinação de termos, hierarquização de termos em categorias de assuntos equivalentes aos seus conceitos, incremento nas relações de equivalência e associativas, entre outros.

Os tesouros conceituais são indicativos de linguagens controladas e documentárias que retratam tais características. Construídos pelas perspectivas teóricas e metodológicas da área de Organização e Representação do Conhecimento em Ciência da Informação, do campo científico da Terminologia e das diretrizes e dos princípios estabelecidos pelas normas internacionais tornam-se sistemas de organização do conhecimento compatíveis entre a necessidade de representação, o mais fiel possível, dos conteúdos dos documentos para a recuperação precisa da informação de usuários especialistas.

⁴ Especificidade do sistema de recuperação da informação: refere-se “a extensão em que o sistema nos permite ser precisos ao especificarmos o assunto de um documento que estamos processando”. A exaustividade do sistema diz respeito a extensão em que analisamos certo documento, em que todos os assuntos discutidos são reconhecidos durante a indexação e traduzidos pela linguagem documentária (FOSKETT, 1973, p. 12).