

SUPTES TEÓRICOS PARA PENSAR LINGUAGENS DOCUMENTÁRIAS¹

Rodrigo de Sales

Resumo:

Este artigo consiste em uma breve revisão na literatura referente aos estudos e pesquisas que vêm sendo realizados no Brasil predominantemente nos últimos cinco anos a respeito da construção de modelos de representação do conhecimento, mais popularmente conhecidos como linguagens documentárias. O propósito é identificar quais as principais teorias vêm fornecendo subsídios para as respectivas pesquisas. Antes de mapear parte do campo epistemológico das linguagens documentárias, será explicitada uma breve contextualização histórica e conceitual a respeito das linguagens documentárias. Almeja-se concluir com a identificação de possíveis linhas ou tendências de pesquisa acerca da construção de modelos de representação do conhecimento.

Palavras-Chave:

Linguagem documentária, Terminologia, Lingüística

THEORETIC SUPPORTS FOR THINKING DOCUMENTARY LANGUAGES**Abstract:**

This article consists of an brief literature revision regarding the studies and researches that are being accomplished in Brazil predominantly in the last five years regarding the construction of models of knowledge representation, more popularly known as documentary languages. The purpose is to identify what are the main theories supplying aid to the respective researches. Before mapping the epistemological field of documentary languages, a brief historical and conceptual contextualization regarding the documentary languages will be made. Longing for concluding with the identification of possible research lines or tendencies concerning the construction of models of knowledge representation.

Keywords:

Documentary language, Terminology, Linguistics

¹ Artigo apresentado como conclusão da Disciplina Modelo de Representação do Conhecimento do Curso de Mestrado em Ciência da Informação da UFSC.

1 INTRODUÇÃO

Com o desenvolvimento do conhecimento científico e com a evolução das tecnologias de informação e comunicação, o conhecimento passou a ser produzido e divulgado com muita rapidez e em enorme quantidade. Em consequência disso, aumentaram as dificuldades no processo de tratamento da informação², principalmente no que diz respeito ao armazenamento e à recuperação. Com isso, potencializa-se o investimento na elaboração de linguagens artificiais que auxiliam as atividades de representação de conteúdos informacionais, viabilizando um armazenamento de informações facilmente recuperáveis. Essas linguagens próprias, denominadas linguagens documentárias – LDs, são constituídas por um léxico de termos específicos de uma determinada área de conhecimento, ou de um determinado contexto, e suas possíveis relações. Atualmente as LDs são utilizadas como instrumentos no processo da análise documentária³. Cada área do conhecimento possui sua própria linguagem de especialidade, e toda LD deve ser ancorada pela linguagem de especialidade (ou linguagem especializada) a qual está submetida. A especificidade das linguagens especializadas restringe seu uso aos conhecedores de determinado conjunto de códigos muitas vezes não conhecido por qualquer sujeito falante, pois as linguagens de especialidade são aquelas criadas e comunicadas entre membros possuidores de um conhecimento específico. As linguagens de especialidade possuem terminologias próprias (terminologias no sentido de conjuntos de termos específicos de um determinado domínio), que não são de conhecimento geral. Ressalta-se que toda linguagem de especialidade é concebida e aceita pelos membros da área de conhecimento que abriga a respectiva linguagem, ou seja, é concebida tanto pelos produtores das informações de um determinado domínio do conhecimento quanto por seus consumidores (usuários).

Linguagens documentárias (LDs) são sistemas de signos que visam a uniformização do uso da linguagem de especialidade, proporcionando uma representação padronizada do conteúdo informacional, bem como uma recuperação da informação mais pertinente. Por esse motivo, as linguagens documentárias são instrumentos fundamentais na atividade de

² O conceito de informação empregado neste artigo será semelhante ao de Buckland (1991, *apud* ALVARENGA, 2001) que consiste na noção de informação como ‘coisa’, como algo usado para informar. Porém, ressalta-se que o termo ‘coisa’, no referido conceito, não remete à noção de substância material, mas sim, à noção de bem tangível ou intangível.

³ Antigamente as LDs eram consideradas produtos do processo de análise documentária, pois aquelas eram geradas nesse processo.

indexação da informação. Dentre os diversos tipos de linguagens documentárias, pode-se citar os *cabeçalhos de assunto*, os *glossários*, os *dicionários*, os *vocabulários controlados*, os *sistemas de classificação decimal*, as *taxonomias*, os *tesauros* e as *ontologias*. Para efeito de esclarecimento, os modelos de representação do conhecimento focados neste artigo são aqueles que se fundamentam como um sistema complexo de termos e relações entre os termos, como os tesauros e as ontologias.

Este artigo visa identificar na literatura da Ciência da Informação como vêm sendo teoricamente embasados os estudos recentes realizados no Brasil sobre linguagens documentárias, ou modelos de representação do conhecimento. Com o propósito não somente de identificar os suportes teóricos, mas principalmente de identificar qual a teoria que melhor fundamenta a realidade das linguagens documentárias, o artigo apresenta uma breve análise sobre a Teoria Geral da Terminologia (TGT) e a Teoria Comunicativa da Terminologia (TCT), por considerar necessária a aproximação da Terminologia e da Ciência da Informação para estudos desta natureza. Para tanto, foram consultados artigos de periódicos e relatórios de pesquisa, especialmente Teses e Dissertações, publicados no Brasil nos últimos seis anos, correspondentes a linguagens documentárias. Para a localização dos artigos e relatórios de pesquisa foi realizada uma busca, livre de estratégias rigorosas, nos sítios das revistas científicas brasileiras da área da Ciência da Informação e nas bases de Teses e Dissertações da Biblioteca Central da Universidade Federal de Santa Catarina, além de consultas a Anais de Congresso. Livros de autores/pesquisadores da temática em questão, foram utilizados principalmente para embasar conceitualmente e teoricamente o respectivo trabalho

2 LINGUAGEM DOCUMENTÁRIA

Linguagens documentárias (LDs) são linguagens artificialmente construídas e constituídas de sistemas simbólicos que visam “traduzir” sinteticamente conteúdos documentais, utilizadas nos sistemas documentários para indexação, armazenamento e recuperação da informação. Para Saracevic (1996), a recuperação da informação foi (e continua sendo) a principal solução para o primeiro grande ‘problema’ da Ciência da Informação, a “explosão informacional”. Anunciada em 1945 por Vannevar Bush, a explosão informacional foi a mola propulsora das discussões incipientes no Pós-Segunda Guerra e a geradora da referida Ciência. Tal afirmação evidencia a importância da

recuperação da informação no universo da Ciência da Informação e, conseqüentemente, a relevância das linguagens documentárias.

O léxico utilizado pela linguagem natural é composto pelo vocabulário de uma determinada língua, ou seja, a relação de todas as palavras faladas por essa mesma língua, enquanto o léxico adotado pela linguagem documentária “é a relação de termos e símbolos empregados para a identificação temática nos sistemas de informação” (CAVALCANTI, 1978, p. 21). Para a mesma autora, linguagem documentária consiste em um conjunto de regras, símbolos e termos pré-estabelecidos com o intuito de indicar assuntos de documentos. Faz-se necessário distinguir linguagem natural (LN) e linguagem documentária (LD). Para Cintra et al. (2002), tanto a LN quanto a LD são sistemas simbólicos instituídos que visam facilitar a comunicação. Porém, a primeira é composta por palavras e seus diversos significados, ao passo que a segunda é composta por palavras que assumem o papel de termos, ou seja, restringem seus significados de acordo com um contexto ou de acordo com uma área específica do conhecimento.

Analisando sob uma ótica social, Cintra et al. (2002) afirma que a função das linguagens documentárias vai além da simples tradução de conteúdos documentais; ela tem a função de possibilitar que os diferentes segmentos sociais tenham adequado acesso aos estoques de conhecimento. Sem as linguagens documentárias, não seria possível cumprir com o caráter público da informação, pois são elas que garantem o compartilhamento social entre estoques de conhecimento e usuários de informação. Numa visão também social Lima (2004) reforça a afirmação acima alegando que a representação e a recuperação eficientes da informação possibilitadas pelas linguagens documentárias potencializam a construção de novos conhecimentos, fato este que contribui para o desenvolvimento da sociedade, pois segundo a autora, os usuários de informações documentárias são produtores de conhecimento em potencial. As afirmações de cunho social descritas acima, demonstram a preocupação de fazer das linguagens documentárias verdadeiros instrumentos capazes de possibilitar que a informação efetivamente cumpra seu papel com a sociedade, ou seja, as linguagens documentárias se tornam imprescindíveis para conectar a informação ao seu público, para daí poder exercer a transformação por meio do conhecimento adquirido.

Lima (2004) faz uso do termo *informação documentária* para denominar o signo documentário que possibilita a articulação entre conceitos e termos de um domínio de conhecimento, melhor dizendo, a *informação documentária* é o signo que, auxiliado por uma linguagem documentária, representa o conteúdo informacional de um documento. Para a autora, a informação documentária é o resultado de um processo de codificação (realizado pelo indexador) e de um processo de decodificação (realizado pelo usuário que acessa a informação). Contudo, é possível afirmar que informações documentárias são unidades significantes codificadas (descritores). Unidades significantes são entendidas neste texto como sendo o código que representa um conceito (significado) tendo como base um referente (objeto), este último proveniente de uma percepção e de uma interpretação social.

Termos, conceitos e características

Termos, conceitos e características são elementos comuns em toda linguagem documentária. Para Dahlberg (1978), filósofa alemã criadora da Teoria do Conceito, a soma total dos enunciados verdadeiros sobre um objeto forma o conceito desse objeto. Segundo a autora, os enunciados são compostos por elementos do conceito, esses elementos são as características do conceito, em outras palavras, características são os atributos predicáveis do objeto.

Linguagens documentárias são formadas por um conjunto de termos, que são considerados por alguns como designações verbais; os termos são considerados pelas ISOs 704 (2000) e 1087-1 (2000) como sendo a unidade mínima da terminologia. Para Lara (2004a) o termo é uma designação que corresponde a um conceito em uma linguagem de especialidade. O conceito é, para o Manual de Elaboração de Tesouros Monolíngües criado por Hagar Espanha Gomes em 1990, um construto mental que representa um objeto individual material ou imaterial. Pode-se afirmar, pautado na Teoria Comunicativa da Terminologia (TCT), que conceito é a representação mental de um objeto, e que, somado a uma unidade lexical que o denomina, forma o “termo”⁴. Os conceitos sempre estarão relacionados entre si, e o que determinará essas relações serão as características inerentes aos conceitos.

⁴ Esta definição de ‘terno’ fornecida por Cabré (1999) é a definição mais coerente segundo a perspectiva deste trabalho.

Em estudo mais aprofundado, Lima (2004) apresenta uma distinção entre termos e descritores, pois não raramente se encontra na literatura um uso indiscriminado desses dois signos que segunda Lima se distinguem em vários aspectos. Embasada por Lariviere, Lima (2004) pontua o seguinte: os termos pertencem exclusivamente às linguagens de especialidade, ao passo que os descritores podem pertencer às linguagens de especialidade e à linguagem comum; os termos são geralmente extraídos dos documentos, ao passo que os descritores geralmente são signos conhecidos dos usuários e em hipótese alguma são criados; os termos remetem a um objeto da realidade, enquanto que os descritores remetem a documentos que tratam dessa realidade; o termo se refere a uma só realidade (monorreferencial), enquanto o descritor é escolhido (preferencialmente) entre outros descritores para representar uma determinada realidade. Lima (2004) aponta também algumas semelhanças entre termos e descritores, a saber: ambos são unidades significantes que representam um conceito dentro de um domínio de conhecimento; e ambos são unidades da linguagem natural e não códigos artificiais.

Outro trabalho que apresenta uma maior preocupação em definir o que é o termo dentro de uma linguagem documentária, porém, menos detalhista em sua descrição, é o trabalho realizado por Tristão; Fachin; Alarcon (2004). Nele as autoras alegam que o termo é o elemento do conceito que sintetiza o próprio conceito, permitindo sua comunicação. Afirmam ainda que o termo denota um referente, ou seja, trata-se de um signo que representa um objeto. A tese de Lima (2004) converge com essa afirmação e complementa que esse objeto/referente é percebido e interpretado numa práxis, numa realidade social. Contrariando a afirmação de Tristão; Fachin; Alarcon (2004), o presente trabalho adota como mais adequada a definição postulada por Cabré (1999), que enfatiza ser o ‘conceito’ um elemento pertencente ao ‘termo’, pois o termo é composto por um conceito e uma unidade lexical que o denomina.

3 SUBSÍDIOS TEÓRICOS ENCONTRADOS NA LITERATURA PARA PENSAR AS LINGUAGENS DOCUMENTÁRIAS

Para um estudo consistente de linguagens documentárias, deve-se lançar mão, além das disciplinas de Ciência da Informação, de duas práticas científicas, a Lingüística e a Terminologia. Enquanto a Lingüística, de uma maneira geral, centra seus esforços

teóricos na “palavra” como elemento de um processo de significação amplo (a palavra por si própria), a Terminologia visa utilizar a palavra como “termo”, elemento de um processo de significação restrito. Enquanto a primeira estuda a palavra como elemento lingüístico, em sua origem, estrutura e significado⁵, a segunda se preocupa apenas com o significado (conceito) do termo. Na Lingüística, o contexto redacional em que se insere a palavra não é a prioridade, tornando assim, seu significado amplo e variável. Na Terminologia, o contexto de redação onde está inserido o termo é prioridade para conceituar seu significado, ou seja, a terminologia faz uso do contexto para demarcar território e delimitar o uso dos termos.

Linguagens documentárias, são constituídas de elementos delimitados por um determinado contexto ou por uma determinada área do conhecimento, melhor dizendo, são constituídas por termos. Com efeito, são subordinadas a uma terminologia, contextualizadas por um conteúdo documental ou por um domínio específico do conhecimento. Porém, a sintaxe e a semântica (objetos da Lingüística) são fatores relevantes na construção das linguagens documentárias. Outra noção lingüística que pode ser utilizada para melhor pensar a essência das linguagens documentárias (LD como sistema relacional de signos), é a noção semiológica de signos, significados e significantes, postulada por Ferdinand de Saussure e revisitada por Roland Barthes. A utilização da Lingüística para o processo de análise documentária data do início da década de 1970, apesar de já ter sido esboçado no final da década de 1950.

Nesta seção são apresentados alguns estudos recentes (relatados em artigos e relatórios de pesquisa) realizados no Brasil correspondentes a linguagens documentárias de modo a ressaltar os suportes teóricos adotados pelos pesquisadores.

A Teoria do Conceito de Ingetraut Dahlberg, que analisa o conceito como a soma total dos enunciados verdadeiros de um objeto, pontuando informações a respeito de suas características, relações, categorias e definições, é a teoria que vem frequentemente fornecendo subsídios teóricos aos estudos de representação do conhecimento na área da Ciência da Informação.

⁵ É importante frisar que esta definição das características da Lingüística se inscreve no estudo etimológico e principalmente na definição estruturalista de Ferdinand de Saussure, que serviu de suporte para Roland

Dentre as teorias da Terminologia que ancoram os estudos mais recentes de organização e representação do conhecimento, merecem destaque a Teoria Geral da Terminologia (TGT) de Eugen Wüster, Teoria da Socioterminologia de François Gaudin e a Teoria Comunicativa da Terminologia (TCT) de Maria Teresa Cabré. A TGT, dotada de uma perspectiva prescritiva e normativa, é responsável pelo fornecimento dos primeiros supostos de trabalhos terminológicos. A Socioterminologia dotada de uma perspectiva descritiva, dando ênfase ao uso social da língua, possibilita maior variação e flexibilidade lexical e conceitual, o que proporciona uma maior aproximação entre mecanismos de informação e usuários. A TCT, embasada pela Socioterminologia, fundamenta-se em aspectos comunicativos das línguas naturais. Paralelo a isso, a Teoria da Gramática Funcional de Simon Dik, também vêm sendo aplicada no campo da Ciência da Informação, postulando fundamentos preciosos aos estudos de linguagens documentárias. Essas perspectivas teóricas serão melhores descritas no decorrer deste artigo.

É relevante ressaltar que existem na literatura trabalhos como os de Mendonça (2000), Café (2003, 2006), Bufrem (2005) e Pereira e Bufrem (2005), para citar alguns, que enfatizam ser necessário uma relação estreita entre Lingüística e Ciência da Informação para melhor entender e construir linguagens documentárias. Por meio de um estudo bibliométrico referente aos temas abordados nos artigos dos fascículos da *Revista Ciência da Informação*, no período de 1972 a 1998, Mendonça (2000) constatou que, em média, apenas 01 (um) artigo contendo como tema a interseção entre lingüística e ciência da informação foi publicado por ano, durante os 26 anos investigados. Para a autora, que sugere a inserção da disciplina *Terminologia* nos currículos dos cursos de graduação e pós-graduação da área da ciência da informação, essa baixa produção literária (que aborda o respectivo tema) compromete a qualidade da organização do conhecimento e da recuperação da informação efetuadas por profissionais da área. Com base no estudo de Mendonça (2000), é seguro afirmar que essa baixa produção literária afeta diretamente o processo de construção de linguagens documentárias, pois tanto a eficiência da organização do conhecimento quanto à eficiência da recuperação da informação estão diretamente ligadas à eficácia da linguagem documentária que será empregada para os respectivos processos. Essa afirmação leva à conclusão de que a lingüística e a terminologia devem fazer parte do processo de construção de linguagens documentárias.

Café (2003), em trabalho que visa contribuir para a formação de base de conhecimento, muito pesquisada no tratamento automático da linguagem natural, propõe uma interpretação sob a luz da Teoria da Gramática Funcional (que abrange análise da sintaxe, da semântica e da pragmática dos termos) para elaborar uma metodologia de construção de *softwares* que se empenhem em reconhecer automaticamente segmentos lexicais em textos. Esse estudo é um exemplo categórico de uma análise qualitativa de termos-descriptores. Nele, a autora, ancorada pela Teoria de Simon Dik, analisa descritores compostos (constituídos por mais de uma unidade lexical) de textos escritos em português do Brasil na área de Biotecnologia de Cultura de Tecidos e Plantas. Esses descritores, denominados pela autora de *unidades terminológicas complexas* (UTCs), são submetidos a uma análise lingüística de natureza humana. A análise de UTCs apresentada por Café (2003) é um esforço intelectual que converge com a atual tendência em recuperação automática da informação, que é a construção de sistemas de tratamento de texto integral. A autora, ao dissertar a respeito das contribuições da Gramática Funcional na delimitação de segmentos descritores de informação, afirma que a relação entre a Lingüística e a Ciência da Informação é indispensável, pois ao lidar com o processamento automático de conteúdos de informação fazendo uso da linguagem natural, e ao elaborar ferramentas de recuperação de informação, é possível evidenciar a importância da relação entre as respectivas áreas.

Campos e Gomes (2003), estudando a organização e a representação dos domínios do conhecimento seguindo os princípios da teoria da classificação facetada de Ranganathan, afirmam que a importância fundamental da organização do conhecimento está ficando na superficialidade dos estudos voltados à representação de conteúdos de informação no ambiente em Rede, pois os princípios das classificações de assuntos são adotados nos dias atuais somente na organização da informação de suportes físicos tradicionais. As autoras ressaltam que

Os princípios apresentados por Ranganathan para a elaboração de uma base teórica sólida, visando a construção de classificação bibliográfica, como pode ser observado, é de fundamental importância para a organização de domínios de conhecimento. A representação de um domínio do saber, se configura como princípio norteador para a organização de documentos e informação. Dessa forma, consideramos que um repensar sobre estudos seminais, como o de Ranganathan, possam trazer novos rumos para conceitos atuais como o de organização de domínios do conhecimento. O desenvolvimento desses estudos, no âmbito de modelos teóricos de representação, permitirá ao profissional de informação a possibilidade de atuar cada dia mais num espaço interdisciplinar

que englobe questões ligadas à epistemologia, à lógica, à teoria cognitiva, à computação e à terminologia. (CAMPOS E GOMES, 2003, p. 162)

Nota-se na literatura que um número considerável de estudos relacionados à linguagem documentária são embasados teoricamente pela Teoria do Conceito de Dahlberg e pela Teoria Geral da Terminologia de Wüester. Segundo Tristão; Fachin; Alarcon (2004), para a compreensão do conceito e de suas relações é imprescindível adotar princípios da Teoria do Conceito, pois se trata de um método indicado para a fixação do conteúdo do conceito, conseqüentemente, determinante para o posicionamento desse dentro de um sistema de conceitos. Esse preceito converge com afirmações proferidas anteriormente e posteriormente por Moreira (2003), Gomes e Campos (2004), Moreira; Alvarenga; Oliveira (2004), Pereira e Bufrem (2005) et al., Pereira e Bufrem (2005), além de utilizarem as teorias do Conceito e da Terminologia, lançam mão da Teoria da Classificação Faceta de Ranganathan para analisarem princípios e métodos de organização e representação de conceitos para as linguagens documentárias.

Com o objetivo de compreender melhor os conceitos terminológicos, buscando transcender o uso normalizado dos termos e dando ênfase à importância da Terminologia na construção de linguagens documentárias, Lara (2004b) afirma que a ciência da informação ao assumir a responsabilidade de organizar e transferir informação deve necessariamente inserir em suas atividades de representação do conhecimento, sobretudo na construção de linguagens documentárias, princípios descritos pela lingüística documentária e, principalmente, pelas teorias da terminologia.

Discutindo sobre estruturas classificatórias subjacentes às linguagens documentárias, Batista (2004), refletindo a respeito das redes de conceitos sob a égide da teoria dos jogos de linguagem de Wittgenstein, do labirinto tipo rede de Umberto Eco, e do modelo do rizoma de Deleuze e Guattari, afirma que, assim como o usuário da língua é convencionalmente treinado a utilizá-la, o usuário da informação pertencente à comunidade científica também depende de linguagens pré-estabelecidas que explorem semelhanças entre conceitos de um determinado domínio do conhecimento, determinando seu uso.

No âmbito da era digital, Gomes e Campos (2004), Moreira; Alvarenga; Oliveira (2004) convergem em afirmar que a maioria dos trabalhos referentes à indexação e recuperação

de informação realizados com recursos automáticos, está operando somente com base na frequência de ocorrências de palavras pertencentes aos documentos submetidos à análise, ou seja, abrangendo somente aspectos estatísticos. O emprego somente de dados estatísticos não abrange os aspectos semânticos dos conteúdos dos documentos, comprometendo a qualidade da indexação e permanecendo muito aquém da indexação realizada por especialistas. Porém, os respectivos autores não mencionam trabalhos que vêm sendo realizados por pesquisadores da área da ciência da informação empenhados em desenvolver métodos de indexação automática considerando elementos semânticos contidos nas informações textuais. Estudos de Kuramoto (1996) e Moreiro, Marzal e Beltrán (2003) são exemplos de esforços voltados à qualidade da indexação e recuperação da informação tomando por base estruturas da linguagem natural, como sintagmas verbais e nominais, buscando identificar aspectos semânticos dos conteúdos informacionais. Souza (2005), em trabalho realizado após as afirmações de Gomes e Campos (2004), Moreira; Alvarenga; Oliveira (2004), avançando os estudos de Kuramoto (1996) e Moreiro; Marzal ; Beltrán (2003) propõe uma nova metodologia para indexação automática utilizando sintagmas nominais. Mesmo que os estudos de indexação e recuperação pautados em sintagmas não sejam ainda uma tendência de pesquisa amplamente difundida, a menção desses trabalhos deve ser incluída nas revisões de literatura, e os seus esforços devem ser continuados, pois se o paradigma da ciência da informação está voltado ao usuário da informação, não há mais como fugir da idéia de que a recuperação da informação ancorada somente em indexações quantitativas não facilitam a recuperação de informações pertinentes por parte do usuário.

Na questão da recuperação eficiente da informação através da normalização terminológica, Gomes e Campos (2004) ressaltam que os autores (sobre tudo os cientistas) também podem contribuir para a precisão da comunicação da informação ao adotarem em seus textos termos normalizados de acordo com linguagens documentárias de sua área de domínio.

Como proposta de elaborar uma metodologia para construção de linguagens documentárias, Cervantes (2004) busca na Terminologia, na Terminografia e na Teoria do Conceito elementos para fundamentar teoricamente sua pesquisa de Mestrado. A autora justifica sua fundamentação afirmando que a Terminologia, estudando signos de designação de linguagens de especialidade, é uma área cuja prática envolve “a criação de

termos, sua análise, processamento, apresentação e uso na comunicação especializada” (CERVANTES, 2004, p. 57), sendo o recolhimento, a constituição, a organização e a difusão dos termos sob a forma de dicionário terminológico, uma incumbência da Terminografia. Cervantes (2004) faz uso também do substantivo Terminótica, para designar o tratamento de termos com recursos automatizados, ou seja, a Terminótica é a Terminologia automatizada. Na Teoria do conceito, Cervantes (2004) busca bases mais consistentes para a compreensão do que é conceito no que tange representação e recuperação da informação.

Em estudo que visa determinar uma metodologia mais adequada para a representação de domínios de conhecimento, levando em conta objetivos de aplicações ligados às áreas da Ciência da Informação, da Ciência da Computação e à da Terminologia, Campos (2004) investiga e discute a problemática representacional nas referidas áreas, fato que permite a representação do conhecimento em diversos domínios, objetivando identificar um núcleo comum de conceitos fundamentais para modelar o conhecimento. Para tanto a autora resgata a teoria da classificação facetada de Ranganathan e a teoria do conceito de Dahlberg (na ciência da informação), o modelo orientado a objetos e a ontologia formal (na ciência da computação) e a teoria da terminologia de Wüester. No trabalho de Campos (2004), a representação do conhecimento, em busca de uma modelização ideal de domínios, é abordada com base no método de raciocínio, nos objetos de representação e na relação entre os objetos de representação.

Convergindo com os pensamentos de Mendonça (2000) e Café (2003), Pereira e Bufrem (2005) também afirmam ser imprescindível a relação entre a lingüística e documentação para aprimorar a organização e a recuperação da informação (consequentemente aprimorar a construção de linguagens documentárias), e acrescentam, embasadas na literatura, que o diálogo deve ser composto pela terminologia, pela ciência da informação e pela ciência da computação. As autoras afirmam que trabalhos como os de Wanderley (1973), Baranow (1983), Medeiros (1986), Miranda (1989), Boulanger (1995), Faulstich (1995 e 1995a) e Peixoto (1995) (PEREIRA; BUFREM, 2005), são enfoques diversificados que tratam o cientista da informação a partir de contribuições da Lingüística, da Terminologia e da Socioterminologia. Bufrem (2005) contempla ainda a literatura voltada à relação indispensável entre a Lingüística e a Documentação com uma resenha crítica do livro *El contenido de los documentos textuales: su análisis y*

representación mediante el lenguaje natural, de José Antonio Moreiro González, que segundo a autora trata-se de um “contexto teórico exaustivamente analisado e discutido, em que se realizam aproximações atuais relativas à análise documental” (BUFREM, 2005, p. 5).

Com os estudos levantados até aqui, fica evidente a indispensável necessidade de considerar aspectos concernentes à Lingüística, à Terminologia e à Documentação para refletir, no âmbito da Ciência da Informação, a conceitualização, a teorização, a elaboração, a aplicação e a avaliação das linguagens documentárias. Contudo as teorias mais freqüentes nos estudos aqui apresentados são a Teoria do Conceito, a Teoria da Classificação Facetada, a Teoria Geral da Terminologia (TGT), a Teoria Comunicativa da Terminologia (TCT) e a Socioterminologia. A próxima seção apresenta mais detidamente os fundamentos da TGT e da TCT, duas perspectivas terminológicas distintas, e até certo ponto antagônicas, que formam duas vertentes de estudo.

4 A TGT E A TCT

Seguindo uma corrente investigativa em evolução no Brasil (Enilde Faulstich e Marisa Bräscher, para citar algumas) e no mundo (François Gaudin e Maria Teresa Cabré), para citar alguns), Café (2006) faz uso da abordagem funcionalista para pensar a respeito da organização do conhecimento e das atuais tendências da Terminologia, ressaltando a evolução das teorias da Terminologia, que tanto tem contribuído para a representação do conhecimento fornecendo subsídios teóricos para a elaboração de linguagens documentárias. Cabe aqui fornecer uma explanação a respeito da perspectiva funcionalista. A ótica funcionalista, diferentemente da formalista, que metodologicamente se estrutura da função sintática para a semântica⁶, entende a linguagem como um instrumento de interação social utilizada para a comunicação de interlocutores reais. A perspectiva funcionalista busca entender a instrumentalidade da linguagem inserida nas situações sociais, como uma atividade cooperativa estruturada. Para tanto, a visão funcionalista parte do princípio que a função pragmática é que rege a semântica e a sintaxe nas expressões lingüísticas. Não se detendo às distintas correntes

⁶ O paradigma formalista, que considera a linguagem como um objeto abstrato baseado no conjunto de sentenças, acredita que só é possível entender o significado do discurso por meio da análise sintática, sendo esta a responsável em reger a semântica e o pragmatismo do texto.

funcionalistas⁷, pode-se afirmar em concomitância com Castilho (1994), Pezatti (1994), Neves (1997) que a gramática sob a égide da perspectiva funcional é uma atividade social.

A Teoria Geral da Terminologia (TGT) foi a primeira a tratar sistematicamente a terminologia como uma disciplina, concebendo assim os primeiros alicerces da teoria moderna. Na década de 1930, o engenheiro austríaco Eugen Wüster (1898-1977), preocupado em criar um instrumento de trabalho que desfizesse a ambigüidade na comunicação tecno-científica, publica em 1938 o *The Machine Tool*, dicionário especializado na área de engenharia. Com uma preocupação puramente metodológica e normativa (ainda não teórica), Wüster traça as primeiras linhas de sua teoria terminológica (a TGT). Décadas depois, mais precisamente em 1968, Wüster deixa de lado a prática metodológica para estudar a fundo a natureza dos termos. Para Cabré (1993) a Terminologia é abordada por Wüster como uma matéria autônoma descrita pela TGT e definida como um campo próprio em relação a outras ciências e outras disciplinas. Melhor dizendo, a Terminologia é considerada pelo criador da TGT como uma disciplina de encruzilhada entre a Lingüística, a Ciência Cognitiva, a Ciência da Informação, a Comunicação e a Informática.

Em síntese a TGT é uma teoria de caráter prescritivo, pautada em uma linguagem ‘ideal’ que desconsidera a flexibilidade efetiva do processo comunicativo no ambiente especializado, e visa a normatização⁸ (padronização) internacional dos termos especializados.

Na década de 1990, com base no funcionamento da comunicação especializada e com uma visão descritiva, surge a Socioterminologia que, fundada por François Gaudin em 1993, estuda as terminologias na análise da língua, privilegiando seu uso social, contrapondo a atribuição formal de rótulos postulados pela Teoria de Wüster. A Teoria de Gaudin abre caminho para outros trabalhos que seguem a influência de estudos lingüísticos e comunicacionais para pensar a Terminologia. Os estudos de Enilde

⁷ Neves (1997) discorre sobre o funcionalismo conservador (que apenas aponta os elementos inadequados do formalismo), o funcionalismo moderado (que além de ressaltar os pontos inadequados do formalismo, propõe uma análise funcionalista da estrutura) e, o funcionalismo extremado (que nega a estrutura e se baseia unicamente na função, desconsiderando a sintaxe).

⁸ Entende-se o termo normatização como referente à ‘tornar norma’, pronta e acabada. Diferentemente do que se entende por normalização, que remete à ‘tornar normal’, recomendação de uso.

Faulstich, da Universidade de Brasília, e de Maria Teresa Cabré, da Universidade Pompeu Fabra de Barcelona, são exemplos dessa tendência teórica. Gaudin (1993), afirmando que a terminologia não pode ser estudada isoladamente, sem se considerar o contexto social, econômico e comercial, constrói os pilares da Socioterminologia. O autor lança mão da sociolingüística, que considera o contexto social, cultural e político, para analisar o emprego da língua francesa, e afirma que a prática terminológica não pode de forma alguma ser dissociada do conhecimento do campo de atuação dessa prática, levando em consideração o texto produzido nesse campo, seu público, sua ação e sua utilidade. Segundo Gaudin (1993), as características da Socioterminologia possibilitam estudar o funcionamento dos léxicos profissionais que efetivamente são utilizados na comunicação entre os pares de um domínio de conhecimento ou profissional. Outra característica marcante da Socioterminologia é se reportar à origem dos termos, verificando não somente sua recepção e aceitação na comunidade especializada, mas principalmente identificando as causas do fracasso ou as razões do sucesso de seu uso efetivo. Tal fato, faz com que a respectiva teoria transcenda a terminologia escrita e normativa. É evidente que a Socioterminologia atua nas práticas lingüísticas e sociais concretas, e para tanto, localiza-se nas leis que unem trabalho e linguagem. Concomitantemente ao desenvolvimento da Socioterminologia surge a TCT.

A criação da TCT ganha forças na segunda metade da década de 1990, momento em que a Teoria Geral da Terminologia recebe inúmeras críticas por seu caráter reducionista e idealista. Não deixando de considerar a importância da TGT para o estudo sistemático da terminologia, mas sim, evoluindo o pensamento terminológico de acordo com as necessidades reais de comunicação do universo tecno-científico, Cabré (1999) ressalta que a perspectiva prescritiva e normativa da TGT limita a comunicação profissional devido à inflexibilidade ao se tratar os termos, seus conceitos e características. Para a autora, esse reducionismo ocorre devido às crenças idealistas de que os conceitos preexistem às expressões; de que o conhecimento tecno-científico é uniforme e universal; de que a estruturação de um campo de conhecimento independe de seu contexto; de que os âmbitos especializados são neutros e consensuais; e a crença de que o termo normalizado apresenta as características mais significativas para todas as situações.

A TCT, diferentemente da TGT que enaltece as diferenças entre Lingüística e Terminologia, visa considerar as concomitâncias existentes entre ambas disciplinas para

traçar uma teoria de base lingüística que analise as unidades terminológicas tendo em conta a dimensão textual e discursiva dos termos. Em vez de restringir seus fundamentos, a TCT busca uma teoria generalizada levando em consideração que a Terminologia é interdisciplinar (integrando aspectos da Lingüística, das Ciências Cognitivas e das Ciências Sociais) e transdisciplinar (atua em todas as disciplinas, pois não há representação e comunicação eficiente do conhecimento sem uma terminologia).

Segundo Cabré (1999), a TCT não considera os termos como unidades isoladas que constituem seu próprio sistema, mas sim, considera-os como unidades que se incorporam no léxico de um falante no momento em que este adquire o *know how* de especialista por meio da aprendizagem do conhecimento especializado.

Na esfera da Terminologia, ficam evidentes duas vertentes teóricas distintas. A primeira, liderada pelos postulados da TGT, desenvolve o raciocínio sob uma perspectiva formal e prescritiva, e, a segunda, liderada pela Socioterminologia e pela TCT, postula seus fundamentos sob a abordagem funcionalista do uso lingüístico.

5 CONCLUSÕES

Os modelos de representação do conhecimento despontam na atualidade como instrumentos capitais nos processos de tratamento da informação, mais especificamente no processo de análise documentária que é responsável pela representação temática da informação. Tal afirmação leva a inferência de que as linguagens documentárias são primordiais para uma eficiente indexação e recuperação da informação.

Conforme exposto neste trabalho, são diversas as teorias que estão fornecendo subsídios para a fundamentação teórica dos estudos de linguagens documentárias, como a Teoria do Conceito, a Teoria da Classificação Facetada, a TGT, a Socioterminologia, a TCT, Sociolingüística e a Lingüística Funcional. Por ser a linguagem documentária um instrumento de controle terminológico, não é descabido afirmar que sua fundamentação teórica deve ser pautada em teorias cujo objeto seja a unidade terminológica, ou seja, o termo. O fato dessas linguagens documentárias serem empregadas em ambientes de comunicação especializada, seja de um determinado campo do saber seja de uma determinada área profissional, as tornam instrumentos requisitados em situações ‘reais’

de comunicação. Dessa forma, uma teoria de base lingüística, que leve em consideração as variáveis possíveis em um processo de comunicação, e que opere sob uma abordagem funcionalista, pautada no pragmatismo do discurso efetivamente proferido e consumido nos ambientes especializados, é imprescindível para fundamentar as reflexões a cerca das linguagens documentárias.

Como visto, a TGT zela pelo formalismos e pela idealização do discurso especializado, ao passo que a ótica funcionalista da TCT se ocupa com a comunicação especializada realmente utilizada. Isso permite inferir que, além de incorporar e complementar a Teoria do Conceito, acrescentando a idéia de que o ‘conceito’ é parte do ‘termo’, e além de avançar os estudos terminológicos apontando as insuficiências da TGT, a TCT progride rumo a uma teoria que dialoga diretamente com a prática. Sendo assim, aponta-se a TCT como uma teoria capaz de fundamentar os estudos de linguagens documentárias de modo que essas operem como instrumentos pertinentes a situações reais da comunicação especializada.

REFERÊNCIAS

ALVARENGA, Lídia. A Teoria do Conceito revisitada em conexão com ontologias e metadados no contexto das bibliotecas tradicionais e digitais. **DataGramZero: Revista de Ciência da Informação**. v. 2, n. 6, dez/2001. Disponível em <http://www.dgz.org.br/dez01/Art_05.htm>. Acessado em: 17.mai.2005.

_____. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 15, 1º sem. 2003. Disponível em <<http://www.encontros-bibli.ufsc.br/regular.html>> Acessado em: 25.ago.2006.

BARTHES, Roland. **Elementos de semiologia**. Tradução de Izidoro Blinkstein. 3. ed. São Paulo: Cultrix, 1974, 116 p.

BATISTA, Gilda Helena Rocha. Redes de conceitos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 9, n. 1, p. 6-17, jan./jun. 2004. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=55>>. Acessado em: 09.mar.2006.

BUFREM, Leilah Santiago. A relação inescusável entre lingüística e documentação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n.19, 1º sem. 2005. Disponível em <<http://www.encontros-bibli.ufsc.br/sumario.htm>>. Acessado em: 04.mai.2006.

CABRÉ, M. Teresa. **La terminología: representación y comunicación**. Barcelona: Institut Universitari de Lingüística Aplicada, 1999. 369 p.

- CABRÉ, M. Teresa. **La terminología**: teoría, metodología, aplicaciones. Traducción castellana de Carles Tebé. Barcelona: Editorial Antártida/ Empúries, 1993. 526 p.
- CAFÉ, Lígia. Aplicação do modelo de predicação sintático-semântico na construção de linguagens documentárias facetadas. In: CUNHA, Miriam Veira da; SOUZA, Francisco das Chagas de. (Orgs.). **Comunicação, gestão e profissão**: abordagens para o estudo da Ciência da Informação. Belo Horizonte: Autêntica, 2006. 176 p.
- _____. Contribuições da Gramática Funcional da delimitação de segmentos descritores de informação. In: **Organização e representação do conhecimento na perspectiva da ciência da informação**. Georgete Medleg Rodrigues e Ilza Leite Lopes (organizadoras). Brasília: Thesaurus, 2003, p. 118-140. (Estudos Avançados em Ciência da Informação, v.2)
- CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ciência da Informação**, Brasília, v. 33, n. 1, p. 22-32, jan./abril 2004. Disponível em <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=77&layout=html>>. Acessado em: 02.mai.2006.
- CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Organização de domínios de conhecimento e os princípios rangenathianos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 2, p. 150-163, jul./dez. 2003. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=140&layout=abstract>> Acessado em: 19.mai.2006.
- CASTILHO, Ataliba Teixeira de. Um ponto de vista funcional sobre a predicação. **ALFA**: Revista de Linguística. São Paulo, UNESP, v. 38, 1994. p. 75-95.
- CAVALCANTI, Cordelia R. **Indexação & tesouro**: metodologia e técnicas. Brasília: Associação de Bibliotecários do Distrito Federal, 1978. 89 p. (Edição Preliminar).
- CERVANTES, Brígida Maria Nogueira. **Contribuição para a terminologia do processo de inteligência competitiva**: estudo teórico e metodológico. Marília, 2004. 183 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade do Estado de São Paulo, Campus de Marília, 2004.
- CINTRA, Ana Maria et al. **Para entender as linguagens documentárias**. 2 ed. São Paulo: Polis, 2002. 92 p. (Coleção Palavra-Chave, 4).
- DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**. Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.
- GAUDIN, François. **Socioterminologie**: des problemes semantiques aux pratiques institutionnelles. Rouen, França: Publications de L'Université de Rouen, 1993.
- GOMES, Hagar Espanha (Coord.). **Manual de elaboração de tesouros monolíngües**. Brasília: Programa Nacional de Bibliotecas de Instituições de Ensino Superior, 1990. 78 p.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. International Standard. **Terminology work: principles and methods**. 2. ed. 2000. (ISO 704).

_____. International Standard. **Terminology work: vocabulary**. Part 1 – Theory and application. 1. ed. 2000. (ISO 1087-1).

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e recuperação de informação textual: os sintagmas nominais. **Ciência da Informação**, Brasília, v. 25, n. 2, 1996. Disponível em <<http://www.ibict.br/cionline/250296/25029605.pdf>>. Acessado em: 09.mar.2006.

LARA, Marilda Lopez Ginez de. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. **Ciência da Informação**. Brasília, v. 33, n. 2, p. 91-96, maio/ago. 2004a. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=304&layout=html>> Acesso em: 5 abril 2006.

_____. Linguagem documentária e terminologia. **Transinformação**. Campinas. v. 16, n. 3, p. 231-240, set./dez., 2004b. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=8>>. Acesso em: 20 abril 2006.

LIMA, Vânia Mara Alves. **Da classificação do conhecimento científico aos sistemas de recuperação de informação**: enunciação de codificação e enunciação de decodificação da informação documentária. São Paulo, 2004. 148 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, 2004.

MENDONÇA, Ercilia Severina. A lingüística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**., Brasília, v. 29, n. 3, set./dez. 2000, p. 50-70. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=259&layout=abstract>>. Acesso em: 2 abr. 2006.

MOREIRA, Alexandra; ALVARENGA, Lúcia; OLIVEIRA, Alcione de Paiva. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. **DataGramZero** – Revista de Ciência da Informação. Rio de Janeiro, v.5, n.6, dez. 2004. Disponível em: <http://www.dgz.org.br/dez04/F_I_art.htm>. Acesso em: 20 abr.2006.

MOREIRA, Alexandra. **Tesouros e Ontologias**: estudo de definições presentes na literatura das áreas das Ciências da Computação e da Informação, utilizando-se o método analítico-sintético. Belo Horizonte, 2003. 150 f. il. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, 2003. Disponível em <[http://opus.grude.ufmg.br/opus/opusanexos.nsf/4d078acf4b397b3f83256e86004d9d55/915f0db8ceb5bb3583256fb0006a1d5e/\\$FILE/mestrado%20%20Alexandra%20Moreira.pdf](http://opus.grude.ufmg.br/opus/opusanexos.nsf/4d078acf4b397b3f83256e86004d9d55/915f0db8ceb5bb3583256fb0006a1d5e/$FILE/mestrado%20%20Alexandra%20Moreira.pdf)>. Acessado em: 16.abr.2006.

MOREIRO, José; MARZAL, Miguel Angel; BELTRÁN, Pilar. Desarrollo de un Método para la creación de mapas conceptuales. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6, 2003, Belo Horizonte. **Anais...** Belo Horizonte. CD-ROM.

NEVES, Maria Helena de Moura. **A Gramática funcional**. São Paulo: Martins Fontes, 1997. 160 p. (Texto e Linguagem).

PEREIRA, Edmeire Cristina; BUFREM, Leilah Santiago. Princípios de organização e representação de conceitos em linguagens documentárias. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 20, 2º semestre de 2005, p. 21-37. Disponível em <<http://www.encontros-bibli.ufsc.br/sumario.htm>>. Acessado em: 04.mai.2006.

PEZATTI, Erotilde Goreti. Uma abordagem funcionalista da ordem de palavras no português falado. **ALFA: Revista de Linguística**. São Paulo, UNESP, v. 38, 1994. p. 37-56.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Disponível em <<http://www.eci.ufmg.br/pcionline/viewarticle.php?id=5&layout=abstract>>. Acessado em: 09.mai.2007.

SAUSSURE, Ferdinand de. **Curso de lingüística geral**. Tradução de José Victor Adragão. 6. ed. Lisboa: Dom Quixote, 1992, 392 p. ISBN 972-20-0056-X

SOUZA, Renato Rocha. Uma proposta de metodologia para indexação automática utilizando sintagmas nominais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 6, 2005, Florianópolis. **Anais...** Florianópolis. CD-ROM.

TRISTÃO, Ana Maria Delazari; FACHIN, Gleisy Regina Bóries; ALARCON, Orestes Estevam. Sistema de classificação facetada e tesouros: instrumentos para organização do conhecimento. **Ciência da Informação**. Brasília, v. 33, n. 2, p.161/171, maio/ago. 2004. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=120&layout=abstract>>. Acesso em: 26 maio 2006.

Rodrigo de Sales

Mestrando em Ciência da Informação – UFSC

Bacharel em Biblioteconomia – UFSC

Bolsista de Desenvolvimento Tecnológico Industrial do CNPq - Nível G

Membro do Núcleo de Pesquisa em Informática, Literatura e Linguística – NUPILL/UFSC.

rodrigo_biblio@yahoo.com.br

Recebido em: 07/11/2006

Aceito para publicação em: jan. 2007