PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Programa de Pós-Graduação em Letras
Doutorado em Linguística e Língua Portuguesa

A hierarquia semântica de conceitos em tesauros

Denise Queiroz Novaes

Belo Horizonte
2009

Denise Queiroz Novaes

A hierarquia semântica de conceitos em tesauros

Tese apresentada ao Programa de Pós-Graduação em Letras da Pontifícia Universidade Católica de Minas Gerais, como requisito para obtenção do título de Doutor.

Orientador: Prof. Hugo Mari

Belo Horizonte
2009

FICHA CATALOGRÁFICA
Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

Novaes, Denise Queiroz
N935h A hierarquia semântica de conceitos em tesauros / Denise Queiroz Novaes. Belo Horizonte, 2009.
142f.: il.

Orientador: Hugo Mari
Tese (Doutorado) – Pontifícia Universidade Católica de Minas Gerais.
Programa de Pós-Graduação em Letras.

1. Tesauros. 2. Linguística. 3. Conceitos. 4. Categorização (Linguística). 5. Ciência da informação I. Mari, Hugo. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Letras. III. Título.

CDU: 025.44

Denise Queiroz Novaes
A hierarquia semântica de conceitos em tesauros

Tese apresentada ao programa de Pós-Graduação em Letras da Pontifícia Universidade Católica de Minas Gerais, como requisito para obtenção do título de Doutor.

_____
Prof. Dr. Hugo Mari (Orientador) PUC Minas

_____
Profa. Dra. Maria Aparecida Moura – UFMG

_____
Profa. Dra. Helena Magalhães Gramisceli – UFMG

_____
Prof. Dr. Manoel Palhares Moreira – PUC Minas

_____
Prof. Dr. Paulo Henrique Mendes Aguiar – PUC Minas

_____
Profa. Dra Ana Maria Cardoso Pereira – PUC Minas

Belo Horizonte, 10 de novembro de 2009.

Dedicatória e Agradecimentos

Dedico este trabalho especialmente à minha família e aos amigos sinceros, que sempre acreditaram mais em mim do que eu mesma.

Dedico também à incomparável equipe de professores da Ciência da Informação, PUCMINAS, pela oportunidade de convivência durante a qual desenvolvi outros olhares sobre as possibilidades da linguagem.

Aos profissionais de Letras, que compreendem o importante papel que desempenhamos, como profissionais do texto, na cibercultura.

Ao Prof. Hugo Mari, por acreditar na ideia e não desistir de mim.

À PUCMINAS pela oportunidade de formação.

E à ELE, que sempre está comigo e é parte de tudo que faço.

"Ciências Humanas e Ciências Exatas
vibram em consonância e
intensificam o resultado final".
Zamparoli e Camlong

"O pensamento não conhece tudo,
a linguagem não expressa tudo,
o documento não registra tudo".
Paul Otlet

RESUMO

Este trabalho pretende contribuir o debate sobre a interface linguística nos instrumentos tecnológicos de organização da informação. O *corpus* de análise é constituído por tesauros. Para fundamentação da análise foi apresentado o interrelacionamento das áreas afins entre Língua Portuguesa e Linguística com a Ciência da Informação. As contribuições das Teorias Linguísticas, da Semiótica e da Semântica são apresentadas como base para a categorização e a formação dos campos semânticos, bem como para o estabelecimento das noções de conceito, haja vista que o tesauro é uma lista composta de substantivos, hierarquicamente ordenados, de modo a orientar os processos de organização de documentos pelos profissionais de gestão da informação, principalmente, os bibliotecários. Dessa forma, foram também apresentadas as Teorias Documentárias como a Terminologia, Lexicografia, Lexicologia, Terminografia, e também a Linguística Documentária que já demonstra claramente a interseção existente entre as referidas áreas. A noção de conceito foi revista na perspectiva do profissional de Ciência da Informação, bem como os princípios éticos e tecnológicos no trabalho de indexação. A partir da análise de três (3) tesauros, o Tesauro Brasileiro da Educação (Brased), o Tesauro de Ciência da Informação (TCI) e, mais detalhadamente, o Tesauro do Folclore e Cultura Popular Brasileira (Ministério da Cultura), propõe-se a reflexão sobre os critérios linguísticos que permeiam a organização dos dados em um tesauro. Foram analisados os princípios organizadores indicados pelas normas ISO bem como as orientações específicas da área de Ciência da Informação. Conclui-se que critérios linguísticos de ordem semântica podem auxiliar o processo de organização e contribuir para a indexação dos documentos.

Palavras Chave: Tesauro. Linguística. Conceito. Categorização. Hiperonímia. Hiponímia. Meronímia. Holonímia. Ciência da Informação.

ABSTRACT

This work aims to contribute to the debate about the linguistic interface of the technological tools used to organize information. The corpus of this work typically includes thesauri, while the analysis is supported by the interrelation of Linguistic and Portuguese with Information Science. Linguistic, Semiotic and Semantic theories are used as the basis to categorize and create semantic fields, as well as set notions of concept, having in mind that a thesaurus comprises a list of hierarchically ordered nouns aimed at guiding the file organization processes executed by information management professionals, more specifically, librarians. Other document theories are included as well, like Terminology, Lexicography, Lexicology, Terminography and Document Linguistics, which already indicates the intersection between the above mentioned areas. The notion of concept was reviewed under the perspective of the Information Science professional, and the ethical and technological principles involved in the indexation job. From the investigation of three thesauri, the Brazilian Thesaurus of Education (Brased), Information Science Thesaurus (TCI), and in a more detailed the Brazilian Popular Culture and Folklore Thesaurus (Ministry of Culture), this work proposes a reflection about the linguistic criteria which permeate the organization of data within a thesaurus. It also investigates the organizing principles set by the ISO norms and the specific guidelines of Information Science. The conclusion indicates that the semantic linguistic criteria can both help the organization process and contribute to document indexation as well.

Keywords: Thesaurus, Linguistics, Concept, Categorization, Hyperonymy, Hyponymy, Meronymy, Holonymy, Information Science.

RESUMEN

El presente trabajo pretende contribuir con el debate sobre la interfaz linguística en los instrumentos tecnológicos de organización de la información. El *corpus* de análisis está constituido por tesauros. Para fundamentar el análisis, se presentó el interrelacionamiento de las áreas afines entre la Lengua Portuguesa y la Lingüística con las Ciencias de la Información. Las contribuciones de las Teorías Lingüísticas, de la Semiótica y de la Semántica son presentadas como base para la categorización y la formación de los campos semánticos, así como para el establecimiento de las nociones de concepto, teniendo en cuenta que el tesauro es una lista compuesta de sustantivos, jerárquicamente ordenados, de gestión de la información, principalmente, los bibliotecarios. De ese modo, también se presentaron las Teorías Documentarias como la Terminología, Lexicografía, Terminografía y también la Lingüística Documentaria, que ya revela claramente la intersección existente entre las referidas áreas. La noción de concepto fue revista en la perspectiva del profesional de Ciencias de la Información, así como los principios éticos y tecnológicos en el trabajo de indexación. A partir del análisis de tres (3) tesauros, el Tesauro Brasileño de la Educación (Brased), el Tesauro de Ciencias de la Información (TCI) y, más detalladamente, el Tesauro del Folclore y Cultura Popular Brasileña (Ministerio de Cultura), se propone la reflexión sobre los criterios lingüísticos que permean la organización de los datos en un tesauro. Fueron analizados los principios organizadores indicados por las normas ISO así como las orientaciones específicas del área de Ciencias de la Información. Se concluye que criterios lingüísticos de orden semántico pueden auxiliar el proceso de organización y contribuir para la indexación de los documentos.

Palabras Clave: Tesauro; Lingüística; Concepto; Categorización; Hiperonimia; Hiponimia; Meronimia; Holonimia; Ciencias de la Información.

LISTA DE FIGURAS

Figura 01 - Tela de busca no Tesauro Brasileiro da Educação (Brased)..... 24
Figura 02 - Matriz Conceitual do Tesauro Brasileiro da Educação (Brased)..... 25
Figura 03 - Triângulo Semiótico..... 41
Figura 04 - Triângulo Conceitual de Dahlber..... 83
Figura 05 - Áreas de relacionamento - Tesauro em Ciência da Informação – TCI..... 113
Figura 06 - Visão geral do Tesauro de Folclore e Cultura Popular Brasileira – MC..... 120
Figura 07 - Subitem expandido: Literatura Popular - Tesauro de Folclore e Cultura Popular Brasileira - MC..... 122
Figura 08 - Subitem expandido: (pela forma de circulação) - Tesauro de Folclore e Cultura Popular Brasileira – MC..... 122
Figura 09 - Orientações sobre o termo Literatura Popular - Tesauro de Folclore e Cultura Popular Brasileira – MC..... 124
Figura 10 - Renque expandido de Literatura Popular - Tesauro de Folclore e Cultura Popular Brasileira – MC..... 125

LISTA DE QUADROS

Quadro 01 – Subáreas do tesauro do Tesauro Brasileiro da Educação (Brased)..... 26
Quadro 02 - Níveis de categorias, segundo Rosch..... 55
Quadro 03 - Linguagem natural: vantagens e desvantagens..... 90
Quadro 04 - Vocabulário controlado: vantagens e desvantagens..... 91
Quadro 05 - Hierarquia das classes em cadeias - Tesauro Brasileiro da Educação (Brased).....110-111
Quadro 06 - Hierarquia da classe de Arquivologia - Tesauro em Ciência da Informação – TCI..... 115

LISTA DE ABREVIATURAS

ABNT Associação Brasileira de Normas Técnicas

Brased Tesauro Brasileiro da Educação

CDD Classificação Decimal de Dewey

CDU Classificação Decimal Universal

CI Ciência da Informação

DRT *Discourse Representation Theory*

FST *Fuzzy Set Theory*

IA Inteligência Artificial

INEP Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

ISO *International Organization for Standardization*

LC Linguagem Controlada

LN Linguagem Natural

NBR Norma Brasileira

PCNEM Parâmetros Curriculares Nacionais para o Ensino Médio

PLN Processamento de Língua Natural

RTM Teoria da Representação Mental

TC Teoria dos Conjuntos

TCI Tesauro de Ciência da Informação

TCT Teoria Comunicativa da Terminologia

TGT Teoria Geral da Terminologia

TI Tecnologias da Informação

TP Teoria dos Protótipos

TST Teoria Sociocognitiva da Terminologia ou Socioontologia

Unesco Organização das Nações Unidas para Educação, Ciência e Cultura

SUMÁRIO

1 INTRODUÇÃO 132
1.1 Resumo dos objetivos 20
1.2 Hipóteses preliminares 21

2 INTERSEÇÃO: LINGUÍSTICA E GESTÃO DA INFORMAÇÃO..... 22

3 TEORIAS LINGUÍSTICAS 31
3.1 Semiótica e Semântica 31
3.2 Breve percurso das abordagens semânticas 33
3.3 A Semântica Formal no universo das discussões semânticas 39
3.4 Campos Semânticos 47
3.5 Categorização 49
3.6 Conceito 58
3.7 Abordagens complementares sobre categorização 64

4 TEORIAS DOCUMENTÁRIAS..... 68
4.1 Terminologia 698
4.2 Lexicologia..... 754
4.3 Lexicografia..... 765
4.4 Terminografia 787
4.5 Linguística Documentária..... 77
4.6 Visão de conceito na perspectiva da CI e princípios organizadores dos tesauros 821
4.7 Princípios éticos e garantias no trabalho do indexador..... .87

5 ANÁLISE DA HIERARQUIA SEMÂNTICA DE TESAUROS..... 94
5.1 Metodologia de organização de tesauros. 954
5.2 Por uma análise linguística de critérios organizadores..... 108
5. 2. 1 Metodologia empregada..... 108
5. 2 2 Análise da hierarquia semântica do Thesaurus Brasileiro da Educação (Brased) - INEP..... 109
5. 2. 3 Análise da hierarquia semântica do Tesauro em Ciência da Informação - TCI..... 113
5. 2. 4 Análise da hierarquia semântica do Tesauro do Folclore e Cultura Popular Brasileira..... 110

6 REFLEXÕES FINAIS 130

REFERÊNCIAS..... 136

1 INTRODUÇÃO

Há algum tempo, Bakhtin, filósofo russo, já acenava que as ciências exatas pensam o homem *fora* do texto e as ciências humanas veem o homem *no texto*, uma vez que já tomava a língua como um fato social originado da necessidade de comunicação. Mas tal distinção postulava uma dicotomia que não encontra o mesmo respaldo na atualidade. Talvez fosse mais adequado dizer que as ciências humanas tentam se estabilizar a partir das ciências naturais, na tentativa de registrar essa relação do homem com o mundo. Dessa forma as ciências humanas instalam um necessário desequilíbrio, na problematização da própria realidade, movimentando a pesquisa e a construção de novos conhecimentos; no entanto, muitas vezes necessitam de instrumentos de apoio das ciências exatas que propiciem levantamentos de dados e o ajuste ao formalismo rigoroso dos registros científicos. Então, podemos afirmar que ciências exatas e humanas hoje dialogam num mesmo texto, propiciando o uso de recursos que tentam acompanhar a infinitude do pensamento humano.

Na atualidade, grande parte das diferentes formas de organização de bancos de dados na informática visa a armazenar material - muitos deles, linguísticos - e disponibilizá-lo ao usuário/autor, a fim de que tenha uma gama de recursos para suas produções na constituição do seu discurso. Também para o estudioso dos gêneros textuais, conhecer as formas de organização do léxico certamente é um elemento importante (poderíamos afirmar que sempre foi), pois sabemos que as escolhas lexicais demonstram as formações ideológicas que sustentam os objetivos enunciativos de seus autores, bem como as inovações e estéticas que os caracterizam.

Essa relação, Informática e Letras, impõe, talvez, um novo papel ao linguista. Além de ser um potencial utilizador dos recursos da informática na organização e manipulação de dados linguísticos, como na análise de regularidades lexicais e sintáticas, passa a ser cobrada uma contribuição maior do ponto de vista linguístico para a implementação de sistemas informatizados mais ajustados às nossas necessidades enunciativas. Acreditamos, a partir de uma compreensão mais competente da estrutura e do funcionamento da linguagem, e do conhecimento de suas possibilidades, que seja possível um melhor ajuste de instrumentos de busca e de editoração, como o *Google* e o *Word*, por exemplo, às necessidades de seu usuário, para a obtenção, representação e uso do conhecimento linguístico através do computador.

Partindo do pressuposto de que a construção social de conhecimentos passa necessariamente pelo registro em forma de gêneros textuais, seja ele mediado por quaisquer

formas de representação - oral, escrita, gestual, táctil, sonora e/ou visual -, e comunicado por meio de uma das diversas possibilidades midiáticas, o profissional de Letras, aquele que pensa sobre as mais variadas formas de organizar as ideias em um texto passível de ser compartilhado, não pode mais furtar-se a ter um conhecimento sobre as soluções propostas pela informática.

Os editores de texto Microsoft Word, BROffice Writter, Adobe Acrobat e PageMaker, entre outros, trouxeram novos recursos ao ato de redigir, ampliando as possibilidades de produção, sem prejuízo algum do ato de criação. Tais programas permitem um reestruturar constante na produção, com um sistema de busca e opções de substituição de palavras, ferramentas de revisão e correção, sumarização, comparação, mesclagem de textos, e, ainda, com possibilidade de gravar todas as versões, que podem funcionar como rascunhos até se chegar à versão final. Além disso, as possíveis modificações efetuadas não danificam a apresentação visual como os indesejáveis rabiscos de cancelamento ou as infindáveis folhas inutilizadas, mantendo o texto num padrão formatado, - se este for o desejo do autor. Pois, se o produtor quiser, também estão à sua disposição inúmeros recursos de tipos de letras, cores, molduras, bem como de associação de imagens, sons, como já acontece nos belíssimos videopoemas. Isso quebra, de certa forma, até mesmo a resistência da perspectiva literária, que muitas vezes lamenta a "perda" do perfil do manuscrito, pois também é possível registrar imagens em letras cursivas, preservando-se o desenho das letras que tanto personalizam o autor e mantendo esses documentos por tempo indeterminado. Podemos afirmar que a capacidade imaginativa do homem fica cada vez mais livre para criar, menos preocupada com parte das tarefas repetitivas, como as de formatação, resolvidas previamente pelos editores de texto.

Consequentemente, as formas de ler também foram ampliadas, textos e hipertextos[1] circulam na internet, promovendo leituras lineares e não-lineares, mas igualmente ricas. Mecanismos de busca como o *Google, Cadê, Quatrocantos*, entre outros, trazem uma gama inesgotável de gêneros textuais que contém todo tipo de informação, exigindo cada vez mais a formação de um leitor crítico. Tantas possibilidades de trocas de informação estimulam o leitor a participar de produções de textos coletivas como nos fóruns de discussão, a manifestar seu ponto de vista sobre uma publicação em um *site*, uma vez que as possibilidades de

[1] Hipertexto, no seu nível mais básico, é um manipulador de bases de dados, que permite ligar páginas informativas usando *links* que os associam. Num nível mais amplo, é um ambiente de software em que se realiza trabalho colaborativo, comunicação, e aquisição de conhecimentos. As características desse software estimulam o cérebro para armazenar e recuperar informação, fazendo uso de *links* para um acesso rápido e intuitivo na produção de sentido. O surgimento dessa nova modalidade textual pode ser atribuído ao surgimento dos meios de escrita digitais, tornando possível a criação dessa rede de informações que pode mesclar imagens sons e texto.

comunicação tornam-se cada vez mais acessíveis, sem levar em consideração as possibilidades de inclusão social, como o programa DOSVOX, desenvolvido para deficientes visuais, que "lê" o texto para o usuário, entre outros instrumentos que registram os textos pelo tom de voz do falante, viabilizando, também, a produção textual por portadores de deficiência motora nas mãos.

Como o profissional de Letras, um linguista, pode ficar à margem desse processo?

Portanto, não há dúvidas de que o avanço da informática abriu novas e ilimitadas possibilidades em diversas áreas do conhecimento humano. O efeito mais imediato pôde ser observado nas ciências exatas na pesquisa de obtenção de dados de natureza numérica e estatística, propiciando comparações e permitindo leituras dessas informações em um lapso de tempo antes impossível de serem realizadas. Mas sabe-se que várias outras áreas beneficiaram-se dos novos instrumentos de tecnologias da informação. Dos estudos de Letras e Linguística, tanto no registro do texto falado quanto no do texto escrito, aos estudos de literatura comparada, aos de tradução, bem como aos do ensino de línguas, hoje é inegável a contribuição da informática para constituir variadas formas de dizer, o que abre novas perspectivas de análise do discurso. Hoje temos ampliadas as possibilidades de compreensão, haja vista o numeroso índice de estudos sobre o hipertexto.

Inicialmente, houve tentativas restritas em prol da sintonia real entre os recursos que a informática propicia e os estudos tradicionais da linguagem. Tal intenção está refletida nos programas para busca, organização e quantificação de dados linguísticos, que são utilizados nas análises quantitativas e nos permitem diferentes leituras sobre os dados obtidos (excel, minitab, corpógrafo, motores de buscas diversos...). Já é comprovado que grande parte dos estudos linguísticos utiliza programas específicos, para efetuar análises mais profundas e controladas, de maneira objetiva, científica e interdisciplinar, além de análises da tessitura lexical, textual e discursiva que contribuem para refinar o tratamento e análise de línguas naturais. Podemos citar: analisadores sintáticos, sumarizadores, tradutores automáticos, corretores ortográficos, dentre muitos, além dos editores de textos, que oferecem recursos cada vez mais diversos.

Diante disso, podemos afirmar que as tecnologias da informação provocaram inegáveis mudanças nos processos tradicionais de produção textual e na comunicação. Na atualidade, a produção e disseminação eletrônica de documentos e a recuperação da informação concentram-se, de maneira geral, na extração do conteúdo diretamente dos textos completos dos documentos digitalizados, muitos deles já disponibilizados eletronicamente na *web*. Essas mudanças afetaram grande parte dos processos de produção de textos bem como

do tratamento dessas informações para fins de organização social do conhecimento, principalmente na indexação[2]. Uma indexação é considerada de qualidade quando consegue equilibrar a recuperação da informação a um alto nível de precisão ao atendimento da busca do usuário/leitor. Este tem sido o grande desafio para a qualidade dos motores de busca de informações na *web*. Alguns utilizam como critério geral um número estabelecido de palavras para a indexação de textos de *sites*, o que ocorre automaticamente quando detectado o número limite; em outros ambientes, os critérios podem variar por ordenação alfabética, como nas bibliotecas digitais, ou ainda por campos semânticos nas ontologias[3], como se pretende aqui discutir.

Podemos afirmar que a informação, base do conhecimento, seria o conjunto de conhecimentos reunidos sobre determinado assunto. Rafael Capurro e Birger Hjorland, estudiosos que discutem a amplitude do termo informação na sociedade contemporânea, demonstram as implicações interdisciplinares do termo em áreas como a biologia, psicologia, filosofia, entre outras, e chegam a concluir pela impossibilidade de uma definição que esteja dissociada de um dado domínio:

> A informação pode ser identificada, descrita e representada em sistemas de informação para diferentes domínios de conhecimento. É claro que surgem problemas para determinar se uma coisa é informativa ou não para determinado domínio. Alguns domínios têm alto grau de consenso e critérios de relevância explícitos. Outros domínios têm paradigmas diferentes, conflitantes, cada um contendo sua própria visão, mais ou menos implícita, da informatividade dos diferentes tipos de fontes de informação. (CAPURRO; HOJRLAND, 2007, p. 192).

Optamos aqui por nos filiarmos a essa perspectiva, pois o objetivo deste estudo é discutir exatamente tais formas de organização de conteúdos, refletindo sobre as contribuições linguísticas nesse processo humano de inserção de informações nas ferramentas informatizadas. Acreditamos na dimensão da organização por campos semânticos que fundamentam os principais requisitos de seleção e hierarquização de conceitos, levando-se em conta os critérios linguísticos (a saber) empregados para o membramento lexical.

Continuam os autores:

---

[2] Ato de fazer índices para documentos ou de pôr em ordem alfabética qualquer série de palavras ou frases destinadas a auxiliar a localização de informações específicas.

[3] Parte da Filosofia que se ocupa do ser enquanto ser, ou seja, do ser concebido na sua totalidade e na sua universalidade. Tomamos aqui o conceito aplicado às Ciências da Computação e da Informação, em que se entende uma ontologia como um modelo de dados que representa um conjunto de conceitos dentro de um domínio e os relacionamentos entre estes. Uma ontologia é utilizada para realizar inferências sobre os objetos de um determinado domínio.

> [..] Quando representamos dados em nossos sistemas de informação, o fazemos a fim de dar suporte a certas atividades humanas. Não deveríamos simplesmente considerar nossas representações como objetivas, porque isto implica que nunca especificamos completamente as suposições teóricas, sociais e históricas com base nas quais agimos. Todos os tipos de sistemas de informação têm políticas e objetivos mais ou menos explícitos. O que consideramos como informação deveria também ser um reflexo da função social do sistema de informação. (CAPURRO; HOJRLAND, 2007, p. 187). (Grifo nosso).

Sendo assim, quando esses dados são contextualizados e adquirem valor para o leitor, transformam-se em informações que permitem a composição de um conhecimento. Em outras palavras, se o leitor processa esses dados e os considera relevantes, torna-se informação; essa informação, então, será inferida e se tornará conhecimento que, com o tempo se tornará sabedoria. E através dos processos cognitivos realizados pela mente humana, essa informação captada é armazenada na memória[4], a qual permite recuperá-la quando há necessidade. A partir dessa percepção e de posterior categorização da realidade através de conceitos, é possível transmitir os conhecimentos (pela fala, escrita ou quaisquer meios, dentre eles os informatizados), já que foram explicitados e são passíveis de serem comunicados. A compreensão desse processo cognitivo permite o ato de conceituar para entender o comportamento humano, diante da nova informação que a pessoa recebe e diante da informação apreendida que se deseja disseminar.

Para Shannon e Weaver (1971), informação é a medida da liberdade do emissor ao escolher a mensagem a ser veiculada. Essa visão promove a dissociação entre informação e significado, pois a informação é considerada sob o aspecto quantitativo e avaliada a partir da transmissão de mensagens em um sistema de comunicação. Esse conceito imprime um caráter estatístico, pois os autores defendem que informação é a medida da liberdade do emissor. Dessa forma, um instrumento de comunicação ideal é aquele em que a mensagem escolhida e transmitida é exata ou, aproximadamente, a mesma recebida e reconstruída pelo receptor. Evitar possíveis ruídos nesse processo é o desafio de todas as ciências que estudam a comunicação.

É importante ressaltar que os conhecimentos são idiossincráticos, portanto o que podemos fazer é expô-los através da informação. Dessa forma, a utilização de sistemas tecnológicos de organização de conceitos, que podem ser representados através de estruturas conceituais, em instrumentos como o tesauro, torna-se de grande valor para as atividades de

[4] Entende-se aqui a memória (humana) como a faculdade cognitiva de reter idéias, imagens, expressões e conhecimentos já vivenciados e ativados pelas lembranças ou reminiscências; seria uma forma de armazenamento mental de representações do passado. Dessa forma, é a retenção de conhecimentos aprendidos.

explicitação, armazenamento e disseminação dos conhecimentos produzidos pela sociedade e retidos pelos indivíduos.

Esses instrumentos tecnológicos de representação do conhecimento em sistemas de recuperação da informação, conhecidos como "memórias documentárias", são construtos do conhecimento gerado pela sociedade que reúnem cadeias de representação presentes na dinâmica social. Segundo Dodebei (2002), tais modelos derivam da necessidade humana de entender a realidade e são representações simplificadas e inteligíveis do mundo. Nesse sentido, a linguagem documentária[5] cumpre sua função referencial, viabilizando a busca de conhecimento público, que deve estar organizado para socialização e uso.

Cabe-nos esclarecer que a representação documentária é o produto da leitura e análise de um texto, na qual o profissional gestor de informação identifica os conteúdos pertinentes aos propósitos e às características do sistema em que será inserida tal informação. Para o registro de uma memória documentária, deve-se levar em consideração critérios como a padronização e a univocidade, de modo a intermediar o processo comunicativo entre o usuário e a máquina. Essa memória documentária pode ser entendida, então, como o resultado de um processo de ressignificação do analista em função da geração de um documento a ser veiculado por um sistema de informação, no nosso caso, o tesauro. Dessa forma, a palavra utilizada na linguagem natural, e, posteriormente, transformada por uma linguagem documentária (organizada sob critérios acima descritos), é legitimada para representar conceitos; material este frequentemente utilizado pelos autores em suas produções.

No entanto, as características próprias das línguas naturais, principalmente a polissemia, a homonímia e a sinonímia, além da função sintática, interferem significativamente no resultado da busca de informações; ou seja, não garantem que a informação recuperada seja a mais próxima possível da informação desejada, dentro do mesmo campo semântico. Além disso, a representação da informação para a recuperação posterior por múltiplos usuários implica a descrição do conhecimento em forma de síntese para auxiliar na pesquisa de fontes.

Exemplo prático dessa representação pode ser observado no número de ocorrências resultantes de palavras inseridas nos motores de busca como o *Google*. Em uma busca pela palavra "língua", imediatamente, são apresentadas 95.900.000 possibilidades; entre as 10 (dez) primeiras apresentadas já se notam os problemas de polissemia, pois "língua" é

[5] É a linguagem artificial utilizada (nos sistemas documentários) para indexação, armazenamento e recuperação de informações.

inicialmente apresentada como órgão muscular; depois como objeto de gramática; sistema de comunicação; expressões como "papas na língua"; língua de tamanduá; e em quinto lugar aparece língua portuguesa. Essa recuperação de informação exige um refinamento, a fim de se restringir os resultados para atender às necessidades do usuário. Dessa forma, o *Google* inclui na primeira página de amostra a opção: "Pesquisas relacionadas a: língua: linguagem, doenças da língua, anatomia da língua, língua e linguagem, língua humana, língua órgão, língua da boca, língua e fala". Porém, ainda deixa em aberto uma infinidade de possibilidades, que necessitam, talvez, de critérios de hierarquização semântica para sua ordenação. Soluções para amenizar os problemas de localização da informação, da precisão na recuperação da informação desejada podem surgir da pesquisa linguística, pois as abordagens tradicionalistas de tratamento e recuperação de documentos digitais, utilizando-se de embasamentos matemático-estatísticos, não têm sido suficientes para abarcar o enorme volume de informação disponibilizado. E, como os aparatos tecnológicos da informática operam através da memória constituída pelos bancos de dados e por um conjunto de regras linguísticas traduzidas pelas linguagens de programação, tem sido cada vez mais frequentes - e necessárias - as intersecções entre a informática e os estudos de linguagem natural. [6]

A questão que se interpõe nessa *tradução* do conhecimento humano expresso em linguagem natural para linguagem documentária e/ou informatizada confronta-se, principalmente, com questões semânticas, no tocante, por exemplo, aos problemas de ambiguidade, pois a linguagem natural, em si, já é sintetizadora do pensamento (humano), na medida em que, qualquer que seja a forma de representação, nunca será igual à realidade. Na perspectiva da semiótica, o ser humano, em sua necessidade de representar, procura estabelecer signos que exerçam tal função. Porém, como usuários de uma linguagem pré-estabelecida socialmente, nem sempre conhecemos as razões daquela constituição semântica atribuída a determinado signo. Como interlocutores nativos, sabemos sobre alguns usos e, na interlocução, tomamos conhecimento de outros, bem como podemos contribuir para a constituição de novas significações. No entanto, torna-se necessária uma parametrização para nortear as formas de organização. Portanto, considera-se que quaisquer tipos de síntese, entre elas a da palavra, para o acesso tecnológico à informação são formas ainda mais acentuadas de redução de significado. Essa cadeia de reduções semânticas, segundo Dodebei (2002), faz

[6] Na Europa, já temos consistentes trabalhos na área de PLN (Processamento de Língua Natural) como o Projeto da Linguateca, em Portugal (com pólos na Noruega e Dinamarca), que tratam do processamento computacional da língua portuguesa. No Brasil, já temos alguns centros de linguística computacional, como o *nilc*, da USP de São Carlos, que, inclusive, participa da composição do banco de dados de língua portuguesa dos corretores ortográficos como o do *Word*.

parte da base da memória documentária para que a organização do conhecimento em unidades mínimas seja passível de decodificação e utilização.

Nesse contexto, o estudo da estrutura hierárquica de conceitos é uma tentativa de reproduzir o processo de organização social do conhecimento, que certamente traz consigo questões culturais. Dada a complexidade desse processo de representação da memória documentária, há que se considerar a não-homogeneidade de interpretação de um conceito, que pode variar segundo diferentes abordagens (linguísticas e documentárias) e cuja existência só se concretiza no indivíduo. A perspectiva social da representação perpassa necessariamente pela discussão sobre semântica, pluralidade de significados e produção de sentido, na ótica linguística, bem como deve atender aos requisitos de funcionalidade e economia das representações. Conforme reforça Dodebei, é preciso "reduzir informação para obter conhecimento". (2002, p.36).

Estabelecendo campos semânticos e permitindo as relações entre os conceitos, as ferramentas de organização de informação nos permitem usufruir do conhecimento configurado nessas memórias documentárias, propiciando-nos a construção de novas estruturações, novos *projetos de dizer*, lembrando Ingedore Koch (2003). Segundo a autora, no processamento textual, operamos com dois sistemas de conhecimento: a memória semântica, aquela que abrange o conhecimento categorial sobre o mundo e representada pelo léxico de uma língua; e pela memória episódica que armazena nossas vivências pessoais.

Embora a memória não seja objeto de discussão neste trabalho, seria relevante destacar que Aristóteles já distinguia a memória propriamente dita, a *mneme*, como a faculdade de conservar o passado; da reminiscência, a *mamnesi*, faculdade de invocar voluntariamente o passado. Para o filósofo, as impressões sensoriais, fonte básica de conhecimento, são percepções tratadas pela faculdade da imaginação, formando assim imagens que se tornam material para a faculdade intelectual. Portanto, a memória humana é entendida como retenção de um dado conhecimento, mas também como ativadora da imaginação e das capacidades de interpretação, problematização e reinvenção, as quais atuam sobre o que é recordado pelo sujeito.

Também Vigotsky (2005) define a memória humana como a capacidade que o indivíduo tem de conservar e reproduzir as apropriações da realidade. De acordo com o autor, há dois tipos de memória: a *elementar*, que é caracterizada por ser imediata e involuntária; a *superior*, mediata (por instrumentos ou signos) e intencional. Vigotsky afirma que os instrumentos culturais adquiridos desde a infância permitem-nos organizar nossas funções psicológicas, tendo como base o uso de conceitos e abstrações. Então, boa parte dos

conteúdos recordados por um adulto é mediada por signos internos ou externos, pois é pelas relações sociais que o homem compartilha pensamentos, informações, idéias.

De forma similar, podemos concluir que uma memória documentária tem o papel de ativar e ampliar as possibilidades de produção de um usuário. Diante disso, os estudos lexicais adquirem um aspecto relevante nesta pesquisa, pois nosso objetivo é discutir os critérios que norteiam a organização humana de conceitos, sabendo-se que eles se constituem como unidades cognitivas de ativação da memória, princípio este amplamente utilizado nos sistemas informatizados. Um dos desafios da informática é situar o signo (palavra, imagem, som) ideal para estimular o usuário das ferramentas a encontrar o que busca e saber como utilizar o que encontrou.

Há que se ressaltar que o termo "memória", na informática, é utilizado para denotar qualquer forma de armazenamento eletrônico, rápido e temporário; em síntese, seria um local de armazenamento de dados. A informação mantida na memória de um computador, por exemplo, permite, em tese, uma busca mais rápida das informações. Partindo do pressuposto de que, hoje, dado o volume de produção e disseminação de informações, temos a nossa disposição ferramentas que podem agilizar o processo de pesquisa (independentemente de estar arquivada em um computador) e permitir o contato com uma infinita variedade de fontes de informação, tornam-se relevantes as reflexões sobre o processamento linguístico empregado pelo profissional que alimenta o banco de dados, sobre como faz sua escolhas lexicais, especialmente na organização hierárquica de conceitos que, certamente, será um fator de interferência para a estruturação do conhecimento, e, posteriormente, da utilização eficiente desse conhecimento pelo produtor/usuário. Numa linguagem controlada (documentária), é importante que as palavras no vocabulário sejam "normalizadas", para assegurar a coincidência entre as formas das palavras usadas na indexação das informações com a busca do usuário.

1.1 Resumo dos objetivos

Conforme se anuncia no título deste trabalho, o objeto desta pesquisa é o estudo da hierarquização semântica de conceitos, tomando como *corpus* de análise um instrumento de organização de informação, o tesauro.

Em consonância com as questões descritas na introdução, pretendemos demonstrar as relações entre as áreas de conhecimento de Letras e de Ciência da Informação, especificamente no que tange ao processo humano empregado na organização de conceitos

em memórias documentárias como o tesauro, bem como as implicações desse processo no resultado final. Pretende-se, portanto,

- Discorrer sobre a interface linguística no processo de gestão de informações.
- Apresentar e discutir a contribuição das teorias semânticas no processo de hierarquização de conceitos.
- Apresentar e discutir as interferências das teorias documentárias no estabelecimento dos parâmetros de organização.
- Mapear e discutir os critérios linguísticos que norteiam a organização humana de conceitos.

Para tanto, apresentaremos a análise de partes de três tesauros: o Tesauro Brasileiro da Educação, do INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira, com o objetivo de mostrar o universo de pesquisa; a seguir, será apresentado o Tesauro em Ciência da Informação, parte da Tese "Ambiente para geração e manutenção semi-automática de tesauros", defendida pelo Prof. Dr. Manuel Palhares, Professor de Banco de Dados nos Cursos de Ciência da Computação e Ciência da Informação da PUCMINAS, a fim de comparar formatos de organização, pois se trata de um tesauro gerado a partir de outros tesauros; e, finalmente, será analisado mais detalhadamente o Tesauro do Folclore e Cultura Popular Brasileira, que teve a coordenação teórico-metodológica da Profª Drª Maria Luiza de Almeida Campos, Professora de Ciência da Informação/UFF/RJ.

Esse tesauro, em particular, possui um modo de organização peculiar, que permite uma visão completa dos critérios organizadores, com amplo uso de variados recursos além dos textos, como vídeos e imagens, e possibilita uma análise que contempla, ao lado do estudo da hierarquia semântica, outras questões caras à Letras que remetem às discussões sobre gêneros textuais nesse tipo de organização, estas, talvez objeto para discussões futuras.

1.2 Hipóteses preliminares

Algumas hipóteses podem ser formuladas, levando-se em conta o grau de interseção das áreas de Letras e de Ciência da Informação na atualidade. Dentre elas, podemos indicar:

- A falta de clareza dos critérios linguísticos empregados na hierarquização pode gerar ordenações contraditórias de uma mesma palavra.
- É possível empreender uma análise terminológica sem levar em consideração uma análise linguística.

- A Norma ISO, instrumento parametrizador na constituição de tesauros, originalmente constituída na língua inglesa, pode induzir a erros na hierarquização de palavras de língua portuguesa.

2 INTERSEÇÃO: LINGUÍSTICA E GESTÃO DA INFORMAÇÃO

Na Era da Informação[7], as habilidades mais requisitadas estão relacionadas à organização do conhecimento. Portanto, é necessário refletir sobre o ciclo de sua produção, para recuperar o processamento da rede significativa que o compõe. O primeiro momento é o da seleção e coleta de dados que são organizados para dar sentido à informação. Após se considerar as informações existentes sobre determinado tema, é possível estruturar uma visão sobre dado conhecimento que, ao estabelecer relações com outros conhecimentos, poderá gerar novos dados, novas informações e novas abordagens, configurando o ciclo da produção científica e literária, entre outras.

No entanto, a tecnologia cria novas necessidades e altera alguns dos paradigmas de produção de textos já estabelecidos ao longo do tempo, tornando a informação cada vez mais atrelada ao objeto físico que a contém (desde a perspectiva dos gêneros textuais, como a dos processos e instrumentos comunicativos como a internet), dado o índice de abstração que envolve a atribuição de sentidos. Em outras palavras, toda organização de bancos de dados envolve, necessariamente, uma estratégia de organização hierárquico-semântica para ordenar os conhecimentos de modo a facilitar a busca do usuário. Parte-se, geralmente, de uma perspectiva de conceitos gerais para os mais particulares, numa relação hiperonímia (classe mais geral) e de hiponímia (classe mais específica). Uma das propostas deste trabalho é exatamente discutir esses critérios, seu emprego, ou, talvez, a carência deles. Questionamos, ainda, se tais formas de organização atendem às nossas buscas.

As estruturas hierárquicas em bancos de dados convencionais, como os armazenamentos físicos de documentos nos arquivos e bibliotecas tradicionais, convivem com a produção intensa de um mundo de informação dinâmica, em permanente mudança e expansão, o que exige dos operadores uma integração, continuamente atualizada, com o vocabulário dos usuários. As contribuições linguísticas sobre o processo natural de composição lexical, bem como a tentativa de estabilização semântica de um vocábulo - tendo em vista a relatividade desse conceito -, são fundamentais para: a articulação e hierarquização adequada de conceitos, para as questões referentes aos ajustes aos contextos enunciativos e

[7] O século XX foi denominado a Era da Informação. Associado a isso, temos testemunhado vários avanços tecnológicos em diversas áreas. Dentre elas, duas que têm causado significativo impacto sobre o *modus vivendi* das pessoas neste século são Computação e Telecomunicações. Nesse cenário de avanços tecnológicos, deparamo-nos com uma carga de informações cada vez maior.

para a clareza dos critérios de inclusão de um vocábulo num determinado campo semântico (membramento lexical). Tais reflexões tornam-se diferenciais para os indicadores de eficiência de uma dada ferramenta de informática e podem resolver muitos dos problemas de recuperação de informação, haja vista que não seria muito produtivo armazenar, se não fôssemos capazes de recuperar e utilizar tal informação.

Considerando, portanto, a informática como fator importante nesse processo de recuperação/constituição da informação, temos uma enorme gama de instrumentos utilizados para representar o conhecimento de uma dada área do saber. Um deles é denominado Tesauro[8], linguagem documentária que apresenta algumas peculiaridades, pois sua hierarquia de assuntos possui uma relação associativa e sua estrutura não se baseia em conceito de palavras simplesmente, mas, sim, de termos conceituados e relacionados. Segundo Cavalcanti,

> Tesauro é uma lista estruturada de termos associados empregada por analistas de informação e indexadores, para descrever um documento com a desejada especificidade, em nível de entrada, e para permitir aos pesquisadores a recuperação da informação que procura. (CAVALCANTI, 1978, p. 27).

O uso dessas linguagens documentárias (organizacionais), em bases terminológicas, como instrumento de representação/recuperação - como o tesauro-, permite a comunicação entre o documento, a informação e o usuário, uma vez que esse processo ocorre através da linguagem. Situa-se, nesse contexto, o momento da tradução, ou talvez, retextualização do conhecimento humano em linguagem natural para a linguagem documentária, com o objetivo de viabilizar o acesso à informação. Entretanto, toda essa informação registrada necessita de uma interrelação dos conceitos emergentes com a consequente alteração do significado dos conceitos existentes na área, e de suas relações com as outras áreas – princípio que norteia nossa gramática descritiva. Portanto, percebe-se que um conjunto de termos estruturados num tesauro, numa linguagem específica na sua forma e conteúdo, pode permitir a geração de novos termos e/ou alterar significados já existentes, de modo a garantir ao pesquisador, no processo de coleta e seleção da informação e do documento, encontrar a resposta de que necessita numa linguagem padronizada. Nessa perspectiva, a identificação de conceitos estará relacionada à sistematização e representação de uma determinada área do conhecimento, delineada pelo campo semântico, haja vista que um tesauro sempre é produzido para uma área específica de conhecimento, como da educação (tema do tesauro analisado mais a frente), da cadeia alimentícia, da segurança pública, entre infinitas áreas. A Figura 01, a seguir,

[8] Tesauro (latim = *thesaurus*, grego = *thesaurós*) significa tesouro ou armazenamento ou repositório, que utiliza a palavra para indicar um acervo ordenado de informações e conhecimentos.

exemplifica o resultado de busca da expressão *academia de educação física* no Tesauro Brasileiro da Educação (Brased), do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP):

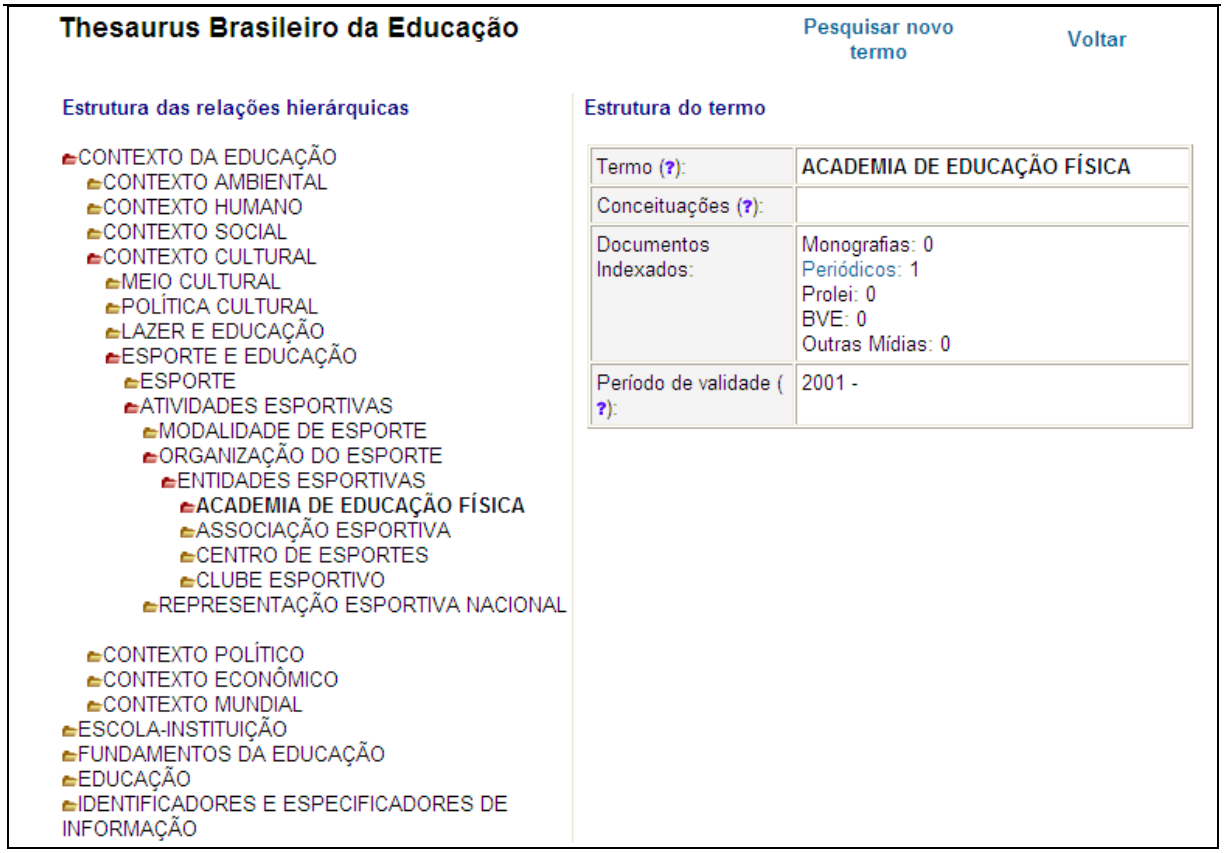

Figura 01 – Tela de busca no Tesauro Brasileiro da Educação
Fonte: INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

A área contemplada por esse tesauro é a educação. Portanto, todas as associações estão restritas a tal área, como se pode notar nas facetas (ou opções) de busca, marcadas pela forma de ordenação (recuada ou não) que caracteriza a inclusão do termo. Numa visão panorâmica dessa ferramenta, observamos a hierarquia semântica estabelecida pelo seu organizador (autor) e questionamos quais seriam os referenciais para estabelecer os parâmetros da formulação, bem como da organização dos conceitos que foram relacionados? Sabemos que há um conjunto de normas (taxonomia) para a organização de tesauros - aspecto que será um dos primeiros caminhos desta pesquisa; no entanto, são frequentes os equívocos a que essa tentativa de normalização tem induzido, resultando, quase sempre, na frustração do atendimento às necessidades de recuperação de informação.

Para tanto, o produtor de textos, nesse caso, aquele que insere as informações na ferramenta, ao analisar um documento, além de examinar os objetivos e o contexto de

produção, no sentido de garantir a etapa inicial de coleta de dados, deve também analisar os aspectos das relações conceituais entre os termos, ou seja, permitir condições básicas de análise estrutural dos vocabulários, para delimitar a equivalência de sentido entre o texto-fonte e sua representação. Portanto, para investir no objeto desta pesquisa, é necessário identificar tais princípios referenciais que norteiam a área de representação, recuperação e organização da informação. Tomando novamente o tesauro do INEP (Figura 01) como referência, nota-se que nele as palavras não são agrupadas segundo a ordem alfabética, como ocorre com os dicionários da língua, mas *de acordo com as ideias que elas exprimem*, ou seja, as palavras são encontradas pelo conteúdo que elas podem veicular; então, são arranjadas estritamente de acordo com seu significado. Geralmente, os tesauros possuem uma matriz conceitual, conforme se pode visualizar na Figura 02 e na citação seguinte, com o objetivo de restringir o campo semântico e facilitar a compreensão do leitor/usuário da ferramenta.

Figura 02 – Matriz Conceitual do Tesauro Brasileiro da Educação
Fonte: INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

A matriz conceitual é acompanhada do seguinte texto:

> Todos os termos do Thesaurus são selecionados e estruturados a partir de uma Matriz Conceitual da área. Para conceber o Thesaurus Brased partiu-se do princípio de que a educação é o processo pelo qual o ser humano (indivíduo e coletividade) desenvolve seu intelecto, suas potencialidades, sua cultura, satisfaz suas necessidades e se torna agente de sua história interagindo constantemente com o meio. A matriz conceitual do Thesaurus Brased, portanto, coloca o homem no centro do sistema educacional. (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2009).

Diante disso são estabelecidos campos ou subáreas a fim de se delimitar a abrangência da área da educação, conforme se visualiza no Quadro 01.

100 - Contexto da Educação	A educação do homem se realiza dentro da realidade global e em interação com esta; fora desta não há educação.
200 - Escola como instituição social	A Escola é a educação institucionalizada; na sociedade politicamente organizada, de fato, encontraremos todas as condições para que a educação do Homem socialmente aconteça.
300 - Fundamentos da Educação	A educação é o principal processo do desenvolvimento humano, que é pluri e interdisciplinar, isto é, muitas ciências fundamentam e integram no processo e a ação educativos.
400 - Educação: princípios, conteúdo e processo	O homem evolui interagindo constantemente com o meio: é a Educação propriamente dita com seus princípios, conteúdo e processo.

Quadro 01 – Subáreas do Tesauro Brasileiro da Educação
Fonte: INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

Mesmo com os critérios de restrição acima expostos, ainda permanece a questão da hierarquia semântica: como estabelecer princípios para a ordenação dos conceitos dentro de cada subárea, diante da rede de significados? Por que "academia de educação física" está incluída na faceta[9] Entidades Esportivas e não na faceta Escola-Instituição? (Figura 01). Alguns indicadores de caráter mais técnico já são conhecidos, como por exemplo, quando um termo é muito longo deve ser subdividido em dois ou mais termos que serão combinados, ou mesmo a definição se o vocabulário será mais superficial ou mais específico (matriz conceitual), pois a preocupação é dar fluência à busca de dados e, portanto, disponibilizar um vocabulário de entrada adequado, ou seja, quando ocorrer a busca do termo, o usuário/leitor deverá ser imediatamente conduzido ao conceito estando este sob um cabeçalho geral ou específico.

Entretanto, na lógica dos enunciados de busca, é perceptível que certos conceitos são mais importantes do que outros, pois há perfis de busca que são formulados por meio de listas de termos, nas quais para cada um se atribui um peso e uma classificação de relevância, combinando termos avaliados como adequados à organização. Podemos afirmar que a análise conceitual, na interpretação do que o usuário está à procura, é a tradução ou a representação de suas necessidades, pois os termos selecionados do vocabulário do sistema tentam corresponder à representação dessa necessidade de informação. Diante disso, é projetada uma rede significativa hipotética como uma forma de tentar atender à busca do usuário. Essa compatibilização do campo semântico com o sistema de busca/armazenamento pode ser

[9] Manifestação das categorias fundamentais. Tem a função de agrupar os renques (classes derivadas), dentro de cada categoria.

definida na alimentação do banco de dados de uma ferramenta de informação, estabelecendo-se critérios de restrição e abrangência para demarcar as relações de pertinência de um termo a um grupo conceitual. Conforme já mencionado, relações como hiperônimo/hipônimo serão úteis para auxiliar a hierarquização dos termos (do mais geral para o particular), além de contribuir para solucionar os problemas de equivalência semântica para auxiliar a desambiguação[10], entre outras necessidades; em outras palavras, constituindo regras para comportamentos linguísticos compatíveis com os usos em linguagem natural.

Podemos afirmar que, de modo geral, em uma *relação hierárquica*, a noção genérica mantém com a noção específica uma relação de inclusão, que se denomina relação hiperonímica-hiponímica. A hiperonímia é uma relação hierárquica de inclusão semântica entre duas unidades lexicais, partindo do genérico (hiperônimo) para o específico (hipônimo), de forma que o primeiro impõe sempre as suas propriedades ao segundo, criando assim, entre eles, uma dependência semântica. Um hiperônimo poderia substituir qualquer um dos seus hipônimos, porém o contrário não é possível; essa característica o diferencia da relação de sinonímia. Poderíamos afirmar, na observação da hierarquia proposta na Figura 01, que o termo "Entidades esportivas" impõe suas propriedades semânticas aos seus hipônimos "Academia de educação física", "Associações Esportivas", "Centros de esportes", "Clube esportivo". Consequentemente, a hiponímia, que é também uma relação hierárquica de inclusão semântica entre duas unidades lexicais, parte do específico (hipônimo) para o genérico (hiperônimo), mas, de forma inversa, possui os seus próprios traços diferenciadores, além de conservar as propriedades semânticas impostas pelo segundo. Por exemplo, o significado de "Clube esportivo" implica o significado de "Entidades Esportivas", porém resguarda características próprias. Portanto, nota-se uma relação de equivalência, mas que implica uma relação de inclusão.

Outra relação semântica que deve ser considerada e, muito comumente verificada em instrumentos de ordenação de informações, é a *relação partitiva*, expressa no duo meronímia/holonímia que tem sido objeto de estudo de diferentes áreas da Linguística, desde Lyons (1980) e Tamba (1991), numa dimensão semântica, bem como de Koch e Marcuschi (1998), no tratamento das anáforas. Embora haja autores que identificam algumas diferenças

[10] Também conhecida como desambigüização. Frequentemente encontramos *links* que, por erro ou ambiguidade, se referem a artigos cujos títulos são homônimos (palavras que têm dois ou mais significados diferentes; ou com grafia semelhante, produzindo o mesmo som). Muitas vezes, uma página pode conter informações sobre o nome abreviado de uma pessoa, por exemplo, que coincide com o nome de outra; ou então sobre um determinado tema que, apesar de não ser homônimo de outro, apresenta aspectos muito distintos para serem explorados no mesmo artigo. Para isso foram criadas as páginas de desambiguação; ou seja, uma página comum a várias entradas com o mesmo nome geral. Essa página aponta para cada uma dessas entradas colocadas em páginas com nomes mais específicos.

na delimitação do conceito e tipos de meronímia, nós a entendemos como uma relação semântica que considera uma unidade léxica como parte de uma outra unidade léxica. Em outras palavras, é a relação semântica entre dois termos, quando um é identificado como parte do outro, ou quando um dos lexemas está contido no outro. É também considerada uma relação de hierarquia semântica entre duas unidades lexicais, porque ao denotar a parte (merônimo), pode criar uma relação de dependência ao implicar a referência a um todo (holônimo), relativo a essa parte. A meronímia ocorre, então, quando o significado de um lexema faz parte ou é uma porção do significado de outro lexema. Por exemplo, as palavras "capa" e "folhas" são merônimos de livro (no entanto, não implicam necessariamente uma hierarquia). Dessa forma a holonímia também seria uma relação de hierarquia semântica entre duas unidades lexicais, de modo que uma denota um todo (holônimo), sem impor obrigatoriamente suas propriedades semânticas à outra, considerada sua parte (merônimo). No caso da palavra "carro", por exemplo, podemos estabelecer uma relação de holonímia com "volante", sem, porém, lhe impor suas propriedades. Essa relação pode ser entendida como uma relação "é parte de": "dedo" como merônimo implica "mão" como holônimo; "porta" e "janela" são merônimos ("parte") do lexema casa. Assim como na hiperonímia/hiponímia, a holonímia seria uma relação inversa à meronímia, isto é, uma relação que ocorre entre dois lexemas quando um contém o outro entre as partes que o constituem. É também entendida com uma relação "é formado por". No exemplo de "casa", esta é holônimo de porta e janela, pois pode se dizer que casa "é formada por" (constituída) porta e janela. Essa relação merônimo/holônimo pode denotar mais definidamente uma relação de inclusão, sem que haja a identificação de atributos característicos de cada classe, como no caso da relação homonímia/hiperonímia.

Além das relações destacadas nesta seção, há também as *relações associativas sequenciais*, ou seja, de equivalência significativa, como no caso da sinonímia, quando duas ou mais unidades lexicais remetem ao mesmo referente; e as *relações associativas não seqüenciais*, de oposição, como a antonímia, que é uma relação de oposição entre o significado de duas unidades lexicais, gerando consequentemente um outro arranjo hierárquico. É importante ressaltar que a antonímia é uma relação associativa complexa, pois há questões como a gradabilidade, complementaridade, multiplicidade e incompatibilidade. Acreditamos, porém, que todas essas relações entre palavras sejam empregadas, mesmo que intuitivamente, na constituição de bancos de dados de ferramentas de gestão de informação, aspecto que será retomado mais detalhadamente nas seções subsequentes.

Ressaltamos que a tradução e articulação pela linguagem de programação[11] não serão contempladas por esta pesquisa; interessa-nos aqui o processamento linguístico natural que é anterior à máquina e essencial para a validação desta. Vale destacar que um vocabulário de entrada bem estruturado reduz o esforço intelectual na busca e na indexação, se essa rede conceitual estiver adequada. Portanto, resgatar esse processo linguístico semântico empregado pelo autor, especificamente relacionado aos critérios de seleção, extração, análise e organização hierárquica dos termos, e tentar documentá-lo, pode contribuir para as reflexões acerca do cálculo do sentido na organização e hierarquização das redes conceituais. Salientamos que nos filiamos aos pressupostos de Mari (2008), que toma o cálculo do significado como procedimentos de ordem mais formal, de concepção logicista, como a análise da predicação, por exemplo, sendo possível a enumeração de "formatos de significado, a partir de categorias atômicas, associadas a relações gramaticais aprioristicamente definidas" (p.78), num dado sistema linguistico. Sendo assim, podemos tomar como exemplo quando é preciso ordenar/incluir o termo "bola" numa estrutura hierárquica de produtos farmacêuticos; para tanto, numa dimensão mais superficial, seria necessário identificar traços diferenciadores para distinguir "bola", de objeto natural ou fabricado, maciço ou oco, redondo em toda a volta, esfera, entre "ocasião propícia para um cometimento; boa estrela"; para, talvez chegar a "comprimido, geralmente calmante ou soporífero", provavelmente mais adequado à sua momentânea demanda, e, quem sabe, excluir "bolo de carne ou bocado de comida envenenada com que se matam cães". (HOUAISS, 2006).

Pesquisar os critérios empregados nesse processo de hierarquização de conceitos é nossa intenção, pois acreditamos que seja de fundamental importância, pois ele se situa na base dos instrumentos de busca de informações. Esta pesquisa pretende comparar e demonstrar os critérios utilizados pelos gestores de informação que buscam normatizar essa organização. Nossa discussão é a de que tais critérios são de ordem semântica, e é nessa dimensão temática que pretendemos trabalhar, na tentativa de tentar trazer subsídios para reconhecer a aplicabilidade da dimensão linguística em conexão com a interface comunicacional, que é um princípio básico das tecnologias da informação (TI); ou, ainda, mapear e dar maior transparência a esse processo, com o objetivo de compreender os mecanismos de leitura e produção de textos mediados pelas TIs.

[11] Conjunto bem definido de signos (palavras, números, expressões etc.), regras sintáticas e semânticas utilizado para escrever programas de computador. Essas regras de encadeamento lógico fornecem uma sequência de instruções (um programa) que podem ser decodificadas e executadas por computador.

Esperamos documentar esse processamento linguístico empregado na estruturação/formulação[12] das redes conceituais, tentando delinear os referenciais norteadores para o processo de hierarquização semântica realizado pelo produtor de texto, na inserção de conteúdos em ferramentas de organização da informação. Pensamos que investir na delimitação de traços para as categorias poderia gerar uma discussão cujo intuito é encaminhar para o desenho hierárquico da rede conceitual, definida na seleção lexical das estruturas hierárquicas. Acreditamos que seria importante rever a teoria dos protótipos, numa tentativa de estabelecer princípios associativos e hierarquizadores. Para tanto, articulamos alguns autores como referência: Rosch, Fodor, Cohen e Murphy, Osherson e Smith, Johnson-Laird, entre outros, de modo a delinear um plano de investigação.

[12] A língua, tomada como mediadora dos processos interativos, é considerada como uma atividade sócio-cultural, através da qual os homens representam seus conhecimentos, estabelecem princípios e negociam sistemas de compreensão, em atendimento às demandas de sua sobrevivência histórica. Portanto, podemos afirmar que é inerente aos estudos linguísticos a análise semântica, uma vez que as possibilidades de significação são produto, muitas vezes, de estruturações sintáticas demandadas pelas situações comunicativas e adaptadas aos perfis contextuais institucionais.

3 TEORIAS LINGUÍSTICAS

3.1 Semiótica e Semântica

Acreditamos que qualquer trabalho que envolva a palavra como referência de estudo, necessariamente, precisa percorrer o caminho da criação e atribuição de sentidos, tarefa desafiadora, mas de grande relevância; por outro lado, infinita e de parâmetros provisórios e instáveis, dada a dinâmica de qualquer língua.

Na medida em que captamos, categorizamos e transformamos dados resultantes de nossa percepção em informações estruturadas e contextualizadas, geramos um sistema de conceitos que representa a realidade percebida e apreendida. Essa capacidade cognitiva de estruturar ideias e reorganizá-las, sempre que necessário, indica nossa capacidade de criar e utilizar conceitos de acordo com nossas necessidades enunciativas. Os conceitos já apreendidos por um indivíduo permitem que ele contextualize os novos conhecimentos e os fixe, proporcionando-lhes sentido e utilidade, essenciais ao aprendizado, e, consequentemente, à produção social do conhecimento.

Sob a perspectiva da semiótica de Pierce (1972), a relação do indivíduo com sua realidade é mediada pela representação, e verbalizada através de signos constituídos socialmente e articulados a processos de produção de sentido. Pierce propõe uma classificação de significantes, ou seja, do aspecto perceptível dos signos em índices, ícones e símbolos. A relação direta com o referente, o *índice*, denota a relação perceptual mediada pelos nossos sentidos; já a relação de semelhança com o referente caracteriza os *ícones;* e a representação *simbólica*, que nos interessa aqui, é convencional, imposta pela sociedade. Podemos citar como exemplo sumaríssimo desse aspecto da teoria peirceana, a relação indicial/perceptiva: visual e/ou olfativa (índice) de um cheiro de fumaça; a iconográfica na representação gestual em libras e/ou em desenhos de pequenas ondas verticais (ícone); e simbológica, arbitrária e convencional, da constituição da palavra *fumaça* em português e *smoke* em inglês (símbolo). Em outras palavras, a partir de sua relação com os objetos, o homem categoriza sua realidade através de símbolos, que podem ser palavras, gestos, imagens, sons.

Temos de nos remeter, também, às contribuições saussureanas na estruturação dos signos, a dualidade de sua constituição: o aspecto sensível, nomeado como *significante*,

produto da percepção sensorial, e o aspecto inteligível, chamado de *significado*, de caráter semântico, são duas faces do signo que não se separam, porém se distinguem. Um significante, por exemplo, pode ter mais de um significado, da mesma forma que um significado pode ter mais de um significante, como por exemplo: *manga* com seus vários significados e o conceito de beleza pode ser representado por outros significantes como *belo* ou *bonito*. A plasticidade linguística na estabilização de um conceito demonstra como os repertórios lexicais do sistema linguístico são ilimitados. No entanto, a delimitação de campos lexicais pode evidenciar traços de sistematicidade que permitem a formatação de redes semântico-lexicais, que denotam regularidades para a viabilização da comunicação. São pressupostos da Lexicologia[13], tomada aqui como um ramo da Linguística Aplicada, que contribuem para o estabelecimento de terminologias em áreas distintas para a circulação do saber.

Consideramos necessário, no entanto, um resgate das perspectivas teóricas empreendidas em torno da semântica a fim de fundamentar e situar os parâmetros que nortearão este trabalho. Acreditamos que a problemática do significado vai além das fronteiras da linguística, pois está fortemente relacionada com questões referentes à aquisição e organização do conhecimento. Sendo assim, podemos afirmar que existem correntes semânticas que tentam estabelecer noções de significado e modelos próprios.

3.2 Breve percurso das abordagens semânticas

Tamba-Mecz (2006) fez recentemente um levantamento sobre a história da semântica e sustenta que 3 (três) correntes teóricas conduziram os estudos semânticos. A primeira corrente conduz a uma orientação histórica, situada pela autora entre 1883 a 1931, configurando um período *evolucionista*, marcado pelas discussões fundantes da *ciência da significação*, proposta por Michel Bréal, até Jorst Trier, com a *Teoria dos Campos Semânticos*. Nesse período, segundo a autora, a semântica tinha por objeto o estudo da evolução das significações nas línguas, comandadas por leis gerais, próprias do fenômeno semântico e produto da observação de fatos de sentido. Dessa forma, diversos modelos explicativos desse período procuravam contemplar a extensão cronológica e geográfica, o

[13] Segundo Haensh, a lexicologia tem por objetivo a descrição do léxico e se ocupa das estruturas e regularidades dentro da totalidade de um léxico de um sistema individual ou de um sistema coletivo.

quadro descritivo e o estatuto teórico da significação. De acordo com Tamba-Mecz, esse *evolucionismo* contribuiu para estabilizar a semântica como uma "ciência autônoma, liberta da tutela ancestral da lógica e da filosofia, ao lhe assinalar um objeto próprio de estudo: a história de suas transformações internas" (p.25-26). Consequentemente, diante das regularidades contínuas, a perspectiva evolucionista mostrou que a história é um meio para a linguística e não um fim, pois os acasos da história acabavam por relativizar os princípios gerais.

A partir de 1916, com a publicação do *Curso de Linguística Geral*, de F. Saussure (2003), começam a conviver os pontos de vista evolucionista e o sincrônico. A perspectiva estrutural demarcaria, portanto, um período misto, situado por Tamba-Mecz, como a história das palavras e do léxico, de 1931 a 1963. Nesse período, ela destaca Ullmann (1973) que integra, em seus estudos, esses dois pontos de vista. Ressalta, ainda, que houve uma tentativa de elaboração de princípios semânticos, de unidades mínimas de sentido e de suas regras de composição para descrever todos os significados lexicais das línguas; além disso considera o surgimento da Teoria dos Campos- a ser desenvolvida mais a frente- e sua associação à análise sêmica, que gerou a utilização de traços distintivos opositivos para compor um campo semântico. Também situa Weinreich, que estabeleceu um paralelismo entre as relações mantidas pelos semas que compõem um vocábulo e as que ligam os termos de um enunciado definitório; e Rastier, que adapta o instrumento sêmico à análise textual. Tamba-Mecz considera que essa corrente estruturalista teve um duplo efeito: contribuir para limitar o objeto da semântica ao léxico, concebido como uma rede de relações de sentidos auto-regulada; e evidenciar a dificuldade de encerrar em si mesmas as estruturas lexicais, articuladas a conhecimentos extralinguísticos.

Cabe aqui uma referência às contribuições do linguista, Algirdas Julien Greimas, nome ligado à semântica estrutural e à fundação da Escola Semiótica de Paris. Greimas (1975) levanta fatores que interferem na estabilização dos sentidos, postulando que toda transcodificação, toda nova articulação de sentido pode indicar um acréscimo deste ou pelo menos uma diferenciação que deve ser identificada e anotada. Ele discutiu a articulação da lógica simbólica com a lógica linguística, bem como da lógica das equivalências. Ou seja, os fenômenos linguísticos devem ser avaliados na amplitude das variantes que interferem na produção de sentido. Para ele, a significação é a transposição de um nível de linguagem a outro, de uma linguagem a outra linguagem diferenciada, e o *sentido é apenas essa possibilidade de transcodificação,* em outras palavras, se revela do enunciado. Greimas situa que se o significado não é possível sem o significante, então eles se inter-relacionam,

completam-se, referem-se apenas nas significações. Dessa forma afirma que os significantes são "elementos ou grupos de elementos que possibilitam a aparição da significação ao nível da percepção" (p.17); e que os significados são o conjunto das "significações que são recobertas pelo significante e manifestadas graças à sua existência." (idem). Greimas opunha-se ao formalismo behaviorista, que distinguia forma de conteúdo. Ele propôs uma semântica linguística, que se ocuparia da análise da significação tal como é fornecida pelo código da língua, mas enfocada nas diferenças produtoras do sentido do texto.

O terceiro período proposto por Tamba-Mecz (2006), e de particular interesse neste trabalho, é o período das teorias linguísticas e do tratamento computacional, de 1960 a 1990, no qual a autora subdivide em três conjuntos de estudos em semântica: *semântica formal; semântica e cognição; e as teorias pragmático-enunciativas do sentido.* No primeiro período estipulado pela autora, a semântica é formalizada no quadro de uma teoria gramatical das línguas naturais (1963-1982), com a gramática gerativa de Chomsky e a gramática universal de Montague (1974). Chomsky propõe um estudo supostamente autônomo e formal, *independente da semântica,* suficiente para dar as regras de boa formação de todas as frases (1957). Em 1963, Jerrold J. Katz e Jerry A. Fodor incluem um componente semântico na componente sintática de base.

Trata-se de uma proposta de cunho estrutural, que tinha dentre seus propósitos fazer leituras de sentenças a partir de amálgamas de itens lexicais de modo a desambiguar sentenças. Essa fusão tão estreita de dois ou mais morfemas possibilita separar somente um segmento no plano da forma, embora possam ser identificados diversos significados no plano do conteúdo, ação questionável em se tratando de línguas diferentes. Os autores estenderam, assim, o domínio estrutural para além da mera análise componencial, sugerindo a inserção das regras de projeção, que dariam feição sintática à proposta. Tinham como meta a leitura semântica de sentenças. Nesse sentido, afirmavam que uma teoria semântica deveria encerrar dois subcomponentes: um dicionário e uma regra de projeção. O primeiro diz respeito ao conjunto de itens lexicais, como entradas associadas às definições, o que implica uma lista de formas teoricamente ilimitadas. A segunda se relaciona com a capacidade de codificar e decodificar a informação veiculada pela frase. Tais regras devem ser conhecidas pelo usuário, e são aplicáveis sobre os itens lexicais, impondo percursos de leitura.

Em 1965, Chomsky adota o ponto de vista desses linguistas, segundo Tamba-Mecz, e propõe a gramática gerativa transformacional, que preserva a autonomia e a primazia do componente sintático, mas a completa, adicionado-lhe um componente semântico interpretativo. A autora afirma que de 1969 a 1972, houve uma ruptura em duas escolas: a

ortodoxa, da gramática gerativa de Chomsky e Katz, que defendem a autonomia da sintaxe e da semântica interpretativa; e a da semântica gerativa, defendida por George Lakoff e James McCawley, que propuseram partir de formas subjacentes identificadas a predicados lógicos para derivar as estruturas sintáticas de superfície e as unidades lexicais. Portanto, do status de semântica lexical passa-se a uma semântica frástica. Ressaltamos, no entanto, que a semântica sentencial não foi resultado dessa crise gerada por Lakoff e McCawley; isso já estava em Katz, de forma mais efetiva do que nesses autores. De toda forma, as significações conceituais, relacionadas à problemática da palavra-signo em relação às coisas e ao sistema semiótico, conforme já situado neste trabalho, ficam parcialmente esquecidas em proveito das significações relacionais, resultantes da combinatória dos significados estruturais gramaticais e dos significados lexicais. Consequentemente, a análise do semantismo interno das unidades lexicais e a análise das relações de sentido interfrásticas são atribuídas à semântica, situando-a na articulação das estruturas lexicais e sintáticas.

Tamba-Mecz ressalta que, em 1974, os trabalhos de R. Montague estabeleceram uma semântica lógica fundada na teoria dos modelos, que representa um programa de investigação da linguagem natural que rompe com a tradição de enraizar a linguística no interior da Psicologia e, consequentemente, como um ramo das ciências formais. Para Montague, a linguagem natural humana não se distinguia essencialmente de uma linguagem artificial construída; por isso seria possível investigá-la dentro de uma disciplina formal como a Lógica e a Matemática. Ele assumiu, como Chomsky o fez, a hipótese de descrever a estrutura da linguagem mediante formalismos precisos, que permitiriam explicitá-la sintaticamente. Porém a ruptura com o programa gerativista estava no fato de que Montague não privilegiou a sintaxe nem lhe concedeu a autonomia que Chomsky lhe atribuiu; sintaxe e semântica representariam componentes da teoria da linguagem a serem construídos de maneira integrada em uma interface explicitamente isomórfica, com similaridades superficiais. Jorge Campos (2004) afirma que, para Montague (1974), é a semântica que determina a direção do processo, impondo restrições sobre a sintaxe, ao complexo formal, o que provocou muitas críticas ao seu trabalho, no sentido de ele ter reduzido a semântica a uma sintaxe combinatória de índices para os itens lexicais. A explicitação formal do significado, portanto, depende primeiro da determinação de sua sintaxe, tendo em vista a posterior interpretação no nível semântico. Tamba-Mecz reafirma que, ainda hoje, a semântica de Montague estimula muitas formas de expansão, especialmente pela sua reconhecida potencialidade para implementação computacional.

Como resultado de tais discussões, Tamba-Mecz (2006) aponta que algumas conquistas desse período situam-se na definição de fronteiras, colocando a semântica formal e sua problemática no nível das relações entre sons e sentidos, formas e significados linguísticos, relações estas examinadas no quadro das frases, como unidades teóricas geradas por uma gramática ou por um conjunto de regras, *independente de qual seja seu contexto de uso*. Dessa forma o interesse se deslocou das relações de sentido para as operações das quais elas resultam. A semântica formal passa a enfocar as propriedades lógico-semânticas das línguas e das linguagens formais. A autora reforça que isso implica um aprofundamento das pesquisas formalizantes, principalmente sob pressões da Linguística Computacional, Inteligência Artificial (IA) e das Ciências Cognitivas. Haja vista que, na década de 80-90, surge uma gama de gramáticas "chamadas de unificação, baseadas nas funções gramaticais associadas às formas de superfície e em uma representação por estrutura de traços. Tem-se, portanto, um tratamento lógico-simbólico unificado e flexível de conhecimentos de todas as ordens, o que facilita a articulação entre sintaxe, semântica e pragmática."(p. 41). Essa perspectiva prevê, então, modelos que, em tese, facilitariam o processamento textual, na medida em que disponibilizam um conjunto organizado de dados para as produções. No entanto, sabemos do alto nível de relativização que envolve essa discussão, pois a tentativa é a de prever uma gama de possibilidades de busca dos usuários.

Nessa fase, Tamba-Mecz cita ainda o advento da DRT, "Discourse Representation Theory", proposta por Hans Kamp, em 1993, e refere-se a Francis Renaud, que elabora um modelo semântico que comporta uma base de conhecimento inspirada na base da IA e em um módulo de raciocínio, para tratar a informação evolutiva e estimular a compreensão humana. Além de situar a ampliação do estudo do léxico, que, segundo a autora, é *um verdadeiro gargalo para o tratamento automático das línguas*, ressalta, ainda, o grande número de pesquisas linguísticas que fazem uso de bancos de dados, *corporas*, cuja consulta e manipulação, viabilizados pelo processamento de dados, enriqueceram, consideravelmente, os conhecimentos lexicais, bem como a criação de dicionários informáticos que registram as propriedades combinatórias de alguns milhares de verbetes, permitindo uma melhor avaliação das estruturas lexicais.

Além da semântica formal, no terceiro período situado por Tamba-Mecz, a perspectiva cognitiva nos estudos semânticos, ou seja, as ligações entre estruturas semânticas e cognitivas, reforçam os estudos sobre a categorização (Rosch), conforme já citado neste trabalho, em busca de uma noção prototípica como uma alternativa à definição lexical componencial. Como último ponto desse período, a autora indica a importância das teorias pragmático--

0enunciativas do sentido, que traduzem a incorporação de determinações provenientes das condições de uso das línguas.

Todo esse percurso tem a intenção de situar a semântica como linha teórica básica deste trabalho. A questão principal que se interpõe é a tentativa de apreensão das relações entre formas e sentido nas línguas. Podemos afirmar que desde o rompimento das teorias semânticas com a preocupação etimológica, que propunha o agrupamento de termos em relações no plano fonético ou no morfológico, houve estudos relativos à compreensão da linguagem, porém dedicados às particularidades do vocábulo, em ordem dispersa e sem preocupação com os métodos, numa perspectiva histórico-comparativa. O ano de 1900 foi certamente o marco para os estudos linguísticos com as propostas de Saussure, pois ele desloca a perspectiva diacrônica, dando início aos estudos sincrônicos, configurando o estruturalismo clássico na linguagem. Nessa fase, nota-se que há a consolidação da Semântica Geral, cujos princípios teóricos, evoluções de sentido e campos semânticos foram estabelecidos. Nos anos que se seguiram (1930-1950), três escolas se consolidaram como estruturalistas clássicas: a Escola de Paris, 1930, com destaque para Martinet e a Sintaxe Estrutural; a Escola de Copenhague, 1940, na qual o modelo da semântica estrutural e da semiótica científica se consolida como forma do plano do conteúdo semântico, com ênfase em Hjelmslev, que, de certa forma, definiu o programa de uma Semântica Estrutural, pois conduziu seus estudos de modo a estudar a forma do plano de conteúdo das línguas naturais, ou seja, a estruturar os fatos; atribuir valores; porém careceu de metodologia científica; e a Escola Americana, 1945-1959, com Bloomfield e Ullman, que publicou, em 1951, “Principles of Semantics”, nos quais aplicou os processos da investigação estrutural à perspectiva semântica. Até que chegamos à semântica linguística, no período gerativo-transformacional, como uma fase preparatória para o pós-estruturalismo. Temos, então, os estudos de Chomsky, que propõem o “inatismo”, partindo do pressuposto de que o indivíduo possui a gramática internalizada e, consequentemente, conhece a gramática de sua própria linguagem. E, a partir de 1970, temos Ducrot afirmando que os únicos dados de que o semanticista dispõe são “as suas ocorrências nas diferentes situações de emprego”. (p. 117). Dessa forma, o sentido é algo que se revela do enunciado. Ele ainda afirma que não existe na realidade nenhum enunciado que ocorra *extracontexto*. Um enunciado pode apresentar um contexto de dimensões extremamente reduzidas, mas ele possuirá, sempre, um contexto – aquele da situação que põe em contato dois interlocutores.

Temos, então, uma linha que parte da palavra, caminha para a estrutura sintática e atinge as dimensões contextuais. Não consideramos essas correntes excludentes, pois

pretendemos demonstrar aqui que a perspectiva das tecnologias da informação necessita de uma linha mais estrutural para permitir a constituição dos bancos de dados, muito embora faça a extração de seus dados de produções contextualizadas. O Compara, por exemplo, é uma ferramenta de informática muito utilizada em traduções, que possui um imenso banco de exemplos (excertos de obras literárias e/ou textos jornalísticos) nos quais determinada palavra aparece empregada. Essas *cenas enunciativas* denotam situações de uso, para que o autor/usuário possa refletir sobre as possibilidades e utilizar os dados de que necessita para constituir suas produções.

3.3 A Semântica Formal no universo das discussões semânticas

Cabe aqui ressaltar o interesse mais específico pela semântica formal, em razão da perspectiva tecnológica que permeia este trabalho. Entendemos como objetivo da semântica formal a tentativa de caracterizar com relativa precisão os significados que podem ser expressos pelas línguas humanas e explicitar como esses significados são compostos pela estrutura sintática. Na perspectiva dos instrumentos de informática, a semântica formal preocupa-se em especificar *rigorosamente* o significado ou comportamento de programas, partes de hardware etc; dessa forma, a sintaxe descreve as estruturas de uma linguagem, e a semântica descreve o significado dessas estruturas, nas linguagens de programação. Conforme já anunciado, o nosso interesse é o de mapear o processo humano na tentativa de estabilização de significados, que é anterior à máquina; porém, como, até o momento, é o homem o elemento responsável pela alimentação do banco de dados, pode-se afirmar que encontramos na semântica formal o respaldo para uma necessária (e provisória) estabilização de significados.

Conforme ressalta Borges Neto (2003), a *natureza não-fisica* dos significados nos leva a considerar a problemática do significado no âmbito das relações; no entanto faz-se necessário *reduzir* os significados a alguma representação que possa ser cientificamente manipulada, que traduza a relação entre o linguístico e o não linguístico. Nesse sentido distinguem-se algumas abordagens: uma delas é a de que o significado está nas relações entre as expressões linguísticas e o mundo externo, que as expressões descrevem (condições de verdade); outra que considera as relações que se dão entre as expressões linguísticas e o uso que os falantes fazem delas (atos de fala); uma terceira que é a relação estabelecida entre as

expressões e os conceitos mentais, propostos pela teoria ideacional do significado (semântica cognitiva, em Jackendoff (1983), por exemplo). Podemos, então, configurar três (ou mais) frentes: uma linha referencial, uma ideacional e outra comportamental, todas com o objetivo de denotar relações que constituem o objeto da semântica: estabelecer relações sistemáticas entre as representações dos conceitos e entre as próprias representações entre si; entre as expressões das línguas naturais e as representações semânticas. Borges situa que a polêmica das abordagens semânticas está na falta de uma caracterização formal de uma linguagem que traduza as representações semânticas, que ele chama de *linguagem auxiliar*, que permita sua interpretação. Em outras palavras seria "uma sintaxe" dessas linguagens, representações semânticas "escritas" numa linguagem formalmente bem caracterizada. Sabe-se que a utilização de linguagens formais como modelo para a linguagem das representações semânticas é possível, porém incompleta, pois há limites da enunciação inacessíveis às linguagens formais, como no caso das inferências. Apesar disso, este é um dos princípios ordenadores dos instrumentos de informática, objeto deste trabalho, porque há a necessidade de se estabelecer uma rede lógica entre os termos, como as relações de oposição e equivalência, por exemplo, a fim de ordenar hierarquicamente os conceitos. Assumimos aqui como necessária uma perspectiva de aporte lógico-matemático que, inegavelmente, se encontra na estrutura linguística; mas já prevendo as polêmicas, entendendo-a como uma forma de representação, uma linguagem "auxiliar" conforme nomeada por Borges. O caminho da formalização nos encaminhará aos modelos para o estudo da semântica das línguas naturais, que reforçam o ponto de vista da semântica formal em nossa discussão.

Nessa perspectiva, apresentamos as concepções de Oliveira (2001) que propõe a definição de três correntes semânticas distintas: *a semântica formal, a semântica da enunciação e a semântica cognitiva.* Segundo a autora, a *semântica formal* descreve o problema do significado a partir da estrutura lógica das sentenças. Nesse sentido, Frege (1978) afirma que o estudo científico do significado só é possível se diferenciarmos os seus aspectos subjetivos para focar nos objetivos; para ele, *objetivo* é tudo aquilo que pode ser compartilhado em sua acessibilidade por mais de um sujeito, caracterizando-se por ser independente desse ou daquele sujeito particular, enquanto o que é *subjetivo* depende de um sujeito, de um portador. Dessa forma, ele exclui da semântica o estudo das representações individuais que uma palavra pode provocar, delegando à psicologia tal reflexão. Estabelece, então, para a semântica o estudo daquele significado garantido pelo acordo social de dada comunidade. Para tanto, Frege diferencia sentido de referência com a analogia do telescópio apontado para a lua. Tomando a lua como *referência*, situa que a existência dela independe

daquele que a observa; e essa visão comum a todos é entendida por ele como *sentido*. Sendo assim, o sentido nos permite alcançar um objeto no mundo e a linguagem cumpriria esse papel de viabilização. Então, o conceito de verdade seria externo à linguagem, permitindo-nos tratar mais objetivamente aquele objeto e adquirir conhecimento sobre ele, a partir de uma possível ordem no mundo que dê conteúdo à linguagem.[14] Podemos afirmar, portanto, que a semântica formal possui um caráter supostamente objetivista que, segundo Oswald Ducrot, é ilusório, pois nós falamos para construir um mundo e, a partir daí, tentamos convencer o nosso interlocutor de nossa verdade.

Ducrot discute a relação estruturalismo e enunciação, propondo uma *semântica da enunciação* (também conhecida como semântica argumentativa). Ele afirma que:

> A descrição semântica de uma língua, considerada como conjunto de frases ou de enunciados, não só não pode ser acabada, como não pode ser empreendida de forma sistemática, se não mencionar, desde o início, certos aspectos da atividade linguística realizada por essa língua. (DUCROT, 1977, p. 291).

Esses aspectos por ele mencionados referem-se à linguística de "fala" que remete a Searle, com a Teoria dos Atos de fala. Ducrot emprega, também, a noção de autoreferência de Benveniste (1989) como característica da atividade de fala e atribui aos enunciados uma eficácia especial: "a própria materialidade do que é dito traz em si um poder sagrado". Partindo da enunciação como um *ato único*, propõe uma unidade semântica que intitula "a significação do enunciado".

A esta altura, consideramos pertinente delinear a proposta de Oliveira, que se filia a esse paradigma, partindo do pressuposto de que a linguagem constitui o mundo e que não existe ordem no mundo que não seja dada a partir da linguagem e da história. Nesse caso, a referência torna-se uma ilusão criada pela linguagem, pois é interna ao jogo discursivo e, portanto, o constitui. Entendemos a referência como o objeto que seja representado pelo signo, que está fora de *nós*; já o significado está em *nós*, pois é uma construção social. Por exemplo, ao tentarmos responder qual seria o significado da cor "vermelha", podemos construir duas possíveis respostas: uma que remete ao seu referente, uma amostra do vermelho; e outra que envolve definições relacionadas à categoria do vermelho como pertencente às cores primárias, ou seja, que remetem ao sentido que o vermelho tem.

[14] Cabe aqui ressaltar que os sistemas computacionais tratam o termo *semântica formal* como um requisito para a especificação, rigorosa e precisa, da sintaxe e semântica de linguagens de programação. O formalismo contribui para garantir a robustez de seus modelos de hardware. Tal abordagem será retomada à frente.

Cabe ressaltar, porém, que o referente de um signo pode ser outro signo ou pode ser também o próprio signo; por isso uma *árvore*, por exemplo, não depende da linguagem para existir, mas certamente a linguagem é usada para apontá-la. Então expressões referenciais, retomando Ducrot, permitem ao locutor designar, para o destinatário, um ou mais objetos particulares do universo do discurso (real ou imaginário). E a estrutura proposicional (o enunciado), a *ordem*, precisa da linguagem para existir.

Numa obra marcante da semântica do Século XX, "The Meaning of Meaning" (1923), Ogden e Richards ilustram com um célebre triângulo (Figura 04) essa relação triádica da representação, já tratada no início deste capítulo, em que, na base do triângulo, se encontram o símbolo, no lado esquerdo, e o referente, no lado direito, e no topo o pensamento ou referência. Como na base do triângulo não há uma relação direta entre símbolo e referente, a relação entre estes dois é indireta, mediada pelo pensamento ou referência que se encontra no topo.

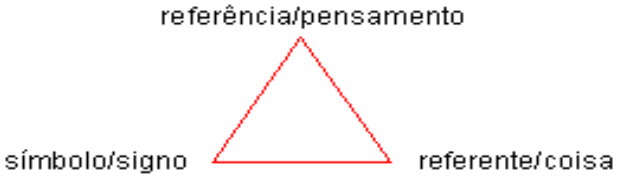

Figura 03 – Triângulo Semiótico
Fonte: OGDEN; RICHARDS, 1956, p.11.

Desde a obra de Ogden e Richards muita investigação foi feita no âmbito da semântica, e várias propostas de "novos" triângulos foram retomadas por diversos autores. Cabe aqui, no entanto, uma distinção básica entre significado e referência, feita inicialmente por Frege (1978), que partindo da relação de igualdade, mostra que o significado de um nome ou signo é apreendido por quem conhece a língua ou o conjunto dos signos em que esse signo se enquadra, pois a língua é um produto cultural, na mesma medida em que a cultura constitui-se através da língua. E, conforme já dito, normalmente, um signo tem um significado e a esse significado corresponde uma referência; mas nem sempre a um significado corresponde uma referência, pois o mesmo significado e a correspondente referência têm em diferentes línguas diferentes expressões. Frege frisa que o significado não é uma representação subjetiva; para ele o significado é objetivo; no entanto, a representação que uma pessoa faz de um objeto pode ser diferente das representações que outras pessoas têm do mesmo objeto. A representação de uma árvore em forma de um desenho, por exemplo, pode variar de pessoa para pessoa; porém o significado de árvore, em contrapartida, é comum a todos aqueles que o apreendem.

Além da perspectiva da enunciação, retomando novamente Ducrot (1977), árvore pode adquirir vários outros significados: vegetal lenhoso, grafo conexo sem ciclos, diagrama que representa a estrutura de constituintes da oração ou estrutura genealógica. Então, a ideia expressa pela proposição (enunciado) constitui o seu significado, isto é, o valor do pensamento expresso na proposição depende da referência dos seus elementos: esse valor é justamente o valor de verdade da proposição. Segundo Ducrot, há certos elementos da gramática de uma língua que têm como função indicar a força argumentativa dos enunciados e o sentido para o qual apontam, denominados por ele de *operadores argumentativos*. Afirma o autor que toda língua possui, em sua gramática, mecanismos que permitem indicar a orientação argumentativa dos enunciados: "a argumentatividade está inscrita na própria língua". (p. 306).

Dessa forma, a *semântica da enunciação* sustenta-se nos conceitos de polifonia (a pluralidade de vozes na constituição do discurso) em razão de sua orientação argumentativa. Nessa linha, Bakhtin (2002) contribuiu com o conceito de dialogismo, sustentando que todo enunciado contém uma dialogia interna, fundamental na constituição dos sentidos. Em outras palavras, caracteriza a linguagem como essencialmente dialógica; para ele, a palavra é um ato de duas faces, determinado tanto por quem o emite, quanto por quem o recebe. Diante disso, a semântica da enunciação considera as diversas representações do sujeito da enunciação no enunciado. Um dos méritos da semântica da enunciação foi possibilitar a descrição de fenômenos que resistem a um tratamento formal, e que certamente contribuem para as redes conceituais que são estabelecidas no discurso.

Como último paradigma, Oliveira (2001) situa também *a semântica cognitiva* como um modelo que contesta a ideia de que a linguagem estaria numa correlação direta com o mundo real. Sendo assim, considera fundante a perspectiva formal, tomando a concepção da linguagem como representação do pensamento, conforme já demonstrado aqui; mas ela se tornaria insuficiente à luz da semântica cognitiva, na medida em que descreve a forma semântica de uma proposição, tomando como referência a informação linguística que se mantém constante nas várias possibilidades de encadeamentos. Em outras palavras, a semântica cognitiva opõe-se à abordagem formal, principalmente no aspecto da representação, pois, segundo a autora, o pensamento é estruturado por *esquemas de imagens* expressos pelas nossas ações e pelos nossos pensamentos, frequentemente viabilizados por meio de "metáforas, mapeamentos entre domínios conceituais distintos." (p. 27). O significado, nessa direção, emerge de dentro para fora, da experiência do corpo em interação com o meio que nos circunda, portanto é motivado.

Para essa abordagem trazemos as discussões de George Lakoff e Mark Johnson (2002), que defendem a ideia de que o significado é natural e experiencial, sendo construído nas nossas interações físicas. Lakoff define como realismo experiencialista, de onde adquirimos sistemas imagético-cinestésicos com a associação da percepção sensório-motora ao significado linguístico. Dessa forma, nossas ações no mundo nos permitem apreender grande parte dos conceitos que armazenamos na memória; mas há outras conceitualizações que derivam de mecanismos de abstração como os da metáfora e da metonímia enfocados pelos autores. Eles demonstram que a metáfora é um processo cognitivo que permite deslocarmos esquemas apreendidos na interação física para domínios mais abstratos, de experimentação indireta, fruto de processos argumentativos nos jogos linguísticos. Como acontece no caso da palavra *touro:* pela experiência física e na interação social, atribuímos a essa palavra o significado de um *animal*, macho bovino não castrado; já na perspectiva metafórica, pode ser entendido como símbolo de *força muscular*, em construções como *Ele é um touro!;* ou ainda, na astrologia, pode ser entendido como um *signo do zodíaco*, dentre outras possibilidades. A metáfora viabiliza a articulação entre os mecanismos de inferência que permitem a localização de um termo fonte para um domínio alvo, estruturado em dada sentença, ou situação específica. Sabe-se, porém, que toda essa discussão sobre o significado de touro não decorre apenas de representações internas dos sujeitos, pois há um construto social aprendido decorrente das comunidades as quais estão inseridos. Mas há que se ressaltar que a perspectiva cognitiva leva em consideração as referências armazenadas na memória, resultantes das relações experenciais conforme já dito. Isso reforça também o aspecto não excludente das abordagens aqui apresentadas.

Frege (1980) sustenta que um conceito pode ser caracterizado por um conjunto de atributos de definição, que ele chamou de **intensão**. Podemos afirmar que seriam as propriedades definidoras dos termos; ou, ainda, o conjunto de propriedades essenciais que determinam a aplicabilidade de um termo. Desse modo a intensão pode remeter a várias extensões. Por exemplo, a *intensão de flor* seria representada pelas pétalas, pelo caule fino e pequeno, e pelo perfume entre outros atributos (elementos do conjunto flor); já a *extensão de flor* delinearia um subconjunto margarida, cravo, rosa etc. Então a *intensão* é composta pelo conjunto de atributos que definem as necessidades para ser um membro daquele conceito, e pela *extensão*, que é o conjunto de entidades que são membros daquele conceito, ou seja, por um número de exemplares que o formam. Em outras palavras, a **extensão** seria a classe de entidades a que uma palavra se aplica, ou o conjunto de objetos da realidade extralinguística a que uma expressão faz referência, a classe das coisas às quais ele é corretamente aplicado.

Em outro exemplo, uma *maçã* possui determinadas propriedades identificáveis pelos nossos esquemas sensoriais (cor, formato, cheiro etc – intensão) e que nos permitem alcançar os objetos no mundo (extensão). Mas nem sempre a intensão se refere a um indivíduo, pode ser também a um conjunto: a intensão de maçã como *fruto* pode definir todo o conjunto, como "um termo botânico que designa um órgão gerado por vegetais que produzem flor, a partir da fecundação e desenvolvimento do ovário." (HOUAISS, 2006). Então, nesse conjunto se incluem laranja, goiaba, mamão, berinjela e carrapicho, pois todos se desenvolvem a partir do ovário, o que seria um atributo para o pertencimento. A questão é que nem todo *fruto* é *fruta*; a categoria *fruta* refere-se à "parte comestível da planta que não é formada pelo ovário, mas sim pelo pedúnculo floral" (HOUAISS, 2006), sendo, portanto, órgãos vegetais que contêm reservas nutritivas, são comestíveis e adocicados, de sabor agradável. Dessa forma, no conjunto *fruta* não caberiam carrapicho nem berinjela, mesmo que esta seja comestível.

Há também que se ressaltar que nem sempre categorizamos por meio do estabelecimento de propriedades necessárias e suficientes, pois há evidências psicológicas de que somos capazes de empreender relações de inferência e deslocamentos de sentido, com base nas estruturas sintáticas. Por exemplo, é comum a colocação do desenho de uma maçã em placas de motel (de preferência, mordida), por se relacionar à questão do desejo que remonta a referenciais contidos na Bíblia.

Nessa direção, a semântica cognitiva demonstra que aprendemos categorias de nível básico diretamente, porque derivam de uma experiência perceptiva primária, que nos remontam às relações propostas por Peirce (1972), já sinalizadas nesta seção e a serem desenvolvidas num capítulo à parte (categorização). Aprendemos inicialmente uma categoria de nível básico como *bola* e, só posteriormente, por processos de metonímia, seremos capazes de incluí-la em categorias genéricas como *brinquedos,* e em particulares como *material esportivo para futebol*, bem como processos metafóricos como *bola de neve,* em construções como *os problemas se tornaram uma bola de neve*. Mas não elimina a hipótese contrária de apreensão de outros significados pela experiência social, como no caso de aprendermos algumas denominações de alimentos em cardápios (ou de bola como *veneno*).

Portanto, experiências exemplares são mais reveladoras bem como organizadoras das categorias. Pode-se afirmar que, quando classificamos, nem sempre recorremos ao estabelecimento de condições necessárias e suficientes, como aquela demonstradas no exemplo da maçã, mas buscamos referências exemplares que permitam associações. Daí a formulação da hipótese de que os conceitos se estruturam por protótipos, que se relacionam

com outros membros hierarquicamente organizados numa relação de pertinência. Inscrevem-se aqui trabalhos de Eleanor Rosch, que serão retomados mais a frente.

Ludwig Wittgenstein (1979) também problematizou a questão das categorias, partindo do pressuposto de que uma propriedade necessária não delimita uma classe. Ele usou como exemplo a palavra *jogo*, que se refere a diversas modalidades como vôlei, amarelinha, futebol... Fica difícil encontrar a propriedade comum que definiria a classe; ele cita a propriedade *divertimento* ou *competição,* mas demonstra que ambas são igualmente genéricas. Então, ele afirma que os usos de uma mesma palavra se assemelham constituindo uma família; mas lembrando que, para pertencer a uma mesma família, não é necessário que os membros compartilhem a mesma propriedade. Já Gilles Fauconnier, citado por Oliveira (2001), parte da hipótese de que na interpretação dos enunciados formamos espaços mentais, estruturas conceituais que descrevem como os falantes atribuem e manipulam a referência. Em convergência com a semântica cognitiva, Fauconnier demonstra que o significado não está na linguagem; ele entende que a linguagem é como um método que permite a identificação de uma estrutura cognitiva subjacente. Sendo assim, propõe que durante a interpretação construímos domínios ou espaços mentais nos quais estão as referências. Daí a noção de pressuposição estar associada à transferência de significados de um espaço mental a outro. Na linha do exemplo da palavra *jogo*, na expressão *Não entrarei nesse jogo de mentiras*, anula-se a possibilidade de divertimento, supostamente estabelecida num primeiro espaço mental e abre-se a um segundo espaço, associando a palavra *jogo* à noção de rede, nesse caso, de intrigas, provavelmente.

É importante reiterar que os três modelos semânticos mostram que uma abordagem não necessariamente nega a outra, mas que podem complementar-se. A categorização, escopo deste trabalho, interessa aos três modelos, mas principalmente, à *semântica formal* e à *semântica cognitiva*; à *semântica da enunciação* são necessários demais elementos que nos permitam uma análise mais abrangente do fenômeno enunciativo. Pretendemos aqui tentar demonstrar que a semântica formal, como um instrumental lógico para descrever a linguagem, é essencial para esta pesquisa, pois viabiliza uma perspectiva paramétrica do sistema linguistico; sem ela, seria impraticável a estruturação de modelos interativos informatizados que operam com uma hipótese pré-formatada de pressuposições para atender ao usuário que utiliza um sistema. Por outro lado, a semântica cognitiva, que nos interessa mais de perto, fornece a bagagem descritiva necessária ao produtor de textos, alimentador de um banco de dados, na organização das informações. E seria impossível negar, conforme discutido anteriormente, que é da semântica enunciativa que advêm muitas das possibilidades de

atribuição de significados. Mesmo porque, apesar de este trabalho enfocar a perspectiva do léxico, é na estrutura enunciativa que são constituídos os referenciais semânticos para organizações conceituais.

3.4 Campos Semânticos

Seguindo na perspectiva semântica, consideramos importante trazer as noções da teoria dos campos semânticos proposta por John Lyons (1980). Lyons filia-se às abordagens de Jost Trier que considera o vocabulário de uma língua como um sistema integrado de lexemas religados do ponto de vista do sentido. Segundo Trier, esse sistema se encontra em fluxo constante de produção e abandono de formas bem como de atribuição de sentidos, mudando de acordo com as necessidades sociointerativas. Dessa forma,

> Os campos são realidades vivas intermediárias entre as palavras individuais e a totalidade do vocabulário; como partes de um todo, partilham com as palavras a propriedade de estarem integrados numa estrutura mais vasta e, com o vocabulário, a propriedade de se encontrarem estruturados em termos de unidades mais pequenas. (TRIER apud LYONS, 1980, p. 207).

Lyons afirma que o conceito-chave para Trier é a ideia de articulação orgânica e interdependente. No entanto, Lyons diferencia campo lexical de campo conceitual, além de discutir a questão também da variação num recorte temporal (diacrônico/sincrônico). Como campo lexical, define-se o conjunto de lexemas de que uma língua se serve para, em função das relações de sentido existentes entre eles, abranger e estruturar uma área conceitual. Seria o termo que se refere à organização estrutural de uma área conceitual por um sistema linguístico particular, partindo do pressuposto de que se encontra uma substância de significado não estruturada subjacente ao vocabulário de todas as línguas. É o conjunto de lexemas que, organizados em função das relações de sentido existentes entre si, abrange uma determinada área de significação, estruturada num campo conceitual. No caso do campo conceitual, as relações de parentesco viabilizam linguisticamente as relações num campo lexical, que inclui, por exemplo, lexemas como *pai, mãe, filho*, etc.

Portanto o campo semântico difere dos dois anteriores por ser constituído pela área de significação de uma palavra ou de um grupo de palavras. Campo lexical seria o conjunto das palavras de uma língua, o próprio léxico, e campo semântico seria o conjunto dos empregos

dessas palavras, pois, em cada contexto de uso, podem adquirir nova carga semântica. Para descrever o campo semântico da palavra *luva*, por exemplo, incluiremos nele todas as possibilidades semânticas como: *luvaria, luveiro, assentar como uma luva, atirar a luva, cair como uma luva, de luva branca, deitar a luva, macio como uma luva*. Até este momento, ninguém demonstrou que todo o vocabulário esteja coberto por campos semânticos, haja vista a dinamicidade da própria língua; mas temos os dicionários analógicos que trabalham nessa direção (AZEVEDO, 1983; SARGENTIM, [19-]). Mas, certamente, podemos afirmar que essa linha vem conduzindo princípios de organização de conhecimentos; talvez o desafio maior seja mostrar qual termo da língua não se vale de um campo semântico.

No entanto, a partir da teoria dos campos semânticos podemos tomar o vocabulário enquanto sistema de cadeias parciais articuladas entre si. Pode-se afirmar, então, que o vocabulário não é um inventário rígido de componentes independentes e isolados, mas uma série de elementos combináveis que, estruturados em campos, constituem a totalidade do léxico.

Diante de tais colocações é importante situar as distinções entre sentido e significado. Mari propõe a seguinte distinção:

> [..] ao significado reserva-se uma caracterização conceitual primeira de um item lexical (de um sintagma, de um enunciado). O significado, então, nada mais é do que certa matriz de traços atômicos (recolhidos na diversidade das propostas teóricas da Semântica que vem a partir do estruturalismo), ou um amálgama de matrizes, os quais servem para caracterizar o conteúdo nocional de um determinado signo, ou de uma unidade superior a ele. (MARI, 2008, p. 62).

Essa concepção vai ao encontro das noções discutidas em torno dos campos semânticos, embora ofereça uma visão mais específica associada aos traços semânticos, levando em consideração as possibilidades de uso sob uma análise paradigmática como natureza, composição, finalidade, entre outras (matriz conceitual), bem como a perspectiva da estruturação sintática. Nesse sentido, o autor ressalta que o significado seria o "construto conceitual que antecipa os usos específicos" (p. 63), retomando, assim, a perspectiva da pressuposição já discutida neste trabalho.

Para sentido, resguardadas as dimensões flutuantes de tal discussão, Mari (2008) propõe situá-lo como modo de apresentação e percepção de um objeto, na linha de Frege, numa primeira instância, e em outra, na de Wittgenstein, que traz a dimensão associada a situações concretas de uso. A partir daí, conclui que o sentido se constitui após a significação

de um termo, pois haveria traços primitivos na composição do significado que seriam expandidos para sentido pelas condições de uso, estabelecidas em contextos diversos.

Em busca de exemplos, poderíamos problematizar a palavra *santo* utilizada na associação com o nome do indivíduo santificado que se iniciasse por vogais (Santo Antônio) – haja vista que *são* emprega-se antes de nomes iniciados por consoante (São Pedro). Na dimensão do significado, *santo* é aquele que pertence a uma religião, e que a professa de forma exemplar, sendo, portanto, canonizado. Boa parte de usos dessa palavra (santo), principalmente em contextos religiosos, assume essa significação primeira. Mas, num breve levantamento, podemos tê-la com sentido oposto (irônico), como aquele que finge ser santo ("santo do pau oco"); como algo eficaz (santo remédio); aquilo que não pode ser violado (santo sepulcro); e até mesmo os santinhos de campanhas eleitorais. Dessa forma, percebe-se a confluência das abordagens aqui já discutidas, com a de Lakoff sobre as extensões metonímica e metafórica do termo em foco.

3.5 Categorização

Estabelecidos os princípios da semântica, as noções de campos semânticos e as possíveis distinções entre significado e sentido, vamos retomar as questões relativas à categorização para fundamentar nossas discussões na perspectiva da hierarquia semântica.

Retomando a questão da categorização, vamos iniciar pelo modelo clássico proposto por Aristóteles que é considerado por alguns como a hierarquia perfeita do mundo. Nele, as categorias são definidas pelas propriedades comuns a todos os seus membros. Dessa forma, nenhum membro pode ser considerado melhor exemplo da categoria que os demais. As categorias são definidas apenas pelas propriedades inerentes aos seus membros; elas devem ser independentes de quem estiver categorizando, ou seja, ela não deve ser influenciada pelos movimentos do corpo humano, pela neurofisiologia humana, pelas capacidades humanas de percepção, pelo aprendizado, pela lembrança, pela comunicação nem por fatores sócio-culturais. A teoria clássica fornece uma explicação direta para o fato de separamos membros e não-membros de uma categoria. Os membros de uma categoria são os objetos, ou exemplares, que exibem as características necessárias e suficientes, conforme já indicado anteriormente neste trabalho, que definem aquela categoria; aqueles que não exibirem essas características são considerados não-membros, possibilitando, também, uma formação intuitiva do conceito.

Segundo esse modelo, formamos conceitos quando temos vários membros de um grupo, e determinamos as características únicas que dividem esses membros em classes separadas. Assim, fica demarcado o que constitui ou não um exemplar de uma categoria, com fronteiras claramente demarcadas. Na hierarquização das categorias, aquelas mais específicas incluem todos os atributos das categorias mais gerais. Pode-se tomar como exemplo um conceito mais específico, *canário* (definido como *emplumado, animado, bípede, pequeno*), e o superordenado (mais geral) *pássaro* (definido como *emplumado, animado, bípede).* Além da crítica ao tratamento dos conceitos como sendo definidos por um jogo de características necessárias e suficientes que eram descobertas empiricamente, o ponto frágil dessa abordagem seria ela se tratar de uma proposta de representação e não de processo.

Quando temos objetos com características mais definidas, esse modelo pode ser satisfatório; mas nem sempre é possível captar todos os aspectos significativos do comportamento conceitual, uma vez que nem sempre podemos afirmar que os conceitos dependem de uma conjunção de atributos necessários. Ademais, em qualquer circunstância que formarmos um conceito, ele terá de representar dados (traços, categorias, propriedades...) que sejam adequados a descrever/representar a classe de objetos em que se inclui. O conceito de cadeira precisa ter alguma categoria que seja comum a todos os objetos da classe (assento, encosto): nada impede, porém, que tipos específicos de cadeira possam 'deformar' alguma categoria em favor de sua inclusão na classe (como cadeira de balanço ou a cadeira do "papai", por exemplo).

Porém, conforme já demonstrado aqui, não é tarefa simples determinar os atributos que definam a maioria dos conceitos, pois toda tentativa de dividir os conceitos em atributos essenciais e auto-suficientes é fundamentalmente mal concebida, pois alguns conceitos simplesmente parecem atributos difusos ou com atributos correlatos a outros conceitos que os defina, mesmo porque, quando essas diferenças vão diminuindo, uma forma pode se sobrepor à outra, por diversas razões. Por exemplo, o campo semântico do conceito de *enganar*, pode ser trapacear, burlar, engabelar, fazer de bobo, trair, entre outras possibilidades. A questão é que a *crença em algo não verdadeiro* é um atributo que une todas elas; por outro lado, cada elemento desse campo possui um atributo mais específico: trapaça envolve má fé; burlar pode configurar uma fraude; engambelar significa livrar-se da obrigação; "fazer de bobo" pode ser uma dissimulação; e trair, cometer um adultério. Nota-se que a tentativa de categorizar esbarrou, nesse exemplo, em limites muito tênues. Diante dessa questão, pode-se dizer que qualquer forma de definição sempre esbarrará em dificuldades possivelmente intransponíveis.

Podemos, ainda, ter casos mais confusos nos quais as características estejam mascaradas por termos subordinados, bem como casos em que haja uma insuficiência geral na especificação das características definidoras para a maioria dos conceitos. Por exemplo, qual seria a categoria mais adequada para classificarmos, hoje, a palavra *moderno*? Será que os atributos necessários desse vocábulo ainda significariam algo inovador, haja vista que temos, hoje, a palavra "hodierno" para indicar algo que é atual, além da palavra "contemporâneo" ou "atualidade". Sabemos que são vários os fatores que interferem nesse processo, como, por exemplo, a idéia de tempo, de estranheza que pode integrar também o conceito. De certa forma, essa questão torna o campo semântico um princípio organizador, mesmo no caso de objetos que não se incluam em um dado conceito, possivelmente se incluirão em outros campos diferentes.

Outro ponto discutível seria a captação dos aspectos significativos do comportamento conceitual, talvez mais sensíveis à semântica enunciativa. De toda forma, o modelo clássico teve, e ainda tem, uma grande repercussão em várias áreas do conhecimento.

Numa linha mais cognitiva, podemos citar Lakoff quando afirma que

> A maioria de nossas palavras e conceitos designam categorias [...] Categorização não é um processo que deve ser estudado superficialmente. Não há nada mais básico do que a categorização para o nosso pensamento, percepção, ação, e discurso. Cada vez que nós vemos algo como "um tipo" de coisa, por exemplo, uma árvore, nós estamos categorizando. [...] A compreensão de como categorizamos é o ponto central para a compreensão de como nós pensamos, funcionamos e, consequentemente, um ponto central para a compreensão daquilo que nos faz humanos. (LAKOFF, 1987, p. 5).

Para o autor não existe nada mais básico do que a categorização para o nosso pensamento, nossa percepção, ação e fala. Todas as vezes que vemos alguma coisa como um tipo de coisa, ou como parte de alguma coisa, estamos categorizando. Isso ocorre, principalmente, pelas características oriundas das similaridades e diferenças existentes entre conceitos, dentro de determinado contexto.

A elaboração de categorias geralmente é referida como uma metodologia na qual conceitos formam novas categorias pelas características inerentes a eles próprios. Essas deficiências na definição dos atributos levaram estudiosos a refletirem sobre a lista geral de características, proposta pelo modelo clássico que a maioria - ou muitos - dos membros possuem, e comprovar que a lista é a descrição da categoria e não dos membros em particular. Talvez tenhamos aí um critério de membramento, haja vista que as noções de conceito

permaneceram na base da organização, como nos propomos a demonstrar nas subseções seguintes.

Eleanor Rosch (1975) colocou em xeque os pressupostos da visão clássica, a partir da década de 70, baseada no pressuposto de que, se no modelo clássico as categorias são definidas somente pelas propriedades que todos os membros da classe possuem, então, nenhum membro pode exemplificar a categoria melhor que outro. Rosch desenvolveu um estudo sobre as categorias naturais para o reconhecimento da estrutura hierárquica como um complexo dentro do qual conjuntos de categorias são organizados de modo a formar taxonomias, ou seja, sistemas nos quais os elementos são ligados uns aos outros por inclusão de classes. Quanto mais uma categoria apresenta um nível de inclusão elevado no sistema taxonômico, mais a categoria inclui outras categorias, e igualmente, mais elevado é o nível de abstração da categoria. Essa noção de abstração de uma categoria definida segundo a quantidade de categorias que ela pode incluir, corresponde a uma experiência subjetiva, pois ela demonstrou que os conceitos não são fixos, mas formam categorias cujos membros apresentam semelhanças de família. Segundo a autora, não existe um conjunto de critérios fixos, necessários e suficientes estabelecidos por um conceito, para que algo possa pertencer a uma categoria. Os membros de uma categoria não são criados de forma igual, e os melhores exemplos de uma categoria são chamados de *protótipos*. O membro de uma categoria é percebido como mais típico se ele possui um número maior de atributos comuns com os outros membros da categoria à qual ele pertence *(semelhança familiar)*. De acordo com a autora, para a maioria das categorias naturais forma-se um protótipo a partir do qual se categorizam os objetos encontrados no dia-a-dia. Esse protótipo é uma representação abstrata que possui os valores médios das características, definindo sua pertinência a uma dada categoria. Nota-se, portanto, que há limitações dessa linha teórica quanto ao reconhecimento de categorias.

Essa *visão prototípica* no modo de encarar a formação de conceitos ficou também conhecida como visão probabilística. De acordo com essa teoria, os indivíduos, ao entrarem em contato com os múltiplos e variados exemplos de uma categoria, abstraem deles apenas as características que ocorrem com maior freqüência, que, de certo modo, nos remeteria ao modelo clássico. Mas, a partir dessas características abstraídas, formam uma representação sumária da categoria, denominada protótipo. Uma vez formado, o *protótipo* é utilizado para decidir a inclusão ou exclusão de membros na categoria. Se algum exemplo é suficientemente similar ao protótipo, ele será classificado como membro dessa categoria; se não o for, será rejeitado. Rosch propõe, portanto, uma concepção prototípica ou probabilística na articulação

de conceitos. A autora baseou-se em Wittgenstein, que sugeriu o princípio de semelhança entre categorias, formando famílias, de modo que cada item tivesse um ou mais elementos em comum com alguns outros, mas que nenhum elemento precisasse ser comum a todos os itens. Segundo Rosch, Wittgenstein rejeita a noção de que todos exemplares de um conceito têm uma série de atributos em comum, como defende a teoria clássica, e propõe que aquilo que une os exemplares dentro de um mesmo conceito é a semelhança familiar entre os mesmos.

Rosch (1975) sugeriu, então, a organização de categorias em torno de um conjunto de propriedades ou conjuntos de atributos correlacionados que são característicos ou típicos, rejeitando, assim, a noção de atributos definidores. No entanto, sabe-se que os elementos que decidem sobre a inclusão de um membro em uma família são atributos definidores desse elemento, pois um elemento só pode pertencer a uma família se houver ao menos "um" atributo que seja comum a algum elemento que define a inclusão nessa família. Dessa forma, pode-se afirmar que, se no modelo clássico há dificuldades para explicar os efeitos da tipicidade, outros modelos como o da Rosch não explicam satisfatoriamente a organização das categorias. Poderíamos supor que existam categorias que se ajustam melhor ao modelo clássico e outras, ao modelo de protótipo. Uma combinação do aspecto central de um termo com um processo de identificação do conceito pode ser ainda considerada a mais eficiente na categorização como um processo cognitivo.

Essa concepção prototípica, da mesma forma que a clássica, identifica, então, conceitos com conjuntos de propriedades. Essas propriedades constituem um protótipo, de tal maneira que a aplicabilidade de um conceito a uma entidade depende do grau de similaridade que existe entre a entidade e o protótipo do conceito. Portanto, há um enquadramento de exemplares mais ou menos próximos do protótipo, com limites pouco definidos. A autora demonstrou a existência de categorias difusas, uma alternativa à categoria em si. Dentro dessa perspectiva, o termo "cadeira", por exemplo, tem certas propriedades funcionais que formam uma parte específica dessa categoria. Rosch realizou trabalhos experimentais para observar os efeitos da tipicalidade, isto é, do conceito que representa o objeto como sendo aquele que reúne todas as principais características que o tornam representante de uma categoria (um caráter de tipicalidade). As pesquisas de Rosch sobre as categorias naturais permitiram uma sistematização da definição das categorias mal-delimitadas, substituindo, assim, o ponto de vista clássico no estudo dos conceitos mentais.

Numa tentativa de melhor esclarecer, pode-se afirmar que o princípio fundamental desse modelo sustenta que as categorias são organizadas em torno de protótipos centrais. Um item é considerado como membro de uma categoria não por se saber que ele possui um

determinado atributo ou não, mas por se considerar o quanto as dimensões desse membro se aproximam das dimensões ideais para ele. Em outras palavras, um exemplo representativo de uma classe seria aquele que compartilhasse com os outros membros da categoria do maior número de características e que, por outro lado, compartilhasse de poucas características (ou nenhuma) com elementos provenientes de fora da classe. Esta talvez seja uma das limitações da teoria dos protótipos. Um canário, por exemplo, é protótipo de um pássaro pelas características similares entre os exemplares da classe como tamanho, tipo de plumagem, bico, dentre outras; mas uma galinha, apesar de também ser considerada ave -por outros critérios. reordenadores -, não pode ser considerada como um exemplar da categoria pássaros.

Rosch (1975) mostrou que quanto mais prototipicamente um membro de uma categoria é julgado, maior será a quantidade de características comuns que ele terá com os outros membros dessa categoria, e menos atributos comuns com membros de categorias contrastantes. Em resumo, protótipos são justamente aqueles membros de uma categoria que mais refletem a redundância da estrutura de uma categoria como um todo. De acordo com o modelo de protótipo, conceitos são representados por um grupo de características. Um novo membro é categorizado como um tipo de conceito se é suficientemente similar ao seu protótipo.

O agrupamento de conceitos em uma dada categoria se daria, segundo a pesquisadora, não pela alternância dos traços binários, ou seja, pertence ou não pertence, mas pela semelhança com o protótipo, em que o membro condensasse os traços mais característicos da categoria. Na mesma linha exemplar, um pintarroxo é uma ave mais prototípica (na propriedade de voar) do que uma galinha ou um pinguim para a categoria pássaro; o que o faz ser mais facilmente reconhecido como uma ave (asas e penas) e menos provável que seja classificado erroneamente como membro de outra categoria (tal como um mamífero ou um peixe).

Os pressupostos da teoria clássica confrontados pela autora residem na definição absoluta das categorias as quais são estabelecidas em termos de um conjunto de traços necessários e suficientes. Então, segundo esse modelo, um elemento pertence a uma categoria determinada somente se possuir todos e cada um dos traços que a definem; a falta de algum desses traços significaria a sua exclusão automática da categoria. Além disso os traços são binários, conforme citado acima, as coisas possuem ou não possuem um traço, pertencem ou não pertencem a uma categoria; estas com limites bem definidos de pertencimento. Assim, negam-se os casos ambíguos, pois todos os membros de uma categoria têm o mesmo status. Nesse ponto as duas teorias se diferenciam, pois se os elementos [+] pertencem ou [-] não

pertencem a uma categoria, não existiriam os graus de pertinência, isto é, não haveria elementos prototípicos que sejam membros melhores do que outros, conforme proposto por Rosch. Se conhecemos a categoria a que um elemento pertence, podemos perceber, com certeza, que alguns de seus atributos, condições necessárias e suficientes para a pertinência à categoria são concorrentes. Se a estrutura das categorias configura um conjunto de semelhanças de família (Wittgenstein) e não em traços mínimos essenciais compartilhados e se, conforme já sinalizado em exemplos anteriores, as categorias tendem a se fundirem em outras em razão de possuírem limites difusos; mesmo os conhecendo, isso não garante que possamos estabelecer, com absoluta certeza e unanimidade, que os membros pertençam ou não a ela. Por isso, a teoria clássica não é suficiente para demonstrar as categorias, mesmo porque novos membros são criados, enquanto outros caem em desuso. Concluímos, portanto, de acordo com Rosch, que as categorias são estruturas difusas e de uma grande variabilidade. Mas, de acordo com Wittgenstein, somos capazes de diferenciá-las não pelos traços abstratos e essenciais, supostamente comuns a todos os membros, mas sim mediante um ensinamento baseado em exemplar; no entanto, ele não desenvolveu sistematicamente as idéias de *semelhança familiar* e de ensinamentos mediante exemplares. Rosch foi um entre outros pesquisadores que deram forma a uma perspectiva semântica alternativa à teoria clássica.

A visão prototípica de Rosch com a opção pelos atributos característicos permite os reagrupamentos e dá uma maior flexibilidade ao estabelecimento das categorias. É sabido que o protótipo, em si, já traz um conjunto de categorias que servirá de base para a decisão de pertinência de membros em um conjunto. Mas a autora admite que os membros de uma categoria não a representam igualmente bem, sendo que aqueles que mais se assemelham ao protótipo constituem melhores exemplos do que os menos semelhantes. Por outro lado, defende que os limites das categorias são mal definidos, uma vez que os membros de uma categoria podem, em maior ou menor grau, partilhar características de outras categorias. Nessa versão padrão formulada por Rosch, o protótipo é considerado o exemplar mais adequado, o melhor representante ou caso central de uma categoria; daí tem-se que os conceitos são representados por um ou alguns exemplos particulares, em vez de por um protótipo único. Então, o exemplar idôneo comumente associado a uma categoria contém os aspectos *graduais de proximidade ao protótipo* dentro da categoria que são definidos pela frequência de uso ou atribuição entre os sujeitos (recorrência, perfil estatístico).

Em nossa convivência diária, todos nós vamos armazenando experiências exemplares sob as quais formamos nossos modelos de compreensão; a partir deles formulamos nossas hipóteses e interagimos em sociedade. Mas nossos mecanismos de previsão nem sempre são

eficazes e, por isso, estão em constante atualização e passíveis de reformulação. Provavelmente nunca *erramos* no reconhecimento de uma árvore, de um pássaro, de um peixe; porém, podemos *errar* no reconhecimento de uma sibipiruna, de um melro, de um mandi. Talvez, seja este o ponto frágil de qualquer tentativa de categorização; no entanto sua função paramétrica é inegável.

Mas, para Rosch, pela sua perspectiva e por sua área de formação, a psicologia, o protótipo atua como ponto de referência cognitiva (*cognitive reference point*) para os processos de classificação dos elementos de nossa experiência. Seus experimentos de verificação dos prototípicos derivados do modelo permitiram que chegasse às seguintes conclusões:

a) Os membros prototípicos são categorizados mais rapidamente que os membros não-prototípicos.

b) Os membros prototípicos são os que as crianças aprendem primeiro.

c) Os membros prototípicos são os primeiros mencionados, quando solicitamos aos falantes que listem todos os membros de uma categoria.

d) Os protótipos servem de ponto de referência cognitiva.

e) Geralmente, quando o que se pede é a enumeração dos primeiros membros de uma categoria, os protótipos aparecem mencionados em primeiro lugar.

A partir de então, Rosch (1978) estabelece a existência de uma dimensão horizontal, ou seja, uma organização intercategorial hierárquica. A autora propõe três níveis dentro das categorias: supraordenado, básico e subordinado. Ao estabelecer essa hierarquia, ela determinava um nível privilegiado dentro da categoria: o nível de base. Por exemplo:

SUPRAORDENADO	Arma	Fruta	Móvel
BÁSICO	Arma de fogo	Maçã	Cadeira
SUBORDINADO	Revólver	Maçã argentina	Cadeira de balanço

Quadro 02- Níveis de categorias, segundo Rosch.
Fonte: Dados da pesquisa

Utilizamos essa hierarquia para representar, mentalmente, as relações de inclusão de classes nas categorias (*cadeira* dentro da categoria *móveis*)[15]. O nível básico é aquele em que os conceitos têm maior número de atributos distintivos e é o mais econômico cognitivamente (há menos partilha de atributos de conceito). Em seus estudos, Rosch pediu a alguns falantes

[15] Entende-se por categoria o conjunto de membros representados pelo conceito, o que configura uma classe.

que listassem todos os atributos de itens em cada um dos três níveis existentes dentro de uma hierarquia (ex. *móveis - poltronas - espreguiçadeiras*). Descobriu que poucos atributos foram listados para as categorias supra-ordenadas (como *móveis*) e muitos atributos foram listados para as categorias nos outros dois níveis. No nível inferior, atributos muito semelhantes foram listados para as diferentes categorias. As categorias de nível intermediário (ou básico) (ex. *cadeira*) salientaram-se por um equilíbrio entre a informatividade (número de atributos de um conceito) e a economia (resumo dos atributos importantes que distingue as categorias entre si). Ou seja, no nível superior, sacrifica-se a informatividade e, no nível inferior, sacrifica-se a economia.

Essas abordagens de Rosch serão úteis ao nosso estudo, pois observamos que, associando os conceitos por características semelhantes, constituímos famílias, já configurando campos semânticos, conforme já discutido neste trabalho, que permitem operar com relações de inclusão ou não. Se associarmos, por exemplo, as palavras bola, chuteira, campo, time, jogador, é porque as consideramos com características similares que constituem o campo semântico *futebol*. Tal princípio é norteador das normas internacionais de organização de informação, como será exposto na próxima seção (a das teorias documentárias). Além disso, a estruturação dos níveis acima discutidos pode viabilizar critérios mais consistentes para a organização das informações.

Finalizando esta seção, gostaríamos de ressaltar que, apesar desse modelo de protótipo ser apresentado como uma alternativa ao modelo clássico, muitos críticos questionam se ele pode realmente substituí-lo. O protótipo é ainda uma categoria muito escorregadia. Um dos argumentos contrários à Rosch é que nem todas as categorias possuem características prototípicas, apresentando alguns limites ao grau de generalidade do modelo, pois a existência de grandes flexibilidades em pertencer ou não a determinada categoria mais abstrata relativiza a estrutura de protótipo, diferentemente das categorias concretas. Outro ponto é que o modelo de protótipo fica limitado ao conhecimento que as pessoas possuem sobre as relações entre categorias, uma vez que alguns atributos variam mais facilmente do que os outros. Mais do que isso, o modelo de protótipo não explica por que as categorias são coerentes. Poderíamos supor que haja algum mecanismo de semelhança que seja responsável pela coerência da categoria, ou seja, os elementos são agrupados em categorias, porque têm certos atributos comuns. Mas este não é o único mecanismo que utilizamos, podemos formar categorias que possuem um tênue embasamento com atributos compartilhados, mas que, mesmo assim, são coerentes. Mais à frente, discorreremos sobre abordagens complementares ao conceito de categorização.

3.6 Conceito e Categorização

Antes de seguir em frente, faz-se necessário estabelecer uma distinção entre as noções de conceito e de categoria, muitas vezes utilizadas como sinônimos.

Podemos afirmar que o *conceito* seria a representação de um objeto pela reunião de suas características gerais, e expresso por termos ou predicados. Pelo conceito somos capazes de agrupar características comuns ao um conjunto de seres da mesma espécie, diferenciando-os dos seres pertencentes a outra(s) espécie(s).

Mas descrever um *conceito* é uma tarefa que pressupõe a relativização de algumas variantes, como na perspectiva geográfico-ambiental que implica numa variabilidade de objetos, nas mudanças de estado, de forma e de aparência que podem ocorrer. Além disso a essência do conceito, conforme já dito acima, é o princípio da similaridade, de aspectos comuns entre os objetos. No entanto, há que se ressaltar que os conceitos podem funcionar como princípios de classificação, pois sua função é também possibilitar a recognição, com a (re)classificação de exemplares pela diferenciação entre os objetos. Levando em consideração tais variantes, situamos o proposto em Houaiss, segundo o qual *conceito* é "uma noção abstrata contida nas palavras de uma língua, para designar, de modo generalizado e, de certa forma, estável, as propriedades e características de uma classe seres, objetos ou entidades abstratas". (HOUAISS, 2006).

Porém para nos aprofundarmos na noção de conceito, consideramos relevante abordar brevemente os pressupostos de Peirce (1972), alguns deles já expostos no início desta seção, e suas reflexões sobre as origens do conceito. Para ele, antes de o conceito representar, por convenção simbólica, um conhecimento sobre a realidade, os signos são produzidos pela mente e sua relação com o ambiente, ao acaso. Peirce afirma que o acaso é uma força presente no universo; isso explicaria a contínua evolução do conhecimento humano e a diversidade de elementos passíveis de estudos. Ele também ressalta a perspectiva da continuidade, do constante movimento que teve como origem uma série indefinida de fenômenos. Podemos constatar no nosso cotidiano, quando atualizamos qualquer um de nossos conhecimentos, sempre que nos deparamos com uma realidade não prevista antecipadamente, como a força do acaso mostra-se no contínuo da vida. Nesse sentido, as ideias que ainda não representam nenhum objeto concreto são potenciais e passam a se misturar com ideias semelhantes, cruzamentos de fenômenos, até se tornarem uma regra mental, resgatando um conhecimento sobre o objeto associado. Sempre que os conceitos científicos, por exemplo, são alterados em

função de novas informações, eles não são aceitos imediatamente, pois isto contraria o hábito mental já sedimentado. É necessário um tempo de validação até que seja possível interpretar o dado objeto segundo algum parâmetro determinado. Dessa forma, mesmo o conceito científico tem algo de imprevisto, de incontrolável, constantemente passível de atualizações.

Numa primeira instância, o conceito teria uma função organizadora, que permitiria conhecer o real ou o que poderia ser uma realidade para o senso comum e o discurso científico. Um conceito bem estruturado seria a segurança de uma teoria científica, sustentada por um arranjo conceitual bem delimitado. Pelos conceitos podemos falar dos fenômenos, sem recorrer a eles para comunicar nossas ideias, além de expressar as relações empíricas entre fatos que ocorrem, ora ao mesmo tempo, ora em um espaço de tempo que nos autoriza a indicar que alguns fatos são procedentes de outros anteriores. Os conceitos nos permitem avaliar o que é essencial em uma dada realidade; atribuímos, também, a eles a expectativa de que algumas coisas são regulares e podem assim ser representadas. Desse modo, dadas *certas* regularidades delimitáveis, podemos crer na eficácia de um conceito.

Não é tarefa simples delimitar um conceito, pois nos deparamos com desvios que alargam ou restringem sua abrangência. Um conceito não é uma unidade ou uma quantidade definida de conhecimento, pois é o produto de uma relação habitual que temos com a realidade e com a comunicação intersubjetiva. Se fosse possível alcançar o significado último de um conceito, teríamos um engessamento da capacidade de produzir conhecimentos novos.

Segundo Vigotsky (2005), ao longo do desenvolvimento cognitivo, a formação de conceitos passa por três fases básicas: a primeira é o *Sincretismo,* durante o qual a criança não forma classes entre os diferentes atributos dos objetos; ela apenas os agrupa de forma desorganizada formando amontoados. Uma criança, nesse período, quando solicitada a formar grupos com diferentes objetos (plantas, animais, objetos de cozinha etc.), poderá colocar juntos objetos que não possuem relação entre si como, por exemplo, animais e objetos de cozinha. Nessa fase, ela agrupará ao acaso ou por contiguidade no tempo ou no espaço. A segunda fase é o *Pensamento por Complexos,* durante a qual o agrupamento não é formado por um pensamento lógico abstrato, mas sim por ligações concretas entre seus componentes que podem ser os mais diferentes possíveis. Nessa fase, a criança pode, por exemplo, agrupar por qualquer relação percebida entre os objetos, ou por características complementares entre si. Num estágio mais evoluído dessa mesma fase, ela começa a se orientar por semelhanças concretas visíveis e formar grupos de acordo com suas conexões perceptivas. A criança, nesse estágio, é capaz de agrupar os animais em um grupo e as plantas em outro. Nesse estágio, o terceiro, denominado *Pseudoconceito,* os resultados alcançados são semelhantes aos obtidos

no pensamento conceitual. No entanto, o processo mental empregado não é o mesmo que ocorre no pensamento conceitual.

No percurso do desenvolvimento, de acordo com Vigotsky, o uso das formas de pensamento mais primitivas (sincréticas e por complexos) diminui gradualmente e começam a se formar os conceitos. A formação de conceitos exige abstração, isolamento de elementos e o exame dos elementos abstratos separados da experiência concreta, cuja representação é viabilizada pela linguagem que é o instrumento do pensamento conceitual.

Dessa forma, podemos afirmar que um conceito é uma ideia, um quadro mental sobre o grupo de coisas que têm algo em comum. Os conceitos são formados pela generalização da experiência de cada um sobre um particular assunto ou coisas; é um processo de classificação, de enumeração de características. Uma criança, por exemplo, forma o conceito de cachorro, a partir do seu cão de estimação, ao descobrir que ele tem quatro pernas, pelo longo, orelhas alongadas e cauda; além de perceber que é seu amigo. Baseado nessa noção, ao ver outro cão, é capaz de identificá-lo, classificando-o por características que já consegue reconhecer. À medida que vive novas experiências com cães, seu conceito vai se modificando, completamentando-se, até mesmo com o domínio da forma lexical correspondente.

Portanto, conceito seria um tipo de representação, fruto da experiência sociointerativa, cuja representação cognitiva constitui o núcleo e a especificidade das diferentes línguas, pois cada uma pode possuir uma forma de representação léxica. Ampliando essa dimensão, pode-se, inclusive, afirmar que as diferenças lexicais entre as línguas, por exemplo, podem ser resultantes de (ou resultar em) conceitos diferentes. Mesmo resguardada a dimensão histórica de uma língua que carrega em seu bojo traços de sua origem, como no caso de povos colonizados como nós, notam-se diferenças não só semânticas como também sintáticas entre povos que utilizam uma mesma língua.

Já por *categoria* entende-se o conjunto de membros representados pelo conceito. Dessa forma, quando falamos do *conceito alimento*, estamos nos referindo a uma representação mental; quando falamos da *categoria alimento*, estamos nos referindo a todos os tipos de alimentos que já existiram, existem atualmente e existirão no futuro.

Sabemos que a categorização é essencial à atividade cognitiva. De acordo com Mari,

> Categorizar é, no plano perceptivo, executar um conjunto de operações orgânicas, capazes de possibilitar um conjunto de operações (neurofisiológicas), capazes de possibilitar a compreensão de um dado objeto, de uma dada ação; no plano metalinguistico-representativo, categorizar é explicitar algum tipo de procedimento suficiente para justificar o que conhecemos – uma teoria, por exemplo. (MARI, 2002, p. 74)

Entende-se, portanto, que haja uma perspectiva de cunho cognitivo, sensório-perceptivo, e outra decorrente da nossa necessidade de representar em algum padrão racional o que é registrado pelos nossos sentidos. Essa dualidade explica, talvez, como nossa memória semântica se organiza em categorias, conforme mencionado anteriormente, possibilitando assim a articulação entre os referenciais necessários ao processamento textual no ato enunciativo.

Mari (2002) baseia-se nos postulados de Lakoff, citado na seção anterior, que discorreu a importância da categorização, conceituando-a genericamente como uma forma de razão, e afirma que qualquer espécie de raciocínio, do mais simples ao mais complexo, nos leva a realizar uma categorização. Segundo ele, a categorização é o meio de que dispomos para ordenar o nosso comportamento, e sem o qual "seríamos incapazes de identificar quaisquer repetições, quaisquer regularidades estruturais ou funcionais e a nossa experiência seria uma sucessividade de atos desconexos". (LAKOFF apud MARI, 2002, p.71).

Lakoff retoma a Teoria dos Protótipos, proposta por Rosch, e situa a categoria como estrutura de atributos (traços) graduáveis e com limites difusos. Então os graus de relevância dos atributos dentro das categorias se organizam em torno de um centro cognitivo exemplar, no qual tais entidades são ordenadas, e também incluídas ou excluídas, pelos falantes, produzindo o que se conhece como protótipo.

Categorizar é, portanto, agrupar entidades (objetos, ideias, ações, etc.) por semelhança. Dessa forma, podemos afirmar que as categorias possuem uma estrutura interna, centrada em protótipos ou estereótipos, conforme já esclarecido, e outros exemplares são definidos como mais ou menos periféricos, dependendo do grau de compartilhamento de características cruciais com o protótipo central. Trata-se de um mecanismo que simplifica a interação individual com o ambiente, facilitando o armazenamento e a recuperação da informação, além de reduzir a demanda da memória humana. Na categorização, o reconhecimento das similaridades e diferenças leva à criação de um conhecimento novo, pelo agrupamento de entidades, de acordo com as similaridades e diferenças observadas. Segundo Roch (1975), o objetivo da categorização é reduzir as infinitas diferenças entre estímulos para proporções utilizáveis comportamentais e cognitivas. Sendo assim, a categorização básica é o nível mais geral e inclusivo no qual as categorias podem delinear estruturas correlacionais do mundo real. Por fim, é um mecanismo de ordenação para estabelecer ordem por meio de fenômenos agrupados e relacionados; em outras palavras, divide o mundo da experiência em grupos ou

categorias cujos membros sustentam algumas semelhanças imediatas dentro de um dado contexto.

Sabemos que, para organizar o conhecimento de qualquer área, da forma de representação à recuperação desse conhecimento, estudam-se, primeiramente, os conceitos que compõem esse campo do conhecimento e as relações entre eles. Mas a influência do contexto de uso é essencial na categorização, de modo que, qualquer descrição individual na classificação que o desconsidere, pode mostrar-se muito limitada. Portanto, categorias e hierarquias de categorias podem resultar em critérios mais seguros para organizar o conhecimento e facilitar a adequada recuperação. O processo de categorização certamente está em ênfase em áreas da ciência da informação e da ciência cognitiva, no que concerne à estratégia de se classificarem as coisas, fatos e fenômenos existentes, que são objetos da cognição.

Retomando a questão da estruturação de conceitos, podemos afirmar que ele visa ao compartilhamento do conhecimento. Um conceito em si é um elemento de significação que permite a comunicação, no nosso caso, verbal. Ele denota um referente com suas características e estabelece uma equivalência entre a realidade e a representação, com o propósito de delimitar o uso de tal termo em um discurso. Essa perspectiva nos encaminha às bases terminológicas que tratam dos conceitos como elementos da estrutura do conhecimento, na medida em que estão representados no léxico de uma língua. Segundo Benveniste,

> A constituição de uma terminologia própria marca, em toda ciência, o advento ou o desenvolvimento de unia conceitualização nova, assinalando, assim, um momento decisivo de sua história. Poder-se-ia mesmo dizer que a história particular de uma ciência se resume na de seus termos específicos. Uma ciência só começa a existir ou consegue se impor na medida em que faz existir e em que impõe seus conceitos, através de sua denominação. Ela não tem outro meio de estabelecer sua legitimidade senão por especificar seu objeto denominando-o, podendo este constituir uma ordem de fenômenos, um domínio novo ou um modo novo de relação entre certos dados. O aparelhamento mental consiste, em primeiro lugar, de um inventário de termos que arrolam, configuram ou analisam a realidade. Denominar, isto é, criar um conceito, é, ao mesmo tempo, a primeira e última operação de uma ciência. (BENVENISTE apud KRIEGER, 2004, p. 17)

Benveniste ressalta o aspecto cognitivo da representação do conhecimento científico, mostrando a relevância da função denominativa do componente lexical das línguas para a fixação e a circulação do saber científico e técnico. Um léxico temático torna-se um componente linguístico para a configuração de campos semânticos de comunicações especializadas, a fim de transmitir conteúdos próprios de cada área.

No início desta seção, levantamos alguns dos processos de construção de conceitos e ressaltamos como esse processo é um produto social, mediado pela língua. Saussure, citado por Genouvrier e Peytard (1974), reforça esse caráter social, quando conceitua língua como um conjunto de *convenções necessárias, adotadas pelo corpo social* de modo a permitir o exercício da faculdade de linguagem entre os indivíduos. Genouvrier e Peytard ressaltam, porém, que a distinção saussureana impõem problemas complexos, haja vista que a língua é um modelo abstrato que ninguém domina totalmente. Os autores demonstram, ainda, o quanto deve ser relativizada a idéia de que em um dicionário estaria registrado todo o léxico de uma língua.

Cada um de nós emprega uma parte considerável do léxico de sua língua materna, o que torna possível o processo de compartilhamento. Dessa forma, uma relativa precisão conceitual é condição necessária para um intercâmbio comunicativo eficaz na transmissão de conhecimentos. Portanto, o conjunto de unidades lexicais típicos de uma área, por exemplo, determina e viabiliza a circulação do conhecimento. No entanto, a estabilização de um conceito leva em conta contratos de ordem sociolinguística e pragmática em benefício de um dado campo semântico, que interferem nas realizações sintagmáticas, e geram, consequentemente, um relativo controle.

Essa necessária categorização exige situar-se em um contexto para que as informações adquiram sentido de comunicação. Por isso, a hierarquização de conceitos é tão importante para representar o conhecimento detido por um indivíduo, indicando a relação entre conceitos mais gerais e conceitos mais específicos, de acordo com o domínio do conhecimento em questão. Trata-se da forma como os conceitos são organizados na própria estrutura cognitiva, ou seja, representam o modo como o conhecimento está armazenado e é compreendido.

Segundo Pierre Levy (1993), ao se adotar a hierarquia entre conceitos como princípio de organização de mapas conceituais, possibilita-se maior facilidade de aprendizagem de novos conhecimentos, já que estudos sobre a compreensão de textos mostram que, para captar seu conteúdo corretamente é indispensável que os leitores apreendam a macroestrutura conceitual presente na escrita. A hierarquia também pode limitar a navegação e flexibilidade de formar novas conexões entre os conceitos, como ocorre, por exemplo, em buscas feitas em dicionários, que são organizados alfabeticamente e não permitem possíveis conexões com outros conceitos; no entanto, dicionários enciclopédicos costumam remeter a conceitos correlatos. Tem-se, portanto, que uma estrutura de conceitos sem uma hierarquia que indique quais são os conceitos centrais e mais importantes não seria capaz de traduzir a estrutura

cognitiva e de representar o domínio que um indivíduo possui de um conhecimento, como uma estrutura textual nos permite, em princípio, inferir. Dessa forma, as limitações da hierarquização são minimizadas diante da possibilidade de desorientação causada pela inexistência de uma organização dos conceitos, segundo sua generalidade/especificidade.

A teoria clássica sobre a formação de conceitos tradicionalmente empregou conceitos precisos descritos por regras lógicas definidas. Uma regra lógica trata da ligação que une os valores de duas dimensões, sendo mais usada a regra conjuntiva (e), disjuntiva (ou), e condicional (se-então). Então, um conceito é definido por uma característica ou uma combinação de características feitas através de certas regras de combinação. No entanto, outros modelos vêm sendo discutidos. Madeira (1987), por exemplo, acredita que as categorias podem ser estudadas sob o enfoque artificial ou natural. Segundo o autor, as categorias naturais são aquelas determinadas pelas culturas ou pelo próprio sujeito, através do contato cotidiano com os objetos do mundo natural. Elas são organizadas hierarquicamente em taxonomias e definidas mais em relação a sua dimensão contínua do que a sua dimensão descontínua. Largura e cor, por exemplo, são dimensões contínuas, pois podem caracterizar determinado objeto, diferenciando-o em seus atributos e possibilitando sua inclusão em uma dada categoria. Em uma categoria natural, os membros não são todos iguais, havendo alguns que se mostram mais típicos do que outros da mesma categoria, além das características não serem perfeitamente independentes umas das outras, como nas categorias artificiais. Percebe-se, portanto, que o princípio de ordenação lógica ainda permeia as relações, embora haja uma vinculação sociocultural que estabelece os vínculos entre os conceitos.

3.7 Abordagens complementares sobre categorização

Conforme já citado, o filósofo Jerry Fodor (1983) questiona a existência de estados mentais e sua eficácia causal, sem acreditar que existam duas substâncias, mente e matéria, interagindo uma com a outra. Em sua abordagem, ele acredita que deve haver uma linguagem do pensamento, *a linguagem da mente*. Se os sistemas cognitivos envolvem representações, se as operações cognitivas envolvem a manipulação de representações simbólicas, então essas representações devem existir em algum lugar e serem manipuladas de alguma maneira. Ele ainda afirma que a linguagem do pensamento deve ser um veículo extremamente rico para poder executar os muitos processos cognitivos, como a percepção, raciocínio, aprendizagem

da língua e de valores semelhantes, que os seres humanos são capazes de fazer. Portanto, as representações mentais devem ser concebidas como estados individuais com estrutura e identidade próprias, como fórmulas transformadas por processos computacionais ou de pensamento. Para tanto, propõe uma *teoria representacional da mente*, de acordo com a qual os estados intencionais (crenças, desejos, etc.) possuem uma relação com as proposições que formam a linguagem mental interna. Segundo o autor, essa linguagem mental possuiria propriedades análogas às das linguagens lógicas; sendo assim, existiriam estados mentais que são as instâncias (os *tokens)* dos tipos (*types)* invariantes dessa linguagem - analogamente à relação instância/tipo presente na dualidade sintaxe/semântica existente em lógica formal. Quanto aos processos mentais, eles são sequencias causais de estados mentais, assim como as relações entre as instâncias das representações mentais que são uma instância das relações sintáticas que ligam os diferentes tipos da linguagem formal interna. De acordo com Fodor, "um fluxo de pensamento é uma sequencia causal de tokens de representações mentais que exprimem as proposições que são objeto dos pensamentos."(1987, p.17).

Outra teoria que discute essa questão é a dos modelos mentais de Johnson-Laird (1980) a qual diz que nossa habilidade de explicar a realidade está intimamente relacionada com nossa compreensão daquilo que é explicado, e para compreender qualquer fenômeno ou estado de coisas, precisamos ter um modelo funcional dele. Esses modelos são análogos aos eventos que acontecem no mundo real, embora incompletos e nem tão fiéis; mas permitem ao sujeito compreender fenômenos e eventos, atribuir causalidade, tomar decisões e controlar sua execução. Alguns desses modelos são adquiridos através da interação com o meio e com outras pessoas, pelas experiências sensoriais; outros, através da transmissão de conhecimento e da cultura. Segundo o autor, se temos um modelo do fenômeno na cabeça, geralmente esse modelo tem uma estrutura similar ao processo que ele modela. Johnson-Laird afirma que os modelos mentais são mais simples que os reais e podem não ser completos ou técnica e cientificamente corretos, mas existem para explicar ou ajudar a entender situações do mundo real. Tais abordagens corroboram a perspectiva cognitiva de todo o processo que certamente embasa as propostas de organização do conhecimento.

Benjamin Cohen e Gregory L. Murphy, em artigo intitulado "Models of concepts" (1981), assim como Rosch (1978), discutem a noção de conceito sob a perspectiva da inclusão. Segundo eles, o conceito é a representação de parte do conhecimento que é usada para categorizar e compreender um domínio de objetos, eventos ou processos. Eles retomam Frege e associam o conceito a um conjunto de implicações predicativas, configurando-o como um construto lógico. Essa fundamentação lógica insere a pesquisa semântica em duas

dimensões: uma extensional e outra na teoria dos conjuntos. Para esses autores, um conceito é modelado como um conjunto de condições para a definição de um membramento na classe designada por um termo. Assim como Fodor, os autores acreditam numa modelagem a partir de estruturas computacionais para a representação do conhecimento. Pode-se afirmar que, numa primeira instância, eles também dialogam com Rosch na ideia de família. Os autores recuperam a teoria clássica, na qual cada conceito corresponde a algum conjunto ou a uma coleção de entidades, cujo membramento se faz para todos ou para nenhum. Ao definir instâncias, tem-se a formação de palavras e a formação de conceitos, sendo que a primeira refere-se à aplicação de regras da língua e a segunda ao processamento mental, viabilizado pela língua. As dimensões dessa categorização se definem, no plano representacional, pelo traço que o torna membro, partindo do pressuposto de que qualquer conceito possa ser decomposto em partes menores, recorrentes na construção de outros conceitos; já no plano extensional, define-se pelo membro do conjunto (pertinência clássica), que mantém relações inferenciais, hierarquizadas com demais membros do conjunto.

Segundo Cohen e Murphy (1984), na dimensão representacional, um conceito é modelado por uma lista finita de nomes ou descrições primitivas de exemplares individuais ou por uma definição baseada numa lista finita de condições definidoras ou traços que são partilhados por todos os exemplares, conforme já citado. No entanto, a perspectiva da pertinência e não pertinência é resultante de um critério de valor binário, no qual os traços podem ser decorrentes da experiência sensível.

Daniel N. Osherson e Edward E. Smith, em artigo publicado na “Cognition”, vol. 09, em 1981, intitulado “On the adequacy of prototype theory as a theory of concepts”, afirmam que a importância conceitual da teoria dos protótipos (TP) associa-se ao fato de que as entidades não se submetem, com precisão, a uma extensão conceitual nem dela decorrem concisamente. Ou melhor, permitem avaliar uma classe como um contínuo de gradação entre seus membros. Nesse caso, um objeto materializa um conceito apenas na condição de ser ele semelhante ao protótipo de um conceito; a fronteira entre membramento e não-membramento na extensão de um conceito é, portanto, difusa (FST-Fuzzy Set Theory), opondo-se à TC (teoria dos conjuntos). Os autores consideram dois aspectos conceituais relevantes para adoção da teoria dos protótipos (TP): o primeiro relativo à combinação de conceitos, ao processo pelo qual os conceitos complexos são formulados a partir de outros mais simples; o segundo refere-se às condições de verdade para o pensamento, às circunstâncias nas quais um pensamento correspondente a uma proposição declarativa é verdadeiro. O protótipo é concebido como um membro médio da categoria, porque contém os atributos mais frequentes

dos seus membros. No entanto, ainda é uma abstração idealizada a partir dos indivíduos da categoria.

Esses autores criticaram o modelo de Roch, porque consideram que ele não explica as maneiras pelas quais os conceitos complexos (por exemplo, o conceito "peixe de aquário") são aparentemente compostos de conceitos simples. Em sua teoria, eles sustentam que a categorização necessita de um conceito nuclear, na linha do modelo clássico de conceitos e na qual esse aspecto central deve ser combinado com um processo de identificação. Osherson e Smith defendem que uma teoria híbrida é importante porque "a habilidade para construir pensamentos e conceitos complexos a partir de algum armazenamento básico de conceitos parece se encontrar perto do coração da atividade mental humana". (1981, p. 55)

Johnson-Laird (1980) distingue entre procedimento de identificação de um conceito (chave) de fatos centrais de um conceito. Dessa forma, propôs que a análise dos conceitos somente baseada nos atributos não podia explicar aspectos significativos da categorização, sem considerar o conhecimento complexo que as pessoas possuem sobre conceitos. Eles afirmam que as pessoas podem aprender a identificar objetos através de testes perceptuais que não são necessariamente centrais para o conceito. Sendo assim, um conceito é equivalente a uma teoria sobre algum domínio e como a maioria das teorias, ela será um pouco difusa. Já Osherson e Smith (1981) sugerem que a teoria dos protótipos pode ser verdadeira quanto à identificação de procedimentos nos quais os objetos se ajustam a testes perceptuais em maior ou menor grau. Para esses autores, o conceito é tomado como uma representação organizada de conhecimentos gerais e de crenças sobre quaisquer objetos ou eventos denotados pelas palavras, sobre o que elas são ou fazem, sobre o que pode ser feito com elas, como elas são relacionadas, com o que elas se relacionam.

Eleanor Rosch (1999), em um artigo disponível no site da Universidade de Berkeley, no qual rediscute as noções de conceitos, articula grande parte dos autores aqui citados. A autora baseia-se em Fodor (1998) e discute o modelo de Teoria da Representação Mental (RTM), quando o autor afirma que não há cognição sem representação. Ela afirma que os conceitos possibilitariam o relacionamento da perspectiva do significado real e o mundo referencial, e funcionariam como símbolos para articulação sintática (operações formais) que possibilitam a interação. Dessa forma, Rosch trata os conceitos como unidades reguladoras. Nesse texto, ela discute se essa questão é filosófica ou cognitiva, retoma as discussões sobre a teoria clássica, cita Wittgenstein e curiosamente indica autores que, de certa forma, questionaram seu trabalho.

Pode-se afirmar, portanto, que a Teoria dos Protótipos e seus princípios possuem um amplo campo de aplicação, pois, não só abrange a área da semântica léxica, mas também é aplicável a grande parte dos fenômenos que impliquem uma categorização.

4 TEORIAS DOCUMENTÁRIAS

4.1 Terminologia

A Terminologia é um campo de estudos que toma o léxico como objeto. Compreende os termos técnico-científicos, representando o conjunto das unidades lexicais típicas de uma área científica, técnica ou tecnológica. Possui uma dimensão aplicada à produção de glossários e dicionários técnicos, dentre outros instrumentos de organização como o tesauro, foco de interesse deste trabalho.

Podemos tomar a Terminologia como uma área multidimensional, pois agrega tanto as questões relativas à natureza do léxico especializado quanto o seu papel nas comunicações profissionais; além de contemplar, consequentemente, as ações pragmáticas no manejo e controle dos termos técnicos, as chamadas *realizações*, vinculadas ao ideal de padronizar a comunicação profissional. Possui interfaces com a lexicologia, a lexicografia e a terminografia, que serão apresentadas mais a frente, bem como com a semântica, já discutida no capítulo anterior.

A produtividade linguística de que trata a Terminologia ocorre em universos das ciências, das técnicas e das distintas atividades de trabalho profissional. Terminologias como a da Matemática, Química, Linguística ou Jurídica são representativas de conhecimentos especializados. Dessa forma a fonte de dados são os empregos lexicais em publicações científicas, artigos, teses, dissertações, textos especializados em geral. Esse conjunto de termos também é chamado de *tecnoletos* ou língua de especialidade.

Esse tipo de léxico temático, como um componente linguístico, realiza suas funções essenciais: a de representação e a de transmissão do conhecimento especializado. Nesse sentido, a Terminologia cumpre, então, um papel auxiliar na solução das ambiguidades e nas questões de polissemia, comuns no léxico geral da língua, em busca de uma precisão conceitual. Esse expediente contribui para o intercâmbio comunicativo que se intensifica numa sociedade globalizada. Maria da Graça Krieger e Maria José Bocorny Finatto afirmam que “essa funcionalidade operada pelo léxico especializado na transmissão de conhecimentos, na transferência de aparatos tecnológico” (2004, p.16), situa a relevância da Terminologia como base da organização e divulgação da informação. As autoras ressaltam, ainda, que essa busca por estabelecer uma padronização terminológica nas linguagens técnicas é própria das

políticas linguísticas articuladas sobre a crença de que o uso recorrente de um termo garante a univocidade da comunicação especializada. Exemplo dessa política seria o do Comitê Técnico 37, *Terminologia: princípios e coordenação,* da Organização Internacional de Normalização, ISO[16]. Esse comprometimento com um conjunto de princípios organizadores contribui para a estabilização e controle de vocabulário.

O campo de estudos da Terminologia surgiu a partir da segunda metade do século XX como um ramo da Linguística Aplicada, principalmente na dimensão prática, pois um dos princípios ordenadores é a fraseologia[17] e a descrição terminológica que exigem o desenvolvimento de uma análise descritiva baseada em empregos linguísticos nos domínios científicos especializados. Em outras palavras (e no meu ponto de vista), os estudos linguísticos fundamentam, e até determinam, num *continuum*, grande parte do estabelecimento de um léxico especializado. Sendo assim, se a perspectiva linguística trata do conteúdo de uma palavra, contemplando a língua em sua amplitude de usos e possibilidades, a perspectiva terminológica dela extrai os conceitos, alinhando-o tanto como objeto teórico quanto descritivo e/ou aplicado, oferecendo dados sobre as estruturas produtivas das terminologias em um âmbito determinado. Krieger e Finatto sustentam que a Terminologia cada vez mais se inscreve nos estudos linguísticos, uma vez que se baseia em estudos sobre a constituição e comportamento dos termos, desde sua gênese ate a descrição de suas relações nas diversas áreas de conhecimento.

A Teoria Geral da Terminologia (TGT) teve como precursor Eugênio Wüster, que estabeleceu as bases da disciplina com uma feição autônoma e multidisciplinar, situada na convergência da linguística, da lógica, da ontologia, das ciências da informação e das diferentes áreas do conhecimento científico. O princípio fundamental da TGT assenta-se no componente conceitual: sua formação, suas características, suas relações dentro de sistemas conceituais; a descrição de conceitos, mediante definições, a atribuição de termos a conceitos ou de conceitos a termos, a essência dos termos e sua formação. Possui, portanto, um caráter metodológico, de natureza prescritiva e normalizadora em detrimento da apreensão quanto aos modos de funcionamento dos léxicos terminológicos. A TGT, também conhecida como

[16] As normas ISO (International Organization for Standardization) são técnicas internacionais estabelecidas por um organismo internacional de normalização para aplicação em âmbito mundial. Essas normas são desenvolvidas nos seus comitês técnicos (ISO/TC), que são organizados numa base temática com representantes dos seus membros. As representações são nacionais. A aprovação das normas ISO é feita mediante votação entre os seus membros. A participação brasileira nos trabalhos de normalização da ISO é efetuada através da ABNT. A adoção de uma norma ISO como Norma Brasileira recebe a designação NBR ISO.

[17] Parte da gramática que se dedica ao estudo da frase; compilação de construções de uma determinada língua.

Teoria Clássica, é reconhecida como um passo importante para o esclarecimento da essência das linguagens de especialidade.

Segundo Krieger (2000), a história da Terminologia está associada à consciência dos cientistas sobre o papel das nomenclaturas científicas na fixação e na circulação do saber científico e técnico, inerente ao próprio caráter dos termos técnico-científicos, considerados como ideais de expressão de monorreferencialidade, de monossemia e de exclusividade denominativa. A autora relata que os estudos fundadores da TGT tomavam os signos linguísticos de valor monossêmico, buscando caracterizá-los pela monorreferencialidade, a fim de veicular o significado específico de uma área, bem como estabelecer uma única referência com o universo exterior comunicativo. Nessa perspectiva, considera-se um termo como unívoco ou monovalente aquele que, em um contexto discursivo determinado (situação comunicativa), especificamente numa frase, tenha um significado localizado, embora possa ser polissêmico. Essa distinção, monossemia e monovalência, ou univocidade em sentido estrito, permite limitar a exigência teórica da monossemia em terminologia como a uma única condição econômica: *que os termos sejam 'monovalentes', sem serem necessariamente 'monossêmicos'* (WÜSTER apud KRIEGER, 2000, p. 213). Na concepção wüsteriana, a terminologia expressa conceitos e não significados, questão que será retomada à frente. Essa perspectiva relaciona-se ao posicionamento de que a parte significante do termo é designação de um conceito. Tem-se, então, a denominada relação biunívoca entre conceito e termo, postulando-se que, de acordo com a teoria clássica, o conceito precede a denominação, além de reforçar que a onomasiologia caracteriza o processo designativo das ciências e das técnicas. O direcionamento do conceito à denominação, tanto em relação à produtividade lexical quanto na metodologia proposta para as atividades de organização terminográfica dos léxicos especializados tem sido apresentado como o divisor entre as atividades de terminologia e as de lexicografia, a ser tratada no tópico seguinte.

Krieger postula que a TGT não ampliou seu universo explicativo, porque deixou de considerar os aspectos comunicativos e pragmáticos inerentes ao léxico das linguagens especializadas, limitando seu objeto de análise às unidades unívocas normalizadoras próprias dos âmbitos científicos-técnicos. Dessa forma, reduziu a atividade terminológica à recolha de conceitos e termos para a normalização dos termos, fixando noções e denominações normatizadas. Consequentemente, os objetivos foram limitados para assegurar a univocidade da comunicação profissional, fundamentalmente no plano internacional. Conclui-se, portanto, que se sobrepuseram, nesse momento, as concepções normativas que auxiliaram a expandir e a consolidar a organização internacional da terminologia, sob a feição de uma linguagem

controlada, como os regramentos da ISO, determinados pelo seu Comitê Técnico de Terminologia, conforme já citado. Esse reducionismo diante do funcionamento da linguagem tornou-se um dos focos principais das críticas à TGT. De modo geral, atribui-se à TGT o baixo prestígio alcançado pelo campo de estudos dos léxicos especializados na pesquisa linguística.

Continuando numa linha cronológica, a autora revela que, nos anos 80, Alain Rey propõe a compreensão da produção terminológica à luz de um ponto de vista descritivo, demonstrando que "tratar de terminologia técnico-científica é tratar de uma questão da linguagem e não de um construto ideal e homogêneo a serviço de uma comunicação restrita ao âmbito de especialistas e isento de polissemia e de ambiguidades conceituais". (2000, p. 214). Mais recentemente, nos anos 90, Krieger ressalta que se fortaleceram os estudos fundamentados na complexidade que envolve o funcionamento das terminologias, tanto como qualquer outra unidade da língua natural, dando ênfase aos aspectos linguísticos e pragmáticos dos termos e promovendo interfaces com questões sociolinguísticas, filosóficas e tradutológicas bem com a demanda das tecnologias digitais que movem a Terminologia em direção aos estudos descritivos.

Nessa linha, destacam-se as contribuições de Maria Teresa Cabré (1999) que propôs uma nova teoria estruturada nos aspectos comunicativos das linguagens especializadas em detrimento dos propósitos normalizadores: a Teoria Comunicativa da Terminologia (TCT) a qual contempla a dimensão da variação linguística, além de estipular que, tanto o conhecimento especializado, quanto os textos especializados, como as unidades terminológicas podem ocorrer em diferentes níveis de especialização e serem descritas em diferentes níveis de representação. A TCT considera os termos como unidades linguísticas, enfatizando a função da língua como instrumento de comunicação; assim recupera a noção de termo como unidade de forma e conteúdo que, em determinadas condições discursivas, adquirem valor especializado. Cabré entende que as unidades terminológicas fazem parte da linguagem natural; para tanto não considera os termos como entidades autônomas que formam um léxico especializado. Nessa dimensão, os termos são unidades conceituais com capacidade de referência, desde que analisados em seu contexto de uso em situações determinadas. Os termos para a TCT representam e transferem o conhecimento especializado, em graus e modos distintos, como em situações diversas. Pode-se afirmar que a TCT atribui dimensão textual e discursiva à terminologia.

Além da TCT proposta por Cabré, as novas vertentes da Terminologia são socialmente orientadas como a Socioterminologia, de François Gaudin, citado por Krieger (2000), que

enfatiza a inoperância do ideal normalizador para a produção terminográfica e alerta para a necessidade de se efetivar um diálogo interdisciplinar entre as áreas de conhecimento constitutivas da Terminologia. Para Gaudin, as relações entre signo e realidade são estabelecidas nas situações sociais onde acontecem; por isso propõe a substituição do conceito de domínio pelo de *episteme* (interpretável), considerando-o como mais adequado para descrever a realidade do trabalho científico e das práticas linguísticas que o sustentam.

A Socioterminologia enfatiza a importância das relações semânticas para um melhor entendimento da estrutura interna do léxico, substituindo as definições formais pelas descrições mais produtivas do significado das palavras, uma vez que leva em consideração a identificação e a categorização das variantes linguísticas dos termos em diferentes tipos de situação de uso da língua. Para que o especialista em terminologia desenvolva seu trabalho de pesquisa, é necessário atentar para os critérios básicos de variação terminológica no meio social, bem como critérios etnográficos, porque sabemos que as comunicações entre os membros de uma dada comunidade podem gerar termos diferentes para um mesmo conceito ou mais de um conceito para o mesmo termo. Portanto, a perspectiva descritiva prepondera sobre uma perspectiva supostamente prescritiva, incluindo, as diversas camadas de termos que se articulam a um termo padrão, contextualizando-o, conforme as circunstâncias de comunicação e os grupos de usuários. Avalia-se, dessa forma, o grau de estabilidade ou mutabilidade da variação, avaliando se as variáveis têm efeito positivo ou negativo sobre a emergência de usos linguísticos alternativos, tentando prever seu comportamento regular e sistemático. Essa perspectiva linguístico-social fundamenta-se na análise das condições sociais e linguísticas de circulação dos termos, reforçando a aceitação da variação lexical e a negação da univocidade terminológica.

Gaudin relaciona as metáforas e os contextos históricos de uso das palavras, em contraposição à monossemia wüsteriana. Demonstra, ainda, que a relação entre o signo e a realidade é construída no âmbito das interações verbais. A referência denota, portanto, um caráter social e, por essa razão, não é possível propor a biunivocidade entre conceito e termo. O autor recupera os conceitos de dialogismo e polifonia, de Bakhtin, propondo observar a polifonia enunciativa (as múltiplas vozes num texto) e reavaliar as tipologias que separam os textos em categorias segundo o grau de sua cientificidade.

Já a Teoria Sociocognitiva da Terminologia ou Socioontologia (TST), proposta por Rita Temmerman (2000) destaca o papel dos modelos cognitivos, procurando mostrar as relações entre os processos de categorização e a linguagem. Essa corrente mais atual baseia-se na semântica cognitiva e questiona a centralidade da padronização em detrimento de uma

descrição realista dos significados dos termos tal como aparecem nos textos. Para Temmerman, a definição de conceito da terminologia tradicional, como unidade de pensamento constituída por abstração de propriedades, é muito restritiva, já que poucos conceitos existem de forma objetiva. O mundo é melhor compreendido no marco dos modelos cognitivos nos quais se relacionam diferentes unidades de compreensão ou entendimento estruturadas de forma prototípica. Essa é a razão pela qual propõe substituir a noção de conceito por *unidade de compreensão* (entendimento) observando que grande parte das unidades ditas conceituais tem uma estrutura prototípica e que podem, dessa forma, denominar categorias.

A TST vê um termo como um construto que se instaura no contexto da comunicação especializada; assim um termo não existe a priori, denominando um conceito pré-existente no mundo real, mas se constitui em processos de conceitualização e categorização sociocultural. Esse processo inerente à língua, consequentemente, é mediado por modelos cognitivos muitas vezes metafóricos que facilitam a compreensão da realidade; portanto propõe a unidade termo/conceito como uma estrutura prototípica observável no contexto da comunicação da ciência e da técnica. O conhecimento científico corresponde, então, a um padrão sociocognitivamente modelado. Novamente, a base para a delimitação do conteúdo é o texto no qual está inserido o termo, portanto o conceito não é universal, nem imutável; expressa um conjunto de elementos de natureza linguística que se consubstanciam num texto que possui não apenas uma dimensão linguística, mas também pragmática, discursiva e comunicativa. Tem-se, então, que o ponto de partida da descrição terminológica é o termo, e não o conceito. Cada termo é classificado como entidade, atividade ou como pertencente a uma categoria coletiva, considerando-se sua intenção e sua extensão. A TST põe em relevo as relações entre os estudos da terminologia e da linguística textual, reforçando a ideia de que os termos não podem ser compreendidos fora de seu ambiente natural - os textos -, verdadeiros conjuntos dinâmicos de elementos linguísticos, pragmáticos, discursivos e comunicativos.

Temmerman dá ênfase à categoria, pois a considera adequada para descrever a unidade de compreensão. A categoria tem um núcleo e uma estrutura delimitada, mas está em processo de contínua reformulação. As categorias têm estruturas prototípicas que recolhem informação intercategorial e intracategorial. Percebe-se sua existência no discurso, o que permite identificar mais módulos de informação do que via características diferenciadoras ou posição numa classificação lógica ou ontológica. A dimensão intracategorial induz a uma definição nuclear (por exemplo, parte de), informação histórica, procedimental etc.; na intercategorial podem ser estudados o domínio, a perspectiva e a intenção do modelo

cognitivo usado. Sendo assim, dependendo do nível e tipo de especialização do emissor e do receptor, a informação considerada relevante para uma definição varia, o que torna mais flexível a estrutura conceitual, já que as descrições do significado dependem tanto do tipo de unidade de compreensão, como dos participantes da comunicação. Consequentemente, a teoria admite a sinonímia e a polissemia no processo de compreensão e de comunicação justificadas pela funcionalidade, pela flexibilibilidade (os modelos cognitivos se alteram constantemente) e pela diversidade dos processos de categorização.

Pode-se afirmar que as correntes terminológicas apresentadas têm procurado refletir sobre as necessidades decorrentes das alterações no pensamento da ciência em geral, procurando meios para solucionar as restrições relativas aos seus fundamentos originais. As proposições teórico-metodológicas da TGT que forneceram as primeiras coordenadas do saber e das práticas em Terminologia estão, atualmente, em busca de uma adequação do ponto de vista linguístico e comunicacional dos termos técnico-científicos. Diante disso, o léxico especializado passa a ser compreendido como componente das línguas naturais, não imunes, portanto, aos efeitos decorrentes do funcionamento da língua.

4.2 Lexicologia

A Terminologia é um dos ramos da Lexicologia e ambas apresentam um aspecto prático em oposição ao teórico: a terminografia e lexicografia respectivamente. A título de delimitação preliminar de campos, segue-se uma breve distinção entre as áreas.

Lexicologia, do grego *lexis*, palavra, é o estudo sistemático do *léxico*, de ou para as palavras. Tendo, portanto, como objeto a palavra, a categorização lexical e a estruturação do léxico estudam as relações deste com os outros sistemas da língua, mas sobretudo as relações internas do próprio léxico. A Lexicologia abrange domínios como a formação de palavras, a etimologia, a criação e importação de palavras, a estatística lexical, e se relaciona necessariamente com a fonologia, a morfologia, a sintaxe e em particular com a semântica. Nesse âmbito, as relações semânticas de sinonímia, antonímia, hiponímia, hiperonímia interessam mais especificamente à lexicologia.

Trata-se de uma disciplina linguística que se ocupa do estudo do léxico, ou seja, das palavras de uma língua de um modo geral. Estuda todos os aspectos relacionados com as unidades de primeira articulação, ou seja, na perspectiva saussureana, as unidades dotadas de

duas faces, significante e significado. O léxico é o nível linguístico que mais facilmente emerge na consciência dos locutores, pois está diretamente relacionado com a significação no mundo em que vivemos. Ocupa-se dos aspectos formais e semânticos das unidades lexicais de uma língua, uma vez que está intimamente relacionada aos aspectos morfológicos, no tocante à problemática da derivação das palavras, da categorização léxico-gramatical, bem como à estruturação dos sintagmas, além das relações de significado.

Segundo Andrade (2001), a Lexicologia propõe-se a pesquisar o universo de todas as palavras de uma língua, vistas em sua estruturação, funcionamento e mudança, com o objetivo de conceituar e delimitar a unidade lexical de base – a lexia -, bem como elaborar os modelos teóricos subjacentes às suas diferentes denominações. Para tanto toma a palavra como um instrumento de construção e detecção de um sistema de valores como reflexos de uma cultura, além de analisar e descrever as relações entre a expressão e conteúdo das palavras e os fenômenos dela decorrentes.

Tradicionalmente, os lexicólogos baseiam-se, em suas análises, numa perspectiva morfológica, conforme já mencionado, considerando, portanto, o estudo dos neologismos. Apesar de se tratar de uma disciplina com limites imprecisos e indefinidos, dado o universo ilimitado dos repertórios lexicais, há que se ressaltar que os campos lexicais têm evidenciado traços de sistematicidade ao componente lexical.

4.3 Lexicografia

A Lexicografia se ocupa da feitura de dicionários, ocupa-se da análise da significação das palavras. Portanto as contribuições da Lexicologia são fundamentais para a Lexicografia, que também pode ser entendida como um ramo da Lexicologia aplicada. Pode-se afirmar que é a ciência dos dicionários, uma atividade antiga e tradicional. As primeiras obras eram meras listas de palavras para auxiliar o leitor de textos da antiguidade. Os séculos XVI e XVII marcam a Lexicografia, com a elaboração dos primeiros dicionários monolíngues e bilíngues.

Ao longo dos anos, a descrição do léxico foi efetivamente realizada pela Lexicografia, porém com uma práxis pouco científica. Mais recentemente, o fazer lexicográfico tomou uma feição científica, fundamentada numa teoria lexical. Até o século XIV, por exemplo, as palavras eram organizadas não em ordem alfabética, mas em grupos de significados parecidos. Esse tipo de dicionário sobreviveu aos nossos dias: são os tesauros – objeto deste

trabalho-, dicionários de sinônimos ou de ideias correlatas. Foi Noah Webster (1758-1843), lexicógrafo americano, que estabeleceu critérios fundamentais para separar as palavras dicionarizáveis, analisar sua freqüência no interior dos discursos individuais e coletivos, do presente e do passado, para, em seguida, apreender os valores significativos nucleares e virtuais, explicitando definições e construindo uma rede de acepções, configurados em um verbete. Pode-se dizer que os autores dos dicionários são tributários das coerções do uso vivo das palavras em meio às comunidades linguísticas, porque registram a norma lexical corrente na sociedade, porém entende-se aqui que os dicionários são materiais de consulta e funcionam como parâmetros de possíveis usos das palavras de uma língua. É certo que os falantes de uma língua não necessariamente consultam um dicionário para interagir em sociedade, porém se o fizerem poderão ampliar sua capacidade enunciativa.

Um dicionário é constituído de entradas lexicais, ou lemas que identificam um termo da língua ou um referente do universo extralinguístico. Essa lista total de lemas constitui a sua macroestrutura e o verbete, sua microestrutura. Os verbetes têm como eixo a definição da palavra em contextos de uso orais ou escritos, e/ou demonstrados com exemplos em frases. Os significados e os usos referidos são aqueles registrados e documentados em contextos realizados, frequentemente acompanhados de informações sociolinguísticas bem como remissões a outras unidades lexicais associadas.

Há diversos estudos no sentido de acompanhar não só as formas que vão sendo consagradas pelo uso, mas também os avanços tecnológicos do século XXI; a versão digital dos tesauros é um deles. Atualmente, as bases de dados lexicais são, em geral, utilizadas como repositório central de informações lexicais de uma determinada língua. As informações armazenadas podem ser de natureza sintática, semântica, pragmático-discursiva, fonético-fonológica ou morfológica, ou ainda podem expressar relações entre itens lexicais de uma mesma língua (como em um tesauro) ou entre itens lexicais de línguas distintas (como em um dicionário bilíngüe). Além de facilitar a manipulação e manutenção das informações, essas bases são uma forma estruturada de armazenar os dados. Inicialmente, os computadores eram utilizados apenas para acessar exemplos de usos das palavras. A partir da década de 80, esse cenário foi sendo alterado e, aos poucos, os computadores passaram a ter maior participação no processamento de palavras e no processo de criação de dicionários, principalmente com o desenvolvimento de bases de dados de informações lexicais.

4.4 Terminografia

Também conhecida como Lexicografia Especializada, a Terminografia é a disciplina Linguística intimamente ligada à Terminologia, que se ocupa da descrição das propriedades linguísticas, conceituais e pragmáticas das unidades terminológicas de uma ou mais línguas, a fim de produzir obras de referência, tais como dicionários, glossários, vocabulários em formato papel ou eletrônico, bases de dados terminológicas e bases de conhecimento especializado. Segundo Krieger (2004), a Terminografia não se confunde com a Lexicografia, uma vez que toma o termo como objeto de descrição e aplicação e não a palavra, como faz a Lexicografia; procura, ainda, definir seu conteúdo, considerando seu uso profissional. Por sua vez, o objeto aplicado exige o dimensionamento dos fundamentos teóricos necessários à identificação da terminologia, o reconhecimento da variedade de suas formas, bem como os princípios de análise do funcionamento dos termos, a fim de proceder ao registro como referência especializada.

Os estudos terminográficos oferecem subsídios para o estabelecimento de princípios metodológicos, orientando o tratamento a ser dado aos elementos constituintes do universo de informações a ser trabalhado. Na Terminografia, por exemplo, os termos podem ser registrados no feminino ou plural, não sendo reduzidos à forma canônica como na Lexicografia (masculino, singular). Outra característica é a especificação de uma área de conhecimento cujo léxico é repertoriado, permitindo a correlação de termos e conceitos de uma determinada área.

A função normalizadora da Terminografia consiste numa medida de controle dos usos de um determinado vocábulo. No entanto, as políticas linguísticas normalizadoras apontam, ainda, para as teorias clássicas da Terminologia. Mas, já há instrumentos que registram os sinônimos e as variações, ou seja, formas concorrentes que o discurso especializado emprega.

4.5 Linguística Documentária

A Linguagem Documentária é um instrumento por meio do qual se realiza a mediação entre sistemas ou conjuntos informacionais e os usuários, exercendo a função de ponte entre eles. A organização da informação via linguagem documentária é essencial, quando se

necessita de mecanismos que desempenhem o papel de filtros para a recuperação, dado o grande volume de informações recuperadas na internet. Nessa perspectiva, a Linguística Documentária, subdomínio da Ciência da Informação (CI), propõe códigos para o tratamento e a recuperação da informação, e recorre à Terminologia, visando a garantir bases referenciais e metodológicas para a organização de campos temático-funcionais, bem como para a estruturação dos campos lógico-semânticos das linguagens documentárias, desempenhando, assim, um papel complementar à norma documentária de elaboração de tesauros (ISO 2788-1986)[18].

As interfaces entre Terminologia e Linguística Documentária, no Brasil, tiveram início a partir da proposta clássica, através das normas terminológicas internacionais (ISO 704-2000 e ISO 1087/1-2000, dentre outras)[19] e dos textos teóricos fundamentais. Essas contribuições teóricas e metodológicas resultaram no aperfeiçoamento das metodologias de construção da linguagem documentária, uma vez que as normas terminológicas são, comparativamente à norma de elaboração de tesauro (ISO 2788), mais completas, permitindo uma melhor compreensão dos conceitos, das relações entre os conceitos e da modelagem dos sistemas conceituais. Já o uso da terminologia concreta - dicionários técnicos especializados, glossários - significou a possibilidade de conferir referência concreta à interpretação dos descritores dos tesauros. Pode-se afirmar que a Linguística Documentária é uma área de estudos interdisciplinar que congrega conceitos da Linguística Estrutural, porque considera alguns elementos, dentre eles, os dados de imprenta[20], como dados formais de identificação, bem como na identificação dos conteúdos e significados, referentes a uma dada temática com o objetivo de classificá-los por matérias ou por disciplinas; tal procedimento certamente alia as teorias do discurso, principalmente na questão da determinação dos gêneros textuais, que também é um princípio organizador; da perspectiva da Linguística Gerativa amplia o campo das possibilidades de combinações, na tentativa de estabelecer parâmetros de indexação de materiais. Já da semântica, a linguagem documentária busca um princípio orientador no tocante à delimitação de áreas e subáreas, com o objetivo de organizar, sistematizar e disseminar a informação, apoiando-se nas ciências da comunicação e das tecnologias inteligentes de modo a promover a divulgação e o acesso aos documentos, como o

[18] O Brasil não dispõe de norma brasileira própria. No âmbito internacional existe a "Norma Internacional de Elaboração de Tesauros Monolíngues (ISO 2788)", que, apesar dos esforços empreendidos no sentido de apresentar princípios para uma normalização terminológica consistente, verifica-se, ainda, muitos problemas relativos à unidade de representação, ou seja, ao termo de indexação.

[19] Os termos são considerados pelas ISOs 704 (2000) e 1087-1 (2000) como sendo a unidade mínima da Terminologia.

[20] Conjunto de informações, normalmente impresso na parte inferior da página de rosto de um livro, contendo nome da editora, cidade e ano de publicação.

PERGAMUM (sistema informatizado) utilizado em Bibliotecas, para o gerenciamento de dados, que permite a busca de informações sobre diversos tipos de materiais disponíveis, bem como a localização destes para consultas posteriores.

No entanto, o trabalho documentário de organização da informação para a recuperação não obterá sucesso se restringir-se ao uso das referências da terminologia clássica, em universos que não constituem formalmente um domínio claramente delimitado ou que se encontram em transformação, caracterizando-se por inúmeras relações interdisciplinares, conforme discutido em seções anteriores, bem como quando precisamos solucionar problemas relativos a conjuntos de informações que têm graus menores de ordenação ou não são analisáveis a partir de parâmetros consensuais de organização. Podemos concluir sobre a importância de se estabelecerem estratégias para tratar a heterogeneidade, a diversidade e a multiplicidade dos pontos de vista da produção, do uso e do consumo informacional.

Na atividade documentária, é comum encontrarmos domínios temáticos para os quais não existem vocabulários formalizados, situações em que se deve recorrer ao uso de procedimentos terminológicos e terminográficos contemporâneos para a exploração de *corpus* particular e para a identificação das redes relacionais entre os termos. Esse procedimento tem se mostrado fundamental para o melhor conhecimento dos universos a serem trabalhados documentariamente, visando tanto à construção de instrumentos de intermediação como os tesauros, quanto à organização direta de conjuntos informacionais, como é o caso dos conteúdos de sites. Confirma-se, portanto, a aproximação de uma terminologia mais associada à Linguística Textual, considerando a variação denominativa e conceitual. A identificação da estrutura e tipologia textual, baseada na *microestrutura*, que corresponde à análise sobre a realidade física de um texto e seus símbolos de significação, as palavras que o compõem; na *macroestrutura*, que se configura como o tópico representativo hierárquico e coerente da unidade textual, configurado pelas estruturas sintático-semânticas de um texto; e na *superestrutura*, que seria um modelo convencionado para a produção do texto (Van Dijk), encontra o respaldo necessário, inclusive, ao desenho das condições de produção de um texto, que devem ser consideradas no planejamento e inserção de conteúdos nas ferramentas de gestão de informação. A observação do fenômeno terminológico no interior da linguagem permite recuperar a função básica da língua, a comunicação, deslocando a ênfase do caráter representacional da língua para o comunicativo. Assim as formas apriorísticas de categorização são substituídas por categorias funcionais, deslocando-se a ênfase da noção de documento como suporte do conhecimento, para a de sua função.

Nessa linha, a informação documentária se define na relação produção/recepção, considerando o pólo da produção (ou da enunciação/autor), sem centrar-se no pólo da recepção (ou da enunciação/leitor). Além disso, não se pode desconsiderar o usuário como sujeito da interpretação (o enunciatário), nem as marcas intencionais da enunciação. O problema se desloca, então, da enunciação e da recepção para a *mensagem*, enfatizando a característica mediadora da atividade documentária: modo de veiculação do bem simbólico que é a informação. A ênfase é dada ao conteúdo, embora não sejam totalmente desconsiderados os papéis do leitor e do autor nesse processo. Há que se admitir que os sujeitos interlocutores ficam mais opacos, dado o caráter generalista e globalizante de uma tecnologia de informação. Certamente está prevista uma "cena enunciativa", pois, conforme já dito, o objetivo é veicular informação, na medida do possível bem ajustada ao usuário, para atender às suas demandas. Para tanto são consideradas as variantes como os fatores de contextualização de uma informação, ampliando suas possibilidades de utilização.

A linguagem documentária tem os "descritores" como denominações que funcionam com chaves autorizadas de acesso à pesquisa dos termos que, por sua vez, viabilizam a comunicação especializada. O termo é uma unidade léxica que assume um valor semântico próprio de uma determinada área de conhecimento; já o descritor seria uma unidade de indexação própria da linguagem documental, e nessa medida, artificial, pois não pode apresentar ambiguidade e deve ser normalizada, cumprindo uma função de metarrepresentação e de metatransferência do conhecimento. (KRIEGER; FINATTO, 2004). Há que se ressaltar, ainda, que o vocabulário documental não equivale à terminologia, mas constitui-se um referencial de busca da informação.

A demonstração dessas teorias que fundamentam a documentação pode comprovar como se aproximam a atividade linguística e a atividade documentária. Em que pese o conjunto de convenções normalizadoras que visam ao controle do vocabulário, de modo a facilitar o acesso aos acervos, é na linguagem que se sustentam os sistemas de informação documental, é a perspectiva linguística que norteia a seleção prévia das informações a serem organizadas.

Dessa forma, na seção seguinte, será apresentada uma tentativa de análise de base semântica de cadeias e renques de tesauros, a fim de levantar a discussão sobre critérios mais eficazes para a organização da informação de modo a atender melhor às demandas dos usuários.

4.6 Visão de conceito na perspectiva da CI e princípios organizadores dos tesauros

O tesauro, conforme já apresentado, é uma ferramenta de organização de informação em linguagem documentária, de modo controlado e dinâmico, ou seja, sistêmico, que contém termos relacionados genérica e semanticamente, cobrindo um domínio específico do conhecimento. É, por definição, especializado, pois se ocupa de um assunto bem delimitado, haja vista que um número muito grande de conceitos seria de difícil sistematização bem como para a busca de informações. Questões relativas à homonímia, sinonímia, hiperonímia/hiponímia, metonímia, entre outras ocorrências linguísticas, são um problema recorrente na construção de um tesauro.

Para minimizar alguns desses problemas, os profissionais de Ciência da Informação utilizam o método de *"Categorização",* que busca pensar o domínio do tesauro de forma dedutiva, ou seja, determinar as classes de maior abrangência (facetas) dentro da temática escolhida. Ao analisar o domínio segundo a categorização, eles acreditam que fica mais fácil e seguro determinar os termos que devem ou não integrar o tesauro. O sistema de conceitos é previamente planejado em gráficos feitos e refeitos, até que se chegue à solução relativamente definitiva. Portanto, segundo esse método, os conceitos devem ser reunidos em classes gerais ou facetas.

O método de Categorização ou Facetação foi desenvolvido por Ranganathan na década de 1920 para a construção da Tabela de Classificação Bibliográfica. Na elaboração de tesauros, essa metodologia fornece os princípios para agrupar conceitos de mesma natureza em Classes Gerais ou Facetas para construir cadeias e renques. A relação hierárquica é uma relação lógica. Nela, o conceito superordenado é conhecido como *conceito genérico (no tesaurus, TG-termo genérico),* quando inserido no sistema de conceitos; enquanto o conceito subordinado é conhecido como *conceito específico (TE-termo específico)* e reúne conceitos que têm características comuns entre si. São relações de dois tipos: subordinação, formando *cadeia*, uma série vertical de conceitos; coordenação, formando *renque*. Se os conceitos que estão sendo estruturados não forem capazes de representar a cadeia completa, então devemos introduzir os conceitos intermediários necessários. O renque é constituído de conceitos subordinados a um mesmo conceito, ou seja, conceitos coordenados entre si; podemos dizer que são conceitos "irmãos". A organização dos conceitos é conduzida pelo processo de indução/dedução, de acordo com o contexto ao qual se refere o tesauro. Por exemplo, num tesauro sobre plantas, macieira e limoeiro são elementos da classe Árvore Frutífera, formada

pela característica de divisão: tipos de árvores frutíferas. Os renques configuram, portanto, classes formadas a partir de uma única característica: serem frutíferas. Já a cadeia é uma sequencia de conceito que têm uma característica a mais ou a menos, conforme a cadeia seja ascendente ou descendente. No nosso exemplo, ficaria assim:

* Plantas (TG)
  **Árvores frutíferas (TE)
    ***macieira
    ***limoeiro

Nele, podemos visualizar que a relação entre planta (TG) e Árvores frutíferas (TE) configura uma cadeia (formato edentado); já macieira e limoeiro exemplificam um renque.

Na verdade, aplicar a categorização é analisar o domínio a partir de recortes conceituais que permitem determinar a identidade dos conceitos (categorias) que fazem parte daquele domínio. Mas há três princípios básicos norteadores para sua construção, que envolvem garantias:

a) literária: os termos devem pertencer à bibliografia de sua área específica, assim como dela devem ser retirados;

b) do usuário: os termos, além de pertencerem à bibliografia, devem ser empregados pelos usuários da área;

c) estrutural: a própria estruturação do tesauro sempre relacionada à tipificação ou subdivisão em partes.

Após a escolha dos termos, os construtores do tesauro escolhem suas facetas. Entenda-se por faceta a visão desejada para o instrumento. Por exemplo, se fôssemos construir um tesauro para o curso de Letras, poderíamos optar pelas facetas: professores, disciplinas, áreas do conhecimento etc. Outros poderiam preferir: disciplinas de cada semestre, disciplinas optativas etc. Escolhidos os termos, são criados e/ou recuperados conceitos equivalentes a cada um dele.

A noção de conceito no universo da CI é a de uma unidade de conhecimento, porque os textos aos quais um tesauro se destina são registros do conhecimento e tal vocabulário deverá representá-los. Assim, o tesauro se constituirá como um mapeamento sobre determinada área do conhecimento, no qual os conceitos estão sistematizados. Deduz-se, pois, que mesmo isolado, o termo carrega consigo o significado. A decisão na formulação de conceitos é um passo fundamental para estruturar os relacionamentos dos termos no tesauro.

No universo da CI, utiliza-se, como princípio na identificação de um conceito, o modelo de I. Dalhberg, que é uma releitura do triângulo semiótico, já demonstrado neste trabalho (subseção 3.3), no qual são estabelecidas três condições necessárias expressas no triângulo a seguir:

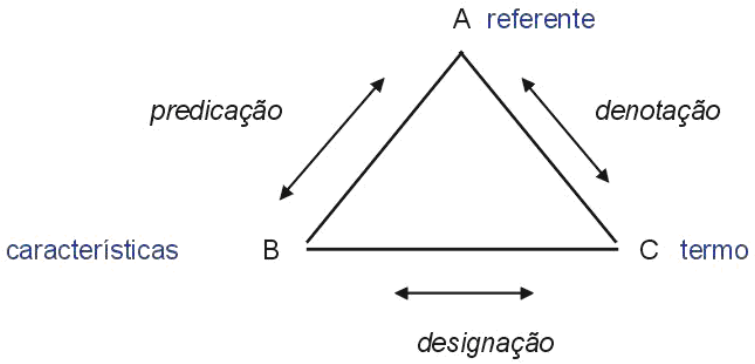

Figura 04 – Triângulo Conceitual de Dahlberg
Fonte: DAHLBERG, 1978, p. 102.

O triângulo acima pode ser entendido como o conjunto das propriedades significativas atribuídas a um referente (característica), pelos membros de uma área de assunto, e sintetizado num signo linguístico que é o termo. Tem-se, portanto, que entre o signo linguístico e o referente não há relação direta e que esta se estabelece somente através do conceito. Conforme propõe Dalberg,

> Se o conhecimento pode ser considerado a totalidade de proposições verdadeiras sobre o mundo, existindo - em geral - nos documentos ou nas cabeças das pessoas, então o conhecimento pode parecer existir também em todas as afirmações verdadeiras (em todos os julgamentos) e em todas as proposições científicas que obedecem a um postulado verdadeiro. (DALBERG, 1978, p. 103).

Já a característica seria o elemento constitutivo, formador do conceito. A análise de um objeto, um referente, implica a identificação de suas propriedades. Pode-se dizer, ainda, que as propriedades dos referentes/objetos podem corresponder às características do conceito. O termo, como o terceiro elemento do triângulo, é a designação do conceito e pode ser constituído por uma palavra ou por um grupo de palavras; seria a menor unidade de representação do conceito, indivisível na indexação e nos tesauros. Diferentemente da palavra, o termo tem seu significado assegurado, mesmo fora de contexto, ou seja, isoladamente, haja vista o efeito padronizador de um tesauro.

Na ausência de uma das condições do triângulo de Dahlberg, não se terá o conceito. No âmbito de uma área de assunto, em um tesauro, não poderia haver dúvida quanto ao

conceito que o termo comunica, ou seja, é necessário existir uma monorreferencialidade; em outras palavras, o termo deve evocar o que tem de ser evocado.

Podemos, então, indicar aqui uma norma ISO 704, norma de certificação que deve ser considerada na estruturação dos tesauros. Nela, o conceito é tomado como uma unidade abstrata criada a partir de uma combinação única de características. Assim, os conceitos são representados pelos termos, que são designações verbais, pois, na dimensão da organização dos tesauros, não são associadas imagens[21]. O termo é, na norma, considerado como a unidade mínima da terminologia (ISO 704; ISO 1087-1). Essa norma ISO 704 caracteriza uma definição como uma unidade com intensão e extensão únicas, ou seja, uma definição é um enunciado que descreve um conceito permitindo diferenciá-lo de outros conceitos associados, podendo ser formulada de duas maneiras básicas: *definição por compreensão (*ou *por intensão),* ou ainda, *definição intensional*, que compreende a menção ao conceito genérico mais próximo (o conceito superordenado) –já definido ou supostamente conhecido–, e às características distintivas que delimitam o conceito a ser definido; e *definição por extensão* ou *extensional*, que descreve o conceito pela enumeração exaustiva dos conceitos aos quais se aplica (conceitos subordinados), que correspondem a um critério de divisão (ISO 1087-1).[22]

O conceito é, segundo Campos (Tutorial), uma unidade do conhecimento constituído por características que refletem as propriedades significativas relevantes atribuídas a um objeto ou a uma classe de objetos e expresso comumente por signos linguísticos. O conceito, portanto, é tomado como objeto formal, como construto mental.

Dessa forma, quando se comparam os conceitos, verifica-se que há características comuns entre eles (ISO 704), o que possibilita relações hierárquicas e/ou ontológicas. A relação hierárquica pode ser estabelecida entre conceitos ou classes, mediada pela lógica de semelhança, na qual há um conceito superordenado (TG) e um (ou mais) conceitos subordinados (TE), conforme já apresentado no início desta subseção. Já a relação ontológica se define entre conceitos que representam objetos relacionados por contigüidade no espaço ou no tempo, ou por conexão de causa-efeito (ISO 704). São ontológicas a relação partitiva e a associativa entre o objeto considerado no todo e suas partes e entre as partes propriamente

[21] As imagens que aparecem em tesauros são ilustrativas e não fazem parte do vocabulário controlado.

[22] Nessa linha, e no nosso ponto de vista, a norma ISO legisla no sentido de formatar o conceito, a fim de ajustá-lo ao instrumento organizador, normatizando-o, de modo a permitir uma recuperação adequada de informações. Neste momento, seria importante reafirmar que as proposições feitas nesta seção baseiam-se nas perspectivas da CI para a elaboração de tesauros; elas podem se tornar objetos de discussão, e/ou contrargumentação, neste trabalho, mais a frente.

ditas. E, por fim, há a relação ontológica de dependência entre dois conceitos concernentes a objetos com proximidade temporal ou espacial (ISO 704). Essas relações serão retomadas na seção seguinte mais detalhadamente.

O desafio da linguagem documentária é inscrever a unidade do conhecimento, o termo, num sistema de informação organizado, que apresenta temática específica. O tesauro é um tipo de vocabulário controlado e estruturado para o tratamento de informação, de determinado domínio científico. Percebemos que, quanto mais a temática do tesauro se aproxima do domínio do conhecimento, mais especificidade existe entre termo e descritor, mas, também, maiores tornam-se os problemas. Há casos de discrepâncias e equívocos que podem comprometer a recuperação da informação no domínio estudado, conforme será demonstrado nas análises a seguir.

A linguagem documentária, além de ser a base para a estruturação do conhecimento dentro das linguagens de especialidade, através da sistematização dos conceitos, e de servir de canal para a transferência do conhecimento, também constitui a base para a formulação de textos técnicos (redação técnica), para a tradução de textos de especialidade (tradução ou interpretação técnicas) e para a descrição, armazenamento e recuperação da informação especializada (documentação técnica).

Cabe esclarecer que os tesauros são instrumentos importantes para consulta de bibliotecários e demais gestores de informação no momento da indexação dos documentos. No entanto, para essa tarefa já havia uma Tabela de Classificação Decimal Universal (CDU) que ainda hoje é empregada como elemento parametrizador dessa classificação, criada antes da proposta teórica de Ranganathan, apresentada no início desta seção. Trata-se de um sistema de organização numérica, anterior à criação dos tesauros.

A primeira utilização de uma notação decimal como um código, de modo a expressar os conceitos de classificação documentária é atribuída ao físico André Marie Ampèr. Tal código foi amplamente divulgado pelo bibliotecário Melvil Dewey, com publicação oficial em 1876, conhecido como Classificação Decimal de Dewey (CDD), um sistema de classificação mais utilizado internacionalmente. Porém, várias evoluções e expansões foram feitas, especialmente por Paul Otlet e Henri La Fontaine que, em 1907, publicaram, em forma de catálogo sistemático, a Classificação Decimal Universal (CDU), com aproximadamente 33.000 subdivisões. Eles partiram do princípio de que quanto mais números decimais fossem utilizados, mais específica seria a organização da informação, alcançando, assim, um maior grau de detalhamento na organização dos assuntos. A CDU permite o uso de sínteses, ou seja,

a composição de números compostos para indicar assuntos inter-relacionados. Dessa forma o usuário pode buscar assunto de forma mais específica.

A tabela principal é identificada como notação primária e compõe-se de 10 classes:

0 Generalidades. Ciência e Conhecimento. Organização. Informação. Documentação. Biblioteconomia. Instituições. Publicações.

1 Filosofia. Psicologia.

2 Religião. Teologia.

3 Ciências Sociais. Estatística. Política. Economia. Comércio. Direito. Administração Pública. Forças Armadas. Assistência Social. Seguro. Educação. Folclore.

4 Vaga

5 Matemática e Ciências Naturais.

6 Ciências Aplicadas. Medicina. Tecnologia.

7Arte. Belas Artes. Recreação. Diversões. Esportes.

8 Linguagem. Linguística. Literatura.

9Geografia. Biografia. História.

A base da CDU é constituída por essas nove classes específicas e uma classe geral, seguindo o mesmo esquema da CDD, diferenciando-se na supressão de dois zeros à direita, portanto, apresenta-se somente com um algarismo arábico. Há que se observar que a classe 4, que seria da Linguística, foi incorporada na classe 8, de Literatura (em 1963), deixando-se, então, a classe 4 vaga para futuras expansões. Portanto, a classe 8, por exemplo, *Língua. Linguística. Literatura,* tem como subclasse a **80** *Questões gerais relativas à linguística e à literatura. Filologia.* Os números de classificação são frações decimais e seguem casas consecutivas, obedecendo a uma estrutura hierárquica; então as tabelas são divididas hierarquicamente, tanto na parte numérica como na parte conceitual, partindo do geral (da classe principal) para o específico (as subclasses).

Essa breve referência deve-se à constatação de que os princípios da CDU certamente interferem também no processo de organização de dados nos tesauros, haja vista que são instrumentos geralmente construídos sob a coordenação de profissionais da Ciência da Informação.

4.7 Princípios éticos e garantias no trabalho do indexador

Antes de partirmos para a análise linguística de tesauros, cabe aqui uma referência a questões de acesso, cultura e ética na obtenção, organização e representação do conhecimento em instrumentos de gestão de informação. Tratar de dimensões éticas implica o conhecimento dos atributos físicos dos documentos, sobre o conteúdo das obras, seus suportes, acesso às informações em níveis global e local, diversidade de línguas, indivíduos que pertencem às várias culturas, grupos étnicos e religiões. No atual contexto de potencial produção científica e notáveis opções tecnológicas, é necessário refletir sobre os princípios teóricos que apóiam um sistema, tendo em vista a globalização, já que o acesso pode ocorrer de qualquer parte do planeta, mesmo considerando-se que nem todos os países conhecem as possibilidades das tecnologias da informação.

Clare Beghtol (2002a, p. 509) relata que "as discussões sobre cultura se referem a valores compartilhados, história, língua, memória coletiva, atitudes sociais, preferências e práticas, entre outros". A autora reforça que os indivíduos podem ser multiculturais, multilíngües e participar de atividades multidisciplinares. Dessa forma defende que o fundamento da hospitalidade cultural é importante para promover um debate ético entre indivíduos e culturas, principalmente pelo perfil de preservação, mas também pela comunicação do legado de uma determinada comunidade ou nação.

Segundo a autora, na organização e representação de conhecimento, a *garantia cultural* atende ao pressuposto de que sistemas de diferentes culturas necessitam de diferentes tipos de informação. Em outras palavras, que podem ser úteis para os indivíduos em alguma cultura, essencialmente se forem baseados nas crenças e valores daquela cultura; por isso, podemos afirmar que se trata de artefatos culturais. Beghtol, diante dessa necessidade de encontrar soluções para as questões éticas no projeto, construção, manutenção e revisão de sistemas de organização e representação do conhecimento para acesso global à informação, propôs um modelo de análise que contempla três níveis: a) nível de descoberta, em se que procede a uma análise do sistema sob alguns valores éticos; b) nível teórico, momento no qual as principais teorias éticas são identificadas e refinadas; e c) nível de aplicação, em que as principais teorias são aplicadas aos resultados de pesquisa do nível de descoberta. Mas também reforça a necessidade de se estabelecerem políticas e procedimentos éticos, a fim de dar transparência ao processo; dessa forma as culturas profissional e pessoal do usuário bem

como do profissional da informação garantem o estabelecimento adequado de campos, termos, categorias e classes nesses sistemas.

Sabemos, no entanto, que qualquer garantia cultural a ser adotada pode ser apropriada para alguns grupos. Essas diferenças culturais merecem reflexões, mas essas questões são certamente minimizadas no corpus em discussão nesta pesquisa, os tesauros. Sua estrutura sistemática já restringe um campo semântico que limita a dimensão semântica da área contemplada por ele.

Dos estudos teóricos sobre organização da informação, seria importante também destacar Elaine Svenonius (2000). A pesquisadora discute sobre questões relacionadas aos fundamentos intelectuais da área de organização do conhecimento, enfocando os itens que são organizados, os princípios, os objetivos e a utilização de linguagens bibliográficas, bem como seus diversos componentes como vocabulário, semântica, sintaxe e pragmática. Discute sobre a natureza e os propósitos das linguagens de indexação que são empregadas para descrever o conteúdo de que trata o documento. Um ponto que nos interessa mais de perto é o paralelo que a autora faz entre a linguagem de indexação e a linguagem natural quanto a vocabulário, semântica e sintaxe. Para tanto ela discute sobre o cuidado necessário para transformar a linguagem natural em uma linguagem de indexação, no que tange aos problemas advindos da sintaxe de linguagens de indexação, como no caso dos termos sinônimos, ordem de citação, regularidade versus complexidade, automação, pré-coordenação versus pós-coordenação, e sintaxe natural versus sintaxe artificial.

É importante esclarecer que os termos que constituem o vocabulário de uma linguagem controlada (LC) podem ser extraídos dos próprios documentos, na própria linguagem do autor em linguagem natural (LN), que são os *termos derivados*. A LN pode ser entendida como sinônimo de discurso comum, pois linguagem usada habitualmente na fala e na escrita, porém na base de dados corresponde aos termos do título e resumo dos documentos. Há também os *termos assinalados* que são os termos utilizados para normalização bibliográfica, denominados “metadados” de organização da informação. Eles constituem os vocabulários controlados e estabelecem as relações hierárquicas entre os componentes da LC, entre eles as palavras-chave, os descritores, termos de indexação e os códigos de classificação.

Svenonius (2000) afirma que a LC é derivada da literatura que pretende descrever, portanto torna-se necessária a delimitação de suas fronteiras, tendo definido a LC como: “Uma linguagem normalizada que seleciona da linguagem natural um conjunto restrito de palavras e frases, tratando esse conjunto semanticamente para o estabelecimento das relações

com outros termos.”(p. 134). Dessa forma, a autora ressalta que o controle do vocabulário é um elemento essencial para a organização da informação, pois é uma linguagem artificial produzida com o objetivo sistematizar o vocabulário do usuário para um vocabulário padronizado. Diante disso, a autora acredita que a linguagem natural não deve ser utilizada para organizar a informação efetivamente, pois a sinonímia e homonímia podem causar dispersão e desordem. Propõe, a partir daí três passos para a criação de um vocabulário controlado: “o primeiro passo é a escolha da forma autorizada para o nome; o segundo é tirar a ambigüidade do nome, isto é, torná-lo único, caso já não o seja; e o terceiro é relacionar a forma autorizada do nome com as outras variações”. Como exemplo, Svenonius cita o caso dos nomes compostos, como João Antônio, que retornará ocorrências de diferentes autores que possuem partes desse mesmo nome. Por isso, reitera que “A eficácia de um sistema de acesso à informação é uma função direta da inteligência posta em organizá-lo”. (2000, p. ix). Svenonius alerta para a necessidade de padronização e reforça a função do tesauro como instrumento imprescindível na recuperação da informação.

Lopes (2002) desenvolveu uma revisão de literatura na área de organização de informação e propôs quadros comparativos para assinalar as vantagens e desvantagens de uso de LN e LC em bases de dados. Apesar de não haver uma comparação absoluta entre os dois quadro, optamos por incluí-los do mesmo modo como foi apresentado pelo autor, por entendermos que as informações neles contidas possibilitarão conhecermos as concepções de língua subjacentes ao processo de hierarquização e que podem, talvez, auxiliar-nos na compreensão referente às escolhas feitas pelos indexadores na estruturação dos conceitos.

O Quadro 03, por exemplo, apresenta limitações quando considera desvantajoso o uso de LN, sob o argumento de que exigiria um maior “esforço intelectual” na identificação dos termos. Se levarmos em consideração as questões relacionadas às garantias culturais, conforme assinala Beghtol, essa afirmação perde sua sustentação. Outra questão que não fica clara é a possível “perda de confiança” no uso de LN; tal argumento não justifica a negativa da revocação. A questão relacionada aos custos é inegável, assim como a qualidade de uma memória documentária que atende adequadamente aos usuários. Necessário torna-se, porém, estabelecer padrões mediadores (talvez, de perfil linguistico) entre a garantia literária e a garantia cultural, conforme se observa no Quadro 03.

VANTAGENS	DESVANTAGENS
1. Permite o imediato registro da informação em uma base de dados, sem necessidade de consulta a uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos, as grafias alternativas, os homônimos etc.
2. Processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle.	Haverá alta incidência de respostas negativas ou de relações incorretas entre os termos usados na busca (por ausência de padronização).
3. Termos de entrada de dados são extraídos diretamente dos documentos que vão constituir a base de dados.	Custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios.
4. Temas específicos citados nos documentos podem ser encontrados.	Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para cada base de dados (ex: nomes comerciais de substâncias químicas não ocorrem no Chemical Abstracts).
5. Elimina os conflitos de comunicação entre os indexadores e os usuários, pois ambos terão acesso aos mesmos termos.	Perda de confiança do usuário em uma possível resposta negativa

Quadro 03: Linguagem natural: vantagens e desvantagens
Fonte: LOPES, 2002, p. 48.

Conforme se observa no quadro acima, o princípio é o da funcionalidade, ou seja, de dar condições favoráveis à consulta do usuário. Podemos afirmar que, de modo geral, essa proposta de comparação tem o intuito de demonstrar que a LN tanto *facilita* o registro imediato daquela informação, dispensando, talvez, uma pesquisa mais específica e agilizando o processo de comunicação entre usuário e indexador, quanto *dificulta* no sentido de que, sem a devida padronização, poderá ocorrer um alto índice de respostas negativas às buscas do usuário, implicando a provável falta de confiança no sistema.

Já segundo o Quadro 04, a LC garantiria um satisfatório índice de acerto na busca do usuário, possibilitando uma estrutura hierárquica mais definida, facilitando, também, o trabalho do indexador. O argumento contra sustenta-se nos custos de produção e manutenção bem como da atualização que, no nosso ponto de vista, no balanço custo benefício, seria irrelevante dada a eficiência do sistema. Podemos afirmar que qualquer tentativa de reunião de vocábulos supõe uma pesquisa permanente para se avaliarem as novas possibilidades de emprego de um termo, ou mesmo pela demanda de novos termos que podem invalidar ou ampliar a significação, sendo passível de ocorrer em LN ou em LC.

VANTAGENS	DESVANTAGENS
1. Controle total do vocabulário de indexação, minimizando os problemas de comunicação entre indexadores e usuários.	Custos: a produção e manutenção da base de dados terão despesas maiores com a equipe de indexadores. Será necessário ainda manter pessoal especializado na atualização do *thesaurus*.
2. Com o uso de um *thesaurus* e suas respectivas notas de escopo, os indexadores podem assinalar mais corretamente os conceitos dos documentos.	O vocabulário controlado poderá não refletir adequadamente os objetivos do produtor da base, caso esteja desatualizado.
3. Se bem constituído, o vocabulário controlado poderá oferecer alta recuperação e relevância e, também, ampliar a confiança do usuário diante de um possível resultado negativo.	Um vocabulário controlado poderá se distanciar dos conceitos adequados para a representação das necessidades de informação dos usuários.
4. As relações hierárquicas e as remissivas do vocabulário controlado auxiliam tanto o indexador, quanto o usuário na identificação de conceitos relacionados.	Necessidade de treinamento no uso dos vocabulários controlados tanto para os intermediários, quanto para os usuários finais.
5. Redução no tempo de consulta à base, pois a estratégia de busca será mais bem elaborada com o uso do *thesaurus*.	Desatualização do vocabulário controlado poderá conduzir a falsos resultados.

Quadro 04: Vocabulário controlado: vantagens e desvantagens
Fonte: LOPES, 2002, p. 47.

Podemos reforçar que o Quadro 04 mostra que as vantagens assinaladas consideram aspectos mais voltados à velocidade de acesso e busca, bem como a um "provável" controle do vocabulário. Já as desvantagens apresentadas por Lopes são decorrentes da manutenção de quaisquer instrumentos de tecnologia de informação: atualização constante de dados e de pessoal especializado, e o risco de não atender a demanda dos usuários, fato inerente à capacidade humana de reestruturar significados e mudar seus objetivos de busca, além da diversidade de cada usuário em si mesmo.

Dias e Naves (2007) esclarecem que o tratamento da informação é uma tarefa intelectual, por excelência, do trabalho do indexador e envolve fatores diversos como o nível de conhecimento prévio sobre o assunto, sua formação e experiência, além de questões relativas aos processos de leitura e compreensão de textos que implicam aspectos linguísticos, cognitivos e lógicos. Os autores ressaltam, ainda, que:

> A análise documentária é definida como um conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos sob formas destinadas a facilitar a recuperação da informação. Essa passagem de um texto original para um tipo de representação é uma operação semântica, mesmo que não obedeça a nenhuma regra precisa e varie em função de cada organismo e do analista. Este seleciona as

> palavras-chave, normalmente de forma intuitiva, em função de sua ocorrência e do seu interesse para a instituição. (DIAS; NAVES, 2007, p. 11)

Dias e Naves ainda lembram que, geralmente, as normas de documentação não são elaboradas no próprio âmbito das instituições para as quais aquele tesauro é elaborado. É comum que sejam desenvolvidos e mantidos por organismos criados especificamente para tal finalidade.

Esse aspecto possibilita reflexões que também reforçam o posicionamento de Beghtol (2002a) no sentido de que um sistema deve estar preparado para a possibilidade de acréscimo de novos conceitos e também para a garantia cultural que os acompanha, uma vez que esse sistema deve ser permeável às outras garantias culturais. Essa opção já pode ser observada no tesauro do INEP - objeto de análise no próximo capítulo-, no qual há um campo de preenchimento de sugestões de inserção de novos termos, aberta aos usuários. Essa "hospitalidade cultural" exige a "habilidade de uma notação de classificação para incorporar novos conceitos em lugares apropriados e permitir o estabelecimento de relações exatas entre conceitos antigos e novos" (p. 519), além de acomodar de maneira ideal as várias garantias de diferentes culturas e refletir apropriadamente as suposições de qualquer indivíduo, grupo ou comunidade". (p. 905). A autora acredita que isso estimule o profissional, pois se trata de uma garantia ética de respeito aos indivíduos e às culturas, de modo a assegurar, além do acesso, o respeito à diversidade.

É necessário, no entanto, distinguir, entre as tarefas do indexador, aquela mais atinente às discussões desta pesquisa. Além do aspecto morfológico sinalizado na própria designação - fazer índices -, esse profissional é responsável pela catalogação de assuntos contidos nos documentos, tarefa que envolve uma leitura de caráter técnico. Dias e Naves reclamam, em seu livro "Análise de Assunto", que muito pouco é encontrado sobre como os indexadores decidem qual é o assunto de um documento, como eles decidem sobre o que ele trata. Eles referem-se às discussões de Van Dijk e Kintsch, trazendo referenciais da Linguística Textual para a indicação de uma análise proposicional, na qual se aplicariam as macrorregras (microestruturas, macroestruturas e superestruturas), para a identificação do sentido geral de um texto que é atividade fundamental para a sumarização.

Dada a complexidade da tarefa de compreensão de textos, podemos afirmar que tal discussão sugere reflexões sobre os gêneros textuais, como elementos auxiliares na identificação do assunto de um documento, bem como das palavras que o representam nos sistemas. Dias e Naves consideram o processo de identificação da estrutura e tipologia textual como processos anteriores às operações documentárias de análise e de condensação como a

elaboração de resumos, enunciados, palavras-chaves que norteiam as buscas, bem como as ações de representação daquela informação. Mas temos aqui aspectos relativos a esse sujeito e sua capacidade subjetiva de interpretar, relacionadas a questões éticas no gerenciamento das informações, conforme ressalta Beghtol.

Como não é nosso objetivo tratar especificamente dos processos de leitura (missões futuras), tentaremos enfocar as questões relacionadas às extrações de conceitos. Dias e Naves (2007) reiteram que o corpo conceitual é a base de qualquer campo da ciência; por isso mostram que sua representação forma um conjunto de termos relativo a determinado campo. Os tesauros, como coleções especializadas de sintagmas, tendo o substantivo como elemento central, por sua própria natureza, designam conceitos, e, portanto, pressupõem uma operação de corte num universo ilimitado e o interrelacionamento entre as palavras nele relacionadas.

Os autores finalizam seu livro referindo-se a uma das leis de Ranganathan: "poupe o tempo do leitor", alegando que os sistemas informatizados de buscas não levam em conta esse princípio e consideram como diferencial a quantidade de sites indexados. No ponto de vista do usuário, se fosse aplicado um processo mais seletivo, o número de documentos seria reduzido drasticamente, possibilitando, talvez, um índice de revocação mais próximo da necessidade deste. Essa é uma das qualidades da ferramenta em estudo nesta pesquisa.

Outra colocação importante de Dias e Naves é ressaltar a importância do profissional de indexação, pois "nada se fará de produtivo nessa área sem se conhecer os processos humanos de tratamento da informação, que deverão constituir necessariamente a base para o desenvolvimento dos respectivos processos automatizados." (2007, p.105). Continuam os autores reiterando que o estudo de padrões de processamento dos indexadores pode auxiliar no aperfeiçoamento da doutrina, bem como contribuir para a automatização do processo.

As discussões desta subseção reforçam a clara interseção entre a linguística e a indexação, e, especialmente a semântica, pois o significado e sua representação são os vetores das buscas automatizadas de documentos.

5 ANÁLISE DA HIERARQUIA SEMÂNTICA DE TESAUROS

5.1 Metodologia de organização de tesauros

A Unesco (Organização das Nações Unidas para Educação, Ciência e Cultura) define o tesauro como um instrumento que se aplica em duas frentes: como uma estrutura de termos semanticamente relacionados em uma área específica de conhecimento; e como um dispositivo de controle terminológico, para o tratamento e a recuperação de informações.

> É um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento [...] É um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema (linguagem de documentação, linguagem de informação) mais restrita. (UNESCO, 1973, p.6)

O tesauro, também conhecido com um "armazém de palavras", como uma lista estruturada de termos é útil para identificar o(s) assunto(s) de um documento. Nesse caso, os termos selecionados são inseridos em uma relação de índices e possibilitam uma posterior recuperação do documento. Segundo Campos (2001), a definição dos termos pode ser realizada através dos métodos dedutivo ou indutivo. No dedutivo, as áreas de conhecimento são divididas em categorias. Para cada área, os termos são definidos através do conhecimento dos especialistas dessa área. A estrutura (termos e relacionamentos) é definida à medida que os documentos da área em questão são analisados. Já no indutivo, a estrutura é definida através da análise detalhada das palavras relacionadas ao *campo de atuação*. A organização/definição dos termos inicia-se a partir de um termo específico e, para este, são localizados os seus termos genéricos.

Sabemos que a estrutura de um tesauro é composta de termos e relacionamentos existentes entre os termos (hierárquicos, de similaridade ou associativos). O relacionamento entre os termos é o mecanismo que torna possível localizar um documento, mesmo que não se utilize o mesmo termo utilizado no momento da indexação. Isso só é possível, se os termos – de busca e de indexação – possuírem algum tipo de relacionamento.

Dessa forma, a norma ISO 2788[23], Norma Internacional de Elaboração de Tesauros Monolíngues, publicada pela "International Organization for Standardization", estabelece um padrão para a criação de tesauros. Ela aponta três tipos de relacionamentos entre os termos: relacionamento hierárquico, de equivalência e o associativo. O relacionamento hierárquico é utilizado para realizar a ligação entre um termo superior (genérico) a um termo subordinado (específico). O relacionamento é realizado através dos termos:

• **NT (*Narrower Term*)** – utilizado para ligar o termo genérico ao termo específico. Representa a relação hierárquica do maior para o menor. Ex. água NT mineral.

• **BT (*Broader Term)*** – liga o termo específico ao termo genérico. Representa a relação do menor para o maior. Ex. mineral BT água.

O relacionamento de equivalência pode ocorrer entre termos *parassinônimos*, unidades lexicais com expressões verbais distintas, mas com o mesmo recorte conceitual, ou seja, a mesma referência cognitiva, que possibilitam referências conotativas diferentes, como acontece com "falar" e "dizer"; entre termos que são *quase-sinônimos*, por se apresentarem como sinônimos parciais, uma vez que não podem ser permutados em todos os contextos de uso, como se observa em "roubar" e "furtar"; e entre as *variantes lexicais* (carro/automóvel). Essa relação de equivalência é representada pelos termos:

• **USE** – utilizado para ligar o termo não preferencial ao termo preferencial. Ex. água-ardente USE cachaça.

• **UF (*Used For)*** – realiza a ligação do termo preferencial ao termo não preferencial. Ex. cachaça UF água-ardente.

A indexação e a recuperação de um documento devem ser realizadas através dos termos preferenciais. Os termos não preferenciais são utilizados para realizar o direcionamento da rotina a um termo preferencial.

Outro relacionamento é o associativo que é representado através da sigla **RT** (*Related Term*). Ocorre entre termos que não possuem uma hierarquia entre si e tampouco são equivalentes, mas possuem uma oposição entre seus significados. Ex. nutrição RT desnutrição. O relacionamento associativo pode ser:

• **Simétrico** – O relacionamento ocorre nas duas direções. Ex. professor RT aluno e aluno RT professor.

[23] A ISO é uma rede de organismos nacionais de normalização de 157 países (maio/2008) e a sua secretaria central é baseada em Genebra, na Suíça. O nome ISO, para evitar problemas de abreviações em diferentes línguas, vem da palavra "iso" em grego, que quer dizer igual. Uma norma ISO é um documento, estabelecido e aprovado por consenso, que provê, para uso comum, regras, guias e/ou características para uma atividade ou seus resultados, com o objetivo de alcançar o grau de excelência num dado contexto.

Assimétrico – o relacionamento só ocorre em uma direção. Nesse caso, a ocorrência de um implica existir o outro (Ex. briga RT raiva), pois, em tese, para ocorrer uma briga, é necessário que exista a raiva, porém a ocorrência da raiva pode não resultar em uma briga.

Optamos por iniciar a reflexão sobre os critérios organizadores, tomando como referência a Norma ISO, por acreditarmos no perfil supostamente parametrizador. No entanto, após algumas pesquisas e entrevistas com profissionais da área, descobrimos que, na prática, esse instrumento regulador interfere, em parte, na hierarquização dos conceitos. Há problemas com relação ao acesso à Norma, além de não haver uma tradução autorizada, em português. Isso gerou um número considerável de produções que visavam a orientar os profissionais da área na gestão da informação. Um bom exemplo disso é o Tutorial para Elaboração de Tesauro Documentário, implementado por Maria Luiza Campos, profissional de referência na organização de informação, que, juntamente com Hagar Espanha Gomes e Dilza Fonseca Motta, investe na formulação de métodos, bem como manuais instrucionais para a construção de tesauros.

Campos e Gomes (2004) empreenderam uma análise da referida Norma ISO 2788, com o objetivo de discutir os princípios ali estabelecidos e orientar o gestor de informação numa perspectiva crítica na alimentação do banco de dados. As autoras ressaltam que, embora o tesauro seja considerado uma linguagem de indexação, pois pode ser usado como um sistema de recuperação de informação, também pode ser considerado uma tecnologia independente, o que implica um necessário planejamento, bem como uma fundamentação teórica que resulte no controle da sua construção e manutenção.

Elas demonstram que, apesar da importância do tesauro, o Brasil não dispõe de norma própria; portanto, utiliza-se daquela disponível em âmbito internacional (2788). Tais normas são de consulta restrita (somente no *site* da ISO) e não são permitidas cópias delas; além do mais, não dispõem de uma tradução autorizada. Em que pesem os esforços empreendidos no sentido de delinear princípios para uma normalização terminológica consistente, verificam-se problemas relativos à unidade de representação; em outras palavras, ao termo de indexação. Muitos deles relativos a inserções de novos termos, de forma indevida, decorrentes de decisões tomadas para a solução de um caso específico, não generalizável. Segundo as autoras, esse controle terminológico é o que permite maior eficácia na integração de informações. Dessa forma, os princípios para organização de tesauros com base em conceitos se tornam fundamentais, porque a estruturação dos termos possibilita a interoperabilidade[24]

[24] "Habilidade de dois ou mais sistemas (computadores, meios de comunicação, redes, software e outros componentes de tecnologia da informação) de interagir e de intercambiar dados de acordo com um método

semântica. Algumas das reflexões de Campos e Gomes (2004) sobre a aplicação da norma ISO serão aqui expostas sob a perspectiva da linguagem documentária[25]. Na próxima seção, essas reflexões serão comparadas ao processamento hierárquico dos conceitos, numa perspectiva linguística. Foram aqui selecionadas aquelas que se referem à definição de "termos", pois acreditamos que coadunam os objetivos desta tese, relativos à hierarquização de conceitos.

Seria importante salientar, ainda, que, num tesauro, a recuperação da informação não acontece pelas palavras tais como estão nos textos. No *Google*, por exemplo, quando buscamos a palavra "comida", temos a recuperação de possibilidades como "comida chinesa". Porém, num tesauro, não cabe adjetivação, pois nele são indexados apenas *substantivos* (classe de palavras). Segundo as autoras; portanto, não se trata apenas de uma fatoração sintática ou morfológica.

Por isso, a unidade de descrição no tesauro não deve ser a *palavra*, mas o *conceito*, materializado no *termo*. Cabe aqui esclarecer que, para as autoras, o *termo* pode ser constituído por uma palavra ou por um grupo de palavras e *representa um conceito* e, como tal, indivisível na indexação e nos tesauros. Elas trabalham sob a perspectiva de que o *termo* tem seu significado assegurado, *mesmo fora de contexto, ou seja, isoladamente*. Reforçam, ainda, que a análise do conceito se dá a partir do referente e não da palavra, porém reconhecem que é a expressão verbal que funciona como elemento de manipulação, então tal análise é feita via *termo*.

A partir das discussões anteriores, podemos dizer que, na perspectiva linguística, o conceito seria um tipo de representação, fruto da experiência sociointerativa, cuja representação cognitiva constitui o núcleo e a especificidade das diferentes línguas, pois cada uma pode possuir uma forma de representação léxica. Já na perspectiva das teorias documentárias, o conceito adquire uma perspectiva mais voltada ao processo comunicativo, também tratado, obviamente, na perspectiva linguística, mas, antes de tudo, contemplando facilidade de recuperação da informação, a partir daquele conceito; por isso é, ao mesmo tempo, mais padronizada na tentativa de reduzir ao mínimo a ocorrência de ambigüidades. Como no tesauro a pesquisa já estaria restrita a uma área específica do conhecimento, ele elimina grande parte das milhares de ocorrências trazidas numa consulta a instrumentos de

---

definido, de forma a obter os resultados esperados." (ISO); tem por meta a consideração de todos os fatores para que os sistemas possam atuar cooperativamente, fixando as normas, as políticas e os padrões necessários para consecução dos objetivos propostos.

[25] Essa escolha se deve ao fato de acreditar que uma análise feita por profissionais da informação estamparia as concepções daquela área, para o pretendido confronto com a linguística.

busca como o *Google*. Uma pesquisa sobre o termo "comida" num tesauro sobre alimentação, dentre as opções disponibilizadas ao usuário estariam: nutrientes e vitaminas; nos instrumentos de busca, certamente estaria "aquilo que se come", bem como "cozinha" e "culinária", e até mesmo "pessoa com quem se tem relações sexuais costumeiras ou pessoa que se entrega sexualmente, em geral de maneira passiva", "bexiga de rês" ou "ganho ilícito proveniente de arranjo ou negócio escuso; comedeira".

Na perspectiva documentária, o conceito é visto como uma unidade indivisível formada pelo referente[26], pela característica e forma verbal. Afirmam as autoras que

> Se a palavra tem seu significado evidenciado no contexto frasal, o termo, por seu turno, carrega consigo o significado/conteúdo. O termo assim é constituído por uma forma verbal e um conteúdo conceitual. A homonímia fica, em geral, eliminada nos tesauros, porque estes cobrem domínios específicos do conhecimento e é no contexto do domínio que o termo evoca, para a comunidade do domínio, o conteúdo que tem que [sic] evocar. (CAMPOS; GOMES, 2004, p.4-5).

Para minimizar ocorrências como a homonímia e a polissemia, são estabelecidas regras de relacionamento entre palavras, entre elas, as relações conceituais, hierárquicas (ou de implicação), partitivas e/ou de oposição a serem detalhadas mais à frente. Dessa forma, os termos seriam agrupados pelas características dos conceitos (atributos) obtidas pela própria formulação do enunciado. Segundo Dodebei (2002), podem ser diferenciadas duas categorias: *intrínsecas*, que são aquelas inerentes ao objeto e que constituem sua essência; *extrínsecas*, aquelas que dependem de fatores externos e que podem ser geral ou individualizantes. Assim, podemos dizer que o conhecimento das características conceituais permite maior índice de acerto nos ajustes à ordenação classificatória. No entanto, conversas informais com profissionais elaboradores de tesauros dão conta de que, quando não é possível eliminar consideravelmente as questões de ambiguidade entre os termos, e se isso compromete a hierarquização, tais referências são automaticamente excluídas.

Entre as importantes contribuições da Terminologia para a elaboração de tesauros, conforme apresentado em seção anterior, deste trabalho, estão a tentativa de estabilizar um conceito e suas relações com outros conceitos, como num sistema, ou seja, como eles se definem uns em relação aos outros, pois, um conceito precisa de outros conceitos para poder ser apreendido. Para dominarmos os conceitos, precisamos conhecer as características de cada um dos objetos a que eles se referem, as quais são, por sua vez, também conceitos. Por isso,

[26] Parece-nos que o entendimento das autoras com relação ao referente está relacionado aos elementos do mundo extralinguistico ao qual remete o signo linguistico, num determinado contexto sociocultural e de discurso.

um instrumento de organização como o tesauro, implica, além de uma competente equipe para sua elaboração, uma freqüente atualização para a incorporação de novas possibilidades de inserção de termos relacionados. Por exemplo, em um tesauros de educação, o conceito de alfabetização certamente seria revisto a partir do contraste com a recente discussão sobre parâmetro do *letramento*.

Campos e Gomes (2004) indicam como exemplo, as expressões "conhaque" e "conhaque de gengibre"; caso se considerasse apenas a forma verbal, *conhaque* seria o termo geral e *conhaque de gengibre* seria o termo específico. No entanto, se houver uma análise das características dos objetos a que os termos se referem, pode-se concluir que conhaque é *um tipo de bebida alcoólica fermentada-destilada a partir do vinho*; e que o conhaque de gengibre é *um destilado de cana-de-açúcar adicionado de gengibre, como substância aromática.* Dessa forma, pode-se estabelecer uma relação associativa mais adequada entre os termos, necessária para a construção de um tesauro. As autoras reforçam, ainda, que a análise do referente é o que possibilita estabelecer *relações lógicas* (de superordenação/ subordinação), o que requer utilizar a análise das propriedades do objeto/referente, e *relações ontológicas* (partitivas e associativas), quando se observa sua relação com outros termos, como, por exemplo, sua função na realidade. Esse aspecto será objeto de discussão, conforme sinalizado anteriormente, na próxima seção.

Nessa versão da Norma ISO 2788, em análise por Campos e Gomes (2004), o conceito foi definido como uma *unidade de conhecimento.* O termo seria a síntese, pois traz consigo as características do objeto/referente. De acordo com a Norma ISO 1087 - Vocabulário da Terminologia, "termo é definido como designação de um conceito definido numa língua especializada, por uma expressão linguística". Diante disso, o termo seria essa materialização, expressa por uma palavra ou por um grupo de palavras.

Um dos objetivos do tesauro é, segundo a Norma ISO 2788, assegurar "uma prática consistente num serviço de indexação ou entre diferentes serviços", como, por exemplo, numa rede. Já nesse conceito, as autoras apontam uma inconsistência, pois a Norma situa como "linguagem de indexação, o conjunto controlado de termos selecionados da linguagem natural usados para representar, de modo resumido, o assunto dos documentos". Questionam também o vocábulo "controlado", dadas as limitações de um tesauro, pois nele não é possível visualizar o termo num contexto linguistico determinado, o que relativiza o conceito. Além disso, Campos e Gomes afirmam que não há uma clara definição do que é o "termo", nem clareza na definição do que seja "assunto", na linguagem de indexação.

É curioso observar no discurso das autoras uma referência constante à inconsistência das análises linguísticas, em detrimento de uma análise terminológica que garantiria maior eficácia nas definições. São conhecidas as dificuldades de delimitação de fronteiras nas questões sobre sentido; isso não justificaria, porém, que pressupostos linguísticos não fossem contemplados de forma mais efetiva. Em nossas reflexões primárias, consideramos pouco provável que se possa fazer uma análise dessa ordem, sem levar em conta as abordagens linguísticas, conforme se verá na seção seguinte. Há aspectos importantes, principalmente em relação à determinação dos campos semânticos que poderiam contribuir para os ajustes organizacionais necessários às hierarquias formuladas nos tesauros.

Há, na Norma ISO 2788, uma definição de tesauro, como sendo "o vocabulário de uma linguagem de indexação controlada e organizada formalmente, explicitando "a priori" relacionamentos entre os conceitos (por exemplo, mais geral e mais específico)." Campos e Gomes colocam em equivalência os "vocabulário" e "linguagem de indexação", como um conjunto de termos, considerando a definição circular. Assinalam, ainda, que permanece a falta de definições, por exemplo, do que se entende por "conceito"; elas avaliam até que ponto seriam as Normas realmente orientadoras, haja vista que o profissional da informação estabelece relações hierárquicas na ordenação dos termos.

A Norma define *termo de indexação* como "representação de um conceito preferencialmente sob a forma de uma palavra ou expressão." Notamos que, embora não haja definição, na Norma, do que seja conceito[27], admite-se que o termo pode ser representado por uma palavra ou uma expressão, acrescentando uma nota que diz: "um termo de indexação pode consistir em mais de uma palavra e então é conhecido como termo composto." Essa postura, de acordo com as autoras, coloca em xeque o entendimento sobre a unidade de representação dos tesauros documentários, já que a própria Norma ISO 2788 estabelece que "termo composto é um termo de indexação que pode ser fatorado morfologicamente em componentes separados, cada um dos quais pode ser expresso ou re-expresso como um substantivo capaz de servir independentemente como termo de indexação." Na análise de Campos e Gomes, a Norma privilegia o aspecto verbal do termo de indexação e, segundo elas, se considerarmos que o termo representa o conceito na "fatoração"[28], os constituintes do

[27] Há que se ressaltar que, na perspectiva da CI, o conceito seja entendido como a representação intelectual de um objeto concreto ou abstrato, conforme apresentado na seção 4.6 deste trabalho.

[28] A partir de uma referência matemática, é aqui entendida como uma forma de decomposição dos termos. Nos tesauros é chamada de fatoração semântica e se aplica quando se reconhece mais de um conceito em uma forma verbal composta. Há também regras de fatoração sintática de expressões que representam declarações de assunto, principalmente na fase de extração dos termos da literatura corrente, e, posteriormente, na análise de definições conceituais, sendo determinado os termos autorizados, ou descritores.

termo poderiam ser isolados, comprometendo a significação. Apesar de concordarmos em parte com tal reflexão, podemos inferir que há um equívoco da parte das autoras no que diz respeito a privilégios, mesmo que estejam trabalhando com uma noção de forma lexical.

A título de ilustração, numa estrutura hierárquica de "alimentos", o termo composto "pé-de-moleque" deveria ser excluído, porque nem o termo "pé" e nem o termo "moleque", isoladamente, seriam passíveis de inclusão numa lista organizada sobre alimentos. Tal pertinência poderia ser viabilizada numa outra lista que contivesse uma classe geral que comportasse o termo "pé", como parte do corpo; ou numa lista em que "pé" indicasse cada unidade de uma determinada planta, como de árvore frutífera na qual poderíamos incluir "pé de laranja", ou numa de hortaliças, em que caberiam "pé-de-couve", "pé-de-alface, "pé-de-almeirão". Dessa forma, o termo "moleque" poderia ser incluído numa lista de gírias, como sinônimo de criança.

Nesse aspecto, concordamos com as autoras, mas questionamos, a partir do exemplo comentado, que se possa privilegiar parte do termo; mas, se assim fosse, simplesmente, o emprego dessa norma seria inviável na análise de boa parte dos termos compostos.

Algumas partes dessa Norma (2788), apresentadas a seguir, podem demonstrar uma falta de clareza nos parâmetros orientadores, tão peculiares a instrumentos reguladores como os da ISO. Na língua portuguesa, um substantivo composto é formado por mais de um radical ou mais de uma palavra, como "aguardente", "floricultura", ou "couve-flor" e "pôr do sol". Geralmente, esses substantivos podem ser formados por um substantivo seguido de outro substantivo (ex: couve-flor), por um substantivo e um adjetivo (ex: bóia fria), por um verbo e um substantivo (ex: porta bandeira), um substantivo e um elemento invariável (ex: abaixo assinado) ou por palavras repetidas (ex: pisca-pisca). O elemento em acréscimo ao substantivo, geralmente, o caracteriza, particularizando-o; isso certamente torna inviável a secção. Parece-nos que o processo de tradução da Norma, não levou em conta as especificidades de cada língua (ou levou em conta apenas a língua inglesa...). Vejamos uma extensa nota sobre os termos compostos, na qual não se recomenda a fatoração:

> As partes da grande maioria de termos compostos podem ser distinguidas como segue:
> a) O foco, isto é, o substantivo componente que identifica a classe geral de conceito à qual o termo como um todo se refere.
> Exemplo 1: o substantivo componente 'indexes' (índices) no termo composto 'printed indexes' (índices impressos).
> Exemplo 2: o substantivo 'hospitals' (hospitais) na expressão 'hospitals for children' (hospitais para crianças).

> b) A diferença ou o modificador, isto é, um ou mais componentes que servem para diminuir a extensão do foco e assim especificar uma de suas subclasses.
> Exemplo 1: o adjetivo 'printed' (impresso) no termo composto 'printed indexes' (índices impressos). (...publicados)
> Exemplo 2: a preposição + substantivo 'for children' (para crianças) no termo composto 'hospitals for children' (hospitais para as crianças).
>
> [...]O foco e suas diferenças podem se apresentar com palavras separadas, como em 'dining rooms' (sala de jantar) ou 'soup spoons' (colheres de sopa) ou podem vir concatenadas numa palavra, como em 'bedrooms' (quartos, dormitórios, "sala de dormir") ou 'teaspoons' (colheres de chá)....[...] [29] (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Campos e Gomes avaliam, por exemplo, que a nota *a*) indica que não deve haver fatoração, se um dos elementos do termo composto designa o termo específico da classe, como no caso de "hospitals for children", assim como "carro de passeio", nos quais há dois substantivos, um especificando o outro. Caso haja a separação dos termos, eles deveriam ser indexados em classes diferentes, ou até mesmo em tesauros diferentes, conforme já apresentado em exemplo anterior ("pé de moleque"). Elas ainda consideram que, dessa vez, houve um privilégio em relação ao conceito, orientando a opção por uma das formas considerada mais geral, o substantivo hospital, em detrimento da locução adjetiva ("de criança"), visão um tanto limitada do papel do adjetivo o qual permite determinar, particularizar um dado termo, muitas vezes alterando consideravelmente sua significação, como no exemplo de "pé-de-moleque".

Já para a nota *b*), acreditam as autoras que o problema está na manutenção do mesmo exemplo para duas situações: em *a*) o foco caracteriza o todo, em *b*) a diferença especifica. Isso torna opaca a orientação. Acreditamos que mais problemática é a questão de se considerar o "modificador"-no português, o adjetivo-, como elemento que *reduz* a extensão do foco. Tomemos como outro exemplo, o termo "salão" modificado para "salão de beleza" ou "salão de jogos" ou, ainda, "dança de salão"; nele a locução adjetiva especifica o termo "salão", caracterizando, assim, um novo membro para o conjunto, ou outra classe, como no caso de "dança de salão". Há uma recomendação sobre palavras separadas ou aglutinadas nos termos compostos, "girassol" e "televisão", porém em nada contribui ao perfil orientador que uma norma deve ter. Seria necessária, talvez, uma discussão sobre a noção de *sintagma*, em relação ao que a Norma entende como *termos compostos*.

No capítulo 6 da Norma, os termos de indexação são classificados do ponto de vista conceitual, ou seja, os termos designam:

[29] Tradução livre de diversos vocábulos e/ou expressões das Normas ISO 2788, daqui em diante.

1) Entidades concretas

* coisas e suas partes físicas

* materiais

2) Entidades abstratas

* ações ou eventos

* entidades abstratas e propriedades de coisas, materiais ou ações

* disciplinas ou ciências

* unidades de medida

3) Entidades individuais, ou classes-de-um, expressas por nomes próprios.

Numa perspectiva analítica, tal classificação está aqui entendida como um arranjo (até arbitrário em alguns momentos), cuja funcionalidade prescritiva tem o objetivo de nortear futuras classificações. No entanto, faz-se necessário refletir sobre os critérios empregados para o estabelecimento de cada classe, de modo a se estabelecer uma conduta padronizada, típica de instrumentos da ISO. Nesse sentido, Campos e Gomes reiteram que a falta da definição de conceito inviabiliza a classificação e induz a uma abordagem linguística do termo, "ao incluir um capítulo sobre a forma verbal, fornecendo orientação sobre o uso de substantivos, adjetivos, advérbios e verbos e, ainda, sobre o uso de singular e plural." (2004, p.10).

O capítulo 7 da Norma ISO 2788 trata de fatoração e nos interessa em particular, o item 7.1

> A fatoração semântica aplica-se a termos simples e compostos. Um termo que representa uma noção complexa é re-expressa na forma de um termo mais simples ou da definição, cada um dos quais pode também ocorrer em outras combinações para representar uma série de diferentes conceitos. Esta técnica não é recomendada. **(Grifo nosso).** (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Essa *fatoração semântica* implica uma seleção de características do conceito, as quais podem ser também conceitos estruturados no sistema, podendo ocorrer em outras combinações. Conforme se verifica na citação, curiosamente, a Norma não recomenda essa técnica. Segundo as autoras, no entanto, seria uma técnica adequada, por assegurar um campo maior de precisão, em áreas tão específicas, típicas do tesauro; do ponto de vista conceitual, esse fatoramento pode fazer sentido. O exemplo citado pelas autoras é o a expressão "leite de soja" que, no âmbito de um instituto de tecnologia alimentar, não é muito utilizado; porém duas de suas características o são: os conceitos *soja* e *extrato proteico*.

Além da fatoração semântica, há, na Norma, a fatoração sintática: “ela pode ocorrer na análise morfológica em componentes separados, cada um dos quais pode ser aceito como um termo de indexação por si só. Não recomenda esta técnica para termos sincategoremáticos[30]”, - que, na verdade, nem seria possível. Na perspectiva de organização de dados em um tesauro, que exige relação entre os termos, se fatorarmos o termo composto, na perspectiva da norma, o conceito representado pelo termo deixa de existir, permanecendo apenas os seus componentes. Quanto à questão da não recomendação de uso da técnica nos termos sincategoremáticos, deixa vaga a orientação, pois se este denota um conceito, torna-se indivisível. Além do mais, um termo só poderia ser sintecategoremático, se considerada sua relação com outros elementos da sintaxe, sendo talvez desqualificado como um termo inadequado à indexação.

Notamos que se alternam na Norma itens de orientação linguística e outros de base mais terminológica, porém, numa perspectiva contrastiva e não complementar, como pretendemos demonstrar neste trabalho. No item seguinte, em 7.2, orienta-se em casos de não-fatoração, como nos seguintes:

> a) termos muito familiares no uso comum. Exemplo: 'data processing' processamento de dados).
> b) perda de sentido' quando se combinam os termos. Exemplo 'food + plants' pode significar tanto 'food for plants', como 'plants as food' (alimento + plantas pode significar tanto ‘alimento para as plantas’, como ‘plantas como alimento’). (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Campos e Gomes (2004) questionam a ambiguidade do termo “familiares”, pela falta de parâmetros de inclusão (traços distintivos), no sentido de se estabelecer como considerar um termo como familiar: se na perspectiva da frequência de uso comum à área daquele tesauro ou da relação conceitual entre os ternos, como no caso de “aguardente” e “planalto”, palavras consideradas de uso bastante comum, ou “topa-tudo”, num tesauro de móveis, e “bicho de pé”, num tesauro de dermatologia. Comentam, ainda, que a expressão "food for plants" (alimento para as plantas) é um conceito, pois denota uma classe de substâncias para nutrir as plantas. Mas "plants as food" (plantas como alimento) já delimita uma classe específica de plantas, com certas características que a tornam comestíveis. O que elas querem

[30] Os termos que significam de modo fixo e determinado (termos que exercem a função de sujeito ou de predicado, que são representados pelas categorias substantivos, os adjetivos, os pronomes pessoais, os pronomes demonstrativos e os verbos - exceto os auxiliares), são qualificados de categoremáticos; aqueles cuja significação não é definida são chamados de sincategoremáticos (incluem-se aí os advérbios, as conjunções e as preposições). Estes podem associar-se àqueles, alterando a significação de ambos ou contribuindo tanto para a formação quanto para determinação das condições de verdade das proposições.

demonstrar é que um pode representar um *conceito* e o outro uma *classe* de inclusão, dependendo da área de abrangência do tesauro em que estão sendo inseridas, atribuindo-se essa propriedade a qualquer um dos dois termos. Poderíamos até mesmo concluir que a Norma torna-se circular.

Na Subseção 7.2.1 da Norma, encontra-se uma orientação mais rigorosa indicando termos compostos que não devem ser separados, como

> a) nomes próprios.
> b) termos em que a diferença perdeu seu significado original. Exemplo: 'lawn tennis' (tênis de grama/quadra) e 'trade winds' (ventos alísios).
> c) Termos que contêm uma diferença que sugere uma semelhança como um símile, com uma coisa ou evento não relacionado. Exemplo: 'tree structures' (estrutura arbórea), 'wing nuts' (asa/nozes: cavilha de orelha, borboleta).
> d) Termos que não podem ser re-expressos (por exemplo, definidos), sem o uso de um substantivo extra que esteja presente no termo composto apenas por implicação. Exemplo: 'free escapes' (escapes for people from fire). (área livre para escape; fuga para pessoas em caso de incêndio- escada de incêndio).
> e) Termos que contêm nomes sincategoremáticos. Exemplo: 'artificial limb' (membros artificiais); 'fossil fishes' (fósseis de peixes/ peixes que são fósseis). (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

As autoras problematizam que seria irrelevante a busca de um significado original, bastando conhecer a qual objeto o termo se refere. Notamos que há uma tentativa de designação de conceitos, mesmo que a orientação se refira, claramente, à questão de isolar ou não as palavras constitutivas do termo. Há que se ressaltar, também, que os exemplos dados para as formulações nessa subseção são inadequados, pois não representam a realidade dos fatos linguísticos. Pensamos ser necessário um ajuste referente a questões de tradução, não somente do ponto de vista da palavra em si, mas das relações filológicas (item a) bem como as sintagmáticas (item b), pois implicam o comprometimento da instrução, como um modelo paramétrico e padronizador na língua portuguesa.

A Seção 7.3.1 trata de termos que devem ser fatorados sintaticamente:

> a) um termo composto deve ser fatorado, se o foco se refere a uma propriedade ou parte (inclusive materiais) e a diferença representa o todo ou o possuidor daquela propriedade ou parte. Exemplo: 'aircraft engines' (motores de aeronaves); 'hospital personnel' (pessoal hospitalar); 'soil acidity' (a acidez do solo).
> b) ao contrário, o nome do todo pode ser modificado pelo nome de sua parte ou propriedade. Exemplo: 'acid soils' (solos ácidos), 'concrete bridges' (concretas pontes); 'skilled personnel' (pessoal qualificado). (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Nesse item, Campos e Gomes assinalam que a Norma orienta no plano da língua inglesa, conforme já indicado acima, e, portanto, sem aplicação na dimensão da língua

portuguesa do Brasil. Trata-se de um termo composto por dois substantivos, um determinando o outro. Com a tradução das expressões, o resultado se enquadraria nos casos em que a Norma não recomenda fatorar; em "motor de avião", sabemos que se trata de um motor *para* avião, como ocorre em português com "copo de leite", a flor, e "copo com leite", copo com o líquido; "creme de cabelo" e "creme de leite". Já no caso dos adjetivos, elementos modificadores, não devem ser separados dos substantivos em função da descaracterização do termo, pois, no tesauro, conforme esclarecido, só cabem substantivos; assim sendo, os adjetivos seriam eliminados, comprometendo a significação.

A subseção 7.3.2 da Norma trata de expressões relativas a ações ou eventos:

> a) o nome de uma ação transitiva não deve ser modificado pelo nome do objeto sobre o qual a ação se efetua. Exemplos: 'office management' (gerência, gabinete de gestão); 'steel casting' (vazamento de aço); 'textile printing' (impressão têxtil, textura).
> b) ao contrário, o nome de uma coisa ou material pode ser modificado pelo nome de uma ação efetuada sobre ela/ele. Exemplo: 'cast steels' (aço fundido), 'extruded plants' (expulsão, saída forçada de plantas, máquina de aparar grama); 'printed textiles' (impresso têxteis, impresso no tecido). (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Reforçamos que, nesses itens, apesar da viabilidade linguística na fatoração do termo, haverá perdas significativas em relação ao conceito. Curiosa é a expressão/sintagma "ação transitiva", pois seria uma questão relativa à predicação verbal, a transitividade dos verbos, talvez, pois o substantivo seria o objeto que sofre a ação. Então teríamos em português: "salva-vidas", "pára-raio". Além disso, seria necessário rever a expressão "nome de uma ação", pois as ações podem ser designadas por substantivos ou verbos. Ainda com relação a ações ou eventos, as autoras apresentam uma última inconsistência do caráter internacional da Norma ISO, na qual ressaltam haver um modelo de análise linguística próprio de cada língua:

> a) o nome de uma ação intransitiva não deve ser modificado pelo nome do realizador da ação. Exemplos: 'bird migration' (ave migratória); 'plant growth' (crescimento de plantas); 'metal deterioration' (deterioração do metal).
> b) ao contrário, o nome de uma coisa pode ser modificado pelo nome da ação intransitiva na qual ela está ou foi envolvida. Exemplos: 'degraded plants' (plantas degradadas); 'migrating bird ' (ave migratória); 'moving pavements' (pavimentos móveis). (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. 2788, 1986.)

Nesse caso, sendo uma ação intransitiva, poderíamos considerar o termo "recém-nascido" ou "bem viver"; mas parece-nos desnecessária essa sutil separação entre ações transitivas e intransitivas, ou pouco aplicável.

Numa análise mais geral, poderíamos afirmar que um dos maiores problemas apontados pelas autoras foi a falta de clareza do que seja um conceito, e se conceito é entendido como termo. Dessa forma, não caberiam discussões sobre fatoramento sintático ou morfológico, pois o termo que denota o conceito ficaria sem estrutura no tesauro, já que o conceito representado por um grupo de palavras seria sumariamente eliminado. Assim, as autoras defendem que, na construção de tesauros, sejam empregadas a Teoria do Conceito e a Teoria da Terminologia, como fundamentos teóricos para a classificação e ordenação dos dados, porque consideram que a Norma 2788 estaria mais fundamentada na forma lexical, tornado-a inconsistente em razão das peculiaridades de cada língua. Ressaltam que o tesauro exige uma estrutura de conceitos, ajustada a domínios específicos de conhecimento. Finalizam a análise afirmando que

> Ferramentas semânticas como tesauros precisam ser construídas respeitando princípios teóricos consistentes, abarcando domínios de estudos e pesquisa em língua portuguesa, para serem utilizadas em bibliotecas digitais e virtuais, em sistemas para gestão de conhecimento, para viabilizarem processos de integração de informações entre pesquisadores, como auxílio para as ferramentas de busca de um modo geral, e principalmente como um instrumento para a melhoria do tratamento e da recuperação de informação na rede. Além disso, o uso de conceitos como base para o estabelecimento dos termos em um tesauro permite maior facilidade para o estabelecimento de equivalentes em outras línguas. Atualmente, o crescimento da utilização de arquiteturas distribuídas, especialmente no ambiente da Web, de interfaces abertas de acesso a bancos de dados, de tecnologias de mediadores e de padrões de formato para troca de dados contribuiu para que informações originalmente isoladas sejam disponibilizadas para serem utilizadas de maneira integrada. Tesauro assume papel fundamental nesta integração, viabilizando a interoperabilidade semântica de sistemas distribuídos heterogêneos. (CAMPOS; GOMES, 2004, p.15-16).

Diante de algumas das inconsistências apresentadas nesta seção, por profissionais de referência na área, e, após uma apresentação de linhas teóricas que fundamentam o trabalho documentário, pretendemos, à luz da análise de alguns exemplares de tesauros, mapear os critérios de caráter linguistico utilizados, a fim de auxiliar a suprir a necessidade premente de parametrização de condutas na organização da informação. É importante atentar para o fato de que esse mapeamento dos critérios hierarquizadores, talvez, possa ser empregado em outras diferentes instâncias. Mas é preciso deixar claro que não se trata de estabelecer critérios porque já os utilizamos intuitivamente, mas, sim, seria descrevê-los, dar clareza sobre os seus usos e ajustá-los às necessidades enunciativas da atualidade, muitas delas amparadas em tecnologias da informação.

Apresentar a Norma foi uma medida comparativa. O que pretendemos demonstrar é que mesmo sendo um instrumento regulador, não há muita consistência com referência aos

critérios como os que foram ressaltados. Apesar de haver uma certa engenharia proposta pelo próprio instrumento, pouco se fornece ao profissional que irá examinar documentos, extrair a essência temática, fazer levantamentos de palavras sobre aquela área e promover a organização hierárquica. Podemos afirmar que há pouca clareza nesse sentido; ou há manuais e tutoriais que tratam a questão de uma forma mais mecânica, ou a referência de uma Norma de certificação pouco esclarecedora.

5.2 Por uma análise linguística de critérios organizadores

5. 2. 1 Metodologia empregada

Para esta análise foram selecionados três tesauros: o Thesaurus Brasileiro da Educação (Brased), do INEP, já pré-analisado no capítulo 02; o Tesauro em Ciência da Informação que foi construído a partir de outros tesauros; e o Tesauro de Folclore e Cultura Popular Brasileira, tomado como corpus central das análises desta pesquisa. Há que se ressaltar que foram inseridas, neste trabalho, apenas partes dos referidos tesauros, haja vista que são imensos conjuntos de informações, cujos endereços eletrônicos são de fácil acesso e livre consulta. Na medida do possível, foram recortadas: a matriz conceitual norteadora, uma imagem que oferece uma impressão geral (ou mais ampla) do tesauro e cadeias e renques para efetuarmos a análise da hierarquia.

Os aspectos principais a serem observados situam-se na identificação do tipo de relação estabelecida entre os termos, e, principalmente, na tentativa de compreender os critérios utilizados, de modo a trazer reflexões linguísticas para esse processo. Pretendemos, ainda, mostrar até que ponto as relações de hiperonímia/hiponímia e de meronímia/holonímia interferiram na hierarquização.

5. 2. 2 Análise da hierarquia semântica do Thesaurus Brasileiro da Educação (Brased) – INEP

Embora já tenha sido objeto de análise no Capítulo 2, deste trabalho, cabe aqui ressaltar alguns aspectos da hierarquia dos conceitos. As orientações que constam na apresentação do tesauro do INEP são de caráter geral, aplicáveis a qualquer tesauro:

> a) No *thesaurus* cada termo corresponde a um conceito. Uma vez aceito, esse termo torna-se um "descritor" ou um "indexador". Caso o termo não seja aceito como "descritor", ele pode ser aceito como "remissivo", isto é, remete para um termo autorizado (descritor).
> b) Todos os termos estão relacionados entre si. Nenhum termo pode figurar no *thesaurus* sem estar relacionado a algum outro, sendo essa relação determinada pelo seu significado.
> c) A conceituação dos termos e as relações entre eles são definidas pelo sistema ontológico da área e pelo estudo de cada termo.
> (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA-INEP)

Não há muitas orientações descritivas no *site*, além de a hierarquia ser estruturada dentro da lógica da matriz conceitual, conforme exposto no Capítulo 2, com quatro campos temáticos: Contexto da Educação, Escola como Instituição Social, Fundamentos da Educação e Educação: princípios, conteúdos e processos. Esclarece-se, ainda, que os termos são ordenados de acordo com as relações lógico-ontológicas - de hierarquia, de equivalência e de associações - existentes entre eles. Há orientações sobre o que é termo, conceituações, documentos indexados e prazo de validade, porém há poucas facetas que contêm conceitos.

O campo temático escolhido para análise foi o do Contexto da Educação, enfocando o termo Academia de Educação Física, para o qual temos a seguinte estrutura das relações hierárquicas, conforme o fragmento da Figura 01, já apresentado anteriormente:

- CONTEXTO DA EDUCAÇÃO
  - CONTEXTO AMBIENTAL
  - CONTEXTO HUMANO
  - CONTEXTO SOCIAL
  - CONTEXTO CULTURAL
    - MEIO CULTURAL
    - POLÍTICA CULTURAL
    - LAZER E EDUCAÇÃO
    - ESPORTE E EDUCAÇÃO
      - ESPORTE
      - ATIVIDADES ESPORTIVAS
        - MODALIDADE DE ESPORTE
        - ORGANIZAÇÃO DO ESPORTE
          - ENTIDADES ESPORTIVAS

- **ACADEMIA DE EDUCAÇÃO FÍSICA**
- ASSOCIAÇÃO ESPORTIVA
- CENTRO DE ESPORTES
- CLUBE ESPORTIVO
- REPRESENTAÇÃO ESPORTIVA NACIONAL
- CONTEXTO POLÍTICO
- CONTEXTO ECONÔMICO
- CONTEXTO MUNDIAL
- ESCOLA-INSTITUIÇÃO
- FUNDAMENTOS DA EDUCAÇÃO
- EDUCAÇÃO
- IDENTIFICADORES E ESPECIFICADORES DE INFORMAÇÃO

"Academia de educação física" está contida imediatamente em vários subconjuntos, como o das "Entidades esportivas", e no de "Organização do esporte" como uma classe mais geral; mantendo a relação hiperonímia e hiponímia (numa relação gênero e espécie), está também subordinada às classes "Atividades Esportivas" e "Esporte e Educação", mais amplas, até chegar ao "Contexto Cultural", superordenado por "Contexto da Educação". Também nesse tesauro, são ordenados apenas termos, em relações associativas e hierárquicas de pertinência.

O exemplo acima já foi problematizado no Capítulo 2, desta tese. Percebemos que "Entidades Esportivas" seria a classe geral e as demais estão ali colocadas em padrão de equivalência: "Academia de Educação Física, Associação Esportiva, Centro de Esportes; Clube Esportivo", embora saibamos que cada uma delas guarde peculiaridades.

O tesauro do INEP apresenta, ainda, uma listagem com os Identificadores e Especificadores de Informação que classificam e complementam as informações relacionadas aos quatro primeiros campos, conforme quadro a seguir, no qual se situa a área escolhida.

100 CONTEXTO DA EDUCAÇÃO	**200 ESCOLA**	**260 Profissionais da Educação**	**300 FUNDAMENTOS DA EDUCAÇÃO**
110 Contexto ambiental	**210 Pesquisa da educação**	261 Administração de Recursos Humanos	**310 Cultura e educação**
111 Meio Ambiente	211 Realidade da Educação	262 Funções Docentes	311 Cultura
112 Problemas Ecológicos	212 Informação e Educação	263 Funções Não-Docentes	312 Cultura e Religião
113 Política Ambiental	213 Educação Comparada	264 Magistério	313 História das Civilizações
114 Meio Ambiente e Educação	**220 Estatística da educação**	**270 Instituições de ensino**	**320 Filosofia e educação**
120 Contexto humano	221 Estatística	271 Natureza Jurídica das Instituições	321 Filosofia
121 Pessoa	222 Estatísticas do Contexto Educacional	272 Escolas	322 Processo Filosófico
122 Família	223 Estatísticas Educacionais	273 Instituições de Educação Básica	323 Filosofia Fundamental
123 Comunidade	224 Estatísticas Gerenciais	274 Escolas para Formação Profissional	324 Filosofia Prática
130 Contexto social	225 Estatísticas dos Profissionais da Educação	275 Instituições de Ensino Superior	325 Filosofia das Ciências
131 População	226 Indicadores Econômicos em Educação	276 Instituições de Educação Especial	**330 Trabalho e educação**
132 Realidade Social	**230 Política da educação**	277 Centros de Educação de Adultos	331 Trabalho Humano
133 Condições Sociais	231 Política	**280 Administração escolar**	332 Psicologia do Trabalho
134 Política Social	232 Educação como Ação Política	281 Administração das Instituições de Ensino	333 Relação Educação/Trabalho
140 Contexto cultural	233 Política Mundial da Educação	282 Gestão Administrativa da Escola	334 Administração do Trabalho
141 Meio Cultural	234 Direito Educacional	283 Gestão Pedagógica da Escola	**340 Psicologia e educação**
142 Política Cultural	**240 Administração da Educação**	284 Organizações Escolares	341 Psicologia
143 Lazer e Educação	241 Administração	285 Registro Escolar	342 Psicologia da Personalidade
144 Esporte e Educação	242 Administração da Educação Escolar	**290 Economia da educação**	343 Psicologia do Desenvolvimento
150 Contexto político	243 Planejamento da Educação	291 Economia	344 Psicologia Social
151 Meio Político	244 Produtividade e Avaliação da Educação	292 Financiamento da Educação	345 Psicologia Cognitiva
152 Constituição Nacional	**250 Educandos**	293 Administração Financeira	346 Psicologia da Dinâmica
153 Governo da Nação	251 Tipos de Educandos	294 Infra-Estrutura Escolar	347 Psicologia da Educação
160 Contexto econômico	252 Vida do Educando		**350 Sociologia e educação**
161 Meio Econômico	253 Assistência ao Educando		351 Sociologia
162 Mercado de Trabalho			352 Estrutura Social
163 Direito do Trabalho			353 Processo de Interação Social
170 Contexto mundial			354 Sociologia da Educação
171 Estrutura Política Mundial			**360 Comunicação e educação**
172 Relações Internacionais			361 Comunicação
173 Problemática Mundial			362 Linguagem e Educação
174 Política Mundial			

370 Saúde e educação	**450 Curso e Curriculo**	**920 Identificadores de conteúdo**
371 Biologia Educacional	451 Currículo	921 Áreas do Conhecimento
372 Excepcionalidade	452 Curso	922 Cursos
373 Medicina e Educação	453 Plano de Ensino	923 Profissões e Ocupações
374 Psicologia Clínica	454 Conteúdos Curriculares	924 Atividades Planejadas
375 Educação Física	455 Atividades Curriculares	925 Manifestações Sócio-Culturais
400 EDUCAÇÃO	456 Avaliação do Currículo	**930 Especificadores de forma de conteúdo**
410 História da educação	**460 Processo de ensino-aprendizagem**	931 Documentos Não Oficiais
411 História Geral da Educação	461 Aprendizagem	932 Documentos Oficiais
412 História da Educação Brasileira	462 Orientação Educacional	**940 Especificadores de suporte de informação**
420 Filosofia da educação	463 Ensino	941 Língua do Documento
421 Princípios da Educação	464 Orientação Pedagógica	942 Publicações
422 Crítica da Educação	465 Organização do Trabalho Intelectual	943 Publicações em Multimeios
423 Pedagogia e Educação	471 Produtividade escolar	
424 Finalidades da Educação	472 Avaliação Escolar	
430 Educação escolar	**480 Meios de ensino**	
431 Escolarização	481 Tecnologia Educacional	
433 Educação Profissional	482 Equipamentos Didáticos	
434 Educação Superior	483 Material Didático	
435 Formação dos Profissionais da Educação	**900 IDENTIFICADORES E ESPECIFICADORES DE INFORMAÇÃO**	
440 Modalidades de educação	**910 Identificadores de denominação**	
441 Educação Formal	911 Personalidades	
442 Educação Não-Formal	912 Entidades	
443 Educação Informal	913 Lugares	
444 Educação Especial	914 Indicadores de Tempo	
445 Educação de Jovens e Adultos	915 Grupos Humanos	
446 Educação a Distância		
447 Modalidades de Educação Profissional		
448 Educação de Grupos Atípicos		

Quadro 05 – Hierarquia das classes em cadeias - Tesauro Brasileiro da Educação (Thesaurus Brased)
Fonte: INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA

Foram eleitos, dentro da subárea do "Contexto da educação", 07 contextos (de 110 a 174): ambiental, humano, social, cultural, político, econômico e mundial. Ao selecionar este último, encontramos "Estrutura Política Mundial, Relações Internacionais, Problemática Mundial e Política Mundial". Uma clara rede de pertinência, que empregou o item 900 do Quadro 05, como base para a formulação da hierarquia, fica evidente.

Poderíamos afirmar que, inicialmente, a relação predominante é a de hiperonímia/hiponímia, pois são notadas as classes principais como amplas e genéricas. No entanto, continuando no exemplo do parágrafo anterior, poucos itens contidos nessa hierarquia possuem atributos que os reorganizam dentro de um mesmo universo, numa relação de hiponímia. Dessa forma poderíamos pensar que as relações entre os referidos "Identificadores e Especificadores de Informação" são de um outro tipo que seria a meronímia/holonímia, tratada como relação partitiva, conforme já apresentado no Capítulo 02. Há que se ressaltar que, numa relação partitiva, estão associados o todo e suas partes; ou também pode ocorrer um relacionamento funcional. Dessa forma, podemos entender que "Estrutura Política Mundial, Relações Internacionais, Problemática Mundial e Política Mundial" são partes de um todo que é o contexto mundial, não necessariamente estabelecendo uma relação de hierarquia. O mesmo pode ser observado no holônimo "Contexto humano" (120) que contém os merônimos "Pessoa, Família e Comunidade" (121,122,123) como subclasses. Além disso, seria importante considerar que essas relações não foram feitas em termos de hiperonímia/hiponímia em razão da natureza dos elementos (sintagmas criados por proximidade semântica ou por relações cognatas). Essas relações são legítimas quando construídas a partir de termos simples ou de sintagmas fixos na língua.

Podemos inferir, portanto, que os conceitos de um domínio naturalmente se relacionam entre si como um sistema; e uma linguagem documentária como o tesauro pode evidenciar tal relação. Porém, o que limita a análise desse tesauro é a pouca quantidade de informação para cada termo, de modo a possibilitar o necessário aprofundamento nessa discussão. Há poucas *Notas de Aplicação,* que são os pequenos textos que contêm a explicação do conteúdo do conceito, pois através da definição de um conceito é possível esclarecer os relacionamentos propostos entre os conceitos que fazem parte de um domínio. No entanto, verificamos que diferentes tipos de relacionamento são utilizados na hierarquização de conceitos em um tesauro.

O Tesauro Brasileiro da Educação (Thesaurus Brased) foi criado para facilitar a pesquisa em educação, como uma lista de consulta para profissionais indexadores e atende satisfatoriamente a seus propósitos. Acreditamos que o uso de conceitos dos termos

específicos da área, permite, por meio das relações entre eles, a melhor compreensão do contexto global e interdisciplinar da educação, possibilitando pesquisas em maior profundidade.

5. 2. 3 Análise da hierarquia semântica do Tesauro em Ciência da Informação

Em busca de um outro formato de estruturação de tesauros, escolhemos o Tesauro em Ciência da Informação (TCI). Trata-se de um tesauro constituído a partir de outros tesauros, o que poderia evidenciar, em tese, tipos diferenciados de relações entre os conceitos.

Para nortear a constituição do TCI, foram delimitadas áreas de relacionamento tais como: Administração, Arquivologia, Museologia, Ciências Cognitivas, Ciências da Computação, Economia, Educação, Linguística e Sociologia, conforme Figura 05. Notamos que, ao definir essas relações, possibilita-se um grande número de relações no domínio da CI.

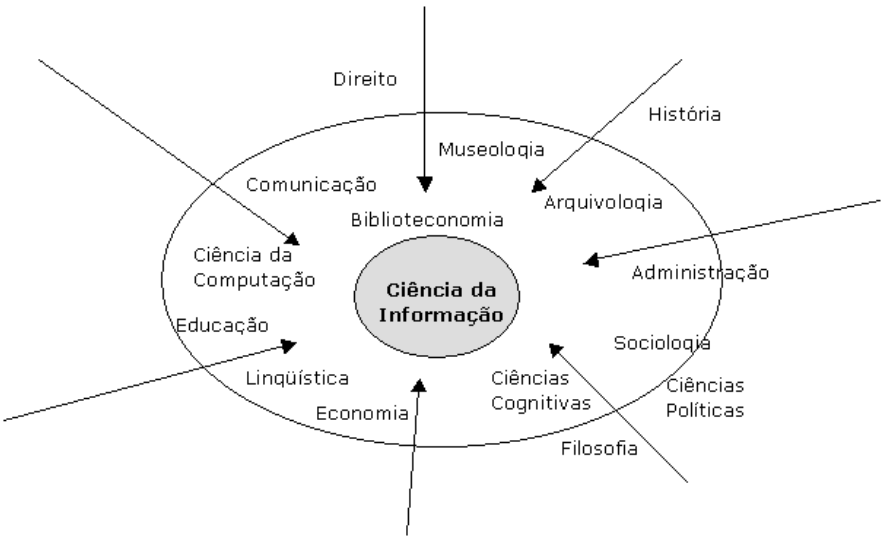

Figura 05 – Áreas de relacionamento –Tesauro em Ciência da Informação - TCI
Fonte: MOREIRA, 2005.

Acreditamos, porém, que a particularidade desse tesauro, construído a partir de outros tesauros, conforme acima indicado, pode relativizar a padronização dos critérios hierarquizadores. Segundo informa o histórico de construção disponível no *site*, "os termos foram separados de suas hierarquias originais e estas foram mantidas a parte para a redistribuição dos mesmos."(MOREIRA, 2005). Foram escolhidas as facetas dos tesauros de origem, gerando as seguintes: Arquivologia, Campos e disciplinas, Ciência da Informação,

Conhecimento e Informação, Documentação, Fontes de Informação, Informação e operações em bibliotecas, Mídia física e de comunicação, Museologia, Organizações, Pesquisa e métodos analíticos, Pessoas, profissionais, grupos informais, Tecnologia da informação, Unidades de informação. Cabe esclarecer que não é possível uma visão global desse tesauro, pois isso pode variar de acordo com o leiaute escolhido pelo elaborador, além do universo de palavras indexadas, cujos bancos de dados podem variar de tamanho. O universo do TCI abarca 1694 termos. Dessa forma foi escolhida, para análise, a faceta Arquivologia, conforme Quadro 06. Podemos observar que a relação de subordinação é demarcada pelo número de pontinhos azuis; assim, a classe geral está delimitada com um ponto; a subclasse imediata, com dois pontos; a terceira subclasse com três, a quarta, com quatro, até a quinta, com cinco pontos.

Tesauro em Ciência da Informação • Arquivologia •• Automação de arquivo •• Ciclo vital ••• Arquivo corrente ••• Arquivo intermediário ••• Arquivo permanente •• Documento arquivístico ••• Acesso ao documento arquivístico ••• Agrupamento documental •••• Documento simples •••• Fundo de arquivo •••• Grupos de fundo •••• Série documental ••• Deterioração de documentos ••• Tipo de documento •••• Ata •••• Ato •••• Carta •••• Censo •••• Livro contábil •••• Livro de registro •••• Memorando •••• Protocolo •• Gestão de documentos •• História de arquivo •• Instrumento de consulta ••• Instrumento de controle •••• Registro de controle	••• Instrumento de descrição •••• Catálogo de arquivo •••• Censo de arquivo •••• Guia de arquivo ••••• Censo-guia ••••• Guia de fonte •••• Índice (arquivo) •••• Inventário (arquivo) •• Patrimônio documental •• Política arquivística ••• Cooperação (arquivística) ••• Gestão de arquivo ••• Legislação arquivística ••• Planejamento de arquivo ••• Rede (de arquivo) ••• Regulamento de arquivo ••• Sistema arquivístico •••• Sistema nacional de arquivo ••• Terceirização (de serviço arquivístico) •• Serviço arquivístico ••• Microfilmagem ••• Serviço de extensão arquivística •••• Atividade cultural arquivística ••• Serviço de referencia arquivística •• Teoria arquivística ••• Princípio de procedência ••• Princípio de respeito à ordenação prévia •• Tipos de arquivos ••• Arquivo administrativo •••• Arquivo central •••• Arquivo de gestão ••• Arquivo audiovisual ••• Arquivo da nobreza	••• Arquivo de mídia de comunicação ••• Arquivo familiar ••• Arquivo fotográfico ••• Arquivo pessoal ••• Arquivo vertical •• Tratamento de arquivo ••• Armazenamento de documento ••• Custódia de arquivo ••• Descrição arquivística •••• Normas de descrição arquivística ••••• ISAD ••• Entrada de documento •••• Traslado de fundo (arquivo) ••• Identificação documental ••• Organização de arquivo •••• Classificação de fundo ••••• Quadro de classificação •••• Ordenação de documento ••• Princípio de procedência* ••• Valoração documental •••• Descarte (Arquivologia) ••••• Tabela de temporalidade •••• Seleção de documento arquivístico •••• Tabela de avaliação • Campos e disciplinas

Quadro 06 – Hierarquia da classe de Arquivologia - Tesauro em Ciência da Informação - TCI
Fonte: MOREIRA, 2005.

Há nele informações de que, também, foram gerados relacionamentos de equivalência e relações associativas. Os relacionamentos de equivalência correspondem à relação ‘Use’ ou ‘Usado para’. As orientações a seguir constam na estrutura da ferramenta e, conforme já ressaltado brevemente na seção 5.1, deste Capítulo, visam a dar mais clareza ao modo como foi organizado, para que seja possível estabelecer as relações:

> SN – Scope Note (Notas de escopo) Equivalem a uma nota esclarecedora do termo.
> UF (USADO PARA) – Relacionamento utilizado quando um descritor é recomendado no lugar de termos não descritores. Corresponde à inversa da relação USE. Exemplo: AACR – Usado para Anglo-American Cataloging Rules.
> USE – Relacionamento que associa um não descritor a um descritor. Essa associação pode ser de sinonímia ou de equivalência do termo dentro do contexto do tesauro. É uma relação inversa à relação UF. Exemplo: Anglo-American Cataloging Rules USE AACR.
> BT - Broader term (termo genérico) – Pode ser traduzido como termo descritor superior ao termo em questão na hierarquia do tesauro. Observe que termos descritores na condição de BT são também chamados na literatura de termos de topo. Existem tesauros que aceitam relacionamento poli-hierárquico. O TCI é mono-hierárquico, ou seja, cada termo possui apenas um termo a ele superior. Quando o mesmo termo pertence a mais de uma hierarquia, no TCI utilizou-se do artifício do ‘*’ ao final do termo para diferenciar um termo do outro.
> NT – Narrower Term (termo específico) – Como o próprio nome indica, é um específico e está sempre associado a um BT. Um NT pode ter outros NTs abaixo dele, para os quais torna-se um BT. (Grifo nosso) (MOREIRA, 2005).

Partindo, então, para a análise, ao acessar a faceta “Arquivologia”, as informações sobre o termo selecionado não apresentam conceitos e, sim, termos associados como Automação de arquivo, Ciclo vital, Documento arquivístico, Gestão de documentos, História de arquivo, Instrumento de consulta, Patrimônio documental, Política arquivística, Serviço arquivístico, Teoria arquivística, Tratamento de arquivo, Tipos de arquivos. Como subclasse do item “Tratamento de arquivo”, aparece a seguinte hierarquia:

••Tratamento de arquivo

••• Armazenamento de documento

••• Custódia de arquivo

••• Descrição arquivística

•••• Normas de descrição arquivística

••••• ISAD

No acesso ao termo genérico *Tratamento de arquivo* (BT), surgem as opções de termos específicos (NT): Custódia de arquivo, Descrição arquivística, Identificação documental, Entrada de documento, Armazenamento de documento, Organização de arquivo, Princípio de procedência*, Valoração documental. O asterisco marca a presença do referido termo em outra hierarquia. Já a subclasse “Descrição arquivística”, contém “Normas de

descrição arquivística", que contém "ISAD", International Standard of Archival Description, uma norma internacional. Podemos afirmar que esse tesauro tem um perfil mais sistematizado, de caráter mais eletrônico; em outras palavras, a seleção e hierarquização dos termos parecem baseadas no uso das tecnologias de informação. Os objetos de natureza linguística foram manipulados por técnicas de indexação e algoritmos sustentados por regras linguísticas primárias (com poucos traços descritivos) nas linguagens de programação. No entanto, nos propicia pensar na relação de hiperonímia/hiponímia e/ou holoníma/meronímia, no pertencimento do termo aos conjuntos. O termo genérico "Arquivologia" possui dois termos relacionados (RT): Arquivologia, como área de conhecimento, e Arquivista. Ao acessá-los, notamos que há um novo arranjo na relação hierárquica em função da nova relação de pertencimento, conforme podemos observar:

Termo Arquivologia (área de conhecimento)
BT Ciências Sociais Aplicadas
NT Organização de Arquivos
RT Arquivologia
Faceta[31] CD - Campos e disciplinas

Termo Arquivista
BT Profissional de informação
RT Arquivologia
Faceta PP - Pessoas, profissionais e grupos formais
(MOREIRA, 2005)

Os termos, ao serem inseridos em outras facetas, são remetidos a outras relações associativas: *Arquivologia* passa de termo geral (BT) "Tesauro em Ciência da Informação" e, como NT (específico) "Automação de arquivos" e os demais já citados, para o BT "Ciências Sociais Aplicadas", uma vez que está associada à faceta "Campos e disciplinas", com NT Organização de Arquivos. Dessa forma temos a seguinte estrutura:

Termo Arquivologia
BT Tesauro em Ciência da Informação
NT Automação de arquivo
Ciclo Vital
Documento arquivístico
Gestão de documentos
História de arquivo
Instrumento de consulta

[31] A classificação facetada contém ambas as categorias fundamentais – assuntos básicos e suas facetas –, e estas, por sua vez, contêm termos "isolados", que são os diversos assuntos, em âmbito mais específico, dentro de qualquer faceta. Um assunto básico pode estar sozinho, por exemplo, "literatura", dentro de uma faceta no assunto "literatura inglesa".

Patrimônio documental
Política arquivística
Serviço arquivístico
Teoria arquivística
Tratamento de arquivo
Tipos de arquivos
(MOREIRA, 2005)

De acordo com o Quadro 06, verificamos que, em algumas das subclasses acima relacionadas, há outras subdivisões, como ocorre em "Ciclo Vital", em que se incluem "Arquivo corrente, Arquivo intermediário, e Arquivo permanente"; também em "Política Arquivística", incluem-se "Cooperação (arquivística), Gestão de arquivos, Legislação arquivística, Planejamento de arquivo, Rede (de arquivo), Regulamento de arquivo, Sistema arquivístico (no qual se inclui Sistema nacional de arquivo) e Terceirização (de serviço arquivístico)". Podemos notar que predomina a relação holônimo/merônimo nas classes mais gerais; já nas classes mais específicas, constatamos a relação hiperônimo/hipônimo mais definida, conforme ocorreu com "Arquivo administrativo" que comporta "Arquivo central" (classe mais geral) antes de "Arquivo de gestão" (classe mais específica).

Parece-nos que os atributos de inclusão para a relação holonímia/meronímia são mais especificados, fato que se observa no exemplo já indicado, no qual os atributos para inclusão de termos em "Política Arquivística" poderiam ser, por exemplo, o compartilhamento e a regulação. No entanto, esses atributos não se aplicam aos elementos hierarquizados em "Ciclo Vital"; podemos teorizar que o atributo seja a sistematização.

Numa análise final, consideramos que o campo semântico está claramente definido, pois a própria definição de áreas e facetas, de certa forma, já permite um parâmetro classificatório. Já nas questões referentes às relações entre os conceitos também ficam restritas no TCI, pois nem todos os termos indexados possuem Notas de Aplicação; dessa forma fica inviabilizada a busca da possível configuração de um protótipo.

5. 2. 4 Análise da hierarquia semântica do Tesauro de Folclore e Cultura Popular Brasileira

Em busca de um instrumento que permitisse uma análise mais detalhada e consistente sobre a hierarquia semântica de conceitos, encontramos a segunda versão do Tesauro de Folclore e Cultura Popular Brasileira, publicado pelo Centro Nacional de Folclore e Cultura Popular do Instituto do Patrimônio Histórico e Artístico Nacional, do Ministério da Cultura. Esse tesauro despertou nossa atenção porque, além dos termos classificados - como em todos os tesauros -, foram associados outros recursos multimídias que permitem ao usuário complementar a consulta com trecho de vídeo ou som, fotografia ou ainda ler artigo publicado sobre o assunto pesquisado. Possui, também, uma apresentação colorida e convidativa, em forma de gráfico, que permite uma visão geral do universo de palavras associadas e propicia uma análise que nos parece melhor atender aos nossos propósitos.

A equipe multidisciplinar que produziu esse tesauro, além dos profissionais da área de Ciência da Informação, compõe-se de antropólogos, sociólogos, dentre outros profissionais que validaram as áreas temáticas. Gostaríamos de ressaltar que, embora haja menção aos revisores de textos, não há referência alguma a profissionais de Letras para validação, por exemplo, dos diversos gêneros textuais classificados, formas de expressão, que caracterizam toda manifestação folclórica.

O Tesauro de Folclore e Cultura Popular Brasileira possui as seguintes categorias, conforme Figura 06: Alimento, Artefato, Associação, Atividade produtiva, Atividade ritual, Prática religiosa e Farmacopéia popular, Indivíduo, Matéria-prima, Literatura oral, Medicamento, Construção artesanal, Sistema de crença, Lugar e Tempo.

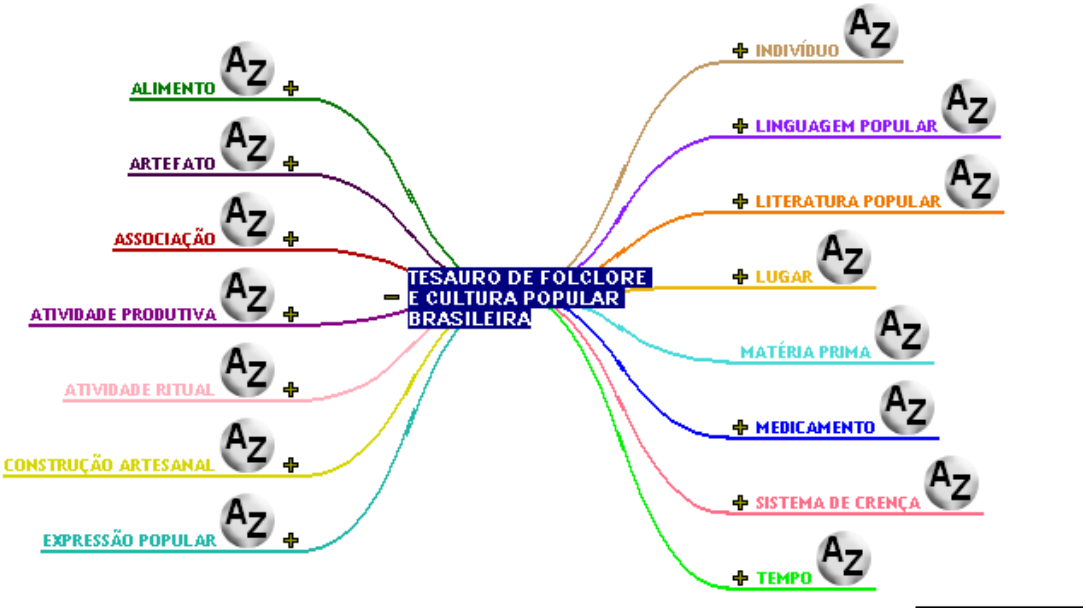

Figura 06 – Visão geral do Tesauro de Folclore e Cultura Popular Brasileira (MC)
Fonte: COELHO, 2009.

Na introdução, parte importante num tesauro, pois é onde são expostas as redes de relações propostas naquele instrumento, está estabelecido que o domínio específico de conhecimento é o folclore e a cultura popular brasileiros, áreas de caráter interdisciplinar. Na sequência, já, de início, está exposto o conceito de folclore como: "os modos de agir, pensar e sentir de um povo, ou seja, como expressões da cultura desse povo, e, em consonância com o que preconiza a Unesco, considera equivalentes os termos folclore e cultura popular." (Grifo nosso). Essa noção orientadora pode explicitar os critérios de pertinência dos termos ali inseridos; talvez, isso permita a rede de relações que se estabelece e que diferencia um tesauro de um dicionário. Inclusive, indicam a base utilizada: os acervos da Biblioteca Amadeu Amaral e do Museu de Folclore Edison Carneiro, usando termos já consagrados dentro dessa temática. Aqui se entende que podemos configurar a delimitação de um campo semântico, porém, não somente uma lista de palavras, dada a característica relacional de um tesauro.

Outra informação importante são as relações estabelecidas. Há um esclarecimento prévio de que os termos podem ser constituídos de palavras ou grupos de palavras. Um dos critérios para diferenciar a forma e o significado de termos que possuem a mesma grafia, mas significados diferentes, foi o de acrescentar um qualificador para delimitar o escopo semântico, representado entre parênteses, como no caso de balaio (artefato) e balaio (dança). Já a expressão "balaio de gato", por exemplo, não seria um complicador num tesauro, uma vez que o campo semântico é delimitador, e, se fosse o caso poderia ser inserida numa

possível faceta de gírias ou expressões populares. Daí são caracterizadas as relações entre os termos que evidenciam os conteúdos semânticos:

> **Relação de equivalência** — entre termos sinônimos ou quase sinônimos; expressa pela abreviatura USE/UP;
> **Relação hierárquica** — superordenação e subordinação de dois termos; expressa pela abreviatura TG/TE; (termo geral e termo específico)
> **Relação partitiva** — entre termos que são, um, o todo e, outro, a parte; expressa pela abreviatura TGP/TEP;
> **Relação associativa** — entre termos com proximidade de significado (causa/efeito, produto/processo, material/produto); expressa pela abreviatura TA/TA.
> (COELHO, 2009).

Há, também, esclarecimentos sobre os dois tipos de relacionamento empregados. Os termos são apresentados em ordem *sistemática e classificada* - perfil de maior interesse para esta tese -, na qual "os termos estão classificados de forma hierárquica por meio de categorias e classes formando um mapa conceitual do domínio de folclore e cultura popular brasileira"; e em ordem *alfabética*, na qual encontramos cada termo relacionado, independente de sua classificação hierárquica, mas dispostos em ordem alfabética com seus respectivos conceitos e demais informações associadas. As notas explicativas que acompanham os termos são de duas naturezas:

> **Notas de aplicação** — apresentam a definição mais apropriada para o termo em questão; expressas pela abreviatura NA;
> **Notas de indexação** — apresentam questões relacionadas ao uso daquele termo no contexto institucional; expressa pela abreviatura NI. (COELHO, 2009)

Uma das categorias selecionadas foi a de *Literatura popular*, com o objetivo de analisar a hierarquia semântica implementada no renque que possui diversos subitens, conforme Figura 07:

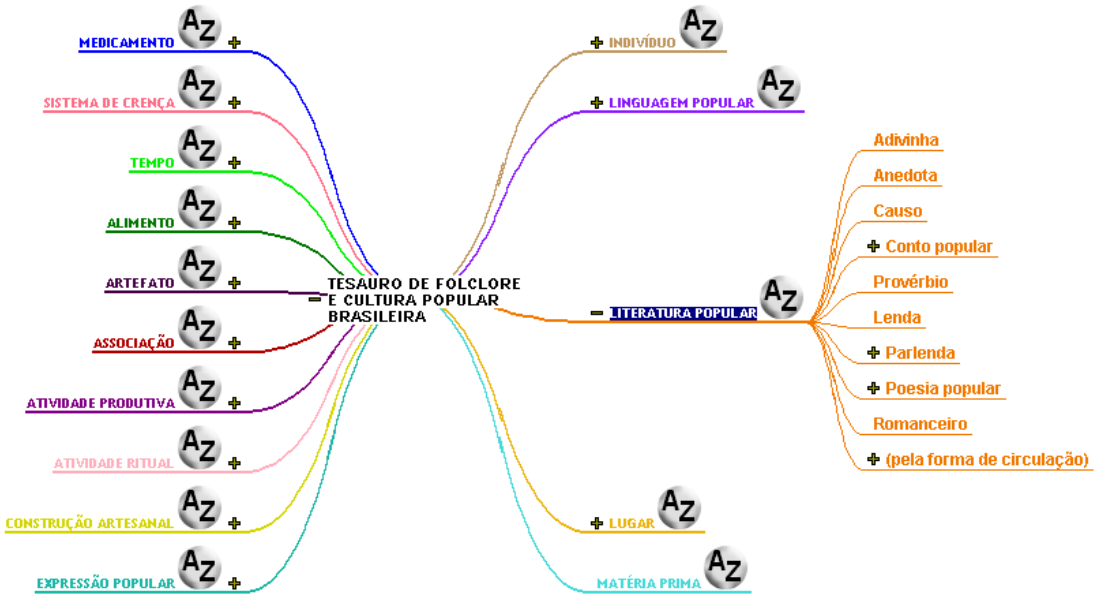

Figura 07 – Subitem expandido: Literatura Popular
Tesauro de Folclore e Cultura Popular Brasileira (MC)
Fonte: COELHO, 2009.

Podemos observar que, ao solicitarmos a abertura do renque, clicando no sinal (+), são expostas 10 (dez) opções subordinadas ao termo Literatura popular, consideradas, por nós, como *gêneros textuais*: Adivinha, Anedota, Causo, Conto Popular, Provérbio, Lenda, Parlenda, Poesia popular, Romanceiro, e uma alternativa de abertura apresentada como "(pela forma de circulação)". Abrindo essa opção, em específico, temos as opções de Literatura de cordel e Pasquim, conforme mostra a Figura 08.

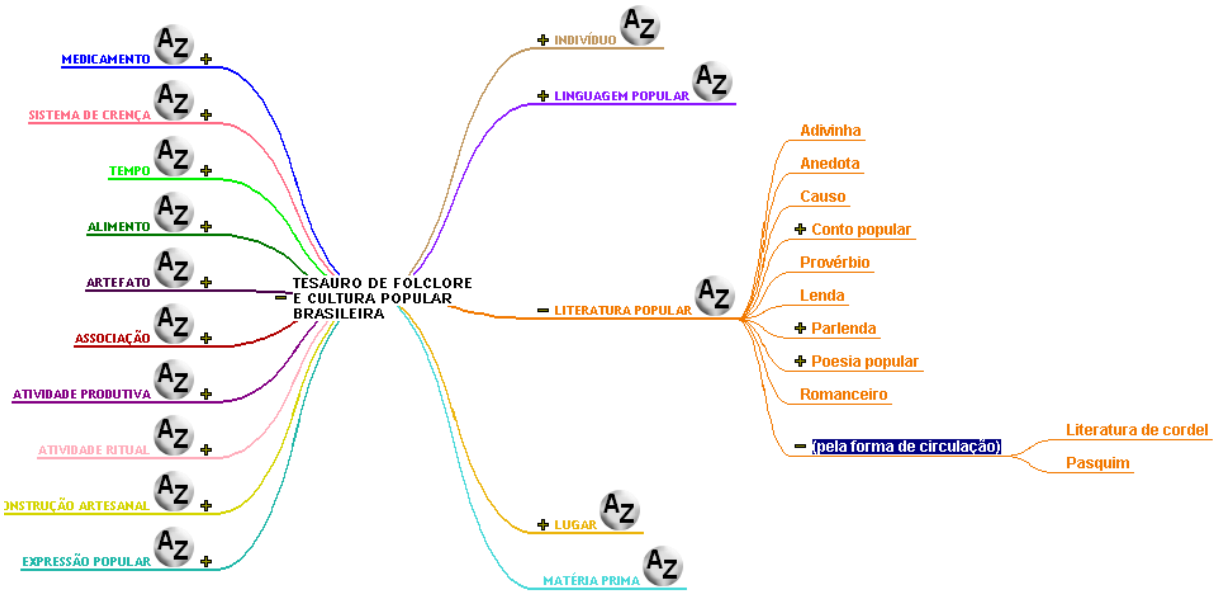

Figura 08 - Subitem expandido: (pela forma de circulação)
Tesauro de Folclore e Cultura Popular Brasileira – MC.
Fonte: COELHO, 2009

Essa opção nos faz questionar sobre os princípios que nortearam a classificação dos gêneros textuais já apresentados, uma vez que estão subordinados a um atributo de "forma de circulação". Podemos levantar a hipótese de que o objetivo foi associá-los pelo modo de construção em versos, que seria um *atributo de inclusão* - que implica também o modo de circulação-; porém, há outros gêneros já exibidos na classe anterior que também são compostos em versos, como a poesia popular e a parlenda, o que invalidaria essa subclasse. Outra hipótese mais pertinente seria a tentativa de *aproximar* Pasquim e Literatura de cordel pelo modo de difusão, pois a exposição em versos, rimada ou não, geralmente é cantada; ou então *diferenciá-los,* porque singulariza a Literatura de cordel que é veiculada em forma de folhetos vendidos nas feiras, originalmente pendurados em barbantes ou cordéis. Então, podemos afirmar que a relação hiperônimo/hipônimo estaria comprometida, mas a de merônimo/holônimo, não. Porém, se levarmos em consideração o tom crítico de um Pasquim, ele possibilitaria outros relacionamentos.

Esse exemplo também nos faz pensar nas categorias propostas por Dalberg (1978), a partir dos modelos de organização propostos por Aristóteles e depois revistos por Ranganathan, pois acreditamos na hipótese de que foram considerados outros aspectos como a forma e a finalidade.

É importante demonstrar que na apresentação alfabética, os referidos gêneros estão listados no mesmo ambiente, conforme vemos na Figura 09:

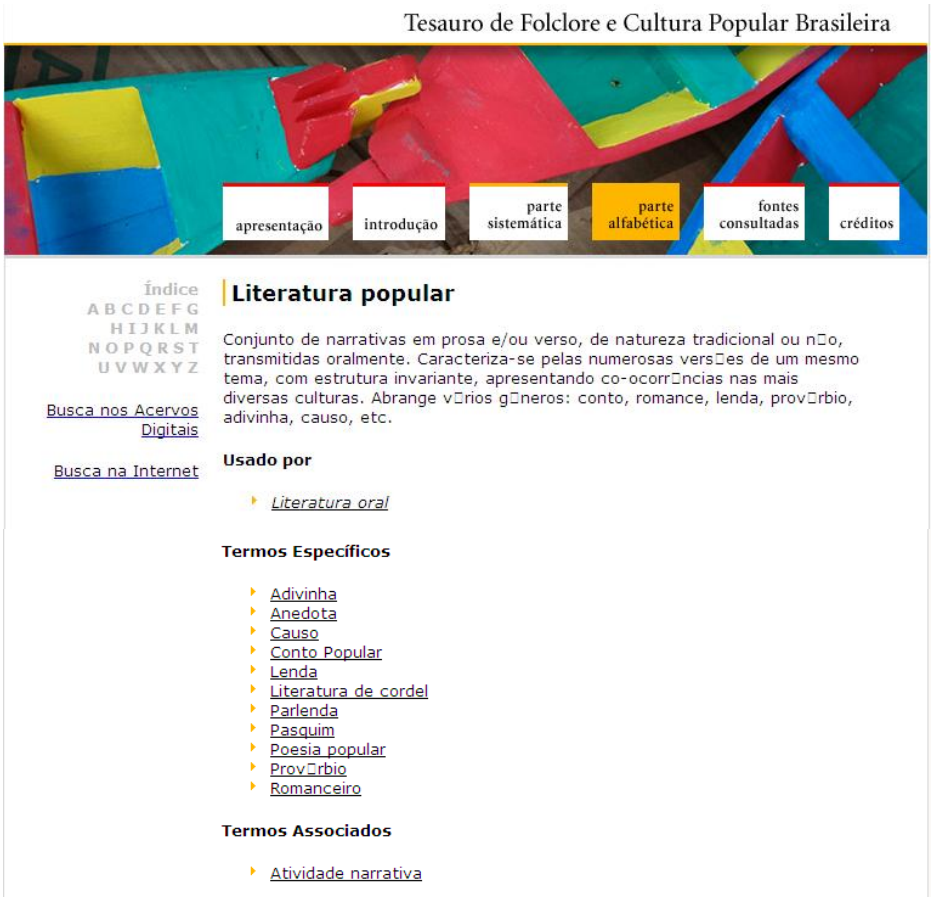

Figura 09 – Orientações sobre o termo Literatura Popular
Tesauro de Folclore e Cultura Popular Brasileira – MC.
Fonte: COELHO, 2009

Conforme mostra a Figura 09, entende-se como Literatura popular[32], que é o termo geral analisado, o

> Conjunto de narrativas em prosa e/ou verso, de natureza tradicional ou não, transmitidas oralmente. Caracteriza-se pelas numerosas versões de um mesmo tema, com estrutura invariante, apresentando co-ocorrências nas mais diversas culturas. Abrange vários gêneros: conto, romance, lenda, provérbio, adivinha, causo, etc. (COELHO, 2009).

Ressalta-se que essa acepção também é tomada por outro termo geral relacionado, Literatura oral, que remete ao mesmo conceito de Literatura popular. A seguir, temos os Termos Específicos: Adivinha, Anedota, Causo, Conto Popular, Lenda, Literatura de cordel, Parlenda, Pasquim, Poesia popular, Provérbio, Romanceiro. Além de haver um Termo Associado: Atividade narrativa, mas o renque continua subdividindo-se em outras possibilidades, conforme Figura 10:

[32] Todas as acepções foram retiradas da parte alfabética do próprio tesauro em análise.

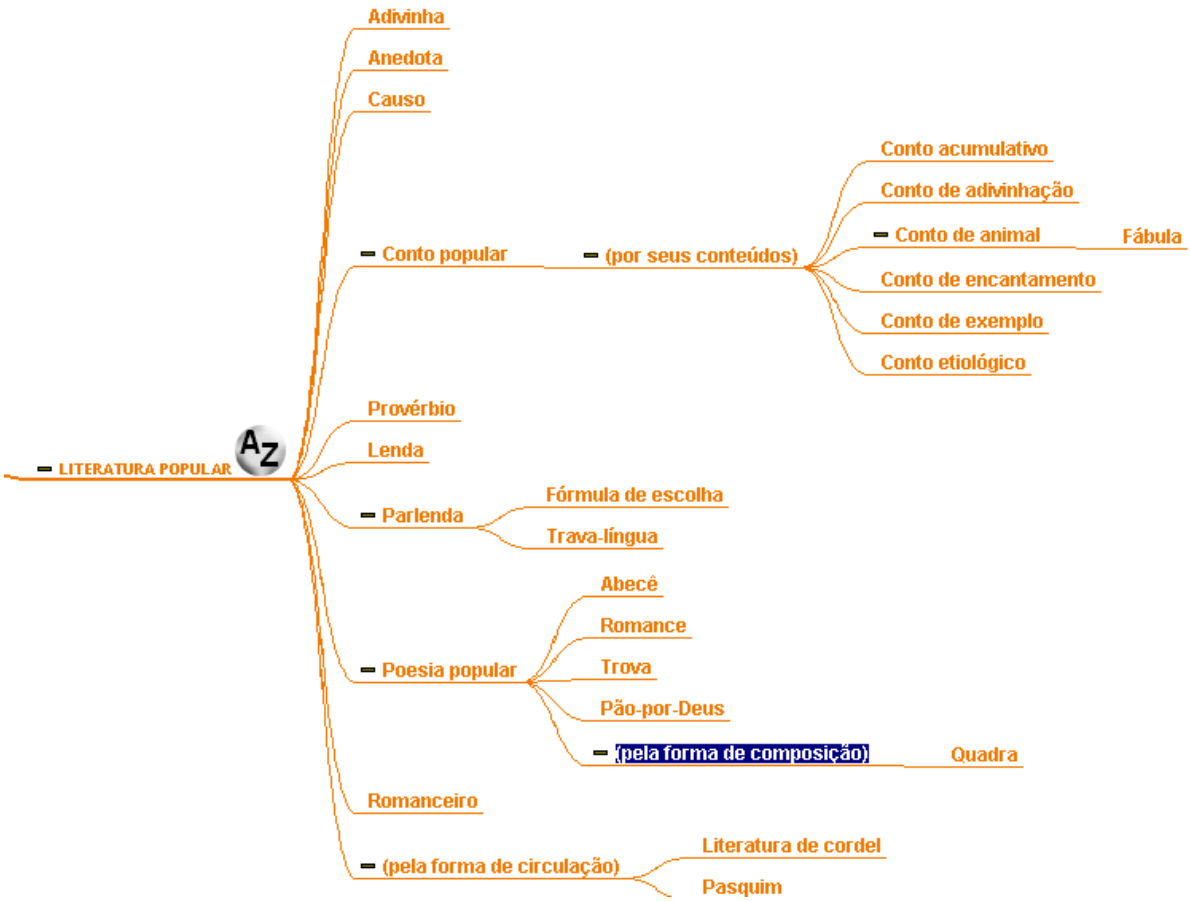

Figura 10 – Renque expandido de Literatura Popular
Tesauro de Folclore e Cultura Popular Brasileira - MC
Fonte: COELHO, 2009

Numa análise preliminar, poderíamos afirmar que a relação hierárquica estabelecida toma como referência a relação generalização e especificação entre elementos de mesma natureza, dado o universo conceitual preestabelecido, acima explicitado. Retomando os pressupostos teóricos aqui propostos, a relação de pertinência se faz pela aproximação de características próprias de cada termo. Um bom exemplo seria pensar a relação do renque de poesia popular, inicialmente descrita como "Literatura popular em versos livres e/ou providos de rimas que, em geral, descrevem cenas e episódios do cotidiano, cantados e difundidos pela tradição oral". Em sua relação hierárquica, estão subordinados alguns formatos poéticos muitos deles específicos da cultura popular, como "Pão-por-Deus" e o "Abecê", associados a tipos mais cristalizados como "Romance" que, talvez, nem se encaixasse nessa estrutura; porém, para justificar a relação de pertinência, seria necessário levantar a origem da nomenclatura numa perspectiva mais geral relacionada à prosa, de caráter narrativo, e relativizá-la numa perspectiva regional, mais nordestina, de composição poética do

romanceiro popular, ou mesmo um poema em versos para ser cantado. Talvez, por tentar atender a essa variação, é que consta na acepção do termo Romance: "Poesia popular na forma de narrativas lírico-dramáticas. São fragmentos das longas epopéias medievais conservadas na memória popular, transmitidos pela oralidade, por meio de manuscritos ou editados na forma de folhetos de cordel.", uma vez que o termo geral, por um lado, toma-o como Literatura popular e, por outro, o assegura como Literatura Oral, conforme já comentado.

A questão é que tal termo é relacionado com outro, Atividade Narrativa, termo associado à Atividade Ritual. Conforme se observa, a Atividade Narrativa é descrita no tesauro em análise, como

> Atividade ritual que compreende as tradições transmitidas pela memória, o imaginário, as poéticas orais e edições populares, contemplando o texto cultural como um contínuo, sem o limitar aos registros estéticos. A atividade narrativa significa o procedimento verbal, mas não implica a inexistência de escrito, que venha fixar o que é feito pela palavra falada. (COELHO, 2009).

Além disso, o termo Romanceiro aparece como categoria equivalente à poesia popular na apresentação gráfica, com a seguinte descrição: "Literatura popular que congrega coletânea de romances". Independente da análise de aspectos característicos para delinear a relação de pertinência à família, conforme proposto por Rosch, nota-se que os critérios orientadores são de diferentes ordens.

Ora os "romances" pertencem ao conjunto "Poesia popular", ora ao "Romanceiro", o que torna improdutiva essa categoria. No entanto, ao relacioná-la à Atividade Ritual, ganha contorno de Literatura Oral, termo genérico. É preciso pensar que relação foi empreendida, entre as já apresentadas neste capítulo; se for a de equivalência não caberia, nem tampouco a de hierarquia; porém se pensarmos na idéia partitiva, todo e parte ou na de associação, tal relação se justificaria. No entanto, isso deixa muito abertas as possibilidades de organização, que podem gerar outros equívocos como no caso da opção "pela forma de composição", na qual se situa a "Quadra". Pergunta-se, então, qual foi o critério utilizado nas designações anteriores, como o "Abecê?", por exemplo, conforme se verifica em sua descrição como "Poesia popular que utiliza o alfabeto na sua sequência natural. Suas estrofes se iniciam com a primeira letra do alfabeto, seguindo até a última". Essa "forma de composição", na qual se inclui a Quadra como "Poesia popular composta de estrofe de quatro versos setissílabos, com rima toante ou assonatada no segundo e quarto versos.", em que diferiria da orientação da forma como começam as estrofes no Abecê? E o caso de "Pão-por-Deus", descrito como

“Poesia popular contendo mensagem em verso, quadra, trova, escrita em papel multicor, recortado geralmente em forma de coração, com declaração de amor ou qualquer solicitação.”, que inclui a própria Quadra? E a Trova, “Poesia popular ligeira que utiliza a quadra em sua composição poética em quatro versos heptassílabos, com rima consoante ou assonante, entre o primeiro e o terceiro versos e entre o segundo e o quarto.”, não estaria exposta uma forma de composição? Além disso, podemos afirmar que algumas das acepções parecem contemplar uma perspectiva de representação escrita, o que relativizaria o termo genérico Literatura Oral.

Outro exemplo nesse mesmo tesauro é o renque dos Contos Populares. Esse conjunto possui como critério de ordenação, “os conteúdos”; os elementos pertinentes a esse conjunto são os Contos Acumulativos, de Adivinhação, de Animal, de Encantamento, de Exemplo e Etiológico. Retomamos aqui a problemática já assinalada sobre o termo genérico relacionado como equivalente, Literatura oral. “Contos Populares” é aqui tomado como *Literatura popular constituída de narrativa em prosa, de acontecimentos considerados fictícios, narrados com finalidade lúdica.* O primeiro elemento desse conjunto são os “Contos Acumulativos” cuja descrição não sinaliza aspecto algum relacionado ao conteúdo a saber: *Conto popular construído por palavras ou frases encadeadas, ações ou gestos, que se articulam em seriação ininterrupta.* Aos demais tipos de contos, correspondem as descrições que remetem ao conteúdo, como os de Adivinhação relacionados à *solução de enigmas, charadas, tradução de gestos e/ou decifração da origem de certos objetos*; os de Animal em que os *papéis principais, heroi e adversário, são representados por animais* - neste se inclui a fábula como *conto de animal constituído de narrativas revestidas de cunho moral cujos personagens são animais*; de Encantamento, caracterizado pelo *elemento sobrenatural ou fantástico, em que intervêm seres fabulosos, animais antropomórficos, objetos mágicos, etc*.; de Exemplo, em que os *fatos representam os valores de conduta moral de determinados grupos culturais, com intenção doutrinária*; e o Etiológico que *explica e justifica a razão de ser de um aspecto, forma, propriedade, caráter e disposição de um ser animal, vegetal ou mineral.* Ainda assim, uma análise mais atenta invalidaria ou, talvez, reunisse em outra subcategoria a questão dos conteúdos relativos a animais.

Conforme sinaliza Campos (2004), a questão da representação do conhecimento por modelos conceituais tem o desafio de estabelecer as relações de domínio, e busca na interface com outras áreas uma solução. A autora chega a refletir sobre as relações de equivalência, constatando que são estabelecidas relações de sinonímias entre as diversas formas de denominação; ressalta, porém, que essa relação se estabelece no plano da língua e não tem

sido apresentada nos modelos que compõem as relações entre os objetos. Podemos afirmar, portanto, que as ferramentas empregadas na organização dos dados carecem de *regras linguísticas* adequadas que orientem na hierarquização dos termos.

Conforme tentamos demonstrar aqui, nota-se que há diferentes orientações para a hierarquização que nem sempre ficam claras para o usuário. Assim sendo, quem usa a ferramenta *não reconhece aquela linguagem* como sua e, dessa forma, o processo de facilitação da comunicação pode ficar comprometido, não atendendo à sua necessidade enunciativa. Por isso reforçamos a urgência de refletirmos sobre tais regras que orientam tanto a estruturação dos pontos de função de um instrumento como a organização do banco de dados, de modo a ampliar e aperfeiçoar a capacidade de atendimento às demandas do usuário.

Prossegue a autora, ponderando que

> Nas teorias e metodologias que nos propomos a analisar, verificamos que a área da ciência da informação, apesar de ter teorias bem fundamentadas sobre o conceito e relações conceituais, é fraca em modelos que auxiliem a elaboração de representações gráficas. Tanto na teoria da classificação facetada quanto na teoria do conceito, não é apresentado nenhum modelo para expressar graficamente as relações conceituais. Acreditamos que isso se deva ao fato de que essas teorias têm por objetivo a elaboração de linguagens documentárias que, apesar de possuírem uma parte sistemática, não têm os conceitos representados em forma gráfica, mas em forma de uma lista endentada de termos, com uma notação que, de certa forma, deixa evidentes os grupos de termos afins. (CAMPOS, 2004, p.31)

Indo mais além, ponderamos que essas operações de hierarquização são, em boa parte, executadas segundo representações precisas e explícitas, implementadas por meio de programas escritos em uma linguagem de programação; evidenciamos, pois, a necessidade de se reverem tais regras que embasam os programas de mineração[33] das palavras.

Como acreditamos aqui que o processo de hierarquização é anterior à máquina, podemos afirmar que o conjunto de procedimentos expostos neste trabalho auxiliou no conhecimento de grande parte dos processos de organização de informação empregados na hierarquização dos conceitos em um tesauro. Foi importante verificar a complexidade da tarefa dos profissionais da informação diante de todo compromisso conceitual que está envolvido numa tarefa de hierarquização.

Finalizando esse capítulo de análise, podemos concluir que os estudos linguísticos podem contribuir para orientar tal tarefa, especialmente no tocante à clareza das relações de

[33] Metodologia para predição e classificação em grandes massas de dados observacionais; tem como objetivo o processamento de informação textual, para extrair índices numéricos significativos a partir do texto, de modo a tornar essa informação acessível aos programas disponíveis nos sistemas de mineração de dados. Podem ser analisadas palavras, agrupamentos de palavras, ou mesmo documentos entre si através das suas similaridades ou de suas relações com outras variáveis de interesse num projeto de mineração de textos.

hiperonímia/hiponímia e de holonímia/hiponímia, pois tentamos demonstrar isso nas análises empreendidas. Perdura a crença na probabilidade de serem estabelecidos critérios de inclusão ou de exclusão nas classes, pois a hierarquização impertinente num instrumento como o tesauro pode comprometer o uso dos termos e induzir a relações equivocadas.

6 REFLEXÕES FINAIS

"Todo modelo reflete a complexidade do real".
Edgar Morin (1990)

O tesauro é um instrumento de organização da informação que representa a estrutura conceitual de uma área específica, através de uma rede de termos relacionados entre si, predominantemente por três tipos relações: as *hierárquicas* (gênero/espécie), que acreditamos que sejam embasadas nas relações de hiperonímia e de hiponímia; as de *associação* não-hierárquica (teoricamente infinitas), como a partitiva (um todo e suas partes), a funcional (entre categorias: coisa-processo; material-produto), que configuram, no nosso modo de ver, relações de holonímia e meronímia; e as *enumerativas*, que são relações de equivalência e/ou de oposição (antônimos/sinônimos), que determinam as seleções dos melhores termos que representam dado conceito.

Cabe reforçar, no entanto, que um estudo mais aprofundado dessas relações certamente configura-se como uma tarefa a ser realizada pela linguística de modo mais efetivo, pois elas carecem de maiores esclarecimentos, quando situadas no âmbito de um tesauro. O ritmo incessante de produção e disseminação de informação exige, hoje, mecanismos cada vez mais eficazes, com alto índice de precisão. Para atingir tal excelência, a dimensão linguística poderia contribuir para mais clareza dos princípios organizadores.

Há que se ressaltar que as relações lógico-semânticas são o objeto do tesauro, pois este tem como fonte aquelas palavras que fazem parte de um conjunto de referências compartilhadas por grupos de usuários específicos (o que configura um sistema de significação da área específica). Nesse contexto, as relações hierárquicas implicam a subordinação de conceitos e podem ser apresentadas sob forma de árvore (séries verticais de conceitos). Nesse tipo de relação, notamos que a caracterização dos atributos interfere na classificação e nos critérios de inclusão dos termos, estabelecendo o vínculo da classe ou categoria e seus membros ou espécies. As características comuns podem revelar a hierarquia, pois são aquelas que descrevem a essência e as partes de um referente, configurando as cadeias e renques. Mas podemos afirmar que a apresentação hierárquica mostra as relações lógicas (inclusão) como das associativas e partitivas (ontológicas). Notamos que os parâmetros para o estabelecimento dessas relações associativas são pragmáticos, o que não quer dizer que sejam eminentemente empíricos. Esse tipo de relação permite a análise da definição para verificar outros termos relacionados ao termo selecionado. Na relação entre o objeto considerado como um todo e suas partes, o conceito superordenado se

refere ao todo e os subordinados às partes; tem-se, então, que o conceito superordenado é o conceito abrangente e o conceito subordinado é o conceito partitivo em relação ao conceito abrangente. Essas relações foram observadas nos três tesauros analisados.

Já as relações de equivalência viabilizam a seleção de um único descritor para representar o conceito e visam, portanto, a possibilitar o funcionamento do vocabulário. Nela podemos ter a distinção entre termo preferido (como por exemplo, *morte*) e preteridos (como *falecimento*). Essa escolha implica a compatibilização entre linguagens, de modo a permitir o acesso ao sistema (compatibilização entre a linguagem dos documentos, linguagem dos usuários, linguagem da instituição e a linguagem documentária). Permite, ainda, o deslocamento genérico, pois um termo mais específico pode ser deslocado para um mais geral, funcionando como meio de economia do sistema. Porém, não se trata de uma equivalência no sentido estrito, mas funciona como tal. Esse tipo de relacionamento ficou mais evidenciado na análise do TCI. Há que se ressaltar que a política de indexação[34] de um sistema pode determinar que os antônimos, por exemplo, sejam considerados quase-sinônimos, razão pela qual são tratados como termos equivalentes, e não relacionados. Do ponto de vista linguístico, essas noções são complexas e mereceriam estudos mais aprofundados, de modo a contribuir para haver maior precisão, mais rigor e critério para definir tal princípio numa política de indexação. No entanto, essas decisões no tratamento da informação supõem atualizações permanentes, conforme se modifiquem as condições institucionais e/ou se altere o conhecimento humano, pois isso pode afetar a linguagem natural e, consequentemente, as linguagens documentárias.

Um dos principais objetivos de um tesauro é possibilitar o tratamento e a representação do conteúdo dos documentos através de termos (unidades que representam conceitos da área), que são fonte de pesquisa para o trabalho de indexação de documentos bem como para a recuperação destes. O estabelecimento das redes relacionais também leva ao controle terminológico, alcançado por meio de modificadores, definições e notas de escopo, uma vez que possibilita fixar o lugar dos termos em relação aos demais. Essas relações entre os termos de um tesauro formam a rede paradigmática. Dessa forma, os tesauros podem agilizar a consulta aos dados estocados, facilitando a manutenção e atualização, além de

[34] É um conjunto de critérios pré-estabelecidos que define regras mínimas quanto aos procedimentos a observar no processo de indexação. Têm caráter racionalizador e condiciona os resultados da estratégia de busca. Consequentemente, resultam na economia de tempo, de recursos materiais e financeiros; portanto é uma decisão administrativa indispensável a um sistema de recuperação de informação. Por exemplo, o Sistema de Informação e Conhecimento do Senac (2002) estabelece em seu manual, uma política de indexação nos seguintes termos, o item 4.2: "Nossa política de indexação se refere a dois aspectos importantes que influem no desempenho de um sistema de informação a saber: generalidade e especificidade".

padronizar os procedimentos de organização da informação das equipes e até conferir confiabilidade ao sistema de informações.

A Norma ISO 2788, embora insuficiente, tem sido uma referência para orientar a estruturação de tesauros. Mas numerosos trabalhos, principalmente os de Campos (2004a, 2004b) e Dodebei (2002), têm como propósito auxiliar o profissional na hierarquização dos conceitos na elaboração de tesauros, preenchendo uma lacuna existente entre os requisitos do sistema informatizado de organização e os processos de seleção, em grande parte, humanos, que determinam a hierarquia final. Verificamos que essa lacuna corresponde à compatibilização de traços distintivos claramente estabelecidos, de modo a indicar parâmetros de inclusão e organização dos dados; muitos deles de ordem linguística. Notamos que há diferentes orientações como a Teoria do Conceito e Teoria da Classificação Facetada, tratadas como complementares, que tentam auxiliar no levantamento de requisitos para compreensão dos conceitos, balizando questões norteadoras do estabelecimento das definições do tipo: o que é?; para quê serve?; de que partes é constituído?; ou em qual dimensão deverá ser incluído?; ou ainda a qual delas deverá ser relacionado?

Dessa forma, podemos concluir que há muito por fazer no campo dos estudos linguísticos em relação à formulação parâmetros orientadores, bem como no estabelecimento de regras linguísticas do PLN, que também podem servir de referência à estruturação das próprias linguagens de programação. Assim, as ferramentas de TI mais aptas a atender cada vez melhor as demandas do usuário, estarão mais próximas dos usos em língua natural. Muitas dessas regras já estão na base de várias ações previstas nos *softwares* que utilizamos; atrás de cada uma delas, há uma estruturação linguística traduzida na linguagem que aquele programa entende. Haja vista que a Inteligência Artificial começou com a linguística computacional, no acionamento de um braço mecânico através de comandos formulados em linguagem natural...

Porém, acreditamos que todo o estranhamento dessa interface no universo de Letras esteja no próprio objeto da linguagem tão infensa a estruturas limitantes, em se tratando de uma língua. Grande parte das regras linguísticas em LN já são objetos de polêmicas no interior da própria língua, dependendo de sua natureza e função; que dirá em uma interface tecnológica com seus modelos estruturantes. Porém, as tecnologias de informação, cada vez mais, tornam-se instrumentos de mediação das interações e estão à nossa disposição para executarmos aquelas tarefas de cunho repetitivo, possibilitando-nos visões antes impossíveis e, consequentemente, incitando outras relações de raciocínio e de produção de sentido as quais somente a capacidade da mente humana pode empreender.

O tesauro é um desses instrumentos, pois permite uma visão de conjunto que auxilia na dimensão da relação todo e parte, para melhor interpretação do seu contexto específico (área de conhecimento). Esse diferencial de organização vai ao encontro de nossas formas básicas de organização cognitiva, portanto, podem funcionar como elementos de conforto e induzem a avanços nas nossas conclusões. E, como ignorar a dimensão linguística desse processo? Ela sempre esteve lá, aquém dos próprios linguistas, realizada intuitivamente por profissionais de outras áreas que, muitas vezes, como bons usuários da língua – que somos todos -, partiram em busca do estabelecimento de princípios linguísticos de modo a adequar as ferramentas para o uso de todos nós, até dos próprios linguistas.

Dos tesauros analisados, aqueles que contêm as notas de aplicação sobre os conceitos propiciam análises mais "ricas", do ponto de vista linguístico, pois permitem o levantamento dos traços que justificam os parâmetros empregados para a efetivação da inclusão de termos. Certamente, os princípios de categorização necessitam de um melhor tratamento, pois também funcionam como critérios de organização. Além de ser necessário, ainda, trabalhar nos protótipos das classes dos objetos que, associados aos campos semânticos, poderiam fornecer um direcionamento bastante satisfatório para os ajustes das relações de pertinência. Observamos, por exemplo, que os princípios morfológicos da constituição do léxico são muito pouco utilizados, o que é justificado pelos profissionais de CI, pela própria natureza de hierarquização dos conceitos expressos pelos termos.

Das categorias de análise do conceito de Aristóteles, substância para *entidades*, quantidade, qualidade e relação para as *propriedades*, e de tempo, posição e espaços para *dimensões*, às linhas teóricas desenvolvidas por Ranganathan, que propôs cinco categorias: **e**ntidade, **m**atéria, **e**nergia, **e**spaço e **t**empo, o homem vem refletindo sobre formas de organizar os conhecimentos muito antes do advento da computação. Nessa linha teórica, a Terminologia, representada pelas concepções de Dahlberg, retoma os princípios aristotélicos e propõe, para as *entidades*, as categorias fenômeno, objeto geral e objeto material; para as *propriedades*, contagem e mensuração, a qualidade e a comparação; e para as *dimensões*, tempo, posição e espaço, aqui, já numa perspectiva de recuperação dos dados pelas tecnologias da informática.

As relações entre categorias, as relações hieráquicas e as associativas bem como as de equivalência, ressaltadas no início desta seção e tratadas em várias partes deste trabalho, podem traduzir parte das relações que empreendemos em nossas redes neurais, fundamentadas em nossas memórias semânticas, para interagirmos em sociedade. E, certamente, estão associadas aos nossos processos de compreensão, ponto que não foi

contemplado por esta tese, porém indicativo para investigações futuras mediadas por questões como: de que forma são selecionados, por exemplo, os termos que representam um dado documento?

Gostaríamos de trazer o discurso do Prof. Bento Carlos, professor da Universidade Estadual Paulista, no Departamento de Letras Modernas e pesquisador do NILC, Núcleo Interistitucional de Linguística Computacional, onde desenvolve estudos na área de PLN, Processamento de Língua Natural, cujo conteúdo traduz um pouco nosso ponto de vista. Ele situa essa nova frente de estudos como uma "antropomorfização":

> A antropomorfização da máquina pode ser justificável do ponto de vista da inteligibilidade das explicações. Essa estratégia discursiva, no entanto, acaba gerando também interpretações que contribuem para formar a imagem de que pesquisar PLN é adotar uma visão mecanicista do homem, ou ainda, que o PLN é, por definição, um campo desprovido de conceitos e termos precisos para descrever seus objetos. A precária troca de trabalhos sobre o PLN entre os linguistas e os projetistas de sistemas de PLN fica também evidente nas raras publicações traduzidas para o português. Os textos, que em sua massiva maioria estão publicados em inglês, quando são traduzidos, além de trazerem termos bizarros, apresentam imprecisões e confusões conceituais. (SILVA, 2006, p.12)r

E, continua o autor, em outro ponto de sua fala, quando ressalta que o estudo do PLN nos coloca diante de uma outra dificuldade: "uma metalinguagem fragmentada, um hermetismo terminológico e uma multiplicidade de termos exóticos". (SILVA, 2006, p.13).

Concluímos, portanto, afirmando que temos ainda muito a pesquisar para esclarecer o papel da língua na interface com as novas tecnologias da informação. Até mesmo documentos oficiais como os Parâmetros Curriculares Nacionais para o Ensino Médio (PCNEM), já preveem o desenvolvimento de competências relacionadas à área de "Linguagens, Códigos e suas Tecnologias". Propõem, por exemplo, que o aluno utilize os recursos computacionais, que saiba manuseá-lo e conheça a nomenclatura de seus diferentes componentes. De acordo com os PCNEM, a informática visa a propiciar que o aluno seja capaz de "conectar os inúmeros conhecimentos com suas aplicações tecnológicas" (1999, p. 106). São as novas gerações que são educadas, tendo como referência a construção e compartilhamento de conhecimentos via TI's.

É interessante notar que já temos, inclusive, um "Thesaurus da Língua Portuguesa do Brasil" iniciado pelo Prof. Valdir Jorge, professor formado em Letras, pela USP, que se dispôs a apresentar à "comunidade lusófona e lusófila uma ferramenta de trabalho que considera importante" (JORGE, 2004). Em contato com o Prof. Valdir, ele informou que não fez mais atualizações no referido tesauro e que, hoje, atua como administrador de banco de

dados e programador na Universidade de Concórdia, em Montreal. Indicou-nos, no entanto, um outro projeto colaborativo, do qual participa na atualidade e nos convida, como profissionais linguistas, a colaborarmos: o Wikicionário, um dicionário universal de conteúdo livre, no qual a língua portuguesa está representada.

Há que se ressaltar também que, no *site* da Universidade Federal de Santa Catarina, há uma listagem dos tesauros existentes nas mais diversas áreas, no mundo todo, que foi disponibilizado pelo Centro de Ciências da Educação; trata-se da "Biblioteca Virtual nas áreas de Biblioteconomia e Ciência da Informação – Thesaurus". Nessa listagem, somam-se, hoje, um total aproximado de cinquenta (50) tesauros relacionados, muitos com uso de variados recursos, de modo a facilitar as consultas, observando-se, porém, que a maior parte deles está em língua estrangeira ou foi elaborado fora do Brasil. Universidades como a PUCMINAS deveriam dispor de instrumentos dessa ordem que facilitariam a indexação do grande volume de documentos aqui produzidos e armazenados. Esta é uma tarefa que consta das metas da bibliotecária da PUCMINAS, Helenice Rêgo, da Biblioteca Pe. Alberto Antoniazzi.

Como o tesauro é um dispositivo de controle terminológico, para auxiliar a construção de estruturas conceituais é necessário o uso de *softwares* para sua construção, de modo a facilitar o armazenamento, manipulação e apresentação dos termos definidos e suas relações, aspecto que não foi contemplado neste trabalho. No entanto, o objetivo do tesauro e a escolha criteriosa de um *software* para sua implementação são importantes decisões prévias. Porém, definir tesauro apenas como uma linguagem documentária utilizada para recuperação de informação, pode ser muito sintetizador, pois notamos que se trata de um instrumento que, atualmente, se expande em aplicações. Observando seu percurso histórico, alguns deles apresentados neste trabalho, podemos afirmar que um século de história e aplicações talvez justifiquem a crescente importância desse instrumento de representação da informação que, quando construído com um consistente suporte metodológico e uma ampla equipe multidisciplinar, torna-se um aliado eficaz para a recuperação de informações.

REFERÊNCIAS

ANDRADE, Maria Margarida de. Lexicologia, terminologia: definições, finalidades, conceitos operacionais. In: OLIVEIRA, Ana Maria Pinto Pires de; ISQUERDO, Aparecida Negri. **As ciências do léxico:** lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: UFMS, 2001.

ARTENCIO, Luciane Maria. **Princípios de categorização nas linguagens documentárias**. 2008. 129f. Dissertação (Mestrado em Comunicação) – Universidade de São Paulo, São Paulo. Disponível em <http://www.teses.usp.br/teses/> Acesso em jan 2009.

AUSTIN, D. & DALE, P. (1993) **Diretrizes para o estabelecimento e desenvolvimento de tesauros monolíngues**. Trad. De Bianca Amaro de Melo. Brasília: IBICT/SENAI.

AZEVEDO, Francisco Ferreira dos Santos. **Dicionário analógico da língua portuguesa (idéias afins).** Brasília: Coordenada: Thesaurus, 1983.

BARBOSA, Maria Aparecida. **Léxico, produção e criatividade:** processos do neologismo. São Paulo: Global, 1981.

BAKHTIN, M. M. **Marxismo e filosofia da linguagem:** problemas fundamentais do método sociológico na ciência da linguagem. 10. ed. São Paulo: Annablume, Hucitec, 2002.

BEGHTOL, C. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation,** London, v. 58, n. 5, p. 507-532, 2002a.

BEGHTOL, C. Universal concepts, cultural warrant, and cultural hospitality. In: LÓPEZHUERTAS, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21st century:** integration of knowledge across boundaries. Würzburg: ERGON-Verlag, 2002b. p. 45-49.

BENVENISTE, Émile.**Problemas de Linguística geral II**. Campinas: Pontes, 1989.

BIBLIOTECA VIRTUAL NAS ÁREAS DE BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO/THESAURUS. Disponível em http://www.ced.ufsc.br/bibliote/virtual/thesauros.html Acesso em jul/2009

BLIKSTEIN, Isidoro: **Kaspar Hauser ou a fabricação da realidade**. São Paulo: Cultrix, 1990.

BORGES NETO, J.. Semântica de Modelos. In: MÜLLER, Ana Lúcia; NEGRÃO, Esmeralda Vailati & FOLTRAN, Maria. José (orgs). In: **Semântica Formal**. São Paulo: Contexto, 2003.p. 9-46.

BRASIL. Secretaria de Educação Média e Tecnológica. **Parâmetros curriculares nacionais:** ensino médio. Brasília: Ministério da Educação, 1999.

CABRÉ, María Teresa. *La terminología: representación y comunicación*. Barcelona: Institut Universitari de Linguística Aplicada de la Universitat Pompeu Fabra, 1999.

CÂMARA JÚNIOR, J. Mattoso. **Princípios de linguística geral:** como introdução aos estudos superiores da língua portuguesa. 6. ed. com atualização ortográfica Rio de Janeiro: Padrão, 1980.

CAMPOS, Maria Luiza de Almeida. **Modelização de domínios de conhecimento: uma investigação de princípios fundamentais.** Ciência da Informação. Brasília, v33, n.1, 2004.

CAMPOS, Maria Luiza Almeida; GOMES, Hagar Espanha. **Tesauro e normalização terminológica**: o termo como base para intercâmbio de informações. DataGramaZero - Revista de Ciência da Informação - v.5 n.6 dez/2004.

CAMPOS, Maria Luiza Almeida. **Linguagem Documentária**. Niterói: EdUFF, 2001.

CAMPOS, Jorge. Metateoria Linguística. (Considerações ao Nível da Filosofia da Ciência). **Revista da ADPPUCRS.** FALE, Porto Alegre, nº. 5, p. 25-32, dez. 2004.

CAPURRO, Rafael; HOJRLAND, Birger. O conceito de informação. **Perspectivas em ciência da informação.** Belo Horizonte: UFMG, Escola de Biblioteconomia, Quadrimestral. v.12, n.1, p.148-207, jan./abr. 2007. Trad. Ana Maria Pereira Cardoso, Maria da Glória Achtschin Ferreira, Marco Antônio de Azevedo.

CAVALCANTI, C. R. **Indexação e tesauro: metodologia e técnica**. Brasília, ABDF, 1978.

COELHO, Marisa Colmago. **Tesauro do Folclore e Cultura Popular Brasileira**. Disponível em http://www.cnfcp.gov.br/tesauro/sistematica.html Acesso em jan 2009.

COHEN, B & MURPHY, G. L. *Models of concepts.* **Cognitive Science**, 8, 1984.

DAHLBERG, Ingetraut. **Teoria do conceito**. Ciência da Informação v. 7, n. 2, p. 101-07, 1978.

DICK, Maria Vicentina de Paula do Amaral. O sistema onomástico: bases lexicais e terminológicas, produção e frequência. In: OLIVEIRA, Ana Maria Pinto Pires de; ISQUERDO, Aparecida Negri. **As ciências do léxico:** lexicologia, lexicografia, terminologia. 2. ed. Campo Grande: UFMS, 2001.

DODEBEI, Vera Lucia Doyle. **Tesauro**: linguagem de representação da memória documentária. Niterói; Rio de Janeiro: Intertexto; Ed. Interciência, 2002.

DUCROT, Oswald. **Princípios de semântica Linguística**: (dizer e não dizer). São Paulo: Cultrix, 1977.

ECO, Umberto. **Semiótica e filosofia da linguagem.** São Paulo: Ática, 1991.

FIORIN, José Luiz. Sendas e veredas da semiótica narrativa e discursiva. **DELTA** [on line] 1999, vol.15, n.1.

FODOR, Jerry. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press, Cambridge, 1987.

FODOR, Jerry. *Representations*: Philosophical Essays on the Foundations of **Cognitive Science,** MIT Press, Cambridge, 1983.

FODOR, Jerry. *Concepts: a potboiler*. **Cognition,** 50, 1994.

FREGE, G. **Sobre o conceito e o objeto**. In: Lógica e Filosofia da Linguagem. São Paulo: Cultrix, 1978.

FREGE, Gottlob. **Sobre o Sentido e a Referência.** In: Lógica e Filosofia da Linguagem. São Paulo, Cultrix, 1978.

GENOUVRIER, Emile; PEYTARD, Jean. **Linguística e ensino do português.** Coimbra: Almedina, 1974.

GREIMAS, Algirdas Julien. **Semântica estrutural: pesquisa de método.** 2. ed. São Paulo: 1976.

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida; MOTTA, Dilza Fonseca da.**Tutorial para elaboração de Tesauros. (online)**. Disponível na Internet http://www.conexaorio.com/biti/tesauro

GUIMARÃES, Eduardo. **História da semântica:** sujeito, sentido e gramática no Brasil. Campinas: Pontes, 2004.

HOUAISS, Antônio. **Dicionário eletrônico Houaiss da língua portuguesa:** versão 2.0. Rio de Janeiro: Objetiva, 2006. 1 CD-ROM

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). **Tesauro Brasileiro da Educação (Brased).** Disponível em <http://www.inep.gov.br/pesquisa/thesaurus> Acesso em jul 2009.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *British standard guide to establishment and development of monolingual thesauri: 2788*. London : BSI, 1986. Disponível em <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=31696> Acesso em out 2008

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Principles and methods of terminology:* 704. 2. ed. [S. l. : s. n.], 2000. <http://www.iso.org/iso/home.htm> Acesso em nov 2008

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Terminology work – principles and methods:* 1087-1. Part 1: theory and application = *Travaux terminologiques – vocabulaire.* Partie 1: théorie et application. [S. l. : s. n.], 2000. Disponível em <http://www.iso.org/iso/home.htm> Acesso em nov 2008

ISO 5964. *Guidelines for the establishment and development of multilingual thesauri*, 1985. Disponível em: <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=12159> Acesso em 6 jun. 2008.

JACKENDOFF, R. *Semantics and Cognition.* Cambridge MASS: MIT Press, 1983.

JOHNSON-LAIRD, P. N. *Mental Models in Cognitive Science*. **Cognitive Science**, 4, 1980.

JORGE, Valdir. **Thesaurus da Língua Portuguesa do Brasil.** 2004. Disponível em http://alcor.concordia.ca/~vjorge/Thesaurus/ Acesso em jul/2009

KATZ, Jerraud J. e FODOR, Jerry A. In: Lobato, Lúcia Maria Pinheiro (org.). **A semântica na Linguística moderna: o léxico**. Rio de Janeiro: Francisco Alves, 1977.

KATZ, Jerrauld J.. O escopo da semântica. In: DASCAL, M (org.) **Semântica**. Campinas: Instituto de Estudos da Linguagem da UNICAMP, 1982.

KLEIBER, Georges. *La Semántica de los Prototipos: Categoria y sentido léxico*. Madrid: Visor, 1995.

KOCH, Ingedore Grunfeld Villaça. **Desvendando os segredos do texto.** 2. ed. São Paulo: Cortez, 2003.

KOCH, Ingedore Grunfeld Villaça. **A coesão textual.** 15.ed. São Paulo: Contexto, 2001.

KOCH, I. V.; Marcuschi, L. A.. **Processos de referenciação na produção discursiva**. D.E.L.T.A., 14 (Especial): 169-190, 1998.

KRIEGER, Maria da Graça. **Terminologia revisitada** (Terminology Revisited). In: D.E.L.T.A., Vol. 16, N.º 2, 2000 (209-228)

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

LAKOFF, George. 1987.

LAKOFF, George; JOHNSON, Mark. **Metáforas da vida cotidiana.** São Paulo: Educ; Campinas: Mercado de Letras, 2002. (As faces da vida linguística)

LEVY, Pierre. **As tecnologias da inteligência**. Rio de Janeiro, Ed 34, 1993.

LÍNGUA. Google Disponível em <http://www.google.com.br/search?hl=pt-BR&q=lingua&meta=&aq=1&oq=l%C3%ADngua> Acesso em 10 abr 2009.

LOBATO, Lucia Maria Pinheiro (org). **A semântica na Linguística moderna:** o léxico. Rio de Janeiro: F. Alves, 1977.

LOPES, I. L. **Uso das linguagens controlada e natural em bases de dados:** revisão de literatura. Ciências da Informação, v.31, n.1. p. 41-52, 2002.

LYONS, John. **Semântica.** Lisboa: Presença: Martins Fontes, 1980. 2v.

MADEIRA, M. J. P. **Processos cognitivos**: O desenvolvimento de conceitos mentais. Psico, 1987. p. 162-165.

MARI, Hugo. **Os lugares do sentido**. Campinas: Mercado das letras, 2008.

MARI, Hugo. Categorização. In: MARTINS, Roberto P.; MAR, Hugo (orgs.). **Universos do conhecimento**, Belo Horizonte: FALE/UFMG, 2002. p. 67-89.

MARI, Hugo. Dos fundamentos da significação à produção do sentido. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.1, n.1, p.93-109, jan./jun. 1996.

MONTAGUE, R. English as a formal language. MONTAGUE, R. *Formal philosophy: selected papers of Richard Montague*. THOMASON, R. H. (ed.). News Haven: Yale University, 1974.

MOREIRA, Manoel Palhares. **Ambiente para geração e manutenção semi-automática de tesauros**. 2005. 197f. Tese (Doutorado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2005.

MOREIRA, Manoel Palhares; MOURA, Maria Aparecida. **Construindo tesauros a partir de tesauros existentes**: a experiência do TCI - Tesauro em Ciência da Informação. DataGramaZero - Revista de Ciência da Informação - v.7. n.4 ago/06. Disponível em: <http://www.dgz.org.br/ago06/F_I_art.htm> Acesso em: jul. 2009.

MOREIRA, Manoel Palhares. **Tesauro em Ciência da Informação.** Disponível em <http://www.inf.pucminas.br/ci/tci/> Acesso jan 2009.

OGDEN, C.K. & RICHARDS, I.A.: *The Meaning of Meaning*, New York, Hartcourt, Brace & Co., 1956.

OLIVEIRA, Roberta Pires de. **Semântica formal:** uma breve introdução. São Paulo: Mercado de Letras, 2001. (Coleção "Ideias sobre linguagem")

OLIVEIRA, Roberta Pires de. Semântica. In: MUSSALIM, Fernanda; BENTES, Anna Christina. **Introdução à linguística: domínios e fronteiras.** 2. ed. São Paulo: Cortez, 2001. p. 17-46.

OSHERSON, D. N. & SMITH, E.E.. *On adequaçy of prototype theory as a theory of concepts.* In: **Cognition,** 9, 1981.

PEIRCE, Charles Sanders. **Semiótica e filosofia**. São Paulo: Cultrix, 1972.

PEIRCE, Charles Sanders.; FREGE, Gottlob. **Escritos coligidos.** 2. ed. São Paulo: Abril Cultural, 1980. (Os pensadores)

PINHO, Fabio Assis. **Aspectos éticos em representação do conhecimento**: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol. 2006. 132f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências da UNESP, Marília. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/PesquisaObraForm.do?select_action=&co_autor=18734> Acesso em: jul. 2009

PORTO, André da Silva. Quatro espaços semânticos e modais. In: **As bases filosóficas do construtivismo matemático de Wittgenstein**. 2002. 211f. Tese (Doutorado em Filosofia) – Pontifícia Universidade Católica do Rio de Janeiro. P.8-48. Disponível em <http://www.maxwell.lambda.ele.puc-rio.br> Acesso em: fev 2009

ROSCH, Eleanor. *Cognitive representations of semantic categories*. **Journal of Experimental** Psychology, 1975. p.192-223.

ROSCH, Eleanor & LLOYD, Barbara B. (eds.). *Cognition and Categorization*. Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1978.

ROSCH, Eleanor. *Reclaiming Concepts.* In R. Nunez & W.J. Freeman (Eds.) (1999). Reclaiming cognition: The primacy of action, intention and emotion. Thorverton, Eng.: Imprint Academic. Published simultaneously as a special issue of The Journal of Consciousness Studies, 6, No. 11-12, 61-77. Disponível em <psychology.berkeley.edu/faculty/profiles/erosch.html> Acesso em: maio 2008.

SARGENTIM, Hermínio G. **Dicionário de idéias afins.** São Paulo: IBEP, [19-].

SAUSSURE, Ferdinand de. **Curso de Linguística geral**. 25. ed. São Paulo: Cultrix, 2003.

SHANNON, C.; WEAVER, W. *The mathematical theory of communication*. Urbana, Chicago, London: University of Illinois Press, 1971.

SVENONIUS, Elaine. *The intellectual foundation of information organization*. Cambridge, MA: MIT Press, 2000.

TEMMERMAN, Rita. *Towards new ways of terminology description: the socio-cognitive approach.* Amsterdam: John Benjamins, 2000.

SILVA, Bento Carlos Dias da. **O estudo Linguistico-Computacional da Linguagem**. Letras de Hoje. Porto Alegre. v. 41, nº 2, p. 103-138, junho, 2006. Disponível em <http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/viewFile/597/428> Acesso em: jan 2009

SISTEMA DE INFORMAÇÃO E CONHECIMENTO DO SENAC - SERVIÇO NACIONAL DE APRENDIZAGEM COMERCIAL. Rio de Janeiro, 2002. Disponível em <http://www.dn.senac.br/cedoc/Indexacao.pdf> Acesso em set 2009.

SMITH, E. E. & MEDIN, D. L. *Categories and Concepts*. Cambridge, MA: Harvard University Press, 1981.

TAMBA-MECZ, Irene. **A semântica**. São Paulo: Parábola, 2006. Tradução: Marcos Marciolino.(Coleção "Na ponta da língua")

TAMBA, Irene. *Organisation hiérarchique et relations de dépendance dans le lexique*. **L'information Grammaticale.** 50:43-47, 1991.

ULLMANN, Stephen. **Semântica:** uma introdução a ciência do significado. 3. ed. Lisboa: Fundação Gulbenkian, 1973.

UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973.

VIGOTSKY, L. S. **Pensamento e linguagem.** 3. ed. São Paulo: Martins Fontes, 2005.

ZAPPAROLI, Zilda Maria; CAMLONG, André. **Do léxico ao discurso pela informática.** São Paulo: Edusp, FAPESP, 2002.

WITTGENSTEIN, Ludwig. **Investigações filosóficas.** 2. ed. São Paulo: Abril Cultural, 1979. (Os pensadores)