

Universidade Federal do Rio Grande do Sul
Instituto de Letras
Programa de Pós-Graduação em Letras
Área: Estudos da Linguagem
Especialidade: Teorias do Texto e do Discurso
Linha de pesquisa: Lexicografia e Terminologia: Relações Textuais

**DOCUMENTAÇÃO, TERMINOLOGIA E LINGÜÍSTICA:
UMA INTERFACE PRODUTIVA**

Vera Maria Araujo Pigozzi de Araujo

Dissertação de Mestrado em Estudos da Linguagem
apresentada ao Programa de Pós-Graduação em
Letras da UFRGS como requisito parcial para a
obtenção do título de Mestre em Letras.

Porto Alegre
2006

Universidade Federal do Rio Grande do Sul
Instituto de Letras
Programa de Pós-Graduação em Letras
Área de Estudos da Linguagem
Especialidade: Teorias do Texto e do Discurso
Linha de pesquisa: Lexicografia e Terminologia: Relações Textuais

**DOCUMENTAÇÃO, TERMINOLOGIA E LINGÜÍSTICA:
UMA INTERFACE PRODUTIVA**

Vera Maria Araujo Pigozzi de Araujo

Dissertação de Mestrado em Estudos da Linguagem
apresentada ao Programa de Pós-Graduação em
Letras da UFRGS como requisito parcial para a
obtenção do título de Mestre em Letras.

Orientadora: Profa. Dra. Anna Maria Becker Maciel

Porto Alegre
2006

Aqueles que passam por nós não se vão sós, não nos deixam sós. Deixam um pouco de si, e levam um pouco de nós.

As tempestades, a bruma e a neve vão incomodá-lo algumas vezes. Pense, então, em todos aqueles que já passaram por isso e diga para si mesmo: o que eles conseguiram eu também posso conseguir.

Meu amigo é, muitas vezes, aquele que tem uma idéia contrária à minha, pois me engrandece. Ele me obriga à superação.

Crescemos cada vez que cumprimos uma tarefa.

Antoine de Saint-Exupéry

DEDICATÓRIA

Dedico este trabalho àqueles que passaram em minha vida e marcaram a minha personalidade pelo seu exemplo de luta, superação e dedicação e que em momentos diferentes desta dissertação, pelo seu exemplo, me deram força para continuar nesta difícil tarefa que decidi assumir. Ao meu pai, aos meus sogros, à minha irmã Sílvia, que como Bibliotecária abriu caminho para que eu seguisse nesta profissão. À minha grande amiga e colega Bibliotecária Maria Lizete, uma grande incentivadora para que eu fizesse o Mestrado, meu muito obrigada e minha eterna saudade. Um dia nos reencontraremos.

Dedico ainda este trabalho a duas grandes amigas e colegas Bibliotecárias: Maria Dinah Penz Bergmüller, por ter acreditado em mim desde o início da minha vida profissional e a Iara Conceição Neves, por ter me passado conhecimentos que me valeram para a vida toda.

Dedico também à minha mãe, à minha família, meus filhos, nora e genros que me deram todo o amor e apoio que precisei para produzir este trabalho.

Em especial dedico ao meu marido e ao meu neto. Ao meu marido, companheiro de mais de 40 anos, que com uma dedicação implacável me acompanhou nesta minha grande e aparentemente interminável caminhada; e ao meu neto, por me fazer perceber que a vida é um eterno renascer e que apesar de todos os dissabores, ainda vale a pena viver com alegria, força e muita energia.

AGRADECIMENTOS

À minha Orientadora, que muito contribuiu para o êxito deste trabalho. Suas idéias, ainda que muitas vezes contrárias às minhas, me conduziram a reflexões cujos resultados superaram às minhas expectativas iniciais. Meu propósito, aqui, não é fazer-lhe um agradecimento, pois esse não pode ser expresso em palavras. O que desejo é mostrar a minha admiração e o meu reconhecimento à sua sabedoria, dedicação, energia e perspicácia que só os grandes Mestres possuem. Já dizia Saint-Éxupéry: “É a experiência que dita as leis.”

Meu reconhecimento e eterna admiração pelos grandes Mestres com quem eu tive o prazer de conviver e aprender nesta minha trajetória como bibliotecária, como funcionária da BSCSH e como aluna deste Curso: Maria da Graça Krieger, Maria José Bocorny Finatto, Cleci Regina Bevilacqua, Luciene Juliano Simões, Sandra Sirangelo Maggio, Valdir Nascimento Flores, Félix Valentin Bugueño Miranda, Mário Klassmann e Antônio David Cattani.

Meu agradecimento às minhas colegas e amigas bibliotecárias que contribuíram com suas reflexões na composição desta dissertação, em especial à: Maria Cristina Bürger, Maria Lizete Gomes Mendes (In Memoriam), Maria Teresa Ferlini Machado, Inês Maria de Gasperin, Letícia Strehl, Maria Hedy Lubisco Pandolfi, Rejane Raffo Klaes e Vera Lúcia Linhares Dias.

Agradeço ainda a todos os colegas da BSCSH e aos bolsistas que trabalharam comigo na minha trajetória inicial, no tempo da execução dos projetos FAPERGS e CNPq.

RESUMO

Esta pesquisa tem como objetivo principal propor uma metodologia que agilize a construção de uma ferramenta no campo da Documentação. Trata-se da geração de uma base de dados terminológica com sustentação na terminologia utilizada pelo especialista em sua área de domínio. Ela se apóia nos pressupostos teóricos da Teoria da Enunciação, da Teoria Comunicativa da Terminologia e da Socioterminologia. Com esse referencial acredita-se ser possível assegurar a efetiva comunicação entre os Sistemas de Recuperação de Informação e os usuários, sendo o bibliotecário o mediador do processo comunicativo que tem origem no autor do texto indexado. Buscou-se o suporte da Terminografia e da Linguística de *Corpus* pela possibilidade de coletar, tratar e armazenar um grande volume de informações de uma determinada área do saber.

Palavras-chave: Análise documentária. Base de dados terminológica. Documentação. Linguística de *Corpus*. Sistemas de Recuperação da Informação. Socioterminologia. Teoria da Enunciação. Teoria Comunicativa da Terminologia. Terminografia. Terminologia.

ABSTRACT

The main objective of this study is to propose a methodology by which the implementation of a tool in the field of Documentation can be optimized. It deals with a terminological database built on the terminology used by the specialist in his area. Its is backed by the underlying assumptions of the Theory of Enunciation, of the Communicative Theory of Terminology and of the Socioterminology. It is believed that, in this theoretical framework, it is possible to ensure effective communication between the Information Retrieval Systems and users, being the librarian the mediator of communicative process originated in the author of the indexed text. The research draws from resources of Terminography and *Corpus* Linguistics in order to operationalize the process of collecting, managing and storing a huge amount of information of a given knowledge field.

Keywords: Documentary Analysis. Terminological Database. Documentation. *Corpus* Linguistics. Information Retrieval Systems. Socioterminology. Theory of Enunciation. Communicative Theory of Terminology. Terminography. Terminology.

LISTA DE ILUSTRAÇÕES

QUADRO 1 - Arquitetura da base de dados terminológica	124
QUADRO 2 - Terminologia da base de dados terminológica em LINGÜÍSTICA TEÓRICA APLICADA	126
TABELA 1 - Comparativo entre UT/UFE e DESCRITORES/TERMOS	129

LISTA DE SIGLAS

- ABNT – Associação Brasileira de Normas Técnicas
- ABRALIN – Associação Brasileira de Lingüística
- AD – Análise documentária
- ANPOLL – Associação Nacional de Pós-Graduação e Pesquisa em Letras e Lingüística
- AR – Autor Relacionado
- BIREME – Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde
- BNC – British National *Corpus*
- BSCSH – Biblioteca Setorial de Ciências Sociais e Humanidades
- BTQ – Banco de Terminologia do Quebec
- CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- CDD – Classificação Decimal de Dewey
- CDU – Classificação Decimal Universal
- CEPRIL – Centro de Pesquisa, Recursos e Informação em Linguagem
- CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico
- CSH – Ciências Sociais e Humanidades
- ENANPOLL – Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Letras e Lingüística
- FAPERGS – Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul
- FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo
- GTs – Grupos de Trabalho
- INFOTERM - International Information Centre for Terminology
- ISO - International Organization for Standardization
- IULATERM – Grupo de Terminologia do IULA (Institut Universitari de Lingüística Aplicada)
- LAEL – Programa de Estudos Pós-Graduados em Lingüística Aplicada e Estudos da Linguagem
- LC – Library of Congress
- LD – Linguagem documentária
- LLBA – Linguistic and Language Behavior Abstracts
- MOD – Modificador
- NBR – Norma Brasileira Registrada
- PUC-SP – Pontifícia Universidade Católica de São Paulo
- PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul

RI – Relações Interdisciplinares
SABI/UFRGS – Sistema de Automação de Bibliotecas / UFRGS
SciELO – Scientific Electronic Library Online
SIBi/USP – Sistema Integrado de Bibliotecas / USP
SRI – Sistema de Recuperação de Informação
TCT – Teoria Comunicativa da Terminologia
TE – Termo Específico
TEP – Termo Específico Partitivo
TEs – Termos Equivalentes
TG – Termo Genérico
TGP – Termo Genérico Partitivo
TGT – Teoria Geral da Terminologia
TRs – Termos Relacionados
UFE – Unidade Fraseológica Especializada
UFEI – Unidade Fraseológica Especializada em língua inglesa
UFRGS – Universidade Federal do Rio Grande do Sul
UNESCO - Organização das Nações Unidas para a Educação, a Ciência e a Cultura
UNICAMP – Universidade Estadual de Campinas
USE – Unidade de Significação Especializada
USP – Universidade de São Paulo
UT – Unidade Terminológica
UTI – Unidade Terminológica em língua inglesa

SUMÁRIO

1 INTRODUÇÃO	12
1.1 Antecedentes	16
1.2 Estado da arte	18
1.3 Hipóteses de trabalho	24
1.4 Objetivo geral	24
1.5 Objetivos específicos	24
1.6 Público-alvo	25
1.7 Estrutura da Dissertação	25
2 LINGUAGEM E SENTIDO	28
2.1 Introdução	28
2.2 Linguagem	29
2.3 Reflexões sobre o sentido, significado e significação	32
2.3.1 <i>Sentido sob diferentes pontos de vista</i>	34
2.3.2 <i>Sentido na visão da Lingüística Textual</i>	37
2.3.3 <i>Sentido no âmbito desta pesquisa</i>	40
3 DOCUMENTAÇÃO	45
3.1 Introdução	45
3.2 Biblioteconomia e Documentação	45
3.3 Informação e Conhecimento	48
3.4 Sistema nocional	51
3.5 Linguagem documentária	54
3.6 Processo de indexação	59
3.7 Sistemas de Recuperação da Informação	63
3.8 Vocabulário controlado	71
3.9 Novo paradigma	74
4 TERMINOLOGIA	77
4.1 Introdução	77
4.2 Terminologia	78
4.3 Histórico	84
4.4 Termo	93
4.5 Relação termo e descritor	98

4.6 Fraseologia	99
4.7 Comunicação especializada	102
4.8 Terminologia e Informática	103
4.7.1 <i>Linguística de Corpus</i>	103
4.7.2 <i>Terminografia</i>	108
4.7.3 <i>Bancos e bases de dados terminológicos</i>	111
5 PROPOSTA PARA UMA BASE DE DADOS TERMINOLÓGICA	115
5.1 Considerações preliminares	115
5.2 Seleção da área	116
5.3 Constituição do <i>Corpus</i>	116
5.4 Seleção das ferramentas	119
5.4.1 <i>Ferramentas de análise dos textos</i>	119
5.4.2 <i>Instrumentos de gerenciamento de dados</i>	120
5.5 Coleta e análise de dados	120
5.5.1 <i>Levantamento de listas</i>	120
5.5.2 <i>Seleção de palavras representativas dos grandes domínios da área</i>	121
5.5.3 <i>Produção de concordâncias</i>	121
5.5.4 <i>Análise das concordâncias</i>	122
5.5.5 <i>Identificação das combinações recorrentes nas concordâncias</i>	122
5.5.6 <i>Seleção das USEs e dos modificadores</i>	123
5.5.7 <i>Categorização das USEs e modificadores</i>	123
5.6 Arquitetura da base e inserção dos dados	123
5.6.1 <i>Organização dos relacionamentos</i>	125
5.6.2 <i>Preenchimento dos registros terminológicos</i>	125
6 ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS	126
7 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS	135
REFERÊNCIAS	143
APÊNDICE A	149
APÊNDICE B	150
ANEXO A	153
ANEXO B	159
ANEXO C	161
ANEXO D	162
ANEXO E	163

1 INTRODUÇÃO

Actualmente estamos inmersos en una sociedad en la cual se está produciendo no ya la revolución de la Informática, sino la revolución de la Información, y ello está suponiendo cambios profundos y radicales en la búsqueda de información previa necesaria tanto para la investigación como para la tomada de decisiones (AGUADO, 1995, p. 22).

Esta Dissertação visa propor uma metodologia que possa contribuir para a geração de bases de dados terminológicas no âmbito da análise documentária (AD). Para fins deste trabalho entende-se por base de dados terminológica um sistema de informações armazenadas em computador, representativo do conhecimento de uma área de especialidade e organizado segundo as relações de sentido que os conceitos mantêm entre si. (O tópico 4.7.3 retoma essa questão).

Dentre seus objetivos está o de fortalecer o processo de comunicação entre os Sistemas de Recuperação de Informação (SRIs) e a sua comunidade usuária, tendo o bibliotecário como mediador. Hoje se constata que essa comunicação se efetua de forma precária, pois os resultados obtidos no processo de busca e recuperação da informação nem sempre são representativos das necessidades dessa comunidade.

Acrescenta-se, a esses propósitos, a intenção de colocar à disposição do bibliotecário um referencial teórico e metodológico para ser aplicado na construção de novas bases terminológicas. A expectativa que se tem, no âmbito da informatização do conhecimento, é de que o impacto dos resultados dessa proposta se faça sentir sobre os pontos de vista teórico e prático.

Do ponto de vista teórico tem-se em vista oferecer um referencial que contemple as atividades desenvolvidas no campo da Documentação. Pretende-se, com a reflexão e a proposta aqui apresentadas, um alcance nos diferentes segmentos do processo de pesquisa informacional, que se inicia com o controle terminológico, passa pela indexação do documento e finaliza na busca e recuperação da informação. Nesse sentido, a intenção é propiciar melhores resultados à comunicação do conhecimento produzido pelos especialistas da área em estudo e a transferência da informação.

Do ponto de vista prático, dirige-se à construção de ferramentas no campo das Ciências Sociais e Humanidades (CSH), oferecendo, a título de projeto piloto, a geração de uma base de dados em um de seus domínios, a Linguística Teórica e Aplicada. Por observar-se, na literatura consultada, a falta de consenso quanto à classificação das áreas de conhecimento que compõem o grupo das CSH, considerou-se para as finalidades deste estudo

as seguintes áreas: Ciências Sociais (Antropologia, Política e Sociologia), Filosofia, História, Artes, Estudos da Linguagem e Literatura. Contribuiu decisivamente para esse posicionamento a experiência pessoal desta Bibliotecária como funcionária da Biblioteca Setorial de Ciências Sociais e Humanidades (BSCSH) da Universidade Federal do Rio Grande do Sul (UFRGS) no período de 1977 a 2003.

Em conseqüência, fez-se a opção pela exclusão, no conjunto daquelas ciências, das seguintes: Economia, Geografia, Educação, Comunicação, Administração, Estudos Ambientais e Arquitetura; justifica-se tal entendimento pelas peculiaridades das diferentes áreas. Não existe, por parte da autora deste trabalho, a convicção de que elas também possam ser beneficiadas pelos resultados desta pesquisa.

Não há consenso quanto à classificação das áreas de conhecimento, entretanto os estudiosos da ciência concordam que as exatas, da saúde e sociais se diferenciam umas das outras sob diferentes aspectos. Isso, tanto em relação à natureza de seus processos de pesquisa e de comunicação, como quanto às suas práticas profissionais, ao tipo de conhecimento produzido e à forma como o saber é estruturado e organizado.

Foi determinante para a escolha da área, além da experiência pessoal, o tipo de linguagem que as Ciências Sociais e Humanidades utilizam na divulgação do conhecimento produzido. Por ela não ser compacta e codificada como a empregada pelas Ciências Exatas e Biológicas, tal linguagem dificulta significativamente o desenvolvimento do processo de indexação. (O tópico 3.6 aborda essa questão). Enquanto essas últimas tomam o experimento científico como critério de cientificidade, as primeiras desenvolvem o saber científico a partir do conhecimento empírico. Velho (1997) salienta que o grau de autoria múltipla é muito mais alto nas Ciências Naturais do que nas Ciências Sociais e Humanidades, o que reflete uma outra particularidade dessa área, a do conflito em contraposição ao consenso por parte de grupos de pesquisa.

A relevância deste estudo aparece quando se constata os seguintes aspectos: a) escassez de ferramentas disponibilizadas ao bibliotecário para a indexação temática dos documentos em CSH; b) falta de especificidade temática oferecida por essas ferramentas e pelos catálogos eletrônicos nessas áreas de conhecimento; c) inconsistência no uso da Linguagem documentária (LD) utilizada pelo bibliotecário na construção dessas bases. Em estado da arte (tópico 1.2) tecem-se alguns comentários a respeito.

Decorrente desses três fatores e considerando a relação entre a informação que as bibliotecas armazenam e aquela que é oferecida para a recuperação em catálogos eletrônicos, fato já identificado na literatura produzida na área a esse respeito, compartilha-se, aqui, da idéia de que a recuperação da informação torna-se, sob o ponto de vista da precisão, um processo de baixa representatividade.

Em pesquisa efetuada na Internet no dia 11 de março de 2005 sobre a existência de vocabulários controlados, tendo como expressões de busca ESTUDOS DA LINGUAGEM e LINGÜÍSTICA, verificou-se a ausência de um instrumento no âmbito da análise documentária, que apresentasse uma terminologia específica e consistente. Dos vocabulários controlados existentes na área foco desta pesquisa, observou-se, inclusive, a ausência de uma política de atualização. Essa constatação, aliada ao saber adquirido por esta Bibliotecária, enquanto funcionária da BSCSH, com a construção de uma base de dados utilizando o *Microsoft Access* nesse campo do conhecimento, priorizou a seleção desse domínio para o desenvolvimento do projeto piloto aqui apresentado.

Por outro lado, olhando sob o viés do avanço tecnológico, há a considerar que a pesquisa documental apresenta resultados altamente positivos, como a informatização dos sistemas de bibliotecas e o acesso às informações de seus acervos por meio de redes e sistemas de informações. Nas bibliotecas, com os catálogos eletrônicos, grande volume de informação circula em velocidades fantásticas por diferentes e distantes espaços geográficos.

Como resultado da informatização de seus acervos registram-se as mudanças que ocorreram no processo de busca e recuperação da informação, a possibilidade de intercâmbio de dados bibliográficos *on-line* entre os diversos sistemas de informação, e a integração de diferentes tipos de coleções documentais e de diversas áreas do conhecimento em um mesmo catálogo eletrônico, constituindo-se em um sistema de informações conectadas. A esse respeito, ressalta-se que tal recurso não era oferecido pelas antigas formas de armazenamento da informação, como os catálogos impressos.

O usuário, nesse contexto, passa a ter maior autonomia em virtude das novas formas de identificar, localizar e obter informações e documentos. Entretanto, essa modalidade de pesquisa amplia significativamente os problemas relacionados aos SRIs, às linguagens documentárias e ao processo de indexação.

A partir dessas considerações e das reflexões desenvolvidas no decorrer desta Dissertação, espera-se contribuir para a geração de bases de dados terminológicas e o

desenvolvimento de catálogos eletrônicos mais eficazes. Desse modo, sob um novo paradigma, se deseja oferecer uma comunicação facilitada entre os SRIs e os usuários, pelo acesso ao conjunto de informações produzidas em sua respectiva área de conhecimento nos diferentes níveis de especialização do saber, de forma rápida, precisa e consistente.

A procura de subsídios para a identificação de princípios, que sustentem um referencial teórico e que ofereçam uma metodologia mais eficiente, tem sido motivo de estudos e discussões em diferentes áreas de conhecimento, particularmente, no que se refere aos processos de indexação, busca e recuperação da informação. Os avanços da Informática, associados à rápida evolução e disseminação do conhecimento técnico e científico e às tendências para a interdisciplinaridade são razões significativas para que proliferem reflexões na área biblioteconômica.

A partir de uma linguagem documentária mais flexível e adequada às necessidades de pesquisa do usuário e de uma nova forma de organização do conhecimento, constata-se que a Documentação enfrenta o grande desafio de construir um Sistema de Recuperação da Informação. Sob um novo paradigma, em que o conhecimento não se limita mais a uma estruturação hierárquica, pois a interdisciplinaridade e a multidisciplinaridade tornaram-se uma presença marcante, faz-se necessário ir ao encontro dessa nova realidade que se apresenta e que está se caracterizando como a sociedade da informação e do conhecimento.

Conforme nos afirma Aguado (1995), o grande desafio a ser enfrentado em relação ao acesso à informação conduz a uma situação bastante peculiar, pois se relaciona ao excesso de oferta que os sistemas oferecem. Assim sendo, não decorre mais da dificuldade em localizar informações pertinentes, mas da necessidade de examinar, dentre um conjunto imenso de informações disponíveis, aquelas que são realmente relevantes para o conhecimento que se deseja construir.

Como conseqüência, a insatisfação do usuário revela-se não apenas pelas dificuldades que ele enfrenta devido à falta de precisão nos resultados obtidos no momento da recuperação da informação, como também pelo pouco conhecimento da ferramenta que utiliza, especialmente quanto à linguagem que emprega e aos recursos que oferece. Nesse novo panorama observa-se que o usuário de hoje, mesmo que preparado para o uso dos recursos de Informática, não domina as linguagens de indexação, o que o impede de obter do sistema todo o potencial que está à sua disposição.

A incompatibilidade existente entre a linguagem que os Sistemas de Informação fazem uso com a linguagem de especialidade empregada pelos pesquisadores nos textos publicados e a linguagem expressa pelo usuário é um fato que exige reflexões e um trabalho a desenvolver. Mesmo consciente de que isso é apenas um dos pontos que afeta a eficácia dos SRIs em meios eletrônicos, o enfoque desta proposta restringe-se à geração de bases de dados terminológicas no âmbito da análise documentária, com o propósito de minimizar as falhas que esses sistemas vêm apresentando.

Reflexões nesse sentido vêm sendo feitas por esta Mestranda em decorrência de sua experiência profissional. Foi dessa inquietação que nasceu a motivação para a realização do presente trabalho. Com o intuito de enfrentar esse desafio considerou-se necessário encontrar um referencial teórico que pudesse conduzir para uma proposta metodológica a ser aplicada na geração de bases de dados terminológicas, visando oferecer maior especificidade temática e consistência no uso de descritores/termos, que são unidades de informação representativas do conteúdo temático de um documento, atuando como signos de comunicação. Acredita-se que tal propósito possa ser alcançado na medida em que a comunicação for estabelecida entre os SRIs e a comunidade usuária, que tanto pode ser constituída por pesquisadores, como por alunos de cursos de pós-graduação, de graduação, ou alguém que simplesmente busque informações sobre uma determinada área do conhecimento.

1.1 Antecedentes

A motivação que levou esta Mestranda ao desenvolvimento desta pesquisa é o resultado de mais de 26 (vinte e seis) anos de vivência nesta Universidade (UFRGS). Seu trabalho sempre esteve ligado às tarefas de classificação e indexação de documentos. Quando ainda eram utilizados os catálogos manuais, apesar de se observar que a forma de acesso preferida pela comunidade acadêmica era pela autoria e pelo título do documento, já se percebia a dificuldade em recuperar um documento pelo assunto. Naquela época a indexação temática já se mostrava como uma das tarefas mais complexas desenvolvidas pelo bibliotecário.

Tal realidade era constatada até mesmo em áreas em que o bibliotecário tinha a seu dispor vocabulários especializados, pois a linguagem utilizada não era do conhecimento do usuário. Nas áreas das CSH essa situação sempre se apresentou mais complexa pela falta de vocabulários apropriados. Com a informatização dos catálogos e da crescente especialização do conhecimento, o problema se agravou ainda mais.

Acrescenta-se, ao acima exposto, o fato de que em pesquisa no Prossiga, no *site* Bases de dados Brasileiras na Internet¹, não foi identificada nenhuma base de dados com as expressões de busca: ESTUDOS DA LINGUAGEM e LINGÜÍSTICA como a que aqui se vai propor.

Essa motivação teve origem também na insatisfação do usuário quando não localizava, nos catálogos de assunto daquela Biblioteca, temas específicos. Assim, a partir da idéia de disponibilizar ao usuário o acesso às informações temáticas específicas, foram desenvolvidos alguns projetos que tiveram o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

Com o propósito de compatibilizar a linguagem utilizada para representar o conteúdo temático de um documento e aquela que era usada como ponto de acesso pelo usuário, esses projetos privilegiaram a indexação dos artigos de periódicos, por representarem a atualização do conhecimento e a especificidade temática almejada pelo usuário. Os projetos iniciais selecionaram periódicos de duas áreas de conhecimento: ESTUDOS DA LINGUAGEM e LITERATURA. A prioridade atribuída a esses domínios deveu-se, em muito, ao fato desta Bibliotecária ser lotada no Instituto de Letras e ter uma familiaridade maior com essas áreas, o que facilitaria na construção de uma nova ferramenta.

Pela falta de instrumentos adequados para a indexação de documentos nessas áreas, esses projetos evidenciaram a necessidade de criar ferramentas que auxiliassem o bibliotecário em suas tarefas, oferecendo-lhes uma linguagem específica e consistente. Verificou-se, então, a importância em desenvolver bases de dados que auxiliassem o bibliotecário no controle dos termos utilizados como descritor/termo.

Inicialmente essas bases de dados utilizaram como *software* o *DBase*. Assim que equipamentos mais modernos foram adquiridos, mudou-se para o *software Access*, por serem mais amigável e oferecer uma maior produtividade. Como fruto desses projetos foram implementadas bases de dados nas áreas de ESTUDOS DA LINGUAGEM e LITERATURA. A experiência adquirida com o desenvolvimento desses projetos foi sempre muito valiosa e instigadora, seja pelos contatos mantidos com os coordenadores, sempre um pesquisador da área, como com os bolsistas, alunos do Curso de Letras, que revelavam as suas necessidades de pesquisa e as da comunidade que representavam.

¹ Endereço eletrônico: <<http://www5.prossiga.br/basesdedados/asp/buscaavancada.asp>>.

Com a perda constante de funcionários a responsabilidade desta Bibliotecária, como indexadora, estendeu-se para todas as áreas de conhecimento da Biblioteca. Com a informatização do acervo as tarefas de indexação tornaram-se ainda mais complexas, exigindo um controle maior no âmbito da consistência temática. Assim sendo, deu-se início a uma nova base de dados muito mais abrangente, abarcando todas as áreas das CSH.

Com os avanços oriundos da Informática a maioria das bibliotecas universitárias já tem seus documentos processados em catálogos eletrônicos. Em consulta feita a esses catálogos, como Mestranda do Programa de Pós-Graduação do Instituto de Letras, foi constatado que a falta de especificidade dada ao conteúdo temático de um documento e a inconsistência entre os descritores/termos utilizados permanecem ainda como graves obstáculos à recuperação da informação. Acresce uma outra dificuldade a essa situação que tem origem no excesso de informações que tais catálogos oferecem, propiciando alta revocação e uma baixa precisão, ou melhor dito, uma recuperação com um número excessivo de documentos, dos quais poucos são relevantes para responder à consulta do usuário.

1.2 Estado da arte

Com o propósito de confirmar ou não a necessidade de desenvolver estudos e propor metodologias no âmbito dos Sistemas de Recuperação da Informação, e assim observar as condições de pesquisa da comunidade acadêmica e o comportamento dos procedimentos de busca e recuperação da informação em catálogos eletrônicos, procedeu-se a algumas pesquisas.

Selecionando Pesquisar palavras no campo assunto e utilizando a opção palavras adjacentes (sim), obteve-se o resultado de 2.651 registros de documentos para a expressão de busca ANÁLISE DO DISCURSO. Do resultado obtido constatou-se a preservação do conceito, uma vez que a indexação foi feita pelo termo que o representa. Entretanto, o resultado não ofereceu a precisão necessária quanto à recuperação de ANÁLISE DO DISCURSO enquanto teoria, verificando-se a necessidade de especificar o termo indexado para se obter a precisão desejada. Nesse caso, uma das possibilidades que se apresenta é a de contextualizar a expressão de busca, utilizando o modificador² TEORIA ou PRÁTICA, por exemplo; outra possibilidade consistiria em acrescentar um outro descritor/termo: TEORIA DO DISCURSO.

² Nesse trabalho entende-se modificador conforme Bräscher (1999, p. 8). Ver tópico 5.3.4.7 Categorização das USES e dos modificadores.

Fazendo uso de recursos semelhantes, torna-se possível distinguir os textos que tratam da teoria daqueles que tratam da prática em ANÁLISE DO DISCURSO.

Em um catálogo *on-line* consultado constatou-se que foram indexadas pelo descritor/termo DISCURSO (ANÁLISE), as obras: <Análise do discurso: entornos do sentido> e <A análise do discurso: uma revisão teórica> No entanto, nessa mesma base, o descritor/termo ANÁLISE DO DISCURSO foi usado para indexar obras sobre o mesmo assunto. Tal fato revelou duas situações: a de inconsistência pelo uso de dois descritores/termos para um mesmo conceito e a de uma indexação em que não houve a preocupação de preservar o conceito, pois foi feita a partir de palavras e não de termos.

Ao fazer pesquisa pelo título, tendo como expressão de busca ANÁLISE DO DISCURSO e utilizando a opção Palavras adjacentes (sim), obteve-se um resultado de 111 registros. Utilizando a opção Multicampo em Pesquisar palavras e preenchendo os campos de título e assunto com a mesma expressão ANÁLISE DO DISCURSO, obteve-se um resultado de 97 registros. Ao identificar os registros que omitiram essa informação, enquanto descritor/termo como ponto de acesso aos documentos, observaram-se situações diferenciadas: o texto foi indexado sobre outros aspectos, mas não sob ANÁLISE DO DISCURSO; a indexação foi feita em um nível mais genérico, utilizando descritores/termos como LINGÜÍSTICA e LINGUAGEM E LÍNGUAS; em casos em que o texto tem como fundamentação teórica a ANÁLISE DO DISCURSO, constatou-se falha nos procedimentos de indexação pela omissão desse descritor/termo como ponto de acesso.

Escolhendo a alternativa Pesquisar palavras no campo assunto e selecionando a opção palavras adjacentes (sim), obteve-se o total de 20 registros como resultados de busca sobre gramáticas da língua portuguesa. Dentre os resultados que se obteve, utilizando como expressão de busca PORTUGUÊS GRAMÁTICA observou-se o uso de diferentes descritores/termos: GRAMÁTICA : PORTUGUÊS; PORTUGUÊS : GRAMÁTICA (mesmo descritor/termo, mas na forma invertida); LÍNGUA PORTUGUESA : GRAMÁTICA; LÍNGUA PORTUGUESA e GRAMÁTICA em campos diferentes; DICIONÁRIOS : PORTUGUÊS : GRAMÁTICA. Registra-se, aqui, mais uma vez, um caso de falta de consistência pois foram atribuídos dois descritores/termos para um mesmo conceito: PORTUGUÊS e LÍNGUA PORTUGUESA.

Entende-se por consistência temática o controle sobre a terminologia usada na indexação de um documento. Ela não se opõe ao princípio da variação, pois quando

necessário indicar o uso de variantes para representar um mesmo conceito, o recurso de remissivas deve ser empregado em uma linguagem controlada, .

Em uma pesquisa utilizando a expressão de busca LÍNGUA PORTUGUESA e GRAMÁTICA, a opção Pesquisar palavras no campo assunto e palavras adjacentes (sim) foram encontrados documentos como: <Todo o mundo tem dúvida, inclusive você>. Essa obra não se constitui em uma gramática, embora trate das dificuldades da língua portuguesa no âmbito gramatical. Sob essa mesma concepção encontram-se, igualmente, vários textos que foram indexados pelo descritor/termo LÍNGUA PORTUGUESA : GRAMÁTICA mas que tratam apenas de aspectos gramaticais. Nesses casos o uso de um modificador, conforme entendido nesse trabalho, daria uma maior precisão na linguagem de indexação e, conseqüentemente, na recuperação da informação, podendo ser assim representado: ASPECTOS GRAMATICAIS ou QUESTÕES GRAMATICAIS

Em pesquisa realizada em duas bibliotecas de um mesmo catálogo eletrônico percebeu-se que o documento <Novo guia ortográfico> teve o descritor/termo ORTOGRAFIA omitido na verbalização de assuntos de uma das bibliotecas, pois foi utilizado como ponto de acesso ao documento os seguintes descritores/termos: LÍNGUA PORTUGUESA : GRAMÁTICA, PORTUGUÊS : GRAMÁTICA e GRAMÁTICA : PORTUGUÊS. Comprova-se, nesse caso, a falta de especificidade temática e o impedimento da recuperação deste documento em uma busca cuja expressão fosse ORTOGRAFIA e LÍNGUA PORTUGUESA, registrando-se, neste caso, uma situação de silêncio do SRI. O descritor/termo apresentado pela outra biblioteca foi LÍNGUA PORTUGUESA : GRAMÁTICA : ORTOGRAFIA. Na comparação entre as duas bibliotecas constatou-se que a falta de consistência se apresenta em nível de sistema.

Os resultados dessas pesquisas constatarem que os Sistemas de Recuperação da Informação, assim construídos, estão se mostrando ineficientes por diferentes motivos: a) por não contextualizarem um determinado descritor/termo ou não preservarem o conceito expresso em um documento, b) por tratarem a informação em um nível genérico e assim não oferecer um descritor/termo específico como ponto de acesso, c) por não respeitarem o princípio da univocidade, que defende a idéia de uma única forma para representar um mesmo conceito.

Considerando que são os vocabulários controlados que dão sustentação às atividades de indexação e asseguram a eficácia dos Sistemas de Recuperação da Informação, é preciso considerar que tais vocabulários devem possibilitar a seleção de descritores/termos que sejam, simultaneamente, unidades de representação do conhecimento expresso pelo autor e

expressões de busca da área de abrangência temática do vocabulário controlado. Assim, é preciso que, na sua construção, seja dado um tratamento adequado à seleção dos descritores/termos, de modo a possibilitar a construção de catálogos eletrônicos com informações consistentes e específicas. Um outro ponto importante a considerar é que um vocabulário controlado deve ser representativo da rede conceitual de um domínio de especialidade.

Muitos dos instrumentos utilizados na indexação de documentos, nas áreas das CSH, apresentam uma extensa lacuna terminológica em relação ao universo temático de sua área de especialidade, notadamente quanto à especificidade. A esse respeito faz-se referência à: EUROVOC, Library of Congress Classification Outline, OECD Macrothesaurus, UNESCO Thesaurus: hierarchical list, Vocabulário controlado do SIBi/USP. Infelizmente, não foi possível conhecer o *Thesaurus of Linguistic Indexing Terms* utilizado pela base de dados LLBA – *Linguistic and Language Behavior Abstracts* –, por ser de acesso restrito.

Diferentes estudos já constataram que uma informação indexada apenas pelo descritor/termo genérico inviabiliza a recuperação de um documento e de uma temática ao nível de sua especificidade. Nesse caso, a indexação temática deixa de ser representativa das necessidades de pesquisa daquele usuário que precisa recuperar informações de natureza específica.

A gravidade decorrente desses vazios terminológicos³ é atestada pelo fato de que a exclusão de um descritor/termo, em um vocabulário controlado, inviabiliza a representação e a conseqüente recuperação de uma determinada informação em um Sistema. Tais vazios ocorrem, prioritariamente, por duas razões: um procedimento inadequado em relação à coleta dos candidatos a descritores/termos e a falta de uma constante avaliação e atualização do vocabulário controlado.

Em face da necessidade de se oferecer um novo modelo de pesquisa, imposto pela informatização da informação documental armazenada nas bibliotecas e pela nova forma como o conhecimento se organiza, não apenas a partir de relações hierárquicas, mas também de relações interdisciplinares, multidisciplinares e transdisciplinares, torna-se necessário reavaliar os procedimentos que vêm sendo adotados no processo de representação, busca e recuperação da informação. Tal modelo deve permitir ao usuário total independência para

³ Vazio terminológico, conforme mencionado no tópico 3.7, Sistemas de Recuperação da Informação, refere-se a ausência de uma unidade de informação como ponto de acesso de um sistema.

consultar os catálogos eletrônicos e buscar as informações de que necessita, dispensando, se desejar, a presença do bibliotecário.

De maneira especial torna-se preciso rever os procedimentos adotados nos processos de indexação e de busca e recuperação da informação temática, que são aqueles que apresentam o maior número de dificuldades, como atestam os estudos realizados nesse sentido. Os problemas parecem ter origem na diversidade com que a informação vem sendo tratada pelos bibliotecários das diferentes bibliotecas em relação às diferentes áreas do conhecimento.

A partir de observações como essas, torna-se urgente desenvolver estudos com o objetivo de adequar a linguagem documentária a ser utilizada em catálogos eletrônicos de assunto com aquela que o autor utiliza em seus textos e o usuário em suas expressões de busca. Dessa forma, torna-se necessário oferecer ao indexador uma ferramenta de trabalho que o torne capaz de criar condições favoráveis para a comunicação a ser estabelecida entre os SRIs e os usuários, e assim buscar a almejada univocidade terminológica.

Nesse direcionamento esta Dissertação tem a pretensão de propor uma metodologia, na forma de projeto piloto, para a geração de bases de dados representativas de uma área de especialidade, procurando atingir alto grau de especialização e adequado equilíbrio entre o índice de revocação e o de precisão. Tais conceitos são abordados no tópico 3.7.

Reforçando ainda mais a motivação para o trabalho aqui desenvolvido, destaca-se que a consulta aos catálogos eletrônicos tem na recuperação de assuntos uma das formas mais procuradas pela comunidade acadêmica. Tal fato é atestado pela bibliografia especializada e constatado pela própria experiência da Mestranda em atividades exercidas em biblioteca universitária.

É do conhecimento daqueles que trabalham em bibliotecas que têm acervo específico na área de ESTUDOS DA LINGUAGEM ou LINGÜÍSTICA TEÓRICA E APLICADA, a dificuldade em acessar um vocabulário controlado que ofereça ao indexador a possibilidade de indexar os documentos, fazendo uso de uma terminologia específica, uniforme e consistente. Tal situação faz com que cada indexador utilize um vocabulário diversificado, o que leva os usuários a fazerem consultas terminologicamente diferenciadas às bases e a obterem resultados de baixa produtividade.

Para fins desta pesquisa, considera-se que a terminologia de uma área especializada se evidencia nos resumos que os próprios autores elaboram para a divulgação de seus trabalhos.

As palavras-chave, por sua vez, sintetizam o conhecimento produzido e se propõem a ser o elo de comunicação entre quem produz o saber e quem dele faz uso.

Reforçando esse entendimento vale mencionar Holzem e Wable (2001). Da leitura de seus textos, interessou para as finalidades desse estudo a idéia de que nos resumos produzidos pelos autores de um trabalho científico, no caso por ela citado, o das teses, é possível localizar a terminologia que veicula em seu contexto de uso.

Sobre os resumos, Holzem (1998) os percebe como um documento secundário para um público não especializado, mas que mesmo contendo uma informação secundária é fundamental para a seleção da informação especializada. Ela ressalta que esse tipo de texto não é destinado a ser avaliado entre os seus pares, mas que cumpre uma exigência, no caso das teses, institucional; ele se dirige a um público amplo e a sua difusão se dá por meio de bases de dados.

Por cumprir a determinados propósitos o resumo apresenta regularidades em sua estrutura, o que o caracteriza como um tipo de texto de divulgação. Wable e Holzem (2004⁴) assim se expressam a respeito:

A partir des régularités observées au niveau de sa macrostructure et d'une analyse linguistique de la position énonciative du locuteur, nous poserons l'hypothèse d'un genre textuel particulièrement intéressant pour le travail terminologique au sens ou il offre un cadre de description relativement méthodique de l'avancée des connaissances scientifiques.⁵

Embora as intenções do trabalho desenvolvido por Holzem e Wable (2001) e este não sejam as mesmas, pois o interesse dela está em registrar os neologismos, há a considerar em comum o fato de ambos se proporem a assegurar a veiculação da terminologia realmente em uso produzida em meios acadêmicos.

Com a preocupação em registrar a terminologia produzida pelas comunidades acadêmicas e assim registrar o avanço do conhecimento científico em língua francesa, Holzem e Wable (2001) propõem um modelo de ficha terminológica para o registro de informações a partir dos seguintes campos: termo, termo em inglês, sigla, termos genéricos, termos específicos, termos relacionados, autores relacionados e termos sinônimos.

⁴ Documento eletrônico.

⁵ Tradução da autora: A partir das regularidades observadas em nível de sua macroestrutura e de uma análise lingüística da posição enunciativa do locutor, nós levantamos a hipótese de um gênero textual particularmente interessante para o trabalho terminológico, no sentido em que ele oferece um quadro de descrição relativamente metódico do avanço dos conhecimentos científicos.

Para os propósitos da pesquisa aqui desenvolvida chamou a atenção desta Mestranda o uso de: termo em inglês e autores relacionados. A opção por incluí-los na estrutura da base terminológica sugerida nesta Dissertação teve como razões os seguintes fatos: a ferramenta proposta no âmbito da análise documentária passou a ter um alcance internacional e, a presença de autores relacionados, nesse caso, com as teorias de uma área, ampliam as possibilidades de indexar um documento e de contextualizá-lo com os teóricos que a representam. Esses dois aspectos por ela considerados, embora já se mostrassem sugestivos pela análise do *corpus*, auxiliaram na decisão tomada. a partir da leitura de seu texto.

1.3 Hipóteses de trabalho

A informação veiculada pelos catálogos eletrônicos vigentes não proporciona resultados precisos e adequados às necessidades dos usuários, quanto à consistência dos dados e a especificidade temática da área de pesquisa.

As ferramentas disponibilizadas ao bibliotecário para a indexação de documentos, como tesouros e listas de cabeçalhos de assunto, não privilegiam a linguagem utilizada pelos especialistas.

A Lingüística de *Corpus* e a Terminografia se apresentam como um novo recurso alternativo a ser utilizado na geração de bases de dados terminológicas.

Bases de dados terminológicas construídas a partir dos resumos elaborados pelos autores dos documentos constituem-se em uma ferramenta adequada à interface SRI/usuários.

1.4 Objetivo geral

Propor uma metodologia para a geração de bases de dados terminológicas, em formato eletrônico, nas áreas de conhecimento das Ciências Sociais e Humanidades.

1.5 Objetivos específicos

Assegurar a efetiva comunicação entre os Sistemas de Recuperação de Informação e os usuários, tendo o bibliotecário como mediador do processo de comunicação que se dá entre o autor do texto e o usuário.

Identificar a terminologia de uma área de especialidade conforme verbalizada pelo autor do documento e entendida pela Terminologia⁶.

Registrar a terminologia de um domínio em bases de dados terminológicas.

Viabilizar o uso de uma linguagem consistente, precisa e específica para uso em catálogos eletrônicos.

1.6 Público-alvo

Esta Dissertação considera como o seu público-alvo prioritário o bibliotecário documentalista das bibliotecas universitárias, mas tem no usuário de redes e sistemas de informação a motivação da pesquisa. A intenção é aumentar o grau de satisfação desse usuário, propiciando-lhe resultados de busca e recuperação de informações representativos de seus desejos e necessidades.

1.7 Estrutura da Dissertação

Os pressupostos teóricos que fundamentam esta Dissertação se encontram em três áreas do conhecimento: Documentação, Terminologia e Lingüística. Nos princípios básicos de tais áreas do saber, que têm em comum a linguagem como objeto de estudo, procura-se o apoio necessário para sustentar a proposta de uma metodologia que torne mais eficiente a busca e a recuperação da informação. Em uma perspectiva mais ampla, de caráter multidisciplinar e transdisciplinar, busca-se também os subsídios da Informática, da Lingüística de *Corpus* e da Terminografia.

Ela é estruturada em capítulos e em tópicos e compõe-se de um Quadro referencial, de uma Metodologia, Análise dos dados e Discussão dos Resultados e, finalizando, apresenta as Considerações finais e Perspectivas futuras. Na Introdução apresenta em linhas gerais: a proposta de trabalho desenvolvida, os objetivos, o público a quem se destina, as hipóteses de trabalho, a motivação que deu origem a este trabalho e a situação em que se encontram os Sistemas de Recuperação da Informação, no contexto da pesquisa acadêmica.

O capítulo que focaliza o estudo da Linguagem e do Sentido discute os aspectos relevantes da teoria lingüística, principalmente no que diz respeito ao estudo do significado e de suas representações. Seu entendimento é essencial para que se estabeleça a interface com

⁶ Neste trabalho entende-se que Terminologia grafada com T maiúsculo refere-se à disciplina e com t minúsculo ao conjunto de termos de uma área.

as áreas de saber da Documentação e da Terminologia. O capítulo sobre a Documentação se ocupa de temas e noções que interferem na construção dos SRIs, nas linguagens documentárias e no processo de indexação.

O capítulo, dedicado à Terminologia apresenta um breve percurso histórico e enfatiza alguns conceitos como termo e fraseologia. Faz referência a duas teorias de cunho lingüístico textual: a) a Teoria Comunicativa da Terminologia (TCT), desenvolvida por Teresa Cabré e o grupo IULATERM, Grupo de Terminologia do IULA (Institut Universitari de Lingüística Aplicada) e b) a Socioterminologia, que teve em François Gaudin um de seus principais representantes. Apresenta também a contribuição da área da Informática para a Terminologia.

Para aplicar o referencial teórico desenvolvido nos primeiros capítulos desta Dissertação, este estudo propôs uma metodologia na forma de um projeto piloto, na área da LINGÜÍSTICA TEÓRICA E APLICADA, também entendida por muitos como ESTUDOS DA LINGUAGEM. Esta proposta implica em: a) constituição de um *corpus* representativo dessa área de conhecimento, b) identificação e registro da terminologia utilizada, pelos pesquisadores, em resumos de diferentes tipos de textos (periódicos, teses e dissertações e trabalhos apresentados em congressos e seminários) e, c) tratamento terminográfico dos dados, com vistas à geração de bases de dados terminológicas no âmbito das CSH.

A fim de comparar a terminologia coletada nos resumos dos pesquisadores com aquela empregada pelos bibliotecários e expressa pelos descritores/termos disponibilizados em Redes e Sistemas de Informação, procedeu-se a uma consulta nos seguintes catálogos eletrônicos SABI/UFRGS, PUCRS, UNICAMP, USP – e internacionais: LIBRARY OF CONGRESS (LC) e BIBLIOTHÈQUE ET ARCHIVES CANADÁ (AMICUS). A partir de então se procedeu à Análise dos dados e Discussão dos Resultados, apresentando-se uma síntese dos resultados mais significativos.

Em Considerações finais são destacados os fatores considerados relevantes para propor uma aproximação entre Documentação, Terminologia e Lingüística, ressaltando também a importância dos conhecimentos advindos da Lingüística de *Corpus* quando se pensa em organizar e registrar o conhecimento de um determinado domínio e, assim, construir bases de dados terminológicas com fins de análise documentária. E, com a intenção de dar continuidade ao trabalho aqui desenvolvido, apresentam-se algumas possibilidades de estudos a serem desenvolvidos futuramente.

Em Apêndice apresenta-se uma lista contendo os endereços eletrônicos das bases de dados consultadas e uma pequena amostra da construção da base de dados terminológica. Em Anexo, uma amostra das listas de palavras que constituíram o *corpus*, conforme apresentadas pelo *software WordSmith Tools* e sob diferentes formas de apresentação.

2 LINGUAGEM E SENTIDO

2.1 Introdução

Este capítulo é dedicado ao estudo da linguagem e do sentido, a partir dos conhecimentos oriundos da Lingüística e da Semântica, visando estabelecer uma interface com a Documentação e, dessa forma, estabelecer algumas aproximações com a linguagem documentária.

A Lingüística foi, dentre as Ciências Sociais e Humanidades, uma das primeiras a se constituir como ciência, com método e objeto próprios e bem definidos. Por se ocupar com o estudo científico da língua e se propor a responder perguntas sobre o funcionamento da linguagem interage com diversas áreas do conhecimento, tais como a Antropologia, a Sociologia, a Psicologia e a Lógica, dentre outras. Fruto dessa interação surge a Sociolingüística e a Psicolingüística. Com o advento da Informática a Lingüística também se faz presente nessa área, haja vista o desenvolvimento de campos de atividade e pesquisa em que as duas se entrelaçam, como a Lingüística Computacional e a Lingüística de *Corpus*.

A Terminologia, zona fronteira que perpassa pelas ciências humanas, exatas e biológicas, conforme já afirmava Wüster, o engenheiro austríaco que lançou os fundamentos do que viria a se constituir num campo de investigação cada vez mais dinâmico na Lingüística contemporânea, também mantém fortes vínculos com a Lingüística.

Nos últimos anos tem sido cada vez maior o inter-relacionamento de outras disciplinas com os Estudos da Linguagem, o que contribui para que a Lingüística assuma um lugar de destaque, não apenas entre especialistas da área, como também das áreas afins. As questões discutidas por essa ciência não interessam apenas aos lingüistas, exceto quando vistas em um sentido muito restrito, pois a área, inicialmente delimitada por Saussure, se expande cada vez mais e se relaciona com outros domínios sem uma relação de exclusividade com a linguagem verbal.

A aproximação da Lingüística com a Documentação é hoje uma realidade, pois ambas estão envolvidas com as questões que dizem respeito ao funcionamento da linguagem; sendo a língua o objeto de estudo da Lingüística e a informação, o da Documentação. Esse interesse mútuo decorre das várias dimensões da operação da língua em relação à construção do saber e do saber-fazer lingüístico, em especial quanto ao conhecimento expresso por meio da

linguagem e ao processo de construção do sentido. Por essa razão é importante desenvolver um estudo que coloque em destaque as relações que se estabelecem entre o conhecimento que tem origem nessas duas áreas, de modo a permitir a compreensão dos fenômenos lingüísticos que estão presentes na análise documentária.

Desse modo, nada mais natural para um profissional da Biblioteconomia do que o interesse pelas palavras, pelo seu uso e, sobretudo, pelo que contribui para a formação do sentido dentro do contexto da Informação e da Documentação. Conhecer os mecanismos de funcionamento da linguagem para melhor compreender aqueles que são utilizados na estruturação de uma linguagem documentária, enfrentar e resolver as dificuldades que se apresentam no trabalho documentário, tanto no processo de indexação temática do documento quanto no de busca e recuperação da informação, são focos de interesse desse profissional.

Ressalta-se que a preservação do sentido é uma das grandes dificuldades que se apresentam no âmbito da Documentação. Isso torna necessária a adequação entre a linguagem que representa um domínio especializado e a linguagem documentária, de modo que os Sistemas de Recuperação da Informação propiciem a circulação do saber com eficácia.

2.2 Linguagem

Nesta reflexão sobre a linguagem é focalizada a língua e a fala, concebidas como um sistema e, como tal, organizadas em uma estrutura. Este entendimento se faz necessário para dar embasamento teórico ao capítulo seguinte, que se concentra na linguagem documentária (tópico 3.5), também vista como um sistema estrutural.

O lingüista suíço Saussure foi quem pela primeira vez demonstrou que uma língua não é apenas uma coleção de objetos lingüísticos como sons e palavras; ao contrário, é um sistema altamente estruturado, onde cada elemento se define em grande medida pela maneira como está relacionado com outros elementos. A visão *estruturalista* da língua dominou desde então o pensamento lingüístico (TRASK, 2004, p. 222-223).

A obra de Saussure marcou profundamente o século XX, sobretudo pelo que permitiu em relação às discussões futuras, uma vez que suas idéias provocaram uma verdadeira revolução na história da Lingüística. Pode-se afirmar que, praticamente, nenhuma área do conhecimento permaneceu imune à influência da teoria saussuriana, abrindo-se para novos pontos de vista, descortinando outros desdobramentos teóricos, ampliando horizontes e inovando metodologias. Até hoje são desenvolvidos estudos a partir da sua obra, seja para criticá-la, seja para colocar-se em sua defesa, ou até mesmo para criar novas teorias.

Um dos princípios fundamentais que sustentam a Lingüística Moderna foi estabelecido por Saussure, ao argumentar que “a língua é um sistema de signos que exprime idéias” (SAUSSURE, 1988, p. 24). Expressa por um código, que permite estabelecer a comunicação entre um emissor e um receptor, ela tem seu funcionamento regulado por certo número de regras e de restrições. O reconhecimento do conceito de signo lingüístico é fundamental para que se possa entender o funcionamento de qualquer linguagem como um sistema simbólico e estruturado.

Para este trabalho, tal entendimento se faz necessário para a compreensão do signo documentário. Na linguagem documentária ele é utilizado para representar a informação contida nos documentos, visando a sua recuperação pelo usuário dos Sistemas de Informação e, à semelhança da linguagem natural, consolidar o processo de comunicação entre um SRI e os seus usuários (LIMA, 1998). Conforme Lopes (1997), Saussure afirma, com extrema modernidade, que a língua só funciona como instrumento de comunicação quando se constitui, também, como construção do saber a ser comunicado.

Um conceito que igualmente se deve a Saussure, e representa uma importante contribuição teórica, é o do signo lingüístico dotado simultaneamente de forma e sentido. Para ele, o signo é uma entidade psíquica de duas faces interdependentes e inseparáveis, significante e significado, intimamente unidas, de tal modo que um reclama o outro (SAUSSURE, 1988).

Destacam-se como relevantes para este trabalho duas características do signo lingüístico: a arbitrariedade e a linearidade. Sobre a arbitrariedade, o lingüista suíço afirma que não há nenhuma razão para que qualquer forma lingüística seja associada a um significado, particularmente. Em relação ao caráter linear do significante o lingüista entende que, sendo de natureza auditiva, ele se desenvolve na cadeia do tempo e, como tal, adquire as características próprias do tempo, representando uma extensão que é mensurável em uma só dimensão. Com base nessa noção é possível afirmar que as palavras estabelecem relações entre si em um discurso, fato que exclui a possibilidade de se pronunciarem dois elementos ao mesmo tempo, uma vez que eles se alinham um após outro na cadeia da fala.

Tais combinações, apoiadas na extensão, são chamadas de sintagmas ou relações sintagmáticas, que existem *in praesentia*, pois repousam em duas ou mais palavras presentes numa série efetiva. Elas são entendidas como aquelas relações que unem os elementos da língua sob o ponto de vista da sucessividade, ou seja, da ordem linear da cadeia falada, resultando sempre de duas ou mais unidades consecutivas, como por exemplo, re-ler, contra

todos e vida humana. Assim, pode-se dizer que a palavra em um sintagma só adquire valor porque se opõe ao que a precede e/ou ao que a segue; portanto, ao situar-se em contextos diferentes, poderá assumir sentido diverso daquele que lhe é dado enquanto signo isolado.

Sob esse prisma, no sistema da língua nada existe senão diferenças. Todo o mecanismo da língua se articula não somente em relações sintagmáticas como também em relações associativas. Tais relações resultam da língua interior de cada indivíduo e da união de palavras *in absentia* em uma série mnemônica virtual. Assim, uma palavra qualquer está sujeita a inúmeras associações, que não se apresentam nem em número definido nem em uma ordem determinada.

Os processos de combinação e de seleção, que ocorrem respectivamente no eixo sintagmático e paradigmático, situam-se nos diferentes níveis de análise da língua. Desse modo, as unidades da língua dependem do que as rodeia na cadeia falada ou das partes sucessivas de que elas próprias se compõem, como resultado da combinação de dois elementos solidários. Tal dependência leva à formulação do princípio da solidariedade sintagmática, em que o todo vale pelas partes e as partes só têm valor em virtude de seu lugar no todo.

Em resumo, pode-se dizer que, ao depreender as noções de sistema, de unidades lingüísticas e de valores, Saussure lançou as bases de um estudo estrutural da língua, do qual derivam regras sobre as quais repousam os processos de escolha e de combinação que conformam as relações sintagmáticas e paradigmáticas.

Opondo-se à língua, Saussure estabelece a fala como a parte secundária da linguagem. Nada existe de coletivo na fala; ela é individual e seu estudo é psico-físico, sendo suas manifestações sempre individuais e momentâneas. Essa é a mais famosa dicotomia saussuriana. Língua e fala são interdependentes, mas absolutamente distintas. A língua é indispensável para que a fala seja inteligível e produza todos os seus efeitos, e a fala é necessária para que a língua se estabeleça, mas é a fala que faz a língua evoluir.

Ao entender a língua como um sistema de signos foi consolidado a idéia de língua como uma estrutura. Sabendo-se, no entanto, que está sempre em mudança, como admitir que ela seja um sistema estruturado de sons de fala, palavras, formas gramaticais e estruturas sentenciais. Essa dúvida permaneceu entre os lingüistas por diversas gerações e só foi entendida a partir da idéia de variação lingüística, inicialmente relegada para segundo plano.

Essa noção começou a se tornar relevante e constituir-se em objeto central de investigação com os sociolinguistas, liderados por William Labov, na década de 1960. A partir de então se passou a ter consciência de que a variação é vital no comportamento lingüístico e, além disso, que a variável sociolingüística é central na descrição da fala.

Entende-se por variação a existência de diferenças no modo como uma língua é usada em uma comunidade de fala. Uma mesma língua não é usada de maneira totalmente homogênea no interior de uma mesma comunidade. A variação apresenta fortes correlações com variáveis como a classe social, a idade e o sexo, pois os homens não falam como as mulheres, nem os jovens falam do mesmo modo como os idosos. Até mesmo uma mesma pessoa não fala de forma igual quando está falando com os amigos, ou quando está em uma reunião de trabalho.

A fala tem sido compreendida pelos lingüistas sob três concepções diferentes: a) como meio, enquanto veículo para a língua; b) como comportamento lingüístico das pessoas que falam; c) como enunciados reais produzidos por pessoas reais em situações reais. No entanto, todos concordam que uma porção significativa da fala se desenvolve segundo regras e apresenta uma estrutura, configurando um evento de fala que envolve a presença de participantes que assumem determinados papéis em contextos bem definidos.

Hoje, o fenômeno lingüístico é visto a partir de fatores intimamente vinculados à vida social, cultural e psíquica do indivíduo. Não é mais possível compreender-se os estudos lingüísticos apenas no âmbito das dicotomias saussurianas, mas sim, a partir dos sentidos que se criam, se cruzam e se mantêm em circulação e se expressam de maneira não uniforme, de acordo com os elementos constitutivos da comunicação. Nesse contexto, a reflexão sobre a compreensão das linguagens com as quais trabalhamos no dia-a-dia é uma necessidade que se impõe.

2.3 Reflexões sobre o sentido, significado e significação

É incontestável que a questão do sentido, significado e significação desempenha um papel nuclear na linguagem. Ainda hoje muito se discute a respeito em uma perspectiva científica. Muito pouco se fez no âmbito do significado ou do conteúdo semântico das formas lingüísticas, mas no plano sintático, morfológico e fonológico os estudos lingüísticos desenvolvidos apresentaram um avanço significativo.

Em um determinado momento evidenciou-se a impossibilidade de examinar os mecanismos gramaticais das línguas sem estabelecer uma correlação com os processos de veiculação do sentido. Foi somente a partir de tal constatação que o ramo da Lingüística que estuda o significado, a Semântica, passou a ser objeto de pesquisa sob os mais diversos pontos de vista e a exercer influência decisiva nas investigações sobre a linguagem.

Marques (2001) observa que definir a natureza, as tarefas e a amplitude da Semântica é uma questão bastante difícil. Há na Lingüística a consciência de que a linguagem é significação e de que é impossível e insensato estudá-la sem considerar a sua essência, que diz respeito à troca de conteúdos ou valores semânticos e à circulação de informações objetivas e subjetivas entre indivíduos.

As discussões acerca da natureza do significado são controvertidas e desenvolvem-se a partir de diretrizes variadas e procedimentos metodológicos complexos e, por vezes, até mesmo contraditórios. A multiplicidade de formas de abordagem semântica conduz à necessidade de fazer escolhas sobre um caminho a seguir quanto à compreensão do significado.

As respostas à pergunta ‘o que é significado’ são múltiplas e divergentes. Os especialistas sequer concordam quanto à terminologia empregada. Significado, sentido e significação recebem interpretações diferentes, que variam segundo as correntes de pensamento, a época e as finalidades, e até mesmo de acordo com a área de conhecimento em que são empregadas.

Marques (2001) propõe aos lingüistas que assumam um comportamento semelhante ao dos físicos quando se preocupavam em resolver a questão ‘o que é matéria’. Ela defende a idéia de que os lingüistas devem ser capazes de explicar os diversos fenômenos relativos ao significado empiricamente e de descobrir e prever explicações para os novos problemas a eles relacionados, não se limitando a estabelecer inter-relações entre esses fenômenos. A autora aponta para o fato de que, enquanto os físicos não tinham respostas para a questão geral ‘o que é matéria’, na medida em que só dispunham de alguns dados empíricos relativos a ela, eles se preocupavam em reduzir o problema a aspectos parciais empiricamente observados. Tal procedimento tem sido prática corrente na ciência. Com base nesse argumento ela propõe que os lingüistas se preocupem em resolver questões relacionadas com aspectos parciais do problema enquanto não forem capazes de oferecer respostas para ‘o que é significado’. Assim, deveriam se ocupar com investigações do tipo: o que é a multiplicidade de significado, o que é igualdade de significado, o que é a pressuposição semântica, por exemplo. Tais respostas,

afirma ela, certamente contribuiriam para a construção de uma teoria semântica de base científica (MARQUES, 2001).

Hoje, várias análises têm sido empreendidas na tentativa de responder a essa questão e construir uma teoria semântica. Tais análises, mesmo que ainda fragmentárias e parciais, configuram uma busca rumo ao estabelecimento de critérios capazes de definir a natureza, as tarefas e o objeto da Semântica.

O estudo do significado ultrapassa, hoje, na Lingüística, os limites da competência gramatical dos falantes, restrita à sentença e seus constituintes, e tenta explicar dados da chamada competência comunicativa, que transcendem o plano gramatical estrito. Começa a incluir fenômenos ligados ao uso concreto da língua, em textos falados ou escritos, circunstancial e contextualmente condicionados (MARQUES, 2001, p. 22).

Muitas vezes os lingüistas deixam de lado aspectos contextuais e não dão atenção aos fatores que interferem na produção e interpretação de formas lingüísticas e de enunciados discursivos, privilegiando o plano puramente abstrato-conceitual. Somente quando conseguem transpor tais fronteiras é que passam a considerar os aspectos histórico-sociais, culturais e psicológicos inerentes à faculdade humana da linguagem, e a abordar a linguagem humana em sua plenitude.

2.3.1 Sentido sob diferentes pontos de vista

A existência de diferentes pressupostos teóricos a respeito do significado e sentido, a diversidade de fatos semânticos sem características comuns e o fato de muitos autores considerarem impossível atribuir um conceito científico à noção de significado são fatores que explicam porque a Semântica foi muitas vezes marginalizada no quadro da ciência lingüística.

Sendo o significado um conceito nuclear em Semântica, é preciso que ele seja expresso com a finalidade de dar origem e justificar uma teoria e, além disso, que seja definido a partir de dados empíricos. Também urge chegar-se a um conceito de significado a partir de dados lingüísticos, intuídos como semânticos, firmando um conceito abrangente de significado. Só assim será possível construir uma teoria capaz de definir o seu objeto e de explicar, de modo integrado e coerente, os princípios que organizam e relacionam os fatos de natureza semântica. O conceito atribuído ao significado deve ser capaz de explicar os diversos fenômenos da natureza semântica que estão presentes na língua.

Considerações de caráter semântico sobre a linguagem têm sido feitas desde a Antigüidade, mas é com Bréal, no século XIX, que o ramo da Lingüística que estuda o

significado se constitui como disciplina; sua obra é um marco importante na relação do sujeito com a linguagem. Para ele, não há como tratar a linguagem sem considerar a significação. Afirma que é preciso considerar as palavras nas relações que estabelecem umas com as outras, tanto no conjunto do léxico como no contexto das frases.

Bréal (1992) vê a variação, a polissemia, a restrição e a expansão do sentido como fenômenos da significação decorrentes da própria natureza da linguagem, das intenções daquele que a usa e do contexto socio-histórico cultural em que ocorrem. Por isso seu pensamento tem grande relevância para esta Dissertação, que focaliza a complexidade da preservação do sentido na interface bibliotecário/usuário.

Ao questionar sobre a multiplicidade de sentido e sobre o fato dela não produzir nem obscuridade nem confusão, Bréal (1992, p. 184) faz a seguinte observação: “É que a palavra chega preparada pelo que a precede e pelo que a rodeia, comentada pelo tempo e o lugar, determinada pelos personagens que estão em cena”.

Na perspectiva saussuriana a questão da significação é exclusivamente tratada no âmbito lingüístico. A noção de valor de um signo é colocada como um fato que ocorre no interior do sistema da língua, uma vez que um signo só possui valor, isto é, um signo só existe em suas relações com os outros elementos do sistema. Dessa maneira, o significado de um signo é o que os outros não são. Portanto, um elemento de um sistema não tem nenhum sentido até que ele seja referido a outro elemento do mesmo sistema ou ao sistema como um todo.

Mesmo sem ter a pretensão de desenvolver estudos no campo da Semântica, ao introduzir a idéia de valor, Saussure se posiciona teoricamente em relação ao sentido, e afirma que é uma grande ilusão considerar uma palavra simplesmente como a união de um som e um conceito. Tal entendimento implicaria em isolá-la do sistema do qual faz parte, isto é, em acreditar ser possível, a partir das palavras, construir o sistema pela sua soma; entretanto, o que ocorre é exatamente o contrário, pois é a partir do todo solidário que se obtém, por análise, os elementos que ele encerra.

Ao definir a língua como um sistema de signos, em que todas as palavras são solidárias e o valor de uma unidade lingüística resulta da presença simultânea de várias palavras e da posição que ela ocupa no interior do sistema lingüístico, Saussure afirma que a língua não é determinada por nada que lhe seja exterior. Assim, ao excluir a exterioridade da língua e tornar a questão do significado estritamente lingüístico, no âmbito do valor, ele fez

surgir dificuldades para o tratamento de fenômenos cruciais ao estudo da significação, como a polissemia e a sinonímia, relações de sentido em que a importância atribuída ao contexto é fundamental.

Para Dubois (1997) a polissemia se insere num sistema duplo de oposições: entre polissemia e homonímia e entre polissemia e monossemia. A homonímia pode consistir na identidade fônica (homofonia) ou na identidade gráfica (homografia) de duas palavras que não têm o mesmo sentido de um modo geral.

Polissemia e homonímia permanecem como fonte de muita discussão teórica. Não há dúvida de que fazem parte do léxico de uma língua, mas os lingüistas ainda não chegaram a um consenso sobre como distingui-las. Dentre os critérios já sugeridos estão o da etimologia e o da proximidade da relação entre os itens.

Na oposição entre polissemia e monossemia, entende-se a polissemia, ou significado múltiplo (LYONS, 1987; CRYSTAL, 2000) como sendo uma propriedade da palavra; ela é representada por uma forma e diversas significações. A monossemia, entretanto, por tratar de unidade do léxico especializado, o termo, tem predominantemente um único significado.

A sinonímia, por sua vez, pode ser entendida sob duas acepções diferentes: sinonímia absoluta e quase-sinonímia. Ocorre a sinonímia absoluta quando duas palavras podem ser substituídas em todos os contextos; quando a possibilidade de substituição se dá em enunciados isolados, sem que haja diferença na significação do enunciado como um todo, tem-se a situação de quase-sinonímia.

A discussão sobre a constituição do sentido, sob a ótica das semânticas formais de inspiração lógica, é vista como uma relação com o mundo, com os objetos. Para esses semanticistas a unidade semântica não é o signo, mas o enunciado, uma vez que o sentido do enunciado é obtido a partir de um bloco de significação, formado pela contribuição que as palavras dão para o sentido da sentença.

Por outro lado, ao conceituarem o sentido, esses autores consideram que fica estabelecida uma relação com os objetos, isto é, com um estado de coisas. Sob esse aspecto o sentido é entendido a partir do conceito de verdade, que é definido com fundamentação em reflexões oriundas da Lógica, visto que admitem a existência de leis expressas por essa ciência, que também são válidas para a linguagem.

Dentre os caminhos que vêm buscando repor a discussão sobre o sujeito no estudo da linguagem há aqueles que têm sido desenvolvidos em Filosofia Analítica. Essa corrente exerce uma grande influência sobre as reflexões que se estabelecem a respeito da significação.

Nesse contexto menciona-se Austin, filósofo que defende a idéia de que a linguagem deve ser tratada fundamentalmente como uma forma de ação e não de representação da realidade. Para ele, o sentido de um enunciado não pode ser estabelecido apenas através da análise de seus elementos constituintes, pois são as condições de uso do enunciado que determinam os seus significados. A teoria de Austin se opõe à posição assumida pelos lógicos, que afirmam que toda a proposição é verdadeira ou falsa.

Dentre os teóricos que abordaram a questão do sentido, buscando repor os objetos excluídos por Saussure, convém lembrar Benveniste, que desenvolve a Teoria da Enunciação com base na inclusão do sujeito e na distinção que faz entre semiótico e semântico. No final deste capítulo, no tópico 2.3.3., consta um estudo mais detalhado sobre a Teoria da Enunciação.

2.3.2 Sentido na visão da Lingüística Textual

Ultrapassando os limites das palavras e das frases surgem os estudos sobre o texto como unidade de análise, no campo dos Estudos da Linguagem, consolidando uma nova área da Lingüística, a da Lingüística Textual. Tais estudos têm por objetivo reintroduzir as noções de sujeito e a situação de comunicação.

O texto, enquanto objeto particular de investigação, não é mais visto como um produto acabado; ele passa a ser estudado dentro de seu contexto de produção e a ser compreendido como uma forma específica de manifestação da linguagem, resultado de operações comunicativas e processos lingüísticos em situações sociocomunicativas.

Pela diversidade atribuída à concepção de texto na Lingüística Textual e pelas variadas correntes que surgem, encontra-se na literatura uma denominação bastante variada⁷. A Lingüística do Texto comporta manifestações no âmbito da Semântica, da Pragmática e da Gramática. Cabe à Semântica do Texto esclarecer o que se deve entender por significação de um texto e como ela se constitui; à Pragmática do Texto, dizer qual é a função de um texto no

⁷ A esse respeito faz-se referência a: Textologia, segundo R. Harweg; Teoria do Texto, conforme S. J. Schmidt; Análise do Discurso para M. Pêcheux e Z. S. Harris; Translingüística de acordo com R. Barthes; Hipersintaxe, no entender de B. Palek; Teoria da Estrutura do Texto – Estrutura do Mundo, na opinião de J. Petöfi (FÁVERO; KOCH, 2005, p. 11).

contexto extralingüístico; e à Sintaxe do Texto, expressar o que está em sua volta. Intimamente relacionada à Sintaxe do Texto, está a Fonética do Texto, que se ocupa das características e dos sinais fonéticos da configuração sintática textual.

A importância atribuída à Linguística do Texto é justificada pelo fato dela se propor a verificar o que efetivamente determina e possibilita afirmar que um texto seja um texto, identificando, para tal fim, seus princípios de constituição, os elementos responsáveis pela sua coerência e as condições em que a sua textualidade se manifesta.

O tratamento do texto no seu contexto pragmático adquire, sob esse enfoque, particular importância para os estudos desenvolvidos sob esse domínio. Conforme nos relatam Fávero e Koch (2005), há autores que consideram a inserção da Pragmática nos estudos da Linguística do Texto um fator importante para a sua evolução em direção a uma teoria pragmática do texto.

Sob essa ótica o texto é o ponto de partida para o ato de comunicação, inserido em uma situação específica comunicativa; o ato de comunicação, aqui, é entendido como uma forma específica de interação social. Desse modo, a competência que constitui a base empírica da Teoria do Texto deixa de ser a competência textual e passa a ser a competência comunicativa, expressão usada para se referir à capacidade do falante de empregar adequadamente a linguagem nas diversas situações de comunicação.

Jakobson ao redefinir e ampliar suas reflexões sobre as funções da linguagem contribui para o estudo do discurso, no âmbito da Linguística. Ele reconhece na linguagem as funções: referencial, expressiva, conativa, fática, metalingüística e poética e afirma que: “A linguagem deve ser estudada em toda a variedade de suas funções” (JAKOBSON, 2003, p. 122). Destaca a importância que a função referencial representa na comunicação verbal e considera que ela é determinante para completar o sentido de um enunciado, ligando o contexto lingüístico e o universo extralingüístico.

Sobre a noção de sentido, conforme entendida pela Análise do Discurso, vale acrescentar algumas considerações. O sentido, assim como o sujeito, constitui-se pela interpelação ideológica, o que faz com que as palavras recebam seu sentido da formação discursiva na qual são produzidas. O sentido não está contido na palavra, mas é “[...] determinado pelas posições ideológicas que estão em jogo no processo sócio-histórico no qual as palavras, expressões e proposições são produzidas” (PÊCHEUX, 1988, p. 160); por isso, novas condições de produção serão geradoras de novos sentidos. A univocidade, a

universalidade e a transparência são ilusões que escondem a multiplicidade do sentido. O sentido é múltiplo. Não há um sentido fixo.

Seguindo a linha funcionalista, M. A. K. Halliday (1976) propõe a existência de três funções: a ideacional, a interpessoal e a textual. A primeira corresponde à função cognitiva ou referencial da linguagem e expressa um conteúdo. Representa a manifestação da experiência que o falante tem do mundo real, inclusive do mundo interior de sua própria consciência. Através dessa função, falante e ouvinte organizam e incorporam na língua as suas experiências, cognições e percepções.

A segunda função, a interpessoal, se refere à oposição que o locutor assume perante o ouvinte no processo de enunciação. Ela serve para estabelecer e manter relações sociais, sendo simultaneamente pessoal e interacional, e permite que o falante utilize a linguagem como um recurso para interagir em um evento de fala, assim como para organizar e expressar o mundo interno e externo a ele.

A terceira, a função textual, permite a estruturação do texto e possibilita que a linguagem contextualize as unidades lingüísticas, de forma que possam ser operadas, tanto no co-texto⁸ como no contexto. Ela capacita o falante e o escritor a construir textos, estabelecendo relações coesivas entre uma oração e outra, e o ouvinte e o leitor a distinguirem um texto de um conjunto aleatório de orações.

A Lingüística Textual concentra sua atenção nos diferentes propósitos dos textos e na identificação das propriedades lingüísticas formais, que distinguem tipologicamente os textos entre si; tais propriedades definem a textualidade de um texto, aspecto ao qual ela dá um destaque especial. Os textos são considerados como sendo a unidade da língua e têm uma função comunicativa caracterizada pelos princípios de coesão e coerência.

Ao domínio da coesão pertencem os mecanismos lingüísticos, freqüentemente de natureza gramatical, que servem para criar conexão e estrutura. Alguns tipos de conexão são alcançados, de forma explícita, por mecanismos lingüísticos visíveis, como as anáforas.

Além dos mecanismos de coesão há aqueles que criam a estrutura de um texto, mas sem apresentar uma natureza explicitamente gramatical, uma vez que o foco é determinar o

⁸ Co-texto aqui se refere aos ambientes lingüísticos de um texto, e contexto, aos ambientes situacionais.

grau em que um determinado texto ou discurso faz sentido. A coerência de um texto se revela pelo seu alto grau de conexidade.

2.3.3 Sentido no âmbito desta pesquisa

Pela diversidade com que o tema do sentido é abordado na literatura da área tornou-se difícil identificar um aparato teórico para dar conta da questão do significado neste trabalho. O que se deseja é criar um instrumento que estabeleça a comunicação entre o SRI e o usuário, a partir da preservação do sentido de um descritor/termo desde o momento da indexação do documento pelo bibliotecário até o da recuperação da informação pelo usuário.

Em busca desse modelo, foram identificados, na Teoria da Enunciação de Benveniste, vários elementos para a formulação dos conceitos que se relacionam com a questão do sentido no âmbito da análise documentária. Essa teoria implica na presença de um 'eu', o locutor, e de um 'tu', o alocutário, respectivamente o indivíduo que profere o enunciado e o que recebe a mensagem. Entre o locutor e o alocutário (no caso específico deste estudo entre o autor do texto e o bibliotecário) se estabelece uma relação de intersubjetividade e com ela uma situação de comunicação. Nessa, se veiculam informações que têm origem no locutor e são recebidas pelo alocutário, e são expressas em um aqui e em um agora.

Ao falar sobre a linguagem, Benveniste (1988) argumenta que há certas categorias de expressão que são comuns a todas as línguas e que um modelo constante subjaz a elas. Ele destaca que, embora as formas que revestem essas categorias sejam registradas nas descrições, suas funções só aparecem claramente quando são entendidas no uso da linguagem e na produção do discurso. O autor afirma que tais formas se constituem como independentes de toda a determinação cultural e expressam a experiência subjetiva dos sujeitos que se colocam e se situam na e pela linguagem. Para esse modelo, que só pode ser percebido e estudado no exercício da linguagem e na produção do discurso, o lingüista formula a Teoria da Enunciação.

Desse modo, segundo Benveniste, é pela enunciação que ocorre a conversão individual da língua em discurso e o seu funcionamento se dá através da ação de um sujeito ou indivíduo. A partir da manifestação individual que atualiza a língua é possível detectar, em seu interior, os caracteres formais da enunciação, definindo-a no quadro formal de sua realização. Nesse contexto a língua é empregada visando expressar uma certa relação com o mundo, o que torna a referência ao mundo parte integrante do ato de enunciação.

Das categorias fundamentais do discurso, Benveniste (1988) menciona a de pessoa e a de tempo. Ele declara que elas são interligadas e que a presença da subjetividade na linguagem é algo evidente. A intersubjetividade, considerada como a condição essencial para que a linguagem humana se torne possível, se manifesta na categoria de pessoa que é expressa na relação ‘eu’ e ‘tu’ em oposição a ‘ele’. ‘Eu’ e ‘tu’ são as pessoas do discurso por excelência e são constituídas dentro de uma relação de subjetividade que exclui ‘ele’, a não-pessoa.

No nível enunciativo o ‘ele’ tem um estatuto referencial e não remete a um sujeito. Como falantes de uma língua, temos a possibilidade de nos posicionarmos no discurso ora como ‘eu’, ora como ‘tu’, alternadamente. Por isso, todo homem se coloca em sua individualidade enquanto ‘eu’ em oposição a ‘tu’ e a ‘ele’. Aquele que fala se posiciona sempre como ‘eu’, mas assume em relação a ‘tu’, uma posição de troca.

Assim, em toda a língua e a todo o momento, aquele que fala se apropria desse eu, este eu que, no inventário das formas da língua, não é senão um dado lexical semelhante a qualquer outro, mas que, posto em ação no discurso, aí introduz a presença da pessoa sem a qual nenhuma linguagem é possível (BENVENISTE, 1989, v. 2, p. 68-69).

É pela inserção do discurso no mundo que é possível a instauração da categoria do tempo. Em relação à temporalidade Benveniste distingue três categorias: o tempo lingüístico, o tempo físico e o tempo cronológico. Para fins deste estudo é o tempo lingüístico que será considerado, por caracterizar o tempo específico da língua. “O que o tempo lingüístico tem de particular é o fato de estar organicamente ligado ao exercício da fala, o fato de se definir e de se organizar como função do discurso” (BENVENISTE, 1989, v. 2, p. 74).

Ao estudar a natureza dos pronomes e a distinção instituída entre discurso e história, Benveniste tem como objetivo mostrar que os aspectos discursivos da linguagem se referem às relações que se estabelecem entre os interlocutores ‘eu’ e ‘tu’, na e pela linguagem, o que viabiliza fazer dela um uso referencial.

Dessa maneira, ao focalizar o discurso em seus estudos, Benveniste passa do nível da palavra para o nível do texto. Para ele a comunicação lingüística se torna possível apenas no âmbito do discurso, especialmente no quadro da língua assumida pelo homem e sob a condição da intersubjetividade.

Fazendo referência à noção saussuriana do signo, Benveniste declara que

[...] é necessário ultrapassar a noção saussuriana do signo como princípio único, do qual dependeria simultaneamente a estrutura e o funcionamento da língua. Esta ultrapassagem far-se-á por duas vias:

- na análise intralingüística, pela abertura de uma nova dimensão de significância, a do discurso, que denominamos semântica, de hoje em diante distinta da que está ligada ao signo, e que será semiótica;

- na análise translingüística dos textos, das obras, pela elaboração de uma metassemântica que se construirá sobre a semântica da enunciação (BENVENISTE, 1989, v. 2, p. 67).

Dos objetos excluídos por Saussure no campo da Lingüística – o referente, o mundo, o sujeito e a história –, Benveniste se ocupa particularmente com a questão da subjetividade na língua.

Na concepção saussuriana, conforme já mencionado, a língua é um sistema de signos auto-suficiente que prescindem do referente para se explicar, pois os signos, por si só, dão conta da significação. Ao invés de um sistema de signos justapostos, Saussure propõe uma rede de signos que se combinam e, por seu inter-relacionamento, criam significados. Benveniste procura esclarecer a questão do sentido através da noção de forma e sentido na linguagem e, a esse respeito, observa que os aspectos da forma se apresentam de modo concreto, enquanto que as manifestações de sentido parecem ser livres e imprevisíveis.

Dos modos de significância que a língua combina, Benveniste faz distinção entre o semiótico e o semântico. Em relação a essa dupla significância, ele assim se expressa.

Ela é investida de uma DUPLA SIGNIFICÂNCIA. [...] o semântico toma necessariamente a seu encargo o conjunto dos referentes, enquanto o semiótico é, por princípio, separado e independente de toda referência. A ordem semântica se identifica ao mundo da enunciação e ao universo do discurso.

A língua é o único sistema em que a significação se articula assim em duas dimensões. Os outros sistemas têm uma significância unidimensional [...] (BENVENISTE, 1989, v. 2, p. 64-66).

Partindo de suas considerações sobre a língua e os dois modos distintos de significância, o semiótico e o semântico, Benveniste define o signo como sendo uma unidade semiótica dotada de significação e a palavra como uma unidade semântica dotada de sentido. Uma distinção interessante que faz a esse respeito é que, para o semiótico é preciso reconhecer as unidades lingüísticas, os signos, e para o semântico trata-se de compreendê-lo no âmbito do discurso.

É no uso da língua que um signo tem existência, fora dela ele não existe; um signo só é identificado no interior e no uso da língua. É na relação em oposição com outros signos que um signo se define; assim, é no interior da língua, no nível intralingüístico, que ele se identifica como unidade semiológica. Enquanto a Semiótica tem a função de significar, cabe à Semântica comunicar, ato que se dá pela palavra, pela frase e não pelo signo. O modo semântico refere-se à língua enquanto produtora de mensagens e considera o sentido

globalmente. “Somente o funcionamento semântico da língua permite a integração da sociedade e a adequação ao mundo, e por conseqüência a normalização do pensamento e o desenvolvimento da consciência” (BENVENISTE, 1989, v. 2, p. 229).

A passagem do semiótico ao semântico representa uma mudança radical de ponto de vista; enquanto “[...] a semiótica se caracteriza como uma propriedade da língua; a semântica resulta de uma atividade do locutor que coloca a língua em ação” (BENVENISTE, 1989, v. 2, p. 230). Enquanto o signo se dá no âmbito da realidade intrínseca da língua e tem o significado como parte integrante, a palavra e a frase permitem a comunicação com o que é exterior à língua e seu sentido faz referência à situação do discurso e à atitude do locutor.

Assim, em uma primeira constatação, observa-se que o sentido na acepção semiótica se define por uma relação de paradigma, enquanto que na acepção semântica se realiza sob a forma do sintagma.

O sentido da frase é de fato a *idéia* que ela exprime; este sentido se realiza formalmente na língua pela escolha, pelo agenciamento de palavras, por sua organização sintática, pela ação que elas exercem umas sobre as outras. Tudo é dominado pela condição do sintagma, pela ligação entre os elementos do enunciado destinado a transmitir um sentido dado, numa circunstância dada. Uma frase participa sempre do “aqui e agora” [...] (BENVENISTE, 1989, v. 2, p. 230).

Sobre o sentido, Benveniste faz a seguinte observação: “O sentido de uma frase é sua *idéia*, o sentido de uma palavra é seu emprego (sempre na acepção semântica). A partir da *idéia*, a cada vez particular, o locutor agencia palavras que nesse emprego tem um ‘sentido’ particular” (BENVENISTE, 1989, v. 2, p. 231). Acresce a essa distinção a necessidade de incluir a noção de referência, necessidade que se evidencia quando, fora do contexto, nem sempre é possível compreender o sentido que resulta da junção das palavras. “Se o ‘sentido’ da frase é a *idéia* que ela exprime, a ‘referência’ da frase é o estado de coisas que a provoca, a situação de discurso ou de fato a que ela se reporta e que nós não podemos jamais prever ou fixar” (BENVENISTE, 1989, v. 2, p. 231).

Por isso, a frase é um acontecimento diferente a cada vez que é proferida; enquanto seu sentido é percebido por uma compreensão global no âmbito da *idéia*, na sua totalidade, a forma é obtida pela dissociação analítica do enunciado até chegar à sua unidade mínima, a palavra. A palavra é o instrumento da expressão semântica e materialmente o signo do repertório semiótico. O sentido da palavra é obtido em relação ao contexto de situação; é por meio da palavra que a mensagem é definida, delimitada.

Retomando a subjetividade, Benveniste entende a enunciação como o ato individual de utilização da língua, o ato de produzir um enunciado. A relação que o locutor estabelece com a língua determina os caracteres lingüísticos da enunciação, que pode ser estudada sob diversos aspectos. “Na enunciação consideraremos, sucessivamente, o próprio ato, as situações em que ele se realiza, os instrumentos de sua realização” (BENVENISTE, 1989, v. 2, p. 83) porque, antes da enunciação, a língua é apenas uma possibilidade.

Assim, ao procurar entender o processo de comunicação no âmbito da Linguagem Documentária e a partir da Teoria da Enunciação percebe-se que ele tem que ser considerado a partir de dois momentos distintos: a do processo de indexação e a do processo de busca e recuperação da informação.

No primeiro momento os elementos (eu, tu, aqui e agora) são representados, respectivamente, pelo autor do texto, bibliotecário, texto e processo de indexação. Em um segundo momento os elementos passam a assumir a seguinte representação: SRI, usuário, catálogo eletrônico e processo de busca e recuperação. Os dois momentos podem ser assim representados:

- » autor ⇒ bibliotecário ⇒ texto ⇒ processo de indexação
- » SRI ⇒ usuário ⇒ catálogo eletrônico ⇒ processo de busca e recuperação da informação.

O vínculo que se estabelece entre o processo de indexação e o SRI é que a linguagem utilizada no processo de indexação está expressa em um SRI.

Assim considerando percebe-se que a comunicação a ser estabelecida entre os SRIs e os usuários tem que considerar também a comunicação que se dá entre o autor do texto e o bibliotecário. Segundo esse entendimento, o bibliotecário deixa de ser visto como um canal, mas passa a ocupar um papel relevante no processo da comunicação que se dá no âmbito da Linguagem Documentária, ao invés daquele que, tradicionalmente, lhe era reservado. Tal perspectiva, que coloca o bibliotecário como um dos atores da comunicação, se constitui em um diferencial do projeto piloto aqui desenvolvido.

3 DOCUMENTAÇÃO

A *informação* não é um *dado*. Ela se constrói no encontro de duas dinâmicas: a dinâmica de quem ‘emite’, de quem ‘enuncia’ (o enunciador) e a dinâmica de quem ‘recebe’ o enunciado (o enunciatário). Ela ocorre sempre num espaço onde as posições de quem ‘fala’ e de quem ‘ouve’ são intercambiadas, num jogo de forças permanente.

Aí começa a *linguagem documentária* (CINTRA et al., 2002, p. 10).

3.1 Introdução

O propósito deste capítulo é o de abordar questões que fundamentem as reflexões sobre os processos de indexação e de busca e recuperação da informação. Inicialmente, este estudo se propõe a esclarecer dois conceitos, Biblioteconomia e Documentação, algumas vezes utilizados como variantes. O tópico seguinte focaliza o processamento e transmissão da informação, e a construção e comunicação do conhecimento.

Ao inserir o sistema nocional como parte desta reflexão, tem-se como propósito a compreensão das possibilidades de relacionamentos que se estabelecem entre os descritores/termos no âmbito das relações lógico-semânticas. A organização conceitual de uma área de especialidade é de fundamental importância para o desenvolvimento de uma linguagem documentária. A relevância atribuída a ela deve-se ao fato dela se constituir em um sistema de informações com função comunicativa. A LD faz uso de uma série de procedimentos para operacionalizar o processo de indexação, do qual depende a eficácia oferecida pelos Sistemas de Recuperação de Informação em seus processos de busca e recuperação.

Finalizando esse capítulo é feita referência aos vocabulários controlados, tesouros e sistemas de classificação bibliográfica, que visam à recuperação temática de um documento. A compreensão desses tópicos torna-se relevante quando se pensa em propor uma metodologia para a geração de bases de dados terminológicas na esfera da análise documentária.

3.2 Biblioteconomia e Documentação

É um momento de transição este que passamos – novos perfis profissionais estão surgindo, novas habilidades estão sendo requeridas. Com certeza, um novo profissional da informação surgirá. Se o bibliotecário terá ou não um lugar neste contexto, vai depender, em parte, da sua capacidade de integração, de sua especificidade como especialista no manejo da informação (LUCAS, 1996, p. 72).

Biblioteconomia é a ciência e a técnica de planejar, implementar, administrar e organizar bibliotecas e sistemas de acesso e recuperação de informação. Ocupa-se do conjunto

de conhecimentos teóricos e técnicos indispensáveis para armazenar, recuperar e disseminar informações de forma ágil, eficaz e dinâmica, em qualquer tipo de veículo ou formato. É a atividade mais antiga que trata da organização de documentos; tem origem efetiva na preservação das unidades do conhecimento registrado, sendo marcada pela intensa disseminação das bibliotecas. O profissional especializado é o bibliotecário.

A palavra Biblioteconomia foi usada pela primeira vez em 1839, na obra intitulada *Bibliothéconomie: instructions sur l'arrangement, la conservation et l'administration des bibliothèques*, publicada pelo livreiro e bibliógrafo Léopold-Auguste-Constantin Hesse. Foi somente no século XIX que as técnicas e práticas do bibliotecário começaram efetivamente a serem sistematizadas. No Brasil a profissão é regulamentada pela lei 4.084, aprovada em 30 de junho de 1962.

Segundo o Art. 6º, da lei 4.084,

São atribuições dos Bacharéis em Biblioteconomia, a organização, direção e execução dos serviços técnicos de repartições públicas federais, estaduais, municipais e autárquicas e empresas particulares concernentes às matérias e atividades seguintes:

- a) o ensino de Biblioteconomia;
- b) a fiscalização de estabelecimentos de ensino de Biblioteconomia reconhecidos, equiparados ou em via de equiparação;
- c) a administração e direção de bibliotecas;
- d) a organização e direção dos serviços de documentação;
- e) a execução dos serviços de classificação e catalogação de manuscritos e de livros raros e preciosos, de mapotecas, de publicações oficiais e seriadas, de bibliografia e referência (BRASIL, 1962).

Documentação é a área do conhecimento especializada em pesquisar, desenvolver e utilizar os mais eficazes métodos para tratar a informação, visando sua recuperação e disseminação, independente do suporte utilizado para os registros dos documentos. Caracteriza-se pelo tratamento que dá ao conteúdo temático, pela diversidade quanto ao tipo de registro de informação com que trabalha e pelo uso otimizado das inovações tecnológicas. A importância que deve ser atribuída às atividades documentárias diz respeito não apenas ao acesso às informações, mas igualmente à seleção, organização e conservação do conjunto de conhecimentos técnico-científicos produzidos.

Durante um período a Biblioteconomia e a Documentação apresentaram um desenvolvimento muito semelhante. Elas surgiram em conseqüência das mesmas necessidades e empregavam processos e instrumentos comuns, como as fichas de 7,5 por 12,5 cm, por exemplo; seus objetivos eram praticamente idênticos e, em muitos casos, deviam seu progresso aos mesmos teóricos.

Por algum tempo essas áreas foram tratadas de forma única; no entanto, interesses particulares começaram a dividi-las em dois grupos separados. Os documentalistas passaram a evitar os instrumentos e até mesmo as palavras adotadas pela Biblioteconomia, embora muitas vezes seguissem caminhos já trilhados e descartados pelos bibliotecários. A cisão entre Biblioteconomia e Documentação tornou-se cada vez mais profunda e a divergência refletiu-se na segmentação das associações de classes, onde foram freqüentes os casos de dissidências.

Com a Documentação surge a preocupação com o acesso ao conteúdo dos documentos. Os antigos bibliotecários, representados pelo erudito e bibliófilo desde a Biblioteca de Alexandria, mantinham-se ocupados com o registro do conhecimento erudito e eram dominados pela forma de organização imposta pela tradição filosófica que marcou o mundo antigo e medieval.

Em fins do século XIX Otlet e La Fontaine sistematizaram e desenvolveram a Documentação enquanto disciplina distinta da Biblioteconomia. A palavra Documentação foi então cunhada para significar de forma mais ampla aquilo que anteriormente era denominado como Bibliografia. Pela importância de seu trabalho, Otlet vem sendo considerado o precursor e o fundador da Documentação e da própria Ciência da Informação.

A Ciência da Informação é a “área de conhecimento que cuida do tratamento da informação e gerência dos sistemas e serviços de informação, cuidando das questões ligadas ao fenômeno da explosão da informação; à diversificação dos suportes de informação e necessidade crescente de desenvolvimento de tecnologias de informação, dentre outros aspectos” (SANTOS, 2003, p. 57). Tem como propósito gerar, socializar e democratizar o conhecimento produzido. É uma área em expansão e de interesse para diversos campos do conhecimento. Enquanto área de conhecimento ela visa o estudo dos processos relativos à produção, organização, transmissão e uso da informação. Faz uso de aportes interdisciplinares oriundos da Ciência da Computação, Linguística, Comunicação, Ciência Cognitiva, Psicologia, Matemática e Lógica, dentre outras.

Armazenar e recuperar informações é um permanente desafio enfrentado pela Documentação desde as décadas de 50 e 60 do século XX, em virtude do avanço crescente do conhecimento científico e tecnológico, assim como, do desenvolvimento das tecnologias de informação. Tal situação vem provocando reflexões entre os teóricos da área e exigindo a mudança de procedimentos metodológicos, a começar pela substituição do conceito de recuperação bibliográfica, que foi substituído pelo de recuperação da informação.

Nesse contexto, a Documentação se confrontou com a mudança de paradigma que antes era o documento e agora é a informação, hoje considerada de importância fundamental para o desenvolvimento sócio-econômico-cultural de uma sociedade. Os procedimentos informatizados passaram a ser as ferramentas básicas de trabalho do bibliotecário, tornando o processamento, o gerenciamento, a recuperação e a disseminação da informação mais eficiente e eficaz.

Apesar de tudo, mesmo sem negar a diferença entre o trabalho do bibliotecário e o do documentalista e sem ignorar a divergência profissional que se fez notar durante muito tempo, atualmente a Biblioteconomia e a Documentação são entendidas como sendo uma só profissão. Sob esse enfoque, bibliotecário-documentalista é o profissional que concebe, organiza e administra estruturas de documentação e informação. Para tal, estabelece e aplica critérios de organização e funcionamento dessas estruturas, como por exemplo, bibliotecas e centros de informação e documentação. No âmbito das suas tarefas avalia, adquire e trata os diversos suportes documentais com o objetivo de facilitar o acesso à informação.

Hoje os cursos de Biblioteconomia visam formar profissionais da informação para atuar em um amplo espectro de unidades de informação, desde as tradicionais bibliotecas públicas, escolares e universitárias, até os centros de informação empresariais. Preocupam-se com a formação de profissionais que compreendam o papel da unidade de informação no processo de transformação da sociedade, estando habilitados a identificar demandas de informação e propor soluções inovadoras.

Até a década de 30, como decorrência de uma forte influência francesa, o bibliotecário recebia uma formação humanista ligada à cultura e às artes. Na década de 30 ele passou a receber uma formação mais técnica, sob influência norte-americana. Na década de 80, com a reformulação curricular nos cursos de Biblioteconomia, ele passa a ter um perfil de agente cultural e de informação. No início da década de 90, com o crescimento editorial e o avanço das novas tecnologias de informação, sua formação volta-se para o gerenciamento de unidades de informação, assistido pelos mais modernos recursos digitalizados. Assim, o bibliotecário, que no início do século XX era visto como o guardião de livros, é modernamente reconhecido como o profissional da informação.

3.3 Informação e Conhecimento

Dentre os novos paradigmas da Ciência da Informação perpassa o conceito de sentido e, com ele, a necessidade de dar sentido à informação. Em outras palavras, o sucesso ou o

fracasso de um Sistema de Informação depende de como os interlocutores percebem e compreendem o que está sendo transmitido. A partir da década de 80, com a formação das primeiras redes de conteúdos digitais, a informação transforma-se operacionalmente em um valor de mercado, independente do suporte, levando os países em desenvolvimento a procurarem estratégias para a visibilidade econômica e científica.

Altamente polissêmica, informação, conforme afirma Araujo (1995), é registrada em mais de 400 definições por pesquisadores de diferentes domínios, com visões e conceituações diversificadas. Para fins deste trabalho, ela é tratada no âmbito específico de qualquer conhecimento produzido na e pela comunidade científica; está diretamente ligada ao desenvolvimento de uma área do saber, uma vez que todo conhecimento começa por algum tipo de informação.

Entre informação e conhecimento é possível estabelecer algumas diferenças. A informação é particular e atomizada, ou seja, reduzida a dimensões pequenas, enquanto o conhecimento é estruturado, coerente e frequentemente tem caráter universal; a informação é temporária e transitória, mas o conhecimento tem duração significativa.

Discutir a questão do sentido e do significado implica em refletir sobre o papel da linguagem e da comunicação em relação ao processamento da informação e construção do conhecimento. É no ato de reflexão que o homem percebe e compreende o mundo; é pela sua capacidade de refletir que ele processa constantemente informações.

O processo de construção do conhecimento resulta da interação entre os homens, e deles com o ambiente cultural em que vivem. O indivíduo, inserido em um tempo e um espaço determinado, constrói seus valores e seus sentidos por meio da linguagem, constituindo-se em agente de uma determinada cultura. Nela ele produz seu próprio conhecimento, não como agente isolado, mas como resultado de uma construção coletiva, fruto da interlocução com as diversas leituras que já tenha feito da realidade em que vive. É como resultado das experiências vividas e das práticas comunicativas que o homem constrói seu saber.

Considerar o sujeito como agente de seu tempo/espaço e de sua comunicação é percebê-lo como sujeito da cultura, mediado pela linguagem e que através dela produz palavras, sentidos e valores. Sujeito que não é a única fonte de sua ação como não o é de seu dizer. O que diz tem a ver com o que já ouviu dizer, com o que vai dizer, com o que pode dizer, com o que quer dizer (TEIXEIRA, 1995, p. 42).

O fato de o conhecimento se encontrar materializado pelo uso da linguagem e de um suporte assegura sua navegação para além do espaço e do tempo em que foi criado e, dessa

forma, fora de seu contexto de criação, uma vez que sua dimensão espacial é extremamente dinâmica. Assim sendo, é sempre possível acessar e utilizar uma informação em um contexto diferente daquele em que foi produzida, particularidade que permite que seja recontextualizada. Com a modernidade torna-se necessário criar formas de controle e difusão do saber produzido e armazenado.

O saber assim constituído decorre das prioridades estabelecidas em relação àquilo que o indivíduo precisa ou deseja conhecer. Mas para que ocorra a construção do conhecimento exige-se que se estabeleça um processo de comunicação, que implica na presença de um emissor, de um receptor e de canais de transferência da informação, que interagem como mediadores da ação comunicativa.

Quando o conhecimento é comunicado, isto é, quando a informação chega ao receptor, ocorre um processo de interpretação e compreensão por parte do indivíduo que a recebe, em virtude da interação que se dá entre o saber acumulado e as experiências vividas. Tal fato interfere na construção do conhecimento, seja pelas relações que estabelece, ou mesmo, pelas exclusões que faz; e como resultado dessa ação interativa, o novo conhecimento construído, quando registrado e comunicado, gera uma nova informação. “Desta forma, o conhecimento é gerado nas ‘ações interativas’, mas pode ser comunicado apenas por meio de ações de interlocução” (GOMES, 2000, p. 64).

Um conhecimento construído é normalmente retomado nas práticas comunicativas, sendo que a cada retomada há uma nova possibilidade de ressignificação. Salienta-se, no entanto, que a distância estabelecida entre o emissor e o receptor pode representar um obstáculo à sua compreensão, decorrente de uma série de fatores culturais. Nesse contexto a ciência busca um padrão universal para a transferência da informação.

É sabido que as informações transmitidas não são completamente absorvidas pelo sujeito receptor, pois ocorre sempre um confronto com aquelas anteriormente interiorizadas. A partir do momento em que o sujeito constrói seu próprio conhecimento, uma nova informação poderá ser gerada; as idéias transmitidas são sempre reconstruídas e recontextualizadas.

O desenvolvimento científico e tecnológico tem sido responsável por uma massa enorme de informações geradoras de novos conhecimentos. Ao serem registrados em um suporte físico, as informações se instituem em documentos, o que permite que o conhecimento não se perca e possa ser compartilhado. Enquanto documentos, elas recebem

tratamento adequado para propiciar a criação de novos conhecimentos, a sua divulgação e a socialização.

Embora as tecnologias da informação e da comunicação possam interferir na construção de um novo saber, a interpretação e a ressignificação sempre ocorre no sujeito receptor. “Enquanto a escrita preserva a informação fixada pelo texto, o sentido será reconstruído pelo sujeito leitor por meio de sua interpretação, sem a intervenção direta e imediata do emissor da mensagem” (GOMES, 2000, p. 65).

Na comunicação oral, onde a emissão e a recepção da informação ocorrem no momento da enunciação, a relação estabelecida entre o sentido e a interpretação do fato comunicado está menos sujeita a um ambiente de interlocução mais crítico e dinâmico do que na comunicação escrita.

Com a escrita, e mais recentemente com as novas tecnologias, o volume de informações armazenadas fora do contexto da emissão ampliou as possibilidades de acesso ao conhecimento e criou condições de fixação da informação em um ambiente externo ao da memória biológica do homem. Tal fato amplia a possibilidade de descontextualizar as informações pelo fato do Sistema viabilizar diferentes combinações de termos no processo de busca e recuperação efetuado pelo usuário. Assim, em uma pesquisa em que o usuário utiliza como expressão de busca os termos *Análise e Discurso*, por exemplo, o resultado da pesquisa não obrigatoriamente oferecerá documentos que tratem exclusivamente sobre *Análise do Discurso*.

Sob esse ponto de vista, a partir da Informática, o acesso à informação ganha em operacionalidade e velocidade, mas ao mesmo tempo perde em precisão. Com o surgimento da Informática, o processamento, o armazenamento e a recuperação da informação sofrem um impacto de grandes proporções. O suporte da informação torna-se mais flexível, registrando, além da palavra escrita, a imagem e o som, sendo possível, inclusive, realizar animação e fazer associações com outros documentos do universo hipertextual.

3.4 Sistema nocional

O sistema de conceitos é fruto de uma visão do domínio estudado, podendo estruturar-se diferentemente, de acordo com os critérios empregados (BARROS, 2004, p. 108).

O universo de conhecimento de uma área de especialidade é constituído por um conjunto de noções que privilegiam determinados enfoques. Essas noções são referidas por unidades lingüísticas que se constituem nos descritores/termos que, organizados de forma sistêmica, configuram o sistema nocional de uma área.

Neste capítulo, termo e descritor são usados como equivalentes no âmbito da Biblioteconomia. Esse entendimento difere daquele adotado pela Terminologia que vê o termo na perspectiva mais ampla que será adotada neste trabalho a partir do quarto capítulo. A Terminologia privilegia o termo como unidade de conhecimento, unidade lingüística e unidade de comunicação. A Ciência da Informação, através da Documentação, e mais especificamente da linguagem documentária, entende o descritor ou termo como unidade de informação e signo de comunicação.

A organização nocional de um campo de conhecimento faz-se necessária para a compreensão das situações de relacionamento que ocorrem entre os descritores/termos e, também, para a construção de ferramentas que permitam a eficácia do processo de tratamento e recuperação da informação. A ausência de um sistema nocional sistematicamente organizado compromete o processo de indexação⁹, que fica sujeito ao entendimento que cada indexador tem da área a ser indexada e do momento em que opera com a informação.

Embora não se tenha dúvida sobre a relevância de um sistema nocional para a organização do conhecimento de uma área de especialidade, não existe consenso quanto à melhor maneira de estruturá-lo, ainda mais na época atual, que assiste ao inter-relacionamento de áreas especializadas que se multiplicam em progressão geométrica e nas quais os conceitos são numerosos e variados.

É inegável que a organização do sistema nocional de uma área de especialidade constitui-se em um parâmetro básico de sustentação das linguagens documentárias. Entretanto, nesse contexto de conhecimento globalizado e onde o computador permite que os mais complexos sistemas de relacionamentos entre conceitos sejam estruturados, vale refletir sobre os tipos de relação a serem estabelecidas. A esse respeito Sager (1990) assim se posiciona: “The simplistic view of the past that concepts are adequately represented by three types of relationships (generic, partitive, other) has been generally abandoned.”¹⁰

⁹ Para maiores informações sobre o processo de indexação, ver tópico 3.6.

¹⁰ Tradução da autora: A visão simplista do passado de que os conceitos são adequadamente representados por três tipos de relações (genéricas, partitivas, outras) tem sido geralmente abandonada.

Tradicionalmente entende-se que um sistema nocional é constituído por uma seqüência de noções distribuídas em um eixo vertical e um eixo horizontal. A partir desse ponto de vista, reconhece-se que a estrutura básica de uma linguagem documentária seja estabelecida por relações hierárquicas e não-hierárquicas. As do tipo hierárquicos podem ser genéricas, específicas ou partitivas; elas marcam relações de gênero e de espécie, assim como de parte e todo, e se definem segundo critérios de subordinação e superordenação.

Nas relações genéricas, as noções são dispostas em um eixo vertical e representadas por uma seqüência de noções subordinadas. Nesse tipo de relação, a noção genérica é aquela que é superordenada e que comporta as mesmas características daquelas que lhe são subordinadas. Nas relações específicas subordinadas, as noções específicas compartilham das mesmas características da noção que lhes é imediatamente superordenada, mas possuem, pelo menos, uma característica distintiva a mais, que serve para diferenciar os conceitos específicos de igual nível de abstração. A superordenação caminha das diferenças para as semelhanças, ou seja, da espécie para o gênero, enquanto a subordinação faz o caminho inverso, indo das semelhanças para as diferenças, do gênero para a espécie.

Na relação partitiva, a noção superordenada diz respeito a um objeto considerado como um todo, e as noções subordinadas referem-se a eles enquanto partes de um todo. Embora sejam representadas de forma semelhante às relações genéricas, elas não se confundem, porque a característica a ser observada é a da relação entre o todo e as suas partes, sendo que o conceito atribuído à parte depende daquele conferido ao todo.

Os relacionamentos enumerativos podem ser considerados como uma modalidade da relação de tipo partitiva; dizem respeito às relações que ocorrem entre uma categoria geral de objetos e um caso especial de tal categoria. Nesses casos, os elementos subordinados hierarquicamente, mesmo não se constituindo em parte de um todo, representam casos específicos do descritor/termo genérico.

As relações não-hierárquicas configuram-se como relacionamentos complexos que não podem ser classificados segundo os critérios de ordem e subordinação lógica. Nesse tipo de relação, enquadram-se as relações associativas e de equivalência. As relações associativas não-hierárquicas indicam a ligação entre descritores/termos que pertencem a campos semânticos distintos, mas próximos. As relações de equivalência remetem o conjunto dos não-descritores ou não-terms para o conjunto de termos adotados pelo sistema.

A importância atribuída ao estudo das relações de equivalência no âmbito do trabalho documentário é que, em uma linguagem construída, a cada unidade de informação deve corresponder um único sentido referencial. Dessa maneira, ao intensificar-se o processo de controle sobre a variação do significado, possibilita-se um rigor maior no tratamento da informação e melhores resultados no processo de recuperação. A transformação da unidade de significação em unidade de informação é a característica primordial de uma linguagem documentária.

3.5 Linguagem documentária

Por linguagem documentária entende-se “[...] uma linguagem convencional utilizada por uma unidade de informação para descrever o conteúdo dos documentos, com o objetivo de armazená-los e recuperar as informações que eles contêm” (SANTOS, 2003, p. 147). A LD se constitui em um Sistema de Informação que, assim como a linguagem natural, tem como propósito a comunicação. Sua função comunicativa é restrita aos contextos documentários e à comunicação que se dá entre o autor do documento e o bibliotecário, e entre o sistema e o usuário. A representação dessa dupla situação de comunicação é apresentada no item 2.3.3.

A aproximação entre o conhecimento lingüístico e o conhecimento documentário é evidenciada por diversos aspectos. Dentre os relacionamentos que se podem estabelecer entre a linguagem natural e a linguagem documentária, destaca-se o caráter comunicativo da linguagem e o fato de que o caráter sistêmico é uma característica que se encontra presente nas duas linguagens, pois tanto uma unidade lingüística como uma unidade documentária não pode ser entendida em separado. “De fato, cada unidade só pode ser ‘lida’ na sua relação com as demais unidades componentes do sistema” (CINTRA et al., 2002, p. 16).

O bibliotecário, ao fazer uso da linguagem documentária, tem como propósito favorecer a socialização da informação, assegurando que o saber produzido por uma determinada área não se perca. Na tarefa de divulgar esse conhecimento ele enfrenta sérias dificuldades, pois, ao indexar um documento, precisa preservar o sentido pretendido pelo autor do texto, de forma que as informações ali contidas possam ser recuperadas pelo usuário com alto grau de precisão e o menor nível de ruído possível. Por ruído entende-se: “Tudo o que dificulta a comunicação, interfere na transmissão e perturba a recepção ou a compreensão da mensagem. [...] Todo fenômeno que ocasiona perda de informação durante o transporte da mensagem entre a fonte e o destinatário” (RABAÇA; BARBOSA, 1987, p. 522).

Compartilhando-se a idéia de que o autor do texto imprime um conjunto de valores às palavras pelas quais se expressa e que geralmente eles não coincidem com os do bibliotecário de indexação, evidencia-se a necessidade de buscar mecanismos que evitem desvios de sentido no processo de comunicação do conhecimento e de transferência da informação. A complexidade dessa tarefa tem gerado inúmeras discussões entre os lingüistas, relacionadas prioritariamente com duas questões: o que é a linguagem e o que é o sentido. De modo especial, no contexto desta Dissertação, estas questões se tornam particularmente relevantes no que se refere à preservação do sentido na linguagem documentária.

A linguagem é entendida por muitos como um fenômeno complexo, “que não se esgota no estudo das características internas à língua, em termos de propriedades formais do sistema lingüístico, mas se abre para outras abordagens que considerem o contexto, a sociedade, a história” (PETTER, 2004, p. 23). Na opinião de Saussure, “é ao mesmo tempo, um produto social da faculdade da linguagem e um conjunto de convenções necessárias, adotadas pelo corpo social para permitir o exercício dessa faculdade nos indivíduos. Tomada em seu todo, a linguagem é multiforme e heteróclita [...]” (SAUSSURE, 1988, p. 17). E para Benveniste (v. 1, 1988, p. 286)

A linguagem só é possível porque cada locutor se apresenta como *sujeito*, remetendo a ele mesmo como *eu* no discurso. Por isso, *eu* propõe outra pessoa, aquela que, sendo embora exterior a ‘mim’ torna-se o meu eco – ao qual digo *tu* e que me diz *tu*. A polaridade das pessoas é na linguagem a condição fundamental, cujo processo de comunicação, de que partimos, é apenas uma conseqüência totalmente pragmática.

Sobre o sentido, conforme foi aqui anteriormente comentado, Saussure busca na noção de valor uma explicação para a questão da significação, enquanto Benveniste formula a mesma questão com fundamentação na Teoria da Enunciação e os filósofos da linguagem desenvolvem suas investigações a partir do conceito de verdade. Bakhtin (1995, p. 95) afirma que a “[...] palavra está sempre carregada de um conteúdo ou de um sentido ideológico ou vivencial.” Conforme revisão de literatura sobre o sentido, apresentado no tópico 2.3 desta Dissertação, fica clara a diversidade com que os teóricos abordam essa questão.

Da mesma forma como ocorre na linguagem natural, o sentido das palavras na LD está estritamente condicionado pelas relações que elas mantêm entre si, sujeitas, inclusive, às condições extralingüísticas. A LD expressa a linguagem que identifica o conjunto de uma área de conhecimento; ela não tem o propósito de exibir o conteúdo de um texto específico, mas de representar o campo conceitual ao qual se refere, sendo de natureza eminentemente referencial. Tem como função o tratamento da informação, disponibilizando o conhecimento

armazenado nos acervos aos diferentes segmentos de usuário e viabilizando a socialização da informação. Seu principal objetivo é alcançar eficiência nos processos de indexação e de busca e recuperação da informação, visando uma aproximação com alto grau de eficácia entre os Sistemas de Recuperação da Informação e os usuários.

A construção de uma LD é um processo bastante complexo que precisa considerar alguns aspectos, tais como: a) um sistema padronizado, por si só, não é garantia de sucesso na recuperação da informação; b) nas comunicações especializadas também se reconhece a existência de variação terminológica; c) o apagamento da diversidade denominativa não evita, obrigatoriamente, os ruídos no processo comunicacional; d) a não coincidência terminológica ocorre com relativa frequência, uma vez que o usuário não domina a linguagem de indexação da qual faz uso o bibliotecário.

A LD é uma linguagem construída que, sob determinadas condições, estabelece relações entre os descritores/termos; não admite que duas ou mais palavras se refiram a um mesmo conceito, nem que uma mesma palavra seja utilizada para designar vários conceitos. Ela está presa a uma série de convenções, como por exemplo, a de que atribuir a uma unidade lingüística o estatuto de descritor/termo é convencionar que seu uso está autorizado. Utiliza uma série de procedimentos e de instrumentos para operacionalizar a tarefa de indexar documentos, seja sob os aspectos formais, autoria, título, local, editora, data, seja sob os aspectos de conteúdo. É sob o aspecto de conteúdo temático que reside o grande desafio enfrentado pelo profissional da Biblioteconomia e que desperta o interesse desta Mestranda.

[...] o grande desafio está sempre na apreensão e divulgação da informação especializada no sentido de que os conteúdos dos documentos sejam identificados e adequadamente registrados por meio do processo de indexação, cuja qualificação maior é dar representatividade aos conceitos veiculados na documentação analisada. Esse processo, de acordo com a área, é feito com a ajuda de um vocabulário limitado e controlado, unívoco e coerente, que dá sustentação à linguagem documentária (KRIEGER, 2004, p. 60).

Para que cumpra sua razão de existir a LD deve estar apta a ser permanentemente atualizada, constituindo-se em um instrumento dinâmico. Como tal, a LD deve permitir a incorporação de novos descritores/termos, não apenas em função do avanço científico e tecnológico, mas também do estabelecimento de novas relações entre descritores/termos, conforme descrito no tópico que trata de Vocabulário controlado, 3.8.

A compreensão que envolve o desenvolvimento de uma linguagem documentária como uma ferramenta a ser utilizada em catálogos eletrônicos torna-se cada vez mais relevante, especialmente pelo fato da nossa sociedade ser reconhecida como a sociedade da

informação. Uma das alternativas buscada para a construção de linguagens documentárias, como condição para se obter resultados positivos na busca e recuperação da informação, é a que tem sido oferecida pela Terminologia, por intermédio das terminologias especializadas.

Vale lembrar que, isoladas, as palavras não têm significado ou têm todos os significados possíveis. É só no discurso, ou seja, no uso, que as palavras assumem significados particulares. Como, via de regra, os elementos das LDs são desvinculados dos contextos onde aparecem, pode-se correr o risco de que as palavras que as integram assumam todos ou nenhum significado. Por meio das terminologias de especialidade, as palavras passam a ser termos, assumindo significados vinculados a sistemas de conceitos determinados. Confere-se, desse modo, referência às palavras, que passam a significar segundo determinados sistemas nocionais, assegurando interpretações pertinentes (CINTRA et al., 2002, p. 40).

Uma linguagem documentária construída com a finalidade de cobrir um domínio específico do conhecimento tem mais probabilidades de viabilizar a representação de assuntos pela sua especificidade. Nesse direcionamento, a tendência à especialização tem levado a abandonar uma intenção primeira de cobrir todo o universo do conhecimento. Da constituição da LD, em relação ao modo como foi concebida em um sistema nocional de um domínio de especialidade, vai depender o grau de especificidade temática a ser atribuído a um documento. Na área das Ciências Sociais e Humanidades é grande o número de linguagens documentárias construídas com o objetivo de abarcar uma extensa área do saber.

A utilização da linguagem natural na construção das linguagens documentárias pode gerar um produto que reflete a complexidade própria da linguagem, devido a fenômenos como a ambigüidade, a sinonímia e a variação. A aproximação que a LD estabelece com a teoria, expressa pelos estudos terminológicos sobre esses fenômenos, tem o propósito de buscar soluções para minimizar essas dificuldades.

Na linguagem natural a ambigüidade é resolvida com certa facilidade pelo contexto; na LD, no entanto, utiliza-se o recurso de modificadores com a finalidade de contextualizar o sentido. Modificador, para alguns autores, é um dos componentes de um descritor/termo composto que serve para restringir a extensão do foco, isto é, do componente que identifica a classe geral de conceitos à qual se refere o descritor/termo composto, como por exemplo, no caso dos adjetivos *semântico* e *lingüístico* em *valor semântico* e *valor lingüístico*. Mas para Marisa Bräscher (1999, p. 8) modificador se refere “[...] aos termos de uso freqüente e de significado geral, que geralmente expressam ações ou atributos e que são utilizados de forma combinada com descritores, esclarecendo ou delimitando o significado dos mesmos.”

A palavra é sempre fonte de significação. Quando na comunicação lingüística torna-se possível mais de uma significação, diz-se que a palavra ou a sentença é ambígua. A ambigüidade decorre tanto da plurissignificação como da polissemia e homonímia. Segundo Cintra et al. (2002), a diferença entre plurissignificação, polissemia e homonímia é que a primeira resulta da organização sintático-semântica de enunciados e acontece no discurso, enquanto que as outras duas são fenômenos específicos do vocábulo.

Em uma linguagem documentária a situação de ambigüidade deve sempre ser evitada, ou pelo menos neutralizada, de modo que se alcance a almejada monossemia. Por monossemia entende-se os termos que apresentam apenas um significado. Enquanto a linguagem natural se caracteriza pela dinamicidade com que se transforma e evolui a cada momento, a LD é construída com a finalidade de significar de maneira precisa. Ao contrário do comportamento polissêmico da palavra na linguagem natural, a LD tende à monossemia e favorece o princípio da univocidade, ainda que admitida a presença de variações.

Na linguagem natural as palavras têm função interpretativa e criadora, enquanto na LD elas têm função informativa, significam sob determinadas condições e se amoldam a cada realidade contextual. Quando se faz referência à linguagem documentária como vocabulário controlado, pressupõe-se que ela resulte de mecanismos interpretativos próprios, que seja de natureza monossêmica e que se apresente como unidades de linguagens de especialidade.

O funcionamento lingüístico de um item lexical na LD deve estar atrelado ao seu respectivo sistema nocional e definir-se por suas relações com os demais. Quando as relações que se estabelecem entre os descritores/termos são de equivalência, observa-se uma situação de sinonímia, fato lingüístico que depende do contexto muito mais do que relações de sentido como hiponímia e antonímia. Na hiponímia está implícita a noção de inclusão ou subordinação e, na hiperonímia, a de superordenação. Tais relações permitem verificar se um descritor/termo pertence ou subordina-se a outro mais geral. Os descritores/termos coordenados que pertencem a uma mesma série são reconhecidos como co-hipônimos.

Nas linguagens documentárias a sinonímia visa remeter o usuário de um descritor/termo não-selecionado, e por isso não-preferencial, para um descritor/termo selecionado e preferencial. Ela ocorre quando há possibilidade funcional de substituição entre dois descritores/termos, compreendendo tanto a sinonímia absoluta como a quase-sinonímia. Uma vez que tais linguagens têm por função compatibilizar a linguagem de especialidade com a utilizada pelo usuário, por meio de descritores/termos preferenciais, a abordagem da questão dos sinônimos e quase-sinônimos torna-se essencial.

A relação de hiponímia, por sua vez, permite explicar os diferentes relacionamentos hierárquicos que não se enquadram na classificação gênero e espécie, todo e parte ou parte e parte. É o caso de descritores/termos em posição de contraste; seu agrupamento por superordenação só se torna possível quando são utilizados elementos provenientes de outras partes do discurso. A hiponímia pode manifestar-se de diferentes maneiras, o que explica o fato de freqüentemente não ser possível aplicar o esquema lógico todos e alguns, por exemplo.

3.6 Processo de indexação

Ao indexador raramente é dado o luxo de poder ler um documento do começo ao fim. A exigência de indexar determinada quantidade de itens por dia haverá de lhe impor que se satisfaça comumente com uma leitura que estará longe de ser completa. [...] As partes a serem lidas atentamente são as que apresentam a maior probabilidade de dizer o conteúdo no menor tempo: título, resumo, sinopse e conclusão (LANCASTER, 2004, P. 24).

A indexação de um documento pode ser entendida sob duas formas: pelo registro dos dados referentes ao autor, título e demais informações de caráter bibliográfico e pela identificação do assunto de que trata, o que corresponde, respectivamente, ao processo de catalogação descritiva e ao processo de indexação temática. Neste trabalho focaliza-se esse último e tem-se como objetivo viabilizar a recuperação de um assunto sempre que solicitado pelo usuário de um Sistema de Informação.

Indexação “é uma operação que consiste em extrair os elementos que caracterizam o conteúdo do documento para se obter uma síntese mediante a atribuição de um ou mais termos, com a finalidade de recuperar a informação” (SANTOS, 2003, p. 122).

Lancaster (2004) considera que o processo de indexação de assuntos envolve duas etapas principais: análise conceitual e tradução. Para o autor, análise conceitual nada mais é do que a identificação dos assuntos de que trata um documento; compreende a análise, síntese e representação da informação e tem por objetivo recuperá-la e disseminá-la. É o procedimento que dá início ao processo e visa coletar, selecionar e registrar os conceitos considerados relevantes para serem recuperados durante o processo de busca e recuperação da informação. A tradução consiste na conversão da análise conceitual de um documento em um determinado conjunto de descritores/termos.

Segundo a NBR 12676 da Associação Brasileira de Normas Técnicas (ABNT) a análise temática compreende três etapas: a que se refere ao exame do documento quanto à sua forma física, uma vez que a identificação de seu suporte físico determina estratégias de análise

diferenciadas; a seleção dos conceitos que serão indexados; e a tradução dos conceitos selecionados para a linguagem de indexação adotada pelo SRI.

Quando o indexador pode fazer uso de uma linguagem bem estruturada, com as relações lógico-semânticas claramente estabelecidas e com uma terminologia precisa, é grande a possibilidade de representar adequadamente o conteúdo temático de um documento nos diferentes níveis de especificidade. O uso de variantes terminológicas (ver tópico 4.3) como termos de indexação, quando expressas pelo uso de remissivas, aumenta ainda mais a possibilidade de sucesso na busca e recuperação de informações.

Uma boa indexação viabiliza resultados de busca representativos das necessidades dos usuários; para tal, é preciso que os descritores/termos atribuídos ao documento representem devidamente os assuntos que ele contém. Segundo Lancaster (2004, p. 83), “define-se de modo muito pragmático a ‘boa indexação’ como a indexação que permite que se recuperem itens de uma base de dados durante buscas para as quais sejam respostas úteis, e que impede que sejam recuperados quando não sejam respostas úteis.”

Uma indexação tem qualidade quando igualmente apresenta coerência. Qualidade e coerência em indexação não podem ser entendidas como tendo o mesmo sentido, o que significa dizer que ter coerência não necessariamente é ter qualidade. Um documento pode estar bem representado tematicamente e de forma diversa por diferentes indexadores, caracterizando qualidade, mas falta de coerência. Para que tenha coerência e qualidade é necessário que os documentos que tratem do mesmo assunto sejam representados por uma única forma e em sua potencialidade, de modo que se evite perda de informações no momento da busca e recuperação de um determinado tópico temático.

Lucas, ao pesquisar sobre a leitura do bibliotecário para fins de indexação, observa que

[...] não existe um conjunto ‘correto’ de termos de indexação para documento algum. A mesma publicação pode ser indexada de forma bastante diferente em diferentes centros de informação, e deve ser indexada de modo diferente, se os grupos de usuários estiverem interessados nesses documentos por diferentes razões (LUCAS, 1997, p. 50).

Um procedimento que pode auxiliar na prática de indexação é a consulta a outras bases especializadas da mesma área de conhecimento. Entretanto, há um ponto essencial a ser observado: o de saber sob que aspectos o texto pode interessar ao usuário. Um dicionário de língua portuguesa, por exemplo, tanto pode interessar por ser um dicionário de língua como

por ser um dicionário monolíngüe. Um romance pode interessar tanto por ser uma obra de literatura como por ter como característica ser um romance histórico ou autobiográfico.

O que se percebe no processo de indexação é que, por ser a análise temática uma tarefa altamente subjetiva, diferentes indexadores ou até mesmo um mesmo indexador, em momentos diferentes, apresentam pontos de vista diferenciados em relação ao conteúdo de um documento, aos conceitos que devem ser considerados relevantes para representá-lo e aos descritores/termos selecionados para identificá-lo.

O procedimento de indexação, mesmo que considere o usuário imediato como prioridade na sua política de indexação, deve igualmente considerar os diferentes tipos de usuários que acessam o sistema. Entende-se como usuário imediato aquele que faz parte da instituição a que o documento pertence e por política de indexação uma decisão administrativa que está condicionada à existência de algumas variáveis, como por exemplo: a) características e objetivos da organização; b) identificação da clientela; c) recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

É na política de indexação que se define a cobertura de assuntos (centrais e periféricos); seleção e aquisição de documentos segundo a extensão da cobertura do sistema em áreas de assunto de seu interesse e a qualidade dos documentos, nessas áreas de assunto, incluídos no sistema; o processo de indexação estabelecendo os níveis de exaustividade e especificidade, a capacidade de revocação e precisão e a escolha da linguagem; a estratégia de busca; a forma de saída; o tempo de resposta do sistema; e a avaliação do sistema de modo a determinar até que ponto o sistema satisfaz as necessidades dos usuários.

A linguagem de indexação afeta o desempenho de um sistema de recuperação de informação tanto na estratégia de busca quanto na indexação; portanto, a partir de estudos do sistema, deve-se optar entre linguagem livre ou linguagem controlada e linguagem pré-coordenada ou pós-coordenada.

Ao descrever o conteúdo temático de um documento o indexador faz uso de um ou mais descritores/termos, que tanto podem ser extraídos de um vocabulário controlado como do próprio documento, nesse caso, caracterizando o uso de um vocabulário livre. Tais descritores/termos podem ser designados como termos de indexação, descritores e palavras-chave. Para os propósitos deste trabalho interessa discutir apenas a indexação que faz uso de um vocabulário controlado.

O grande desafio que envolve o processo de indexação é o de oferecer ao usuário documentos que sejam realmente relevantes para ele, que contribuam para satisfazer suas necessidades de pesquisa e, desse modo, a construção do conhecimento. O que se deseja ao indexar um documento é que a linguagem de indexação empregada pelo bibliotecário venha a ser utilizada como estratégia de busca pelo usuário do sistema, estabelecendo entre ambos a comunicação almejada.

A representação do conteúdo temático de um documento pode ser feita de forma seletiva ou exaustiva, sendo que quanto maior a representação, mais pontos de acesso oferece. São os descritores/termos atribuídos pelo indexador que se constituem em pontos de acesso no processo de pesquisa efetuado pelo usuário ao sistema.

Conforme Lancaster (2004) considera-se uma indexação seletiva aquela que inclui até cinco descritores/termos e exaustiva a que inclui mais de cinco. A indexação seletiva visa representar somente o conteúdo temático principal do documento e, a exaustiva, o conteúdo temático do documento de maneira mais completa. Vale destacar que, quanto maior o número de descritores/termos utilizados para indexar um documento, mais acessível ele se tornará e um maior número de vezes será recuperado.

Exaustividade é uma medida de extensão que se refere à profundidade de análise temática de um documento, objetivando a indexação do assunto a ser especificado; e seletividade é o procedimento que determina a indexação de um documento apenas pelos tópicos considerados mais relevantes e representativos para constituir a unidade de informação. A princípio, uma indexação exaustiva representa melhor o assunto de um documento; mas dependendo das características do texto, ele pode estar muito bem representado por uma indexação seletiva.

Quanto aos critérios de escolha de um termo de indexação, registra-se serem altamente subjetivos, mesmo quando apoiados em uma ferramenta de trabalho, pois resultam da escolha do bibliotecário; quanto à eficácia desse processo é senso comum que ele depende basicamente da qualificação do profissional que executa essa tarefa e dos recursos que a ferramenta de indexação lhe oferece.

Na indexação, o saber que o profissional acumula é determinante. Nesse sentido há a considerar diferentes fatores, tais como: a) o conhecimento da área específica que está sendo indexada, b) a cultura do indexador sobre o mundo, c) a sua capacidade de saber ler, interpretar e fazer inter-relações com outros textos já indexados, d) o seu conhecimento da

linguagem e das estruturas discursivas do domínio de especialidade que estiver indexando, e) sua capacidade de compreensão de leitura em idiomas estrangeiros.

Em conformidade com o ideal de conciliar a linguagem usada pelo autor do documento a ser indexado, com aquela utilizada pelo usuário na pesquisa, a Teoria Comunicativa da Terminologia e a Socioterminologia oferecem subsídios teóricos capazes de dar suporte para que novas linguagens de indexação venham a ser construída. O tópico 4.3 aborda essas duas teorias.

3.7 Sistemas de Recuperação da Informação

Os Sistemas de Recuperação da Informação tratam da representação, do armazenamento, da organização e da localização dos itens de informação. Um item de informação é geralmente constituído de textos, tais como documentos diversos, páginas *web*, livros, etc., embora possa conter outros tipos de dados como fotografias, gráficos e figuras.

No contexto deste trabalho, os catálogos eletrônicos de assunto são entendidos como Sistemas de Recuperação da Informação com o objetivo principal de realização de processos de comunicação, oferecendo ao usuário o acesso às informações potencialmente contidas nos documentos, com a finalidade de maximizar o seu uso. Entendem-se como variantes também as seguintes denominações: SRI, catálogos eletrônicos, sistemas de informação ou, ainda, sistemas.

A interação entre o usuário e o sistema é fundamental para que se estabeleça a comunicação entre ambos e se alcance resultados significativos no processo de busca e recuperação das informações. Para que tal ocorra o bibliotecário precisa conciliar a linguagem utilizada pelo autor do documento a ser indexado e aquela que o usuário presumivelmente formulará em suas expressões de busca.

Guinchat e Menou (1983) afirmam que a eficácia de um SRI pode ser observada tanto sob o ponto de vista das linguagens de indexação como no âmbito do conteúdo de um documento. No primeiro caso a insuficiência do sistema pode ser verificada pelo fato dos descritores/termos não serem suficientemente específicos, pela hierarquia ser insuficiente e/ou pelas relações entre os termos serem inadequadas. No segundo, ela pode ser constatada pela falta de especificidade ou exaustividade, seja pela omissão de conceitos importantes ou pelo emprego inadequado de termos. Esses autores salientam que dois fatores estão fortemente relacionados à eficácia de um SRI: o grau de precisão e o de revocação que o sistema oferece.

Tal como os autores antes citados, Lancaster (2004) reconhece que há várias medidas para avaliar a eficácia de um SRI, mas que precisão e revocação ainda parecem ser as medidas mais adequadas.

Em bases de dados muito grandes torna-se [...] progressivamente mais difícil alcançar um nível de revocação aceitável com um nível de precisão satisfatório, uma situação que chegou a um ponto crítico quando se procura informação na internet (LANCASTER, 2004, p. 4).

Precisão é a capacidade do sistema de impedir a recuperação de documentos não relevantes. Revocação é a capacidade do sistema em assegurar a recuperação de documentos relevantes. Assim, por exemplo, se dos 20 (vinte) documentos que o usuário recupera apenas cinco são de seu interesse, o sistema apresenta um baixo índice de precisão mas um alto índice de revocação, pois há um grande número de documentos recuperados não pertinentes ao interesse de pesquisa do usuário.

Quando se deseja melhorar o coeficiente de revocação, caso ele tenha sido muito baixo ou excessivamente alto, ao refazer-se a busca no sistema é preciso fazer uso de uma linguagem que seja, ou mais genérica ou mais específica, dependendo do propósito que se deseja alcançar. É importante registrar que aumentando a revocação a pesquisa perde em precisão, e que quando se ganha em precisão há uma perda na revocação. Muitas das falhas observadas no processo de busca e recuperação da informação provêm das tarefas de indexação temática de um documento.

Quanto mais precisa é a busca maior será o risco de não se recuperar documentos que tenham sido descritos em um nível mais genérico, mas que mesmo assim possam referir-se ao tema pesquisado, pois quando o número de documentos pertinentes diminui, o silêncio aumenta. O contrário ocorre quando os termos de busca não são tão precisos; nesse caso, quanto maior o número de documentos localizados, maior será o risco de recuperar documentos não pertinentes, aumentando a imprecisão e o ruído.

Entre o grau de especificidade e o grau de precisão que o sistema oferece há uma relação direta. Quanto maior o grau de especificidade, menor o grau de revocação e maior o de precisão; entretanto, quanto maior o nível de generalidade, maior o grau de revocação e menor o de precisão. Guinchat e Menou (1983) afirmam que precisão e revocação são qualidades que se opõem.

O nível de seletividade e de precisão será sempre maior em bibliotecas especializadas. Isso significa dizer que quanto mais especializada for a clientela, maior a necessidade de que

a indexação seja mais seletiva e precisa. A esse respeito cabem duas considerações: como a informação indexada está disponibilizada na Internet, o usuário pode ser tanto um especialista quanto um acadêmico ou até mesmo um leigo que se interesse pelo assunto; entretanto, como o processo de indexação é altamente complexo, é praticamente inviável que se proceda a uma indexação exaustiva.

Araujo (1995), ao tratar das falhas apresentadas pelos SRIs na busca e recuperação da informação, observa que quanto maior o sistema, maior a possibilidade de perda de informação, motivada pela desordem que se instala em seu interior. Tal fato a leva a sugerir que o tamanho do sistema seja reduzido, como uma das alternativas para maximizar o uso da informação e atenuar seus efeitos negativos.

Kuramoto, ao discutir os problemas decorrentes do tratamento e recuperação da informação textual, propõe como abordagem alternativa o uso de sintagmas nominais como termos de indexação. O autor defende a idéia de que “[...] a construção de uma base de dados textual, contendo documentos pertencentes a um só domínio do conhecimento poderá diminuir, ou mesmo evitar, a ocorrência de ambigüidades proporcionando melhor precisão aos resultados de uma busca, o que tornará os SRIs mais eficazes” (KURAMOTO, 1996, p. 185).

Há ainda a considerar, como um dos fatores que geram falhas em um SRI, o fato de se estabelecer uma única forma como ponto de acesso para cada conceito de um domínio. Tal procedimento traz prejuízos para o processo de recuperação da informação, por não considerar o fenômeno da variação, que é próprio não apenas da linguagem natural como também das linguagens especializadas.

Por outro lado, quando o sistema não controla o uso de variantes utilizando o recurso de remissivas, o usuário dificilmente recupera o total de documentos indexados, pois ao localizar textos sobre uma das formas verbalizadas dificilmente fará uma nova pesquisa com uma outra expressão de busca para obter mais documentos. Na maioria das vezes ele não percebe a presença de variação terminológica no sistema. É o caso, por exemplo, de sistemas que utilizam como descritores/termos: *Linguística do Texto* e *Linguística Textual* e *Gramática Gerativa* e *Gramática Generativa*.

Originalmente os estudos sobre os SRIs consideravam o usuário simplesmente como um dos integrantes do sistema e não como a sua razão de ser. Sob esse olhar, ele era alguém

que precisava se adaptar aos recursos que o sistema lhe oferecia; o fato de uma determinada informação servir de modo diferente para cada um deles não era levado em consideração.

Até recentemente os sistemas eram planejados em função de fatores externos ao usuário tais como as tecnologias utilizadas para a sua implementação e o conteúdo informacional a ser inserido. Nessas considerações, as características específicas do consulente, assim como, sua maneira pessoal de perceber, utilizar a informação e produzir conhecimento não eram relevantes para a eficácia do sistema. Até mesmo a maioria daqueles sistemas considerados amigáveis não tinham como proposta resolver os problemas de informação do usuário, mas apenas oferecer-lhes condições facilitadas de uso.

Desde a década de 1980 os estudos desenvolvidos sobre os SRIs mostram a percepção de que a informação só tem sentido quando está inserida em algum contexto, e que o valor que os indivíduos dão a ela é sempre particular e personificado. Estudos recentes começam a manifestar a preocupação em interpretar as necessidades individuais de informação; mesmo assim, muito pouco se tem investigado sobre como as informações recuperadas são usadas pelo usuário.

Conforme afirma Ferreira (1996, p. 220), a “[...] informação é conceitualizada como o sentido criado em um momento específico no tempo e no espaço, por um ou mais indivíduos.” O indivíduo correlaciona, analisa, cria e confere sentido a uma determinada informação; as necessidades de informação mudam no tempo e dependem daquele que a busca. Assim, um SRI deve ser flexível o suficiente para permitir ao usuário, a partir das escolhas feitas, adaptar suas necessidades de pesquisa ao processo de busca e recuperação da informação que o sistema oferece. E, ainda, deve ser orientado de acordo com a natureza das necessidades de pesquisa do usuário e com seus padrões de comportamento na busca e no uso da informação. A sociedade vem se tornando cada vez mais dependente da informação e os mesmos sistemas que filtram, transmitem e distribuem informação, falham por não operarem segundo essa nova orientação.

Os pressupostos básicos que subjazem aos estudos sobre informação centrados no usuário podem ser assim sintetizados: a) a necessidade de informação deve levar em consideração a perspectiva da individualidade do usuário; b) a informação a ser acessada deve estar contextualizada na situação real de onde ela emergiu; c) o uso da informação é determinado pelo usuário.

A partir dessas três constatações considera-se que a informação que o usuário deseja encontrar em um Sistema de Informação, o uso que fará da informação que o sistema lhe oferece e o modo como o sistema pode realmente oferecer-lhe a informação desejada são os aspectos essenciais a serem considerados. Mas, acima de tudo, não se pode esquecer que a busca da informação é um processo dinâmico de tomada de decisão em que o usuário é o agente, quer dizer, é aquele quem seleciona ou rejeita as informações recuperadas.

É fundamental para esta discussão ressaltar a divergência que ocorre entre a linguagem utilizada na indexação, com a finalidade de representar e recuperar as informações registradas nos documentos, e aquela que o usuário faz uso no processo de busca e recuperação da informação. A esse respeito fazem-se duas observações: na maioria das vezes a linguagem utilizada pelo bibliotecário durante o processo de indexação não é de conhecimento daquele que utiliza o sistema; a preocupação com o tratamento da informação é predominantemente quantitativa, em detrimento da informação potencial, isto é, daquela que é adequada às necessidades dos usuários do sistema.

Para o propósito de melhorar a eficácia dos SRIs acredita-se que a Terminologia oferece preciosa contribuição. O crescente interesse da área biblioteconômica pelos estudos terminológicos tem propiciado considerações que sugerem um novo embasamento teórico para a construção dos sistemas. Tais reflexões estão diretamente relacionadas à possibilidade de que os termos técnico-científicos sejam utilizados como descritores, como termos de indexação ou como unidades de informação, viabilizando uma comunicação mais eficiente entre o usuário e os SRIs.

Um planejamento para o desenvolvimento de um SRI faz-se necessário para que se evitem as falhas no que se refere à recuperação de informações. Com esse propósito é preciso aproximar, tanto quanto possível, as linguagens de indexação e de busca e, conseqüentemente, as possibilidades de recuperação que o sistema oferece, das necessidades de pesquisa do usuário.

Planificar contribui para minimizar os problemas relacionados com uma alta revocação e uma baixa precisão, e para evitar ao máximo a possibilidade de deixar de oferecer documentos de interesse do usuário, por uma decisão errada quanto à seleção dos tópicos a serem indexados ou pelo uso de uma linguagem de indexação inadequada.

A especificidade e generalidade, a serem atribuídas aos assuntos de um documento, merecem ser consideradas no planejamento dos instrumentos de recuperação da informação.

Muitas destas ferramentas apresentam uma extensa lacuna em relação ao universo temático de sua área de especialidade, motivada pela falta de especificidade terminológica em seu campo de conhecimento. Especificidade e generalidade dizem respeito ao tópico a ser indexado e se opõem às de exaustividade e seletividade, que estão vinculadas ao tratamento a ser dado ao tema de um documento e não a um determinado item apenas. Especificidade e generalidade se opõem entre si, tal como exaustividade e seletividade.

As decisões a serem tomadas nesse sentido determinam se o tópico indexado vai utilizar um descritor/termo geral ou específico. São importantes, porque trazem conseqüências diretas no processo de busca e recuperação da informação, pois é sabido que uma informação indexada apenas pelo descritor/termo genérico inviabiliza a recuperação de um documento e de uma temática ao nível de sua especificidade. Como resultado de um procedimento inadequado em relação à coleta dos candidatos a descritores/termos e pela falta de uma constante avaliação e atualização do vocabulário controlado, ocorre o que se chama de vazio terminológico, ou seja, a ausência de uma unidade de informação como ponto de acesso de um sistema.

Com base na literatura consultada e mediante observações empíricas, pode-se afirmar que o usuário tem se mostrado insatisfeito com os SRIs. Dentre os aspectos que interferem na eficácia do sistema, com prejuízos para a comunicação com o usuário, é importante abordar as conseqüências advindas da explosão bibliográfica, que levaram os SRIs a atingir proporções que ultrapassaram seus limites de crescimento, fugindo ao controle dos seus gestores. Dentre esses gestores devem ser considerados não apenas o bibliotecário, mas todos os profissionais que lidam com a informação.

Em relação à função dos bibliotecários nesse contexto, Lucas (1996, p. 69) assim se manifesta: “[...] sua eficácia não será julgada em termos de quantidade e qualidade da informação fornecida, e sim a partir do tempo economizado para os usuários. O bibliotecário será, acima de tudo, o ‘refinador’ humano da informação, com a função de criar informação com valor agregado para serviços específicos – informação é conhecimento com valor agregado.”

O excesso de informação processada e armazenada em SRIs gera a recuperação de um número desmedido de documentos, excedendo à tolerância do usuário e, conseqüentemente, interferindo na sua capacidade de transformar em conhecimento a informação recuperada. Fundamentalmente, é possível constatar que o impacto que a informação processada e armazenada em sistemas automatizados representa para o usuário se caracteriza,

prioritariamente, por dois aspectos: pelo acesso a grandes volumes de informações e pela velocidade com que se dá a transmissão da informação, por diferentes e distantes espaços geográficos.

Muitos estudos ainda precisam ser desenvolvidos no sentido de mostrar como as novas tecnologias podem auxiliar no desenvolvimento de sistemas mais eficazes, maximizando o seu uso e minimizando os efeitos negativos provocados pela sua operacionalização de forma desorganizada. A preocupação que predominou nos sistemas automatizados, em sua fase inicial, foi a de acompanhar o crescimento exponencial da informação com a preocupação pela quantidade em detrimento da qualidade, sem questionar as possíveis conseqüências advindas desse comportamento. Nos últimos anos, um volume crescente de informações tem sido registrado nos catálogos eletrônicos, nos mais diversos domínios.

Dentre as alternativas que têm sido propostas para adequar os sistemas às necessidades de pesquisa do usuário está a que prioriza o pensar qualitativamente, de modo a viabilizar uma forma mais eficaz de gerir o conteúdo informacional de um documento. A informação que hoje está disponibilizada em SRIs automatizados encontra-se dispersa, por falta de procedimentos que a considerem sob o ponto de vista qualitativo. O uso inteligente da tecnologia da informação deve contribuir não apenas para elevar ao máximo o uso de documentos, mas até mesmo para buscar novas formas de atender às demandas de uma sociedade pós-moderna, que se caracteriza fundamentalmente por uma informação intensiva e pelo conhecimento especializado.

Nesse novo panorama o sistema deve se preocupar em permitir a recuperação da informação potencial e em evitar a presença de respostas imprevisíveis e indesejadas nos resultados de busca, aproximando documento e informação à demanda dessa nova sociedade que se constitui. Como resultado da automação de bibliotecas inseridas no quadro de redes e sistemas de informação, há a registrar: a) as mudanças que ocorrem no processo de busca e recuperação da informação, advindas de novas formas de identificar, localizar e obter documentos; b) a possibilidade de intercâmbio de dados bibliográficos *on-line* entre os diversos sistemas de informação; c) a integração de diversos tipos de coleções documentais e de diferentes áreas do conhecimento em um mesmo catálogo eletrônico.

Com a informatização da informação, a concepção sobre os SRIs tem sofrido grandes transformações. Nessa nova realidade é preciso saber como acessar informações de modo a recuperá-las de forma fácil, rápida e precisa. Metodologias que viabilizem uma melhor

interface entre usuários e sistemas de informação necessitam ser propostas pelos profissionais da informação. Todo esse quadro que se desenha no âmbito da informação automatizada implica em um olhar diferenciado para a gestão da informação nas bibliotecas, com a transição de uma perspectiva local para a global, e quanto à temática, da generalidade para a especificidade, indo a busca do equilíbrio entre a revocação e a precisão.

As falhas que os Sistemas de Recuperação da Informação vêm apresentando são, fundamentalmente, decorrentes da precária comunicação entre bibliotecário e usuário e do excesso de informação processada e armazenada de forma desorganizada. Sob esse ponto de vista eles vêm falhando em seu objetivo principal, pois não têm conseguido maximizar o uso da informação nem tampouco atender eficazmente às demandas dos usuários.

Nessa nova relação entre o usuário e os SRIs o processo de comunicação entre os dois interlocutores apresenta sérios problemas. Durante muito tempo os Sistemas de Recuperação de Informação foram considerados como um bom instrumento, alcançando um grau de eficiência satisfatório nas pesquisas. Ocorre que, naquela época, o usuário obtinha as informações desejadas através de estratégias de busca, na maioria das vezes estruturadas pelo bibliotecário, que conhecia os recursos de pesquisa que a ferramenta oferecia.

As novas tecnologias informatizadas, entretanto, provocaram um novo paradigma em relação à busca e recuperação da informação e tiveram grande interferência na linguagem documentária. Hoje, quando se pensa na relação que se deve estabelecer entre as linguagens documentárias e os catálogos eletrônicos, o que se torna essencial é compatibilizar e maximizar os recursos oferecidos por ambos.

Nesse contexto, a concepção de um vocabulário controlado para servir a um Sistema de Recuperação da Informação deve viabilizar a possibilidade de indexar e recuperar informações em diferentes níveis de especificidade. Essa decisão vai depender prioritariamente do tipo de informação que o documento veicula e da autoridade que ele representa para a área de conhecimento que cobre.

É preciso, igualmente, definir limites e identificar a área onde um tratamento em profundidade seja essencial e onde uma abordagem superficial da informação será satisfatória, considerando as especificidades próprias dos diferentes materiais a serem indexados. Um resumo de trabalho apresentado em congressos e simpósios, por exemplo, não deve receber o mesmo tratamento dado a um artigo de periódico, pois têm características textuais bem diversificadas. Desse modo, a exaustividade e a especificidade a serem dadas no momento da

indexação devem estar relacionadas não apenas com o sistema a que vão servir, mas também com o documento que está sendo indexado.

Uma outra característica dos SRIs informatizados é que o auxílio do bibliotecário no processo de busca e recuperação da informação torna-se dispensável. Isso não significa dizer que não haja mais lugar para esse profissional nesse novo contexto, pois ele ainda permanece como o mais capacitado para utilizar potencialmente os recursos que o sistema oferece.

3.8 Vocabulário controlado

[...] o tipo de vocabulário controlado (esquema de classificação, cabeçalhos de assuntos, tesouros) não é o fator mais importante na etapa de tradução da indexação. Muito mais importantes são o alcance (abrangência) e a especificidade do vocabulário (LANCASTER, 2004, p. 22).

A eficácia oferecida pelo SRI depende fundamentalmente do processo de indexação, que utiliza uma linguagem documentária e resulta da análise temática dos documentos. A inexistência de um vocabulário controlado e de uma política de indexação compromete fortemente a execução desse processo. Enquanto a política de indexação considera as necessidades de informação dos usuários de um sistema, o vocabulário controlado estabelece a forma de representar tematicamente uma determinada área de especialidade.

Vocabulário controlado é fundamentalmente uma lista de descritores/termos autorizados, visando à recuperação do conteúdo temático de um documento; ele estabelece a organização lógica dos conceitos de uma área do saber. Inclui, em geral, uma forma de estrutura semântica que se destina especialmente a controlar sinônimos, distinguir homógrafos e agrupar termos afins registrado em obras é de uso interno das bibliotecas. Tem o propósito de estabelecer os descritores/termos a serem utilizados pelo bibliotecário no processo de indexação.

É uma ferramenta colocada à disposição do bibliotecário indexador para auxiliá-lo na verbalização dos conceitos extraídos de um documento, e reflete a preocupação que acompanha esse profissional com relação à representação temática. Propicia as condições necessárias para transpor, para a linguagem documentária, a linguagem natural expressa no documento, e para representar de forma resumida as informações materializadas nos textos.

São considerados vocabulários controlados: os tesouros, os esquemas de classificação bibliográfica como a Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU) e a Library of Congress (LC) e as listas de cabeçalhos de assunto; todos

procuram apresentar os descritores/termos tanto na ordenação alfabética como na sistemática, mas se diferenciam em relação a sua estrutura.

Os tesouros têm como objetivo principal o controle do vocabulário. Tiveram origem nas classificações facetadas, que têm como principal característica a análise e a subdivisão dos assuntos em facetas ou focos, considerando-se facetas os aspectos sob os quais um determinado assunto é focado. A tendência recente na construção de tesouros é a de voltar-se para domínios cada vez mais especializados, dirigindo-se a universos bem determinados. A estrutura teórico-conceitual dos domínios específicos é determinada pelo conjunto de descritores/termos do domínio nuclear, sendo representada pela área de especialização propriamente dita e pelos domínios periféricos ou áreas complementares.

A flexibilidade dos tesouros vincula-se a um princípio de utilidade. Desse modo, pode-se construir, para um campo particular do conhecimento, tantos tesouros quantos forem necessários. Cada um deles procurará organizar um dado universo nocional, de acordo com o ponto de vista que se imprime ao domínio, para responder a diferentes necessidades (CINTRA et al., 2002, p. 57).

Nos tesouros os domínios de especialidade são estruturados em classes; o número de classes depende da quantidade de aspectos escolhidos para organizar a área de conhecimento. A classe é determinada pelo conjunto de objetos que possuem as mesmas características. Em alguns tesouros os descritores/termos que equivalem às classes são reconhecidos como *top terms* e não constituem descritores/termos, pois sua função é simplesmente identificar as classes escolhidas para agrupá-los.

Nesses vocabulários as relações hierárquicas que se estabelecem entre os descritores/termos são identificadas pelas expressões termo genérico (TG) e termo específico (TE); alguns tesouros usam ainda termo genérico partitivo (TGP) e termo específico partitivo (TEP) para indicar as relações do tipo todo e parte. Além das relações hierárquicas os tesouros apresentam relações associativas não-hierárquicas; essas relações são identificadas pela expressão termos relacionados (TRs). Ocorre ainda nos tesouros um tipo de relação entre descritores/termos que opera no nível da sinonímia, e que é identificado nas LDs como relações de equivalência.

Os sistemas de classificação bibliográfica, por sua vez, são de natureza enciclopédica e visam cobrir o universo global do conhecimento. Foram elaborados com o objetivo de organizar os acervos de bibliotecas e facilitar o acesso dos usuários à informação contida nos documentos. Podem ser numéricos ou alfanuméricos, sendo usados para representar

tematicamente os documentos e reuni-los por grupos segundo os assuntos que abrangem, dando-lhes um lugar certo na coleção documental e uma localização relativa.

A macroorganização de um sistema de classificação fundamenta-se na organização lógico-hierárquica de suas unidades; a delimitação das classes de assuntos é feita a partir de determinados pontos de vista. Ainda que as classificações bibliográficas se constituam em um importante instrumento de trabalho para o bibliotecário e sejam fundamentais para a organização do acervo de uma biblioteca, elas apresentam alguns problemas estruturais, principalmente em decorrência das áreas de conhecimento não se apresentarem mais de forma tão clara. Hoje existem inúmeros temas interdisciplinares e transdisciplinares que não se encaixam nitidamente em nenhuma das dez classes das tabelas de classificação.

Uma classificação bibliográfica como a CDD ou a CDU até pode alcançar graus de excelência na acomodação das obras de um acervo; entretanto, notações de alta especificidade caracterizam uma interface pouco amigável para o usuário das bibliotecas. Nesses casos eles se defrontam com notações extensas para identificar o conteúdo de um determinado documento; o resultado é que, caso não tenham sido muito bem treinados, perdem tempo e, com muita frequência, só localizam o documento com a ajuda do bibliotecário. Por notação, entende-se o conjunto de símbolos destinados a representar o conteúdo temático de um documento e a indicar a sua localização nas estantes, catálogos e tabelas de classificação. Exemplificando: Análise documentária – 025.4.05.

Tesauros e sistemas de classificação são ferramentas que se estruturam a partir de conceitos formando um sistema. Enquanto os tesauros se desenvolvem a partir do plano das idéias e do plano verbal, permitindo uma boa recuperação temática, os sistemas de classificação se fundamentam no plano das idéias, no plano verbal e no plano notacional, diferenciando-se, aqui, dos tesauros. Pelo plano notacional os documentos são classificados e recuperados de forma eficaz.

Tanto os tesauros como os sistemas de classificação bibliográfica são regulados por uma sintaxe. Nos tesauros a sintaxe se apresenta de forma mais elaborada e se verifica pela utilização dos operadores booleanos; nos sistemas de classificação bibliográfica a sintaxe está presente pelo uso de alguns sinais, como + (mais), / (barra) e : (dois pontos). Tais recursos têm a finalidade de permitir a delimitação mais precisa de um assunto por meio da combinação de seus elementos.

As listas de cabeçalhos de assunto têm uma estrutura similar à dos tesauros, é de base alfabética, mas difere daqueles por não distinguir claramente as relações hierárquicas das associativas. As listas de cabeçalhos de assuntos e os tesauros diferem também pelo fato das primeiras adotarem a pré-coordenação e os segundos, a pós-coordenação. As listas de cabeçalhos de assunto foram criadas para uso em catálogos de assunto e têm como característica a entrada direta. O sistema de cabeçalhos de assunto foi desenvolvido na Biblioteca do Congresso, em Washington (GOMES; MARINHO, 1984).

Os índices que fazem uso de uma linguagem pré-coordenada combinam ou coordenam os termos no momento da indexação. Aqui se enquadram, além das listas de cabeçalhos de assunto, os sistemas de classificação, como: CDD, CDU e LC. As linguagens de indexação pós-coordenadas combinam ou coordenam os termos no momento da busca; delas fazem uso os tesauros.

3.9 Novo paradigma

Sintetizando o que já foi exposto, fazem-se necessárias as seguintes considerações. A experiência demonstra que grande parte da insatisfação do usuário decorre fundamentalmente de dois fatores: da falta de conhecimento da ferramenta que utiliza e do uso de mais de um descritor/termo ou de uma palavra para representar o mesmo conceito ou a mesma informação; tal inconsistência ocorre entre diferentes bases de dados e, até mesmo, em uma mesma base.

Um SRI precisa ser construído a partir de uma linguagem consistente para evitar perda de informação. Se a recuperação de um mesmo tema for feita por descritores/termos diferentes, sem o uso de remissivas, a eficácia do sistema ficará altamente comprometida e os resultados de pesquisa não irão refletir o que ele pode oferecer. Sabe-se que a variação no sistema é inevitável, mas tem de ser controlada pelo uso de remissivas. Mesmo partindo do princípio de que o uso do termo técnico-científico consagrado pelo especialista será a base do SRI, é sempre importante considerar que a variação está presente em qualquer área do conhecimento e, como tal, tem que estar presente também em um SRI.

Em um novo paradigma de acesso à informação, os SRIs devem priorizar as necessidades de um usuário diversificado, cujas perguntas que submetem ao sistema podem ser genéricas ou específicas, podendo estar relacionadas com o acesso ao documento nas estantes ou com o acesso às informações que o documento registra.

Ao propor o desenvolvimento de vocabulários controlados é preciso considerar também a especificidade terminológica da área, seu dinamismo e a contextualização do descritor/termo, que tanto pode ser com o teórico que a representa ou com algum outro descritor/termo que possa oferecer maior precisão sobre a informação que está sendo registrada.

Vale lembrar que para o uso eficaz de um SRI e o conseqüente acesso à informação é importante que o usuário tome conhecimento da linguagem utilizada para a indexação, que tanto pode ser feita pela busca ao índice, que a maioria das bases referenciais oferece, como por um *link* ao vocabulário utilizado para o controle dos termos. As linguagens documentárias, por sua vez, devem permitir uma flexibilidade tal, de forma que possam ser utilizadas para indexar diferentes tipos de documentos, em diferentes níveis de especialização de uma área do conhecimento.

O que se considera fundamental é que o descritor/termo esteja sempre contextualizado e que as expressões de busca sejam representadas por uma única forma, sendo a variação considerada pelo uso de remissivas; que a verbalização temática de um documento seja representativa do conhecimento quanto à sua generalidade e especificidade; e que o usuário tenha acesso à linguagem utilizada pelo sistema. A discussão sobre a pré-coordenação e a pós-coordenação não é objeto deste trabalho. O SRI que está sendo proposto não pretende se posicionar a esse respeito, até mesmo porque essa decisão poderá ser tomada a *posteriori*, pela instituição que dele fizer uso.

A idéia é oferecer um sistema de uma área especializada de domínio que seja flexível, de fácil atualização e uso; que utilize uma linguagem uniforme e consistente capaz de responder às questões genéricas e específicas, atendendo às necessidades de pesquisa desse novo usuário que se apresenta às redes e sistemas de informação.

Num momento em que o usuário tem tantos recursos a seu dispor, não se justifica que ele tenha tanta dificuldade para recuperar informações com rapidez e precisão. Nesse novo contexto de informatização da informação é preciso que os SRIs estabeleçam uma efetiva comunicação entre o bibliotecário e o usuário, e que propiciem a possibilidade dele desenvolver suas pesquisas e acessar os acervos das bibliotecas e centros de documentação com uma independência cada vez maior.

Na área das Ciências Sociais e Humanidades, a falta de instrumentos que ofereçam uma linguagem estruturada e com especificidade terminológica sempre foi um fator agravante

para o desenvolvimento das tarefas de indexação e, conseqüentemente, para a eficácia de um SRI. Tal fato é entendido pela dificuldade que a terminologia dessa área apresenta, uma vez que é altamente polissêmica e fortemente marcada pelo dinamismo terminológico; à medida que a ciência avança surgem novas teorias e novos conceitos são consagrados.

4 TERMINOLOGIA

Cada vez mais, a terminologia assume relevância na e para a sociedade atual. Isto porque, em seus diferentes aspectos, vincula-se muito diretamente aos paradigmas de desenvolvimento econômico, tecnológico e cultural do mundo contemporâneo. Esses paradigmas, por sua vez, estão intimamente relacionados ao processo de globalização, ao acelerado desenvolvimento da ciência e da tecnologia e às novas tecnologias informáticas, com as implicações daí decorrentes, como as exigências de organização e divulgação da informação (KRIEGER, 1998, p. 19).

4.1 Introdução

Este capítulo se propõe a apresentar os fundamentos teóricos da Terminologia e o seu redimensionamento no decorrer dos últimos anos, afirmando-se como uma importante área de saber, fortemente marcada pela interdisciplinaridade. Para tanto traça um percurso histórico enfocando três teorias: a Teoria Geral da Terminologia (TGT), a Teoria Comunicativa da Terminologia (TCT) e a Socioterminologia.

A TCT e a Socioterminologia, mesmo tendo as suas especificidades, compartilham do reconhecimento pleno da linguagem e do papel que os léxicos especializados desempenham na comunicação; elas desenvolvem suas reflexões embasadas nos estudos lingüísticos, especialmente nos relacionados às teorias do texto e do discurso; elas têm no léxico especializado dos sistemas lingüísticos a sua matéria prima e ambas entendem o termo sob uma dimensão lingüística, como uma unidade comunicativa e uma estrutura complexa e poliléxica. A tradicional TGT, no entanto, tem uma visão essencialmente cognitiva e favorece o aspecto conceitual das terminologias.

Dos três objetos que constituem o foco de interesse da Terminologia (termo, fraseologia e definição), os que importam um maior entendimento para os propósitos deste trabalho são o termo e a fraseologia especializada. Dando continuidade à revisão de literatura, considera-se relevante acrescentar algumas considerações sobre a comunicação especializada.

Ao abordar a interface da Terminologia e Informática teve-se como propósito fazer referência às aplicações geradas pela aproximação desses dois domínios. A Lingüística de *Corpus* e a Terminografia oferecem a possibilidade de coletar, tratar e armazenar um grande volume de informações de uma determinada área do saber. A Lingüística de *Corpus* apresenta-se como uma das aplicações dos recursos informatizados e uma importante ferramenta para o tratamento de dados em linguagens documentárias. A Terminografia é reconhecida pela sua prática na elaboração de produtos, como os bancos e bases de dados, a partir do registro de termos de uma área de especialidade. Esses são, dentre os produtos

terminográficos os que mais têm se beneficiado pela relação estabelecida entre Informática e Terminologia.

4.2 Terminologia

A Terminologia, por ser um termo de caráter polissêmico, tem sido concebida como disciplina ou campo de estudos teóricos e aplicados, bem como um conjunto de termos de uma área específica do conhecimento.

Segundo Pavel e Nolet (2002, p. XVII)

En su primera acepción, la palabra **terminología** significa un ‘conjunto de palabras técnicas pertenecientes a una ciencia, arte, autor o grupo social determinado’ [...]

En un sentido más restringido y especializado, el mismo *término* designa una ‘disciplina lingüística dedicada al estudio científico de los **conceptos** y **términos** utilizados en los **lenguajes de especialidad**.’¹¹

Entende-se por linguagem de especialidade ou linguagem especializada aquela que é utilizada em um campo determinado do conhecimento, baseando-se em um vocabulário e em usos lingüísticos específicos desse domínio, com o propósito de propiciar uma comunicação sem ambigüidades.

Cabré (1993) ressalta que nem todos os especialistas concordam em considerar a Terminologia como uma disciplina autônoma, nem tampouco como uma matéria teórica. Para alguns, ela é uma prática vinculada às necessidades sociais e interesses políticos e/ou comerciais; para outros, é realmente uma disciplina científica que toma emprestado de outras ciências, uma série de fundamentos teóricos.

Sager (1993) observa que o interesse pelo estudo terminológico já se manifestava entre os primeiros semânticos, mas que somente no século XX a Terminologia passa a ocupar o lugar de uma disciplina autônoma. Ele a concebe como um conjunto de práticas que tem evoluído em torno da criação de termos, de sua compilação e explicação, e de sua apresentação em formato impresso ou meios eletrônicos; por mais arraigadas que estejam essas práticas, não constituem uma disciplina.

¹¹ Tradução da autora: Em sua primeira aceção, a palavra **terminologia** significa um ‘conjunto de palavras técnicas pertencentes a uma ciência, arte, autor ou grupo social determinado’ [...] Em um sentido mais restrito e especializado, o mesmo *termo* designa uma ‘disciplina lingüística dedicada ao estudo científico dos **conceitos** e **termos** utilizados nas **linguagens de especialidade**’.

O autor afirma que os aspectos essenciais desse campo de conhecimento são explicados mais apropriadamente no contexto lingüístico das Ciências da Informação e da Lingüística Computacional, e acrescenta que não existe uma documentação substancial que dê suporte à proclamação da Terminologia como uma disciplina independente.

Outro pioneiro dos estudos modernos da Terminologia no Canadá, Rondeau (1984), afirma que a Terminologia, como tal, não se trata de um fenômeno recente; o que é novo, mas nem por isso alarmante, é o crescimento progressivo das terminologias. Desde a Antigüidade é possível situar a presença de línguas de especialidade de uma área do conhecimento, haja vista, por exemplo, a terminologia utilizada pelos filósofos gregos, pelos comerciantes cretas e na arte militar. Tão logo o homem utiliza a linguagem para se comunicar, cria palavras para expressar conceitos e denominar objetos e processos de diferentes domínios.

O que é atual, ou seja, da segunda metade do século XX, conforme mencionado por Krieger e Finatto (2004), é o aparecimento de um campo de estudos dedicado à Terminologia, em uma tentativa de fugir da polissemia e das ambigüidades próprias do léxico comum. Ao descreverem conteúdos específicos, as terminologias estão mais propensas a alcançarem a univocidade e a precisão conceitual. Tal propósito resulta do esforço em propiciar uma adequada comunicação do conhecimento e, conseqüentemente, um eficiente processo comunicacional entre os especialistas. Os termos, por transmitirem conteúdos específicos de seu campo de conhecimento, estão a serviço da comunicação especializada; o comportamento sistemático é uma característica marcante das terminologias contemporâneas.

A precisão conceitual que caracteriza as comunicações especializadas favorece a univocidade e torna-se uma condição necessária para um eficiente intercâmbio comunicativo, tanto no âmbito jurídico ou comercial ao firmar relações contratuais, bem como em situações múltiplas e diversas de intercâmbio científico, tecnológico e cultural (KRIEGER, 2004). Uma característica marcante das terminologias contemporâneas é o comportamento sistêmico.

Ao ser revisitado criticamente, conforme observam Krieger e Bevilacqua (2005), a Terminologia avançou porque se permitiu percorrer novos caminhos e aparelhar-se teórica e metodologicamente para enfrentar situações inicialmente não previstas, desencadeadas fundamentalmente pela proliferação dos termos técnico-científicos e pela crise do conhecimento que, quanto à sua estrutura, não mais se submete a um fechamento, mas se constitui em redes. A proliferação de novas terminologias veio revelar a inexistência de fronteiras rígidas entre o léxico geral e o especializado, tornando-se cada vez mais complexa a tarefa de reconhecimento de unidades terminológicas especializadas.

Com o propósito de superar as barreiras e compreender as transformações porque vêm passando o léxico especializado e as linguagens que o comportam, foi possível conferir à Terminologia uma identidade própria, de forma a constituir-se como uma efetiva área de conhecimento e, nessa medida, de reflexões e práticas inter-relacionadas. Fundamentando-se em princípios da Lingüística e dialogando com diferentes campos de especialidade a Terminologia buscou, em novas fontes, o auxílio para identificar e explicar a gênese e o funcionamento de seus objetos. A esse respeito, acrescenta-se o fato dela abrir-se para outras áreas de estudos sobre a linguagem, como as desenvolvidas pelas teorias funcionalistas, cognitivas e por aquelas que a focalizam no texto e no discurso, tal como a Lingüística de *Corpus* e as metodologias de pesquisa baseadas em *corpus*.

A Terminologia tem avançado muito nos últimos anos; ela tem vivenciado uma mudança tanto em relação à teoria como à prática terminológica. O reconhecimento do descompasso que se evidencia entre a teoria e a prática tem demonstrado que a proposta tradicional de exclusividade denominativa e de monossemia não se confirma ante o funcionamento real da linguagem. Os estudos desenvolvidos neste sentido têm demonstrado que nem sempre a correspondência que se estabelece entre a denominação e a noção é de univocidade e monorreferencialidade, mesmo em se tratando de uma mesma área de conhecimento.

Conforme afirma Cabré (1998), neste novo panorama em que a Terminologia se coloca é significativo o desenvolvimento das áreas que a configuram como interdisciplinar. Mas o fenômeno terminológico decorre, também, em função de aspectos socioeconômicos e sociopolíticos, tais como: do desenvolvimento da mídia, das relações políticas internacionais e do comércio internacional. Todos esses fatores determinam o aparecimento e a proliferação de multinacionais, assim como a necessidade de padronizar e normalizar as terminologias.

Sager (1993) acredita que por ter a Terminologia um caráter eminentemente interdisciplinar é necessário delimitá-la e relacioná-la com outras disciplinas, e que mesmo sendo de cunho essencialmente lingüístico e semântico ela encontra uma forte motivação no campo de estudos relacionados com a comunicação.

Wüster, ao estabelecer as bases teóricas de sua teoria, já assinalara essa interdisciplinaridade ao perceber sua relação com a Lingüística, a Lógica, a Ontologia e as Ciências da Informação. O elemento comum a essas disciplinas diz respeito ao fato de que todas se ocupam das relações entre os conceitos e os termos. Lembra-se, aqui, que a área da

Documentação é um campo de conhecimento que se insere no quadro das Ciências da Informação.

Para Cabré (1998) a Terminologia, enquanto disciplina aplicada, deve oferecer uma metodologia de trabalho coerente com os princípios da teoria e ser suficientemente flexível para adequar-se às situações e aplicações diversificadas. Esta nova teoria que se faz necessária, segundo Cabré (2001), resulta da necessidade de considerar uma diversidade de facetas que deve integrar seus fundamentos e suas aplicações. Tal diversidade se evidencia nas percepções que diferentes grupos de profissionais têm da Terminologia, tais como: lingüistas, cientistas de variadas áreas, documentalistas, normalizadores, redatores e tradutores, dentre outros.

Até recentemente os lingüistas percebiam a Terminologia apenas pelo seu caráter prescritivo. As unidades terminológicas eram artificiais e como tal não estavam sujeitas aos fenômenos da linguagem natural. Essa situação começa a mudar com o desenvolvimento de pesquisas que não olham mais os termos como simples palavras estáticas que figuram nos dicionários, mas como unidades léxicas especializadas e dinâmicas, que formam parte do léxico de um falante enquanto especialista de uma área de conhecimento.

Segundo os especialistas as unidades terminológicas fazem parte da linguagem natural e da gramática que as descreve. Nada mais são do que unidades de organização do conhecimento especializado de uma área do conhecimento e, como tal, unidades de expressão e comunicação usadas de acordo com as situações estritamente profissionais. Entre os especialistas estão os usuários que admitem graus diferentes de variação semântica e denominativa no uso dos termos, segundo as circunstâncias discursivas em que são utilizados.

Mas para os documentalistas e normalizadores internacionais a Terminologia é um conjunto de unidades padronizadas por consenso entre as pessoas que delas fazem uso, a fim de reduzir a ambigüidade e assegurar a univocidade e precisão comunicativa. Nesse contexto, os princípios da univocidade e monossemia das unidades terminológicas alcançam seu sentido pleno.

Essa pluralidade de concepções, afirma Cabré (2001), é consequência lógica de sua multifuncionalidade e do caráter poliédrico e multidimensional de suas unidades. Ainda segundo a autora, enquanto campo de investigação científica, a Terminologia se caracteriza pela sua interdisciplinaridade e transdisciplinaridade, e enquanto disciplina aplicada serve a

grupos profissionais diferentes e a finalidades diversas, sempre conservando sua função de representação e de comunicação do conhecimento especializado.

Cabré (1998) entende que a Terminologia, por ser de caráter interdisciplinar e transdisciplinar, deve esboçar um modelo teórico que seja capaz de descrever as unidades terminológicas em toda a sua complexidade, localizando-as em uma teoria mais ampla, de caráter multidimensional.

A Terminologia é uma disciplina interdisciplinar, na medida em que é constituída pelo conjunto das contribuições advindas de diferentes ciências, como: Ciências da Linguagem, Ciências Cognitivas e Ciências da Comunicação. Como consequência da sua interdisciplinaridade, a unidade terminológica se define como multidimensional e, como tal, é concebida como lingüística, cognitiva e comunicativa, isto é, como unidade léxica, unidade de conhecimento e unidade comunicativa, configurando seu caráter poliédrico.

Nesse entendimento há a possibilidade de identificar na teoria terminológica três dimensões: a cognitiva, a lingüística e a comunicativa. A dimensão cognitiva é a que relaciona a forma lingüística ao conteúdo conceitual; a dimensão lingüística é a que examina as formas existentes e potenciais de apresentação das terminologias; e a dimensão comunicativa é a que analisa com atenção o uso das terminologias. Assim considerado, o estudo da Terminologia sob a dimensão cognitiva requer uma compreensão da estrutura do conhecimento, de forma a obter-se uma imagem coerente da natureza, do comportamento e da interação entre conceitos e termos associados. A perspectiva lingüística exige que os termos sejam vistos como unidades da língua natural com todas as suas propriedades e funções; a visão comunicativa encara a contribuição do léxico para a realização do discurso no universo especializado.

Presente em todas as disciplinas, a Terminologia é transdisciplinar, uma vez que todas as áreas de especialidade possuem e utilizam um repertório de termos para representar seus conhecimentos. Assume também uma feição multidisciplinar, porque para ela convergem disciplinas lingüísticas, disciplinas das Ciências Cognitivas e das Ciências da Informação. É esse caráter multidisciplinar que leva alguns estudiosos a julgar que ela não é uma disciplina autônoma.

Vale ressaltar que o interesse pelas terminologias não é mais exclusivo dos especialistas de um domínio do conhecimento, seus usuários primeiros e que sempre tiveram a consciência da necessidade e importância em dominar tal conhecimento. Ele se manifesta

tanto em alunos de graduação e de pós-graduação que precisam adquirir o domínio sobre a sua área de interesse, como também em profissionais envolvidos com o uso da linguagem, dentre eles os documentalistas, os tradutores, os intérpretes, os redatores e os profissionais da mídia. O próprio cidadão comum, como leigo em qualquer dos campos especializados, é afetado em seu cotidiano pelas terminologias, tendo em vista que a sociedade sofre o impacto da acelerada produção do conhecimento.

O interesse cada vez maior da sociedade pelo conhecimento científico tem sido impulsionado pelos meios de comunicação de massa e pelas novas tecnologias da informação. Nos dias atuais a ciência é objeto de larga divulgação; o conteúdo especializado técnico-científico, de uso restrito aos profissionais da área, tornou-se de interesse do público não especializado, uma vez que é o objeto cotidiano de notícia na mídia.

Pela facilidade de acesso que tem à informação especializada, a sociedade também sofre o impacto da acelerada produção do conhecimento. As terminologias, pelas suas existências e circulação em cenários comunicativos diversos, evidenciam que cumprem prioritariamente uma dupla função. A primeira é, sem dúvida, a de fixação do conhecimento técnico-científico; a segunda, não menos importante, é a de promover e facilitar a comunicação desse conhecimento, de modo pontual, por meio da comunicação direta entre os especialistas, ou ainda, por divulgá-lo pelos diferentes veículos de informação. Também não se pode esquecer que, do ponto de vista econômico e comercial, o uso adequado da terminologia especializada é reconhecido como uma das razões que ampliam as condições de competitividade no mundo globalizado.

Com efeito, o surgimento e o considerável e crescente aumento de novos termos é uma conseqüência direta do desenvolvimento extremamente rápido das ciências puras e aplicadas, assim como da tecnologia e das relações internacionais, políticas e comerciais. Tal criatividade requer novas denominações para as novas descobertas. Por isso, o aparecimento de um grande número de produtos, procedimentos, técnicas e conceitos têm levado os seus criadores e usuários a criarem denominações com o propósito de assegurar o intercâmbio das informações.

Os cientistas, conscientes da dinamicidade e transitoriedade das terminologias, impulsionados pelo acelerado avanço da ciência e da tecnologia e pela internacionalização das ciências, dedicam-se à busca de estratégias que assegurem a univocidade da comunicação em âmbito nacional e internacional. Tal fato tem contribuído para a multiplicação de terminologias e para a conseqüente criação de um número significativo de termos. Como

resultado dessa nova realidade que se tornou evidente nos últimos anos, é cada vez maior o interesse pela padronização das terminologias, que também sofrem a influência do processo de globalização que marcou o final do século XX.

A crescente preocupação com a utilização e a tradução das terminologias justifica-se pela necessidade de estabelecer, entre as nações, não apenas uma comunicação mais eficiente, mas também uma adequada transferência de tecnologia e a realização mais produtiva e eficaz de contratos comerciais e acordos socioculturais em âmbito internacional.

É todo esse conjunto de fatores que provoca a expansão dos léxicos especializados e a evolução de uma consciência sobre o papel que eles desempenham na comunicação. É em decorrência desse novo panorama cultural que vai se fortalecendo no decorrer do século XX que, nas mais distintas áreas do conhecimento, desenvolvem-se reflexões e constroem-se teorias a respeito da comunicação especializada.

4.3 Histórico

Fazendo uma breve retomada histórica, cumpre lembrar que as bases teóricas da Terminologia estavam intimamente relacionadas com propósitos pragmáticos, em vista do objetivo de favorecer a comunicação das ciências no âmbito internacional. Por isso, dentre os pioneiros dos estudos terminológicos que surgiram a partir da segunda metade do século passado, distinguem-se aqueles que privilegiaram o enfoque cognitivo do fenômeno do termo, daqueles que salientaram a visão do funcionamento lingüístico das terminologias.

Dentre os que privilegiaram o enfoque cognitivo do termo estão as escolas clássicas, como a de Viena, a de Praga e a escola russa. Nessas prevalece uma perspectiva normativa sobre as terminologias, fato que se contrapõe às linhas teóricas de base descritiva sobre o léxico especializado e que ganhou impulso com o desenvolvimento da Lingüística. Os estudos desenvolvidos nessas escolas prestaram uma relevante contribuição para a construção das bases teóricas da disciplina.

Sob uma orientação de valorização da dimensão cognitiva do termo, visando à sua padronização, é que surgiram importantes reflexões. Dentre elas, a Teoria Geral da Terminologia, teoria desenvolvida por Eugen Wüster, engenheiro eletro-técnico austríaco, considerado o fundador da Terminologia moderna. Foi ele quem propôs os princípios que se tornaram o pilar referencial dos estudos terminológicos.

Seu propósito era normatizar as terminologias, com o fim de facilitar o uso e assim garantir a perfeita comunicação técnica e científica em âmbito internacional. Sob seu ponto de vista tal objetivo seria alcançado em decorrência da univocidade obtida pela normatização. De acordo com o princípio da univocidade, um conceito é designado por só um termo e um só termo se refere a um só conceito. Nessa ótica fica rejeitada a idéia da variação lingüística, da sinonímia, da polissemia e da homonímia.

Wüster e outros terminólogos, como os russos D. S. Lotte e E. Drezen, preocuparam-se em estabelecer orientações metodológicas para o tratamento das unidades terminológicas com base no princípio de que os termos são denominações de conceitos. A partir desse entendimento tornou-se claro, para os três teóricos, que os elementos essenciais da comunicação profissional são os conceitos e os signos a eles associados, cuja precisão deve ser assegurada por meio de léxicos padronizados.

Com a TGT Wüster visava estabelecer um campo de conhecimento com fundamentos epistemológicos e objeto próprio de investigação, o termo técnico-científico. Mesmo havendo consenso entre os teóricos sobre a relevância de seus estudos para o estabelecimento da Terminologia como disciplina, Wüster declara, na abertura do Simpósio do INFOTERM em 1975, que tal reconhecimento deve-se a quatro homens, além dele próprio: ao alemão A. Schloman, por ter sido o primeiro a considerar o caráter sistemático dos termos; a Ferdinand Saussure, por ter sido o primeiro teórico a registrar a sistematicidade das línguas; ao russo E. Drezen, por seu pioneirismo ao destacar a importância da normalização e a impulsionar a organização da ISO; e ao inglês J. E. Holmstrom que, da UNESCO, estimulou a difusão internacional das terminologias.

Dentre os princípios de maior relevância dessa teoria destaca-se aquele que afirma que os termos expressam conceitos e não significados, pois os significados são lingüísticos e variáveis conforme o contexto discursivo e pragmático; os conceitos, por sua vez, são atemporais, paradigmáticos e universais. A partir de então os termos não são entendidos como elementos das línguas naturais, mas como unidades de conhecimento que comportam denominações. Os conceitos científicos, por sua vez, são identificados por meio de rótulos, isto é, por meio de etiquetas denominativas criadas com determinadas peculiaridades que possibilitam evitar as ambigüidades que caracterizam o léxico comum. Desse modo, na perspectiva clássica wüsteriana as unidades lexicais especializadas estão isentas de polissemia, uma vez que não comportam diversidades conceituais; elas ocupam um lugar determinado em uma hierarquia lógica do conhecimento.

Em razão de suas proposições prescritivas e normalizadoras, a TGT tem sido muito criticada, pois seus princípios, que refletem uma visão redutora e idealista da teoria, conduzem ao apagamento dos aspectos comunicativos e pragmáticos, e envolvem apenas o léxico temático. Ao final do século XX, entretanto, a Terminologia passa a ser marcada por investigações de base lingüístico-comunicacional. Nesse período surgem novas concepções sobre os termos e seu funcionamento, bem como discussões sobre a tarefa que compete à teoria terminológica. Daí em diante, a valorização do componente cognitivo, que marcou a TGT, cede lugar aos estudos de fundamentação lingüística.

Na década de 80 do século XX, Alain Rey (1992, p. 3), um dos pioneiros nesse novo percurso teórico que trilha a Terminologia, chama a atenção para um novo paradigma. Ele inicia sua tese de doutorado sobre a Terminologia com a seguinte afirmação: “A l’origine des réflexions sur le nom et la nomination, base de la terminologie, se trouve toute la réflexion sur le langage et le sens.”¹²

A partir desse novo paradigma a Terminologia começa a se ocupar também com as questões lingüísticas e não mais se limita a tratar o termo de forma idealizada, restrito ao âmbito do especialista; a unidade terminológica passa a ser compreendida sob um ponto de vista descritivo, funcionando como qualquer unidade do léxico da língua natural. Surgem novas teorias e diversos estudos sob esse novo enfoque, abordando a complexidade lingüística e comunicacional das terminologias sem abandonar seu aspecto cognitivo.

O percurso revisionista, trilhado pelos teóricos que se opuseram aos princípios estabelecidos pela TGT, caracterizou uma ruptura epistemológica significativa na história da Terminologia. As manifestações desse rompimento estão presentes nas novas concepções teóricas assumidas a partir de então. Os termos passam a serem entendidos como unidades terminológicas que funcionam num modelo de comunicação. A linguagem se apresenta como dinâmica e complexa, levando a uma descrição das terminologias com base em seu comportamento nos textos especializados e ao conseqüente reconhecimento de que a polissemia também está presente no universo das comunicações científicas e técnicas.

As proposições inovadoras, que se desenvolveram sob o enfoque do funcionamento lingüístico dos termos, discordam das teorias clássicas também por se recusarem a aceitar a dicotomia pensamento e linguagem. Sob essa nova visão o foco prioritário é o termo,

¹² Tradução da autora: Na origem das reflexões sobre o nome e a denominação, base da Terminologia, se encontra toda a reflexão sobre a língua e o sentido.

entendido como unidade lingüístico-comunicacional; o componente conceitual passa a ter interesse na medida em que repercute sobre a própria identificação do termo.

As novas teorias, consideradas sob um ponto de vista amplo, mesmo ocupando lugares distintos na Terminologia, têm em comum dois pontos principais. O primeiro é o fato de reconhecerem o pleno funcionamento da linguagem no contexto da comunicação; o segundo é o de criticarem a postura redutora na apreensão do fenômeno terminológico. Suas reflexões se fundamentam no sensível avanço dos novos desenvolvimentos na área da Lingüística. Vale salientar que a Lingüística tem sido a grande responsável pelo redimensionamento teórico por que tem passado a Terminologia, ocorrido principalmente a partir dos anos 90 do século XX.

Os novos direcionamentos têm motivado o desenvolvimento de estudos terminológicos sob um olhar descritivo e estão sendo embasados fundamentalmente na reflexão lingüística, textual e comunicacional, tendo no léxico especializado a sua matéria prima. A partir desse novo conjunto de saberes, que vem dando à Terminologia uma nova identidade, diversas e diferentes investigações têm sido feitas com a finalidade de apreender o fenômeno terminológico.

É importante registrar que o acelerado avanço da Informática tem prestado uma enorme contribuição aos estudos de Terminologia; inúmeros programas de computador têm sido concebidos para permitir a recuperação automática de terminologias. Os recursos informatizados têm permitido o desenvolvimento de estudos com base em *corpora* lingüísticos sobre a geração de bancos terminológicos; a coleta, a seleção, o registro e a análise de um grande volume de dados lingüísticos têm sido possível com relativa facilidade. De outro lado, vale lembrar que os estudos terminológicos trazem uma contribuição inestimável para a Informática, de tal modo que se estabelece um intercâmbio produtivo entre as duas áreas.

A esse respeito, já em 1970, o pioneiro Wüster estava convencido de que as informações contidas nos glossários, reunindo o conjunto de termos técnico-científicos conhecidos, seriam armazenadas em computadores e a informação terminológica estaria acessível a todos, nas diversas partes do mundo, através de grandes redes cibernéticas. Tal imagem, que parecia uma antevisão futurista, é hoje realidade na *internet*, graças à qual se encontram disponibilizados aos usuários os bancos de dados terminológicos de grande porte e dicionários *on line*, sem falar nas bibliotecas virtuais, que disponibilizam catálogos e obras.

A Terminologia wüsteriana, de orientação prescritiva e normativa, recomenda o procedimento onomasiológico de identificação dos termos, segundo o qual, primeiramente são identificadas as noções ou aspectos conceituais e, a partir daí, buscam-se as formas lexicais, como meras etiquetas denominativas desses conteúdos, tais como rótulos que se colocam em frascos alinhados em uma prateleira. Sob esse olhar, o domínio dos conceitos é anterior à língua e, portanto, existe independentemente do processo de denominação. O termo, como unidade de conhecimento, fundamenta a idéia de dissociação entre denominação e conceito e justifica a crença na exclusividade do princípio onomasiológico para o seu reconhecimento.

A visão que prioriza o aspecto lingüístico não acolhe essa dissociação, pois entende que o acesso ao plano cognitivo do conteúdo se faz via componente lexical. Por isso, aqueles que defendem a revisão da teoria tradicional entendem que o método semasiológico e o método onomasiológico coexistem no processo de formação das terminologias. Um conceito só começa a existir quando é referido por uma forma lingüística, o termo. Um conceito completamente anônimo não se desenvolve no intelecto do cientista; de alguma maneira ou de outra ele vai usar da sua capacidade lingüística para chamá-lo à existência. Reforçando esse entendimento há a considerar as situações em que um novo produto e um novo processo, antes mesmo de receberem uma denominação específica, costumam ser identificados genericamente como máquina ou objeto.

No quadro de revisão e atualização dos estudos terminológicos é importante fazer referência à Teoria Comunicativa da Terminologia, TCT, da qual Cabré é a principal teórica; ela está à frente de um grupo de pesquisadores do Instituto de Lingüística Aplicada da Universidade de Pompeu Fabra, de Barcelona. Entre os principais pontos questionados, são mencionados: a prioridade da padronização dos termos, a concepção da unidade terminológica, separando conceito e significado; a supervalorização da função denominativa e a pouca consideração pelas estruturas morfológicas e aspectos sintáticos das unidades lexicais.

Para os defensores da revisão da teoria clássica, sua insuficiência resulta de seu caráter idealista e reducionista. Aspectos fundamentais do léxico especializado como a poliedricidade denominativa, cognitiva e funcional que se realiza na dupla função representativa e comunicativa, não são considerados. Seus elementos operativos partem de uma concepção idealizada do mundo e da linguagem. A comunicação não é vista como uma atividade *in vivo* em função do contexto sociocultural, lingüístico ou científico em que se

situa. Ao contrário, é um produto de laboratório sujeito aos condicionamentos artificiais, realizada *in vitro*, ignorando a variação inerente à diversidade aplicada determinada pelas características pragmáticas do evento comunicativo em situação real. Além do mais, a TGT é acusada de não levar em consideração a pluralidade tipológica dos trabalhos terminológicos causada pela diversificação das necessidades sociais, nem tampouco a multidimensionalidade e dinâmica constante do conhecimento especializado.

A TCT articula-se pela valorização dos aspectos comunicativos das linguagens de especialidade, bem como pela compreensão de que as unidades terminológicas fazem parte da linguagem natural e da gramática das línguas. Ela se define a partir de alguns pressupostos, tais como: a) a Terminologia se concebe como um campo interdisciplinar construído a partir da teoria do conhecimento, da teoria da comunicação e da teoria da linguagem; b) seu objeto de estudo são as unidades de significação especializada (USEs); c) os termos são unidades léxicas que se compõem de forma ou denominação e significado ou conteúdo, e que têm valor referencial, sendo determinado pelo âmbito, tema, perspectiva de abordagem do tema, tipo de texto, emissor, destinatário e situação; d) o conteúdo de um termo nunca é absoluto, senão relativo, segundo cada âmbito e cada situação de uso; e) os conceitos de um mesmo campo especializado mantêm entre si relações de diferentes tipos, sendo que o conjunto das relações que se estabelecem entre os conceitos constitui a estrutura conceitual de uma disciplina; f) o valor de um termo se estabelece pelo lugar que ele ocupa na estruturação conceitual de uma disciplina, segundo critérios previamente estabelecidos; assim, os termos não pertencem a um campo, mas são usados em um campo de conhecimento com um valor singularmente específico.

Dentre os novos direcionamentos que começam a aparecer é relevante acrescentar a contribuição de François Gaudin (1993), da Universidade de Rouen, França. Sua proposta é a Socioterminologia e seu grande mérito é ter impulsionado, a partir dos princípios da Sociolingüística, o exame do funcionamento dos termos em seu real contexto de ocorrência. Desse entendimento resulta o reconhecimento da presença da variação terminológica e da sinonímia nas comunicações especializadas, fenômenos não admitidos pelos estudos clássicos da Terminologia e da Terminografia de cunho prescritivo.

O reconhecimento de que a variação existe também nas linguagens de especialidade contrapõe-se à ilusão de que é possível controlar seu uso por uma intransigente padronização. Defender tal padronização é negar e apagar o verdadeiro funcionamento da língua em contexto de uso. A experiência da França e do Canadá de fala francesa tem demonstrado que

os modelos normalizadores rígidos produzidos com a intenção de controlar o vocabulário tornam-se inoperantes.

Gaudin (1993) postula a variação como eixo central da sua teoria e critica as escolas clássicas por não a considerarem, impedindo o registro das outras formas denominativas e conceituais que os termos comportam com relativa frequência. Aplicando o princípio da variação em toda a sua dimensionalidade, o termo passa a ser visto como a unidade lexical que sofre todas as implicações sistêmicas e contextuais próprias a qualquer palavra no evento comunicativo da língua.

Como variantes terminológicas consideram-se, dentre outras, as variantes: gráfica, lexical, morfossintática, socio-profissional e topoletal ou geográfica. Segundo Morel e Rodriguez (2001), o fenômeno pode ser determinado por fatores de espaço, de tempo e de condicionamentos sócio-culturais.

Gaudin (1993) alerta para a necessidade de se efetivar o diálogo interdisciplinar entre as áreas de conhecimento que são afetadas por problemas ligados à Terminologia. A esse respeito, critica instrumentos como glossários e dicionários técnicos elaborados segundo as recomendações wüsterianas, por não expressarem a realidade dos usos terminológicos. Nesse contexto, aponta para a necessidade de ser efetivado o diálogo interdisciplinar entre as áreas de conhecimento afetadas pela problemática terminológica. Assim, acredita ser possível, sob uma nova visão, propiciar à Terminologia condições para avançar nos aspectos ligados à teoria terminológica e à prática terminográfica.

Conforme observa Faulstich (1995), a denominação Socioterminologia apareceu pela primeira vez com Jean-Claude Boulanger, no início da década de 80 do século XX. Segundo Gaudin (1993a), pouco tempo depois é mencionado por Pierre Lerat e Monique Slodzian; mas é com Yves Gambier que começam as primeiras manifestações para que a Terminologia se transformasse rapidamente em uma Socioterminologia, pois até então aparecia na literatura apenas como um neologismo da língua. No caso da Socioterminologia, pode-se afirmar que o termo precedeu a seu conteúdo nocional.

Enquanto disciplina eminentemente prática se fundamenta na análise das condições sociais e lingüísticas de circulação dos termos. Como disciplina descritiva estuda o termo sob a perspectiva lingüística na interação social e busca na Sociolingüística e na Etnografia uma fundamentação teórica. Na Sociolingüística busca critérios de variação lingüística dos termos no meio social e a perspectiva de mudança; na Etnografia, os princípios que se referem às

comunicações entre os membros da sociedade, capazes de gerar conceitos interacionais de um mesmo termo ou de gerar termos diferentes para um mesmo conceito.

A Socioterminologia começa a observar e diferenciar a comunicação que ocorre entre o pessoal de direção, os setores administrativos, de pesquisa, de produção e de comercialização das empresas, por exemplo, como uma forma mais adequada de descrever o termo técnico e científico. Visto sob uma interpretação variacionista, que considera a diversidade presente nas comunicações especializadas, o termo deixa de ser considerado como uma entidade unívoca.

Em qualquer linguagem existe uma gama de variações. Dentre as variações possíveis, a norma social atua para determinar os critérios de seleção de códigos, cujas propriedades fonológicas, gramaticais e léxicas, podem ser funções da situação em que tem lugar a comunicação. De modo geral a variação de nível fonológico e gramatical é mais evidente nas variantes regionais e sociais, sendo, portanto, sem interesse para a Terminologia. É a variação do léxico que se apresenta de forma mais apropriada nas linguagens de especialidade, o subsistema lingüístico selecionado pelo indivíduo, cujo discurso se centra em um campo temático em particular.

Do ponto de vista metodológico a Socioterminologia tem origem em uma atitude descritiva. Com fundamentação na Lingüística, que é uma ciência essencialmente descritiva, propõe que os termos sejam investigados em sua dimensão interativa e discursiva; nessa análise sua preocupação dirige-se ao nascimento, formação, consolidação e inter-relação, sempre sob uma perspectiva lingüística e de interação social. Entretanto, é o texto que considera o seu objeto de estudo privilegiado, pois é pelo discurso especializado que a Socioterminologia manifesta seu maior interesse e é nele que tenta identificar os aspectos que condicionam a escrita dos textos e a sua leitura. É o texto e o discurso que revelam o tipo de terminologia empregado nas comunicações especializadas, o maior ou menor grau de especialização que as comunicações especializadas refletem e o grau de densidade informativa que comunicam.

Como prática de trabalho terminológico a Socioterminologia fundamenta-se na análise das condições de circulação do termo no funcionamento da linguagem. Utilizando os procedimentos etnográficos a pesquisa socioterminológica observa as características da instituição em que a terminologia é gerada e examina as características da terminologia empregada pelo pessoal, segundo o posto que ocupa e a formação profissional e qualificação.

Dessa maneira, ela coloca em foco a competência do falante, os usos lingüísticos e o registro da variação lingüística na Terminologia.

Em síntese, a Socioterminologia propõe um estudo *in vivo* nas linguagens de especialidade; ela reconhece a existência da variação lingüística nos discursos especializados, sob a forma de sinonímia, homonímia e polissemia; e não deixa de conceber o termo como um signo lingüístico no sentido que lhe dá Saussure, ou seja, como uma unidade indissolúvel constituída de conteúdo (significado) e de expressão (significante), sem que um seja prioritário ou preceda o outro.

Segundo as proposições inovadoras de teorias terminológicas como a TCT e a Socioterminologia, que se fundamentam sob o enfoque do funcionamento lingüístico dos termos, a relevância atribuída ao texto é uma realidade e está atrelada ao princípio comunicacional que postulam.

O reconhecimento do texto como o ambiente natural das terminologias especializadas representa uma importante reversão de paradigmas epistemológicos. O quadro referencial de exame do comportamento e da gênese dos termos passa a ser atribuído ao seu contexto de ocorrência. Sob essa ótica a Terminologia avança no sentido de se recusar a admitir o léxico especializado como um constructo idealizado, sendo que os termos passam a ser concebido como elementos da linguagem em funcionamento, com todas as implicações daí decorrentes. Sob o enfoque que valoriza o contexto discursivo das comunicações especializadas, a Terminologia se aproxima das teorias que tomam o texto como seu objeto de análise, tais como da Lingüística do Texto e das teorias pragmáticas e enunciativas.

A partir desse novo direcionamento, os estudos terminológicos, além de reconhecerem o papel da textualidade e do discurso na constituição dos objetos terminológicos, começam também a interessar-se pelas estruturas e tipologias de textos produzidos nas comunicações especializadas. Os teóricos começam a perceber que, com base em uma abordagem textual das ocorrências terminológicas, é possível observar o comportamento do termo sob vários planos e ângulos.

Dentre os teóricos que reivindicam o papel do contexto para o reconhecimento do termo, cabe citar Jennifer Pearson. Ao reconhecer que uma palavra adquire o estatuto de termo em razão do cenário comunicativo em que se apresenta, a autora postula uma abordagem pragmática e enfatiza a importância do contexto na configuração do termo. A esse respeito, assim se manifesta: “We further suggest that all other efforts to define what a term is

and to examine what distinctions, if any, exist between different types of terms are irrelevant if context has not been considered”¹³ (PEARSON, 1998, p. 26).

No âmbito da valorização dos contextos a noção de universo discursivo é de suma importância para a identificação do termo, uma vez que o discurso científico se apresenta sob diferentes modalidades e temáticas. Por essa razão, a valorização do texto como lugar em que o termo se atualiza estabelece o conhecimento de toda a complexidade constitutiva da estrutura textual como requisito metodológico essencial para a Terminologia de fundamentação lingüístico-comunicativa.

Monique Slodzian (2000, p. 74) também faz referência ao “fonctionnement réel des unités lexicales en contexte¹⁴” e destaca a sua importância para as pesquisas informatizadas em diferentes domínios. Ela ressalta o interesse demonstrado pelos especialistas de áreas como Inteligência Artificial, Engenharia de Documentação e Lexicografia, dentre outras, pelo desenvolvimento de técnicas em Lingüística de *Corpus*, em razão da extração semi-automática de termos e contextos. Lembra ainda a importância que o conjunto de textos, como unidade de análise, representa para a pesquisa e acrescenta que esses especialistas acreditam que, fora do contexto de ocorrência e distantes da textualidade, fica difícil dar conta da diversidade de configurações das estruturas terminológicas.

O conhecimento da importância do contexto lingüístico de ocorrência para a identificação de um termo evidencia-se como uma produtiva orientação metodológica para detectar o estatuto terminológico de uma unidade lexical, que se instaura pela sua relação semântica e pragmática com o texto especializado. Como comentado mais adiante, no capítulo que aborda a metodologia desta pesquisa, a relevância do contexto é um dos pressupostos básicos a serem levados em consideração no momento da escolha dos termos no processo de indexação e na arquitetura do sistema de recuperação de informação.

4.4 Termo

O termo, conforme é concebido pelos teóricos, é percebido sob diferentes pontos de vista, tais como: unidade de conhecimento, unidade de compreensão ou unidade de significação especializada (USE). Quanto ao aspecto denominativo, encontra-se uma rica terminologia, pois ele é expresso por diferentes formas: termo técnico-científico, termo,

¹³ Tradução da autora: Todos os esforços para definir o que é um termo e examinar qualquer distinção entre diferentes tipos de termos são irrelevantes se o contexto não for considerado.

¹⁴ Tradução da autora: funcionamento real das unidades lexicais em contexto.

unidade lexical terminológica, unidade lexical especializada, unidade lexical temática e vocabulário.

O que motiva o surgimento de um termo é a necessidade de denominar descobertas científicas e tecnológicas; sendo importante observar que o aparecimento de um termo está sempre relacionado com uma área do conhecimento. Em sentido restrito, entende-se termo como unidade de análise e de tratamento terminológico ou, em outras palavras, como um objeto teórico e com finalidades descritivas e/ou aplicadas. Como objeto teórico, é possível estabelecerem-se questionamentos sobre a sua identidade, determinar diferenças em relação às palavras e sua forma de aquisição, assim como sobre o modo como se organizam na mente humana, ou ainda, emitir discussões sobre o fato de eles integrarem ou não uma gramática.

Para a Terminologia o termo é a denominação de um conceito; é o núcleo do conhecimento especializado; não existe conhecimento especializado sem uma terminologia. “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 [...]” (BENVENISTE, 1989, v. 2, p. 252).

O termo tem uma dimensão lingüística, comunicativa e cognitiva e uma estrutura complexa e poliléxica; ele adquire o estatuto terminológico por integrar um conhecimento especializado, veiculando conteúdos temáticos no campo das ciências e das técnicas. Compreende uma dimensão lingüística, uma vez que configura o comportamento lexical especializado ou temático das línguas, e uma dimensão cognitiva, na medida em que expressa conhecimento especializado. Enquanto componente lingüístico e como unidade terminológica e elemento constitutivo da produção do saber, apresenta propriedades que favorecem a univocidade da comunicação especializada. O reconhecimento das propriedades cognitivas e formais do termo consiste em uma das tarefas mais difíceis do trabalho terminológico.

O termo, associado à sua natureza constitutiva se caracteriza pela monorreferencialidade pois, por veicular preferencialmente o significado de uma área, estabelece uma única referência com o mundo exterior, tendo em vista o domínio que a unidade lexical está inserida. Aqui se coloca a variação como um paradoxo.

Uma característica que o termo apresenta é a da invariabilidade semântica; tal propriedade diferencia a unidade terminológica da unidade lexical não especializada, isto é, da palavra. O significado dos termos, por expressar conteúdos conceituais das ciências e das

técnicas, se define por sua relação a um conjunto de significados de um mesmo domínio. Tal fato implica que ele não possa ser considerado isoladamente, mas que esteja sempre contextualizado em um campo determinado.

O termo é essencialmente um signo lingüístico no sentido definido por Saussure. O que o distingue de outros signos lingüísticos é, antes de tudo, que sua extensão semântica se define preferencialmente por sua relação com o significado e não com o significante. Teoricamente, portanto, para uma determinada noção há apenas uma única denominação; esta particularidade se fundamenta no postulado terminológico da relação de univocidade entre denominação e noção, ou seja, entre significante e significado, segundo Saussure, estabelecendo entre ambos uma relação do tipo reflexiva (RONDEAU, 1984). Adverte-se, no entanto, que este princípio não se aplica em todas as situações, sendo possível encontrar no interior de uma mesma língua várias denominações diferentes para um único conceito, o que constitui uma situação de sinonímia.

Há também a considerar, dentre as características do termo, as que dizem respeito ao seu modo de formação e ao fato da homonímia não se constituir em um risco de ambigüidade, uma vez que cada termo pertence a um grupo semântico determinado. Dessa forma, no plano do discurso um termo constitui um par claramente identificado pelo contexto, ou seja, denominação/noção e enquanto no plano lógico o termo encontra seu lugar em uma estrutura hierárquica nocional, no interior de um domínio. Tal característica é determinante para que a Terminologia priorize a noção sobre a denominação.

Em poucas palavras, a partir das considerações apresentadas é possível afirmar que o estatuto terminológico de uma unidade lexical define-se por sua dimensão conceitual, e o que faz de um signo lingüístico um termo é o seu conteúdo específico em um contexto de uso. “O reconhecimento, sem dúvida incontestável, do componente conceitual na constituição do fenômeno terminológico responde fortemente pelas interpretações de que um termo é uma unidade de conhecimento” (KRIEGER, 2001, p. 69).

Na concepção das novas teorias terminológicas, em que o termo é entendido por seu valor como unidade de conhecimento, como unidade lingüística e como unidade de comunicação, os contextos lingüísticos e pragmáticos são componentes que favorecem a articulação do estatuto terminológico de uma unidade lexical, bem como esclarecem a presença de sinonímia e variações terminológicas no léxico especializado. Sob esse prisma, as unidades terminológicas especializadas e lexicais dos sistemas lingüísticos estão sujeitas aos processos de sinonímia e variação das mais diferentes naturezas. Considerando tanto o léxico

especializado como o léxico comum, o termo faz parte do sistema gramatical geral enquanto unidade de um sistema léxico, estando sujeito às mesmas regras de construção de frases e de constituição do discurso que marcam as unidades lexicais da língua.

Por fazer parte do léxico do falante o termo não constitui um sistema lexical independente. Ele designa conceitos próprios de uma disciplina ou atividade específica. O conjunto de palavras próprias de uma determinada especialidade constitui uma terminologia que lhe é própria. Se considerado sob uma perspectiva formal ou semântica ele não difere muito da palavra, mas se diferencia significativamente quando visto segundo critérios pragmáticos e comunicativos (CABRÉ, 1993).

A peculiaridade mais notável da terminologia, em oposição ao léxico comum, é a de servir para designar os conceitos próprios das linguagens de especialidade; em conseqüência, os termos são conhecidos pelos especialistas de cada domínio e aparecem com uma freqüência elevada nos documentos especializados. Como unidade de significação só receberá o tratamento de termo se trazer um sentido específico dentro de uma área de conhecimento.

A circulação das terminologias em inúmeros cenários comunicativos decorre, em grande parte, dos efeitos da larga difusão do conhecimento científico e tecnológico. Tal movimentação evidencia que não há uma fronteira rígida separando os léxicos especializados do léxico geral.

Na medida em que é afetado pelo dinamismo da linguagem, o termo está sujeito ao processo de terminologização. Através desse processo as palavras da língua comum alcançam o estatuto de termo quando passam pelo fenômeno da ressignificação, assumindo significados especializados, pertinentes a um determinado domínio técnico ou científico. Como conseqüência, elas se tornam unidades lexicais de repertórios terminológicos.

Em decorrência desse mesmo dinamismo as unidades lexicais especializadas, quando participam de mais de uma terminologia, expressam diferentes significados em cada campo do saber. Desse modo, a forma lingüística valor, sob o ponto de vista da homonímia tem significados distintos na Lingüística, na Filosofia, na Economia e nas Finanças, por exemplo. Entretanto, no âmbito do léxico especializado, mesmo com a constatação de que um termo oferece diversas possibilidades de aplicação ou emprego, tradicionalmente compreendia-se que os termos eram afetados pelo processo de homonímia, mas não de polissemia. A recusa à polissemia teve por base o postulado da exclusividade denominativa, que determinava um

significante para cada conceito e, conseqüentemente, termos distintos para conceitos de diferentes áreas.

Hoje a unidade terminológica é concebida como uma unidade lingüístico-comunicativa, que compreende tanto uma face lingüística quanto uma vertente conceitual e tem na comunicação especializada o seu *habitat*. Como tal, integra os sistemas lingüísticos e expressa conhecimento; agrega-se a essas duas características um caráter comunicacional que permite fixar e favorecer a transferência da informação comunicada.

Dentre os critérios adotados para a seleção de termos está a possibilidade de definição, a qual se relaciona com o ideal de univocidade e monosssemia a que se propõe a Terminologia. Aqui, há a considerar que a definição de um termo sempre terá como ponto de partida o contexto em que é usado e que os estudos teóricos e aplicados da Terminologia são validados na medida em que descrevem e recolhem os termos em seus reais contextos de ocorrência, a saber, os textos especializados.

A diversidade de posicionamento a respeito da natureza e concepção de termo ultrapassa a discussão no sentido de compreendê-lo como unidade de conhecimento ou unidade pragmático-lingüística, e atesta-o como uma unidade complexa e multifacetada. Segundo Cabré o termo é “una unidad conceptual y denominativa poliédrica en la que el concepto es percebido desde una perspectiva determinada por vários factores”¹⁵ (CABRÉ, 2001, p. 32).

Quanto à sua estrutura morfossintática e léxico-semântica o termo pode apresentar-se como termo simples, termo complexo e termo composto. O termo simples é constituído por apenas um radical, enquanto que o termo complexo e o termo composto, são por dois ou mais radicais. O termo composto se diferencia do termo complexo pelo alto grau de lexicalização e por ser representado graficamente pela utilização do hífen.

O termo complexo é predominante na comunicação especializada; ele pode assumir tamanho diferenciado em relação ao número de unidades léxicas que o constitui. Ele é usualmente reconhecido como termo sintagmático ou sintagma terminológico. A identificação de um termo sintagmático nem sempre é fácil, mas a Terminologia propõe uma série de critérios que se fundamentam no grau de lexicalização dos sintagmas e que auxiliam o terminólogo nessa difícil tarefa.

¹⁵ Tradução da autora: uma unidade conceitual e denominativa poliédrica na qual o conceito é percebido desde uma perspectiva determinada por vários fatores.

4.5 Relação termo e descritor

Na aproximação com a Documentação, observa-se a relação que se estabelece entre termo ou unidade lexical terminológica na visão do terminólogo e termo ou descritor na visão do documentalista. Mesmo se distinguindo pela natureza e funcionalidade, ambos se aproximam no processo comunicacional e como elemento de representação do conhecimento.

O descritor, diferentemente do termo, unidade lexical terminológica, é um elemento de uma linguagem construída pelos gestores da informação, configurando-se por assumir um carácter artificial; é a unidade representativa de nódulos conceituais dos campos de conhecimento, sendo, portanto, a unidade descritiva básica na qual se sustenta o processo de indexação das fontes documentais. Ele desempenha uma dupla função: constitui-se em elemento central do processo de indexação e é também expressão de busca e recuperação da informação por ser o nó conceitual, isto é, a menor unidade de significação especializada de um domínio.

Tanto descritor como termo são meios para recuperar a informação, mas não se confundem por possuírem estatutos próprios.

[...] tan arriesgado es pretender constituir un vocabulario pluridisciplinar de descriptores como una terminología de varias áreas, o quizás más, porque en terminología la autoridad del término viene dada por la de la fuente, mientras que en la documentación la de los descriptores depende de convenciones propias daquellos que procesan la información (LÉRAT, 1997, p. 118).¹⁶

Igualmente a Documentação e a Terminologia, a despeito de sua natural correlação, mantêm suas identidades, seus propósitos e princípios específicos. Assim, a Documentação entende os termos técnico-científicos como unidades lexicais, enquanto que a Terminologia os entende como unidades terminológicas, mas ambas os compreendem com a função de representação e comunicação do conhecimento especializado.

Cabe aqui fazer distinção entre unidade lexical e unidade terminológica; enquanto unidade lexical o termo é extraído do léxico comum e enquanto unidade terminológica, do léxico especializado. É como unidade lexical que muitos autores da área biblioteconômica concebem o descritor e o termo como sinônimos. Conforme observa Van der Laan, a maioria dos autores “concebiam um descritor ou um termo de indexação como uma unidade do léxico comum à qual eram aplicadas regras de restrições, tais como controle de plural, uso de

¹⁶ Tradução da autora: [...] é tão arriscado pretender constituir um vocabulário pluridisciplinar de descritores, como trabalhar com uma terminologia de diversas áreas, ou talvez mais, porque em Terminologia a autoridade do termo é conferida pela autoridade das fontes, enquanto que na Documentação a autoridade do descritor depende das convenções próprias daqueles que processam a informação.

substantivo, restrição aos sintagmas nominais, através das normas de fatoração [...]” (LAAN, 2002, p. 56).

Gomes (1990), ao fazer referência aos descritores, já os entende como unidades lexicais de uma linguagem de especialidade e não mais como unidades lexicais da língua comum. Ela distingue os termos dos não-termos ou não-descritores, mas assegura que os não-termos devem estabelecer relações de equivalência com os termos. Sob essa nova perspectiva, a autora defende a idéia de que a coleta de descritores deve estar fundamentada na literatura especializada, pois apenas dessa forma o termo ou descritor será representativo da linguagem utilizada pelos especialistas. A autora ainda adverte que os termos devem estar contextualizados, pois só assim não ocorrerá perda de sentido. “No momento em que tomamos um termo o contexto é imprescindível, na medida em que é através dele que determinamos seu único significado” (GOMES, 1990, p. 31).

A partir da idéia de que um vocabulário controlado deve se constituir em um instrumento de indexação e recuperação da informação, e que esse instrumento terminológico precisa ser expressivo e representativo de uma determinada área de especialidade, registra-se a importância de que a coleta de termos tenha origem em produções textuais aceitas e consagradas na área de especialidade a que se referem.

Sobre as configurações prototípicas do termo, conforme mencionado por Krieger (2004), é possível identificar as seguintes: nomes, incluindo substantivos e adjetivos; sintagmas terminológicos; signos verbais plenos; signos verbais reduzidos, tais como siglas, acrônimos e abreviaturas; e signos não verbais, como fórmulas químicas. Dentre as unidades terminológicas os substantivos predominam, mas adjetivos, verbos e advérbios também podem assumir o valor de termo.

Para os objetivos a que se propõe a pesquisa apresentada nesta Dissertação e a ser utilizada no âmbito da linguagem documentária, conforme a proposta aqui sugerida, o termo é entendido em toda essa complexidade e, como afirma Barros (2004, p. 42), ele é um vocábulo “que designa um conceito específico de um domínio especializado.”

4.6 Fraseologia

No âmbito da Terminologia o estudo da fraseologia especializada é recente, mas na língua comum seu interesse vem desde a Antigüidade e é visto a partir das expressões idiomáticas, frases feitas, provérbios, locuções nominais e verbais, sintagmas, estruturas hoje

chamadas de colocações, assim como estruturas típicas de determinado tipo de comunicação, como é o caso das fórmulas protocolares em eventos formais. A seu respeito constata-se uma diversidade de pontos de vista, seja no plano conceitual ou mesmo no denominativo, e uma variedade de critérios sobre a constituição formal dessas estruturas.

Bevilacqua (2001), em seu estudo a respeito das unidades fraseológicas especializadas (UFEs), identifica uma grande variedade de denominações que se diferenciam segundo as concepções com que são entendidas pelos teóricos. Ela relaciona, dentre outras, as seguintes: colocação, colocação léxica especializada, co-ocorrência, fragmento de frase, fraseologismo, unidade fraseológica e unidade fraseológica especializada.

Ettinger (1982), ao estudar as fraseologias no âmbito da Lexicografia, se posiciona a favor da presença dessas unidades nos dicionários monolíngües e bilíngües, pois percebe a importância de estudá-las como unidades pluriverbais lexicalizadas e não apenas como unidades isoladas.

A um lexicógrafo cabe definir o sentido dessas estruturas e determinar sua forma de apresentação na organização das entradas de um dicionário, pois palavras que isoladas têm um determinado significado, quando empregadas em estruturas fixas não representam obrigatoriamente a soma de seus significados; é o caso, por exemplo, das expressões: não tem pé nem cabeça, dançar na corda bamba, nadar contra a maré. Estruturas dessa natureza, cujo significado constitui um todo independente das partes que a formam, não são usualmente registradas em dicionários de língua, como um verbete.

Além dessas expressões, que são muitas vezes consideradas como provérbios ou adágios populares, observam-se outras combinações recorrentes de palavras, tais como feijão com arroz e café com leite, por exemplo, que sem nenhuma razão sintática ou semântica para sua estrutura, ocorrem sempre da mesma maneira. O seu estudo se mostra importante inclusive para o estudo de léxicos bilíngües na comparação entre línguas, mostrando-se de grande importância para a Tradução enquanto disciplina teórica e aplicada.

Hoje se percebe que fenômeno semelhante também ocorre nas linguagens especializadas e que certas unidades lexicais se combinam com outras, sempre da mesma forma, para apontar unidades lexicais terminológicas do mesmo domínio. Desse modo é possível fazer distinção entre as fraseologias da língua geral e as fraseologias especializadas. Essas passam conseqüentemente a ser objeto de estudos tanto da Terminologia teórica como da Terminologia aplicada, nesse caso, visando à produção de instrumentos como glossários,

dicionários e bases de dados. Os estudos sobre as estruturas fraseológicas estão assumindo tamanha dimensão, que já se encontram pesquisadores reivindicando que sejam entendidas como objeto de pesquisas de uma área autônoma, a Fraseologia.

Apesar da dificuldade em definir as unidades fraseológicas especializadas é possível observar a presença de algumas características que se repetem e que permitem identificá-las. Bevilacqua (2001) percebe três tendências principais: a que define as UFE como colocações, ou seja, como unidades resultantes da combinação de duas unidades léxicas, em que uma delas é o núcleo e a outra o colocado; a que as define como combinação sintagmática, cujo núcleo terminológico é um sintagma; e as que as consideram como sendo representadas por expressões usualmente formuladas em um domínio discursivo e que podem chegar a ser uma frase completa.

O seu estudo no âmbito desta Dissertação se justifica uma vez que o *corpus* constituído e que permitiu a coleta da terminologia que veicula entre os especialistas revelou, além da presença das unidades terminológicas, a ocorrência de unidades fraseológicas especializadas, o que inicialmente não havia sido previsto. Levando em consideração a diversidade com que os teóricos entendem essa questão, para os propósitos da pesquisa aqui desenvolvida e a ser utilizada no âmbito da linguagem documentária, entende-se a fraseologia especializada em uma perspectiva ampla. Nessa visão, identifica-se a UFE pela co-ocorrência dos elementos que as constituem, apresentando graus de fixação não obrigatoriamente rígidos, mas não totalmente livres, expressando um significado especializado e uma frequência significativa dentro da área de especialidade.

As fraseologias especializadas são formas de expressão recorrentes, semanticamente vinculadas aos conteúdos da comunicação especializada; diferenciam-se do termo, que se caracteriza por designar objetos e conceitos. Apresentam pelo menos uma unidade terminológica, considerada como núcleo dessa unidade; distinguem-se das unidades terminológicas poliléxicas por não serem definidas como unidades léxicas de um domínio de especialidade, mas como unidades sintáticas, uma vez que não podem ocupar na frase a posição de um constituinte sintático autônomo mínimo; são basicamente nominais, têm caráter denominativo, valor referencial e representam um nó conceitual na estruturação de conceitos de um campo especializado.

Elas podem admitir variação interna e também se apresentarem como semifixas, mas sempre com elevada frequência em determinado discurso especializado. Como se trata de um elemento lingüístico freqüente na constituição da comunicação especializada interessa

sobremaneira aos propósitos desta Dissertação. Currás (1995) ao fazer referência à construção de um banco de dados terminológico já considerava, além de conceitos e termos, o que ela denomina as frases terminológicas.

4.7 Comunicação especializada

O conjunto de termos, por refletir a estruturação conceitual de um campo de conhecimento, constitui-se como a base da comunicação especializada. Assim como a organização do pensamento e a atribuição de conceitos representam a dimensão cognitiva da Terminologia, a comunicação do conhecimento vem a ser o pilar da sua dimensão comunicativa.

Por expressar o conhecimento especializado, o termo configura o léxico especializado ou temático de uma língua. Ele representa o saber produzido por especialistas de um determinado domínio, científico, técnico e tecnológico, e é inerente à comunicação especializada. Esta se diferencia formalmente da comunicação geral sob dois aspectos: quanto ao tipo de texto que produz e quanto ao uso de uma terminologia específica, tendo como produto artigos científicos, teses, resenhas e textos especializados.

O uso de uma terminologia padronizada na comunicação especializada contribui para tornar eficaz a comunicação entre especialistas. Nos textos especializados predominam, além da objetividade e do uso sistemático de termos técnico-científicos, o esforço por alcançar a concisão, a precisão e a adequação. A concisão permite diminuir a possibilidade de se produzir distorções na comunicação; a precisão satisfaz aos propósitos da temática técnico-científica e da comunicação entre especialistas; a adequação à situação comunicativa em que se produz, segundo as circunstâncias de cada situação, permite adaptar-se às características dos interlocutores e ao respectivo nível de conhecimento sobre o assunto.

O processo de comunicação entre especialistas e sobre temas especializados não difere muito do processo geral de comunicação. O esquema comunicativo, conforme foi estabelecido por Jakobson (2003), pressupõe os seguintes elementos: remetente, destinatário, linguagem, realidade e canal; o ato de comunicação pressupõe um outro elemento, a mensagem.

Originalmente a dimensão comunicacional da Terminologia foi entendida a partir do universo das ciências; hoje, é possível afirmar-se que a presença das terminologias vai muito além desse universo. Na medida em que se reconhece a presença dos léxicos temáticos

operando como unidades lexicais a serviço de uma comunicação especializada é preciso admitir, também, que determinadas unidades lexicais adquirem a funcionalidade e o estatuto terminológico nas mais diferentes áreas do saber e que a comunicação especializada se realiza em diferentes níveis de especialização e cultura.

A partir da constatação de que os saberes técnico e científico se expandem cada vez mais e com extrema velocidade, provocando o fenômeno da explosão da informação, ampliam-se, também, os domínios de fixação de vocabulários especializados. Para a eficácia das comunicações especializadas a univocidade comunicacional é efetivamente uma condição a ser alcançada tendo em vista as diversas proposições de intercâmbio que se intensificam com a globalização. Em busca da precisão conceitual, as linguagens de especialidade utilizam termos próprios para circunscreverem conceitos e transmitirem informações específicas. Viabilizar as condições de comunicação constitui-se em uma estratégia cujos suportes são os repertórios terminológicos.

4.8 Terminologia e Informática

O relacionamento da Terminologia e da Informática não é de hoje. Há muito tempo os dois campos se aproximam e buscam um intercâmbio cada vez maior. A Terminologia do século passado ansiava por recursos operacionais, enquanto a Terminologia contemporânea vê na Informática um campo comum de troca de pesquisas, experiências e aplicações (MACIEL, 2005). Por outro lado, a Informática se vale dos aportes da Terminologia quando se trata de selecionar os termos numa hierarquia lógico-cognitiva, tal como nas aplicações de Inteligência Artificial e no Processamento da Linguagem Natural.

Este tópico se propõe a abordar algumas das aplicações dos recursos informáticos no processamento de dados terminológicos e que fundamentaram a pesquisa desenvolvida. Ele se refere particularmente à Lingüística de *Corpus*, à Terminografia e às bases de dados terminológicas.

4.7.1 Lingüística de Corpus

A Lingüística de *Corpus* enquadra-se perfeitamente nos propósitos desta pesquisa, uma vez que tem como objeto de estudo a língua em contexto de uso e, como tal, possibilita a análise da terminologia de uma área específica do conhecimento, a partir da constituição de um *corpus* representativo. Em sua fundamentação teórica está o princípio de que o significado

de uma palavra se configura pelas combinações em que ela se apresenta no evento comunicativo.

Em seus pressupostos e sua metodologia encontrou-se um caminho facilitado para a geração de bases de dados terminológicas construídas a partir da terminologia que veicula em textos especializados, visando buscar e recuperar informações em catálogos eletrônicos. Acredita-se, portanto, que ela se apresenta como um novo recurso a ser utilizado pelos profissionais da Documentação.

Atualmente, vem se mostrando como uma das áreas mais produtivas nos estudos que envolvem a linguagem e que vêm sendo desenvolvidos pelos centros de excelência que se dedicam à pesquisa teórica e aplicada. Permite a constituição e exploração de um *corpus* criteriosamente selecionado e representativo de uma determinada área do conhecimento, viabilizando procedimentos de análise que possibilitam a identificação de uma terminologia em pleno uso.

Um *corpus* é um objeto criado com fins específicos para a pesquisa; ele é formado por um imenso conjunto de textos, inteiros ou segmentos maiores do que a frase, legíveis por máquina, selecionados, organizados e identificados segundo critérios previamente estabelecidos, com o objetivo primordial de serem utilizados para fins de análise lingüística.

Sardinha (2004), o grande divulgador da Lingüística de *Corpus* no Brasil, faz as seguintes considerações sobre a constituição de um *corpus*: não pode conter textos produzidos especificamente para os propósitos de uma determinada pesquisa lingüística; é necessário que eles sejam produzidos por falantes nativos; seu conteúdo tem que ser criteriosamente escolhido; precisa ser representativo de uma variedade lingüística ou de um idioma. Por registrar a linguagem natural realmente falada pelos falantes nativos em situações reais, um *corpus* constitui-se em uma importante fonte de informação.

O quadro conceitual da Lingüística de *Corpus* insere-se em uma abordagem empirista e, dentro de uma visão da linguagem, como sistema probabilístico. Por caráter probabilístico entende-se que a linguagem forma padrões que apresentam certa regularidade e variação sistemática, e que certos traços são mais freqüentes do que outros. Assim sendo, quanto maior for o *corpus* maior a probabilidade de ser representativo, mas fazem-se duas ressalvas: é impossível estabelecer o tamanho ideal de uma amostra e não é suficiente que ele seja representativo, pois ele deve adequar-se aos interesses do pesquisador e ser específico aos propósitos da investigação.

A Lingüística de *Corpus* se desenvolveu a partir da década de 60 do século XX, graças ao uso de computadores em centros universitários de pesquisa. No entanto, considerando o sentido original da palavra *corpus*, conjunto de textos, pode-se fazer referência à sua presença na Antigüidade e na Idade Média, citando, por exemplo, o *corpus* helenístico definido por Alexandre, o Grande, e o de citações da Bíblia; *corpora*, obviamente, coletados, organizados e analisados manualmente pelos monges das famosas bibliotecas medievais. O que leva a pensar que o trabalho do bibliotecário com base em *corpora* é uma tradição que remonta aos tempos medievais e que está na hora de revitalizar.

Como visto, a pesquisa lingüística baseada em um *corpus* não é nova; a novidade está no uso do computador e na sua constituição a partir dos textos digitalizados, tal como no desenvolvimento de aplicativos especialmente preparados para a pesquisa lingüística, sob os mais variados pontos de vista. A importância do computador para a Lingüística de *Corpus* está em viabilizar maior precisão e praticidade aos estudos desenvolvidos a partir de um *corpus*; ele é uma ferramenta indispensável para análise, pela possibilidade de reunir grande volume de textos.

Acrescenta-se à eficiência da operacionalização automatizada ou semi-automatizada a objetividade das buscas, isentas de uma ação naturalmente tendenciosa do pesquisador. Livre do árduo trabalho de lidar com uma grande massa de dados, ele dedica-se exclusivamente aos aspectos intelectuais que orientam os propósitos específicos da pesquisa, tais como: coleta, organização, validação e interpretação dos dados (MACIEL, 2002).

Dentre as diversas funções e utilidades dos estudos de *corpora* eletrônicos estão a construção e gestão de bancos e bases de dados textuais e terminológicos. Três *corpora* eletrônicos podem ser considerados como marcos históricos: Brown *Corpus*, BNC (British National *Corpus*) e Bank of English. O *Brown University Standard Corpus of Present-Day American Language* deu início à organização dos grandes bancos de dados textuais eletrônicos; o BNC foi o primeiro a conter 100 milhões de palavras; e o Bank of English é um *corpus* em crescente expansão.

A proliferação dos microcomputadores na década de 80 do século XX, a popularização dos *corpora* eletrônicos e a oferta de ferramentas de processamento contribuíram decisivamente para o fortalecimento da pesquisa lingüística sob o enfoque da Lingüística de *Corpus*. É importante ressaltar, no entanto, que embora um *corpus* seja uma fonte rica e inesgotável para pesquisa, não é exaustivamente completo nem está livre de erros

de uso, uma vez que os critérios de seleção dos textos dão primazia à autenticidade e à maior abrangência possível de contextos, em detrimento da gramaticalidade.

Na época atual os primeiros que se dedicaram à constituição de um *corpus* foram os dicionaristas e os educadores. A preocupação preponderante era a descrição e o ensino da língua; hoje, a ênfase tem sido dada à pesquisa lingüística não necessariamente com fins lexicográficos ou de cunho pedagógico. Há inclusive a proliferação de centros de pesquisa mantidos por empresas dos mais diferentes ramos, com finalidades comerciais, o que atesta também o potencial econômico que representa.

As bases teórico-metodológicas da Lingüística de *Corpus* podem ser encontradas em J. R. Firth, que é considerado um pioneiro nesta área. Mesmo não tendo acesso aos modernos recursos da Informática ele percebeu seu valor para a Lingüística. A partir da idéia de que há uma regularidade nos tipos de associações a que se submetem as palavras de uma língua, o autor observou que o significado de uma palavra se configura pelo contexto de uso, ou seja, pelas combinações que ela estabelece no evento comunicativo; a recorrência, freqüência e estabilidade dessas combinações mostram que elas não ocorrem ao acaso. Ele enfatizou a importância das relações sintagmáticas e paradigmáticas do léxico, salientando o aspecto sócio-lingüístico da comunicação.

Dentre os britânicos que seguiram os passos de Firth vale citar Fillmore, que têm seu nome associado à organização dos primeiros grandes *corpora* constituídos em língua inglesa; ele observou o inter-relacionamento que deve ocorrer entre os dados empíricos e a introspecção. Em sua ótica, competência e desempenho são características que devem co-existir na pessoa de um lingüista, perfazendo duas faces de uma mesma moeda. A razão para tal afirmação é que são indispensáveis para a análise lingüística, pois uma não pode subsistir sem a outra.

Fillmore, um dos pioneiros dos estudos lingüísticos assistidos pelo computador, fez a seguinte afirmação no famoso I Simpósio Nobel sobre Lingüística de *Corpus*, realizado em Estocolmo em 1992.

I have two main observations to make. The first is that I don't think there can be any corpora, however large that contain information about all of the areas of English lexicon and grammar that I want to explore: all that I have seen are inadequate. The second observation is that every *corpus* that I've had a chance to examine, however

small, has taught me facts that I couldn't imagine finding out about in any other way (FILLMORE, 1992, p. 35).¹⁷

Uma grande discussão tem sido feita sobre a definição do *status* da área, pois alguns a situam como disciplina e outros como uma metodologia. Entretanto, há teóricos que afirmam que ela não é nem uma disciplina e nem uma metodologia, mas uma nova abordagem para os estudos lingüísticos, tendo um *corpus* como base. Os que negam que ela seja uma disciplina argumentam que ela não se dedica a um assunto definido, mas a vários fenômenos focalizados em outras áreas; e aqueles que alegam que ela não é apenas uma metodologia referem-se ao fato dela permitir a produção de novos conhecimentos, uma vez que revela fatos a respeito da língua que não tinham sido observados.

A Lingüística de *Corpus* apresenta como vantagem procedimentos de análise para um grande volume de informações e o fato de que as ferramentas, em sua maioria, se encontram à disposição do usuário na *web*; além disso, os grandes *corpora* referenciais podem ser acessados *on-line*, na maioria dos idiomas.

A proposta metodológica da Lingüística de *Corpus*, conforme é entendida atualmente, fundamenta-se a partir de um *corpus*, de um computador e da análise dos dados pelo pesquisador. O *corpus* é essencial, o computador torna-se imprescindível sob o ponto de vista da coleta, armazenamento e análise de dados, e o pesquisador constitui-se como elemento fundamental por determinar quais são os dados que devem ser coletados e como devem ser organizados para que seja viável o processo de pesquisa.

A Lingüística de *Corpus* se constitui como um processo interativo entre o homem e a máquina, onde conhecimento científico e tecnologia estão presentes e interagem. Há ainda a considerar o uso de recursos disponibilizados na *web* para efetuar a análise dos dados, sendo, um deles, o *WordSmith Tools*, um utilitário que tem sido muito utilizado pelos pesquisadores. Das ferramentas disponibilizadas, algumas são livres de taxa, sendo outras distribuídas comercialmente.

A Lingüística de *Corpus*, por depender de conhecimentos advindos da Informática, estabelece com esta área uma interface perene, pois os profissionais destes dois domínios

¹⁷ Tradução da autora: Tenho que fazer duas observações importantes. A primeira é que eu acho que não existem *corpora*, por maiores que sejam, que contenham toda a informação sobre o léxico e a gramática do inglês que eu gostaria de explorar: todos que eu vi são inadequados. A segunda observação é que todo o *corpus* que eu tive a oportunidade de examinar, por menor que fosse, me ensinou fatos que eu não poderia imaginar encontrar de outra maneira.

precisam manter um diálogo constante. Isoladamente não é possível desenvolver uma ferramenta que se mostre válida e confiável para a análise de dados lingüísticos.

Além da Informática, a Lingüística de *Corpus* também utiliza conhecimentos da área da Estatística. O uso da estatística se fundamenta no caráter probabilístico da língua e decorre da necessidade de avaliar o significado dos dados e seu valor como amostragem do sistema da língua, tanto no que diz respeito à extensão, como no caso dos grandes *corpora* referenciais, como em recortes de uso, ou seja, em *corpora* especializados.

A metodologia para a geração de bases de dados terminológicas que se quer aqui propor encontra na Lingüística de *Corpus* grande apoio. Isso não só devido à necessidade de manipular grandes conjuntos de textos mas, acima de tudo, porque esta proposta também compartilha dos mesmos princípios a respeito da configuração do significado no seu contexto de uso e da importância de fatores sócio-culturais dos interlocutores.

4.7.2 Terminografia

A Terminografia é a face aplicada da Terminologia. Definida como a prática de elaboração de glossários, dicionários técnicos ou terminológicos e bancos e bases de dados, é reconhecida por alguns como Lexicografia Especializada. Ela estuda e registra os termos de uma área de especialidade; difere da Lexicografia, que elabora dicionários de língua ou dicionários especiais e registra unidades lexicais em todas as suas variações morfossintáticas e acepções. A Terminografia, ainda que intimamente ligada à Terminologia, tem sua própria fundamentação teórica; essa, por sua vez, influencia a metodologia adotada na coleta e organização de termos.

Terminografia e Lexicografia diferem entre si por terem objetos e metodologias diferenciadas. Um dicionário terminológico especializado é elaborado dentro de uma visão pragmática, registra exclusivamente a acepção que corresponde à unidade terminológica de uma área do saber. Um dicionário de língua tem uma perspectiva semântica e registra várias acepções diferentes para uma unidade lexical e até mesmo termos a ela relacionados, além de expressões idiomáticas que se referem à mesma unidade lexical.

A Terminografia, segundo Boulanger (2001, p. 13) pode ser definida como:

Trabalho e técnica que consiste em recensear e em estudar termos de um domínio especializado do saber, em uma ou em línguas determinadas, considerados em suas formas, significações e relações conceituais (onomasiológicas) assim como em suas relações com o meio sócio-profissional.

A Terminografia e a Lexicografia se diferenciam quanto ao campo de atuação (língua de especialidade e língua geral), unidade padrão (unidade terminológica e unidade lexical), nível de atualização da unidade lexical (norma de especialidade e sistema) e tipo de obras produzidas (vocabulários técnicos, científicos e especializados, e dicionários de língua e dicionários especializados). Agrega-se, aos aspectos mencionados, o caráter onomasiológico da Terminografia, em razão de sua preocupação primeira de focalizar o plano do conteúdo, sobrepondo-se ao plano do significante no trabalho de identificação das unidades lexicais que assumem o estatuto de termo, embora o aspecto semasiológico também seja contemplado.

Os produtos terminográficos repertoriam o componente léxico temático, o termo, em sua forma plena, como entrada de verbete, refletindo suas condições de uso nas comunicações especializadas. Entretanto, não acolhem as palavras ditas gramaticais que, mesmo presentes nas comunicações especializadas, não representam o conteúdo conceitual de uma área de conhecimento. Sob esse enfoque os produtos terminográficos diferem dos lexicográficos, que usualmente não registram os sintagmas e as locuções como entrada de um verbete, mas apenas como parte dele, uma vez que a entrada é preferencialmente constituída pelo item lexical integrante da expressão sintagmática.

Terminografia e Lexicografia se afastam fundamentalmente quanto ao seu objetivo. A Terminografia tem como propósito oferecer informações específicas de uma área de conhecimento, correlacionando termos e conceitos desse universo terminológico e conceitual. A Lexicografia visa cobrir todas as realizações lingüísticas e semânticas de uma unidade lexical; dessa forma, esforça-se em relacionar todos os significados possíveis que uma palavra ou expressão comporta. Além de registrar usos e sentidos das palavras, oferece informações etimológicas e gramaticais. Modernamente, cresce a tendência em inserir informações sobre a língua nos repertórios terminográficos, tais como categoria gramatical, variações, sinonímias e marcas de uso socioprofissional dos termos repertoriados.

A Terminografia mantém com a Terminologia uma estreita relação; os estudos desenvolvidos sobre os termos dão suporte teórico para a produção das obras terminográficas. Embora seja considerado por alguns estudiosos como sendo um ramo da Terminologia, a Terminografia, segundo outros, possui uma identidade própria. Conforme Barros (2004, p. 68), a Terminografia “é uma disciplina científica que analisa seu objeto de estudo (os dicionários terminológicos), propõe novos modelos de tratamento dos dados, reflete cientificamente sobre seu trabalho, além de construir uma metalinguagem própria e de consolidar uma metodologia de elaboração de dicionários terminológicos.” Enfim,

Terminologia e Terminografia se diferenciam pelo caráter de ciência fundamental da primeira e de ciência aplicada da segunda. Elas se caracterizam por suas práticas diferenciadas, mas complementares no tratamento da unidade terminológica.

Os instrumentos terminográficos são representados pelos dicionários terminológicos – monolíngüe, bilíngüe ou multilíngüe – e bancos e bases de dados terminológicos; eles variam em função de sua estrutura e conteúdo. Mesmo cumprindo finalidades semelhantes quanto à organização e divulgação das terminologias, cada obra terminográfica possui características específicas.

A Terminografia não se restringe a uma visão pragmática de produção de instrumentos terminográficos. Ela se dedica também ao estudo do termo, tendo em vista a necessidade de observar e dimensionar os fundamentos teóricos necessários à identificação das terminologias, ao reconhecimento da variedade de suas formas. Com base nos princípios de análise do funcionamento dos termos, visa o seu registro em instrumentos de referência especializada. Esse entendimento leva à identificação de subsídios para o estabelecimento de princípios e diretrizes, metodológicos e pragmáticos, respectivamente.

Dentre as características da Terminografia destaca-se a sua função normalizadora, que tem o propósito de estabelecer a padronização terminológica, registrando a relação de termos utilizados nas comunicações profissionais. Dessa forma, os produtos terminográficos são instrumentos construídos para registrar e, de certo modo, fixar expressões especializadas criadas pelos especialistas e empregadas na comunicação de uma área de conhecimento e/ou em instituições e empresas. Ao fixar a forma de expressão considerada a mais adequada, entende-se que a ação padronizadora nada mais é do que uma medida de controle terminológico, estando em consonância com a concepção de que a comunicação especializada requer um elevado grau de precisão.

Há instrumentos terminográficos construídos com o propósito de padronizar a terminologia de uma área, enquanto que há outros que favorecem um ponto de vista descritivo. Entre os primeiros estão os glossários produzidos por uma empresa industrial ou comercial e aqueles elaborados para os programas de Inteligência Artificial. Outros instrumentos, entre os quais deve ser citada a base terminológica aqui proposta, descrevem a terminologia de uma área de especialidade e não se limitam a repertoriar apenas os termos que seriam considerados recomendados, mas registram também sinônimos e variações, contribuindo, dessa maneira, para a sua normalização.

Considerando que a finalidade primeira de um produto terminográfico é atender às necessidades de informação dos usuários, aquele que é produzido a partir de uma visão descritiva cumpre melhor a sua finalidade. A reflexão que hoje predomina sobre o fazer terminográfico é exatamente a que busca, nos avanços dos estudos lingüísticos de caráter lingüístico-comunicacional, a sua fundamentação. Este trabalho procura estar em consonância com tais avanços.

A Terminografia de fundamentação lingüístico-textual constrói sua base teórico-metodológica a partir das relações entre o funcionamento da linguagem, das especificidades das comunicações especializadas e do estatuto terminológico das unidades lexicais. Ao postular esses princípios a Terminografia está em harmonia com a recuperação de uma terminologia *in vivo* e não *in vitro*, conforme priorizado pelas teorias clássicas. Com esse juízo, a possibilidade de que um instrumento terminográfico cumpra o seu papel de facilitar a comunicação especializada torna-se bem mais viável.

4.7.3 Bancos e bases de dados terminológicos

Atualmente, são os cientistas da computação, e não os bibliotecários, que têm usualmente criado as ferramentas para se navegar nas redes de informação. Caberá, pois, aos bibliotecários (com um novo perfil), aos novos profissionais da informação, explorar ferramentas de *software* derivadas das pesquisas em inteligência artificial (tão anunciadas mas raramente implementadas) que sejam capazes de recuperar informação ou realizar outras tarefas de forma automatizada (LUCAS, 1996, p. 71).

Dentre os produtos terminográficos que mais se têm beneficiado do relacionamento da Terminologia e da Informática estão os bancos e bases de dados terminológicos. A criação e a gestão dessas ferramentas constituem-se em uma nova faceta da Terminologia aplicada.

O primeiro banco de dados terminológico, conhecido como DicAutom, tem origem em 1963; ele marcou o início de uma nova atividade na área, a Terminótica, que na expressão informal de Gouadec (1990), é apenas o casamento da Terminologia e da Informática. DicAutom, transformando-se mais tarde no EuroDicAutom, é hoje o banco de dados terminológico multilíngüe da União Européia. No Canadá, em 1970, foi criado o Termium e, em 1973, o BTQ, o Banco de Terminologia do Quebec; tais bancos, de finalidade normativa, serviram de modelo a novos projetos e são reconhecidos como os bancos de dados da primeira geração.

Hoje, graças à popularização da tecnologia, os bancos de dados terminológicos se multiplicam em instituições públicas, acadêmicas e comerciais; as bases de dados terminológicas cada vez mais são utilizadas nos ambientes de trabalho de tradutores e

redatores, não só com objetivos de normalizar uma terminologia mas, antes de tudo, pela tendência atual de descrever os termos usados nas áreas especializadas.

Para Tebé (1996), três características particularizam a arquitetura de um banco de dados terminológico: a integração, a estruturação e o grande volume de informações. Com a integração das informações terminológicas, variadas e provenientes de diversas fontes, armazenadas em um único suporte, é possível efetuar o acesso a elas utilizando uma mesma linguagem de consulta. A estruturação realiza-se por meio da ficha terminológica; os dados ali registrados são distribuídos sistematicamente, em campos definidos, formando uma rede de informações que se estrutura por meio de *links*. Em relação ao volume de informações um banco de dados terminológico é formado por um repertório considerável e representativo de termos e textos de várias áreas do conhecimento.

Dentre as observações feitas por Tebé (2001) vale ainda ressaltar que há dois fatores que intervêm no desenho de um banco de dados: a organização interna da informação e a finalidade com que a ferramenta é criada. Quanto à organização interna a maioria dos bancos compartilha alguns princípios comuns, independente da finalidade com que são criados e de seus destinatários.

Nesse sentido, o autor identifica as seguintes aproximações: eles reúnem um grande volume de informações; estruturam-se em torno do binômio conceito/termo; estabelecem uma subestrutura interna por áreas temáticas. Esse último princípio citado por Tebé constitui-se em um recurso para superar os problemas de consulta gerados pela homonímia, quando os binômios conceitos/termos são utilizados para designar uma ou mais áreas e subáreas temáticas.

Quanto à finalidade, a maioria dos bancos responde a diferentes necessidades. No que diz respeito a esse parâmetro, é possível identificar: a) aqueles que se destinam aos tradutores e se preocupam em estabelecer equivalências para os serviços de tradução de grandes organizações multilíngües, públicas ou privadas; b) aqueles que são criados tendo como função principal atualizar o vocabulário especializado de uma língua, como parte do processo de normalização dessa língua; c) aqueles que são criados em contextos de organizações especializadas, motivadas por alguma temática, normalmente de caráter técnico-científico, e que perseguem o desejo de regularizar e harmonizar o vocabulário para facilitar a comunicação entre seus especialistas.

Ainda, segundo Tebé (2001), o atual modelo de banco de dados terminológico tem mostrado insuficiência e contradições graves quanto ao seu funcionamento. Quanto à estrutura interna das informações de um banco, o autor observa que os grandes bancos acumulam problemas que têm origem nas variações de significado de conceitos de áreas afins ou, até mesmo, no interior de uma mesma área de especialidade; apesar de sua estruturação rígida em uma área de especialidade, o que parecia ser uma garantia de máxima precisão na localização e identificação de cada conceito. Ele aponta também para os desajustes detectados no que respeita à finalidade dos bancos de dados quanto à adequação às necessidades de seus usuários e constata que quando esses bancos não são estruturados a partir do perfil dos seus usuários há uma grande possibilidade de gerar ruído e silêncio nas consultas.

Para os propósitos deste trabalho, esclarece-se que se entende base de dados terminológica como um banco de dados de proporções reduzidas, em outras palavras, um sistema de informações armazenadas sistematicamente em computador, representando o conhecimento de uma área de especialidade. Tal base tem por finalidade reunir e organizar os conceitos de um determinado domínio, segundo as relações de sentido que eles mantêm entre si. Ela pretende ser a mais representativa possível dos estudos desenvolvidos na área, na atualidade. Trata-se, assim, de um banco de dados de pequeno porte, constituído por um conjunto digitalizado de unidades lexicais de significação especializada de uma determinada área ou subárea do conhecimento, interligadas conforme critérios pré-estabelecidos e estruturadas em um programa de gerenciamento de dados.

Uma base de dados terminológica constitui-se, na atualidade, em ferramenta referencial relevante para diferentes profissionais, em especial, para o bibliotecário. Pelas vantagens apresentadas – o intercâmbio de dados de forma rápida e eficiente entre os diferentes bancos ou bases e as formas de acesso facilitadas – tornam-se ferramentas altamente indicadas para uso, seja para atender aos tradutores por meio de bases informativas, seja para normalizar ou descrever terminologias. Hoje, várias bases terminológicas – institucionais ou privadas – estão sendo desenvolvidas e colocadas à disposição dos usuários; elas têm sido criadas em função de suas necessidades de informação e, se *on line*, podem ser acessadas por qualquer pessoa, em qualquer parte do mundo.

O projeto de uma base de dados é complexo e precisa considerar fatores humanos e logísticos mas, em primeiro lugar, o público a que se destina e a finalidade com que é criada. Em consonância com objetivos e recursos, torna-se necessário decidir a partir de duas possibilidades: adotar uma estrutura de suporte já pré-fabricado ou construir uma estrutura de

acordo com as peculiaridades do contexto de uso. Tomada essa decisão, surge a necessidade de um segundo posicionamento e que se refere à seleção das fontes de coleta dos termos e do tipo de sistematização a ser adotado. Nesse momento é preciso criar um sistema de categorias segundo o qual os termos serão organizados e registrados, isto é, é necessário decidir que tipo de sistema nocional será privilegiado. As várias etapas de planejamento e a seqüência de procedimentos serão esclarecidas no capítulo seguinte, quando será explicitada a metodologia seguida na proposta aqui apresentada de criação de base de dados terminológica.

5 PROPOSTA PARA UMA BASE DE DADOS TERMINOLÓGICA

5.1 Considerações preliminares

Este capítulo apresenta a metodologia sugerida para a geração de uma base de dados terminológica a ser utilizada no âmbito da análise documentária. Conforme foi possível observar no referencial teórico desenvolvido e como já havia sido atestado pela experiência pessoal, em razão das atividades realizadas na BSCSH como a bibliotecária responsável pela classificação e indexação de documentos, as dificuldades enfrentadas nos processos de indexação, busca e recuperação da informação, sempre se mostraram como um problema a ser discutido no âmbito da análise documentária.

Com o objetivo de oferecer novas reflexões sobre o tema, teve-se como propósito, neste trabalho, desenvolver uma metodologia para a geração de bases de dados terminológicas com fundamentação na Documentação, Terminologia e Lingüística. Várias aproximações já foram feitas nesse sentido, especialmente entre Documentação e Lingüística, e Documentação e Terminologia. Este trabalho, no entanto, procurou ir mais além, não apenas por sugerir uma interface produtiva entre as três áreas, mas por apresentar uma metodologia para a geração de bases de dados terminológicas a partir da constituição de um *corpus* especialmente construído para esse fim pelo profissional encarregado da indexação. Além do mais, rompe-se aqui com um dos princípios consagrados na área, quando se trata da elaboração de tesouros, o da organização hierárquica do conhecimento.

A esse respeito, esclarece-se que em momento algum é negada a sua importância, mas não há como ignorar a sua complexidade. Como se sabe “[...] muitos autores vêm procurando definir um sistema de classificação das inúmeras ciências. Nenhum desses sistemas se mostra absolutamente satisfatório” (GIL, 1999, p. 21).

Tal posicionamento justifica-se, também, a partir das seguintes considerações: nem mesmo os especialistas são unânimes em relação à estrutura hierárquica do conhecimento; na medida em que um mesmo texto pode ser visto sob diferentes pontos de vista, a estrutura hierárquica do conhecimento será sempre relativa; um mesmo termo, ainda que em uma mesma área do saber, pode ser considerado sob diferentes formas hierárquicas. Citando como exemplo o termo signo, logo se percebe que há diversas possibilidades de classificá-lo adequadamente em diferentes estruturas hierárquicas, tais como: Lingüística Estrutural,

Linguística Transformacional, Semântica, Semiologia, Semiótica, Teoria da Comunicação e Teoria da Enunciação.

Na prática, o bibliotecário das áreas das Ciências Sociais e Humanidades, por não possuir uma ferramenta que contemple satisfatoriamente a terminologia da área, vem se valendo, há vários anos, das tabelas de classificação para melhor entender a estrutura hierárquica do conhecimento. E ainda assim, conforme visto na literatura e na prática biblioteconômica, nem mesmo tais tabelas são consistentes, pois a cada nova edição apresentam mudanças estruturais significativas em relação à forma como o conhecimento se organiza.

Finalizando essas considerações e dando continuidade a este estudo, a partir do quadro referencial teórico desenvolvido nos capítulos anteriores, são descritas as etapas metodológicas propostas para a construção de uma base de dados terminológica: seleção da área, constituição do *corpus*, seleção dos textos, seleção das ferramentas, coleta e análise de dados, arquitetura da base e inserção dos dados.

5.2 Seleção da área

Para fins desta pesquisa a seleção da área de especialidade priorizou aquela com a qual esta Mestranda tem familiaridade, a Linguística Teórica e Aplicada, uma das áreas das Ciências Sociais e Humanidades. Considera-se que o conhecimento da área selecionada oferece grandes vantagens quanto ao tratamento a ser dado à terminologia especializada, tanto em relação à seleção das unidades de significação especializada (USEs) como quanto a sua categorização em unidades terminológicas, unidades fraseológicas especializadas e modificadores.

5.3 Constituição do *Corpus*

Trata-se de um *corpus* piloto que não tem a pretensão de ser exaustivo, mas ser apenas uma amostragem da área em foco com o objetivo de exemplificar de modo concreto todas as etapas da metodologia que se quer propor. O *corpus* é formado predominantemente por resumos de trabalhos publicados em periódicos, teses e dissertações, congressos e seminários, no espaço de sete anos, de 1999 a 2005, nos quais a LINGÜÍSTICA TEÓRICA E APLICADA é um dos temas principais. Salienta-se que a familiaridade com a terminologia utilizada nessa área, adquirida pela autora da Dissertação na sua vivência como bibliotecária, foi inegavelmente de grande valia.

A construção de um *corpus* textual levou em consideração dois fatores principais: a sua representatividade, tanto na área como junto à comunidade acadêmica, e a facilidade de acesso. Observe-se que a representatividade aqui referida não está relacionada com a extensão, mas com a população envolvida e com a área temática.

Foram selecionados textos do Portal da CAPES, da Base SciELO, e de três CD-ROMs de anais de eventos acadêmicos: Colóquio Nacional de Letras em Diálogo e em Contexto (2003), II Congresso Nacional da ABRALIN e XIV Instituto Lingüístico (1999), XVII ENANPOLL – Encontro Nacional da Associação Nacional de Pós-Graduação e Pesquisa em Letras e Lingüística (2002) e ainda, em CD-ROM, uma compilação de textos em Lingüística Aplicada (2000).

O Portal de Periódicos da CAPES¹⁸, Portal Brasileiro da Informação Científica e Tecnológica, cobre todas as áreas de conhecimento. Ele oferece acesso ao texto completo de artigos de mais de 9.530 publicações periódicas nacionais e internacionais e a mais de 90 bases de dados com resumos de documentos¹⁹. A maioria dos títulos de periódicos encontra-se com textos na íntegra, com volumes retroativos desde os anos de 1995/96. A comunidade acadêmica de 163 Instituições de Ensino Superior e de Pesquisa do País tem acesso imediato à produção científica mundial. O uso do Portal é livre e gratuito para os usuários das instituições participantes e está em permanente desenvolvimento.

A *Scientific Electronic Library Online – SciELO*²⁰ é uma biblioteca eletrônica que abrange uma coleção selecionada de periódicos científicos no âmbito da América Latina, Caribe e Espanha. Tem como objetivo proporcionar um amplo acesso às coleções de periódicos como um todo, aos fascículos de cada título de periódico, assim como aos textos completos dos artigos. O site da SciELO é parte do Projeto FAPESP/BIREME/CNPq e um dos produtos da aplicação da metodologia para preparação de publicações eletrônicas em desenvolvimento, especialmente o módulo de interface Internet. É um *site* constantemente atualizado, tanto no seu formato como no seu conteúdo, que disponibiliza gratuitamente 298 títulos de periódicos em língua portuguesa e espanhola, apresentando textos completos e resumos com palavras-chave.

A ANPOLL tem como principal característica associar Programas de Pós-Graduação em Letras e Lingüística. Ela reúne os intelectuais da área desde 1984 e tem hoje 61 Programas

¹⁸ O endereço eletrônico do Portal é: <http://www.periodicos.capes.gov.br/portugues/index.jsp>.

¹⁹ Pesquisa efetuada em 15 de fevereiro de 2006.

²⁰ O endereço eletrônico da base SciELO é: <http://www.scielo.br/>.

filiados, fazendo-se reconhecida na comunidade acadêmica pelos Grupos de Trabalho (GTs) temáticos. Promove Encontros Nacionais para discutir a Pós-Graduação Brasileira e congrega os debates dos GTs, assim como articular e divulgar seus trabalhos.

Com o propósito de congrega os profissionais da Lingüística e tendo por objetivo promover, desenvolver e divulgar os estudos e as discussões que se realizam no Brasil entre especialistas em LINGÜÍSTICA TEÓRICA E APLICADA foi instituída a Associação Brasileira de Lingüística, a ABRALIN.

Os textos de Lingüística Aplicada que foram incluídos no *corpus* deste estudo são frutos de uma publicação do Projeto TELA, do Curso de Mestrado em Letras da Universidade Católica de Pelotas. O propósito de TELA é compilar textos completos de teses, dissertações, trabalhos apresentados em congressos, relatórios de pesquisa, periódicos e livros da área.

Sintetizando as características do *corpus* construído para esta pesquisa, pode-se dizer que se trata de um *corpus* escrito, sincrônico e contemporâneo. Por ser uma amostragem, é um *corpus* estático, fechado, de conteúdo especializado, específico e relevante quanto à densidade conceitual. Adverte-se, no entanto, que para os propósitos de geração de uma base de dados desta natureza recomenda-se que o *corpus* seja dinâmico, isto é, que possa ser continuamente monitorizado e atualizado. Quanto à língua, caracteriza-se por ser bilíngüe, português e inglês. Esse fato mostrou-se bastante produtivo e deu uma dimensão maior ao que fora inicialmente previsto, atingindo a um universo mais amplo. Seu tamanho, classificado de acordo com Sardinha (2004), é médio, uma vez que conta com 589.395 *tokens* (número total de palavras) e 19.594 *types* (tipos de palavras diferentes).

Na seleção de textos, para a formação do *corpus*, além da representatividade da fonte de coleta no âmbito acadêmico e do formato digitalizado, decidiu-se sobre o tipo do texto a servir de análise: resumo com palavras-chave. Nesse contexto, consideram-se palavras-chave as unidades léxicas selecionadas pelo autor como portadoras de maior carga temática no seu trabalho e que estão destacadas no final de cada resumo, em seqüência à indicação Palavras-chave, obedecendo às normas consensuais de formatação de um texto acadêmico.

De acordo com o constructo teórico-metodológico adotado, que privilegia a autenticidade da comunicação, as principais razões que motivaram a escolha do resumo acadêmico foram a sua densidade conceitual e o uso de palavras-chave, pois se acredita que elas representam a forma verbalizada pelos autores ao elaborarem seus trabalhos e transmitirem os resultados de suas pesquisas. Nesse sentido, Lancaster (2004) pondera que os

resumos constituem uma das partes importantes do texto para fins de leitura com propósitos de indexação.

Os textos selecionados para a formação do *corpus* foram salvos em um processador de textos, o *Word*, e reunidos em arquivos. O nome dado aos arquivos indica a fonte de origem do *corpus*, seguindo o princípio mnemônico. Para a sua identificação foram criados cabeçalhos contendo as seguintes informações: a origem, a data da coleta e o número de palavras que contêm. Devido à natureza de seu conteúdo temático (um único tema), os arquivos salvos em formato .txt foram reunidos em uma única pasta. Para fazer a análise dos textos, os arquivos foram salvos na estrutura do *software WordSmith Tools* e selecionados para uso através da opção *choose text(s)*.

5.4 Seleção das ferramentas

A opção pelas ferramentas computacionais tem origem nas vantagens que os computadores oferecem na investigação da linguagem, pois asseguram maior consistência na análise dos dados, são mais eficientes e confiáveis em tarefas tediosas e permitem uma maior abrangência na quantidade de dados com que se pode lidar. A seguir são apresentados os instrumentos usados a fim de demonstrar a possibilidade de utilizar ferramentas bastante amigáveis que, além de não exigirem conhecimentos avançados de Informática por parte do bibliotecário, são de fácil acesso.

5.4.1 Ferramentas de análise dos textos

Duas opções se apresentaram de imediato: recursos de acesso gratuito *on-line* e/ou aplicativos comercialmente adquiridos: LAEL e *WordSmith Tools*. Inicialmente utilizaram-se os recursos Concordanciador e Listador de Palavras oferecidos pelos Bancos de Dados e Ferramentas de Análise do CEPRIL/LAEL²¹, disponibilizados gratuitamente *on-line* pelo Centro de Recursos Pesquisa e Informação do Programa de Pós-Graduação em Lingüística Aplicada e Estudos da Linguagem, PUC-SP. Mesmo sendo bastante produtivos e amigáveis, esses recursos apresentam como limitações o período em que o *corpus* enviado pelo pesquisador fica disponível na rede, uma semana, e o tamanho de envio a cada remessa, que não pode ser superior a 2 MB. Acrescenta-se, como limitação, o fato de mostrar a palavra em linha e não em um contexto mais amplo, como acontece com outras ferramentas, tal como o *WordSmith Tools*, que foi usado em um segundo momento.

²¹ Disponível em: <http://www2.lael.pucsp.br/corpora/index>.

A principal ferramenta utilizada foi o *software WordSmith Tools*, um programa de análise de textos, flexível e amigável, e que aproveita os recursos do ambiente *Windows* para a análise de um *corpus*. Coloca à disposição do analista uma série de recursos extremamente úteis para a observação de diferentes aspectos da linguagem e pode ser adquirido *on-line*.

Dos seus recursos, fez-se uso de *WordList* e *Concord*. O primeiro é um listador de palavras, isto é, uma ferramenta que levanta todas as formas lexicais individuais diferentes que ocorrem no *corpus* e/ou em cada um dos seus textos e cria duas listas, uma em ordem alfabética e outra em ordem de frequência. Oferece ainda os dados estatísticos do *corpus* total ou de seus componentes (ver ANEXO A e ANEXO B). *Concord* é uma ferramenta que produz concordâncias. A concordância, quando se trata de análise textual assistida por computador, consiste de uma lista contendo uma palavra específica que corresponde à palavra de busca ou nóculo, juntamente com parte do texto que está a seu redor, o contexto. É um recurso que agiliza os procedimentos de análise de textos, pois oferece como vantagem a possibilidade de identificar todas as ocorrências de uso de um determinado termo em relação ao seu contexto (ver ANEXO C).

Nos dois casos, LAEL e *WordSmith Tools*, recomenda-se que os arquivos estejam em formato *.txt*, o que significa que contêm somente caracteres do teclado como letras, números e símbolos ortográficos, mas sem códigos de formatação específicos como ocorrem em programas como o *Microsoft Word*.

5.4.2 Instrumentos de gerenciamento de dados

O software utilizado para a constituição da base de dados foi o *Access* por já ter mostrado, em experiências anteriores, que satisfaz às necessidades que se impõem para o desenvolvimento desse projeto piloto. Ele é de fácil acesso tanto em relação ao preenchimento dos dados quanto em relação à pesquisa das informações ali armazenadas.

5.5 Coleta e análise de dados

5.5.1 Levantamento de listas

Para dar início ao exame do *corpus*, o primeiro procedimento consistiu em listar todas as palavras do *corpus* a partir da ordem alfabética e da frequência. Nos Anexos A e B, encontra-se uma amostragem dessas listas. Analisando as duas listagens percebeu-se que, para os propósitos da pesquisa, a consulta à lista por ordem alfabética se tornava mais produtiva. Dentre os fatores considerados para esta tomada de decisão, citam-se: a) havia interesse em

efetuar a pesquisa pelas unidades de significação especializada representativas da grande área da Lingüística Teórica e Aplicada; b) a consulta à lista em ordem alfabética favoreceu, em muitos casos, a identificação de variantes e da terminologia bilíngüe.

5.5.2 Seleção de palavras representativas dos grandes domínios da área

A identificação da terminologia teve como ponto de partida as listas de todas as palavras do *corpus*, obtidas pelo recurso *WordList*, por ordem alfabética e de frequência. Mas, antes mesmo da utilização do *WordSmith Tools* estas listas foram geradas pelo Listador de palavras do LAEL, que foi útil para levantar uma grande quantidade de unidades de significação especializada. No entanto, em virtude das limitações já mencionadas, esse *software* foi substituído pelo *WordSmith Tools*.

5.5.3 Produção de concordâncias.

Várias tentativas foram feitas no sentido de agilizar a coleta dos dados. Inicialmente foram selecionadas para análise as palavras mais representativas da área, tais como LINGÜÍSTICA, SEMÂNTICA, PRAGMÁTICA, DISCURSO, LINGUAGEM, TERMINOLOGIA e SOCIOTERMINOLOGIA, dentre outras, para observar as ocorrências de uso e extrair outras unidades de significação especializada. A seguir efetuou-se a pesquisa pela expressão de busca PALAVRAS-CHAVE, com a finalidade de buscar termos mais específicos e dar maior diversidade temática à base de dados. Em ambos os casos, a cada consulta feita ao *corpus*, várias unidades de significação especializada surgiram na tela e foram registradas na base.

A grande dificuldade que se apresentou de imediato foi a de coletar as unidades de significação especializada, as USEs e, simultaneamente, classificá-las em unidades terminológicas e unidades fraseológicas especializadas. A quantidade de informações gerada era imensa. A observação do contexto e da recorrência das combinações relevantes foi objeto de atenção em todo o processo de seleção das USEs, mas tornou-se mais efetiva a partir do momento em que a base já contava com um grande número de unidades de significação especializada. Optou-se, então, pela coleta e registro dos dados a partir de dois momentos distintos, como descrito no subtópico 5.6.2.

Foi a partir das informações coletadas no *corpus* que foram produzidas as concordâncias, utilizando o recurso oferecido pelo Concordanciador (ANEXOs C e D). Observando a unidade lexical no contexto (ANEXO E) se tornou possível identificar as

unidades de significação especializada, categorizá-las e registrar novos campos da base, organizando o conhecimento pelas relações que os termos estabelecem entre si.

A opção pelo campo termos relacionados, abrangendo as relações do tipo hierárquicas e não-hierárquicas, deve-se a três fatores principais: as relações hierárquicas já estão presentes nas tabelas de classificação; o *corpus* não oferece com nitidez a visualização desse tipo de relação; a organização do conhecimento não se limita mais a uma estruturação hierárquica, pois ela é fortemente marcada por relações do tipo interdisciplinar, multidisciplinar e transdisciplinar. Na área das CSH essa característica adquire relevância cada vez maior, situação que se vê retratada na falta de instrumentos disponibilizados ao bibliotecário.

5.5.4 Análise das concordâncias

A análise das concordâncias possibilitou a identificação das USEs e de todas as ocorrências de uso. Ela tornou-se possível pelo acesso que se tem ao contexto em que as USEs se apresentam. Várias relações se mostraram possíveis: aquelas que relacionam teoria e autor da teoria; unidades terminológicas e unidades fraseológicas especializadas em língua inglesa; USEs relacionadas. Foi possível também identificar casos de variação.

5.5.5 Identificação das combinações recorrentes nas concordâncias

Após a identificação das USEs verificou-se a presença de unidades que não se enquadravam no que se entende por unidade terminológica. Foi pela observação que nessa etapa da pesquisa tornou-se possível identificar a presença de estruturas fraseológicas especializadas e, até mesmo, dos modificadores que usualmente acompanham os termos.

Outras USEs ainda foram selecionadas por serem consideradas relevantes na construção da base terminológica e para serem utilizadas em catálogos eletrônicos, tais como: termos relacionados, termos equivalentes e relações interdisciplinares. A seleção das unidades terminológicas e das unidades fraseológicas em inglês foi considerada por constituírem em informação importante para o indexador; encontrar as equivalências entre termos de idiomas diferentes não é tarefa simples, pois exige além do domínio de uma segunda língua, conhecimentos lingüísticos, terminológicos e do assunto abordado. Para o usuário, essa informação amplia as condições de busca pois facilita o acesso à informação empregada pela base de dados terminológica consultada, quando a expressão de busca tem como ponto de partida o termo em inglês.

5.5.6 Seleção das USEs e dos modificadores

O critério adotado para a seleção das USEs teve como princípio a possibilidade de identificar candidatos a termo e modificadores. Entende-se por candidatos a termo toda a unidade lexical que tem grande probabilidade de ser um termo ou uma unidade fraseológica de uma área de especialidade por apresentar traços semânticos ou/e pragmáticos com a área temática. A familiaridade com a terminologia da área se mostrou de fundamental importância nessa etapa de seleção e a recorrência, um motivo de alerta. Os principais pontos de acesso para a seleção das USEs foram as unidades terminológicas representativas das grandes subáreas, como LINGÜÍSTICA, SEMÂNTICA e PRAGMÁTICA, dentre outras, e a expressão PALAVRAS-CHAVE.

5.5.7 Categorização das USEs e modificadores

A categorização das USEs como unidades terminológicas, UTs, unidades fraseológicas especializadas, UFEs, e modificadores, MODs levou em consideração o seguinte entendimento: a) UT é toda a unidade lexical de significação especializada que designa um conceito específico de um domínio temático; b) UFE é identificada pela co-ocorrência dos elementos que a constituem, apresentando grau de fixação e expressando um conteúdo próprio e uma frequência significativa dentro de uma área de especialidade; MODs são as unidades lexicais de uso freqüente, de significado geral e que são utilizadas de forma combinada com as UTs, esclarecendo ou delimitando o seu significado.

5.6 Arquitetura da base e inserção dos dados

Com o auxílio do software de gerenciamento de bases de dados *Access* foi desenhada a base para armazenar e gerenciar os dados coletados, tendo em vista a inserção da terminologia relevante classificada e relacionada. Foram criados diversos campos correspondentes às seguintes categorias: unidade de significação especializada, unidade terminológica, unidade terminológica em inglês, unidade fraseológica especializada, unidade fraseológica especializada em inglês, autor relacionado, modificador, termos relacionados, termos equivalentes, relações interdisciplinares. Para visualizar a configuração da base de dados terminológica, ver o QUADRO 1, cujos campos são a seguir especificados.

QUADRO 1

Arquitetura da base de dados terminológica

N.º	USE	UT	UTI	UFE	UFEI	AR	MOD	TRs	TEs	RIs

Legenda: N.º - número de série; USE - Unidade de Significação Especializada; UT - Unidade Terminológica; UTI - Unidade Terminológica em língua inglesa; UFE - Unidade Fraseológica Especializada; UFEI - Unidade Fraseológica Especializada em língua inglesa; AR - Autor Relacionado; MOD - Modificador; TRs - Termos Relacionados; TEs - Termos Equivalentes; RIs - Relações Interdisciplinares.

a) Unidade de Significação Especializada (USE): unidade lexical composta de uma ou mais palavras que contenha traços semânticos e/ou pragmáticos característicos da área temática, podendo ser a forma lexical plena ou reduzida como sigla, acrônimo, abreviatura e fórmula. Uma USE configura-se como possível candidato ao estatuto de unidade terminológica, unidade fraseológica ou modificador;

b) Unidade Terminológica (UT): unidade de significação especializada, composta de uma ou mais palavras, que contenha um conceito específico da área temática;

c) Unidade Terminológica em língua inglesa (UTI): unidade de significação especializada, composta de uma ou mais palavras, que contenha um conceito específico da área temática, em língua inglesa, na forma como se apresenta no *corpus*;

d) Unidade Fraseológica Especializada (UFE): combinação recorrente de palavras semanticamente vinculadas ao conteúdo da área temática;

e) Unidade Fraseológica Especializada em língua inglesa (UFEI): combinação recorrente de palavras semanticamente vinculadas ao conteúdo da área temática, em língua inglesa, conforme se apresenta no *corpus*;

f) Autor relacionado (AR): o nome do teórico relacionado com a teoria expressa na unidade terminológica;

g) Modificador (MOD): unidade lexical que não se constitui como termo, mas que é de uso freqüente, de significado não especializado e utilizada de forma combinada com uma unidade terminológica, esclarecendo ou delimitando o seu significado;

h) Termos relacionados (TRs): unidade lexical que mantêm algum tipo de relação associativa com a unidade terminológica;

i) Termos equivalentes (TEs): unidades lexicais não recomendadas como preferencial para uso, constituindo-se como uma variação de uma unidade terminológica recomendada;

j) Relações interdisciplinares (RIs): unidades lexicais de diferentes áreas do conhecimento que mantêm algum tipo de relação conceitual entre si.

5.6.1 Organização dos relacionamentos

A expansão do registro de uma USE permitiu não apenas identificar outras unidades terminológicas e unidades fraseológicas especializadas, como também perceber os relacionamentos que ocorrem no âmbito de uma área de especialidade, sendo possível, a partir de então, o registro de termos relacionados, termos equivalentes e modificadores.

5.6.2 Preenchimento dos registros terminológicos

O preenchimento dos registros terminológicos foi feito, inicialmente, à medida que a informação surgia e era identificada para uso. Visando dar maior produtividade à coleta e registro dos dados, optou-se por um novo procedimento, procedendo-se ao preenchimento dos dados da base em duas etapas: a) coleta das USEs; b) coleta das demais informações, como termos relacionados, termos equivalentes, autores relacionados, modificadores e relações interdisciplinares.

O quadro seguinte (QUADRO 2) apresenta uma simples amostra de como os dados foram inseridos em seus respectivos campos.

QUADRO 2
Terminologia da base de dados terminológica em Linguística Teórica Aplicada

Nº	USE	UT	UTI	UFE	UFEI	AR	MOD	TR's	TE's	RI's
1	alfabetização	ALFABETIZAÇÃO	alphabetsisation; literacy	-	-	-	-	ALFABETIZAÇÃO. APRENDIZAGEM. ENSINO. ESCOLARIDADE. ESCRITA. LEITURA. LETRAMENTO ESCOLAR. INTERAÇÃO.	-	-
2	alofonia	ALOFONIA	allophony	-	-	-	-	ALOFONE. FONEMA. FONÉTICA. FONOLOGIA.	-	-
3	análise de marcadores discursivos	MARCADORES DISCURSIVOS	discourse markers	-	-	-	-	-	-	-
4	análise do discurso crítico	DISCURSO CRÍTICO	critical discourse	-	-	-	ANÁLISE	-	-	-
5	aquisição da leitura	AQUISIÇÃO DA LEITURA	acquisition of the reading	-	-	-	-	ALFABETIZAÇÃO. DISLEXIA. ENSINO. ESCRITA. FONOLOGIA. LEITURA. LETRAMENTO. NEUROLINGÜÍSTICA. PSICO-LINGÜÍSTICA.	-	-
6	coerência textual	COERÊNCIA TEXTUAL	coherence	-	-	-	-	LINGÜÍSTICA TEXTUAL. PRODUÇÃO TEXTUAL. TEXTO.	-	-
7	discurso publicitário	DISCURSO PUBLICITÁRIO	advertisement discourse	-	-	-	-	ANÁLISE DO DISCURSO. PUBLICIDADE.	-	-
8	fonética acústica	FONÉTICA ACÚSTICA	acoustic phonetic	-	-	-	-	ENTONAÇÃO. FONÉTICA. FONÉTICA ARTICULATÓRIA. FONÉTICA AUDITIVA. SOM.	-	-
9	gramática gerativa	-	generative grammar	-	-	NOAM CHOMSKY	-	GERATIVISMO. GRAMÁTICA GERATIVA. LINGÜÍSTICA GERATIVA. MORFOSSINAXE. SINAXE GERATIVA. TEORIAS COGNITIVAS DA LINGUAGEM.	gramática gerativa	-
10	processo de alfabetização	ALFABETIZAÇÃO	alphabetsisation; literacy	PROCESSO DE ALFABETIZAÇÃO	alphabetsisation process	-	-	ALFABETIZAÇÃO. APRENDIZAGEM. ENSINO. ESCOLARIDADE. ESCRITA. LEITURA. LETRAMENTO ESCOLAR. INTERAÇÃO.	-	-
11	programa minimalista	MINIMALISMO	minimalist program	-	-	NOAM CHOMSKY	-	GERATIVISMO. GRAMÁTICA GERATIVA. LINGÜÍSTICA GERATIVA. MORFOSSINAXE. SINAXE GERATIVA. TEORIAS COGNITIVAS DA LINGUAGEM	programa minimalista. teoria chomskiana	-

6 ANÁLISE DOS DADOS E DISCUSSÃO DOS RESULTADOS

Conforme foi relatado no capítulo anterior, as informações coletadas no *corpus* foram registradas na base *Access*. Após categorizar as USEs em unidades terminológicas, unidades fraseológicas especializadas e modificadores foi possível observar que o *corpus*, mesmo tendo uma limitação quanto ao tamanho, se mostrou bastante produtivo, seja pela diversidade terminológica apresentada como pelas relações semânticas que possibilitou fazer. De fato, o *corpus* não só conseguiu mostrar a riqueza terminológica da área como também permitiu a extração de unidades em língua inglesa, dando à base uma dimensão bilíngüe, tal como a identificação de relacionamentos entre autores e teorias, entre termos relacionados e termos equivalentes.

Deve ser mencionado que o *site* de busca *Google Scholar*, embora inferior quanto à seleção da qualidade das fontes, por representar um universo de pesquisa bem superior quantitativamente mostrou-se uma ferramenta auxiliar de grande importância para a validação da terminologia da área. Em determinadas ocasiões ajudou a determinar o estatuto terminológico de um termo, especialmente quando se tratava de decidir entre termos equivalentes qual o mais usado. Ele foi útil também para auxiliar na categorização de unidades de significação especializada, quando as informações oferecidas pelo *corpus* não eram suficientes.

Visando avaliar a pertinência da terminologia coletada segundo a metodologia proposta, procurou-se utilizá-la fazendo buscas em diferentes bibliotecas. Desse modo, se pretendeu verificar a adequação dos descritores/termos utilizados pelos bibliotecários em catálogos eletrônicos à linguagem expressa pelos especialistas da área, registrada em resumos de publicações de caráter técnico-científico.

Para tanto, construiu-se uma tabela cujas expressões de busca foram as unidades terminológicas e as unidades fraseológicas especializadas registradas na base de dados terminológica. A seleção destas unidades foi aleatória, mas teve-se cuidado para que fossem representativas da terminologia de uma área, em diferentes níveis de especificidade. Pelo fato de se utilizar como parâmetros catálogos eletrônicos nacionais e internacionais, as expressões de busca foram registradas tanto em língua portuguesa como em língua inglesa.

Os catálogos eletrônicos selecionados para essa análise foram construídos por bibliotecários e apresentam acervo representativo na área de LINGÜÍSTICA TEÓRICA E

APLICADA. São eles: SABi/UFRGS, VERUM/PUCRS, ACERVUS/UNICAMP, SIBi/USP, LIBRARY OF CONGRESS, BIBLIOTHÈQUE ET ARCHIVES CANADA (AMICUS).

A pesquisa feita na base de dados SABi/UFRGS e VERUM/PUCRS foi feita a partir da opção Pesquisar palavras-simples, Campo a pesquisar: Assunto. No caso das unidades terminológicas compostas por mais de uma palavra e das unidades fraseológicas especializadas, a opção utilizada foi palavras adjacentes. Com esse procedimento pretendia-se preservar o conteúdo conceitual de uma expressão de busca.

O banco de dados bibliográficos do Sistema de Bibliotecas da UNICAMP, ACERVUS, é formado por duas bases de dados: monografias (livros e teses) e periódicos. A pesquisa prevista não foi possível pelo fato do sistema apresentar-se inoperante para migração de versão, no momento da pesquisa.

Em consulta ao Banco de dados Bibliográfico da USP (DEDALUS) selecionou-se a opção Busca e adotou-se o seguinte processo de pesquisa: selecionar campos: assunto; selecionar base: todas.

Na consulta feita à LIBRARY OF CONGRESS a pesquisa foi feita pelas opções: basic search e subject browse.

Na base AMICUS, da BIBLIOTHÈQUE ET ARCHIVES CANADA, a opção selecionada foi recherche élémentaire, campo: subject (mot-clé). Nesta base os descritores/termos aparecem em francês e em inglês; a pesquisa foi feita com a expressão de busca na língua inglesa.

A pesquisa foi realizada em 07/02/2006. Os resultados das buscas encontram-se registrados no TABELA 1. A seguir, procedeu-se à análise comparativa dos dados.

TABELA 1

Comparativo entre UT/UFE e DESCRITORES/TERMOS

UT/UFE	UT/UFE (em inglês)	SABi /UFRGS	PUCRS	SIBI/ USP	LC	AMICUS
Lingüística	Linguistics	5153	6454	2120	2120	500
Análise do discurso	Discourse analysis	1401	460	557	1745	452
Aquisição da linguagem	Language acquisition	195	291	83	1099	268
Atlas lingüístico	Atlases linguistics	27	-	6	-	-
Coesão textual	Cohesion (Linguistics)	43	-	13	39	2
Competência e performance	Competence and performance (Linguistics)	-	-	-	55	13
Competência lingüística	Linguistic competence	7	-	2	-	5
<i>Corpus</i> eletrônico	Electronic <i>corpus</i>	-	-	-	-	-
Ergatividade	Ergativity	9	-	-	-	-
Gramaticalização	Grammaticalization	17	-	-	31	1
Intertextualidade	Intertextuality	285	14	-	387	38
Linguagem e cognição	Language and cognition	-	-	20	-	25
Linguagem e subjetividade	Language and subjectivity	2	-	-	-	-
Lingüística Aplicada	Applied linguistics	210	510	636	211	34
Lingüística Descritiva	Descriptive linguistic	13	-	2	-	125

(continua)

(continuação)

UT/UFE	UT/UFE (em inglês)	SABi /UFRGS	PUCRS	SIBI/ USP	LC	AMICUS
Linguística de <i>Corpus</i>	<i>Corpus Linguistics</i>	6	-	-	-	-
Linguística Histórica	Historical linguistics	46	27	251	208	8
Metalinguagem	metalanguage	35	5	20	54	-
Paradigma verbal	verbal paradigm	-	-	-	-	-
Polissemia	polysemy	13	1	8	20	1
Pragmática	pragmatics	124	103	218	532	19
Semântica	semantics	664	453	575	1730	12
Semântica cognitiva	cognitive semantics	7	-	-	-	-
Sociolinguística	sociolinguistics	314	216	529	1159	299
Socioleto	sociolect	1	-	-	-	-
Socioterminologia	socioterminology	7	-	-	-	-
Linguística do texto	textual linguistics	188	210	-	2	
Linguística Textual	textual linguistics	9	23	-	34	
Varição sociolinguística	variation sociolinguistics	13	210	-	-	-

Esta análise parte da seleção de 29 unidades de significação especializada e cinco catálogos eletrônicos. Tinha-se como idéia inicial também utilizar os dados do Sistema de Bibliotecas da UNICAMP, mas isto não foi possível, conforme razões já mencionadas. Assim sendo, fazem-se as seguintes considerações:

a) quando se trata de domínios e subdomínios abrangentes como: LINGÜÍSTICA, ANÁLISE DO DISCURSO, AQUISIÇÃO DA LINGUAGEM, LINGÜÍSTICA APLICADA, PRAGMÁTICA, SEMÂNTICA e SOCIOLINGÜÍSTICA, a coincidência terminológica entre a linguagem utilizada pelos especialistas e os Sistemas se evidencia. Nesses casos, a comunicação entre os SRIs e os usuários se efetiva;

b) há casos que, mesmo em se tratando de domínios e subdomínios abrangentes, a sua representatividade nos catálogos eletrônicos é muito baixa. Nesta situação foram identificados os seguintes descritores/termos: ATLAS LINGÜÍSTICO, *CORPUS* ELETRÔNICO, LINGÜÍSTICA DESCRITIVA, LINGÜÍSTICA DE *CORPUS*, SEMÂNTICA COGNITIVA e SOCIOTERMINOLOGIA. Nesse caso, a falha que se identifica é pela omissão de conceitos importantes, fato já mencionado por Guinchat e Menou no tópico 3.7, quando fazem referência às causas de insuficiência de um SRI. Como resultado, a comunicação entre os SRIs e os usuários não se efetiva com eficácia;

c) a omissão de conceitos em SRI fica evidenciada quando se verificam, no título de um trabalho, expressões utilizadas como descritores/termos dentro do sistema, mas ausentes no campo de assunto daquele registro. Esse foi o caso de SEMÂNTICA COGNITIVA. Analisando os demais descritores/termos utilizados para representar a temática do documento, verificou-se que ele foi indexado por um descritor/termo de nível hierárquico maior, como SEMÂNTICA e LINGÜÍSTICA. Percebe-se aqui mais uma situação de falha, pois uma informação indexada apenas pelo descritor/termo genérico inviabiliza a recuperação de documentos e temáticas quando se necessita localizar informações mais específicas. Nesse caso a situação é agravada pelo fato de tratar-se de um descritor/termo que ocupa um nível hierárquico elevado em um sistema estruturado de conceitos;

d) no que se refere à especificidade temática, comprova-se o que já se previa nas hipóteses deste trabalho. Não se verifica na maioria dos catálogos eletrônicos consultados um tratamento consistente quanto à informação processada em sua especificidade. Descritores/termos como: COESÃO TEXTUAL, COMPETÊNCIA LINGÜÍSTICA, ERGATIVIDADE, GRAMATICALIZAÇÃO, POLISSEMIA, SOCIOLETO E VARIAÇÃO SOCIOLINGÜÍSTICA não são contemplados em todas as bases;

e) observando a questão da consistência, em bases nacionais, verifica-se o uso de dois descritores/termos para um único conceito: LINGÜÍSTICA DO TEXTO e LINGÜÍSTICA TEXTUAL, ANÁLISE DO DISCURSO e DISCURSO (ANÁLISE). Em ambos os casos o Sistema não disponibiliza recurso algum para relacioná-los no momento da busca e recuperação. No caso

dos descritores ANÁLISE DO DISCURSO e DISCURSO (ANÁLISE) a situação ainda se torna mais grave, pois não há a preservação de sentido do conceito indexado;

f) foram apresentadas para análise três unidades de significação especializada, representando uma combinação recorrente de palavras semanticamente vinculadas ao conteúdo da área temática: COMPETÊNCIA E PERFORMANCE, LINGUAGEM E COGNIÇÃO e LINGUAGEM E SUBJETIVIDADE. O tratamento atribuído a unidades dessa natureza não apresenta uma uniformidade entre as bases e a presença de vazios terminológicos parece existir pela falta de precisão atribuída aos conceitos indexados. Em consulta feita ao *Google Scholar* para atestar ou não a sua representatividade em uma área de conhecimento, observa-se que tais unidades têm presença no *site*: COMPETÊNCIA E PERFORMANCE – 146, LINGUAGEM E COGNIÇÃO – 627 e LINGUAGEM E SUBJETIVIDADE – 282;

g) em pesquisa feita ao catálogo eletrônico SIBi/USP pela expressão de busca VARIAÇÃO AND SOCIOLINGÜÍSTICA, localiza-se uma obra; entretanto, há 495 registros sobre Sociolingüística. Tal fato é um forte indício da falta de especificidade com que os catálogos eletrônicos vêm tratando temas específicos.

Ao analisar a informação veiculada pelos catálogos eletrônicos de assunto teve-se como propósito verificar a adequação dos descritores/termos utilizados, em relação à linguagem expressa pelos produtores do conhecimento. Acredita-se que havendo a coincidência entre a linguagem utilizada por ambos, o processo de comunicação a ser estabelecido entre os SRIs e os usuários se efetive com melhores resultados.

Pela análise dos dados verificou-se que os vazios terminológicos ocorrem não apenas em situações de uma indexação específica, mas também quando se trata de uma indexação mais genérica. Esse fato ficou comprovado pela discrepância apresentada nos resultados de pesquisas feitas no campo de assunto e de título com a mesma expressão de busca. Uma das possíveis causas dessa omissão pode ser o resultado da falta de ferramentas que ofereçam, ao bibliotecário que indexa, informações adequadamente estruturadas sobre a terminologia de uma área.

Também foi constatado nos catálogos eletrônicos nacionais que a variação terminológica está presente, mas que não recebe o tratamento necessário para assegurar resultados de pesquisa representativos dos dados que a base armazena. O mesmo não se observa nos catálogos eletrônicos internacionais; neles o controle terminológico existe por

meio do uso de remissivas. A variação em uma base de dados, quando não controlada, dispersa os resultados de pesquisa e gera perda de informação.

Se compararmos as bases pelo critério de consistência, será possível observar um tratamento altamente diferenciado quanto à relevância com que os temas específicos vêm sendo tratados. O descritor/termo COESÃO TEXTUAL, por exemplo, foi assim considerado pelas cinco bases que nos serviram de amostra: 43, 0, 13, 39, 0, e GRAMATICALIZAÇÃO: 17,0, 0, 31, 1. Trata-se, aqui, de dois descritores/termos cujo nível de especificidade não é tão alto e cuja frequência é bastante elevada em textos escritos, conforme atestam *sites* de busca como o *Google Scholar*. COESÃO TEXTUAL, por exemplo, apresenta como resultado de busca um total de 10.800 e GRAMATICALIZAÇÃO, 556.

É importante ressaltar, outrossim, que os catálogos eletrônicos selecionados para esta análise têm um acervo representativo na área da LINGÜÍSTICA TEÓRICA E APLICADA e que o *Google Scholar*, mesmo apresentando como resultado de pesquisa todas as ocorrências em um texto e não apenas sua presença em palavras-chave como identificadoras de um assunto, não invalida a comparação feita que tem como propósito apenas comprovar seu uso.

No que se refere ao tratamento dado às unidades fraseológicas parece haver, no âmbito bibliotecário, uma certa resistência em tratá-las como descritores/termos. Nesses casos, o que se observa com maior frequência é o uso de descritores/termos isolados, tais como: COMPETÊNCIA; PERFORMANCE; LINGUAGEM; COGNIÇÃO; SUBJETIVIDADE e não COMPETÊNCIA E PERFORMANCE, LINGUAGEM E COGNIÇÃO e LINGUAGEM E SUBJETIVIDADE. Ressalta-se, no entanto, que as unidades fraseológicas têm representatividade entre os especialistas de uma área e seu uso já começa a aparecer nas bases de dados internacionais. Assim, tornam-se indispensáveis quando se pensa em oferecer um produto que tenha como principal objetivo transferir informações e comunicar o conhecimento produzido.

Na área analisada – LINGÜÍSTICA TEÓRICA E APLICADA – pode-se perceber que os instrumentos que estão sendo utilizados, para as atividades de indexação temática de um documento, não oferecem uma linguagem que reflita aquela utilizada nos documentos processados. Com frequência a indexação tem sido feita a partir de palavras e não de unidades de significação especializada.

A pertinência da terminologia apresentada pelos catálogos eletrônicos só se verifica em um nível altamente genérico. A grande falha identificada refere-se à falta de

descritores/termos representando temas específicos, salientando-se que esse problema não ocorre apenas quando se trata de uma terminologia de alta especificidade.

Confirma-se, ainda, que a variação terminológica, sem o adequado controle pelo uso de remissivas, também se apresenta como uma das falhas dos SRIs. Vale lembrar que essa última observação não se contrapõe à idéia da presença de variação em uma ferramenta de análise documentária; ela certamente é inevitável, mas precisa ser controlada para que se evite perda de informação nos processos de busca e recuperação.

7 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Para a finalidade desta Dissertação a aproximação entre a Documentação e a Terminologia justifica-se por uma série de semelhanças que se apresentam; ambas têm evoluído teoricamente com vistas à aplicação e adotam uma atitude pragmática com o propósito de resolver problemas de comunicação. No caso da Documentação, mediante a análise do conteúdo dos documentos e em relação à Terminologia, segundo a descrição e a normalização dos processos de formação de termos. Nesse âmbito, tanto uma quanto a outra dependem, em grande medida, de fatos empíricos tais como: o uso lingüístico, as necessidades do usuário e suas preferências.

Enquanto a Documentação focaliza o texto e seus constituintes, o ponto de interesse da Terminologia reside nos termos e nas suas denominações, nos conceitos e na sua representação, mas em se tratando de palavras-chave, índices e tesauros, parece que seus campos se sobrepõem. Indexadores e analistas fazem uso da terminologia, tal como os terminólogos e terminógrafos. Todos esses profissionais estão a serviço da comunicação científica e almejam que o usuário utilize, de modo independente, o resultado de seu trabalho. No que diz respeito às línguas observa-se também algumas similitudes, pois as duas submetem a linguagem à unificação e normalização, visando alcançar melhores resultados no processo de comunicação.

Quanto ao tipo de apoio técnico que necessitam é possível notar alguns pontos de coincidência; uma e outra têm se beneficiado efetivamente da nova tecnologia da informação. A tecnologia a serviço dos catálogos eletrônicos, das bases de dados e das redes de informação é, por sua vez, importante para a Documentação; sem dúvida, é nos métodos automatizados que as duas ciências se auxiliam mais diretamente uma à outra. A extração automática de palavras-chave passa, em princípio, pelo mesmo processo de reconhecimento automático de termos; as rotinas de manipulação de dados no armazenamento e recuperação da informação contêm mais traços coincidentes do que divergentes.

Documentação e Terminologia perseguem os mesmos objetivos gerais e servem a grupos de usuários muito parecidos; um maior entendimento das unidades paradigmáticas das linguagens especializadas apresenta vantagens importantes para a Documentação. Os propósitos do terminólogo em alcançar uma maior unidade, consistência e clareza de expressão na comunicação especializada podem simplificar os trabalhos do documentalista.

Essas duas áreas de conhecimento têm como objeto de estudo os textos especializados e, como função, a representação e a comunicação do conhecimento e a transferência da informação. A aproximação entre esses dois domínios é hoje um caminho a percorrer e que, certamente, contribuirá para que a comunicação a ser estabelecida, entre os SRIs e os usuários, alcance melhores resultados.

Enquanto a Documentação se preocupa com a organização e o tratamento bibliográfico informacional de textos especializados, com o objetivo de facilitar as condições de acesso às informações codificadas nos catálogos eletrônicos, a Terminologia se ocupa com a análise e a descrição das unidades que representam e comunicam o conhecimento.

O fato é que tanto a Documentação como a Terminologia têm como objetivo facilitar a comunicação no âmbito das linguagens de especialidade. A Documentação elabora índices, cujas unidades indexadoras representam o léxico especializado registrado nos documentos. Tais índices têm o propósito de melhorar a compreensão e a representação dessas informações para uma futura recuperação desses documentos. A Terminologia, por sua vez, compila, estrutura e organiza as informações sobre o léxico utilizado em uma área do conhecimento, visando melhorar o fluxo de acesso às informações. Dessa forma, o trabalho produzido pelos terminólogos precisa ser alvo de reflexões pelos profissionais da área biblioteconômica.

Em sua origem, a Terminologia foi concebida como um ramo da Lingüística Aplicada e entre ambas podem ser estabelecidas algumas aproximações. Enquanto a Terminologia tem o conceito como ponto de partida, com o objetivo de estabelecer claramente os limites conceituais das unidades terminológicas, a Lingüística aborda o conteúdo das palavras para referir-se à língua geral; para os terminólogos somente as denominações dos conceitos são importantes, isto é, o léxico da língua, enquanto que para os lingüistas a morfologia e a sintaxe são relevantes; a investigação terminológica considera a língua sob o ponto de vista sincrônico, enquanto a Lingüística a considera sob a ótica sincrônica e diacrônica.

Esta pesquisa buscou subsídios nas três áreas do conhecimento: Documentação, Lingüística e Terminologia. Ela fundamentou sua proposta de geração de bases de dados terminológica na Teoria da Enunciação, de Benveniste, na Teoria Comunicativa da Terminologia e na Socioterminologia. Utilizou conhecimentos advindos da Lingüística de *Corpus* no processo de construção de um vocabulário controlado para uso dos bibliotecários em suas atividades de indexação. E registrou o conteúdo temático dos documentos expresso pelo uso de termos, fraseologias especializadas e modificadores.

Pelo uso da Teoria da Enunciação pretendeu-se enfatizar a importância da preservação do sentido desejado pelo autor do texto até a busca e recuperação da informação. É relevante para esta pesquisa a maneira como Benveniste aborda a questão do sentido, sendo um dos pioneiros a perceber a importância do texto e do discurso para que a palavra signifique; ele observa que, fora do contexto, nem sempre é possível a compreensão do sentido que a palavra representa. Por isso enfatiza a importância de considerar, além do contexto, o papel desempenhado pelo locutor ao utilizar as palavras para expressar suas idéias, comunicar seu conhecimento e experiências, de forma a transmitir informações. O bibliotecário, enquanto interlocutor, é o responsável pela transmissão da informação em nível nacional [e internacional] (HOLZEM, 1998).

Desse modo, o sentido de uma palavra é alcançado quando está relacionado ao contexto de uso, enquanto que a comunicação só se efetiva pela relação de intersubjetividade que se estabelece entre o locutor e o alocutário, em um tempo e lugar determinado. É nessa situação de comunicação que as informações veiculam.

Essa teoria, no contexto deste trabalho, é entendida pela necessidade de se assegurar a comunicação entre o autor do texto e o bibliotecário e entre os Sistemas de Recuperação e o usuário. O propósito é o de preservar o conteúdo temático do documento a ser indexado, transferindo ao usuário a informação expressa pelo autor do documento sem perda de sentido.

O uso da Teoria Comunicativa da Terminologia e da Socioterminologia, como referencial teórico, justificou-se plenamente. Ambas enfatizam a importância do aspecto comunicativo e admitem a variação conceitual e denominativa do termo, enquanto unidade de conhecimento e de comunicação e portador de uma estrutura complexa e poliléxica. Ademais, valorizam o texto como objeto de análise para entender o real significado de um termo a partir do seu contexto de uso.

A opção pelo termo, conforme concebido pela Terminologia, mostrou-se válida fundamentalmente por reconhecê-lo como unidade de significação especializada em uma área do conhecimento e, como tal, expressar o conteúdo conceitual das ciências e das técnicas. O termo, outrossim, se define por sua relação a um conjunto de significados de um mesmo domínio e se caracteriza pela monorreferencialidade tão buscada pela análise documentária.

Ao descreverem conteúdos específicos, os termos estão mais propensos a eliminarem as ambigüidades que ocorrem com freqüência na linguagem natural, tendendo, assim, a

alcançar a precisão conceitual e propiciar uma adequada comunicação do conhecimento entre especialistas e entre eles e os usuários de um modo geral.

Em relação à geração de bases de dados terminológicas, encontrou-se na Terminografia e na Lingüística de *Corpus* valiosa contribuição. Ambas se constituíram em apoio substancial para a fundamentação dos princípios metodológicos adotados. Em especial, é preciso salientar que a operacionalização dos procedimentos terminográficos foi otimizada, em grande parte, pelos subsídios advindos da Lingüística de *Corpus*, permitindo a análise de um considerável volume de informações sob forma automatizada.

Após considerações enfatizando a produtividade obtida pela interface entre Documentação, Terminologia e Lingüística, apresentam-se, aqui, as reflexões geradas a partir da análise dos dados. Essa análise teve como propósito comparar os descritores/termos utilizados pelos bibliotecários ao indexar os documentos e utilizados como pontos de acesso nos catálogos eletrônicos, com o que aqui se sugere e que se entende como Unidade de Significação Especializada.

Pela análise do *corpus* selecionado para esta pesquisa, acredita-se que foi possível tomar conhecimento da terminologia que realmente circula na época atual entre os especialistas da área de especialidade – LINGÜÍSTICA TEÓRICA E APLICADA –. E que foi possível, também, selecionar e registrar termos para a constituição de uma base terminológica, conciliando os critérios de autenticidade e representatividade da terminologia que se faz necessária quando se pretende estabelecer a comunicação entre os catálogos eletrônicos e os usuários.

O uso de ferramentas automatizadas de recolha de termos mostrou-se produtivo, capaz de agilizar a análise de textos e revelar informações essenciais para a geração de uma base de dados terminológica, como a que aqui se está propondo. Com a análise feita a partir desses dados foi possível identificar as regularidades da terminologia expressa pelos especialistas, revelando sua sistematicidade pelas construções e concordâncias.

Analisando as séries de ocorrência de um termo dado foi possível identificar as variações e as unidades fraseológicas mais freqüentes e pertinentes e, a partir de então, estruturar esse conjunto de informações em um registro para compor uma base terminológica. A ferramenta facilitou inclusive o acesso aos neologismos, às definições e explicações relativas ao conhecimento produzido, revelando e relacionando teorias e teóricos da área. E tudo indica que, pela sua estrutura e composição, é de fácil atualização.

A área definida para a constituição deste projeto piloto foi a LINGÜÍSTICA TEÓRICA E APLICADA, hoje também conhecida como ESTUDOS DA LINGUAGEM. O motivo pelo qual se decidiu considerá-la como foco desta pesquisa foi o propósito de dar continuidade a um trabalho que vinha sendo desenvolvido na Biblioteca Setorial de Ciências Sociais e Humanidades.

Essa base de dados em ESTUDOS DA LINGUAGEM tem hoje 13.738 registros. Isso não significa que tenha o mesmo número de descritores/termos, uma vez que ela contempla a variação e admite a reversão dos termos que compõem as cadeias hierárquicas sob as quais se estrutura. Ela foi construída com a supervisão de alguns professores dessa Instituição – Instituto de Letras da UFRGS – e foi o resultado de projetos encaminhados à FAPERGS e ao CNPq.

A idéia inicial era dar continuidade a essa base. Mas novos horizontes se abriram com o conhecimento adquirido no decorrer do Curso de Mestrado e as leituras efetuadas, sendo a formação de um *corpus*, apenas um dentre tantos. O que antes era feito de forma empírica, agora passou a ter sustentação teórica e, como tal, um rigor que se pretende científico.

A base que se propõe com esta Dissertação se sustenta, fundamentalmente, a partir de três conceitos: termo, fraseologia especializada e modificador. A relação hierárquica, antes admitida como essencial, perdeu a sua relevância diante da relação custo/benefício quando se trata de construir uma ferramenta documentária desse gênero. À complexidade de se estruturar uma área de conhecimento a partir das relações do tipo hierárquicas, soma-se a dificuldade operacional da nova estrutura, em que o conhecimento se apresenta fortemente marcado pela inter, multi e transdisciplinaridade. Cada vez mais, o que parece relevante é ter um conhecimento e domínio da terminologia de uma área de especialidade e das relações que se estabelecem entre os termos, de forma a ser possível assegurar a preservação do conhecimento e da informação a ser transferida.

A partir das hipóteses apresentadas no início desta pesquisa, fazem-se as seguintes considerações. Conclui-se pela sua confirmação em relação às três primeiras. De fato, a informação que veicula nos catálogos eletrônicos deixa a desejar quanto à consistência e a especificidade temática. Esta situação possivelmente é o resultado da falta de ferramentas apropriadas para a indexação de documentos, por não privilegiarem uma linguagem específica e nem a linguagem expressa pelos especialistas; essa última, possivelmente, decorre da falta de especificidade que essas ferramentas apresentam. A coincidência terminológica só existe em nível altamente genérico. A Lingüística de *Corpus* e a Terminografia se apresentaram

como um recurso altamente produtivo na geração de bases de dados terminológica, não apenas por permitir análise de um grande volume de informações, mas por revelar relações entre conceitos e a terminologia efetivamente em uso. Quanto à última hipótese, ela só poderá ser comprovada quando as bases de dados terminológicas, assim constituídas, forem efetivamente utilizadas em bibliotecas, em catálogos eletrônicos.

Ao finalizar estas considerações ressalta-se ainda que o essencial no processo de indexação é conhecer a terminologia de um determinado domínio e as relações que se estabelecem entre os termos, hierárquicas ou não, para que seja possível selecionar aqueles que melhor representam os conceitos emitidos pelo autor do documento. A partir desse conhecimento fica facilitada a possibilidade de contextualizar o termo recomendado para uso e selecionar termos de diferentes níveis de especialização, oferecendo uma maior precisão em processos de busca e recuperação da informação.

O que se defende é que, por ser a estruturação do conhecimento em forma hierárquica indispensável em tabelas de classificação, ela pode ser dispensada em ferramentas que visam o controle da terminologia a ser utilizada na representação da informação que um documento oferece, como é o caso dos vocabulários controlados.

Como foi visto na revisão de literatura, no tópico que trata do processo de indexação (3.6), o que é determinante para o bibliotecário que indexa é o saber que ele acumula, a sua capacidade de ler e interpretar e o conhecimento da terminologia da área que indexa. Uma ferramenta como a que se está aqui propondo oferece ao bibliotecário a possibilidade de conhecer a terminologia que veicula entre os especialistas de uma determinada área e de estabelecer relações relevantes entre os termos. Ressalta-se que, mesmo não sendo claramente expressas, as relações hierárquicas relacionadas com um determinado termo estão presentes.

Nesse entendimento, é de fundamental importância que um mesmo conceito seja verbalizado por uma única forma. Além disso, é preciso que um conceito indexado seja contextualizado com o uso de termos relacionados, para que se possam obter resultados precisos, num equilíbrio entre precisão e revocação. Não se pode esquecer tampouco que conceitos específicos e genéricos devem ser expressos sempre que se mostrarem relevantes para representar o conteúdo temático de um documento.

Desse modo, faz-se necessário disponibilizar, ao bibliotecário, ferramentas em áreas específicas do conhecimento e que tenham como proposta satisfazer às necessidades antes

mencionadas. Acredita-se que assim será possível estabelecer uma relação comunicativa entre usuários e SRIs.

Acrescenta-se como vantagem à proposta apresentada, as facilidades que um instrumento dessa natureza apresenta, pois esta depende fundamentalmente de um *corpus* criteriosamente construído, de *softwares* que viabilizem análise de textos. Ao lado desses recursos materiais é indispensável a atuação de profissionais bibliotecários que tenham conhecimento da área a ser indexada e dos princípios expressos pela Terminologia, especialmente os que se referem às unidades de significação especializada: termos e fraseologias especializadas.

A partir desta Dissertação, propõe-se o desenvolvimento de estudos e de pesquisas que visem dar continuidade à reflexão aqui apresentada e, dessa forma, beneficiar a todos os profissionais que estão envolvidos com tarefas que têm como propósito a indexação, a recuperação e a transferência do conhecimento.

Com relação ao que aqui foi exposto, sugere-se que este projeto piloto tenha continuidade, mas que seja desenvolvido em três etapas. A primeira consistirá na ampliação do *corpus* que deu origem a esta base, para que tenha representatividade junto aos usuários em virtude de uma maior adequação e abrangência de cobertura da área. A segunda será a testagem e a avaliação de sua utilização em uma biblioteca universitária. Em prosseguimento propõe-se que se dê continuidade à Base de Dados Terminológica de Lingüística Teórica e Aplicada aqui iniciada, visando seu uso pelos bibliotecários de bibliotecas universitárias e oferecendo-lhes uma ferramenta que seja fundamentada no princípio da comunicação a ser estabelecida entre SRI e usuário. A colaboração e a assistência do especialista da área será de indiscutível importância nessa etapa do trabalho. A aplicação desta metodologia para a geração de bases de dados terminológicas de outras áreas de saber das Ciências Sociais e Humanidades preencherá a carência de ferramentas que hoje estão disponibilizadas ao bibliotecário.

Também como perspectiva futura sugere-se um estudo de tipologia textual, no âmbito de uma política de indexação. Estudos nesse sentido auxiliariam na tomada de decisões a respeito da seleção dos conceitos a serem indexados, a partir de considerações sobre o grau de especialidade de um texto e a densidade conceitual de um termo. Tais procedimentos auxiliariam a determinar o grau de exaustividade e de seletividade com que se deve tratar tematicamente um texto.

Considera-se interessante, ainda, o desenvolvimento de estudos de usuários para identificar o seu grau de satisfação e o modo como as novas tecnologias de informação podem auxiliar no desenvolvimento de sistemas mais eficazes.

Para que as perspectivas futuras possam se tornar realidade é imprescindível promover a capacitação do bibliotecário indexador, acrescentando ao conhecimento adquirido na área biblioteconômica, uma fundamentação em Lingüística e Terminologia, sem esquecer os princípios básicos da Lingüística de *Corpus*. Com uma formação mais abrangente o bibliotecário terá condições facilitadas de enfrentar o grande desafio que envolve a comunicação entre sistemas de informação e usuários.

REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. *NBR 12676:1992 (NB 1419) – Método para análise de documentos: determinação de seus assuntos e seleção de termos de indexação*. 4 p. Disponível em: <<http://www.unilavras.edu.br/biblioteca/normas.php>>. Acesso em: 19/10/2005.
- AGUADO, Pedro Manuel. Los sistemas expertos y la recuperación documental: ejemplos de aplicación. *Scire: Representación y Organización del Conocimiento*, Zaragoza, v. 1, n. 2, p. 21-32, jul./dic. 1995.
- ARAÚJO, Vania Maria Rodrigues Hermes de. Sistemas de informação: nova abordagem teórico conceitual. *Ciência da Informação*, Brasília, v. 24, n. 1, p. 54-76, jan./abr. 1995.
- AUSTIN, J.L. *Quando dizer é fazer: palavras e ação*. Porto Alegre: Artes Médicas, 1990. 136 p.
- BAKHTIN, Mikhail. *Marxismo e filosofia da linguagem*. 3. ed. rev. e atual. São Paulo: HUCITEC, 1995. 196 p.
- BARROS, Lidia Almeida. *Curso básico de Terminologia*. São Paulo: EDUSC, 2004. 285 p.
- BENVENISTE, Émile. *Problemas de lingüística geral*. 2. ed. Campinas: Pontes, 1988-89. 2 v.
- BEVILACQUA, Cleci Regina. Unidades fraseológicas especializadas: elementos para su identificación y descripción. In: CABRÉ, Maria Teresa; FELIU, Judit (Org.). *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Universitat Pompeu Fabra, 2001. p. 113- 141.
- BOULANGER, Jean-Claude. Convergências e divergências entre a lexicografia e a terminografia. In: LIMA, Marília dos Santos; RAMOS, Patrícia Chittoni (Org.). *Terminologia e ensino de segunda língua: Canadá e Brasil*. Porto Alegre: Núcleo de Estudos Canadenses, [200-?]. p. 7-28.
- BRÄSCHER, Marisa. *Curso de elaboração de tesouros*. Brasília, 1999. 23 p. (Polígrafo)
- BRASIL. Lei n. 4.084, de 30 de junho de 1962. Dispõe sobre a profissão de Bibliotecário e regula seu exercício. CONSELHO FEDERAL DE BIBLIOTECONOMIA. Disponível em: <<http://www.cfb.org.br/legislacao/leis/LEI%204084-62.asp>>. Acesso em: 24/12/2005.
- BRÉAL, Michel. *Ensaio de semântica*. São Paulo: Pontes, 1992. 233 p.
- CABRÉ, Maria Teresa. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antartida/Empuries, 1993. 529 p.
- CABRÉ, Maria Teresa. Sumario de principios que configuran la nueva propuesta teórica. In: CABRÉ, Maria Teresa; FELIU, Judit. *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Universitat Pompeu Fabra, 2001. p. 19-25.

- CABRÉ, Maria Teresa et.al. La Terminología hoy: replanteamiento o diversificación. *Organon*, Porto Alegre, v. 12, n. 26, p. 33-41, 1998.
- CINTRA, Anna Maria Marques et al. *Para entender as linguagens documentárias*. 2. ed. rev. e ampl. São Paulo: Polis, 2002. 92 p.
- COLÓQUIO NACIONAL LETRAS EM DIÁLOGO E EM CONTEXTO: rumos e desafios, 2002, Porto Alegre. *Anais....* Porto Alegre: Programa de Pós-Graduação em Letras, 2003. 1 CD-ROM.
- CONGRESSO NACIONAL DA ABRALIN, 2. XIV INSTITUTO LINGÜÍSTICO, 1999, Florianópolis. *Anais...* Florianópolis: UFSC, 2000. 1 CD-ROM.
- CRYSTAL, David. *Dicionário de lingüística e fonética*. Rio de Janeiro: Jorge Zahar, 2000. 275 p.
- CURRÁS, Emília. *Tesauros: linguagens terminológicas*. Brasília: IBICT, 1995. 286 p.
- DUBOIS, Jean et al. *Dicionário de Lingüística*. São Paulo: Cultrix, 1997. 653 p.
- ENCONTRO NACIONAL DA ANPOLL, 17., 2002, Gramado. *Anais...* Gramado: Quarteto, 2004. 1 CD-ROM.
- ETTINGER, Stefan. Formación de palabras y fraseología en la Lexicografía. In: HAENSCH, Günther et al. *La Lexicografía: de la lingüística teórica a la lexicografía práctica*. Madrid: Gredos, 1982. p. 233-258.
- FAULSTICH, Enilde. Socioterminologia: mais que um método de pesquisa, uma disciplina. *Ciência da Informação*, Brasília, v. 24, n. 3, p. 281-288, set./dez. 1995.
- FÁVERO, Leonor Lopes; KOCH, Ingedore G. Villaça. *Lingüística Textual: introdução*. 7. ed. São Paulo: Cortez, 2005. 120 p.
- FERREIRA, Sueli Mara Soares. Novos paradigmas da informação e novas percepções do usuário. *Ciência da Informação*, Brasília, v. 25, n. 2, p. 217-223, maio/ago.1996.
- FILLMORE, Charles J. *Corpus linguistics or Computer-aided armchair linguistics*. In: SVARTVIK, Jan. (Ed.). *Directions in Corpus Linguistics: Proceedings of Nobel symposium 82, Stockholm, 4-8 August 1991*. Berlin: Mouton de Gruyter, 1992. p. 35-59.
- GAUDIN, François. *Pour une socioterminologie: Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Université de Rouen, 1993. 231 p.
- GAUDIN, François. Socioterminologie: du signe au sens, construction d'un champ. *Meta*, Montréal, v. 38, n. 2, 1993a. Disponível em: <www.erudit.org/revue/meta/1993/v38/n2/>. Acesso em 11/03/2005.
- GIL, Antônio Carlos. *Métodos e técnicas de pesquisa social*. 5. ed. São Paulo: Atlas, 1999. 206 p.
- GINEZ DE LARA, Marilda Lopes. O Unicórnio (o Rinoceronte, o Ornitorrinco ...), a Análise documentária e a linguagem documentária. *DataGramaZero: Revista de Ciência da*

Informação, Rio de Janeiro, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgz.org.br/dez01/Art_03.htm>. Acesso em: 06/08/2005.

GOMES, Hagar Espanha (Coord.). *Manual de elaboração de tesouros monolíngües*. Brasília, Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990. 78 p.

GOMES, Hagar Espanha; MARINHO, Marcílio Teixeira. *Introdução ao estudo do cabeçalho de assunto*. 1984. Disponível em: <www.conexaorio.com/bit/cabeçalho/cab_ass.htm>. Acesso em: 08/12/2004.

GOMES, Henriette Ferreira. O ambiente informacional e suas tecnologias na construção dos sentidos e significados. *Ciência da Informação*, Brasília, v. 29, n.1, p. 61-70, jan./abr. 2000.

GOUADEC, Daniel. *Terminologie: constitution des données*. Paris, AFNOR, 1990. 219 p.

GUINCHAT, Claire; MENO, Michel. *Introducción general a las ciencias y técnicas de la información y de la documentación*. Paris: UNESCO, 1983. 477 p.

HALLIDAY, M.A.K. Estrutura e função da linguagem. In: LYONS, John (Org.). *Novos horizontes em Lingüística*. São Paulo: Cultrix, 1976. p. 134-160.

HOLZEM, Maryvonne. La thèse et son auteur: voie d'une nouvelle complémentarité terminologie-documentation. *Terminologies nouvelles*, Bruxelles, n.18, juin, 1988, p. 28-35.

HOLZEM, Maryvonne; WABLE, Thierry. La description terminologique pour rendre compte de l'avancée des connaissances scientifiques en langue française. *Dans Recherche en rédaction professionnelle*, Sherbrooke, CA, v. 1, n.1, hiver 2001.

JAKOBSON, Roman. *Lingüística e comunicação*. São Paulo: Cultrix, 2003. 162 p.

KRIEGER, Maria da Graça. Terminologia em contextos de integração: funcionalidade e fundamentos. *Organon*, Porto Alegre, v. 12, n. 26, p. 19-31, 1998.

KRIEGER, Maria da Graça. O termo: questionamentos e configurações. In: KRIEGER, Maria da Graça; MACIEL, Anna Maria Becker (Org.). *Temas de Terminologia*. Porto Alegre: Ed. da UFRGS, 2001. p. 62-81.

KRIEGER, Maria da Graça; BEVILACQUA, Cleci Regina. *A pesquisa terminológica no Brasil: uma contribuição para a consolidação da área*. 2005. Disponível em: <http://www.riterm.net/revista/n_1/krieger.pdf>. Acesso em: 08/01/2006.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. *Introdução à Terminologia: teoria & prática*. São Paulo: Contexto, 2004. 223 p.

KURAMOTO, Hélio. Uma abordagem alternativa para o tratamento e a recuperação de informação textual: os sintagmas nominais. *Ciência da Informação*, Brasília, v. 25, n.2, p. 182-192, maio/ago. 1996.

LAAN, Regina Van der. *Tesouro e terminologia: uma relação lógica*. Porto Alegre, 2002. 173 p. Tese (Doutorado em Lingüística) – Instituto de Letras, Universidade Federal do Rio Grande do Sul.

- LANCASTER, F.W. *Indexação e resumos: teoria e prática*. 2. ed. rev. atual. Brasília: Briquet de Lemos, 2004. 452 p.
- LEFFA, Wilson (Comp.). *Textos em Lingüística Aplicada*. Pelotas : Universidade Católica de Pelotas, 2000. 1 CD-ROM.
- LÉRAT, Pierre. Lengua especializada y Documentación. In: LÉRAT, Pierre. *Las lenguas especializadas*. Barcelona: Ariel, 1997. p. 117-123.
- LIMA, Vânia Mara Alves. *Terminologia, comunicação e representação documentária*. 1998. 117 f. Dissertação (Mestrado em Ciências da Informação) – Escola de Comunicação e Artes. Disponível em: http://www.teses.usp.br/teses/disponiveis/27/27143/tde-11052004-122839/publico/Term_Comum_Repres_Documentaria.pdf. Acesso em: 25/01/2006.
- LOPES, Edward. *A identidade e a diferença: raízes históricas das teorias estruturais da narrativa*. São Paulo: EDUSP, 1997. 380 p.
- LUCAS, Clarinda Rodrigues. Biblioteconomia: produção e administração da interpretação. *Ciência da Informação*, Brasília, v. 26, n. 1, p. 46-53, jan./abr., 1997.
- LUCAS, Clarinda Rodrigues. A organização do conhecimento e tecnologias da informação: oráculos humanos e inteligência artificial. In: IBICT: Instituto Brasileiro de Informação (Org.). *Organização do conhecimento e sistemas de classificação*. Brasília, 1996, p. 69-72.
- LYONS, John. *Linguagem e lingüística: uma introdução*. Rio de Janeiro: LTC, 1987. 322 p.
- MACIEL, Anna Maria Becker. Lingüística de *Corpus* : bases teórico-metodológicas. In: COLÓQUIO NACIONAL LETRAS EM DIÁLOGO E EM CONTEXTO: rumos e desafios, 2002, Porto Alegre. *Anais....* Porto Alegre: Programa de Pós-Graduação em Letras, 2003. 1 CD-ROM.
- MACIEL, Anna Maria Becker. *Terminologia e Lingüística de Corpus: reconhecimento de especificidades de termos*. 2005. Trabalho apresentado no V ENCONTRO DE CORPORA, São Carlos, 2005.
- MARQUES, Maria Helena Duarte. *Iniciação à semântica*. 5. ed. Rio de Janeiro: Zahar, 2001. 165 p.
- MOREL, Jordi; RODRÍGUEZ, Carlos. Consecuencias metodológicas de la propuesta teórica (II). In: CABRÉ, Maria Teresa; FELIU, Judit. *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Universitat Pompeu Fabra, 2001. p. 37-53.
- PAVEL, Silvia; NOLET, Diane. *Manual de Terminología*. Quebec: Ministro de Obras Públicas y Servicios Gubernamentales de Canadá, 2002. 156 p.
- PEARSON, Jennifer. *Terms in Context*. Amsterdam: J. Benjamins, 1998. 242 p.
- PÊCHEUX, Michel; FUCHS, Catherine. A propósito da Análise Automática do Discurso: atualização e perspectivas. In: GADET, Françoise; HAK, Tony. (Org.). *Por uma análise automática do discurso: uma introdução à obra de Michel Pêcheux*. Trad. de Péricles Cunha. 2. ed. Campinas: UNICAMP, 1997. p. 163-252.

- PETTER, Margarida. Linguagem, língua, lingüística. In: FIORIN, José Luiz (Org.) *Introdução à lingüística*. São Paulo: Contexto, 2004. p. 11-24.
- RABAÇA, Carlos Alberto; BARBOSA, Gustavo Guimarães. *Dicionário de comunicação*. São Paulo: Ática, 1987. 637 p.
- REY, Allain. *La terminologie: noms et notions*. Paris: PUF, 1992. 127 p.
- RONDEAU, Guy. *Introduction à la Terminologie*. 2. ème. ed. Quebec: Gaëtan Morin, 1984. 238 p.
- SAGER, Juan C. *Curso Práctico sobre el Procesamiento de la Terminología*. Madrid: Fundación Germán Sánchez Ruipérez; Pirâmide, 1993. 442 p.
- SAGER, Juan C. *A practical course in terminology processing*. Amsterdam: J. Benjamins, 1990. 254 p.
- SANTOS, Gildeir Carolino; RIBEIRO, Célia Maria. *Acrônimos, siglas e termos técnicos: Arquivística, Biblioteconomia, Documentação, Informática*. Campinas: Átomo, 2003. 277 p.
- SARDINHA, Tony Berber. *Lingüística de corpus*. São Paulo: Barueri, 2004. 410 p.
- SAUSSURE, Ferdinand de. *Curso de lingüística geral*. São Paulo: Cultrix, 1988. 279 p.
- SLODZIAN, Monique. L'emergence d'une terminologie textuelle et le retour du sens. In: B'JOINT, Henri; THOIRON, Philippe (Org.) *Le sens en terminologie*. Lyon: Presses Universitaires de Lyon, 2000. p. 60-80.
- TEBÉ, Carles. Bancos de dados terminológicos. *Terminometro: La Terminologia en España*, p. 65-68, 1996. Numero Especial.
- TEBÉ, Carles. Elementos para la representación de los conceptos especializados en bases y bancos de datos. In: CABRÉ, Maria Teresa; FELIU, Judit. *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Universitat Pompeu Fabra, 2001. p. 203-208.
- TEIXEIRA, Cláudia Hlebetz. Onde os intérpretes da informação? *Informare: Cadernos do Programa de Pós-Graduação em Ciência da Informação*, Rio de Janeiro, v. 1, n. 2, p. 37-44, jun./dez., 1995.
- TRASK, R.L. *Dicionário de linguagem e lingüística*. São Paulo: Contexto, 2004. 364 p.
- VELHO, Léa. *Notas sobre a Pós-Graduação em Ciências Sociais e Humanidades: por que e em que elas diferem das Ciências Naturais?* 1997. Disponível em: <<http://rima.pucsp.br/documentos/texto06.htm>>. Acesso em: 29/08/2005.
- WABLE, Thierry; HOLZEM, Maryvonne. Structure textuelle et communication scientifique. *Communication, Actes de GLAT –p Barcelona, 2004: la production des textes spécialisés : structure et enseignement*. 12-14 mai. 2004. Universitat Politècnica de Catalunya: 137-148. 11 mai. 2004. Disponível em: http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/13/33/sic_00001333_02/sic_00001333.html. Acesso em: 27/07/2005.

WÜSTER, Eugen. *Introducción a la Teoría General de la Terminología y a la Lexicografía Terminológica*. Barcelona; IULA; Universitat Pompeu Fabra, 1998.

APÊNDICE A

Endereços eletrônicos das bases de dados consultadas

Banco de Dados Bibliográficos da USP

<http://dedalus.usp.br:4500/ALEPH/POR/USP/USP/DEDALUS/START>

Bibliothèque et Archives Canada (AMICUS)

<http://www.collectionscanada.ca/amicus/index-f.html>

Catálogo Geral da PUCRS (VERUM)

<http://verum.pucrs.br/ALEPH>

Library of Congress

<http://catalog.loc.gov/>

Sistema de bibliotecas da UFRGS (SABi)

<http://sabix.ufrgs.br/ALEPH/>

Sistema de Biblioteca da UNICAMP (ACERVUS)

<http://www.unicamp.br/bc/HPSB104.HTM>

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE abaixamento datílico (K)
UT ABAIXAMENTO DATÍLICO **UTI**
UFE **UFEI**
AR **MOD**
TR ABAIXAMENTO ESPONDEU. ACENTO. FONÉTICA AUDITIVA. FONOLOGIA MÉTRICA.
FONOLOGIA PROSÓDICA. MORFOLOGIA. PROSÓDIA. SÍLABA. SONORIDADE.
TE abaixamento dactílico
RI

USE abaixamento espondeu (K)
UT ABAIXAMENTO ESPONDEU **UTI**
UFE **UFEI**
AR **MOD**
TR ABAIXAMENTO DATÍLICO. ACENTO. FONÉTICA AUDITIVA. FONOLOGIA MÉTRICA.
FONOLOGIA PROSÓDICA. MORFOLOGIA. PROSÓDIA. SÍLABA. SONORIDADE.
TE
RI

USE abordagem interacional (K)
UT **UTI**
UFE **UFEI**
AR **MOD** ABORDAGEM INTERACIONAL
TR APRENDIZAGEM. ENSINO. LINGUAGEM. LINGÜÍSTICA APLICADA.
TE
RI

USE academic text (K)
UT TEXTO ACADÊMICO **UTI** academic text
UFE **UFEI**
AR **MOD**
TR GÊNERO TEXTUAL. LINGÜÍSTICA DO TEXTO.
TE
RI

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE	ação lingüística (K)		
UT	AÇÃO LINGÜÍSTICA	UTI	linguistic action
UFE		UFEI	
AR	J.L. AUSTIN.	MOD	
TR	ATO ILOCUCIONÁRIO. ATO LOCUCIONÁRIO. ATO PERLOCUCIONÁRIO. ATOS DE FALA. DÉITICOS. FILOSOFIA DA LINGUAGEM. NARRATIVA.		
TE			
RI			

USE	ação verbal		
UT	AÇÃO VERBAL	UTI	verbal action
UFE		UFEI	
AR		MOD	
TR	AGENTE DA PASSIVA. GRAMÁTICA. MORFOLOGIA. SUJEITO. VERBO. VOZ PASSIVA. VOZ REFLEXIVA.		
TE			
RI			

USE	accent (K)		
UT	ACENTO	UTI	stress; accent
UFE		UFEI	
AR		MOD	
TR	ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE			
RI			

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE	accentuation		
UT	ACENTUAÇÃO	UTI	accentuation
UFE		UFEI	
AR		MOD	
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE			
RI			

USE	acento (K)		
UT	ACENTO	UTI	stress; accent
UFE		UFEI	
AR		MOD	
TR	ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE			
RI			

USE	acento de altura		
UT	ACENTO TONAL	UTI	pitch accent; tonal accent
UFE		UFEI	
AR		MOD	
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE	acento de altura		
RI			

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE acento de intensidade

UT ACENTO DE INTENSIDADE **UTI** stress accent

UFE **UFEI**

AR **MOD**

TR ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.

TE

RI

USE acento final

UT ACENTO FINAL **UTI** final stress; final accent

UFE **UFEI**

AR **MOD**

TR ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.

TE

RI

USE acento inicial

UT ACENTO INICIAL **UTI** initial stress; initial accent

UFE **UFEI**

AR **MOD**

TR ACENTO. ACENTO CONTRASTIVO. ACENTO DE ALTURA. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.

TE

RI

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE	acento primário		
UT	ACENTO PRINCIPAL	UTI	primary stress; primary accent
UFE		UFEI	
AR		MOD	
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE ALTURA. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE			
RI			

USE	acento principal		
UT	ACENTO PRINCIPAL	UTI	principal stress; principal accent
UFE		UFEI	
AR		MOD	
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE ALTURA. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE	acento primário		
RI			

USE	acento secundário		
UT	ACENTO SECUNDÁRIO	UTI	secondary stress; secondary accent
UFE		UFEI	
AR		MOD	
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE ALTURA. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.		
TE			
RI			

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE	acento tonal	
UT	ACENTO TONAL	UTI pitch accent; tonal accent
UFE		UFEI
AR		MOD
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TÔNICO. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.	
TE	acento de altura	
RI		

USE	acento tônico	
UT	ACENTO TÔNICO	UTI tonic accent
UFE		UFEI
AR		MOD
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. ACENTUAÇÃO. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.	
TE		
RI		

USE	acentuação	
UT	ACENTUAÇÃO	UTI accentuation
UFE		UFEI
AR		MOD
TR	ACENTO. ACENTO CONTRASTIVO. ACENTO DE INTENSIDADE. ACENTO DIFERENCIAL. ACENTO FINAL. ACENTO FONÉTICO. ACENTO GRÁFICO. ACENTO INICIAL. ACENTO PRINCIPAL. ACENTO SECUNDÁRIO. ACENTO TONAL. DITONGO. DITONGO ABERTO. DITONGO FECHADO. FONEMA. FONÉTICA. FONOLOGIA. FONOLOGIA PROSÓDICA. GRAMÁTICA. HIATO. MÉTRICA. MONOSSÍLABO. MORFOFONOLOGIA. MORFOLOGIA. ORTOGRAFIA. OXÍTONA. PAROXÍTONA. PROPÁROXÍTONA. PROSÓDIA. SÍLABA RITMO. SISTEMA DE ACENTUAÇÃO. TÔNICA. TRAÇO DISTINTIVO. TREMA.	
TE		
RI		

APÊNDICE B - BASE DE DADOS TERMINOLÓGICA :

LINGÜÍSTICA TEÓRICA E APLICADA

USE acoustic phonetic (K)
UT FONÉTICA ACÚSTICA **UTI** acoustic phonetic
UFE **UFEI**
AR **MOD**
TR ENTONAÇÃO. FONÉTICA. FONÉTICA ARTICULATÓRIA. FONÉTICA AUDITIVA. SOM.

TE
RI

USE acquisition (K)
UT AQUISIÇÃO **UTI** acquisition
UFE **UFEI**
AR **MOD**
TR FONOLOGIA. LINGUAGEM. PSICOLINGÜÍSTICA.

TE
RI

USE acquisition of liquids
UT **UTI**
UFE AQUISIÇÃO DAS LÍQUIDAS **UFEI** acquisition of liquids
AR **MOD**
TR AQUISIÇÃO DA LINGUAGEM. FONOLOGIA. PSICOLINGÜÍSTICA. SISTEMA FONOLÓGICO.

TE
RI

USE acquisition of liquids in coda position
UT **UTI**
UFE AQUISIÇÃO DAS LÍQUIDAS EM POSIÇÃO CODA **UFEI** acquisition of liquids in coda position
AR **MOD**
TR AQUISIÇÃO DA LINGUAGEM. FONOLOGIA. PSICOLINGÜÍSTICA. SISTEMA FONOLÓGICO.

TE
RI

ANEXO A

Lista parcial de palavras do corpus (ordem alfabética)

WordSmith Tools -- 22/2/2006 23:30:59

N	Word	Freq.	%	Lemmas
1	A	11.883	2,02	
2	À	473	0,08	
3	Á	9		
4	Â	2		
5	A;	2		
6	AACHEN		2	
7	AAE	4		
8	AAL	4		
9	AARON		2	
10	AAT	3		
11	AAVE	8		
12	AB	7		
13	ABANDONO		2	
14	ABARCA		3	
15	ABBOTT		2	
16	ABBREVIATED2			
17	ABBREVIATIONS			2
18	ABC	2		
19	ABDELHAK		4	
20	ABERRANT		2	
21	ABERTA		3	
22	ABERTO		4	
23	ABERTURA		4	
24	ABILITIES		83	0,01
25	ABILITY		190	0,03
26	ABLAUT		2	
27	ABLE	64	0,01	
28	ABNORMAL		10	
29	ABNORMALITIES			10
30	ABOLISHED		3	
31	ABONADA		2	
32	ABORDA		17	
33	ABORDADA		3	
34	ABORDADAS		2	
35	ABORDADO		5	
36	ABORDADOS		4	
37	ABORDAGEM		99	0,02
38	ABORDAGENS		34	
39	ABORDAM		8	
40	ABORDAMOS		3	
41	ABORDANDO		3	
42	ABORDAR		6	
43	ABORDARAM		2	
44	ABORDARMOS2			
45	ABORDOU		4	
46	ABORIGINAL		8	
47	ABOUT418		0,07	
48	ABOUTNESS		2	
49	ABOVE48			
50	ABR		6	

51	ABRAÇADO	4	
52	ABRAHAM	2	
53	ABRAHÃO	4	
54	ABRANGE	8	
55	ABRANGÊNCIA		6
56	ABRANGENDO2		
57	ABRANGENTE	5	
58	ABRE	9	
59	ABREU	10	
60	ABRIL	31	
61	ABRINDO	2	
62	ABROAD	3	
63	ABRS	2	
64	ABRUPTA	2	
65	ABSENCE	65	0,01
66	ABSENT	22	
67	ABSOLUTA	2	
68	ABSOLUTE	22	
69	ABSOLUTEIVE	2	
70	ABSOLUTO	2	
71	ABSOLUTOS	2	
72	ABSORÇÃO	3	
73	ABSTRAÇÃO	2	
74	ABSTRACT	1.976	0,34
75	ABSTRACTION4		
76	ABSTRACTIONS		4
77	ABSTRACTLY	2	
78	ABSTRACTS	15	
79	ABSTRACT{	199	0,03
80	ABSTRACT{{A	5	
81	ABSTRACT{{IN		21
82	ABSTRACT{{IT		2
83	ABSTRACT{{THE		30
84	ABSTRACT{{THIS		30
85	ABSTRACT{{TO		2
86	ABSTRACT{{TWO		4
87	ABSTRACT{{USING		2
88	ABSTRACT{{WE		3
89	ABSTRATO	2	
90	ABSTRATOS	2	
91	ABSURDE	3	
92	ABU	2	
93	ABUÊNDIA	3	
94	ABUNDANT	2	
95	ABUNDANTE	2	
96	ABVDS	2	
97	AC	3	
98	ACABA2		
99	ACADEMIC	252	0,04
100	ACADÊMICA	10	
101	ACADÊMICAS	7	
102	ACADÊMICO	13	
103	ACADÊMICOS	12	
104	ACADEMICS	7	
105	ACADEMY	4	
106	ACADIAN	2	

ANEXO B

Lista parcial de palavras do corpus (ordem de frequência)

N	Word	Freq.	%	Lemmas
1	THE	21.143		3,59
2	OF	17.779		3,02
3	AND	14.190		2,41
4	A	11.883		2,02
5	IN	11.242		1,91
6	DE	8.940	1,52	
7	TO	7.665	1,30	
8	E	5.538	0,94	
9	THAT	4.699	0,80	
10	LANGUAGE	4.512	0,77	
11	DO	4.105	0,70	
12	AS	3.732	0,63	
13	DA	3.436	0,58	
14	IS	3.344	0,57	
15	FOR	3.151	0,53	
16	WITH	3.103	0,53	
17	S	2.806	0,48	
18	O	2.755	0,47	
19	ON	2.725	0,46	
20	THIS	2.609	0,44	
21	QUE	2.360	0,40	
22	ARE	2.188	0,37	
23	ABSTRACT	1.976	0,34	
24	BY	1.976	0,34	
25	EM	1.764	0,30	
26	NO	1.718	0,29	
27	WERE	1.689	0,29	
28	AN	1.572	0,27	
29	VOLUME	1.515	0,26	
30	KEYWORDS	1.508	0,26	
31	WAS	1.479	0,25	
32	FROM	1.474	0,25	
33	WORDS	1.467	0,25	
34	SE	1.402	0,24	
35	PAGES	1.382	0,23	
36	WHICH	1.377	0,23	
37	BE	1.335	0,23	
38	OR	1.334	0,23	
39	ENGLISH	1.287	0,22	
40	ISSUE	1.272	0,22	
41	WE	1.218	0,21	
42	THEIR	1.212	0,21	
43	STUDY	1.210	0,21	
44	SPEECH	1.206	0,20	
45	IT	1.161	0,20	
46	OS	1.158	0,20	
47	WORD	1.155	0,20	
48	NOT	1.135	0,19	
49	PARA	1.135	0,19	
50	UMA	1.126	0,19	
51	THESE	1.121	0,19	
52	BETWEEN	1.112	0,19	

53	COMO	1.104	0,19	
54	NA	1.100	0,19	
55	LINGUISTIC	1.068	0,18	
56	TWO	1.034	0,18	
57	UM	1.009	0,17	
58	COM	997	0,17	
59	CHILDREN	985	0,17	
60	AUTHOR	983	0,17	
61	LEXICAL	967	0,16	
62	PROCESSING	940	0,16	
63	BRAIN	932	0,16	
64	AT	923	0,16	
65	DOS	901	0,15	
66	RESULTS	891	0,15	
67	BOTH	856	0,15	
68	WRITING	847	0,14	
69	READING	835	0,14	
70	ANALYSIS	824	0,14	
71	MORE	812	0,14	
72	DISCOURSE	809	0,14	
73	THAN	802	0,14	
74	DISSERTAÇÃO	775	0,13	
75	SEMANTIC	763	0,13	
76	JOURNAL	756	0,13	
77	LÍNGUA	743	0,13	
78	DAS	740	0,13	
79	HAVE	725	0,12	
80	AUTOR	723	0,12	
81	I	720	0,12	
82	BUT	719	0,12	
83	POR	716	0,12	
84	RESEARCH	696	0,12	
85	TESE DISSERTAÇÃO	690	0,12	
86	THEY	681	0,12	
87	HAS	666	0,11	
88	PORTUGUÊS	666	0,11	
89	É	664	0,11	
90	USE	663	0,11	
91	LEARNING	662	0,11	
92	LINGÜÍSTICA	642	0,11	
93	OU	636	0,11	
94	PESQUISA	636	0,11	
95	DIFFERENT	634	0,11	
96	SECOND	631	0,11	
97	ANÁLISE	623	0,11	
98	PALAVRAS	616	0,10	
99	ONE	613	0,10	
100	ALSO	608	0,10	
101	MAY	602	0,10	
102	KEY	596	0,10	
103	TASK	595	0,10	
104	CAN	589	0,10	
105	PAPER	586	0,10	
106	STUDIES	572	0,10	
107	PHONOLOGICAL	562	0,10	

ANEXO C

Concordanciador : search word or phrase : linguistic (lista parcial)

N

Concordance

1 to: (1) characterise were to: (1) characterise **cognitive–linguistic** processes within hypok
 2 arning and generalization within and across **cognitive–linguistic** domains in a 62-year-
 3 stica. ABSTRACT: In this article we present **linguistic** evidence showing that Modern Br
 4 ation Theory, this study aims to describe the **linguistic** use of the Verbal Agreement with
 5 Paraíba - VALPB) which provide us with the **linguistic** and social factors that are relevan
 6 . ABSTRACT: The structuralistic view of the **linguistic** phenomenon which arises when t
 7 istic rates by dividing the frequency of some **linguistic** variable of interest by a frequency
 8 approach. The students frequently possess **linguistic** competence, but they are not abl
 9 nalyzed 78 occurrences from the data of the **Linguistic** Variation Project of Paraíba (Proj
 10 ACT: This study aims to investigate how the **linguistic** proximity between Portuguese an
 11 of individual differences in the use of cross-**linguistic** structures. Implications of the curr
 12 ot in both children. It is suggested that cross-**linguistic** cue competition best captures the
 13 best captures the transfer process, and that **linguistic** environment for each child may b
 14 speakers. Key words: Brazilian Portuguese; **linguistic** drift; linguistic change. CONTRI
 15 retical and methodological postulates of the **Linguistic** Variation Theory, this study aims
 16 ur, could result in a polarized concept of the **linguistic** process. Meanwhile in the interac
 17 d. Second, the common practice of deriving **linguistic** rates by dividing the frequency of
 18 enon is seen as a result of a non-polarized **linguistic** con-flict, the sociological impositi
 19 words: Brazilian Portuguese; linguistic drift; **linguistic** change. CONTRIBUTO PARA U
 20 uestionings involved from this study of those **linguistic** phenomena. KEYWORDS: Synta
 21 resents two contents, adjective degrees and **linguistic** varieties, observing that the refer
 22 g a research based on the relation between **linguistic** awareness and performance in a
 23 o improve their performance. KEYWORDS: **linguistic** awareness; English as a foreign l
 24 Traditional Grammar and the Project NURC **Linguistic** Studies, we started developing a
 25 Florianópolis, S.C., in order to reflect on the **linguistic** and textual competence of those
 26 . It has been used William Labov's theory of **Linguistic** Variation on it. KEYWORDS: S
 27 rth-east of Brazil. KEYWORDS: pragmatic; **linguistic** expressions; literary speech. A
 28 e-legitimation, sensibility, identity, desire and **linguistic** turn are terms frequently used to
 29 ABSTRACT: The article relates the results of **linguistic** investigation on academic texts,
 30 the link between the work of some fucntional **linguistic** policies towards the portuguese l
 31 system of Portuguese. Besides, it describes **linguistic** contexts that promote the emerge
 32 ACT: Following the functionalist approach to **linguistic** phenomena, this article analyses t
 33 identification of the social meaning of some **linguistic** and conversational strategies whic
 34 tten Spanish comprehension. KEYWORDS: **linguistic** proximity, interference, input A l
 35 nouns "me", "te", "nos" for the pronoun "se". **Linguistic** aspect happens with object pron
 36 ury Court. The observation of the discursive-**linguistic** procedure allows to show that the
 37 est, Gün R. Semin, and Steven J. Sherman. **Linguistic** Context and Social Perception:
 38 Abstract This article considers the role of **linguistic** context in social memory and soci
 39 ction (action verbs/behaviors). Based on the **linguistic** category model, togetherwith prior
 40 son descriptions formulated at ahigh level of **linguistic** Abstraction (adjectives/traits) vers
 41 lyzes, from a interactionist point of view, the **linguistic** and discursive functioning in nar

ANEXO D

Concordanciador : search word or phrase : keyword (lista parcial)

N	Concordance
1	vocabulary-independent indexing Author Keywords Expected count expected ter
2	search virtual hypothesis copies Author Keywords Fast decoding speech recog
3	ng, the uncritical adoption of this and other keywords can have serious consequence
4	ns of Difference, Frequency, and Sense in Keywords Analysis. Journal of English Li
5	n Deming's PDCA management method. KEYWORDS : second language acquisitio
6	sed on Historic Linguistics and Philology. Keywords – analytical verbal forms, Portu
7	ines some potential problems of analyzing keywords and ways to overcome them. A
8	keywords and ways to overcome them. A keywords analysis may force the researc
9	similar groups is also helpful. Key Words: keywords • corpus • methodology • frequ
10	approach, his beliefs and social practices. KEYWORDS : reading, writing, social inter
11	ce but also teacher's classroom activities. KEYWORDS : letters ; writing ; communic
12	n Deming's PDCA management method. KEYWORDS : second language acquisitio
13	the principles recommended by linguists. KEYWORDS : Alphabetisation; perception;
14	oriented by its own strategies and rules. KEYWORDS : reading; interaction; author;
15	ce of a "homoerotic hegemony" instead. KEYWORDS : soft gay; plastic sexuality; h
16	acknowledge the others as interlocutors. KEYWORDS : Language, theater games,
17	nts in the profess of formation as a whole. KEYWORDS : ideology; imaginary academi
18	the esthetic sense of the reader-student? KEYWORDS : reading; literary reading; lite
19	e empowering roles on learners abilities. . KEYWORDS : English learning; Interaction;
20	come efficient in their field of knowledge. KEYWORDS : teaching-learning; reading;
21	my among what is classified as polysemy. KEYWORDS : semantics; homonymy; poly
22	logical analyses of the Textual Linguistics. KEYWORDS : narrative; semantics; interac
23	haracteristic of a specific cultural group. KEYWORDS : culture, target culture, targe
24	resence by its users (Cf. Castilho, 1995). KEYWORDS : efficient producer; oral com
25	in different socio-communicative contexts. KEYWORDS : textual genres; language us
26	historical-cultural aspect of that language. KEYWORDS : utterance; interculturality; K
27	oral and written Spanish comprehension. KEYWORDS : linguistic proximity, interfere
28	ractices underlying didactic interactions. KEYWORDS : Applied Linguistics, teaching
29	rt a nontraditional institution in their world. KEYWORDS : Amondawa society, educatio
30	tense should be adopted or developed. KEYWORDS : English teaching; present p
31	in English as Foreign Language classes. KEYWORDS : senior citizen; social intera
32	to make a reflection about this problem. KEYWORDS : Linguistics; Applied Linguisti
33	s, the reading skill becomes fundamental. KEYWORDS : Applied Linguistics; ESP.
34	ance of interaction to language learning. KEYWORDS : reflective teaching; teacher
35	prove language acquisition and teaching. KEYWORDS : Interlanguage; writing; "error
36	ents exert above any text producer agent. KEYWORDS : discourse analysis; verbal i
37	earch to the foreign language classroom. KEYWORDS : action-research, learning st
38	rs and its consequence on their teaching. KEYWORDS : text; coherence; informativity
39	ive and communicative needs of its users. KEYWORDS : Functional linguistics; teachi
40	at might be taking place in the classroom. KEYWORDS : testing; communicative appr
41	e notions of the "text" and "text cohesion". KEYWORDS : divulgation discourse; teach

ANEXO E

Search word or phrase : linguistic - forma expandida do registro 1

1 cognitive processes: Evidence for hypokinetic–hyperkinetic linguistic homologues? *Journal of Neurolinguistics*, Volume 18, Issue 5, September 2005, Pages 361-381

Abstract In relation to motor control, the basal ganglia have been implicated in both the scaling and focusing of movement. Hypokinetic and hyperkinetic movement disorders manifest as a consequence of overshooting and undershooting GPi (globus pallidus internus) activity thresholds, respectively. Recently, models of motor control have been borrowed to translate cognitive processes relating to the overshooting and undershooting of GPi activity, including attention and executive function. Linguistic correlates, however, are yet to be extrapolated in sufficient detail. The aims of the present investigation were to: (1) characterise **cognitive–linguistic** processes within hypokinetic and hyperkinetic neural systems, as defined by motor disturbances; (2) investigate the impact of surgically-induced GPi lesions upon language abilities. Two Parkinsonian cases with opposing motor symptoms (akinetic versus dystonic/dyskinetic) served as experimental subjects in this research. Assessments were conducted both prior to as well as 3 and 12 months following bilateral posteroventral pallidotomy (PVP). Reliable changes in performance (i.e. both improvements and decrements) were typically restricted to tasks demanding complex linguistic operations across subjects. Hyperkinetic motor symptoms were associated with an initial overall improvement in complex language function as a consequence of bilateral PVP, which diminished over time, suggesting a decrescendo effect relative to surgical beneficence. In contrast, hypokinetic symptoms were ass

A663 D

ARAUJO, Vera Maria Araujo Pigozzi de
Documentação, terminologia e Lingüística: uma interface
produtiva / Vera Maria Araujo Pigozzi de Araujo. Porto
Alegre, 2006.
f.

Orientadora: Profa. Dra. Anna Maria Becker Maciel.

Dissertação (Mestrado) – Universidade Federal do Rio Grande
do Sul. Instituto de Letras.

1. Bases de dados terminológica. 2. Análise documentária. 3.
Documentação. 4. Sistemas de Recuperação da Informação. 5.
Teoria da Enunciação. 6. Teoria Comunicativa da Terminologia.
7. Socioterminologia. 8. Lingüística de *Corpus*. I. Maciel, Anna
Maria Becker. II Universidade Federal do Rio Grande do Sul.
Instituto de Letras, Porto Alegre, 2006. III Título.