

Universidade Federal do Rio Grande do Sul
Instituto de Letras
Programa de Pós-Graduação em Letras
Área: Estudos da Linguagem
Especialidade: Estudos linguísticos do léxico
Linha de pesquisa: Lexicografia e Terminologia: relações textuais

**SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO
E LINGUAGENS DOCUMENTÁRIAS:**
contribuições dos Estudos da Linguagem

Vera Maria Araujo Pigozzi de Araujo

Tese de Doutorado em Estudos da Linguagem
apresentada ao Programa de Pós-Graduação em
Letras da UFRGS como requisito parcial para a
obtenção do título de Doutora em Letras.

Porto Alegre
2013

Universidade Federal do Rio Grande do Sul
Instituto de Letras
Programa de Pós-Graduação em Letras
Área: Estudos da Linguagem
Especialidade: Estudos linguísticos do léxico
Linha de pesquisa: Lexicografia e Terminologia: relações textuais

**SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO
E LINGUAGENS DOCUMENTÁRIAS:**
contribuições dos Estudos da Linguagem

Vera Maria Araujo Pigozzi de Araujo

Tese de Doutorado em Estudos da Linguagem
apresentada ao Programa de Pós-Graduação em
Letras da UFRGS como requisito para a obtenção do
título de Doutora em Letras.

Orientadora: Profa. Dra. Maria José Finatto
Co-orientador: Prof. Dr. Valdir do Nascimento Flores

Porto Alegre

2013

A663S

Araujo, Vera Maria Araujo Pigozzi de Araujo
Sistemas de Recuperação da Informação e Linguagens
Documentárias : contribuições dos Estudos da Linguagem. /
Vera Maria Araujo Pigozzi de Araujo. Porto Alegre, 2013.
2 v.

Orientadora: Profa. Dra. Maria José Bocorny Finatto.
Co-orientador: Prof. Dr. Valdir do Nascimento Flores.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de PósGraduação em
Letras, Porto Alegre, BR-RS, 2013.

1. Ciência da Informação. 2. Estudos da Linguagem. 3.
Estudos de Terminologia. 4. Linguística da Enunciação. 5.
Vocabulários controlados especializados. I. Finatto, Maria José
Bocorny. II. Flores, Valdir do Nascimento. III. Universidade
Federal do Rio Grande do Sul. Instituto de Letras. Programa de
Pós-Graduação em Letras. IV. Título.

**SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO
E LINGUAGENS DOCUMENTÁRIAS:**
contribuições dos Estudos da Linguagem

Vera Maria Araujo Pigozzi de Araujo

Tese de Doutorado em Estudos da Linguagem
apresentada ao Programa de Pós-Graduação em
Letras da UFRGS como requisito para a obtenção do
título de Doutora em Letras.

Aprovada em: Porto Alegre, 02 de julho de 2013.

Profa. Dra. Maria José Bocorny Finatto

Prof. Dr. Valdir do Nascimento Flores

Profa. Dra. Cristina Dotta Ortega

Profa. Dra. Heloísa Koch Delgado

Profa. Dra. Patrícia Chittoni Reuillard

DEDICATÓRIA

Dedico esta pesquisa para o meu marido, filhos, netos, genros e nora. Eles souberam entender a importância deste trabalho para a minha vida pessoal e profissional, respeitando minhas horas de pesquisa, e contribuindo com seus conhecimentos nas áreas de Informática e de línguas.

Não posso deixar de registrar a influência marcante que meus pais e meu marido tiveram em minha vida, na medida em que me ensinaram a lutar e ir a busca de novas conquistas, sempre.

Dedico esta pesquisa, também, à minha irmã e amiga Sílvia, que foi quem me influenciou para que viesse a ser bibliotecária; à minha irmã e amiga Lia pelo apoio que sempre me deu em todas as minhas decisões; à minha prima e amiga Ana Ferlini que me deu muita força durante esse período, compartilhando comigo de tarefas que sozinha não conseguiria cumprir sem que causasse prejuízos para a escrita desta tese; às minhas amigas e colegas da BSCSH, Lizete e Dóris, que sempre me incentivaram para que viesse a fazer mestrado e doutorado; e à minha amiga Maria Dinah Bergmüller por ter me dado a primeira oportunidade de trabalho nesta Universidade.

Dedico esta pesquisa a todos aqueles que de uma forma ou outra contribuíram para que esta tarefa chegasse ao fim.

AGRADECIMENTOS

Agradeço aos meus colegas da Biblioteca Setorial de Ciências Sociais e Humanidades, em especial ao grupo de bibliotecários do Setor de Processamento Técnico, com os quais tive excelentes discussões e que em muito contribuíram para o desenvolvimento desta pesquisa. Aos colegas Michel, Leonardo, Gesmar e Tatiana meu muito obrigada. À Aninha, um agradecimento especial, por ter me dado a oportunidade de voltar a exercer as atividades de bibliotecária nesta Universidade, no Núcleo de Pesquisa e documentação em Filosofia Balthazar Barbosa Filho, e às colegas do Núcleo, Maria de Lourdes Mendonça e Hedy Lubisco, por terem participado de inúmeras discussões e contribuído com excelentes sugestões para que esta pesquisa pudesse se qualificar.

Um agradecimento muito mais do que especial a minha amiga e colega Cris, Cristina Bürger, por ter me auxiliado na normalização da tese.

Preciso agradecer ainda a contribuição que recebi no decorrer da minha vida acadêmica das Professoras Doutoras Anna Maria Maciel, Cleci Bevilacqua, Maria da Graça Krieger que marcaram a minha vida acadêmica. Às professoras doutoras Patrícia Chittoni Reuillard e Regina Helena van der Laan, professoras que participaram da banca de qualificação da tese, pela sua disponibilidade e oportunas observações.

E, como não poderia deixar de ser, meus mais sinceros agradecimentos aos meus dois grandes mestres, Maria José Finatto e Valdir do Nascimento Flores, meus grandes colaboradores nesta pesquisa. Se esta pesquisa tem valor, certamente devo a eles a grande contribuição que deram, e que particularmente considero ter sido o diferencial desta investigação.

**Mesmo quando tudo parece desabar, cabe a
mim decidir entre rir e chorar, ir ou ficar,
desistir ou lutar; porque descobri, no caminho
incerto da vida, que o mais importante é o
decidir.**

Cora Coralina

**Feliz aquele que transfere o que sabe e aprende
o que ensina.**

Cora Coralina

RESUMO

Esta pesquisa teve como objetivo gerar uma reflexão no campo dos Estudos da Linguagem e da Ciência da Informação com o propósito de investigar a natureza e o funcionamento das Linguagens Documentárias. Foi na Linguística da Enunciação e nos Estudos de Terminologia que se buscou referencial para esta investigação. Os principais teóricos desta pesquisa são: Benveniste, Martins, Cabré, Hoffmann e Ciapuscio. Fundamentado neste referencial foi possível apresentar duas propostas: sugerir um modelo de comunicação capaz de representar a relação que se estabelece entre Sistemas de Recuperação da Informação e usuários a partir de parâmetros enunciativos, e propor a geração de vocabulários controlados especializados sob o referencial enunciativo e terminológico. Estas propostas são fundamentadas nos seguintes conceitos e noções: enunciação, linguagem, língua, estrutura, sentido, significado, referência, contexto, subjetividade, intersubjetividade, locutor e categorias de pessoa, espaço e tempo, no quadro teórico da Teoria da Enunciação; 2) comunicação especializada, linguagem especializada, texto especializado, unidade terminológica, unidade fraseológica especializada e variação terminológica conceitual e denominativa, no domínio dos estudos terminológicos; 3) organização do conhecimento e representação da informação, linguagens documentárias, sistema nocional, vocabulários controlados, análise documentária, tesouros, ontologias, Sistemas de Recuperação da Informação e comunicação documentária. Concluiu-se ser possível estabelecer uma interface produtiva entre os Estudos da Linguagem e Ciência da Informação, com benefício para ambos.

Palavras-chave: Análise Documentária. Ciência da Informação. Comunicação documentária. Linguagens Documentárias. Linguística da Enunciação. Linguística de *Corpus*. Sistemas de Recuperação da Informação. Socioterminologia. Teoria Comunicativa da Terminologia. Estudos de Terminologia.

ABSTRACT

The aim of this research is to generate a reflection in the field of Language Studies and Information Science in order to investigate the nature and the functioning of Documentary Languages. The referential for this investigation comes from Linguistics of Enunciation and from Terminological Studies, mainly of Benveniste, Martins, Cabré, Hoffmann and Ciaspuscio. Based on this referential it was possible to present two proposals: to suggest a communication model able to represent the relationship established between Information Retrieval Systems and users through its enunciative parameters, and propose the generation of domain-specific controlled vocabularies under the enunciative and terminological referential. These proposals are based on the concepts and notions: 1) enunciation, language, langue, structure, meaning, significance, reference, context, subjectivity, intersubjectivity, speaker and categories of person, space and time, in the theoretical framework of the Theory of Enunciation; 2) specialized communication, specialized language, specialized text, terminological unit, specialized phraseological unit, terminological variation conceptual and denominative, in the area of terminological studies; 3) knowledge organization and representation of information, documentary languages, system notional, controlled vocabularies, documentary analysis, thesauri, ontologies, Information Retrieval Systems and documentary communication. It was concluded to be possible to establish a productive interface between Language Studies and Information Science, with benefit to both.

Keywords: Documentary Analysis. Documentary communication. Documentation. *Corpus* Linguistics. Information Retrieval Systems. Information Science. Linguistics of Enunciation. Socioterminology. Communicative Theory of Terminology. Terminology Studies.

LISTA DE ILUSTRAÇÕES

QUADRO 1 – Modelo de comunicação documentária	166
QUADRO 2 – Particularidades da relação que se estabelece entre SRIs e usuários	185
QUADRO 3 – Aparelho formal da enunciação e a relação SRIs e usuários	190
QUADRO 4 – Arquitetura da base de dados	212
QUADRO 5 – Registro dos dados no Excel	214
QUADRO 6 – % de crescimento de ULDEs e ULRIEs no Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo	216
QUADRO 7 – % de ocorrências comuns à coleta manual e à extração automática de ULDEs e ULRIEs no Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo	217

LISTA DE SIGLAS

ADM – Administração
ARQ – Arquivologia
BIB – Biblioteconomia
BIB/CAT – Biblioteconomia/Catálogo
BIB/CLAS – Biblioteconomia/Classificação
BIB/INDEX – Biblioteconomia/Indexação
BIB/RI – Biblioteconomia/Recuperação da informação
BNC – British National *Corpus*
BSCSH – Biblioteca Setorial de Ciências Sociais e Humanidades
CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CAT – Catálogo
CDD – Classificação Decimal de Dewey
CDU – Classificação Decimal Universal
CEPRIL – Centro de Pesquisa, Recursos e Informação em Linguagem
CLAS – Classificação
COMN – Comunicação
CSH – Ciências Sociais e Humanidades
EDIT – Editoração
ExATOlp – Extrator Automático de Termos para Ontologias em Língua Portuguesa
FIL – Filosofia
FRs – Filósofos relacionados
GRÁF – Artes gráficas
Grupo Temma – Grupo de Pesquisa da Escola de Comunicação e Artes da USP
INDEX – Indexação
INF – Informática
INTERN - Internet
ISO – International Organization for Standardization
LAEL – Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem
LC – Library of Congress
LD – Linguagem Documentária
LING – Linguística

LIT – Literatura

LN – Linguagem natural

NPF – Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho

PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul

RA – Relação associativa

RI – Recuperação da informação

SRI – Sistema de Recuperação da Informação

SABi/UFRGS – Sistema de Automação de Bibliotecas / UFRGS

SciELO – Scientific Electronic Library Online

SRI – Sistema de Recuperação de Informação

TA – Termo associado

TC – Termo correlato

TCT – Teoria Comunicativa da Terminologia

TD – Teses e Dissertações

TE – Termo específico

TEXTQUIM – Projeto Textos de Química

TG – Termo geral

TR – Termo relacionado

UFE – Unidade fraseológica especializada

UFRGS – Universidade Federal do Rio Grande do Sul

ULDE – Unidade lexical do discurso especializado

ULRIE – Unidade lexical de representação da informação especializada

UNICAMP – Universidade Estadual de Campinas

UP – Usado para

USE – Unidade de significação especializada

USP – Universidade de São Paulo

UT – Unidade terminológica

WsTools – WordSmith tools

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Questões de pesquisa e hipóteses	19
1.2	Justificativa	21
1.3	Objetivos	23
1.4	Público-alvo	24
2	QUADRO REFERENCIAL	25
2.1	Estudos da Linguagem	25
<i>2.1.1</i>	<i>Teoria da Enunciação de Benveniste</i>	25
2.1.1.1	Aspectos gerais da Linguística da Enunciação.....	26
2.1.1.2	Introdução à Linguística da Enunciação de Benveniste.....	28
2.1.1.3	Língua em Benveniste.....	31
2.1.1.4	A noção de estrutura em Linguística da Enunciação.....	32
2.1.1.5	A forma e o sentido na linguagem.....	35
2.1.1.6	Da (inter)subjetividade na linguagem.....	43
2.1.1.7	As categorias de pessoa e tempo.....	56
2.1.1.8	O Aparelho formal da enunciação.....	60
2.1.1.9	Reinterpretação da Teoria da Enunciação de Benveniste.....	66
<i>2.1.2</i>	<i>Estudos de Terminologia</i>	71
2.1.2.1	Unidade terminológica (UT)	75
2.1.2.2	Unidade fraseológica especializada (UFE)	80
2.1.2.3	Terminologia Textual.....	85
2.1.2.4	Tipologia textual.....	92
<i>2.1.3</i>	<i>Linguística de Corpus</i>	100
2.1.3.1	Princípios gerais da Linguística de <i>Corpus</i>	102
2.1.3.2	Características de um <i>corpus</i>	109
2.1.3.3	Ferramentas para a elaboração de <i>corpora</i>	112
2.2	Ciência da Informação	114
<i>2.2.1</i>	<i>Organização do Conhecimento e Representação da Informação</i>	116
2.2.1.1	Linguagens Documentárias.....	118
<u>2.2.1.1.1</u>	<u>Princípios teóricos</u>	119
<u>2.2.1.1.2</u>	<u>Sistema nocional</u>	125
<u>2.2.1.1.3</u>	<u>Classificação quanto às formas de representação dos conceitos</u>	129

2.2.1.1.3.1	<i>Ordem sistemática</i>	129
2.2.1.1.3.2	<i>Ordem alfabética</i>	132
2.2.1.1.4	Vocabulários controlados.....	133
2.2.1.1.5	Fenômenos linguísticos.....	138
2.2.1.1.6	Análise Documentária.....	140
2.2.1.1.7	Metodologias.....	145
2.2.1.1.8	Semelhanças e diferenças entre taxonomias, tesouros e ontologias.....	146
2.2.1.2	Ontologias.....	148
2.2.1.2.1	Definição, finalidade, características e a relação das ontologias com os mapas conceituais.....	148
2.2.1.2.2	ExATOlp - Extrator Automático de Termos para Ontologias em Língua Portuguesa.....	151
2.2.2	<i>Sistemas de Recuperação da Informação</i>	152
2.2.2.1	Conceituação e características dos Sistemas de Recuperação da Informação.....	153
2.2.2.2	Eficácia dos Sistemas de Recuperação da Informação	156
2.2.2.3	Comunicação documentária	162
2.2.2.3.1	Processo de comunicação especializadas.....	163
2.2.2.3.2	Modelos de comunicação documentária.....	165
2.2.2.4	Caracterização da relação SRIs e usuários.....	167
2.2.2.5	Dado, Informação e Conhecimento.....	170
2.2.2.6	Política de indexação.....	173
3	LINGUAGENS DOCUMENTÁRIAS SOB UM NOVO REFERENCIAL	178
3.1	As Linguagens Documentárias sob o referencial dos Estudos Linguagem	179
3.2	As Linguagens Documentárias sob o olhar da Teoria da Enunciação de Benveniste	181
3.3	A comunicação documentária e os SRIs vistos a partir de parâmetros enunciativos	184
3.4	Aparelho formal da enunciação e a relação SRIs e usuários: uma proposta alternativa aos modelos de comunicação documentária	188
3.5	Vocabulários controlados especializados sob o referencial dos Estudos de Terminologia, da Linguística da Enunciação e da Linguística de	

	<i>Corpus</i>	192
3.6	Vocabulários controlados especializados: uma proposta para a sua elaboração a partir de ULRIEs	194
4	PROPOSTA METODOLÓGICA PARA A CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS ESPECIALIZADOS	201
4.1	Seleção dos <i>corpora</i>	203
4.2	Constituição dos <i>corpora</i>	204
4.3	Seleção das ferramentas	206
4.3.1	<i>Ferramenta de análise de textos</i>	207
4.3.2	<i>Instrumentos de gerenciamento de dados</i>	208
4.4	Coleta e análise de dados	208
4.4.1	<i>Levantamento das unidades lexicais do discurso especializado (ULDEs)</i>	209
4.4.2	<i>Seleção da terminologia representativa de um domínio</i>	210
4.4.3	<i>Produção de concordâncias</i>	210
4.4.4	<i>Análise das concordâncias</i>	210
4.4.5	<i>Representatividade das ULRIEs</i>	211
4.4.6	<i>Categorização das ULDEs em ULRIEs</i>	211
4.5	Arquitetura da base de dados	212
4.6	Inserção dos dados na base e organização dos relacionamentos	213
4.7	Análise comparativa do processo de extração de candidatos a termos: manual X automático	215
4.8	Composição da lista multilíngue das ULRIEs	218
4.9	Considerações finais sobre a proposta metodológica para a construção de vocabulários controlados especializados	218
5	CONSIDERAÇÕES FINAIS E PERSPECTIVAS	220
	REFERÊNCIAS	227
	APÊNDICE A – Glossário de termos da Teoria da Enunciação de Benveniste¹	236
	APÊNDICE B – Glossário de termos da Ciência da Informação	244
	APÊNDICE C – <i>Corpus</i> sobre Filosofia platônica e Platonismo	266
	APÊNDICE D – Análise de texto usando <i>Concordance</i>	271

¹ As definições utilizadas neste glossário são extraídas do Dicionário de Linguística da Enunciação elaborado por Flores et al. (2009).

APÊNDICE E – Visualização do arquivo original (<i>file view</i>).....	272
APÊNDICE F – Lista multilíngue das ULRIES.....	273
APÊNDICE G – Listas de conceitos geradas pela ferramenta <i>Treebolic</i> a partir do <i>corpus</i> de e sobre Platão.....	274
APÊNDICE H – Árvore hiperbólica gerada pela ferramenta <i>Treebolic</i> a partir do <i>corpus</i> de e sobre Platão.....	276
APÊNDICE I – Vocabulário Controlado Especializado de ULRIEs no domínio da Filosofia antiga.....	277
APÊNDICE J – Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo.....	286
APÊNDICE K – Apêndice ao Vocabulário Controlado Especializado de ULRIEs no domínio da Filosofia antiga – Relação de filósofos	295

1 INTRODUÇÃO

Esta tese de doutoramento se inscreve na linha de pesquisa Lexicografia e Terminologia: relações textuais, e é vinculada à área de concentração Estudos da Linguagem. É uma pesquisa que enfatiza a necessidade de cooperação entre Ciência da Informação, Estudos de Terminologia e Linguística da Enunciação visando aprofundar a investigação no âmbito dos Sistemas de Recuperação da Informação² (SRI) e das Linguagens Documentárias.

A investigação explanada nesta tese dará continuidade aos estudos desenvolvidos no Mestrado por esta pesquisadora. A dissertação *Documentação, Terminologia e Lingüística: uma interface produtiva* (2006) teve como principal objetivo propor uma metodologia, visando agilizar a construção de uma ferramenta no campo da Ciência da Informação que satisfizesse a condição de comunicação entre o Sistema e o usuário. Para alcançar este objetivo, tínhamos como proposta a geração de uma base de dados terminológica com sustentação na terminologia utilizada pelo especialista em sua área de domínio.

Esta nova pesquisa, tal como compreendido desde o Mestrado, será construída a partir do entendimento que:

1) os Estudos da Linguagem oferecem uma interface produtiva com os estudos desenvolvidos em Ciência da Informação;

2) os modelos de comunicação utilizados para explicar a forma como ocorre a comunicação documentária não satisfazem, pois não contemplam as especificidades que caracterizam a relação que se constitui entre os SRIs e os usuários;

3) os vocabulários controlados especializados disponibilizados ao bibliotecário para o exercício das atividades de indexação, na área das Ciências Sociais e Humanidades, são insuficientes quanto à representação temática de uma área de domínio;

4) a Linguística de *Corpus* e a Ontologia entendidas como uma metodologia e uma abordagem teórica a ser empregada na construção de vocabulários controlados especializados com fundamentação terminológica, ainda é pouco utilizada.

Recorreremos aos Estudos da Linguagem para refletir sobre diferentes questões que se inserem no quadro teórico das Linguagens Documentárias. Nestes Estudos buscaremos entender o funcionamento da língua, visando criar condições favoráveis para gerar

² A terminologia SRI(s), Sistema(s) de Informação, Sistema(s), Sistema(s) de Recuperação da Informação será usada nesta tese como variantes de um mesmo conceito. Sob esse enfoque Sistema será grafado sempre com S maiúsculo. Para se referir a noção de sistema como conceito, conforme entendido pelos Estudos da Linguagem, a grafia usada é com s minúsculo.

vocabulários controlados especializados, que sejam capazes de estabelecer uma relação de interação mais satisfatória entre os SRIs e os usuários. A razão desta busca, nesta área de conhecimento, justifica-se por ser a linguagem o código utilizado pelos Sistemas de Recuperação da Informação para codificar o conteúdo informacional de um documento e, assim, propiciar a circulação da informação.

Em relação ao modelo de comunicação é importante considerar que, no quadro teórico da Ciência da Informação, ele é entendido com o sentido de uma representação do modo como se dá a comunicação entre os Sistemas de Recuperação da Informação e os usuários. Nesta pesquisa examinaremos esta comunicação a partir de parâmetros enunciativos.

Os vocabulários controlados especializados, por sua vez, serão examinados a partir do referencial que tem origem nos Estudos de Terminologia, Linguística de *Corpus*, Ontologia e Teoria da enunciação, associado ao conjunto de conceitos e noções da Ciência da Informação. Para os propósitos desta pesquisa, **compreendemos por vocabulário controlado especializado um instrumento construído a partir de unidades lexicais de representação da informação especializada (ULRIEs)**, que são representativas das unidades terminológicas (UTs) e das unidades fraseológicas especializadas (UFEs). Essa concepção implica em um referencial fundamentado em princípios dos estudos e práticas da Terminologia.

O referencial que servirá de base para a construção desta pesquisa será construído com o desenvolvimento de duas seções: Estudos da Linguagem e Ciência da Informação.

A seção Estudos da Linguagem será construída a partir de três subseções. São elas:

1) Teoria da Enunciação de Benveniste – fundamentado nesta teoria e a partir de uma apreciação sobre a eficácia dos Sistemas de Recuperação da Informação, construiremos um referencial caracterizando a relação que ocorre entre os SRIs e os usuários, com o propósito de oferecer uma proposta que contemple as especificidades desta relação;

2) Estudos de Terminologia – a investigação neste domínio, associada àquela efetuada em Teoria da Enunciação de Benveniste e Linguística de *Corpus*, nos possibilitará propor um novo referencial teórico e metodológico sobre Linguagens Documentárias, visando à geração de ferramentas como os vocabulários controlados especializados;

3) Linguística de *Corpus* – alicerçado em seu referencial teórico e metodológico pretendemos agilizar a produção dessas ferramentas.

A seção Ciência da Informação se desenvolverá a partir da investigação de duas subseções: Organização do Conhecimento e Representação da Informação, e Sistemas de Recuperação da Informação. A revisão de literatura, feita no domínio da Ciência da

Informação, terá como propósito estabelecer as bases para a construção do referencial que estamos propondo. Dentre as discussões que faremos nesta pesquisa está aquela que questiona o tipo de relação que SRIs e usuários estabelecem entre si: trata-se de uma relação de comunicação fundamentada na ideia de emissor e receptor, ou de uma relação de comunicação intersubjetiva abalizada na ideia de interação, conforme propõe Benveniste? E, ainda, as unidades lexicais de representação da informação especializada, entendidas como as unidades de um vocabulário controlado especializado, oferecem melhores condições de representação da informação do que os cabeçalhos de assunto e os descritores?

Feita a revisão de literatura, abordaremos as Linguagens Documentárias sob um novo referencial. Este tema será desenvolvido em seis seções:

1) As Linguagens Documentárias sob o referencial dos Estudos Linguagem – com o propósito de identificar um referencial no domínio dos Estudos da Linguagem, que possa contribuir para minimizar os problemas que têm origem no uso de unidades lexicais de representação da informação especializada desvinculadas do texto indexado, fazemos uma breve exposição sobre as seguintes questões: processo de construção do sentido, modelos de comunicação documentária e construção de vocabulários controlados especializados;

2) As Linguagens Documentárias sob o olhar da Teoria da Enunciação de Benveniste – sob este enfoque refletiremos sobre o funcionamento da língua, assim como sobre os seguintes conceitos e noções: enunciação, linguagem (língua em ação), língua, estrutura, semiótica, semântica, significado, sentido, subjetividade, intersubjetividade, locutor, referência, contexto e categorias de pessoa, espaço e tempo;

3) A comunicação documentária e os SRIs vistos a partir de parâmetros enunciativos – com a intenção de discutir a comunicação documentária e os SRIs, fundamentaremos nossa investigação em parâmetros enunciativos na visão de Benveniste (1988a, 1989a) e de Martins (1990). Nesta seção serão explicitadas as particularidades que envolvem a relação que se estabelece entre SRIs e usuários. O estudo do referencial advindo da Teoria da Enunciação de Benveniste tem como propósito aprofundar o conhecimento das Linguagens Documentárias quanto ao seu funcionamento;

4) Aparelho formal da enunciação e a relação SRIs e usuários: uma proposta alternativa aos modelos de comunicação documentária – nesta seção além de oferecermos uma reflexão sobre o aparelho formal da enunciação de Benveniste apresentaremos uma proposta alternativa aos modelos de comunicação documentária. Esta proposta é fundamentada nos seguintes conceitos e noções: subjetividade, intersubjetividade, locutor e categorias de pessoa, espaço e tempo (eu, tu, ele, aqui e agora);

5) Vocabulários controlados especializados sob o referencial dos Estudos de Terminologia, da Linguística da Enunciação e da Linguística de *Corpus* – trazemos para o quadro teórico das Linguagens Documentárias os seguintes conceitos e noções: linguagem especializada, texto especializado, unidades terminológicas, unidades fraseológicas especializadas e variação terminológica conceitual e denominativa. Sob este referencial, ainda, reconhecemos as unidades terminológicas e as unidades fraseológicas especializadas como as unidades lexicais de representação da informação especializada, assim como, a necessidade de integrar os processos de indexação e de busca da informação e recuperação de documentos, e de considerar relevante duas características destes processos: a subjetividade e a intersubjetividade.

6) Vocabulários controlados especializados: uma proposta para a sua elaboração a partir de ULRIEs – propomos nesta seção a construção de uma ferramenta que privilegie a linguagem utilizada pelo especialista, que considere o texto como unidade de análise, e que reconheça a variação terminológica e a presença de uma densidade informativa diferenciada em textos especializados. Com o propósito de oferecer uma metodologia mais ágil para a construção de ferramentas que promovam a melhoria dos processos de indexação temática, buscaremos os princípios teóricos e metodológicos da Linguística de *Corpus*, acrescidos dos recursos disponibilizados pela Ontologia.

Nossa proposta, a título de projetos pilotos, é a de desenvolver vocabulários controlados especializados para as áreas da Filosofia. Acreditamos que esta proposta possa ser estendida para outras áreas das Ciências Sociais e Humanidades (CSH³), por se assemelharem quanto ao tipo de linguagem que utilizam na divulgação do conhecimento produzido. A opção pela área de conhecimento Filosofia deve-se ao fato de ser possível testar a proposta desenvolvida em um Sistema de Recuperação da Informação, disponibilizado via *web*, e do fato desta pesquisadora ter familiaridade com o conhecimento produzido neste domínio.

Assim, a partir de parâmetros enunciativos e terminológicos, ofereceremos duas propostas, uma relacionada aos modelos de comunicação documentária e outra aos vocabulários controlados especializados. Essa discussão, que evolui a partir da articulação entre Estudos da Linguagem e Ciência da Informação, propiciará o desenvolvimento de um

³ Observa-se, na literatura consultada, a falta de consenso quanto à classificação das áreas de conhecimento que compõem o grupo das CSH. Para as finalidades desta pesquisa foram consideradas como inseridas no conjunto destas ciências, as seguintes áreas: Ciências Sociais (Antropologia, Política e Sociologia), Filosofia, História, Artes, Estudos da Linguagem e Literatura. Fez-se 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.

referencial enunciativo e terminológico no campo das Linguagens Documentárias e dos Sistemas de Recuperação da Informação. Na sequência, faremos a exposição das questões de pesquisa e das hipóteses.

1.1 Questões de pesquisa e hipóteses

Tendo estabelecido os Estudos da Linguagem como o referencial que sustentará nossa investigação em Ciência da Informação, formulamos as seguintes questões de pesquisa:

- 1) como se caracteriza o processo de comunicação documentária?
- 2) de que modo a Teoria da Enunciação pode ser usada para explicar a relação que se estabelece entre os Sistemas de Recuperação da Informação e os usuários?
- 3) de que forma os Estudos de Terminologia podem contribuir para a construção de vocabulários controlados especializados?
- 4) até que ponto uma ULRIE é mais adequada que um descritor para representar a informação indexada, e assegurar a preservação da referência desta unidade de informação na passagem do processo de indexação para o processo de busca da informação e recuperação de documentos?
- 5) de que maneira a Linguística de *Corpus* e a Ontologia oferecem um referencial teórico e metodológico adequado para a construção de vocabulários controlados especializados?

Entendendo que a relação que se estabelece entre os interlocutores de um SRI é de intersubjetividade; que bibliotecários, usuários e Sistemas são interlocutores dos processos de indexação, assim como dos processos de busca da informação e recuperação de documentos; que as unidades terminológicas e as unidades fraseológicas especializadas expressam o conhecimento especializado; que os vocabulários controlados especializados devem ser representativos das linguagens dos especialistas; que uma unidade de representação da informação especializada precisa retratar conceitos e noções de um domínio; e que a língua necessita ser entendida em seu contexto de uso, formulamos as hipóteses desta pesquisa.

Nossa primeira hipótese será construída a partir da visão de que uma reflexão em Estudos da Linguagem, sobre a natureza e o funcionamento das Linguagens Documentárias, pode contribuir para melhorar a relação que se estabelece entre SRIs e usuários.

- 1) É possível estabelecer uma interface produtiva entre os Estudos da Linguagem e a Ciência da Informação.

Considerando que a relação que o usuário estabelece com o Sistema é insuficiente quanto às suas expectativas de pesquisa, formulamos uma segunda e terceira hipótese.

2) São interlocutores da comunicação documentária não apenas usuário e Sistema, mas também o bibliotecário.

3) A relação que se estabelece entre os interlocutores da comunicação documentária é de intersubjetividade e se insere em um processo de comunicação intersubjetiva.

Utilizando a Teoria da Enunciação formulada por Benveniste como referencial teórico para discutir o modelo de comunicação documentária, pretendemos investigar a validade da quarta e quinta hipótese.

4) Um olhar enunciativo no âmbito dos Sistemas de Recuperação da Informação e das Linguagens Documentárias permite considerar peculiaridades próprias do tipo de relação que se constitui entre os SRIs e os usuários.

5) As noções de linguagem, subjetividade, intersubjetividade, referência, contexto e categorias de pessoa, espaço e tempo estão presentes na relação que se estabelece entre SRIs e usuários.

Tendo em vista que as ferramentas disponibilizadas ao bibliotecário para a indexação de documentos, como os tesouros e listas de cabeçalhos de assunto, não privilegiam a linguagem utilizada pelos especialistas, apresentamos outras duas hipóteses.

6) Um vocabulário controlado especializado será capaz de representar mais fielmente o conteúdo temático de um documento, se construído a partir da concepção de unidades terminológicas e unidades fraseológicas especializadas, e de um *corpus* formado por textos elaborados pelos autores dos documentos.

7) Uma unidade de representação da informação especializada tende a permitir maior representatividade de um domínio, uma vez que ela representa os conceitos e as noções de uma área de conhecimento.

Pensando em construir um vocabulário controlado especializado, que privilegie a linguagem utilizada pelo especialista e que seja capaz de compatibilizar a linguagem por ele empregada com a do usuário, formulamos uma oitava hipótese.

8) A Linguística de *Corpus* e a Ontologia se apresentam como uma abordagem teórica e uma metodologia adequada para ser utilizada na geração de vocabulários controlados especializados.

E, por fim, entendendo que o registro da terminologia de uma área do conhecimento precisa ser criterioso, propomos uma nona e última hipótese.

9) Para o registro da terminologia de uma área de conhecimento, uma base de dados precisa ser concebida sob o referencial terminológico, viabilizando o registro de ULRIES precisas e contextualizadas.

Dando continuidade a esta introdução, apresentaremos a justificativa e os objetivos que norteiam esta pesquisa.

1.2 Justificativa

A relevância desta pesquisa está em propor uma ampla discussão no domínio da Ciência da Informação, no âmbito das Linguagens Documentárias e dos Sistemas de Recuperação da Informação, a partir dos Estudos da Linguagem. Com esse propósito desenvolveremos uma reflexão sobre a eficácia dos Sistemas de Recuperação da Informação em sua relação com os usuários, assim como sobre as Linguagens Documentárias quanto à representação da informação indexada e armazenada nestes Sistemas. Ela evoluirá a partir da articulação da Ciência da Informação com a Linguística da Enunciação, os Estudos de Terminologia e a Linguística de *Corpus*.

Nosso propósito é desenvolver uma investigação sob dois aspectos:

- 1) qual o tipo de relação que se estabelece entre usuários e Sistemas (catálogos eletrônicos);
- 2) que ferramenta deve ser construída para satisfazer as necessidades de pesquisa do usuário (listas de cabeçalhos de assunto, tesouros ou vocabulários controlados especializados).

Nossa investigação evoluirá a partir da seguinte visão: 1) um bom Sistema de Recuperação da Informação depende fundamentalmente da Análise Documentária, que visa não apenas a construção de vocabulários controlados, como também o tratamento da informação contido em um documento; 2) as Linguagens Documentárias devem ser representativas de um domínio e expressar claramente as relações existentes entre suas unidades lexicais de representação da informação especializada; 3) os conceitos e noções de um domínio precisam ser codificados e representados de forma precisa. Sob essa ótica acreditamos que é possível oferecer condições mais propícias para que a interação entre usuários e Sistema ocorra de forma facilitada, e que a satisfação do usuário seja mais significativa.

A insatisfação do usuário, percebida quando trabalhava na Biblioteca Setorial de Ciências Sociais e Humanidades, e constatada na literatura da área, conforme estudos

desenvolvidos por Araujo⁴ (1995) e Biscalchin e Boccato (2011), é consequência dos seguintes fatores: 1) da falta de consistência dos catálogos eletrônicos; 2) do uso predominante de termos genéricos para representar tematicamente os documentos processados, com prejuízo para a recuperação de documentos por temas específicos; 3) da incompatibilidade existente entre a linguagem empregada no Sistema e aquela usada pelo usuário; 4) da falta de remissivas que conduzam o usuário, do termo que ele usa para aquele adotado pelo Sistema; e 5) da falta de treinamentos adequados para o uso dos recursos de pesquisa que a base oferece.

Em consulta feita em catálogos eletrônicos como a Library of Congress, Bibliothèque Nationale de France, base Acervus do Sistema de Bibliotecas da UNICAMP e catálogo SABI-UFRGS (NPF) observamos a diversidade e divergência de descritores utilizados para um mesmo documento. Muitas vezes, inclusive, constatamos que os documentos não são representados pela sua temática.

Visando esclarecer a situação mencionada, fazemos referência ao diálogo *Protágoras* de Platão, que nos catálogos eletrônicos citados anteriormente foi indexado utilizando como descritores ‘Filosofia grega’, ‘Filosofia platônica’ e ‘Platonismo’. Na BSCSH/NPF/UFRGS, entretanto, este mesmo diálogo foi indexado utilizando as seguintes ULRIEs: ‘Filosofia grega’, ‘Filosofia platônica’, ‘Diálogo platônico’, ‘Diálogo socrático’ (nem todos os diálogos de Platão são diálogos socráticos), ‘Ética platônica’, ‘Mito de Epitemeu’, ‘Mito de Prometeu’, ‘Mito platônico’, ‘Sofistas’, ‘Teoria das virtudes’ e ‘Teoria do conceito’.

A indexação feita pela BSCSH/NPF/UFRGS revela o uso de termos genéricos, específicos e relacionados à temática do diálogo. Ressaltamos que ela é o resultado de um projeto de geração de vocabulários controlados especializados, que tem como propósito viabilizar a representatividade de um domínio em SRIs e em seus diferentes níveis de especificidade.

A situação referida acima evidencia a ausência de vocabulários controlados especializados que sustentem as atividades de indexação, e o fato da indexação ser um processo extremamente subjetivo, o que é constatado pela variedade de descritores utilizados para a representação temática de um mesmo documento.

Nesta pesquisa, ainda, partimos do princípio de que uma informação só tem seu sentido assegurado se as condições para a construção da referência forem garantidas; que a construção do conhecimento só é possível quando se tem acesso à informação, e não

⁴ Assunto desenvolvido no item 2.2.2.2 Eficácia dos Sistemas de Recuperação da Informação.

simplesmente a dados; e, ainda, que a confiabilidade de um SRI depende da sua capacidade de viabilizar a circulação da informação com equilíbrio entre a revocação e a precisão.

Um exame sobre Sistemas de Recuperação da Informação e Linguagens Documentárias, a partir do aporte teórico da Teoria da Enunciação de Benveniste, dos Estudos de Terminologia, da Linguística de *Corpus* e da Ontologia, permite gerar uma contribuição importante ao aprofundamento de questões relacionadas à circulação da informação e à produção do conhecimento.

Concluimos esta introdução com o entendimento de que a aproximação da Ciência da Informação com os Estudos da Linguagem é de fácil compreensão, uma vez que estas duas áreas de conhecimento estão envolvidas com as questões que dizem respeito ao funcionamento da linguagem. Conforme afirmam Krieger e Finatto (2004, p. 59) a correlação entre Documentação e Estudos de Terminologia “[...] se estabelece à luz da natureza e funcionalidade das respectivas unidades básicas de cada área, ou seja, o termo e o descritor [...]”. Finalizamos esta justificativa afirmando que a proximidade destas duas áreas de conhecimento viabiliza uma interface produtiva para ambas; ela se caracteriza por uma relação de bilateralidade e, como tal, aponta para a necessidade de colaboração mútua para o avanço de cada uma delas.

1.3 Objetivos

Esta pesquisa tem como objetivo geral investigar a natureza e o funcionamento das Linguagens Documentárias tendo como referencial teórico os Estudos da Linguagem. Esta investigação tem como propósito refletir sobre a eficácia dos Sistemas de Recuperação da Informação e dos vocabulários controlados especializados, assim como oferecer um novo referencial a partir de parâmetros enunciativos e terminológicos, em relação às questões investigadas.

Ao objetivo geral, acrescentamos os seguintes objetivos específicos.

1) trazer para o âmbito da Ciência da Informação o conhecimento produzido pelos Estudos da Linguagem, de forma que as duas áreas possam se beneficiar mutuamente.

2) ampliar o referencial já desenvolvido no âmbito das Linguagens Documentárias na interface com os Estudos da Linguagem, recorrendo aos fundamentos teóricos da Linguística da Enunciação, dos Estudos de Terminologia e da Linguística de *Corpus*.

3) aprofundar a discussão sobre os modelos de comunicação documentária usualmente apresentados em Ciência da Informação.

4) recomendar a Teoria da Enunciação de Benveniste como referencial teórico para caracterizar a relação que se instaura entre SRIs e usuários.

5) entender esta relação como inserida em um processo de comunicação intersubjetiva, marcada pela presença de interlocutores de natureza heterogênea (bibliotecários, usuários e SRIs).

6) viabilizar uma interação mais efetiva entre os SRIs e os usuários, de modo que um item bibliográfico possa oferecer a recuperação de informações e não apenas de dados, como ocorre com a maioria dos SRIs quando desvinculam a informação do texto indexado.

7) identificar parâmetros enunciativos e terminológicos para a construção de vocabulários controlados especializados.

8) oferecer uma nova concepção de descritores, passando a entendê-los como unidades semânticas e como representantes das unidades terminológicas (UTs) e das unidades fraseológicas especializadas (UFEs) e sendo reconhecidos por ULRIEs, ou seja, por unidades lexicais de representação da informação especializada.

9) sugerir uma nova metodologia para a coleta de ULDEs (unidades lexicais do discurso especializado) e seleção de ULRIEs de um vocabulário controlado especializado, a partir dos princípios teóricos e metodológicos da Linguística de *Corpus* e da Ontologia.

10) propor a arquitetura de uma base de dados para o registro da terminologia de uma área de conhecimento, de modo a viabilizar a construção de um vocabulário controlado especializado.

Finalizamos esta introdução apontando o público-alvo desta pesquisa.

1.4 Público-alvo

Esta pesquisa considera como seu público-alvo o bibliotecário documentalista das bibliotecas universitárias, mas visa também estabelecer um diálogo com os linguistas e terminólogos. Tendo os usuários de redes e sistemas de informação como motivação, direcionaremos os resultados desta investigação no sentido de oferecer-lhes maior satisfação na prática de suas atividades de pesquisa.

2 QUADRO REFERENCIAL

Este capítulo será desenvolvido a partir de duas seções: Estudos da Linguagem e Ciência da Informação, e das seguintes subseções: Teoria da Enunciação de Benveniste, Estudos de Terminologia, Linguística de *Corpus*, Organização do Conhecimento e Representação da Informação, e Sistemas de Recuperação da Informação.

2.1 Estudos da Linguagem

2.1.1 Teoria da Enunciação de Benveniste

A leitura da obra de Benveniste (1988a, 1989a) no contexto desta tese visa o entendimento dos seguintes conceitos e expressões representativas de um domínio: linguagem, língua, estrutura, sentido, significado, referência, contexto, subjetividade, intersubjetividade, locutor e categorias de pessoa, espaço e tempo. A compreensão destas noções tem dois propósitos: aprofundar o conhecimento das Linguagens Documentárias, quanto ao seu funcionamento, e formular uma proposta para caracterizar a relação que ocorre entre SRIs e usuários. Os textos selecionados para o desenvolvimento desta subseção fazem parte das obras Problemas de Linguística geral I (1988a) e Problemas de Linguística geral II (1989a) de Benveniste. São eles: *'estrutura' em Lingüística* (1962), *Os níveis da análise Lingüística* (1964), *A forma e o sentido na linguagem* (1967), *Semiologia da língua* (1969), *Estrutura das relações de pessoa no verbo* (1946), *A natureza dos pronomes* (1956), *Da subjetividade na linguagem* (1958), *A linguagem e a experiência humana* (1965) e o *Aparelho formal da enunciação* (1970). Esta subseção finaliza com uma reinterpretação da Teoria da Enunciação de Benveniste, conforme proposta apresentada por Martins (1990) em sua obra *Enunciação e diálogo*. Para uma melhor compreensão da leitura desta subseção por parte de leitores não familiarizados com os conceitos da Linguística da Enunciação, em especial os profissionais da informação, ofereceremos um apêndice contendo um glossário com os termos mais relevantes da área, presentes neste texto. (APÊNDICE A – Glossário de termos da Teoria da Enunciação de Benveniste).

A revisão de literatura desta subseção será feita sob os seguintes itens: Aspectos gerais da Linguística da Enunciação; Introdução à Linguística da Enunciação de Benveniste; Língua em Benveniste; A noção de estrutura em Linguística da Enunciação; A forma e o sentido na linguagem; Da (inter)subjetividade na linguagem; As categorias de pessoa e tempo; O

Aparelho formal da enunciação; Reinterpretação da Teoria da Enunciação de Benveniste; e Teoria da Enunciação de Benveniste: sua contribuição para as Linguagens Documentárias. Iniciaremos esta seção fazendo uma breve exposição sobre a Linguística da Enunciação.

2.1.1.1 Aspectos gerais da Linguística da Enunciação

A Linguística da Enunciação, assim denominada por Flores e Teixeira (2005), compreende diversas teorias, que podem ser incluídas em um único rótulo por apresentarem várias características em comum. Dentre os princípios que permitem que tais teorias sejam reunidas sob uma única denominação, está o de serem filiadas ao pensamento de Saussure, e de se preocuparem com a questão do sentido na linguagem.

A possibilidade de se pensar em um campo que seja reconhecido como Linguística da Enunciação é viável, segundo os linguistas citados anteriormente, uma vez que é possível identificar diferentes teorias que falam da enunciação. Enunciação é um objeto multifacetado que obedece às restrições teórico-metodológicas impostas pelas diferentes teorias da enunciação, e que está na dependência da meta a cumprir, ou seja, da ótica adotada sobre o fenômeno. Encontramos na literatura da área diversas abordagens tomando a enunciação como foco, tais como: enunciação e subjetividade, enunciação e argumentação, enunciação e polifonia, enunciação e metalinguagem, enunciação e comunicação, por exemplo. Nesta pesquisa abordamos a enunciação na relação com os Sistemas de Recuperação da Informação e com as Linguagens Documentárias.

Ainda que se observe que à Linguística da Enunciação se integram diferentes teorias, ressaltamos o fato de que todas têm a enunciação como seu objeto de estudo, e que há traços comuns entre as diferentes abordagens enunciativas, dentre elas a de Bally, Jakobson, Benveniste, Ducrot e Bakhtin. Mas salientamos, no entanto, que a configuração desse objeto é distinta para cada teoria e seu autor e, ainda, que entre as diferentes teorias é possível vislumbrar a unidade na diversidade.

Essas teorias surgiram em momentos diferentes e tiveram um desenvolvimento especial na França. Elas propõem um novo modo de ver a linguagem e se caracterizam por inserir o estudo da subjetividade no estudo da língua, uma subjetividade que emerge no discurso. Vale destacar que na visão dos linguistas mencionados – Flores e Teixeira –, o sujeito não é exatamente o objeto de estudo dessas teorias, mas a representação que a enunciação oferece dele. E ainda, tendo como enfoque o campo de saber é possível falar em

Linguística da Enunciação; entretanto, sob o prisma da análise linguística é de teorias da enunciação que se deve falar.

Nas palavras dos autores que fundamentam o posicionamento assumido nesta pesquisa sobre Linguística da Enunciação, encontramos os seguintes enunciados:

[...] as teorias da enunciação, cada uma ao seu modo, concebem uma forma de analisar os fenômenos circunscritos pela concepção da enunciação que têm (FLORES; TEIXEIRA, 2005, p. 104);

[...] os fenômenos estudados nas teorias da enunciação pertencem à língua, mas não se encerram nela; pertencem à fala na medida em que só nela e por ela têm existência, e questionam a existência de ambas já que emanam das duas (FLORES; TEIXEIRA, 2005, p. 105);

Qualquer fenômeno linguístico de qualquer nível (sintático, morfológico, fonológico etc.) pode ser abordado desde o ponto de vista da linguística da enunciação (FLORES; TEIXEIRA, 2005, p. 106).

Apesar da aparente dispersão, há algo de unificador entre as diferentes teorias, a saber, “[...] a crença na língua como ordem própria que precisa ser atualizada pelo sujeito a cada instância de uso” (FLORES; TEIXEIRA, 2005, p. 106). Entendendo que o estudo da enunciação compreende a língua em sua totalidade, é possível dizer que “O objeto da linguística da enunciação é todo o mecanismo linguístico cuja realização integra o seu próprio sentido e que se auto-referencia no uso. A enunciação é, pois, o que constitui esse processo” (FLORES; TEIXEIRA, 2005, p. 106).

Conforme estes linguistas, a Linguística da Enunciação é a nomeação do múltiplo e não do unívoco; é um ponto de chegada para onde convergem diferentes teorias. A enunciação não é um nível de análise, mas um ponto de vista. Segundo Flores e Teixeira (2005, p. 107) ainda:

A linguística da enunciação toma para si não apenas o estudo das marcas formais no enunciado, mas refere-se ao processo de sua produção: ao sujeito, tempo e espaço. *A linguística da enunciação* deve centrar-se no estudo das representações do sujeito que enuncia e não no próprio sujeito, objeto de outras áreas.

Dentre os teóricos que se inserem nos estudos que se enquadram na Linguística da Enunciação, destacamos Benveniste. A opção por este teórico, no âmbito desta pesquisa, é em razão do entendimento que buscamos para o esclarecimento de diferentes noções relacionadas aos Sistemas de Recuperação da Informação e às Linguagens Documentárias, conforme mencionado na introdução desta subseção.

Benveniste, diferentemente de outros teóricos, não chega a desenvolver um modelo de análise da enunciação; sua obra oferece reflexões referentes às diferentes questões que envolvem a presença do homem na língua. Na teoria linguística desenvolvida pelo linguista

francês, a subjetividade é pensada no próprio sistema da língua; nela se encontra uma forma de sistematização que possibilita o estudo das relações enunciativas. Esta particularidade é relevante porque permite novas leituras a cada enunciação.

A Linguística da Enunciação sob a ótica de Benveniste será abordada na próxima subseção, dando início a uma reflexão sobre a sua teoria e o entendimento das noções por ele propostas, fundamentando, assim, um novo referencial no âmbito dos Sistemas de Recuperação da Informação e das Linguagens Documentárias.

2.1.1.2 Introdução à Linguística da Enunciação de Benveniste⁵

Benveniste é considerado por muitos teóricos o linguista da enunciação e o principal representante da Teoria da Enunciação. Conforme Flores et al. (2009, p. 102) para Benveniste enunciação, é a “Colocação da língua em funcionamento por um ato individual de utilização”. Talvez Benveniste tenha sido o primeiro linguista pós-saussuriano a desenvolver um modelo de análise da língua direcionado à enunciação. Suas reflexões foram desenvolvidas no apogeu do estruturalismo no âmbito das Ciências Humanas, construindo um método rigoroso de análise sobre fenômenos que eram excluídos da investigação científica.

O estruturalismo teve início com Saussure⁶ a partir da clássica dicotomia *langue/parole*, ou seja, língua e fala.

O objeto da linguística foi concebido, na perspectiva estruturalista, como um sistema de relações internas do qual se deveria reter as leis de organização. Disso resultou o objetivo geral de investigar as regularidades do sistema, abstraindo, para tanto, toda a referência a elementos externos ao método (FLORES; TEIXEIRA, 2005, p. 29).

Benveniste toma por base o mesmo estruturalismo de Saussure, mas enfrenta um clima adverso quando propõe incluir os estudos da enunciação e os da subjetividade no objeto da Linguística. Ainda que sob determinados aspectos Benveniste é fiel ao pensamento estruturalista de Saussure, ele inova em sua visão de sujeito, daquele que fala e dá sentido, e por supor sujeito e estrutura articulados. Com a sua Teoria da Enunciação, o linguista instaura

⁵ Émile Benveniste nasceu em 1902, na Síria, e morreu em 1976. Em 1924 naturalizou-se francês. Sua abordagem enunciativa desperta interesse de filósofos e psicanalistas. O linguista francês é responsável por introduzir um conjunto de questões concernentes a uma linguística bastante diferenciada da que era feita até então. Com ele, os temas da subjetividade, intersubjetividade, referência e significação, dentre outros, assumem outras proporções. Para outras informações ver o site: <http://www.ufrgs.br/benvenisteonline/>.

⁶ Ferdinand de Saussure, linguista e filósofo, nasceu em Genebra, em 1857 e morreu em Morges, em 1913. Sua teoria propiciou o desenvolvimento da Linguística enquanto ciência autônoma. Ele entendia a linguística como um ramo da ciência mais geral dos signos, que propôs fosse chamada de Semiologia. Seus conceitos serviram de base para o desenvolvimento do estruturalismo no século XX, e para as reflexões feitas por Benveniste e outros linguistas.

um pensamento diferenciado acerca da linguagem, na discussão de questões como: estrutura, forma, sentido, significado, referência, sistema semiótico, sistema semântico, subjetividade e intersubjetividade, categorias de pessoa, espaço e tempo. Esses temas estão desenvolvidos de forma dispersa em diversos textos escritos pelo teórico, e serão abordados nesta pesquisa por viabilizarem um olhar enunciativo no âmbito dos Sistemas de Recuperação da Informação e das Linguagens Documentárias.

Sua obra reúne 20 artigos, que se encontram reunidos em dois volumes: *Problemas de linguística geral I* (1988a) e *Problemas de linguística geral II* (1989a). Os conceitos emitidos pelo linguista se alteram e se definem diferentemente ao longo de sua obra. Essa diversidade conceitual se explica uma vez que sua teoria foi sendo construída ao longo de suas reflexões, em um período de 40 anos.

Benveniste não desenvolveu um modelo de análise da linguagem.

O que se convencionou chamar de a *Teoria de Enunciação* de Benveniste é, na verdade, derivado da leitura de um conjunto de textos escritos entre os anos 1930 e 1970, que simultaneamente teorizam e analisam a *marca do homem na linguagem*, expressão usada pelo próprio Benveniste no prefácio de *Problemas de linguística geral I* (FLORES; TEIXEIRA, 2009, p. 154).

A teoria benvenistiana, ainda que seja reconhecida como tal pelos linguistas, não foi concebida com essa finalidade por Benveniste. O linguista francês não tinha como propósito escrever uma Teoria da Enunciação, uma teoria acabada como usualmente é observado em outros linguistas. Em sua obra não é possível identificar a proposição de um modelo, a partir do qual se apresentam todas as suas análises. Cada um de seus textos encerra maneiras específicas de olhar sua teoria e de analisar os conceitos neles emitidos.

Há uma grande flutuação conceitual em sua obra e essa característica não pode ser entendida como falta de rigor metodológico, mas como fruto de reflexões pontuais, apropriadas aos propósitos de cada texto, e decorrentes de um percurso seguido pelo linguista ao longo dos anos. Assim, não é possível ler seus escritos como se fossem contemporâneos um do outro, uma vez que resultam de reflexões de natureza diversa e em períodos diferentes de sua evolução intelectual.

Em seus escritos os conceitos se alteram e se definem de forma diferenciada. Uma vez que a leitura de seus artigos revela uma grande oscilação terminológica e conceitual, a leitura a ser feita precisa levar em consideração que cada texto encerra em si modos específicos de analisar e de teorizar sobre os temas abordados.

A análise dos pronomes, por exemplo, não é metodológica e conceitualmente a mesma feita nos textos *Estrutura das relações de pessoa no verbo* (1946) e em *O aparelho formal da*

enunciação (1970). Ainda assim, nos dois textos podem ser percebidos aspectos enunciativos de sua teoria.

A própria noção de enunciação em Benveniste não é nem única e nem unívoca. Esse fato é facilmente comprovável pela leitura dos textos *Da subjetividade na linguagem* e *O aparelho formal da enunciação*. Convém lembrar que há entre os dois textos um intervalo de quase vinte anos.

A noção de enunciação, que igualmente aparece sob uma grande variação conceitual, no texto *O aparelho formal da enunciação* é entendida como um processo, como um ato de natureza fônica e de natureza gráfica, por exemplo. Só neste texto há 68 ocorrências da palavra enunciação. A formulação mais famosa sobre enunciação em sua teoria é: “A enunciação é este colocar em funcionamento a língua por um ato individual de utilização” (BENVENISTE, 1989b, p. 82). Já a passagem: “Este grande processo pode ser estudado sob diferentes aspectos” (BENVENISTE, 1989b, p. 82) revela um fato fundamental sobre enunciação, ou seja, o de que não há apenas uma forma de analisá-la, mas a de que ela pode ser estudada sob diferentes aspectos. Do enunciado “O que em geral caracteriza a enunciação é a acentuação da relação discursiva com o parceiro, seja este real ou imaginado, individual ou coletivo” (BENVENISTE, 1989b, p. 87), podemos inferir que a ideia da relação com o outro, ou seja, da intersubjetividade na enunciação é fundante dessa teoria.

Ainda que sua teoria se caracterize por uma grande diversidade conceitual, podemos inferir da leitura de sua obra que o fundamento de todo o pensamento benvenistiana pode ser formulado sob o axioma *o homem está na língua*. Segundo Flores e Teixeira (2009, p. 156) “*O operador do axioma O homem está na língua é a enunciação. É ela, enquanto dispositivo, que o faz funcionar.*”

Este parece ser o princípio da evidência de sua teoria enunciativa e a proposição que nos conduz à identificação dos conceitos primitivos de sua teoria, isto é, conceitos que mantêm interdependência entre si. Na visão destes linguistas os conceitos primitivos de sua teoria são: linguagem, línguas, locutor, sujeito, subjetividade e intersubjetividade. Esse entendimento decorre do raciocínio que entende que a noção de língua presente neste axioma evoca as noções de linguagem e línguas, e que a noção de homem, por sua vez, nos remete às noções de locutor e sujeito e, estas duas, às noções de subjetividade e de intersubjetividade.

O texto *Da subjetividade na linguagem* atesta esse entendimento.

A intersubjetividade/subjetividade ali estudada inclui a ordem da *linguagem* – o título já atesta isso – a ordem da *língua* – já que a análise conclui em favor de uma generalização sistêmica da oposição *pessoa/não-pessoa* – e a ordem das *línguas* – já que há análises de inúmeras línguas (o francês,

certamente, mas também as línguas do extremo oriente das quais Benveniste era profundo conhecedor) (FLORES; TEIXEIRA, 2009, p. 157).

Retomando o que já foi ressaltado anteriormente, voltamos a afirmar que cada texto encerra em si maneiras específicas de analisar e de teorizar sobre língua, linguagem e línguas.

O entendimento dos conceitos primitivos, segundo Flores e Teixeira (2009), é fundamental para a compreensão da teoria benvenistiana. Assim, por exemplo, para entendermos a afirmação “A ‘subjetividade’ de que tratamos aqui é a capacidade do locutor de se propor como ‘sujeito’” (BENVENISTE, 1988b, p. 286), é preciso saber o que o linguista entende pelos termos locutor e sujeito. Encerramos esta reflexão, esclarecendo que os conceitos primitivos da obra de Benveniste serão abordados nas subseções seguintes.

2.1.1.3 Língua em Benveniste

A noção de língua, assim como as de linguagem e línguas são de grande importância para Benveniste. Esta subseção se restringe a apresentar a noção de língua a partir da leitura do texto *Semiologia da língua*. Esta noção pode ser vista também a partir da leitura das demais subseções desta revisão de literatura, uma vez que ela se estende por quase toda a sua obra. Início esta exposição sobre a noção de língua, com a intenção de aprofundar nosso conhecimento sobre as Linguagens Documentárias, que assim como a língua possui uma estrutura, uma organização interna dentro de um sistema.

Neste texto, o linguista francês concebe a língua como o único modelo de um sistema que é simultaneamente semiótico na sua estrutura formal e no seu funcionamento. A língua apresenta como propriedade o fato de se manifestar pela enunciação e conter referência a uma certa situação – falar é sempre falar de –; ela consiste de signos, que formalmente são unidades distintas; ela é produzida e recebida por todos os membros de uma comunidade, nos mesmos valores de referência; e, ainda, é a única atualização da comunicação intersubjetiva. São essas propriedades que permitem que a língua se apresente como sistema produtor de sentidos, uma organização semiótica por excelência e o único sistema capaz de comportar simultaneamente a significância dos signos e da enunciação.

Pelo caráter essencialmente social de uma língua, Benveniste a concebe como um consenso coletivo, como produto nascido de um convívio social e cultural. Conforme Gomes (2003, p. 46) “A perspectiva *social* de Benveniste é da própria natureza do homem: *é um homem falando com outro homem que encontramos no mundo [...]*”

A língua, ainda, é descrita por Benveniste como de dupla significância: o semiótico e o semântico. O semiótico é independente da referência e remete ao modo de significar que é próprio do signo, unidade do sistema, pura alteridade em relação a qualquer outro elemento. O semântico toma a seu encargo o conjunto de referentes, é produzido pelo discurso e se identifica com a enunciação. É esta nova abordagem da língua que faz com que Benveniste ultrapasse Saussure, que vê a noção de signo como princípio único, do qual depende simultaneamente a estrutura e o funcionamento da língua. Citando Kuhn (2009), lembramos que ela observa que Benveniste usa os termos língua e linguagem, ora com significação própria, ora não.

Entendendo que as Linguagens Documentárias, tal como os SRIs, precisam ser compreendidos a partir da noção de estrutura, selecionamos o texto *'estrutura' em Linguística* para fundamentar esse entendimento.

2.1.1.4 A noção de estrutura em Linguística da Enunciação

No texto *'estrutura' em Linguística* Benveniste se propõe a explicar a noção de estrutura, e a esclarecer o sentido que esse termo vem assumindo entre os linguistas que, de início, o tomaram em uma acepção precisa. A noção de estrutura surgiu como objeto de estudo com um pequeno grupo de linguistas, um pouco antes de 1930, por oposição “[...] a concepção exclusivamente histórica da língua, contra uma lingüística que dissociava a língua em elementos isolados e se ocupava em seguir-lhes as transformações” (BENVENISTE, 1988c, p. 98).

Este termo tornou-se relevante em Linguística brasileira nas décadas de 1960 e 1970, revelando-se como essencial para qualificar a linguística. Aliás, não exatamente o termo estrutura, mas o termo estrutural e, logo depois, os termos estruturalista e estruturalismo, criando-se, assim, um conjunto de designações que foram sendo tomadas emprestadas por outras disciplinas, adaptando-as aos seus valores.

Há, entre os linguistas, concordância em afirmarem que esse entendimento tem origem nos ensinamentos de Saussure, que é considerado o precursor do estruturalismo moderno. O termo estrutura, no entanto, jamais foi empregado por Saussure; ele não aparece em sua obra *Cours de linguistique générale*. Para este linguista, o fundamental era a noção de sistema e, fundamentada nessa ideia, a de que a língua forma um sistema.

Sobre as formulações feitas no *Cours* a esse respeito, Benveniste destaca as seguintes ideias:

[...] la langue est un système qui ne connaît que son ordre propre⁷ (SAUSSURE, 1968, p.43);

[...] la langue, système de signes arbitraires⁸ (SAUSSURE, 1968, p.106.);

La langue est un système dont toutes les parties peuvent et doivent être considérées dans leur solidarité synchronique⁹ (SAUSSURE, 1968, p.124);

[...] c'est une grande illusion de considérer un terme simplement comme l'union d'un certain son avec un certain concept. Le définir ainsi, ce serait l'isoler du système dont il fait partie; ce serait croire qu'on peut commencer par les termes et construire le système en faisant la somme, alors qu'au contraire c'est du tout solidaire qu'il faut partir pour obtenir par analyse les éléments qu'il renferme¹⁰ (SAUSSURE, 1968, p.157).

Este último enunciado revela o primordial da concepção estrutural, mas é importante salientar que é sempre à noção de sistema que Saussure se refere. O próprio Benveniste faz esta observação.

A noção de sistema era bastante familiar aos alunos de Saussure e aceita tanto pelos gramáticos como pelos linguistas, que mesmo antes da publicação do *Cours* já a utilizavam em seus escritos. E, sempre que enunciada, referiam a Saussure como o mestre que lhes havia ensinado que a língua é um sistema rigorosamente organizado. Acrescentando dois outros princípios

[...] igualmente saussurianos, de que a língua é forma, não substância, e de que as unidades da língua não podem definir-se a não ser pelas suas relações, indicam-se os fundamentos da doutrina que iria, alguns anos mais tarde, pôr em evidência a *estrutura* dos sistemas lingüísticos (BENVENISTE, 1988c, p. 100).

A doutrina estruturalista é inicialmente expressa por três linguistas russos, Roman Jakobson, Sergei Karcevsky e Nikolai Trubetzkoy, que designavam como seus precursores Saussure e Baudoin de Courtenay. As ideias dos linguistas russos, desde 1929, começaram a ser formuladas e divulgadas em língua francesa nas teses publicadas em Praga, no Primeiro Congresso dos filólogos eslavos. Essas teorias, que constituíam verdadeiro manifesto, inauguraram as atividades do Círculo Linguístico de Praga. É neste encontro que o termo estrutura aparece ligado à ideia de 'relação no interior do sistema'.

Todas as ideias aí colocadas deixam clara a importância desta relação – sistema e estrutura – e de que não é possível estudar o lugar de uma palavra em um sistema lexical, sem

⁷ Tradução: “[...] a língua é um sistema que conhece somente sua ordem própria” (SAUSSURE, 1972, p. 31).

⁸ Tradução: “[...] a língua, sistema de signos arbitrários [...]” (SAUSSURE, 1972, p. 86).

⁹ Tradução: “A língua é um sistema do qual todas as partes podem e devem ser consideradas em sua solidariedade sincrônica” (SAUSSURE, 1972, p. 102).

¹⁰ Tradução: “[...] é uma grande ilusão considerar um termo simplesmente como a união de certo som com um certo conceito. Defini-lo assim seria isolá-lo do sistema do qual faz parte; seria acreditar que é possível começar pelos termos e construir o sistema fazendo a soma deles, quando, pelo contrário, cumpre partir da totalidade solidária para obter, por análise, os elementos que encerra” (SAUSSURE, 1972, p. 132).

antes haver estudado a estrutura deste sistema. Dentre as enunciações mais explícitas enunciadas neste encontro está aquela que diz que “[...] *estrutura* se determina como ‘estrutura de um sistema’” (BENVENISTE, 1988c, p. 101).

A propósito de uma definição da palavra estrutura, Benveniste cita Lalande, que em sua obra *Vocabulaire technique et critique de la philosophie* diz: “En un sens spécial et nouveau, s’emploie au contraire pour désigner, par opposition à une simple combinaison d’éléments, un tout formé de phénomènes solidaires, tels que chacun dépend des autres et ne peut être ce qu’il est que dans et par sa relation avec eux”¹¹ (LALANDE, 1997, p. 1031-1032). E sobre a noção de forma, nesta mesma obra, Benveniste faz referência à seguinte citação:

Elle consiste à considérer les phénomènes non plus comme une somme d’éléments qu’il s’agit avant tout d’isoler, d’analyser, de disséquer, mais comme des ensembles (*Zusammenhänge*) constituant des unités autonomes, manifestant une solidarité interne, et ayant des lois propres. Il s’ensuit que la manière d’être de chaque élément dépend de la structure de l’ensemble et des lois qui le régissent¹² (LALANDE, 1997, p. 372-373).

Na tentativa de construir uma definição mínima de estrutura, Benveniste conclui o texto ‘*estrutura*’ em *Linguística* com o seguinte enunciado

O princípio fundamental é que a língua constitui um sistema do qual todas as partes são unidas por uma relação de solidariedade e de dependência. Esse sistema organiza unidades, que são os signos articulados, que se diferenciam e se delimitam mutuamente. A doutrina estruturalista ensina a predominância do sistema sobre os elementos, visa a destacar a estrutura do sistema através das relações dos elementos, tanto na cadeia falada como nos paradigmas formais, e mostra o caráter orgânico das mudanças às quais a língua é submetida (BENVENISTE, 1988c, p. 104).

Sobre a noção de sistema, Benveniste declara no aparelho formal da enunciação que na língua está presente o princípio de continuidade e que ela é percebida como uma estrutura, na qual elementos – índices de pessoa, espaço e tempo – são colocados à disposição do sujeito que faz uso dela.

¹¹ Tradução da autora: Em um sentido especial e novo, é usado para designar, em oposição a uma simples combinação de elementos, um todo formado de fenômenos solidários, de modo que cada um depende dos outros e não pode ser o que é a não ser pela sua relação com eles. Nas palavras de Benveniste (1988c, p. 103): “[...] para designar por oposição a uma simples combinação de elementos, um todo formado de fenômenos solidários tais que cada um depende dos outros e não pode ser o que é a não ser pela sua relação com eles”.

¹² Tradução da autora: Ela consiste em considerar os fenômenos não mais como uma soma de elementos que é, trata-se antes de tudo de isolar, analisar, dissecar, mas como conjuntos (*Zusammenhänge*), constituindo unidades autônomas, que manifestam uma solidariedade interna, e tem suas próprias leis. Decorre que o modo de ser de cada elemento depende da estrutura do conjunto e das leis que o governam. Nas palavras de Benveniste (1988c, p. 103): “Essa concepção consiste em considerar os fenômenos não mais como uma soma de elementos que é preciso antes de tudo isolar, analisar, dissecar, mas como conjuntos (*Zusammenhänge*) que constituem unidades autônomas, que manifestam uma solidariedade interna e têm leis próprias. Daí decorre que a maneira de ser de cada elemento depende da estrutura do conjunto e das leis que o regem”.

Concluindo esta subseção é preciso registrar que a distinção entre sistema e estrutura é relevante para o linguista. No conjunto de sua obra, há textos que criticam a confusão usual na época sobre as noções de estrutura e sistema. Críticas neste sentido estão presentes nos textos *'estrutura' em linguística* e *Estruturalismo e lingüística*. A esse respeito, esclarecemos que em seus trabalhos sobre enunciação é o conceito de sistema que mais interessa ao linguista, mas nos trabalhos que tratam de comparativismo, é forte a influência da noção de estrutura.

Buscando a compreensão do conceito de sentido, noção imprescindível nas Linguagens Documentárias, desenvolveremos na subseção seguinte uma reflexão a partir de dois textos de Benveniste: *Os níveis da análise linguística* e *A forma e o sentido na linguagem*. Além das noções forma e sentido, a leitura destes textos nos leva a compreensão, ainda, de outras três: língua, linguagem e referência.

2.1.1.5 A forma e o sentido na linguagem

No texto *Os níveis da análise linguística*, Benveniste discute a noção do sentido a partir da análise da língua enquanto forma. Ele estabelece o discurso como um novo domínio de análise, contrariando a posição assumida por diferentes linguistas, que até então desenvolviam uma análise da língua centrada na forma. Esta exposição tem continuidade no texto *A forma e o sentido na linguagem*. Neste texto o linguista afirma que a significação não é algo acrescentado à língua, mas é constitutivo da língua. E, ainda, faz distinção entre signo e frase, assegurando que possuem funções distintas. Segundo o linguista o signo é uma unidade semiótica constituído de um significante e um significado; a ele devem ser atribuídos forma e sentido. O signo tem sua significação definida pelo uso da língua, isto é, pela comunidade de fala. A frase, por sua vez, diferencia-se do signo por ter função comunicativa. Com o signo tem-se o significado na relação paradigmática; com a frase passa-se ao domínio do discurso, e é na referência à situação de discurso que se alcança o sentido. Em ambos os textos percebe-se uma separação entre o sistema semiótico, o dos signos, e o sistema semântico, o da língua em uso.

Benveniste, no texto *Os níveis da análise linguística*, tomando um enunciado como objeto de análise, e levando em consideração a relação entre os seus elementos em termos de segmentação e substituição, propõe o sentido como a única maneira de determinar que certa unidade é válida. Para ele “O *sentido* é de fato a condição fundamental que todas as unidades de todos os níveis devem preencher para obter *status* lingüístico” (BENVENISTE, 1988d, p. 130). Sobre a noção de nível, o linguista faz a seguinte afirmação:

A noção de *nível* parece-nos essencial na determinação do procedimento da análise. Só ela é própria para fazer justiça à natureza *articulada* da linguagem e ao caráter *discreto* dos seus elementos; só ela pode fazer-nos reconhecer, na complexidade das formas, a arquitetura singular das partes e do todo. O domínio no qual a estudaremos é o da língua como sistema orgânico de signos lingüísticos (BENVENISTE, 1988d, p. 127).

Ao recolocar o sentido como condição para a análise lingüística, o linguista se diferencia da maioria dos teóricos da época, que entendiam a forma como o único aspecto a ser considerado.

A relação forma e sentido é, segundo o linguista francês, fundamental na análise lingüística. “Forma e sentido devem definir-se um pelo outro e devem articular-se juntos em toda a extensão da língua. As suas relações parecem-nos implicadas na própria estrutura dos níveis e na das funções que a elas correspondem, que aqui designamos *constituente* e *integrante*” (BENVENISTE, 1988d, p. 135). Ainda sobre forma e sentido Benveniste (1988d, p. 135-136) formula as seguintes definições: “A *forma* de uma unidade lingüística define-se como a sua capacidade de dissociar-se em constituintes de nível inferior. O *sentido* de uma unidade lingüística define-se como a sua capacidade de integrar uma unidade de nível superior”.

Desse modo, forma e sentido são entendidos como propriedades inseparáveis e necessárias ao funcionamento da língua. “As suas relações mútuas revelam-se na estrutura dos níveis lingüísticos, percorridos pelas operações descendentes e ascendentes da análise e graças à natureza articulada da linguagem” (BENVENISTE, 1988d, p. 136). Destas noções decorrem as seguintes formulações: sentido e significação, dois universos lingüísticos e duas lingüísticas diferentes.

A noção de sentido implica duas instâncias: 1) uma instância em que o sentido é inerente ao sistema lingüístico e às suas partes, ou seja, em que um elemento da língua é entendido como uma unidade distintiva, opositiva, delimitada por outras unidades, e identificável para os locutores nativos ou falantes de uma língua; e, 2) uma instância em que cada enunciado e cada termo do enunciado tem um *referendum*, e que determiná-lo constitui-se em uma tarefa difícil. Esse entendimento parte do princípio de que a linguagem se refere ao mundo dos objetos, a partir de enunciados completos, seja sob a forma de frases que se relacionam com situações concretas e específicas, ou sob a forma de unidades inferiores que se relacionam com objetos gerais ou particulares.

A frase tem natureza distinta dos outros elementos da língua; é um signo e também é constituída por signos. Ela diferencia-se dos fonemas, dos morfemas e dos lexemas (palavras), que existem em número finito. Ela não tem distribuição nem emprego.

Um inventário dos empregos de uma palavra poderia não acabar; um inventário dos empregos de uma frase não poderia nem mesmo começar.

A frase, criação indefinida, variedade sem limite, é a própria vida da linguagem em ação. Concluímos que se deixa com a frase o domínio da língua como sistema de signos e se entra num outro universo, o da língua como instrumento de comunicação, cuja expressão é o discurso (BENVENISTE, 1988d, p. 139).

Sobre a frase, o linguista acrescenta:

A frase pertence bem ao discurso. É por isso mesmo que se pode defini-la: a frase é a unidade do discurso. [...]

A frase é uma unidade, na medida em que é um segmento de discurso, e não na medida em que poderia ser distintiva com relação a outras unidades do mesmo nível – o que ela não é [...]. É, porém, uma unidade completa, que traz ao mesmo tempo sentido e referência: sentido porque é enformada de significação, e referência porque se refere a uma determinada situação. Os que se comunicam têm justamente isto em comum, uma certa referência de situação, sem a qual a comunicação como tal não se opera, sendo inteligível o ‘sentido’ mas permanecendo desconhecida a ‘referência’ (BENVENISTE, 1988d, p. 139-140).

Pela leitura da obra de Benveniste, fica claro que a língua precisa ser vista sob dois universos diferentes, ainda que envolvam a mesma realidade, e sob duas linguísticas diferentes, ainda que se cruzem a todo instante.

No texto *A forma e o sentido na linguagem* Benveniste (1989c) discute os problemas de linguagem a partir das noções forma e sentido. O propósito deste texto é situar e organizar estas noções na linguagem, no âmbito da linguística e fora de qualquer pressuposto filosófico. O domínio dessa discussão é a linguagem comum, mas tudo que se poderá esclarecer sobre esta linguagem será de proveito também para a compreensão da linguagem poética, diretamente ou não.

Em uma primeira aproximação entre forma e sentido, o linguista faz a seguinte colocação: a forma é o arranjo formal dos elementos linguísticos ao nível linguístico relevante; o sentido é o conjunto de procedimentos de comunicação identicamente compreendidos por um conjunto de locutores.

A oposição forma e sentido é reinterpretada por Benveniste no quadro do funcionamento da língua, com o propósito de abordar a questão da significação. Conforme afirma o linguista, o próprio da linguagem é antes de tudo significar; mas ‘o que é significação?’ se questiona. Em resposta a esta questão, ele faz a seguinte observação: trata-se de um dos grandes problemas da linguagem e diz respeito às diversas ciências; e é exatamente por isso, que não é discutida com seriedade por nenhuma delas. Os únicos que se dedicam verdadeiramente à compreensão do que é significação são os lógicos, dentre eles, os da escola de Carnap e de Quine. Mas faz a ressalva de que os lógicos descartaram toda tentativa de

elaborar uma definição direta de significação. “[...] eles substituíram a análise da significação pelo critério objetivo da aceitabilidade, experimentada por meio de testes, conforme os predicados sejam aceitos ou não pelo locutor” (BENVENISTE, 1989c, p. 223).

Benveniste desenvolve seu raciocínio sobre significação de modo diferente do que fazem os lógicos, uma vez que ele busca fundamentação teórica na própria linguística. Ele toma como verdadeiro o fato que “[...] a linguagem é a atividade significante por excelência, a imagem mesma do que pode ser a significação; todo e qualquer modelo significativo que possamos construir será aceito na medida em que se parecer em tal ou tal de seus aspectos àquele da linguagem” (BENVENISTE, 1989c, p. 223).

Dizer que a linguagem significa é aceitar que a significação é de sua própria natureza e não admitir que algo lhe seja dado por acréscimo. Mas Benveniste ressalta que é próprio da linguagem, também, o caráter de se realizar por meios vocais, de consistir num conjunto de sons que se organizam em palavras dotadas de sentido. A título de uma primeira aproximação, Benveniste retoma Saussure para dizer que a língua é um sistema de signos. A noção de signo é introduzida por Saussure no quadro da oposição que ele estabelece entre língua e fala (*langue* e *parole*). Benveniste propõe-se a ultrapassar Saussure na análise da língua como sistema significante; sua apreciação é feita a partir das noções de forma e sentido.

A reflexão na perspectiva benvenistiana, parte do entendimento que “Dizer que a língua é feita de signos é dizer antes de tudo que *o signo é a unidade semiótica*. Esta proposição [...] contém uma dupla relação que é necessário explicitar: a noção de signo enquanto unidade e a noção de signo como dependente da ordem semiótica” (BENVENISTE, 1989c, p. 224).

Para situar a noção de signo no âmbito da linguagem, Benveniste mostra que a linguagem tem por característica o descontínuo e o dessemelhante; por tal razão, ela não se deixa dividir, mas decompor. Os signos enquanto unidades da linguagem se diferenciam um do outro, e se agrupam para formar novas unidades de nível superior num processo contínuo. O limite inferior desta unidade, que é o signo, é o da significação. O signo enquanto unidade semiótica da língua pode ser assim entendida: “[...] entidade livre, mínima em sua ordem, não decomponível em uma unidade inferior que seja ela mesma um signo livre” (BENVENISTE, 1989c, p. 225). Entendendo o signo como uma unidade semiótica bilateral por natureza, Benveniste propõe que ele seja caracterizado pelo duplo ponto de vista da forma e do sentido e como significante e significado.

O significante não é apenas uma sequência de sons exigida pela língua; ele é a forma sonora que condiciona e determina o significado e o aspecto formal do signo. Nessa

perspectiva, o signo se distingue do fonema; conforme esclarece Benveniste, o signo não se decompõe imediatamente em fonemas e, também, uma sequência de fonemas não compõe imediatamente um signo.

Na esfera do significado, afirma o linguista francês, “[...] o signo se define como a unidade semiótica; ele é dotado de significação na comunidade daqueles que fazem uso de uma língua, e a totalidade destes signos forma a totalidade da língua” (BENVENISTE, 1989c, p. 227). Sob o olhar da semiologia “Para que um signo exista, é suficiente e necessário que ele seja aceito e que se relacione de uma maneira ou de outra com os demais signos. A entidade considerada significa? [...] ‘Chapéu’ existe? Sim. ‘Chaméu’ existe? Não” (BENVENISTE, 1989c, p. 227).

Ainda sobre o significado, Benveniste acrescenta:

No plano do significado, o critério é: isto significa ou não? Significar é ter um sentido, nada mais. E este *sim* ou *não* só pode ser pronunciado por aqueles que manuseiam a língua, aqueles para os quais esta língua é *a língua* e nada mais. [...] É no uso da língua que um signo tem existência; o que não é usado não é signo; e fora do uso o signo não existe. [...].

Enunciemos então este princípio: tudo o que é do domínio do semiótico tem por critério necessário e suficiente que se possa identificá-lo no interior e no uso da língua. Cada signo entra numa rede de relações e de oposições com os outros signos que o definem, que o delimitam no interior da língua. Quem diz ‘semiótico’ diz ‘intralingüístico’. Cada signo tem de próprio o que o distingue dos outros signos. Ser distintivo e ser significativo é a mesma coisa (BENVENISTE, 1989c, p. 227-228).

Das reflexões feitas até então sobre o significado do ponto de vista semiótico, decorre três consequências: 1) a relação do signo com as coisas denotadas e a relação entre língua e o mundo não é o objeto de estudo; 2) o signo tem valor genérico e conceitual, não admitindo significado particular ou ocasional e excluindo, assim, tudo o que é individual; e, 3) as oposições semióticas são de tipo binário; a binaridade parece ser a característica semiológica por excelência, na língua. Nesse contexto, afirma Benveniste (1989c), os signos se dispõem sempre e somente em relação paradigmática.

Após caracterizar a estrutura do signo e suas relações, Benveniste passa para a análise da frase. Qual sua função comunicativa na língua? pergunta o linguista. Afinal, é por frases que nos comunicamos. Conforme análise feita pelo linguista, o signo e a frase constituem dois mundos distintos e, como tal, exigem descrições diferentes.

Com esse entendimento, Benveniste instaura na língua uma divisão diferente daquela proposta por Saussure, a da língua e fala. Assim, o linguista propõe para a língua duas espécies, o da forma e o do sentido, e dois domínios, o semiótico e o semântico, ressaltando, no entanto, que ainda que sejam os mesmos elementos que se encontrem em uma e outra

parte, são dotados de estatutos diferentes. Na perspectiva benvenistiana, portanto, há duas maneiras de ser língua: no sentido e na forma, ou seja, a língua como semiótica e a língua como semântica. E, ainda, há duas modalidades fundamentais da função linguística: aquela de significar para a semiótica e a de comunicar para a semântica.

Com a noção de semântica, passamos ao domínio da língua em emprego e em ação, transmitindo a informação e comunicando a experiência. Saímos do significado do signo para o da frase, que é a expressão semântica, por excelência. Com a frase nos deslocamos para a produção do discurso, para aquilo que o locutor quer dizer.

Assim, do semiótico ao semântico, há uma mudança radical de perspectiva.

A semiótica se caracteriza como uma propriedade da língua; a semântica resulta de uma atividade do locutor que coloca a língua em ação. [...] Com o signo tem-se a realidade intrínseca da língua; com a frase liga-se às coisas fora da língua; e enquanto o signo tem por integrante o significado, que lhe é inerente, o sentido da frase implica referência à situação de discurso e à atitude do locutor (BENVENISTE, 1989c, p. 229-230).

De modo diferente do semiótico, em que o sentido se define por uma relação de paradigma, na acepção semântica ele se realiza no sintagma; de um lado a substituição e de outro a conexão. Enquanto o signo é entendido como a unidade semiótica, a palavra é a unidade semântica, a unidade mínima da mensagem e a unidade necessária da codificação do pensamento. O sentido da frase é a ideia que ela exprime.

[...] este sentido se realiza formalmente na língua pela escolha, pelo agenciamento das 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’ [...] Toda forma verbal, sem exceção, em qualquer idioma que seja, está sempre ligada a um certo presente, portanto a um conjunto cada vez único de circunstâncias, que a língua enuncia numa morfologia específica (BENVENISTE, 1989c, p. 230).

Na tentativa de esclarecer o processo pelo qual o sentido se constitui, na perspectiva semântica, Benveniste faz as seguintes considerações. O sentido da frase é sua ideia e o das palavras, seu emprego. A ideia é sempre particular. O sentido de uma frase é diferente do sentido das palavras que a compõem. Ainda que se compreenda o sentido de cada uma das palavras que formam uma frase, é bem possível que não se compreenda o sentido que resulta da junção das palavras.

E, ainda, o sentido da frase se constitui no momento da enunciação quando o locutor exprime sua ideia, o que significa dizer que o sentido da frase se refere à ideia que o locutor tem como propósito exprimir. Ao dizer que o sentido da frase surge do somatório dos significados das palavras, uma vez que a frase é o conjunto de palavras sintaticamente

organizadas, Benveniste possibilita o seguinte entendimento. Sendo a palavra a unidade semântica, ela própria guarda suas referências e significações próprias. Entretanto, quando há o agenciamento de palavras para a construção da frase, a referência passa a ser o próprio ato de enunciação; dele decorre o sentido da frase. É preciso estar atento, no entanto, para a possibilidade das palavras contraírem valores contrários aos que recebem nos empregos usuais.

Desse modo, ao caracterizar a ideia como uma experiência sempre única e expressa por um locutor que agencia palavras e as emprega com um sentido sempre único, Benveniste insere a noção de referência como essencial no processo de construção do sentido. O linguista observa a confusão usualmente feita entre essas duas noções – sentido e referência – e propõe que se estabeleça uma distinção entre elas. Com esse propósito ele desenvolve a seguinte argumentação.

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. [...] A frase é então cada vez um acontecimento diferente; ela não existe senão no instante em que é proferida e se apaga neste instante; é um acontecimento que desaparece. [...] O sentido de uma palavra consistirá na sua capacidade de ser integrante de um sintagma particular e de preencher uma função proposicional. O que se chama de polissemia não é senão a soma institucionalizada, se pudermos falar assim, destes valores contextuais, sempre instantâneos, aptos a se enriquecer e a desaparecer, em resumo, sem permanência, sem valor constante (BENVENISTE, 1989c, p. 231-232).

Revela-se, assim, o estatuto diferente de uma mesma entidade lexical, seja ela considerada um signo ou uma palavra. Dessa constatação, decorrem duas consequências opostas: uma grande variedade de expressões para um locutor expressar sua ideia (a mesma ideia), e indo além das palavras, a restrição de leis impostas ao locutor no agenciamento de palavras, para expressar sua ideia. “[...] há aqui necessariamente uma mistura sutil de liberdade no enunciado da idéia e de restrição na forma deste enunciado, que é a condição de toda a atualização da linguagem” (BENVENISTE, 1989c, p. 232).

Isso é de certo modo consequência da adaptação recíproca, gerada pelo fato das palavras contraírem valores que não possuem em si mesmas, e que podem até mesmo serem contraditórios com os valores que elas possuem em outros lugares. Nesse sentido, pode ser observada a coligação de conceitos logicamente opostos, que são capazes até mesmo de se reforçarem, ao se unirem. É o caso, por exemplo, de ‘eu tenho perdido’, ‘ele vai vir’ e ‘ele deve pagar’.

Assim, o ‘sentido’ da frase está na totalidade da idéia percebida por uma compreensão global; a ‘forma’ se obtém pela dissociação analítica do

enunciado processada até as unidades semânticas, as palavras. Além disso, as unidades não podem mais ser dissociadas sem deixar de preencher sua função. Esta é a articulação semântica (BENVENISTE, 1989c, p. 232).

O pensamento benvenistiano desenvolvido com o propósito de estabelecer a diferença entre o semiótico e o semântico deixa claro o seguinte. O sentido a transmitir – a mensagem – é definido, delimitado e organizado por meio de palavras. O sentido das palavras, no entanto, é determinado na sua relação com o contexto de situação. As palavras são instrumentos da expressão semântica e também, materialmente, os signos do repertório semiótico, ficando assim estabelecida a relação entre sentido e forma. Os signos em si mesmos são conceituais, genéricos e não circunstanciais. Quando utilizados como palavras expressam noções particulares, específicas e circunstanciais, nas acepções contingentes do discurso.

Semiótico e semântico são dois sistemas que se superpõem na língua do modo como a empregamos. O sistema semiótico está na base e diz respeito à organização de signos. Cada um desses signos, segundo o critério da significação tem uma denotação conceitual e inclui, em uma sub-unidade, o conjunto de seus substitutos paradigmáticos. Sobre a base do sistema semiótico, está o semântico. Aqui, a língua-discurso constrói sua própria semântica, com uma significação intencionada, produzida pela sintagmatização das palavras. Benveniste ressalta o fato de que pela semantização das palavras, cada uma delas retém apenas uma pequena parte do valor que tem enquanto signo.

Benveniste conclui sua discussão sobre forma e sentido dizendo que esse duplo sistema – semiótico e semântico – está constantemente em ação na língua e que funciona de modo muito sutil e rápido. Mas ressalta que o fundamental é que o poder significante da língua é sempre presente e é anterior ao dizer qualquer coisa. Na dicotomia proposta por ele, a frase não tem espaço no domínio do semiótico; ela é da ordem do semântico, uma vez que ela não pertence ao domínio do signo. A frase é um enunciado de caráter necessariamente semântico. A partir do momento em que a língua é considerada como ação, como realização, ela supõe necessariamente um locutor e a situação deste locutor no mundo. Estas relações são dadas no semântico.

Associadas à noção de locutor estão as de subjetividade e de intersubjetividade. Por ser a subjetividade uma propriedade da língua e a capacidade do locutor para se propor como sujeito, conforme afirma Benveniste, o estudo desta noção se torna pertinente nesta pesquisa, especialmente quando se busca a compreensão das Linguagens Documentárias e a formulação de uma proposta que caracterize a relação que se estabelece entre SRIs e usuários. Com esse propósito, abordaremos na subseção seguinte essas duas noções – subjetividade e intersubjetividade –, ressaltando que a noção de intersubjetividade está associada à noção de

subjetividade, mas não em termos de precedência, já que a subjetividade emerge de um processo de intersubjetividade.

2.1.1.6 Da (inter)subjetividade na linguagem

Na teoria de Benveniste, falar de subjetividade é falar de linguagem, uma vez que não atingimos o homem (sujeito) separado da linguagem. É na linguagem e pela linguagem que o homem se constitui como sujeito. A revelação da subjetividade na linguagem é entendida a partir da categoria de pessoa. É a categoria de pessoa que permite a Benveniste propor a subjetividade no sistema linguístico.

Nesta subseção, tratamos a noção de (inter)subjetividade na linguagem considerando a distinção entre semiótico (signo) e semântico (palavra/frase), noção já discutida na subseção anterior com a leitura dos textos *Os níveis da análise linguística* e *A forma e o sentido na linguagem*. O entendimento dos conceitos de intersubjetividade e subjetividade são facilitados com a compreensão desta distinção. Relacionadas a estas noções estão as de língua, linguagem, referência e categoria de pessoa.

Analisaremos a distinção semiótico e semântico pela leitura do texto *Semiologia da língua*, e a noção categoria de pessoa com a leitura dos textos *Estrutura das relações de pessoa no verbo* e *A natureza dos pronomes*. Dando continuidade a essa reflexão e visando alargar a compreensão dos conceitos de intersubjetividade e subjetividade, efetuamos a leitura do texto *Da subjetividade na linguagem*. As noções de língua, linguagem e referência são tratadas nestes textos e, deste modo, são abordadas nesta subseção.

O texto *Semiologia da língua* apresenta a perspectiva semântica desenvolvida por Benveniste; ele estabelece oposição entre dois níveis de significação: o semiótico e o semântico. No nível semiótico a unidade de significação é o signo, mas no nível semântico, a unidade de significação é a palavra. Na tentativa de responder à pergunta ‘o que é significação’, o linguista recorre à noção de signo, articulado à de significação no estudo da língua. Fica evidenciado, assim, a ideia de ‘língua como sistema de signos’ e de ‘signo como unidade semiótica’.

Ao estabelecer oposição entre o semiótico e o semântico, Benveniste propõe dois modos de significação. O primeiro modo de significação, o semiótico, corresponde ao nível intralinguístico, em que cada signo é distintivo, significativo em relação aos demais, possuidor de valores opositivos e genéricos, e disposto em uma organização paradigmática. O segundo modo de significação, o semântico, resulta da atitude do locutor quando coloca a

língua em ação. Aqui, o linguista utiliza o critério da comunicação para definir a palavra como a unidade de operações sintagmáticas que se realizam no nível da frase.

É importante ressaltar que o tratamento dado à referência nesses dois níveis é diferente. Na perspectiva semiótica ela está ausente, mas sob a ótica da semântica ela é definidora de sentido, uma vez que ela se caracteriza pela relação estabelecida entre as ideias expressas sintagmaticamente na frase e a situação de discurso. Nesse contexto, a frase adquire o sentido de discurso ou de língua em ação.

Ao propor um nível de significado que engloba referência aos interlocutores, Benveniste

[...] apresenta um modelo de análise da enunciação em que os interlocutores referem e co-referem na atribuição de sentido às palavras. Essa distinção [semiótico e semântico] possibilita o entendimento da categoria de pessoa e dos conceitos de intersubjetividade e de enunciação, básicos em sua teoria (FLORES; TEIXEIRA, 2005, p. 32).

A compreensão da noção de intersubjetividade em Benveniste é facilitada quando aliada ao estudo da estrutura do sistema pronominal e da enunciação.

No texto *Estrutura das relações de pessoa no verbo*, texto dedicado ao entendimento da categoria de pessoa, Benveniste desenvolve uma teoria linguística da pessoa verbal. Segundo o linguista, ela deve ser construída com base na estrutura opositiva entre elas, na medida em que as duas primeiras pessoas, ‘eu e tu’, distinguem-se da terceira, ‘ele’, a partir de duas correlações: a da personalidade (ou pessoalidade) e a de subjetividade.

Benveniste desenvolve seu pensamento, afirmando que

Em todas as línguas que possuem um verbo, classificam-se as formas da conjugação segundo a sua referência à pessoa, constituindo a enumeração das pessoas propriamente a conjugação: distinguem-se três no singular, no plural e eventualmente no dual (BENVENISTE, 1988e, p. 247).

Ele observa que esta classificação é herdada da gramática grega. Mas ressalta o fato de que

Essas denominações [eu, tu e ele] não nos informam nem sobre o conteúdo que ela implica nem sobre as relações que reúnem as diferentes pessoas. É preciso, portanto, procurar saber como cada pessoa se opõe ao conjunto das outras e sobre que princípio se funda a sua oposição, uma vez que não podemos atingi-la a não ser pelo que as diferencia (BENVENISTE, 1988e, p. 248).

Prosseguindo em seu pensamento ele faz o seguinte questionamento: “[...] se a categoria de pessoa é verdadeiramente necessária e congenial ao verbo ou se constitui somente uma modalidade possível, o mais frequentemente realizada, mas não indispensável, como o são afinal muitas categorias verbais” (BENVENISTE, 1988e, p. 248). Com o propósito de refletir sobre esta questão, ele apresenta alguns exemplos de línguas nas quais a

expressão da pessoa é suscetível de faltar ao verbo; o linguista conclui seu pensamento afirmando que todas as línguas possuem pronomes pessoais.

Em suma, não parece que se conheça uma língua dotada de um verbo no qual as distinções de pessoa não se marquem de uma ou de outra maneira nas formas verbais. Podemos concluir, pois, que a categoria da pessoa pertence realmente às noções fundamentais e necessárias do verbo (BENVENISTE, 1988e, p. 249-250).

Com o propósito de discutir a pessoa verbal, Benveniste afirma que “Uma teoria lingüística da pessoa verbal só pode constituir-se sobre a base das oposições que diferenciam as pessoas, e se resumirá inteiramente na estrutura dessas oposições” (BENVENISTE, 1988e, p. 250). Para fundamentar seu pensamento ele retoma os gramáticos árabes. Segundo eles, a primeira pessoa é ‘aquele que fala’, a segunda pessoa ‘é aquele a quem nos dirigimos’ e a terceira pessoa é ‘aquele que está ausente’. O linguista entende que nessas denominações fica clara a verdadeira relação que ocorre entre as pessoas, uma vez que evidencia a diferença que existe entre a terceira pessoa e as duas primeiras.

Nas palavras de Benveniste, temos o seguinte esclarecimento a respeito desta questão:

Nas duas primeiras pessoas, há ao mesmo tempo uma pessoa implicada e um discurso sobre essa pessoa. *Eu* designa aquele que fala e implica ao mesmo tempo um enunciado sobre o ‘eu’: dizendo *eu* não posso deixar de falar de mim. Na segunda pessoa, ‘tu’ é necessariamente designado por *eu* e não pode ser pensado fora de uma situação proposta a partir do ‘eu’; e, ao mesmo tempo, *eu* enuncia algo como um predicado de ‘tu’. Da terceira pessoa, porém, um predicado é bem enunciado somente fora do ‘eu-tu’; essa forma é assim excetuada da relação pela qual ‘eu’ e ‘tu’ se especificam. Daí, ser questionável a legitimidade dessa forma como ‘pessoa’ (BENVENISTE, 1988e, p. 250).

Ao designar aquele que fala, o ‘eu’ instaura o ‘tu’, que é aquele a quem nos dirigimos, e estabelece uma relação com o ‘tu’, excluindo o ‘ele’ desta relação. O ‘ele’ se inscreve nesta relação como um enunciado sobre algo que está fora da instância ‘eu e tu’. Com base nessas reflexões, Benveniste conclui que ‘eu e tu’ são categoria de pessoa e ‘ele’ categoria de não-pessoa, uma vez que o ‘ele’ não é instituído como pessoa no momento em que ‘eu diz eu’.

A relação ‘eu e tu’ é marcada por duas características: a unicidade e a inversibilidade da categoria de pessoa. O ‘eu’ que enuncia e o ‘tu’ ao qual o ‘eu’ se dirige são sempre únicos. O que ‘eu’ define como ‘tu’ pode vir a ser ‘eu’, e quando isso ocorre, o ‘eu’ torna-se ‘tu’. O ‘ele’, entretanto, pode ser uma infinidade de sujeitos, ou nenhum. O ‘ele’, a terceira pessoa, é a forma verbal que tem por função exprimir a não-pessoa; ele não designa especificamente nada nem ninguém; pode vir a ser qualquer sujeito ou nenhum, e esse sujeito, expresso ou não, jamais é proposto como pessoa. O ‘ele’ tem um comportamento diferenciado em diversas línguas, na relação com a primeira e a segunda pessoa. O ‘ele’ é a única pessoa verbal que

admite o verdadeiro plural, uma vez que nem ‘nós’, nem ‘vós’ pode ser entendido como um ‘eu’ ou um ‘tu’ quantificado ou multiplicado. Segundo o linguista o ‘ele’ é a única possibilidade pela qual uma coisa é predicada verbalmente.

Da oposição entre as duas primeiras pessoas do verbo e a terceira, decorre a correlação de personalidade ou de pessoalidade. ‘Eu e tu’ possuem a marca de pessoa e ‘ele’ a marca da não-pessoa. Mas ainda que ‘eu e tu’ sejam caracterizados pela marca de pessoa, não há dúvida que há uma oposição entre elas, no interior da categoria que constituem.

Benveniste assim se expressa sobre esse tema:

Ao par *eu/tu* pertence particularmente uma correlação especial, a que chamaremos, na falta de expressão melhor, *correlação de subjetividade*. O que diferencia ‘eu’ de ‘tu’ é, em primeiro lugar, o fato de ser, no caso de ‘eu’, *interior* ao enunciado e *exterior* a ‘tu’, mas exterior de maneira que não suprime a realidade humana do diálogo; [...] além disso, ‘eu’ é sempre transcendente com relação a ‘tu’. [...] Essas qualidades de interioridade e de transcendência pertencem particularmente ao ‘eu’ e se invertem em ‘tu’. Poder-se-á, então, definir o *tu* como a pessoa não-subjetiva, em face da pessoa subjetiva que *eu* representa; e essas duas ‘pessoas’ se oporão juntas à forma de ‘não-pessoa’ (= ‘ele’) (BENVENISTE, 1988e, p. 255)

Correlação de subjetividade é um conceito fundamental e base da teoria de Benveniste; ela procede dessa relação ‘eu e tu’ – pessoa subjetiva e pessoa não-subjetiva –. Do conceito de subjetividade decorrem algumas noções, tais como: língua, linguagem, signo, palavra, referência subjetiva e referência objetiva.

É fundamentado nos pronomes que Benveniste procura esclarecer a diferença entre língua e língua em uso, entendendo por língua uma estrutura que reúne signos, que se distinguem pelas relações de oposição. Assim sendo, o signo está no âmbito da língua, mas a palavra está na esfera da língua em uso, ou linguagem, ou seja, do discurso. A palavra, ao contrário do signo, é consequência da atualização da língua por um locutor que a agencia e a coloca em uso; ela é caracterizada por dizer respeito a instâncias de discurso únicas e discretas. Sob esta ótica, palavra e linguagem só são possíveis na condição de enunciação.

Com Benveniste a categoria de pessoa adquire um estatuto diferenciado. Segundo o linguista, não basta defini-la em termos de presença e ausência do traço de pessoalidade; é necessário concebê-la também em termos de subjetividade. Ainda que ‘eu e tu’ constituam a categoria de pessoa, somente o ‘eu’ é pessoa subjetiva, uma vez que ‘tu’ é simplesmente pessoa. Sob essa perspectiva o ‘tu’ é a pessoa não-subjetiva comparado à pessoa subjetiva que o ‘eu’ representa, e ainda, ‘eu e tu’ se oporão a ‘ele’, a forma da não-pessoa.

O texto *A natureza dos pronomes* é fundamental na construção da teoria benvenistiana; assim como o texto *Estrutura das relações de pessoa no verbo* aborda a

questão da categoria de pessoa. Nesse texto – *A natureza dos pronomes* – Benveniste ressalta que a distinção entre categoria de pessoa e de não-pessoa repercute além do ‘eu, tu e ele’. E, ainda, que ela não pode ser vista como pertencente a uma classe unitária, como usualmente tem sido considerada em diferentes línguas, tanto do ponto de vista formal como funcional. Trata-se de uma categoria que comporta elementos absolutamente diferentes; alguns se correlacionam à categoria de pessoa e outros a de não-pessoa.

O linguista busca explicação para esta afirmação a partir da noção de enunciação. Ele parte do entendimento de pronomes como um fato de linguagem, justificando sua convicção no princípio de que eles só são um problema de língua, por ser antes um problema de linguagem. A esse respeito, Benveniste faz a seguinte observação (1988f, p. 277): “Uns pertencem à sintaxe da língua, outros são característicos daquilo a que chamaremos as ‘instâncias do discurso’, isto é, os atos discretos e cada vez únicos pelos quais a língua é atualizada em palavra por um locutor”.

Benveniste inicia o texto com uma discussão sobre os pronomes pessoais, ressaltando que a noção de pessoa está presente apenas nas duas primeiras pessoas ‘eu e tu’, estando ausente na terceira pessoa ‘ele’. E conclui esse mesmo texto, afirmando:

Uma análise, mesmo sumária, das formas classificadas indistintamente como pronominais leva assim a reconhecer classes de natureza totalmente diferentes e, em conseqüência, a distinguir, de um lado, a língua como repertório de signos e sistema de suas combinações e, de outro, a língua como atividade manifestada nas instâncias de discurso caracterizadas como tais por índices próprios (BENVENISTE, 1988f, p. 283).

Partindo da análise do ‘eu’, o linguista deixa evidente as diferenças formais e funcionais que existem entre ‘eu’ e um nome referente a uma noção lexical. Ele menciona o fato de ‘eu’ e ‘tu’ poderem estar ausentes em textos como tratados científicos, por exemplo, mas ressalta que dificilmente se pode conceber um texto falado em que tais pronomes estejam ausentes. Essa mesma situação não se verifica com os signos da língua, que se distribuem indiferentemente entre os diferentes tipos de texto. Por si só, esta condição de emprego da língua já é importante para evidenciar o caráter distintivo que os caracterizam.

A esse respeito, Benveniste (1988f, p. 278) assim se expressa:

Cada instância de emprego de um nome refere-se a uma noção constante e ‘objetiva’, apta a permanecer virtual ou a atualizar-se num objeto singular, e que permanece sempre idêntica na representação que desperta. No entanto, as instâncias de emprego de *eu* não constituem uma classe de referência, uma vez que não há ‘objeto’ definível como *eu* ao qual se possam remeter identicamente essas instâncias. Cada *eu* tem a sua referência própria e corresponde cada vez a um ser único, proposto como tal.

Com esta reflexão, Benveniste deixa claro que a realidade a que se refere ‘eu’ e ‘tu’ é a realidade do discurso. Segundo suas palavras (1988f, p. 278-279)

Eu só pode definir-se em termos de ‘locução’, não em termos de objetos, como um signo nominal. *Eu* significa ‘a pessoa que enuncia a presente instância de discurso que contém *eu*’. Instância única por definição, e válida somente na sua unicidade. [...] É preciso, assim, sublinhar esse ponto: *eu* só pode ser identificado pela instância de discurso que o contém e somente por aí. Não tem valor a não ser na instância na qual é produzido. Paralelamente, porém, é também enquanto instância de forma *eu* que deve ser tomado; a forma *eu* só tem existência lingüística no ato de palavras que a profere. Há, pois, nesse processo uma dupla instância conjugada: instância de *eu* como referente, e instância de discurso contendo *eu*, como referido. A definição pode, então, precisar-se assim: *eu* é o ‘indivíduo que enuncia a presente instância de discurso que contém a instância lingüística *eu*’. Conseqüentemente, introduzindo-se a situação de ‘alocução’, obtém-se uma definição simétrica para *tu* como o ‘indivíduo alocutado na presente instância lingüística *tu*’. Essas definições visam *eu* e *tu* como uma categoria da linguagem e se relacionam com a sua posição na linguagem.

Essa referência à instância de discurso se caracteriza por ser constante e necessária, e por constituir o traço que une ‘eu e tu’ a uma série de indicadores de classes diferentes, como pronomes, advérbios e locuções adverbiais. Essa referência, constante e necessária à instância de discurso, é denominada pelo linguista de referência subjetiva.

Por dizer respeito ao instante da enunciação, o momento em que ‘eu’ se diz ‘eu’ se caracteriza como presente. Isso permite dizer que o tempo da enunciação é sempre o presente. Benveniste ressalta, ainda, que o instante da enunciação é delimitado por indicadores de tempo e espaço, ‘aqui e agora’, que só fazem referência ao instante do emprego da língua.

Tomando como exemplos os advérbios ‘aqui’ e ‘agora’ e colocando em evidência a sua relação com ‘eu’, o que o linguista constata é que eles delimitam a instância espacial e temporal coextensiva e simultânea da presente instância de discurso que contém ‘eu’. Assim sendo, a relação entre o indicador de pessoa, de tempo, de lugar e a presente instância de discurso é fundamental. Sendo também pertinente, o fato dessas formas pronominais remeterem à enunciação, cada vez única, e não à realidade e posições objetivas no espaço e no tempo.

Após esclarecer a diferença entre língua e linguagem em ação, Benveniste afirma que alguns pronomes pertencem à sintaxe e, conseqüentemente, à linguagem. Estes pronomes, segundo seu entendimento, são signos plenos que, ao serem colocados em uso, se atualizam em um objeto singular e se referem sempre à mesma representação. Por pertencerem à sintaxe, estão no campo da linguagem, dizem respeito à ‘ele’ e não se inserem na relação ‘eu e tu’. A referência instaurada por esses pronomes é entendida pelo linguista francês como referência objetiva.

Os pronomes ‘eu e tu’, por sua vez, diferentemente de ‘ele’, são signos vazios e indicadores de subjetividade. Eles não têm referência própria, uma vez que é na e pela enunciação que eles passam a ter referência, ou seja, é pelo uso da língua e na instância do discurso que eles se tornam plenos.

A importância atribuída a essas formas é quanto a sua função, o da comunicação intersubjetiva. Conforme expresso por Benveniste, elas se caracterizam por serem

[...] um conjunto de signos ‘vazios’, não referenciais com relação à ‘realidade’, sempre disponíveis e que se tornam ‘plenos’ assim que um locutor os assume em cada instância do seu discurso. Desprovidos de referência material, não podem ser mal empregados; não afirmando nada, não são submetidos à condição de verdade e escapam a toda negação. O seu papel consiste em fornecer o instrumento de uma conversão, a que se pode chamar a conversão da linguagem em discurso. É identificando-se como pessoa única pronunciando *eu* que cada um dos locutores se propõe alternadamente como ‘sujeito’. Assim, o emprego tem como condição a situação de discurso e nenhuma outra (BENVENISTE, 1988f, p. 280-281).

É ao instituir o ‘eu’ como um signo único e móvel, que a linguagem evita o risco de tornar a comunicação impossível. Segundo as palavras de Benveniste (1988f, p. 281) essa ideia está assim expressa.

É essa propriedade que fundamenta o discurso individual, em que cada locutor assume por sua conta a linguagem inteira. O hábito nos torna facilmente insensíveis a essa diferença profunda entre a linguagem como sistema de signos e a linguagem assumida como exercício pelo indivíduo. Quando o indivíduo se apropria dela, a linguagem se torna em instâncias de discurso, caracterizadas por esse sistema de referências internas cuja chave é *eu*, e que define o indivíduo pela construção lingüística particular de que ele se serve quando se enuncia como locutor. Assim, os indicadores *eu* e *tu* não podem existir como signos virtuais, não existem a não ser na medida em que são atualizados na instância de discurso, em que marcam para cada uma das suas próprias instâncias o processo de apropriação pelo locutor.

Na classe formal dos pronomes e internamente diferentes de ‘eu’ e ‘tu’, pela sua função e pela sua natureza, há aqueles reconhecidos como de terceira pessoa. São pronomes que se caracterizam por não remeterem a eles mesmos, mas a uma situação objetiva.

[...] a não-pessoa é o único modo de enunciação possível para as instâncias de discurso que não devam remeter a elas mesmas, mas que predicam o processo de *não importa quem* ou *não importa o que*, exceto a própria instância, podendo sempre esse *não importa quem* ou *não importa o que* ser munido de uma referência objetiva (BENVENISTE, 1988f, p. 282).

A terceira pessoa, conforme afirma o linguista, cumpre a função de representação sintática; correspondendo a uma necessidade de economia, é empregada para substituir um segmento de um enunciado ou até mesmo um enunciado inteiro. Assim, quanto à função, não há nada de comum entre os pronomes de terceira pessoa e os indicadores de pessoa.

Concluindo, por ora, esta apreciação sobre as categorias de pessoa e de não-pessoa, ressaltamos o fato de que neste texto – *A natureza dos pronomes* – Benveniste faz uma distinção entre os signos que fazem parte da sintaxe da língua – categoria de não-pessoa – e os que se situam na instância de discurso – categoria de pessoa –. Os signos que se situam na instância de discurso têm existência linguística no ato de enunciação; eles têm a função de promover a comunicação intersubjetiva. Essa distinção que Benveniste faz, entre a categoria de pessoa e a categoria de não-pessoa, remete a dois tipos de referência, a referência dêitica ‘eu e tu’ e a referência não-dêitica ‘ele’.

Essa diferença reside no tipo de referência que estabelecem. O par ‘eu e tu’ pertence ao nível pragmático da língua, uma vez que é definido na instância de discurso; caracteriza-se por se referir a realidades distintas, sempre que é enunciado. O ‘ele’ pertence ao nível sintático “[...] já que tem por função combinar-se com a referência objetiva de forma independente da instância enunciativa que a contém” (FLORES; TEIXEIRA, 2005, p. 33). Assim, o ‘eu e o tu’ e o ‘ele’ diferenciam-se por sua natureza e função. A não-pessoa tem como propriedade a

[...] de se combinar com qualquer referência de objeto; [...] de não ser jamais reflexiva da instância de discurso; [...] de comportar um número às vezes bastante grande de variantes pronominais ou demonstrativas; [...] de não ser compatível com o paradigma dos termos referenciais como *aqui, agora*, etc. (BENVENISTE, 1988f, p. 283).

A partir dessas considerações é possível afirmar que é a instância de discurso que confere a unicidade no uso das formas ‘eu e tu’. A não-unicidade de ‘ele’, no entanto, está relacionada à sua independência com relação à enunciação. Concebida a noção de pessoa, percebemos que as noções de subjetividade e de intersubjetividade são de grande relevância e complexidade na obra de Benveniste.

Em seguimento ao tema tratado em *A natureza dos pronomes* Benveniste escreve o texto *Da subjetividade na linguagem*. A leitura deste texto se faz necessária para o esclarecimento das noções de subjetividade e intersubjetividade no contexto da comunicação. E, ainda, porque o texto apresenta a visão de Benveniste sobre a função da língua como instrumento de comunicação. O entendimento destas noções é importante para a construção de um novo referencial no âmbito das Linguagens Documentárias.

Neste texto *Da subjetividade na linguagem* Benveniste propõe a noção de subjetividade fundamentada no fato de que o homem está na língua; ele propõe a subjetividade como a capacidade do locutor para se propor como sujeito. Além de abordar as noções de linguagem, subjetividade e intersubjetividade ele critica a concepção de

comunicação que prevalece até então. Ao inserir a noção de referência no contexto da comunicação, o linguista questiona a função da linguagem como ferramenta da comunicação.

Benveniste inicia este texto com as seguintes palavras:

Se a linguagem é, como se diz, instrumento de comunicação, a que deve ela essa propriedade? A pergunta pode surpreender, como tudo o que parece questionar a evidência, mas às vezes é útil pedir à evidência que se justifique. [...] Quanto ao papel de transmissão desempenhado pela linguagem, não podemos deixar de observar, de um lado, que esse papel pode caber a meios não lingüísticos – gestos, mímica – e, de outro lado, que nós nos deixamos enganar, falando aqui de um ‘instrumento’, por certos processos de transmissão que nas sociedades humanas são, sem exceção, posteriores à linguagem e lhe imitam o funcionamento. Todos os sistemas de sinais, rudimentares ou complexos, se encontram nesse caso (BENVENISTE, 1988b, p. 284-285).

Sob este argumento fica claro que em sua opinião não se pode falar em língua como instrumento de comunicação, pois isto vai contra o princípio segundo o qual a linguagem está na natureza do homem. É impossível estabelecer oposição entre ela – linguagem – e o homem, pelo próprio caráter constitutivo da linguagem. Opor o homem à linguagem é colocá-lo em oposição à sua própria natureza. A linguagem é condição para que o homem exista.

Não atingimos nunca o homem separado da linguagem e não o vemos nunca inventando-a. Não atingimos jamais o homem reduzido a si mesmo e procurando conhecer a existência do outro. É um homem falando que encontramos no mundo, um homem falando com outro homem, e a linguagem ensina a própria definição do homem (BENVENISTE, 1988b, p. 285).

É fundamentado no argumento da impossibilidade de se estabelecer oposição entre a linguagem e o homem, que Benveniste nega o aspecto instrumental da linguagem. Pensar a linguagem como instrumento, seria dissociar do homem a propriedade da linguagem. O vaivém da palavra sugere uma troca e isso leva a pensá-la como possuidora de uma função instrumental e não exatamente como um instrumento. Remetida à palavra a função instrumental, Benveniste faz a seguinte consideração:

Para que a palavra assegure a ‘comunicação’, é preciso que esteja habilitada a isso pela linguagem, da qual é apenas a atualização. [...] é na linguagem que devemos procurar a condição dessa aptidão. [...]

É na linguagem e pela linguagem que o homem se constitui como *sujeito*; porque só a linguagem fundamenta na realidade, na *sua* realidade que é a do ser, o conceito de ‘ego’ (BENVENISTE, 1988b, p. 285-286).

Nessa concepção o sujeito “[...] é produto de um jogo de interação dado pelo uso das formas lingüísticas que, pertencentes à língua, possibilitam a passagem de locutor a sujeito num processo de apropriação da língua” (FLORES; TEIXEIRA, 2005, p. 33). Vale ressaltar que a Teoria da Enunciação de Benveniste não fala do sujeito em si, mas da representação

linguística que a enunciação oferece dele. Segundo estes linguistas, o sujeito de que fala Benveniste não reflete uma concepção idealista de sujeito, pois o linguista aborda essa questão do ponto de vista linguístico.

Associada à noção de linguagem e sujeito, estão as de subjetividade e intersubjetividade. Sobre elas Benveniste (1998b, p. 285) faz as seguintes considerações.

A ‘subjetividade’ de que tratamos aqui é a capacidade do locutor para se propor como ‘sujeito’. [...] essa subjetividade [...] não é mais que a emergência no ser de uma propriedade fundamental da linguagem.’. Encontramos aí o fundamento da ‘subjetividade’ que se determina pelo *status* lingüístico da ‘pessoa’.

A consciência de si mesmo só é possível se experimentada por contraste. Eu não emprego *eu* a não ser dirigindo-me a alguém, que será na minha alocação um *tu*. Essa condição de diálogo é que é constitutiva da *pessoa*, pois implica em reciprocidade – que eu me torne *tu* na alocação daquele que por sua vez se designa por *eu*. [...] A linguagem só é possível porque cada locutor se apresenta como *sujeito*, remetendo a ele mesmo como *eu* no seu 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 consequência totalmente pragmática.

Sobre a polaridade ‘eu e tu’ Benveniste ressalta que ela apresenta um tipo de oposição que não se encontra fora da linguagem; na linguagem, *ego* tem uma posição de transcendência sobre o *tu*, entretanto, ‘eu e tu’ não podem ser concebidos sozinhos; eles são complementares e reversíveis. Sob esta ótica, a condição do homem na linguagem é única.

Nas palavras de Benveniste (1988b, p. 286)

A polaridade das pessoas é na linguagem a condição fundamental, cujo processo de comunicação, de que partimos, é apenas uma consequência totalmente pragmática. Polaridade, aliás, muito singular em si mesma, e que apresenta um tipo de oposição do qual não se encontra o equivalente em lugar nenhum, fora da linguagem. Essa polaridade [eu e tu] não significa igualdade nem simetria: *ego* tem sempre uma posição de transcendência quanto a *tu*; apesar disso, nenhum dos dois termos se concebe sem o outro; são complementares, mas segundo uma posição ‘interior/exterior’, e ao mesmo tempo são reversíveis. Procure-se um paralelo para isso; não se encontra nenhum. Única é a condição do homem na língua.

A respeito de linguagem e subjetividade, diz ainda o linguista, que a linguagem é profundamente marcada pela subjetividade. Em relação ao ‘eu e tu’, ele esclarece que tais termos não devem ser tomados como figuras, mas como formas linguísticas que indicam a ‘pessoa’. A propósito das noções de subjetividade e intersubjetividade Benveniste declara: 1) a subjetividade é fundamentada pela categoria de pessoa que está presente no sistema da língua e é dependente da inversibilidade que caracteriza o par ‘eu e tu’. “[...] É numa realidade dialética que englobe os dois termos [eu e tu] e os define pela relação mútua que se

descobre o fundamento lingüístico da subjetividade” (BENVENISTE, 1988b, p. 287). É essa inversibilidade que assegura a intersubjetividade e torna viável falar de categoria lingüística de pessoa; 2) o fundamento da intersubjetividade é constitutivo da pessoa em função da inversibilidade.

Ainda sobre subjetividade e intersubjetividade, Flores e Teixeira (2005) fazem a seguinte observação: a distinção entre o que é da ordem da linguagem e da ordem da língua é muito clara neste texto de Benveniste. Para os linguistas “[...] a intersubjetividade está para a linguagem assim como a subjetividade está para a língua. [...] na linguagem se vê a intersubjetividade como condição da subjetividade” (FLORES; TEIXEIRA, 2005, p. 34). A linguagem tem como função ‘dizer alguma coisa’. A língua é o sistema sob o qual os falantes de uma comunidade estão expostos. É porque existe intersubjetividade que se pode pensar em subjetividade, uma vez que o sujeito, para se propor como tal na linguagem, precisa ser constituído pelo outro.

Nesse mesmo texto, Benveniste retoma o tema da divisão do sistema pronominal em pessoa e não-pessoa. A categoria de pessoa é o fundamento lingüístico da intersubjetividade e a sua referência é ao ‘eu’. Embora possua um lugar na língua, é uma categoria vazia e subjetiva e somente adquire estatuto pleno na e pela enunciação de ‘eu’. A categoria de não-pessoa pertence à esfera não-subjetiva ou objetiva da língua. É constituída por signos que se relacionam a uma realidade objetiva e que fazem referência à própria língua.

Sobre a noção de pessoa, Benveniste ressalta que é inconcebível uma língua sem expressão da pessoa. Pode ocorrer a omissão dos pronomes em algumas línguas, mas jamais a omissão de referências pessoais. Em relação aos pronomes o linguista acrescenta. Eles se distinguem de todas as designações que a língua articula, pois não remetem nem a um conceito, nem a um indivíduo. O ‘eu’, por exemplo, não denomina nenhuma entidade lexical, tal como árvore, que designa um conceito. Nas palavras de Benveniste apresentamos sua concepção sobre ‘eu’.

A que, então, se refere o *eu*? A algo de muito singular, que é exclusivamente lingüístico: *eu* se refere ao ato de discurso individual no qual é pronunciado, e *lhe* designa o locutor. É um termo que não pode ser identificado a não ser dentro do que, noutra passo, chamamos uma instância de discurso, e que só tem referência atual. A realidade à qual ele remete é a realidade do discurso. É na instância de discurso na qual *eu* designa o locutor que este se enuncia como ‘sujeito’. É portanto verdade ao pé da letra que o fundamento da subjetividade está no exercício da língua. [...]

A linguagem está de tal forma organizada que permite a cada locutor *apropriar-se* da língua toda designando-se como *eu* (BENVENISTE, 1988b, p. 288).

Benveniste reconhece nos pronomes pessoais as condições que fundamentam o entendimento da subjetividade na linguagem. Para organizar as relações espaciais e temporais em torno do sujeito, o linguista retoma os indicadores da dêixis, ou seja, os pronomes demonstrativos, advérbios e adjetivos, que são expressos pelos termos ‘isto, aqui e agora’ e suas correlações ‘isso, ontem, no ano passado, amanhã, etc.’ Todos esses pronomes se assemelham por se definirem somente na instância de discurso na qual são produzidos, isto é, sob a dependência do *eu* que aí se enuncia.

Sobre a natureza da subjetividade, ao considerarmos os efeitos do sentido produzidos pela mudança das pessoas em certos verbos de palavras, Benveniste desenvolve o seguinte raciocínio. É preciso que se perceba a natureza da oposição entre as pessoas do verbo. A terceira pessoa, argumenta o linguista, é a forma do paradigma verbal que não remete a nenhuma pessoa, uma vez que se refere a um objeto colocado fora da alocação. O ‘ele’ existe e só se caracteriza por oposição à pessoa ‘eu’ do locutor que, enunciando-a, a situa como não-pessoa. O ‘ele’ tira o seu valor do fato de que faz necessariamente parte de um discurso enunciado por ‘eu’.

Benveniste exemplifica seu posicionamento, afirmando que o enunciado ‘eu juro’ coloca sobre aquele que se enuncia como ‘eu’ a realidade de um juramento. Neste caso, é o próprio ato que compromete o sujeito que diz ‘eu’; a enunciação identifica-se com o próprio ato. ‘Eu juro’ é um compromisso assumido pelo sujeito que enuncia. ‘Ele jura’, no entanto, é apenas uma descrição; neste caso, o verbo jurar assume um valor diferente daquele que exerce em ‘eu juro’. Sob essa proposição, o que se evidencia é que o mesmo verbo adquire valores diferentes quando é assumido por um sujeito ou quando se situa como não-pessoa. “Muitas noções na lingüística, [...] aparecerão sob uma luz diferente se as restabelecemos no quadro do discurso, que é a língua enquanto assumida pelo homem que fala, e sob a condição de *intersubjetividade*, única que torna possível a comunicação lingüística” (BENVENISTE, 1988b, p. 293).

No texto *Da subjetividade na linguagem*, ao fazer distinção entre categoria de pessoa e categoria de não-pessoa, Benveniste faz ainda uma outra distinção, esta entre as categorias da língua e as categorias do discurso. Aqui, tal como no texto *A natureza dos pronomes* também há uma dupla referência, a objetiva, que contempla os signos plenos e que são do paradigma do ‘ele’, e a subjetiva, que contempla os signos vazios e que são do paradigma do ‘eu e tu’. Ao opor o ‘eu e tu’ a ‘ele’ Benveniste opõe a esfera subjetiva da língua, marcada pela categoria de pessoa, à esfera objetiva da língua, marcada pela categoria de não-pessoa.

O domínio da subjetividade se amplia quando Benveniste insere a expressão da temporalidade no ato de enunciação. Seja qual for a língua, a noção de tempo está sempre presente, e sua marca pode aparecer tanto na flexão de um verbo, como em palavras de outras classes, como partículas, advérbios, etc.; trata-se de uma outra característica de estrutura formal.

Passado, presente e futuro são marcas do tempo, mas passado e futuro estarão sempre separados por um presente, e esse presente tem como referência temporal um dado linguístico. O presente é o tempo em que se está – acontecimento descrito – que coincide com o tempo em que se fala – a instância do discurso que o descreve. Este momento é eternamente presente e é determinado pelo locutor para cada uma das instâncias de discurso referidas. O tempo linguístico é *sui-referencial*, isto é, só adquire significação mediante referência à própria enunciação; a temporalidade revela a subjetividade inerente ao próprio exercício da linguagem.

Sobre o texto *Da subjetividade na linguagem*, Kuhn (2009) faz as seguintes ponderações. Considerar a subjetividade como a capacidade do locutor para se propor como sujeito tem diferentes implicações. Uma delas é que a linguagem só é possível porque há um ‘eu’ que se refere a ‘eu’ na instância do discurso, apresentando-se como sujeito. Outra implicação da proposição acima colocada é de que a relação entre as duas noções – linguagem e subjetividade – suspende o que era aceito até então, ou seja, a língua não é um *a priori*, ela só se atualiza na e pela enunciação, que supõe um ‘eu’ que a atualize. Desse modo, essa capacidade do locutor para se propor como sujeito é própria do ato de enunciação, que é marcado por um tempo próprio, o tempo da fala, e que corresponde ao tempo presente. Por conter indicadores que podem ser mobilizados pelo sujeito e empregados em seu discurso, a língua é a possibilidade de subjetividade. E esses indicadores só se constituem como tal pelo uso que o sujeito faz deles. São eles – os indicadores da dêixis – que organizam as relações temporais e espaciais em torno do sujeito; é por meio deles que o homem se instaura na língua.

Sintetizando, podemos dizer que a linguagem é a possibilidade da subjetividade, uma vez que contém as formas linguísticas apropriadas à sua expressão; o discurso, por sua vez, provoca a emergência da subjetividade em razão de consistir de instâncias discretas. A língua, de outro modo, propõe formas vazias, das quais o locutor em exercício de discurso se apropria para se referir a si mesmo. Aquele que se define como ‘eu’, define a um parceiro como ‘tu’. A instância de descrição é assim constitutiva do sujeito. A instalação da subjetividade na linguagem cria a categoria de pessoa na e fora da linguagem, além disso, tem efeitos muito

variados sobre a estrutura das línguas, seja na organização das formas ou nas relações da significação.

Na medida em que estamos considerando a noção de intersubjetividade como oportuna para a proposição de um modelo que esclareça a relação que ocorre entre SRIs e usuários, e que a intersubjetividade não se restringe à categoria de pessoa, mas diz respeito também às noções de espaço e tempo, referência atribuída na e pela enunciação, a leitura do texto *A linguagem e a experiência humana* se torna necessária. A reflexão que se segue na subseção posterior é fundamentada na leitura deste texto.

2.1.1.7 As categorias de pessoa e tempo

No texto *A linguagem e a experiência humana*, Benveniste tem como propósito esclarecer duas categorias do discurso, distintas, mas necessariamente interligadas: a de pessoa e a de tempo. Relacionadas à categoria de pessoa estão ‘eu’, ‘tu’ e ‘ele’ e a categoria de tempo; as manifestações da marca do tempo se tornam visíveis nos verbos, advérbios ou partículas.

Ao se posicionar como ‘eu’ no discurso, em oposição ao ‘tu’ e ao ‘ele’, o homem se coloca em sua individualidade. Este comportamento de se colocar em sua individualidade, ainda que instintivo, na realidade reflete uma estrutura de oposições linguísticas que é inerente ao discurso. Aquele que fala, o locutor, refere-se a si mesmo sempre pelo uso do indicador ‘eu’.

[...] este ato de discurso que enuncia *eu* aparecerá, cada vez que ele é reproduzido, como o mesmo ato para aquele que o entende, mas para aquele que o enuncia, é cada vez um ato novo, ainda que repetido mil vezes, porque ele realiza a cada vez a inserção do locutor num momento novo do tempo e numa textura diferente de circunstâncias e de discursos. Assim, em toda a língua e a todo 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, 1989d, p. 68-69).

Sob esse modo de entender, o pronome ‘eu’, toda a vez que é pronunciado e assumido como ‘eu’ por alguém se transforma, passando de um elemento de um paradigma para uma designação única, e produzindo, a cada vez, uma atualização de uma experiência que é imprescindível em uma língua; não se concebe uma língua sem essa atualização.

A possibilidade de discurso é determinada por essa experiência, que é necessariamente idêntica em sua forma e inerente à forma que a transmite, constituindo a pessoa no discurso e,

consequentemente, toda pessoa, desde que ela fale. Na comunicação, este 'eu' muda de estado, de modo alternativo na relação 'eu e tu', constituindo um processo dinâmico e dialético e propulsor desta subjetividade.

Partilhando da mesma natureza que os pronomes pessoais, que não podem ser ligados nem a um conceito e nem a um objeto, a língua propicia aos falantes um conjunto de formas que são conhecidas como dêiticos, tal como os pronomes demonstrativos. Estes, indicando objetos, organizam o espaço a partir de um ponto central, que é *Ego* e, como tal, o centro e ponto de referência.

Das formas linguísticas reveladoras da subjetividade, as que exprimem o tempo são as mais ricas. Entendendo que o tempo não é expresso apenas pela categoria do verbo, é possível afirmar que a expressão do tempo está presente em todas as línguas, mesmo as que não fazem parte das línguas flexionais, e que ela é compatível com todos os tipos de estruturas linguísticas.

A ideia de que certas línguas ignoram o tempo é muito difundida e motivo de muita confusão. Outro equívoco que é usualmente associado ao sistema temporal de uma língua é o de se pensar que ele reproduz a natureza do tempo objetivo. Com o propósito de esclarecer esta questão, Benveniste faz a seguinte observação. Há efetivamente um tempo específico da língua, mas é preciso reconhecer e distinguir duas noções distintas de tempo: o tempo físico e o tempo crônico. O tempo físico é um contínuo uniforme, infinito, linear, segmentável à vontade. O tempo crônico é o tempo dos acontecimentos, da nossa própria vida, enquanto sequência de acontecimentos; é o tempo da nossa existência pessoal.

O tempo crônico é o tempo vivido e corre sem fim e sem retorno; jamais coincide com o instante que acaba de passar, com o ontem, nem tampouco com a nossa infância. Nossa ligação com o passado, imediato ou longínquo, se dá por meio de pontos de referência que nos permitem situarmos no tempo. O tempo crônico admite um olhar bidirecional sobre os acontecimentos realizados, do passado ao presente e do presente ao passado. Nossa vida é parte desses acontecimentos; ainda que nossa vida vivida corra em um único sentido, nossa visão sobre ela percorre em uma direção ou outra. No tempo crônico os acontecimentos se dispõem em série; eles não são o tempo, mas estão no tempo. Como afirma Benveniste, tudo está no tempo, exceto o próprio tempo.

O linguista francês observa, também, que em todas as formas de cultura e em todas as épocas há um esforço para objetivar o tempo crônico; a objetivação do tempo crônico é uma condição essencial e necessária para a vida dos indivíduos em sociedade. Uma divisão do tempo crônico é baseada na recorrência de fenômenos naturais, como a alternância do dia e da

noite, das fases da lua, dos movimentos das marés e das estações do clima, dentre outros. O calendário é o resultado desses esforços; ele é o tempo crônico socializado.

O tempo medido pelo calendário – dia, mês e ano – é constituído por quantidades fixas, mas essas quantidades são simples denominações do tempo, que não participam em nada da natureza do tempo; em si mesmas elas são vazias de toda temporalidade. O calendário é exterior ao tempo; ele não o acompanha. Ele simplesmente registra os dias, unidades constantes que se agrupam em unidades superiores que são os meses e os anos. Como um dia é idêntico a outro, em um calendário, tomado nele mesmo, não é possível dizer se ele é passado, presente ou futuro. Uma data qualquer, como ‘13 de fevereiro de 1641’, ainda que seja explícita e completa em razão do sistema, não nos informa em que tempo ela foi enunciada.

O tempo crônico fixado em um calendário é estranho ao tempo vivido e não coincide com ele. “[...] pelo próprio fato de ser objetivo, propõe medidas e divisões uniformes em que se alojam os acontecimentos, mas estes não coincidem com as categorias próprias da experiência humana do tempo” (BENVENISTE, 1989d, p. 74).

Distinguindo-se do tempo físico e do tempo crônico, há o tempo linguístico, o tempo da língua, o tempo em que a experiência humana se manifesta, já que o homem se expressa pela língua. Uma coisa é localizar um acontecimento no tempo crônico, e outro é inseri-lo no tempo da língua. “[...] o tempo linguístico manifesta-se irreduzível igualmente ao tempo crônico e ao tempo físico” (BENVENISTE, 1989d, p. 74).

O tempo linguístico diferencia-se do tempo físico e do tempo crônico, por estar organicamente preso ao exercício da fala e por se definir e se organizar como função do discurso. Este tempo, que tem seu centro no presente da instância da fala é simultaneamente gerador e axial. O locutor quando enuncia seu discurso emprega a forma gramatical no tempo presente, situando, desse modo, o acontecimento que enuncia como contemporâneo da instância do discurso que o menciona.

Este presente, enquanto função do discurso não se localiza em nenhuma divisão particular do tempo crônico, uma vez que ele admite todas as divisões do tempo crônico, mas não se refere a nenhuma em particular. Este presente é reinventado toda vez que um locutor enuncia algo, porque será sempre outro momento, um tempo ainda não vivido. Este presente se desloca com a progressão do discurso, mas permanece no presente e constitui a linha divisória entre dois outros momentos produzidos pelo discurso – passado e futuro – e igualmente inerentes ao exercício da fala.

A única expressão temporal que a linguagem dispõe é o presente, que marcado pela coincidência do acontecimento e do discurso é por natureza implícito. A língua não situa o passado e o futuro no tempo do discurso, segundo sua posição própria, mas apenas como pontos de vista para trás ou para frente, a partir do presente. Por necessidade, a língua ordena o tempo a partir de um eixo referencial, e este é sempre a instância do discurso. É impossível pensar uma língua que admitisse o deslocamento deste eixo para o passado ou para o futuro, uma língua em que o ponto de partida da organização do tempo não coincidissem com o presente linguístico.

O presente do discurso é, sem dúvida, o único tempo inerente à língua. Este presente, que é por natureza implícito à língua, determina duas outras referências temporais, que não se relacionam ao tempo, mas às visões sobre o tempo, projetadas para frente ou para trás a partir do ponto presente.

Uma análise das línguas sob o viés do tempo mostra que a forma do passado aparece em todas as línguas, sendo que algumas possuem duas ou mais formas distintas. O mesmo não acontece em relação à expressão do momento futuro. Há muitas línguas que não têm forma específica de futuro; elas frequentemente fazem uso do presente com algum advérbio ou partícula para indicar um momento futuro. Nesta situação enquadram-se sentenças como ‘Amanhã eu vou fazer uma pesquisa sobre tempo linguístico’.

Um aspecto da temporalidade que merece atenção especial é quanto ao modo como se insere no processo da comunicação. “[...] a temporalidade que é minha quando ela organiza meu discurso, é aceita sem dificuldade como sua por meu interlocutor. Meu ‘hoje’ se converte em seu ‘hoje’, ainda que ele não o tenha instaurado em seu próprio discurso, e meu ‘ontem’ em seu ‘ontem’”. (BENVENISTE, 1989d, p. 77-78). É o tempo do discurso que torna possível a condição de intersubjetividade; ele “[...] nem se reduz às divisões do tempo crônico nem se fecha em uma subjetividade solipsista” (BENVENISTE, 1989d, p. 78), ou seja, na ideia de que a única realidade cognoscível é o ‘eu’.

O tempo linguístico comporta suas próprias divisões, em sua própria ordem, de modo diferente do tempo crônico. O ‘hoje, agora, neste momento’, quando pronunciado por um locutor, localiza um acontecimento como simultâneo ao seu discurso. No entanto, o ‘hoje’ no texto escrito, não é o signo que representa o tempo linguístico, e não pode situar o leitor em algum dia do tempo crônico, uma vez que este ‘hoje’ não se identifica com nenhuma data. Por ter sido proferido em um dia qualquer do tempo crônico, ele se aplicará indiferentemente a qualquer dia. Neste caso, se desejarmos torná-lo inteligível fora do tempo linguístico, é

preciso empregá-lo acompanhado do tempo crônico, como por exemplo: ‘hoje, 12 de junho de 1924’.

A mesma situação ocorre com o ‘eu’ quando subtraído ao discurso que o introduz. Por não designar seu locutor real ele é adequado a qualquer locutor. Daí ser necessário acrescentar o nome deste locutor. Nesses casos, o tempo linguístico e o tempo crônico estão presentes em um mesmo enunciado.

A temporalidade linguística, ainda que seja nítida em suas três articulações – passado, presente e futuro –, é muito limitada em cada uma delas. O ‘ontem’, por exemplo, é deslocado para trás e para frente, apenas em dois dias: ontem e anteontem, e amanhã e depois de amanhã. Uma terceira gradação, ‘depois de depois de amanhã’, ainda que aceita não é usual. “O que caracteriza as séries de designações de ordem intersubjetiva, como se vê, é que uma translocação espacial e temporal torna-se necessária para objetivar os signos tais como ‘este’, ‘eu’, ‘agora’, que têm a cada vez um referente único na instância de discurso e somente ele” (BENVENISTE, 1989d, p. 79). A transferência do tempo linguístico ao tempo crônico é efetuada por operadores que funcionam como índices do distanciamento subjetivo. Inserem-se aqui, expressões como: ‘há oito dias’, ‘em três semanas’, etc.

Finalizando esta subseção, retomamos as palavras de Benveniste, que assim se expressa sobre as duas categorias do discurso, a de pessoa e a de tempo.

A intersubjetividade tem assim sua temporalidade, seus termos, suas dimensões. Por aí se reflete na língua a experiência de uma relação primordial, constante, indefinidamente reversível, entre o falante e seu parceiro. Em última análise, é sempre ao ato de fala no processo de troca que remete a experiência humana inscrita na linguagem (BENVENISTE, 1989d, p. 80).

Entendendo que a relação que se estabelece entre SRIs e usuários se caracteriza como um ato de enunciação, e tendo como propósito desenvolver uma proposta que caracterize esta relação, desenvolveremos na próxima subseção a leitura do texto *O aparelho formal da enunciação*. Buscaremos neste texto a compreensão sobre enunciação, a sua estrutura e sua realização, ou seja, sobre o próprio ato, os instrumentos de sua realização e a situação em que se realiza.

2.1.1.8 O Aparelho formal da enunciação

O texto *O aparelho formal da enunciação* tem como propósito principal definir a enunciação no quadro formal de sua realização. Assim, a leitura deste texto é oportuna para os propósitos desta pesquisa, uma vez que nele, Benveniste sistematiza sua Teoria da

Enunciação e formula o aparelho formal da enunciação. O linguista inicia o texto, fazendo distinção entre as condições de emprego das formas e as condições de emprego da língua, o que implica na proposição de dois universos bem diferentes, e em formas diferenciadas de ver, descrever e interpretar as mesmas coisas. A oposição entre a Linguística das formas e a Linguística da Enunciação proposta por Benveniste implica na seguinte concepção: a Linguística das formas pressupõe um objeto estruturado e é responsável pela descrição das regras que dizem respeito à organização sintática da língua; a Linguística da Enunciação subentende a Linguística das formas e tem a enunciação como seu objeto de estudo.

Na sequência do texto, o linguista afirma que o emprego da língua, diversamente do emprego das formas, refere-se a um mecanismo total e constante que afeta a língua toda. Para o linguista a enunciação distingue-se do produto, isto é, do discurso, que é a manifestação da enunciação e que é produzido a cada vez que se fala.

É na passagem da língua em discurso que se dá a semantização da língua. Vista sob esse prisma, a enunciação “[...] é produto de um ato de apropriação da língua pelo locutor, que, a partir do aparelho formal da enunciação, tem como parâmetro um locutor e um alocutário. É a alocação que instaura o outro no emprego da língua” (FLORES; TEIXEIRA, 2005, p. 35). É a relação do locutor com a língua que determina os caracteres linguísticos da enunciação. Antes da enunciação a língua é apenas uma possibilidade da língua, ou seja, é mera virtualidade. Após a enunciação, no entanto, ela é realizada em uma instância de discurso, que tem origem em um locutor e que tem como propósito colocar diante si, um ouvinte, suscitando outra enunciação de retorno.

A ideia de Benveniste (1989b, p. 83) é “[...] esboçar, no interior da língua, os caracteres formais da enunciação a partir da manifestação individual que ela atualiza. Estes caracteres são, uns necessários e permanentes, os outros incidentais e ligados à particularidade do idioma escolhido”. A enunciação, assim entendida, é constituída pelo próprio ato – enunciação –, pelas situações em que ele se realiza – tempo e espaço – e pelos instrumentos de sua realização – a língua –.

Enquanto realização individual, Benveniste (1989b, p. 84) diz que “[...] a enunciação pode se definir, em relação à língua, como um processo de *apropriação*”. O locutor se apropria do aparelho formal da língua e enuncia sua posição de locutor pelo uso de índices específicos – tempo e lugar – e de procedimentos acessórios, que afetam o locutor em sua enunciação, fazendo com que cada instância de discurso seja sempre única. Tão logo alguém se declara como locutor e assume a língua, ele introduz o outro diante de si, sem importar o grau de presença que atribui a esse outro. “Toda enunciação é, explícita ou implicitamente,

uma alocação, ela postula um alocutário” (BENVENISTE, 1989b, p. 84). Sob este ponto de vista, ela não pode ser vista apenas como um ato individual de apropriação da língua realizado em uma instância de discurso.

Na enunciação é preciso considerar, ainda, que a língua é empregada para a expressão de uma certa relação com o mundo.

A condição mesma dessa mobilização e dessa apropriação da língua é, para o locutor, a necessidade de referir pelo discurso, e, para o outro, a possibilidade de co-referir identicamente, no consenso pragmático que faz de cada locutor um co-locutor. A referência é parte integrante da enunciação (BENVENISTE, 1989b, p. 84).

São essas condições que vão orientar o mecanismo da referência no processo de enunciação, criando uma situação sempre única. A noção de referência adquire estatuto central na obra de Benveniste; ela está ausente no nível semiótico, mas não no semântico.

E quando Benveniste (1989b, p. 84) afirma que “[...] a língua se acha empregada para a expressão de uma certa relação com o mundo”, ele pressupõe uma relação mediada pelo sujeito, dependente da enunciação. Para o linguista, o uso da língua é instaurador de sentidos novos a cada novo ato de enunciação, e o referente é o objeto particular a que a palavra corresponde no caso concreto da circunstância de uso.

“[...] Benveniste utiliza a noção de referência para estabelecer o valor semântico daquilo que chama de frase – que poderia ser também chamada de *enunciado* –” (FLORES; TEIXEIRA, 2005, p. 37). A esse respeito Benveniste (1989b, p. 84) diz: “Se o ‘sentido’ da frase é a ideia 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”.

Na medida em que o linguista afirma que não é possível prever nem fixar a referência da frase, é possível afirmar também que ela, a referência, é sempre única a cada instância de discurso. Essa ideia está expressa no texto *A natureza dos pronomes*, quando o linguista diz que a realidade a qual se refere ‘eu’ ou ‘tu’ é a realidade do discurso, e no texto *Semiologia da língua* quando Benveniste (1989e, p. 65-66) afirma que “[...] o semântico toma necessariamente a seu encargo o conjunto dos referentes, enquanto que 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”.

Correspondendo às marcas que situam o locutor no papel central do processo de enunciação, há um centro de referência interna.

Essa descrição sobre o processo de enunciação, ainda que seja relativamente abstrata, se aplica a um fenômeno linguístico que é familiar quanto ao uso, mas que está apenas começando na construção de uma teoria. Em *O aparelho formal da enunciação* essa dupla referenciação – categoria de pessoa e categoria de não-pessoa – parece se desfazer. Neste texto o sistema de referenciação é visto como um elemento constitutivo da língua; a referência integra a língua na sua totalidade, a qual é agenciada por um sujeito e dele depende para ter sentido. O centro de referência passa a ser o sujeito e sua enunciação. Neste texto o linguista propõe que todo o mecanismo de referência seja regido pelo processo de enunciação. Entendendo o ato de uso da língua como sendo constitutivo da enunciação, o linguista propõe um mecanismo de referenciação única, que é o sujeito e sua ação.

A enunciação, quanto a sua estrutura, pode ser caracterizada pelos seguintes segmentos: índices de pessoa – relação ‘eu e tu’ –, índices de ostensão – ‘este e aqui’ – e formas temporais – verbos e advérbios –. Os índices de pessoa, que se constituem na relação ‘eu e tu’, se produzem somente na e pela enunciação. O ‘eu’ significa o indivíduo que profere a enunciação, e o índice ‘tu’ se refere ao indivíduo que se situa na posição de alocutário. Os índices de ostensão – pronomes pessoais e demonstrativos – são numerosos. Eles são da mesma natureza que os índices de pessoa e se relacionam na mesma estrutura da enunciação. Os índices ‘este e aqui’, por exemplo, implicam um gesto que indica um objeto no instante que é pronunciado a instância do termo.

Os pronomes pessoais e os demonstrativos são formas que aparecem como classe de indivíduos linguísticos, sejam eles pessoas, momentos ou lugares. Eles enviam sempre e somente a conceitos. O estatuto desses indivíduos linguísticos resulta do fato de nascerem de uma enunciação, ou seja, de serem produzidos por esse acontecimento que é sempre individual. Esses indivíduos são engendrados, isto é, gerados toda vez que uma enunciação é proferida, designando, assim, a cada vez, algo novo.

As formas temporais, por sua vez, são constituídas pelo paradigma inteiro e se determinam em relação a EGO, o locutor, que é o centro da enunciação. Os tempos verbais têm no presente sua forma axial, uma vez que coincidem sempre com o momento da enunciação. Assim, os tempos verbais, juntamente com os índices de pessoa e de ostensão formam o aparelho formal da enunciação.

Feita a apresentação da enunciação e da constituição do aparelho formal da enunciação, Benveniste se propõe a refletir sobre o tempo, sua necessidade e o que o fundamenta. A temporalidade, diz Benveniste, é produzida na e pela enunciação.

Da enunciação procede a instauração da categoria do presente, e da categoria do presente nasce a categoria do tempo. O presente é propriamente a origem do tempo. Ele é esta presença no mundo que somente o ato de enunciação torna possível, porque, é necessário refletir bem sobre isso, o homem não dispõe de nenhum outro meio de viver o ‘agora’ e de torná-lo atual senão realizando-o pela inserção do discurso no mundo (BENVENISTE, 1989b, p. 85).

Benveniste (1989b) prossegue suas ponderações afirmando que o presente ocupa a posição central na maioria dos sistemas temporais, em diversas línguas. O presente formal, nada mais faz do que explicitar o presente inerente à enunciação, que se renova a cada produção do discurso. Continuidade e temporalidade, que são as marcas do tempo, se produzem no presente incessante da enunciação. O presente da enunciação coincide com o presente do locutor e se delimita, por referência interna, entre o passado e o futuro.

Sob a ótica da enunciação é preciso distinguir os signos que emanam dela, ou seja, os signos que são criados na enunciação em relação ao ‘aqui e agora’ do locutor, como o ‘eu’, o ‘aquele’, o ‘agora’ e o ‘amanhã’, daqueles que têm seu estatuto pleno e permanente na língua. A enunciação, além das formas que comanda, propicia as condições necessárias às funções sintáticas. O enunciador, sempre que faz uso da língua para influenciar de algum modo o comportamento do alocutário, dispõe de um aparelho de funções para alcançar seu propósito.

Sintetizando o que foi exposto nesta subseção, fazemos as seguintes considerações. O aparelho formal da enunciação nada mais é do que a marcação da subjetividade na estrutura da língua, uma espécie de dispositivo que as línguas possuem para que possam ser enunciadas. Ele é fundamental, a base estrutural de uso da língua. Cada vez que o locutor se apropria da língua, ou seja, do aparelho formal da enunciação, ele produz um uso novo e irrepitível. Essa irrepitibilidade decorre do fato de que as categorias de tempo, espaço e pessoa, fundamentais na enunciação, jamais serão eternizadas no uso da língua. Quando Benveniste afirma que a referência é parte integrante da enunciação, ele deixa claro que a referência à enunciação está associada à ideia de um ato individual de utilização da língua, no qual tempo, espaço e pessoa estão presentes, excluindo assim a ideia de referência ao mundo.

Ao articular forma e sentido a partir da noção de enunciação, Benveniste faz coexistir o semiótico e o semântico. Com a noção de enunciação o linguista propõe que o emprego da língua é um mecanismo relativo a toda a língua, sendo o discurso a sua manifestação. A enunciação é o ato de produzir um enunciado, e a língua é o instrumento de que o locutor se utiliza para se enunciar e produzir seu discurso. É no ato de apropriação da língua que o locutor se estabelece em seu discurso e instaura o interlocutor, o espaço e o tempo. É no

interior da língua que o linguista francês busca os caracteres formais da enunciação, isto é, a universalidade do processo de enunciação.

Visando buscar a universalidade do processo de enunciação, Benveniste introduz uma outra linguística ao lado da linguística da forma, uma linguística que se preocupa com o emprego e com as condições de emprego da língua. É a sistematização e a descrição desse aparato linguístico que constitui o que Benveniste chama de aparelho formal da enunciação.

E, para concluir, podemos dizer que o aparelho formal da enunciação é constituído de formas vazias ou dêiticas, que têm a possibilidade de referir somente em uso, viabilizando a enunciação. A configuração do aparelho formal da enunciação pode ser assim descrita. A apropriação da língua se dá por um ato individual que introduz aquele que fala em sua fala. A presença do locutor em sua enunciação cria um centro de referência interno, a partir do qual se criam as demais relações. A emergência dos índices de pessoa no discurso – ‘eu e tu’ – e os índices de ostensão – ‘este, aqui e agora’ – revelariam as formas que compõem o aparelho formal da enunciação.

Há a considerar que certas categorias linguísticas têm sua referência definida apenas em relação ao ‘eu’ no processo enunciativo. A temporalidade, por sua vez, é produzida na e pela enunciação; é nela que se instaura a categoria do presente, decorrendo daí toda a categoria temporal. É a partir do presente que o homem tem a possibilidade de estabelecer outros traços ou recortes de temporalidade.

No contexto da enunciação, Benveniste reconhece a necessidade de muitos outros desdobramentos. É o caso, por exemplo, da enunciação falada e da enunciação escrita. “Esta se situa em dois planos: o que escreve se enuncia ao escrever e, no interior de sua escrita, ele faz os indivíduos se enunciarem. Amplas perspectivas se abrem para a análise das formas complexas do discurso, a partir do quadro formal esboçado aqui” (BENVENISTE, 1989b, p. 90).

A reflexão desenvolvida na subseção seguinte se refere a uma reinterpretação da Teoria da Enunciação de Benveniste. O interesse pela reinterpretação dada por Martins (1990), à Teoria da Enunciação de Benveniste, reside no fato de introduzir o universo da não-pessoa na estrutura da enunciação, tornando o ‘ele’ elemento constitutivo da relação interpessoal. Nesta pesquisa, consideramos que as relações de intersubjetividade que se manifestam nos processos que envolvem esta pesquisa (processo de indexação e processo de busca da informação e recuperação de documentos) considera a linguagem especializada expressa no documento (o universo da não-pessoa) como constitutivo do processo enunciativo.

Na proposta que apresentaremos, como alternativa aos modelos de comunicação documentária usualmente propostos, partimos do pressuposto de que há uma dupla relação de intersubjetividade, a saber: entre bibliotecários e usuários durante o processo de indexação, e entre usuários e Sistema durante o processo de busca da informação e recuperação do documento. Nas duas situações enunciativas, a relação de intersubjetividade se torna plena pelo acesso às informações contidas no documento indexado e a ser recuperado posteriormente.

2.1.1.9 Reinterpretação da Teoria da Enunciação de Benveniste

Ao analisar o processo de enunciação sob a perspectiva benvenistiana, Martins (1990) observa que Benveniste insere a enunciação dentro do enunciado, fato que inviabiliza separar o ato de enunciação do enunciado, ou seja, do produto. Esta impossibilidade é uma das consequências da proposta de Benveniste, que propõe a dissociação do ‘eu e tu’ – universo da interpessoalidade – e do ‘ele’ – universo da objetividade –.

Desse modo, Benveniste não contempla a possibilidade de inserir a presença de fatores externos na relação que se estabelece entre ‘eu e tu’, ou seja, o conteúdo linguístico da própria enunciação. Essa impossibilidade decorre do fato do linguista estudar a enunciação através de suas marcas formais, o que não prevê a interferência constitutiva de um terceiro elemento na relação ‘eu e tu’. Por entender que o enunciado é constitutivo da enunciação, Martins (1990) propõe inserir o conteúdo linguístico do enunciado – o universo da não-pessoa – na própria estrutura da enunciação.

Em sua reinterpretação da teoria benvenistiana, Martins (1990) se alinha a Benveniste quando afirma que o ‘eu’ produz um enunciado, pelo mesmo ato de enunciação em que se coloca como locutor e que institui o ‘tu’ como seu interlocutor, ao qual se opõe pelo traço da subjetividade. Mas se diferencia do linguista quando afirma que a relação ‘eu e tu’ não é independente do conteúdo linguístico da enunciação. Essa relação que se estabelece entre ‘eu e tu’ é definida pelo conteúdo semântico do enunciado.

Com aquilo que enunciam, afirma Martins (1990), ‘eu e tu’ instauram o conteúdo objetivo de sua relação – o universo da não-pessoa –, e é o conteúdo semântico desse universo que vai determinar o caráter da relação interpessoal. Com esse entendimento, a linguista propõe que o sujeito se define pelo tipo de relação linguística que o contexto semântico impõe. Segundo a pesquisadora, ainda, é somente pelo enunciado que a enunciação se materializa.

Ao colocar o enunciado no interior da enunciação, Martins altera o sentido dado por Benveniste em sua teoria, uma vez que o linguista concebe a enunciação e o enunciado numa relação indissociável, o que impede que um terceiro elemento linguístico interfira nessa relação de modo constitutivo.

Martins não contesta a oposição ‘eu-tu/ele’; ela faz um acréscimo à proposta de Benveniste. Ao apresentar a possibilidade de as pessoas ‘eu e tu’ e a não-pessoa ‘ele’ definirem-se mutuamente, ela propõe que a relação intersubjetiva ‘eu e tu’ não seja independente do conteúdo linguístico da enunciação.

A proposição teórica de sua pesquisa é de que o ‘ele’ interfere constitutivamente no processo enunciativo. O respaldo teórico para esta análise está no próprio Benveniste, pela sistematização da Teoria da Enunciação e em Martins (1990), pela reinterpretação que faz ao aparelho formal da enunciação.

Na reinterpretação que Martins (1990) faz dos componentes do aparelho formal da enunciação de Benveniste, ela propõe para o processo de instauração da subjetividade uma relação tríplice ‘eu-tu-ele’ e não apenas uma relação direta ‘eu-tu’. São significativas as considerações feitas pela pesquisadora sobre a possibilidade de uma certa relativização da subjetividade determinada por fatores diversos, que interferem na instauração do processo enunciativo. É sobre a relação ‘eu-tu’ ou ‘eu-tu/ele’ que passaremos a refletir a seguir.

De acordo com Benveniste e, conforme mencionado por Martins (1990, p. 70) o conceito de enunciação pode ser assim exposto: “Sempre que alguém se manifesta linguisticamente, assume o papel de *eu* numa relação interpessoal e conseqüentemente instaura, pelo ato mesmo da enunciação, o *tu* que o pressupõe nessa relação.” Sob essa ótica, é possível dizer que um ato de enunciação se configura quando alguém se manifesta linguisticamente, assumindo a posição de ‘eu’ e estabelecendo uma relação interpessoal com o ‘tu’, que se constitui como seu interlocutor, e que foi convocado por esta relação pelo ‘eu’. O processo de construção do sujeito é entendido por Benveniste como um ato social que se realiza pela linguagem.

Martins constrói sua proposta de enunciação a partir da sistematização apresentada por Benveniste em sua teoria, e da discussão do conceito de diálogo. Enunciação e diálogo são noções vistas pela linguista, tal como por Benveniste, como relações semânticas. Retomando alguns conceitos de Benveniste, Martins (1990) faz a seguinte reinterpretação.

Benveniste admite dois modos de significação: um próprio da língua – semiótico – e outro do uso da linguagem – semântico –. Ele sistematiza esses dois modos de significação de forma teoricamente articulada. Caracteriza os dois modos de significação pela oposição entre

os níveis semiótico e semântico da linguagem. O nível semiótico diz respeito à língua e o semântico é do domínio da língua em exercício, em ação.

No nível semiótico cada signo é dotado de uma significação, que intrinsecamente lhe pertence, mas se distinguindo dos outros signos, uma vez que no plano da significação a língua é um sistema de oposições. Do ponto de vista semiótico a questão da referência está ausente.

No domínio da semântica, o sujeito expressa ideias que são particulares e que estão relacionadas com os objetos e os fatos de seu universo contextual. Ao manifestar-se linguisticamente, o locutor se apropria da língua e constrói seu discurso, fazendo uso das formas que ela a instrumentaliza para a enunciação. A partir do ato de enunciação é possível falar de sentido, que sob essa perspectiva é indissociável do conceito de referência. Essa condição de construção de sentido está implícita na concepção do processo enunciativo. Neste, “[...] o locutor, ao construir seu discurso, institui o outro, estabelecendo uma relação que se dá no mundo, em determinado contexto, ao qual se referem suas palavras, e que só se efetiva na medida em que o outro refere indiretamente, sendo, pois a referência, co-referência” (MARTINS, 1990. p. 71).

O modelo de enunciação desenvolvido por Benveniste propõe que “[...] os significados são constituídos a cada relação enunciativa pela referenciação aos interlocutores e ao contexto da enunciação. Um ato de enunciação é um ato de co-referência, de atribuição conjunta de sentido às palavras” (MARTINS, 1990. p. 71).

O estudo de enunciação, em Benveniste, está estreitamente vinculado com a proposta desenvolvida por ele sobre a estrutura do sistema pronominal, feita com base nos conceitos de categoria de pessoa e categoria de não-pessoa. Essa categorização atribuída aos pronomes fundamenta-se na proposição de um nível pragmático da linguagem em oposição ao nível sintagmático. De acordo com Benveniste, ‘eu’ e ‘tu’ pertencem à categoria de pessoa e estabelecem uma relação intersubjetiva, enquanto que a forma ‘ele’, que se insere na categoria de não-pessoa, é um substituto de elementos que remetem a uma situação objetiva.

A divisão do sistema pronominal, conforme proposto por Benveniste, é relevante para o entendimento da natureza e funcionamento da linguagem sob a perspectiva da descrição do processo de constituição da subjetividade. ‘Eu’ e ‘tu’ não apenas pertencem ao nível pragmático da linguagem como a viabilizam. Essas formas pronominais permitem que o locutor se aproprie da língua e se constitua como sujeito em razão de uma condição de interação linguística. O sujeito se constitui no exercício da linguagem, mas isso não significa

dizer que a linguagem é anterior a língua. ‘Eu’ e ‘tu’ são elementos constitutivos da linguagem, uma vez que é na relação de intersubjetividade que ela se concretiza como tal.

Retomando o que já foi dito anteriormente por Martins (1990) a linguagem é pura virtualidade fora da enunciação. O caráter existencial dessa relação faz com que cada ato de enunciação se constitua sempre como evento único; essa propriedade se transmite ao ‘eu’ e ‘tu’, fazendo com que eles se constituam em elementos de referência sempre nova, e que tenham seu sentido construído a partir da instância de discurso que os contém.

Ao pronome de terceira pessoa Benveniste atribui um estatuto diferente; o ‘ele’ é um substituto de unidades semânticas que referem, fora do circuito da enunciação, ao mundo dos objetos, das ideias, das ações. Na condição de substituto dessas unidades, o ‘ele’ pertence ao nível sintático da linguagem.

Conforme Martins (1990), a divisão do sistema pronominal a partir da categoria de pessoa e não-pessoa tem o mérito de distinguir os elementos necessários para a elaboração de um conceito de enunciação. Essa proposição feita por Benveniste (1988a, 1989a) viabilizou o desenvolvimento do estudo da linguagem sob a ótica da pragmática. Ao propor a divisão pronominal ‘eu-tu/ele’, o linguista estruturou um esquema de categorias relacionadas com a pessoa, como a categoria de tempo, que permitiu propor uma tipologia das enunciações, expressa na oposição discurso e história.

O discurso, segundo a teoria benvenistiana, e conforme ressaltado por Martins (1990), seria do domínio da subjetividade e produzido no diálogo. A história seria do domínio da objetividade. A marca distintiva do discurso e, por consequência do diálogo, seria a presença de ‘eu’ e ‘tu’ no enunciado. A marca distintiva da história seria a narração dos fatos sem a interferência do sujeito narrador; a narração dos fatos seria do domínio do ‘ele’, da não-pessoa, de onde estariam confiscadas as marcas da subjetividade. Vale ressaltar, no entanto, que a bem da verdade Benveniste não aposta com muita convicção nesta tipologia, especialmente no que se refere à história; o linguista afirma que ela se restringe à língua escrita.

Como visto anteriormente, Benveniste estuda a enunciação através de suas marcas formais, a partir da distinção entre categoria de pessoa e de não-pessoa. Sob essa concepção não é possível separar o enunciado da enunciação, pois ela está no interior do enunciado, ou seja, não é possível opor um terceiro elemento linguístico, o ‘ele’, ao ‘eu’ e ao ‘tu’, interferindo na relação, por ser uma relação que se constitui fechada em si mesma.

Conforme Martins (1990), na perspectiva benvenistiana, o enunciado não é capaz de dar uma dimensão histórica e concreta à intersubjetividade e ao diálogo. Por afirmar que a

relação do locutor com a língua determina os caracteres linguísticos da enunciação, não se pode pensar que o enunciado seja determinante da relação interpessoal estabelecida na enunciação, uma vez que a enunciação está no interior dos enunciados no discurso, mas ausente quando se trata de história.

Sob essa ótica, em discurso e história, tal como em pessoa e não-pessoa, estão separadas a intersubjetividade e qualquer realidade linguística fora dela. Sob essa perspectiva teórica, os enunciados que contêm a forma ‘eu’ só podem ser compreendidos por referência a eles mesmos. Entretanto, os enunciados que não contêm a forma ‘eu’ referem objetivamente e devem ser compreendidos por referência ao mundo das coisas e dos fatos. Tais enunciados – enunciados históricos – não têm as marcas da subjetividade, são autônomos em relação às condições pragmáticas da enunciação.

Considerando a oposição ‘eu-tu/ele’ é possível dizer que se trata de uma dissociação que impede que a relação interpessoal seja afetada por um elemento linguístico que se oponha a ela na enunciação. Desse modo, a intersubjetividade e o diálogo se realizam de forma absoluta, uma vez que ocorrem em um circuito fechado.

Quando Benveniste diz (1988b, p. 286) “É ‘ego’ quem *diz ego*”¹³, fica subentendido que a subjetividade é determinada pelo *status* linguístico da pessoa, e que não há uma interferência externa à relação interpessoal, que acene com alguma relativização do sujeito. A enunciação assim entendida, ou seja, a partir da intersubjetividade e por uma relação biunívoca que se dá entre ‘eu’ e ‘tu’, inevitavelmente se cristaliza na reciprocidade e no equilíbrio.

A partir desse entendimento, Martins (1990) formula dois pressupostos teóricos: o de que a enunciação está inserida nos enunciados, e o de que tais enunciados têm características referenciais diversas, decorrentes da presença ou ausência das marcas da subjetividade. Sob esse olhar, o propósito da linguista é o de formular uma proposta teórica sobre a enunciação, que permita pesquisar a relatividade do sujeito e o diálogo. Opondo-se à perspectiva benvenistiana, Martins (1990) acredita que é possível inserir o enunciado na enunciação, concebendo-se a relação enunciativa como um processo em que entram em jogo duas oposições: ‘eu e tu’ e ‘eu-tu/ele’.

13 O original francês “Est ego qui dit ego”, traduzido por “É ‘ego’ que diz ego” poderia, talvez, ser melhor traduzido se, em lugar do relativo que (é ego que diz ego), fosse utilizado o relativo quem (é ego quem diz ego), o que, a nosso ver, esclareceria melhor o sentido da proposição.

O que Martins (1990) pretende com a sua proposta é introduzir o universo da não-pessoa na estrutura da enunciação, tornando o ‘ele’ elemento constitutivo da relação interpessoal. Sob esse entendimento, o enunciado é compreendido como sendo o universo da não-pessoa, independentemente de suas peculiaridades referenciais. O ponto de partida para essa proposição é de que não pode haver uma relação entre ‘eu’ e ‘tu’ fora do universo discursivo em que essa relação se instaura, universo esse instituído na mesma relação linguística em que é objetivado, pelo enunciado.

A enunciação assim formulada passa a ter uma estrutura triangular, e o enunciado assume uma função diversa daquela que tem sido atribuído a ele até então. Nesta reinterpretação feita por Martins (1990), a oposição entre ‘eu-tu/ele’ é incontestável, do mesmo modo como entendia Benveniste. O que diferencia as duas propostas teóricas é que em Martins (1990) ‘eu-tu/ele’ se definem mutuamente pelas relações opositivas que mantêm. Com essa reinterpretação

[...] o *eu* produz um enunciado pelo ato mesmo da enunciação em que se constitui como locutor e, nesse ato, institui o *tu*, seu interlocutor, ao qual se opõe pelo traço de subjetividade. A relação *eu-tu*, entretanto, não é independente do conteúdo lingüístico da enunciação. Essa relação é definida pelo conteúdo semântico do enunciado.

[...] é o conteúdo semântico desse universo que vai decidir o caráter da relação interpessoal. O sujeito não se define como tal por tomar a palavra e dizer *eu* diante de um *tu*, mas pelo tipo de relação lingüística que o contexto semântico determina. (MARTINS, 1990, p. 76).

Assim, na proposta desenvolvida por Martins (1990) é somente pelo enunciado que a enunciação se concretiza e que pode ser discutida na concretude de seus elementos e historicidade. Em outras palavras, é o enunciado que permite recuperar o traço material e concreto que caracteriza a enunciação enquanto acontecimento historicamente determinado.

Finalizando a subseção 2.1.1 Teoria da Enunciação de Benveniste, daremos início a uma reflexão sobre os Estudos de Terminologia.

2.1.2 Estudos de Terminologia

Os Estudos de Terminologia têm como foco de pesquisa a comunicação científica ou técnica, também denominada de comunicação especializada. Esses estudos, em suas diferentes tendências ou escolas de pensamento, tomam como objetos de estudo a unidade terminológica, conceitos e sistemas de conceitos associados a termos, a padronização terminológica, a variação terminológica, a unidade fraseológica, a definição, a linguagem especializada, como um todo, e os textos técnico-científicos.

Nessa área de conhecimento, sobretudo na sua vertente linguístico-descritiva atual, considerando-se tanto as teorias quanto as aplicações terminológicas, entende-se que se deve considerar a comunicação científico-técnica *in vivo*, em sua variabilidade de formas e de sentidos, embora a padronização terminológica permaneça como um ideal a ser administrado. Assim, o uso vivo e mutável das linguagens especializadas prevalece em seu estudo e nos materiais que as repertoriam, ainda que se reconheça a necessidade da implantação e da regulamentação de padrões para uniformizar alguns tipos de terminologias em algumas situações específicas de comunicação. Portanto, tanto a busca de uniformidades quanto o reconhecimento das variações em terminologias estão contemplados pela Terminologia dos dias de hoje.

Naturalmente, o uso de uma terminologia padronizada na comunicação especializada contribui para tornar menos ruidosa a comunicação entre especialistas, e tende a produzir uma convergência nocional e denominativa. Nos textos especializados, especialmente os científicos, deve predominar, além da objetividade e do uso sistemático de termos técnico-científicos, o esforço por alcançar a concisão, a precisão e a adequação. Com tal propósito é necessário que se estabeleça uma série de convenções sobre a sua escrita.

A concisão do texto contribui para diminuir a possibilidade de se produzir distorções na comunicação. A precisão satisfaz aos propósitos da temática técnico-científica e da comunicação entre especialistas. E, segundo as circunstâncias, a adequação à situação comunicativa em que se produz permite uma adaptação às características dos interlocutores e ao respectivo nível de conhecimento sobre o assunto.

Como um sinal da busca da precisão conceitual, vemos nas linguagens científicas palavras próprias ou especialmente cunhadas para circunscrever conceitos e transmitir informações muito específicas para leitores igualmente particulares. Nesse sentido, para viabilizar as condições dessa comunicação, são necessárias algumas estratégias, nas quais os repertórios terminológicos são importantes suportes.

Um exemplo do papel e da natureza desses repertórios, feitos de modo a associar, em postura descritiva, termos variantes e termos normatizados, é o descrito a seguir. Em Portugal, a área que trata de doenças pulmonares associadas a determinados grupos de trabalhadores corresponde ao termo ‘pneumopatias profissionais’. Entretanto, no Brasil, a forma mais empregada e padronizada é ‘pneumopatias ocupacionais’, existindo a variante ‘pneumopatias do trabalho’. Assim, um inventário terminológico inglês-português, que pretenda auxiliar médicos brasileiros e portugueses, além de tradutores de outras línguas para o português, redatores, etc., deve tanto abrigar as formas padronizadas como as variantes de cada

país/comunidade científica. Deve incluir, também, as menos usadas ou até ‘vulgares’, devidamente identificadas como tal, tendo em vista a recuperação da informação para usuários que tenham diferentes perfis de conhecimento prévio do assunto¹⁴.

Originalmente, a dimensão comunicacional da Terminologia foi entendida a partir do universo das ciências exatas, sobretudo das Engenharias. O entendimento que prevalecia à época era que os repertórios terminológicos deveriam abrigar apenas os usos normatizados ou regulamentados das terminologias, sem menção à divergência, à variação denominativa e à polissemia. Hoje é possível afirmar que a presença das terminologias vai muito além desse universo.

Na medida em que é reconhecida a presença dos léxicos temáticos, operando como unidades lexicais a serviço de uma comunicação especializada, é viável admitir que determinadas unidades lexicais adquirem a funcionalidade e o estatuto terminológico em diversas áreas do saber, e que a comunicação especializada se realiza em diferentes níveis de especialização e cultura¹⁵. Desse modo, as terminologias, em algumas situações e áreas de conhecimento, parecem pouco estranhas ao leigo, e em muito se confundem com as palavras comuns do seu dia-a-dia, embora seu significado ‘em uso especializado’ seja bastante peculiar.

Apesar das aproximações que se podem estabelecer entre as linguagens científicas e a linguagem cotidiana, de modo acentuado com o acesso facilitado à informação científica e aos seus termos, a comunicação especializada tende a se diferenciar formalmente da comunicação em geral sob dois aspectos: quanto ao tipo de texto que produz, e quanto ao uso de uma terminologia específica, tendo como produtos mais usuais o artigo científico, a tese, a resenha e o texto especializado, geralmente com inserção em meio acadêmico.

A partir da constatação de que o saber técnico e científico se expande com extrema velocidade, provocando o fenômeno da explosão da informação, ampliam-se os domínios de fixação de vocabulários especializados, que tendem cada vez mais a serem acessados por públicos leigos. Dessa realidade decorre um fenômeno que tende a incorporar, no vocabulário do cotidiano, termos oriundos das ciências e das tecnologias. Entretanto, para a eficácia das comunicações entre especialistas, a univocidade comunicacional ainda é efetivamente uma

¹⁴ Para informações sobre o *Glossário experimental sobre pneumopatias ocupacionais* que tem o português do Brasil como referência, e que traz termos, variantes e definições coletadas em um *corpus* especialmente reunido para a construção desse glossário, consultar o site: <http://www.ufrgs.br/textecc/pneumopatias/>.

¹⁵ Cultura é entendida aqui como as manifestações produzidas pela sociedade. Ela é transmitida predominantemente através da palavra, sendo que sua manifestação no uso da língua leva a uma competência intercultural.

condição a ser alcançada, tendo em vista as diversas proposições de intercâmbio que se intensificam com a globalização. De outro modo, para que o diálogo entre especialistas e leigos seja possível de uma forma qualificada, é preciso encontrar e conhecer formas de dizer que os aproximem em torno de um dado conhecimento.

Considerando os propósitos desta pesquisa, selecionamos como referenciais teóricos a Teoria Comunicativa da Terminologia (TCT) proposta por Cabré e seus colaboradores (1999) e a vertente de Estudos de Terminologia, que pode ser denominada como Terminologia Textual. Relembramos que esta pesquisa se ocupa da eficácia dos SRIs e tem como propósito aprofundar o conhecimento das Linguagens Documentárias, quanto ao seu funcionamento. Ela se desenvolve associada a todo um conjunto de teorias e perspectivas atuais dos estudos desenvolvidos em Terminologia, assim como, a problemática que envolve a padronização e a variação das terminologias.

Entendemos que uma revisão de princípios e de reflexões sobre as linguagens especializadas e a comunicação especializada, feita a partir desses referenciais, tornará possível a construção de Linguagens Documentárias sob um novo olhar. Esse novo olhar se instaura, na medida em que propomos agregar novos elementos teóricos e práticos aos princípios documentários, e aos estudos que vêm sendo desenvolvidos nesta área com fundamentação nas teorias terminológicas. Com esse perfil, acreditamos que essa revisão possa contribuir para a construção de vocabulários controlados especializados mais adequados aos propósitos de uma relação mais eficaz entre SRIs e usuários.

Uma das propostas a ser apresentada, sob esse referencial, é a geração de vocabulários controlados especializados no domínio da Filosofia, uma das áreas de conhecimento das Ciências Sociais e Humanidades. Esta restrição, que implica na exclusão das áreas exatas e biológicas, resulta do fato das linguagens de especialidade destes domínios terem características diferenciadas das áreas alvo desta investigação. A validação deste referencial nas áreas excluídas – ciências exatas e biológicas – precisa ser confirmada através de outras pesquisas.

Tendo como propósito construir um novo referencial no âmbito das Linguagens Documentárias, abordaremos na sequência desta investigação, os seguintes tópicos: Unidade terminológica, Unidade fraseológica especializada, Terminologia Textual e Tipologia textual. É na perspectiva da Teoria Comunicativa da Terminologia que examinaremos as unidades terminológicas, com a preocupação de conceituá-las e caracterizá-las.

2.1.2.1 Unidade terminológica (UT)

Iniciamos nossa revisão de literatura sobre Terminologia, esclarecendo o que entendemos por uma unidade terminológica, visto que há diferentes conceituações, conforme a linha teórica que se escolha. Tomando como referencial Cabré (2002, p. 56-57), afirmamos que

Los términos son *unidades léxicas, activadas singularmente* por sus condiciones pragmáticas de adecuación a un tipo de comunicación. Se componen de forma o denominación y significado o contenido. La forma es constante; pero el contenido se singulariza en forma de selección de rasgos adecuados a cada tipo de situación y determinados por el ámbito, el tema, la perspectiva de abordaje del tema, el tipo de texto, el emisor, el destinatario y la situación.

[...]

El *valor de un término* se establece por el lugar que ocupa en la estructuración conceptual de una materia de acuerdo con los criterios establecidos en un trabajo. Cada ámbito puede ser estructurado desde diferentes perspectivas y en diferentes concepciones, así como cada objeto temático puede participar en más de una estructura con el mismo o diferente valor. Los términos *no pertenecen a un ámbito* sino que *son usados en un ámbito* con un valor singularmente específico.¹⁶

O estudo desenvolvido nesta pesquisa sobre o termo¹⁷ será feito a partir do reconhecimento destas unidades pela identificação de suas características, conforme a opinião de diferentes autores. Dentre os princípios que a teoria terminológica impõe à sua prática, está o que determina que a coleta de termos deve ter origem em textos reais, ou seja, em seus reais contextos de ocorrência, a saber, os textos especializados.

A opção pela concepção de UT da TCT é por considerar que somente uma teoria comunicativa, que leva em consideração o princípio da poliedricidade da unidade terminológica, pode dar conta do caráter essencialmente discursivo das unidades de conhecimento especializado (UCEs).

¹⁶ Tradução da autora: Os termos são unidades léxicas, ativadas singularmente pelas suas condições pragmáticas de adequação a um tipo de comunicação. Eles são compostos de forma ou denominação, e de significado ou conteúdo. A forma é constante; porém o conteúdo se singulariza como uma seleção de traços adequados para cada tipo de situação e determinados pela área, pelo tema, pela perspectiva de abordagem do tema, pelo tipo de texto, pelo emissor, pelo destinatário e pela situação. [...] O valor de um termo se estabelece pelo lugar que ocupa na estrutura conceitual de um assunto, de acordo com os critérios estabelecidos em um trabalho. Cada domínio pode ser estruturado sob diferentes perspectivas e sob diferentes concepções, assim como cada objeto temático pode participar de mais de uma estrutura com o mesmo valor ou não. Os termos não pertencem a um domínio, mas são usados em um campo com um valor singularmente específico.

¹⁷ No âmbito dos Estudos de Terminologia, a expressão termo(s) é usada como variante de unidade(s) terminológica(s), exceto em citações, quando ela pode assumir o valor de descritor, tal como é entendido no campo das Linguagens Documentárias.

Com base nos estudos terminológicos de base comunicativa, Cabré e Estopà (2003) consideram que a unidade terminológica caracteriza-se por ser um objeto multidisciplinar, cognitivo, linguístico e comunicativo. Sob esse entendimento, esse objeto é empregado em textos especializados e pode ser observado a partir de diferentes pontos de vista: 1) do ponto de vista do conhecimento (UCE), 2) do ponto de vista linguístico (USE), e 3) do ponto de vista comunicativo (UCOME), uma vez que essas unidades integram ao mesmo tempo aspectos linguísticos, cognitivos e comunicativos.

Do ponto de vista do conhecimento, as autoras – Cabré e Estopà – são da opinião de que os textos especializados contêm unidades de conhecimento especializado (UCEs), que atuam como veículos do conhecimento especializado e que fazem parte das unidades de conhecimento geral. Desse modo, elas possuem uma dimensão cognitiva, na medida em que expressam o conhecimento especializado. Formalmente, elas tanto podem ser unidades linguísticas – unidades lexicais ou terminológicas, unidades fraseológicas e unidades oracionais –, como unidades não-linguísticas – símbolos, fórmulas, nomenclatura e iconografia –.

A considerar do ponto de vista linguístico, essas unidades têm um significado especializado que podem variar segundo a sua natureza, estrutura e classe gramatical. Nesta ótica elas são reconhecidas como unidades de significação especializada (USEs). As USEs incluem todos os signos encontrados em textos especializados e são usadas em um sentido especializado; elas configuram o comportamento lexical especializado ou temático das línguas. Nessas unidades são encontradas as unidades terminológicas. Convém notar, entretanto, que os termos são considerados como pertencentes às USEs, somente quando eles são da classe dos substantivos e tem a capacidade para se referir ou categorizar a realidade.

Do ponto de vista comunicativo, a UCE torna-se unidade de comunicação especializada (UCOME). Tais unidades, assim como as USEs e as UCEs também podem ser de natureza linguística ou não-linguística.

Por expressar o conhecimento especializado e configurar o léxico especializado ou temático de uma língua, o termo adquire o estatuto terminológico. Ele representa o saber produzido por especialistas de um determinado domínio – científico, técnico e tecnológico – e é inerente à comunicação especializada.

Enquanto componente linguístico e como unidade terminológica e elemento constitutivo da produção do saber, a unidade terminológica apresenta propriedades que favorecem a univocidade da comunicação especializada. Associado à sua natureza constitutiva, ela 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.

A diversidade de posicionamentos, a respeito da natureza e concepção de termo, ultrapassa a discussão no sentido de compreendê-lo como unidade de conhecimento ou unidade pragmático-linguística, e atesta-o como uma unidade complexa e multifacetada. Segundo Cabré (2001, p.32) o termo é “[...] una unidad conceptual y denominativa poliédrica en la que el concepto es percebido desde una perspectiva determinada por vários factores [...]”¹⁸.

Do ponto de vista da função que desempenham no discurso, os termos podem classificar-se em diferentes grupos funcionais, uma vez que funcionalmente participam das mesmas categorias gramaticais do léxico comum. Na categorização dos termos predominam os substantivos designando objetos, mas também podem assumir o valor de termo, adjetivos denominando propriedades, tal como sensível e inteligível, verbos expressando processos, como filosofar, e advérbios.

Até nomes próprios podem ser entendidos como termos, uma vez que circulam nas comunicações especializadas. Essa particularidade aparece na constituição de alguns inventários terminológicos, como ocorre, por exemplo, na área do Direito Ambiental Internacional. “Nessa área, os nomes de Convenções, Tratados, Conferências Internacionais, tal como Convenção para a Proteção de Ozônio [...], adquirem estatuto de termo em razão de seu valor referencial no conjunto dos elementos constitutivos do universo da lei ambiental” (KRIEGER, 2001, p. 77). Na área da Filosofia podemos mencionar alguns nomes próprios assumindo valor de termos, para indexar tematicamente filósofos, tais como Sócrates, Platão e Aristóteles e escolas filosóficas, como Escola de Alexandria, Escola de Atenas e Escola de Éfeso.

No conjunto de suas características formais, os termos apresentam-se tanto como unidades simples quanto complexas. O termo simples é constituído por apenas um radical, e o termo complexo por dois ou mais radicais. Filosofia, Ética, Lógica, Dialética, anamnese, aretê, diairesis, e eudaimonia são entendidos como unidades simples, mas Filosofia antiga, Filosofia platônica, Filosofia política e dualismo antropológico, como unidades complexas. Quanto à sua estrutura morfossintática e léxico-semântica o termo pode se apresentar também como termo composto, sendo, assim como o termo complexo, formado por dois ou mais radicais, mas se diferenciando desse pelo alto grau de lexicalização, e por ser representado

¹⁸ Tradução da autora: uma unidade conceitual e denominativa poliédrica, na qual o conceito é percebido a partir de uma perspectiva determinada por diversos fatores.

graficamente pela utilização do hífen, tal como ocorre com o termo Filosofia greco-romana, médio-platonismo, meta-empírico, não-ser e neo-aristotelismo.

Ainda sob o ponto de vista da forma, os termos apresentam outras configurações sígnicas, entendendo-se como signos verbais plenos as siglas, acrônimos e abreviaturas, e como signos não-verbais, as fórmulas. Somam-se a essas configurações as formas abreviadas, que são representadas no discurso especializado sob uma denominação reduzida de um termo ou de um sintagma terminológico, tal como ocorre no âmbito da filosofia platônica com a unidade terminológica Academia para se referir a Academia platônica.

Os termos reconhecidos como sintagmas terminológicos predominam na comunicação especializada, e são expressos por diferentes denominações: termo complexo, unidades de significação especializada polilexemáticas, termos sintagmáticos e termos-sintagmas. Eles podem ser formados por uma combinação de palavras que obedecem a uma determinada estrutura sintática, e podem assumir tamanho diferenciado em relação ao número de unidades léxicas que o constituem. O conjunto de unidades lexicais que o compõe representa uma única unidade léxico-semântica.

A sua identificação em um discurso especializado nem sempre é fácil, mas os Estudos de Terminologia propõem uma série de critérios, que se fundamentam no grau de lexicalização dos sintagmas, e que auxiliam o terminólogo nessa difícil tarefa. Uma das grandes dificuldades no reconhecimento de um sintagma terminológico é a sua delimitação, especialmente para quem não é especialista da área; a observação de algumas regularidades constitutivas auxilia nessa tarefa.

Conforme Barros (2004, p. 103), o seu reconhecimento se dá pela identificação de algumas de suas características, a seguir enumeradas:

1) a não-autonomia de um componente em relação a outros que compõem a unidade léxico-semântica, sem que haja alteração de sentido. Exemplo: ‘filosofia’ e ‘antiga’ em ‘Filosofia antiga’, ‘filosofia’ e ‘moderna’ em ‘Filosofia moderna’, e ‘filosofia’ e ‘contemporânea’ em ‘Filosofia contemporânea’, por exemplo;

2) a impossibilidade de comutação de um componente sem acarretar mudança de sentido. Exemplo: ‘filosofia social’, significando uma “[...] tentativa de compreender e listar as categorias básicas com que concebemos os aspectos sociais da vida humana” (BLACKBURN, 1997, p. 155) / em contraposição à ‘filosofia da sociedade’; ‘Filosofia moral’ expressando o “Estudo dos conceitos envolvidos no raciocínio prático: o bem, a ação correta, o dever, a obrigação, a virtude, a liberdade, a racionalidade, a escolha” (BLACKBURN, 1997, p. 129) / em contraposição à ‘filosofia dos bons costumes’;

3) a não separabilidade dos componentes. Exemplo: ‘lógica matemática’, com o sentido de um “Termo que designa a lógica formal moderna, em particular os cálculos lógicos suficientemente poderosos para exprimir a matemática” (BLACKBURN, 1997, p. 231) e não ‘lógica e matemática’ ou ‘lógica da matemática’;

4) a particularidade da estrutura interna, determinada, por exemplo, pela integração dos elementos constitutivos, tal como a unidade terminológica ‘Pirro de Elida’, que se refere ao filósofo grego, fundador do ceticismo grego, nascido na cidade de Elida; ou ‘traço de Sheffer’, para expressar as descobertas feitas por Peirce, em 1880, e H.M. Sheffer, em 1913, que concluíram “[...] que as funções de verdade da lógica elementar podiam ser todas definidas a partir de uma única função” (BLACKBURN, 1997, p. 391);

5) a existência de uma definição no âmbito do discurso especializado. Exemplo: ‘Liceu’, para se referir ao “Nome da escola fundada por Aristóteles em Atenas, num passeio coberto ou ginásio, consagrado a Apolo Liceu [...]. Seus membros eram os peripáticos” (BLACKBURN, 1997, p. 226);

6) a sequência sintagmática designa um conceito. Exemplo: ‘linguagem de primeira ordem’¹⁹, significando a “Linguagem na qual os quantificadores contêm apenas variáveis que tomam valores sobre indivíduos (e não, em particular, sobre propriedades desses indivíduos), e em que as funções só têm como seus argumentos variáveis individuais ou constantes” (BLACKBURN, 1997, p. 226);

7) substituições do sintagma interior por um termo lexemático. Exemplo: ‘Academia’ para se referir à ‘Academia de Atenas’.

Quanto ao grau de lexicalização enumeram-se os seguintes critérios de avaliação: imprevisibilidade semântica, maneabilidade, recorrência, co-ocorrência, estabilidade e memorização. Pelo critério de imprevisibilidade semântica o interpretante conhece o sentido de cada palavra do sintagma separadamente, mas não obrigatoriamente compreende o sentido particular do termo sintagmático. Quanto mais o sentido de um sintagma é imprevisível, maiores são as possibilidades de ele ser lexicalizado. Exemplo: ‘filosofia do organismo’, que faz alusão à “Denominação da filosofia da natureza adotada por Whitehead” (BLACKBURN, 1997, p. 154).

¹⁹ Para alguns teóricos, esta unidade pode ser entendida como uma unidade fraseológica, uma vez que se apresenta sob a configuração (S + Prep. + S + S), conforme explica Tagnin. Para Blais esta unidade poderia ser considerada também como uma unidade fraseológica uma vez que não é constituída por apenas um conceito, mas pela reunião de dois conceitos ‘linguagem’ + ‘primeira ordem’, para a formação de um terceiro ‘linguagem de primeira ordem’.

Outro critério importante e que permite verificar o grau de lexicalização de um sintagma é a frequência da co-ocorrência, ou seja, o reencontro da mesma associação de palavras em uma área de conhecimento, repetidas vezes; isoladamente, no entanto, não é um critério de avaliação decisivo. O termo ‘filosofia primeira’, expressão que designa a Metafísica de Aristóteles apresenta 841 ocorrências no Google acadêmico. (Expressão de busca usada na pesquisa: ‘filosofia primeira’ Aristóteles).

O uso frequente de uma construção sintagmática conduz a uma integração semântico-sintática muito forte e à memorização por parte dos usuários. Ressaltamos, porém, que para a identificação do grau de lexicalização de uma unidade terminológica é necessário aplicar mais de um critério.

No âmbito desta pesquisa, uma unidade terminológica será entendida de modo semelhante às palavras, mas se distinguem delas por sua condição de produção e de recepção, e pelo seu modo de significação. É uma unidade discursiva que procede do léxico, e que integra aspectos linguísticos, cognitivos e comunicativos. Ela representa um conceito, que é percebido como unidade especializada de um domínio.

Dando sequência a esta investigação desenvolveremos uma revisão de literatura sobre as unidades fraseológicas especializadas, sob a perspectiva de diferentes pesquisadores. Uma discussão sobre UFEs, no espaço desta pesquisa, justifica-se pela necessidade de compreendê-las como unidades lexicais de representação da informação especializada, uma vez que as unidades terminológicas não são suficiente para representarem as noções de um domínio do conhecimento, no âmbito das Linguagens Documentárias.

2.1.2.2 Unidade fraseológica especializada (UFE)

A discussão sobre unidades fraseológicas especializadas no âmbito dos estudos terminológicos é bastante controversa, tal como ocorre com as unidades terminológicas. Nesta subseção apresentamos a diversidade de pontos de vista identificados em revisão de literatura a respeito das UFEs, sendo que muitos deles são desenvolvidos na comparação com as unidades terminológicas. Temos como propósito assumir todos eles como válidos, uma vez que no âmbito desta pesquisa, o importante é reconhecer todas essas unidades como unidades lexicais de representação da informação especializada, sem a necessidade de caracterizá-las como unidades terminológicas ou como unidades fraseológicas especializadas.

As unidades fraseológicas especializadas são objeto de estudo das fraseologias da língua de especialidade. Elas se diferenciam das unidades terminológicas ou termos por

expressarem um significado que não é dedutível das partes dessa combinação, caracterizando-se pela projeção de um sentido que não depende da soma de seus constituintes. Na língua comum, por exemplo, encontra-se a fraseologia ‘não tem pé nem cabeça’. As fraseologias da língua geral se diferenciam das fraseologias da língua de especialidade, na medida em que essas últimas são portadoras de conhecimento especializado.

Desde os anos 90 têm surgido diversos estudos a respeito dessas unidades. Muitos deles revelam a grande preocupação em definir características e estabelecer fronteira entre os sintagmas terminológicos e as fraseologias especializadas. A prática tem revelado a complexidade de descrição das estruturas fraseológicas, uma vez que sua configuração assemelha-se à dos sintagmas terminológicos. É o caso de estruturas com elemento deverbal, como ‘associação de ideias’, por exemplo. Para alguns pesquisadores trata-se de uma unidade fraseológica especializada que significa “Padrão segundo o qual itens diferentes ocorrem conjunta ou sucessivamente na consciência. As leis dessas associações desempenhariam, no estudo dos processos conscientes, o mesmo papel que as leis da natureza desempenham no estudo dos fenômenos naturais” (BLACKBURN, 1997, p. 27). Para outros, esta mesma unidade pode ser categorizada como uma unidade terminológica, por apresentar uma série de características próprias destas unidades, tais como as enumeradas por Barros (2004) na subseção anterior (2.1.2.1).

Blais (1993) salienta a importância em distinguir termo e fraseologismo. Por termo ela entende a unidade formada por uma palavra (termo simples) ou várias palavras (termo complexo), designando um conceito de forma unívoca no interior de um domínio de conhecimento. E por fraseologismo, a combinação de elementos linguísticos que designam uma combinação de conceitos, tal como ‘história da filosofia’, que resulta da combinação dos conceitos ‘história’ e ‘filosofia’. Segundo ela os fraseologismos se caracterizam por apresentarem configurações diversas, mas nunca chegam a ter a estrutura da frase, situando-se, assim, entre o termo e a frase.

Para Daniel Gouadec (1994), tanto as unidades terminológicas (UTs) quanto as entidades fraseológicas ou fraseologismos, como denomina, são cadeias de caracteres especializados. Entretanto, ele entende que os termos se caracterizam por designar objetos e conceitos, enquanto que as fraseologias são fórmulas ou expressões que servem para expressar um conteúdo próprio de uma área de especialidade, tal como: ‘teoria da substância’, ‘teoria das categorias’ e ‘teoria das ideias’, por exemplo.

Tagnin (2005), ao estudar as expressões idiomáticas e convencionais, aborda a questão das colocações. Este termo foi introduzido pelo linguista britânico J. R. Firth para designar

casos de co-ocorrência léxico-sintática, ou seja, de palavras que se associam e que são consagradas pelo uso sob uma determinada forma. Costumamos dizer ‘cão e gato’, mas não ‘cachorro e gato’, e nem ‘gato e cão’, por exemplo.

Segundo Firth (1969, p. 196)

The statement of meaning by collocation and various collocabilities does not involve the definition of word-meaning by means of further sentences in shifted terms. Meaning for collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words.²⁰

As colocações, sob o aspecto morfossintático, são categorizadas por Tagnin (2005) em: adjetivas (Adj. + S), nominais (S + S), verbais (V + S, V + Adj.) e adverbiais (Adv. + Adj. e V + Adv.). Como exemplos²¹, citamos: ‘condicional contrafactual’ (Adj. + S); ‘competição perfeita’ (S + Adj.); ‘interacionismo reflexivo’ (S + Adj.); ‘interpretação radical’ (S + Adj.); ‘justiça distributiva’ (S + Adj.); ‘método maiêutico’ (S + Adj.); ‘método socrático’ (S + Adj.); ‘ascensão semântica’ (S + S); ‘competição necessária’ (S + S); ‘competência narrativa’ (S + S); ‘referencialmente opaco’ (Adv. + Adj.); ‘referencialmente transparente’ (Adv. + Adj.). A linguagem técnica vale-se muito dessas colocações, que se apresentam, com relativa frequência, na seguinte configuração: S + Prep. + S. É o caso de: ‘associação de ideias’ (S + Prep. + S); ‘bucha de manancial’ (S + Prep. + S); ‘jogo de linguagem’ (S + Prep. + S); ‘lâmina do tinteiro’ (S + Prep. + S); ‘liberdade de expressão’ (S + Prep. + S); ‘mito da linha’ (S + Prep. + S).

Para essa autora, os binômios constituem-se em um tipo especial de colocação e são formados por duas palavras pertencentes à mesma categoria gramatical. Elas são ligadas por uma conjunção ou preposição, tal como: ‘lucros e perdas’; às vezes eles vêm precedidos de preposição, como em ‘da cabeça aos pés’. Quanto ao aspecto sintático, o binômio se caracteriza pela combinabilidade de seus elementos, de forma convencional. No nível semântico os binômios podem se caracterizar por serem idiomáticos ou não-idiomáticos. No âmbito da filosofia há inúmeros binômios que se constituem como unidades lexicais de representação da informação especializada, tais como: ‘bem e mal’ (S + Conj. + S); ‘ato e potência’ (S + Conj. + S); ‘certeza e dúvida’ (S + Conj. + S); ‘conceito e objeto’ (S + Conj. + S); ‘sentido e referência’ (S + Conj. + S).

²⁰ A proposição de significado pela colocação e várias colocabilidades não envolve a definição do significado da palavra por meio de sentenças adicionais em termos deslocados. O significado para colocação é uma abstração no nível sintagmático e não está diretamente relacionado com o conceitual ou com a abordagem ou ideia conceitual para o significado das palavras.

²¹ Os exemplos citados foram retirados da obra de Simon Blackburn – Dicionário Oxford de Filosofia.

Nesta breve revisão de literatura, observamos nos estudos sobre fraseologia uma diversidade de pontos de vista, tanto no plano conceitual como no denominativo, e uma variedade de critérios sobre a constituição formal dessas estruturas. Entretanto, para a maioria dos pesquisadores, as fraseologias especializadas são entendidas como expressões idiomáticas, mais ou menos fixas e recorrentes, com um determinado grau de fixação, e uma frequência significativa em determinado domínio do conhecimento. Elas se configuram pela associação de uma base deverbal e uma base nominal ou termo, com a presença de pelo menos uma unidade terminológica; há quem defenda que os nomes compostos também devem ser estudados no âmbito das fraseologias especializadas (KRIEGER, FINATTO, 2004). Sob esta configuração, citamos: ‘haste de acionamento’ (termo + base deverbal); ‘acesso privilegiado’ (termo + base deverbal).

Conforme revisão das definições propostas por diferentes pesquisadores para as unidades fraseológicas especializadas, Bevilacqua (1999, 2005), mesmo reconhecendo uma diversidade de abordagens envolvendo a sua denominação, definição e critérios para o seu reconhecimento, identifica duas macro tendências: uma de caráter lexicológico e lexicográfico e outra de caráter terminológico. Os critérios adotados para o seu reconhecimento são determinados segundo cada uma dessas visões.

Essas concepções diferenciadas focalizam o problema a partir de critérios distintos. Na de caráter lexicológico e lexicográfico, prevalece um critério formal, sintático, caracterizando uma abordagem lexicológica. Mesmo tendo origem em diferentes teorias e modelos, há uma similaridade nas proposições da primeira tendência, em que predomina a ideia de que a fraseologia é um fenômeno de colocação, bem como de constituição de sintagmas e locuções. A de caráter terminológico dá ênfase à relação com a área de conhecimento, atribuindo uma importância menor aos aspectos morfossintáticos dos constituintes das estruturas fraseológicas.

Além da diversidade conceitual, Bevilacqua (1999, 2005) identifica também uma grande variedade de denominações, que se diferenciam segundo as concepções com que são entendidas pelos pesquisadores. A autora relaciona, dentre outras, as seguintes: colocação, colocação das linguagens especializadas, co-ocorrente, fraseologismo, entidade fraseológica, frasema, unidade fraseológica e unidade fraseológica especializada. Sobre essa variedade terminológica Bevilacqua (1999, 2005) faz duas ressalvas: a primeira é que essa diversidade denominativa nem sempre significa uma diversidade conceitual, e a segunda é que todas se referem a um mesmo objeto de estudo.

Para os que consideram a fraseologia sob uma perspectiva terminológica, há o entendimento de que é necessário estabelecer limites entre os sintagmas ou unidades complexas de base nominal, que podem se caracterizar como termos, e os sintagmas considerados fraseológicos. Dentre os argumentos que utilizam para identificar cada uma dessas unidades – UT e UFE –, citamos dois: do ponto de vista sintático, os termos são predominantemente de categoria nominal, enquanto que as UFEs, de categoria verbal; e do ponto de vista semântico, os termos são definidos como unidades léxicas de caráter denominativo, referindo-se a um conceito, enquanto que as UFEs são de caráter relacional, resultantes da combinação de conceitos.

Sob essa perspectiva as UFEs são definidas como sendo

[...] unidades formadas por um núcleo eventivo, considerado como tal por ser de base verbal ou derivado de verbo (nominalização ou particípio), e por um núcleo terminológico (termo). Entre estes dois núcleos se estabelecem relações sintáticas, mas principalmente semânticas, determinadas pelas propriedades do texto em que são utilizadas. Portanto, são unidades que se conformam no e pelo texto em que são utilizadas. Cumprem, tal como os termos, a função de representar e transmitir conhecimento especializado (BEVILACQUA, 2005, p. 244).

Segundo essa definição, o que fundamentalmente diferencia as UTs das UFEs é que as UTs têm caráter denominativo e valor referencial, representando um núcleo ou nódulo de conhecimento em determinado âmbito, enquanto que as UFEs, por sua vez, são de caráter relacional. As UTs referem-se aos conceitos de um domínio, e as UFEs representam tanto a união de conceitos como as expressões representativas de um domínio com a presença de pelo menos um conceito.

Quanto ao reconhecimento das UFEs, vários critérios têm sido utilizados, mas Bevilacqua (1999, 2005) observa consenso em relação aos seguintes: 1) presença de uma unidade terminológica, 2) estabilidade sintática, 3) estabilidade semântica, 4) índice de frequência relevante, e 5) uso em determinado campo especializado. A aplicação desses critérios facilita o reconhecimento dessas unidades. Mas, por ser essa tarefa muito complexa, a aplicação conjunta de determinados critérios se torna mais eficaz. Um olhar sobre as especificidades do texto ou da área em que são utilizadas é imprescindível nesse processo de reconhecimento.

Concluindo essa discussão, podemos afirmar que há uma série de confluências nos estudos que envolvem as fraseologias especializadas, mostrando, fundamentalmente, uma dupla visão: a que privilegia a relação com o discurso – abordagem terminológica –, e a que prioriza a dimensão sintática ou lexical – abordagem lexical –. Para alguns, trata-se de uma

formulação estereotipada, configurando-se como fórmulas e frases feitas típicas de determinados discursos; mas para outros é uma espécie de fragmento sintagmático, recorrente em textos de mesma temática, tendendo, nesse caso, a ser entendido como colocação, o que lhe determina um caráter mais sintático, ou uma estrutura poliléxica que inclui um termo.

Bevilacqua (2001), ainda que reconheça a dificuldade em definir as unidades fraseológicas especializadas, observa a presença de algumas características que se repetem nas duas abordagens – terminológica e lexical – e que permitem identificá-las. Ela percebe três tendências principais: 1) a que define as UFEs 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; 2) a que as define como combinação sintagmática, cujo núcleo terminológico é um sintagma; 3) a que as considera como sendo representadas por expressões usualmente formuladas em um domínio discursivo e que podem chegar a ser uma frase completa. Bevilacqua (1999, 2005) conclui sua pesquisa afirmando que as UFEs caracterizam uma diversidade conceitual significativa, o que leva a diferentes interpretações na sua categorização. Como resultado, é possível observar que uma mesma cadeia de caracteres é para alguns uma UT, e para outros, uma UFE.

Para dar conta da diversidade de aspectos que envolvem o conjunto dos objetos da Terminologia, de modo a complementar ao que já se percorreu na TCT, procederemos na continuidade desta pesquisa uma revisão fundamentada nas abordagens textuais e em suas concepções. As abordagens textuais, embora já remontem aos anos 80, representam um novo direcionamento para as pesquisas terminológicas. Elas revelam aspectos do comportamento dos termos que podem não ser muito nítidos fora desta perspectiva mais ampla do funcionamento das terminologias.

2.1.2.3 Terminologia Textual

Anterior a uma aproximação mais explícita entre teorias da Terminologia e teorias do texto e do discurso, os procedimentos metodológicos eram diferenciados, especialmente no seu viés mais normativo, associado à figura de Eugen Wüster, e às necessidades de regulamentar terminologias invariantes. Com tais propósitos, a identificação das unidades terminológicas era baseada exclusivamente nas chamadas árvores de domínio, visto que, colocadas em textos concretos e usuais, as terminologias tenderiam a ter algum comportamento variante. Essas árvores de domínio funcionavam como uma espécie de mapa

conceitual de uma área de conhecimento. Elas têm o propósito de auxiliar no reconhecimento da vinculação terminológica, no sentido de situar a pertinência dos termos a uma área.

Árvore de domínio, conforme Krieger e Finatto (2004, p. 56), é entendida como sendo “[...] um construto teórico que desenha a hierarquia temática de cada domínio de saber, pretendendo, com isso, representar o sistema lógico-cognitivo que particulariza os universos do conhecimento especializado”.

Tradicionalmente, a identificação e a seleção dos repertórios terminológicos costumam se orientar a partir de procedimentos de uma árvore de domínio. Este é um recurso importante, mas não é suficiente, uma vez que a noção de domínio é vinculada à de uma estrutura conceitual limitadamente temática. Em consequência, conforme mencionado por Krieger (2001, p. 72), “[...] observam-se as unidades lexicais especializadas num lugar estático, paradigmático, alijadas das contextualizações discursivas.” Assim, apesar da sua funcionalidade, a árvore de domínio, por ser uma estrutura teórica, não se constitui em um recurso metodológico adequado para a pesquisa aplicada em Terminologia, com um propósito descritivo, associado ao uso *in vivo* dos termos ao longo de textos.

Foi com o objetivo de aprofundar o conhecimento sobre a chamada comunicação especializada, e precisar a sua conceituação, que os estudos sobre o texto especializado começaram a despertar o interesse de muitos estudiosos da Terminologia e terminologias. Krieger (2001, p. 72-73) já fazia a ressalva que

Em realidade, assiste-se, hoje, à construção de um amplo cenário de novas e promissoras perspectivas para o estudo dos termos, em muito favorecido pelo desenvolvimento das teorias do texto e do discurso que, ultrapassando os limites da frase, têm maior poder explicativo para dar conta da complexidade que envolve as unidades lexicais especializadas.

É importante ressaltar que, para identificar o funcionamento das comunicações especializadas, sob os enfoques textuais, importantes contribuições advêm das teorias do texto e do discurso, tais como: ‘Linguística textual’, ‘Análise do discurso’, ‘Pragmática’ e ‘Semiótica narrativa greimasiana’. Vale também salientar, que a presença de terminologias em textos especializados não é suficiente para conferir o caráter especializado às comunicações especializadas, e alcançar a sua compreensão. Assim, há toda uma conformação terminológica, sem sentido amplo, desse tipo de texto.

Sob esse novo olhar, que privilegia a inserção textual das terminologias, proliferam uma série de pesquisas sobre as linguagens especializadas. Estas, além de reconhecerem a importância da textualidade e da discursividade na constituição dos objetos terminológicos, começam a manifestar interesse pelas estruturas e tipologias dos textos produzidos nas

comunicações especializadas. Com o propósito de refletir sobre a aproximação entre os Estudos de Terminologia e os estudos sobre o texto especializado, são feitas logo a seguir algumas considerações.

Para os Estudos de Terminologia, a visão de texto como o *habitat* natural das terminologias representa uma mudança de paradigma. Sob o enfoque textualista, que defende a precedência do texto sobre o termo, há três ponderações a serem feitas: 1) o texto é tomado como objeto central de análise; 2) os termos passam a serem concebidos como elementos da linguagem em funcionamento, com todas as implicações daí decorrentes; 3) o contexto discursivo das comunicações especializadas é valorizado.

O reconhecimento da importância do texto, para o estudo dos termos, está diretamente vinculado ao princípio comunicacional postulado pelas novas teorias da Terminologia: a Teoria Comunicativa da Terminologia (TCT) e a Socioterminologia.

A Socioterminologia teve em François Gaudin um de seus principais representantes. 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.

O grande mérito da Socioterminologia é ter impulsionado, a partir dos princípios da Sociolinguística, 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 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. Ambas desenvolvem suas reflexões embasadas nos estudos linguísticos, em especial, nos relacionados às teorias do texto e do discurso. Elas têm no léxico especializado dos sistemas linguísticos a sua matéria-prima, e ambas entendem o termo sob uma dimensão linguística, como uma unidade comunicativa e uma estrutura complexa e poliléxica.

Além de destacar a relevância dos textos, estas teorias postulam o exame do comportamento das unidades terminológicas em seu real contexto de ocorrência. Elas entendem que os termos sofrem os efeitos de todos os mecanismos sintagmáticos e pragmáticos das cadeias discursivas, que dão suporte à comunicação especializada.

Em decorrência das pesquisas informatizadas, e fazendo referência aos estudos desenvolvidos em Linguística de *Corpus*, Slodzian (2000) ressalta que os termos, fora dos contextos de ocorrência e distantes da textualidade, não podem dar conta da diversidade de configurações das estruturas terminológicas.

O reconhecimento de que é preciso considerar os contextos lingüísticos de ocorrência para dar conta da gênese e do funcionamento das terminologias evidencia que o quadro das propriedades semio-narrativas dos diferentes universos de discurso, incluindo tipologias e propósitos particulares das comunicações especializadas, está se impondo como uma produtiva orientação metodológica para detectar o estatuto terminológico de uma unidade lexical que se instaura como tal, dada sua relação semântica e pragmática com o texto especializado (KRIEGER; FINATTO, 2004, p. 108).

Em revisão de literatura ficou evidenciado que, hoje, vários pesquisadores compartilham a opinião de que somente à luz dos fenômenos da textualidade e da discursividade é possível explicar aspectos importantes do comportamento dos termos, sob a ótica de um processo pragmático de comunicação.

Assim visto, cada vez mais as linguagens especializadas são observadas e descritas em grandes acervos textuais, representadas sob o formato de *corpora*. Tratados com apoio informatizado, a Terminologia se aproxima da inserção textual dos termos e do tratamento informatizado de dados lingüísticos, desenhando-se novas formas de estudar os fenômenos das linguagens especializadas, como por exemplo, a mineração de texto²², também conhecida pelo termo inglês ‘text mining’, e a área de Processamento de Linguagem Natural²³(PLN).

²² A mineração de textos surge no contexto da evolução da tecnologia pela necessidade para administrar e armazenar grande volume de informações, digitalização e produção de conhecimentos em formato eletrônico. Ela tem como ponto de partida textos em línguas naturais e como objetivo extrair informações significativas para a formação do conhecimento. O processo de mineração de textos pode ser dividido fundamentalmente em três etapas: 1) preparação dos dados textuais, 2) processamento do texto e 3) avaliação dos resultados. Com a mineração de texto é possível extrair informação relevante de uma grande base de textos, sem ser necessário lê-los previamente.

²³ Processamento de Linguagem Natural, nas palavras de Vieira e Lopes (2010, p. 184) “[...] é uma área de Ciência da Computação que estuda o desenvolvimento de programas de computador que analisam, reconhecem e/ou geram textos em linguagens humanas, ou linguagens naturais. O PLN não é uma tarefa trivial devido à rica ambiguidade da linguagem natural. Essa ambiguidade torna o PLN diferente do processamento das linguagens de programação de computador, as quais são formalmente definidas, evitando, justamente, a ambiguidade.” As autoras acrescentam, ainda, que “De um ponto de vista lingüístico, o foco das pesquisas em PLN pode estar em um de cinco níveis de análise: (a) fonético ou fonológico, (b) morfológico, (c) sintático (d) semântico ou (e) pragmático.” (VIEIRA; LOPES, 2010, p. 185).

Tendo o texto especializado como objeto de investigação, seja ele representado por um texto ou centenas de textos ao mesmo tempo, várias reflexões têm sido feitas. Muitas delas ocorreram ao final do século XX. Entretanto, já há alguns anos se desenvolviam estudos na Alemanha, no leste europeu e na Rússia, com a preocupação de buscar critérios identificadores dos textos especializados considerados em famílias ou grupos. Na sua grande maioria, esses estudos eram motivados pelo princípio de que a análise desse tipo de comunicação não pode estar limitada ao componente temático que cada área do conhecimento toma a si.

Conforme afirma Finatto (2004, p. 352), a Terminologia Textual é um “[...] estudo do texto que tenha termos [...]” e não um “[...] estudo de termos em textos.” Sob essa perspectiva, as unidades de análise não se limitam a termos, uma vez que frases, sintagmas e gêneros textuais, por exemplo, podem ser considerados elementos portadores de especialização.

Ainda referenciando Finatto (2011b, p. 166, 167, 168) ressaltamos o fato de que não há uma única concepção de Terminologia Textual.

São várias as propostas que se colocaram e que seguem se colocando sob esse rótulo. [...] Distingue-se [...] **uma Terminologia Textual**, que caminha dos termos para os textos, mas que segue considerando os termos como seus elementos fulcrais. Não obstante, há também e ao mesmo tempo, a percepção da necessidade de um outro tipo de estudo ‘terminológico’. É aquele que deixa de ser centrado no termo para ser um estudo de textos que tenha termos e outros tantos elementos igualmente passíveis de atenção. [...] Nas duas suas vertentes, adjetivadas e realçadas pelo *textual*, vemos, de um lado, o texto sendo incorporado ao termo – o que é uma necessidade reconhecida, uma lacuna a ser preenchida – e, de outro lado, o texto que incorpora o termo em meio a outros elementos. Cada uma, ao seu modo, permite compreender um pouco melhor o todo heteróclito da linguagem que assume o figurino de linguagem científica, tecnológica ou técnica.

A partir dos estudos desenvolvidos tomando o texto como ponto de partida, há o entendimento de que a abordagem textual permite identificar os aspectos pragmáticos da comunicação especializada, e de que as unidades lexicais assumem posições diferenciadas no contexto deste tipo de comunicação.

Dentre os precursores destes estudos, destacamos Hoffmann (2004), que entende que a linguagem de especialidade não se resume à presença de termos técnicos, mas é constituída pelo conjunto de todos os recursos linguísticos utilizados no âmbito de comunicação.

Segundo Hoffmann²⁴ (2004, p. 81), linguagem especializada

[...] é o conjunto de todos os recursos lingüísticos que são utilizados em um âmbito comunicativo, delimitado por uma especialidade, para garantir a compreensão entre as pessoas que nela trabalham. Esses recursos conformam, enquanto sublinguagem, uma parte do inventário total da língua. Na composição de textos especializados, sua seleção e estruturação estão determinadas tanto pelo conteúdo especializado quanto pela função ou finalidade comunicativa do enunciado, assim como também por uma série de outros fatores objetivos e subjetivos presentes no processo comunicativo.

A escola textualista, mesmo tendo sua origem na escola russa, é fruto de reflexões que se desenvolvem de forma dispersa, não estando vinculada a um país ou grupo de pesquisadores. Tendo como foco o texto, esta escola defende a lingüística do texto especializado.

A contribuição de Hoffmann é fundamental para a Terminologia Textual. Seus trabalhos, produzidos entre os anos de 1982 e 1988 “[...] já salientavam que a linguagem em geral e seus subsistemas especializados perfazem um mesmo sistema que se atualiza em textos” (FINATTO, 2004, p. 354).

Em Hoffmann se encontra “[...] a idéia de que os textos técnico-científicos são objetos lingüístico-comunicativos fundamentais e complexos, que incluem distintos níveis de análise” (FINATTO, 2004, p. 352). Na concepção deste teórico, ainda segundo Finatto (2004), há a precedência do objeto texto sobre o objeto termo. Sob esta perspectiva, depreendemos uma Terminologia Textual focada no estudo do texto que contém termos, contrapondo-se aquela que entende o estudo de termos presente em textos.

Outra contribuição importante a ser considerada, advinda de Hoffmann, reside na sua exploração das dimensões macro e microestruturais do texto especializado, associadas às propriedades básicas da textualidade, usualmente denominadas coesão e coerência em Lingüística textual. Hoffmann, pelo conjunto de suas proposições, tornou-se um marco referencial na investigação dos elementos constitutivos de um texto especializado. Suas ideias também se tornaram fontes de inspiração para novas pesquisas teóricas e aplicadas.

Ainda de acordo com Hoffmann (2004, p. 87-89),

O texto especializado é instrumento e, ao mesmo tempo, resultado da atividade comunicativa exercida em relação a uma atividade especializada sócio-produtiva. Compõe uma unidade estrutural e funcional (um todo) formado por um conjunto finito e ordenado de orações sintática, semântica e pragmaticamente coerentes [...] ou por unidades de valor frasal que correspondem, na condição de signos lingüísticos complexos, a enunciados

²⁴ O texto em alemão *Grundbegriffe der Fachsprachenlinguistik*, originalmente publicado em *Germanistisches Jahrbuch für Nordeuropa* no ano de 1988, foi traduzido com a permissão do autor, Hoffmann, por Ulla Pedde Muss e Maria José Bocorny Finatto, com revisão de Cleci Regina Bevilacqua.

complexos do conhecimento humano e a circunstâncias complexas da realidade objetiva.

[...]

Na visão comunicativa, o texto é o signo lingüístico primário, isto é, sob condições normais, a linguagem se realiza apenas por meio de textos. E isso vale também para o texto especializado. Por isso, deve o texto, e não a palavra ou frase, figurar como ponto central do estudo sobre linguagens especializadas. O que são lidos, traduzidos, resumidos e trabalhados de diferentes modos são os textos. Todas as outras unidades lingüísticas devem ser vistas como seus constituintes, como elementos que mantêm relações diferenciadas entre si, sem as quais a textualidade não se constitui verdadeiramente.

[...]

É no todo do texto que se pode melhor explicar o uso lingüístico especializado, a preferência por determinados recursos lingüísticos, como também interpretações funcionais e valores comunicativos. É com a consideração desse todo que a lingüística das linguagens especializadas deixou finalmente a fase de observar apenas 'particularidades' entre diferentes sublinguagens. Sua postura agora privilegia 'linguagens especializadas em funcionamento' [...].

Concluindo esta subseção, fazemos as seguintes considerações. A Terminologia, em suas reflexões sobre estudos que privilegiem a dimensão textual e discursiva das linguagens especializadas, contribui para explicar a constituição e o funcionamento dos termos.

Reitera-se, dessa forma, que certos problemas de identificação dos termos não podem se restringir aos limites da frase, nem a um recorte paradigmático, mas se explicam à luz dos fenômenos da textualidade e da discursividade e, nesse sentido, dentro de um processo pragmático de comunicação (KRIEGER, 2001, p. 76).

Krieger observa que os estudos fundamentados em fatores da textualidade e da discursividade, concebendo o termo como uma unidade lingüístico-pragmática integrando os processos de comunicação humana, estão consolidando a produtividade de uma nova perspectiva teórica nos Estudos de Terminologia.

Todas as pesquisas que estão sendo feitas no âmbito da Terminologia Textual, demonstram que as investigações sobre o texto especializado têm levado em consideração diferentes abordagens, como aquelas que consideram além dos fatores temáticos e pragmáticos, um conjunto de elementos que são tanto de natureza estrutural, como de natureza lingüística.

Dentre as autoras que trataram da complexidade textual, vale mencionar Ciapuscio. Conforme Finatto²⁵ (2011a)

Essa autora [Ciapuscio] avaliou o grau de abstração conceitual em diferentes tipos de textos que tratavam de uma mesma temática, mas que eram

²⁵ Original obtido na internet.

dirigidos a diferentes perfis de leitores (cientistas, público semi-leigo e leigo). Considerou como fatores distintivos dos graus de especialização desses textos, produzidos por cientistas e por jornalistas que cobrem temas científicos coincidentes, o uso de terminologia específica e a presença de variação terminológica, realizada na forma de sinônimos, paráfrases e explicações.

Há uma série de pesquisas demonstrando o papel da textualidade e dos componentes pragmáticos do processo comunicacional na compreensão de fenômenos complexos, como: a variação denominativa, a variação conceitual, a constituição e o comportamento de sintagmas terminológicos e das fraseologias.

Acrescentamos, ainda, que pesquisas sobre componentes de natureza textual e pragmática trazem novas perspectivas para a descrição e explicação de outros objetos que integram o universo das comunicações especializadas. Dentre esses objetos ressaltamos o exame do comportamento da definição terminológica, e a compreensão do fenômeno da variação. Com efeito, tais estudos evidenciam o poder explicativo das teorias do texto e do discurso, assim como, o papel preponderante da dimensão comunicativa para a compreensão da natureza, do estatuto e do funcionamento das terminologias.

A compreensão do funcionamento das terminologias no texto especializado e o reconhecimento da textualidade e da discursividade é relevante para esta pesquisa, quando se pensa em construir uma ferramenta – vocabulário controlado especializado – que considere o texto como unidade de análise. As abordagens textuais revelam aspectos do comportamento do texto que não são visíveis fora dessa perspectiva.

A subseção seguinte abordará o tema da tipologia textual, ressaltando a visão de Ciapuscio (1998, 2003) sobre esta temática. A autora investiga a variação tipológica no espaço da comunicação especializada, e identifica tanto uma variação conceitual, quanto denominativa.

2.1.2.4 Tipologia textual

O reconhecimento da presença de uma densidade informativa diferenciada, em um texto especializado, é o que justifica o interesse desta pesquisa pelo estudo da tipologia textual. O entendimento de que um estudo sobre tipologias permite identificar características distintas em textos especializados, nos leva à compreensão de que o tratamento a ser dado a um texto, ainda que especializado, pode ser individualizado. Essa percepção tem consequências não apenas na construção de vocabulários controlados especializados, como também, no estabelecimento de políticas de indexação mais flexíveis para um SRI.

A variação tipológica, no espaço da comunicação especializada, aparece nos diferentes tipos de estruturas que apresentam os textos especializados. Ela está presente na estrutura de uma tese, de forma distinta a de um artigo de periódico altamente especializado em determinado domínio, tal como, na estrutura de um texto de jornal ou de uma revista informativa, produzida com o propósito de divulgação de um evento científico ou de difusão ao grande público. A variação da densidade informativa, tal como o uso maior ou menor de terminologias, é decorrente dos propósitos específicos de cada tipo de comunicação, assim como dos diferentes destinatários pretendidos de serem alcançados.

Admitindo que o texto representa a materialização da comunicação especializada, e que a temática e os participantes do ato comunicativo são fatores determinantes da aparência que assume um texto especializado, é possível afirmar que o texto especializado comporta obrigatoriamente tipologias, já que esses dois elementos – temática e participantes – são variáveis.

A preocupação e o interesse em estabelecer uma tipologia é alvo de discussão de várias teorias do texto. Ainda há poucos estudos aprofundando esta questão e até o momento não há consenso entre os linguistas.

[...] solo el nacimiento de esta disciplina a mediados de los años 60 en Europa central determina que la temática *tipos textuales* se convierta en un punto central de interés de los lingüistas: ‘«una teoría de los textos tiene como uno de objetivos fundamentales el establecer una tipología que dé cuenta de todos los textos posibles», afirmaban los fundadores de la disciplina reunidos a tal efecto en Konstanz en 1972²⁶ (CIAPUSCIO, 2003, p. 31).

Ainda que a preocupação por estabelecer uma tipologia de textos em Linguística exista há pelo menos trinta anos, não existe uma tipologia textual de consenso entre os especialistas. Isso ocorre porque há uma diferença conceitual bastante acentuada, entre os linguistas, sobre o que seja um gênero discursivo ou um tipo textual, de modo que texto e discurso não são necessariamente sinônimos. As propostas hoje existentes estão condicionadas às diferentes concepções que veiculam sobre o texto e sobre o discurso.

Sob a perspectiva da variação da densidade informativa, Pearson (1998), com a finalidade de examinar a presença das terminologias, tem como propósito caracterizar a noção de cenário ou situação de comunicação com caráter especializado como um quadro

²⁶ Tradução da autora: [...] somente o nascimento desta disciplina em meados dos anos 60 na Europa central determina que a temática *tipos textuais* se converta em um ponto central de interesse dos linguistas: «uma teoria dos textos tem como um de seus objetivos fundamentais o de estabelecer uma tipologia que dê conta de todos os textos possíveis», afirmavam os fundadores da disciplina reunidos para tal fim em Konstanz em 1972.

conceitual. A autora entende que a comunicação pode ser vista a partir de quatro tipos de cenários comunicativos: 1) comunicação especialista – especialista; 2) especialista – iniciados; 3) especialista mediano – leigo; 4) comunicação professor – aluno.

We believe that is very important that technical terms ‘should have the same meaning for everyone who uses them’ and that the communicative setting is one means of determining whether the conditions are conducive to this. The authors in the first, second and fourth categories described above will endeavour to use terminology in a precise way because the communicative setting requires it. In expert-expert communication (setting 1), there is an assumption that the terminology used is known and understood by the readers except in the case of recently coined terms which may be explained by the author. In communication between experts and initiates (setting 2), many basic concepts are known and understood and readers are expected, through their reading, to acquire and understand additional terminology in order to broaden their understanding of a subject field. In teacher-pupil communication (setting 4), basic concepts are explained with a view to introducing readers to a new or relatively unfamiliar subject field and, as with the previous category, the purpose is didactic and pupils are expected to understand the terminology as explained. In communication between relative experts and the uninitiated (setting 3) terminology is used in a much more popularized way than in any of the other categories. [...] What we are suggesting here is that the terminology used in setting 1, 2 and 4 is likely to be used in a precise way but that the terminology used in setting 3 is used in a less rigid manner and more as part of a general communicative situation. Consequently, we concluded that settings 1, 2 and 4 are reliable sources for potential term candidates, and that when terms are used within these contexts, we can assume that the people using the terms accept the stipulated and agreed meaning associated with these terms (PEARSON, 1998, p. 39)²⁷.

Tendo como foco a relação comunicativa e o tipo de conhecimento focado em cada situação, Pearson (1998) destaca a relação especialista/especialista como sendo a mais confiável de uso terminológico. Esta, em sua opinião, seria a que tem maior probabilidade de

²⁷ Tradução da autora: Nós acreditamos ser muito importante que os termos técnicos tenham o mesmo significado para todos aqueles que os usam, e que o cenário comunicativo seja um meio de determinar se as condições são propícias para isso. Os autores nas primeiras, segundas e quartas categorias descritas acima [comunicação especialista – especialista, especialista – iniciados e comunicação professor – aluno] se esforçarão para usar a terminologia de uma forma precisa, uma vez que o cenário comunicativo exige isso. Na comunicação entre especialistas (cenário 1), há uma suposição de que a terminologia utilizada é conhecida e compreendida pelos leitores, exceto no caso de termos cunhados recentemente que podem ser explicados pelo autor. Na comunicação entre especialistas e iniciados (cenário 2), muitos conceitos básicos são conhecidos e compreendidos, e espera-se que os leitores, através da sua leitura, adquiram e compreendam a terminologia adicional, de modo a ampliar a sua compreensão de um campo de conhecimento. Na comunicação professor-aluno (cenário 4), conceitos básicos são explicados com vistas a levar os leitores a um domínio novo ou não familiar e, como acontece com o cenário anterior, o propósito é didático e espera-se que os alunos entendam a terminologia, tal como é explicada. Na comunicação entre especialistas medianos e os leigos (cenário 3) a terminologia é usada de modo mais popularizada do que em qualquer uma das outras categorias. [...] O que estamos sugerindo aqui é que a terminologia utilizada no cenário 1, 2 e 4 é suscetível de ser utilizado de uma forma precisa, mas que a terminologia utilizada no cenário 3 é usada de uma forma menos rígida e mais como parte de um situação comunicativa geral. Consequentemente, nós concluimos que os cenários 1, 2 e 4 são fontes confiáveis para potenciais candidatos a termos, e que quando os termos são usados nesses contextos, nós podemos assumir como verdadeiro que as pessoas que usam os termos aceitam o significado estipulado e acordado associado com estes termos.

presença de unidades lexicais candidatas a uma unidade terminológica. Nas palavras de Pearson (1998, p. 37) afirmamos que “This particular communicative context [expert-expert communication] is likely to be the one with the highest density of terms²⁸”.

Conforme Krieger e Finatto (2004), Pearson reconhece a importância que o contexto representa para as investigações sobre a identificação do termo, sobretudo, em *corpora* textuais. Nessa ótica é de substancial importância a noção de universo discursivo. Isso porque, manifestada e concretizada em textos de tipologias diversas, apresenta especificidades e propósitos que contribuem para que uma unidade lexical alcance o estatuto de termo, em razão do cenário comunicativo em que se inscreve.

With regard to the notion of communicative setting, we suggest that this may be the most important factor in allowing us to decide whether words are being used as terms or words. It is an area which has been neglected by terminologists because the assumption is that people know instinctively which communicative settings are likely to show a high occurrence of terms. Some researchers in NLP [Natural language processing] have argued that clear syntactic and lexical differences exist between general language and special language situations. (PEARSON, 1998, p. 28)²⁹.

Ao privilegiar esta noção – cenário comunicativo –, Pearson (1998) assume uma abordagem pragmática. Na noção de universo discursivo, ela inclui o discurso científico em suas variadas modalidades e temáticas, como o discurso didático, o discurso jurídico e o discurso político, dentre outros.

Nesta mesma linha de investigação, tendo como enfoque também os graus de especificidade de um texto, há a pesquisa realizada por Ciapuscio (1998, 2003). Tomamos a investigação desenvolvida por esta pesquisadora no âmbito da tipologia dos textos especializados, como fundamentação teórica para esta pesquisa.

A revisão teórica feita por ela não é uma revisão exaustiva, mas permite mostrar que o caminho percorrido no âmbito dos textos especializados é semelhante ao desenvolvido na discussão tipológica em geral. Sua obra *Textos especializados y Terminología* (Ciapuscio, 2003) constitui-se em um marco na teoria terminológica de perspectiva textual, e se limita ao estudo dos textos de especialidade sob a perspectiva do texto como produto. A noção de variação conceitual no âmbito da tipologia textual é introduzida pela pesquisadora.

²⁸ Este contexto comunicativo particular, [comunicação especialista – especialista] é provável que seja o que tenha a maior densidade de termos.

²⁹ No que diz respeito à noção de cenário comunicativo, sugerimos que este pode ser o fator mais importante para que possamos decidir se as palavras estão sendo usadas como termos ou palavras. É uma área que tem sido negligenciada por terminólogos porque o pressuposto é que as pessoas sabem instintivamente quais os cenários comunicativos tendem a mostrar uma alta ocorrência de termos. Alguns pesquisadores em PLN [Processamento de linguagem natural] têm argumentado que há claras diferenças sintáticas e lexicais entre a linguagem em geral e situações de linguagem especial.

Sua pesquisa evidencia, pelas análises que faz, que a dimensão conceitual das comunicações especializadas varia conforme a tipologia textual. Ela constata que mesmo abordando um mesmo tema científico, as terminologias utilizadas geralmente não coincidem. Há muitas variações denominativas segundo o tipo de texto: altamente especializado ou de divulgação geral, por exemplo. E, há também variação quanto ao grau de densidade informativa veiculada, conforme o destinatário: especialista ou leigo, por exemplo.

Segundo Ciapuscio (1998, 2003), sob o olhar das tipologias, os textos podem ser estudados sob o prisma de um maior ou menor grau de especialização, conforme estabelecido nas comunicações especializadas. Dentre os mecanismos determinantes do grau de especialização dos textos, vale citar o tema e o grau de densidade informativa que apresentam, com maior ou menor utilização da terminologia da área de especialidade.

Ao tratar dos graus de especificidade de um texto, Ciapuscio (1998) demonstra que tanto a variação formal como a variação conceitual que os termos sofrem estão em uma íntima relação com os graus de especialização, peculiares a classes textuais distintas, como resumos, artigos, entrevistas e notícias de jornais.

Através das análises que efetua, ela evidencia o fato de que a dimensão conceitual das comunicações especializadas varia conforme a tipologia textual. Ela faz a seguinte afirmação a respeito: “[...] los factores de índole funcional y situacional (interlocutores, clase textual, ámbito discursivo) condicionan la selección, el tratamiento y los límites de la variación formal e conceptual de la terminología³⁰” (CIAPUSCIO, 1998, p.63).

A pesquisadora conclui suas análises afirmando que as terminologias geralmente não coincidem, ainda que abordem um mesmo tema científico. Ela defende a ideia de que o grau de densidade informativa veiculada varia conforme os destinatários do texto: leigos, quase leigos ou especialistas.

Com o objetivo de contribuir com as pesquisas desenvolvidas em Estudos de Terminologia e teorias do texto e do discurso, a partir da temática da tipologização do discurso especializado e dos graus de especialidade dos textos, Ciapuscio (1998, 2003) propõe um modelo para a tipologização dos textos de especialidade, e opta por uma tipologia de caráter cognitivo-comunicativa. Tal enfoque contempla sistemas de multinível ou de várias dimensões ou módulos, que representam os diferentes aspectos do texto, na linha do que se chama de tipologias complexas (Heinemann y Viehweger, 1991; Heinemann, 2000), integradoras (Glaser, 1993) ou modulares (Weise, 1993).

³⁰ Tradução da autora: [...] os fatores de índole funcional e situacional (interlocutores, classe textual, campo discursivo) condicionam a seleção, o tratamento e os limites da variação formal e conceitual da terminologia.

A tipologia proposta pela pesquisadora pretende refletir o conhecimento sobre classes textuais dos falantes, que lhes possibilita a produzir e compreender textos. Este conhecimento tipológico consiste em representações prototípicas nos distintos níveis ou módulos dos textos. Os textos concretos realizam valores de níveis e parâmetros.

Tomando por base a tipologia proposta por Heinemann, Ciapuscio (1998, 2003) distingue os seguintes níveis de texto: funcional, situacional, de conteúdo semântico e formal-gramatical. Tipologizar um texto, enquanto produto, implica realizar um estudo compreensivo nos diferentes níveis; a integração dos traços e valores dos distintos níveis dará como resultado uma caracterização tipológica do texto.

Uma tipologia como a proposta pela pesquisadora permite distinguir, com fundamentação linguístico-textual, graus de especialização e, neste sentido, avança sobre trabalhos que partem de caracterizações externas ou *a priori*, ou que consideram um amplo universo de textos, porém não oferecem sustentação analítico-textual. Por outro lado, a tipologia permite o domínio no âmbito dos textos especializados de exemplares textuais particulares. Sem dúvida, não há que minimizar a dificuldade de estabelecer cortes nítidos, e em certo sentido arbitrários, entre os textos especializados e os não especializados. É necessário ter em mente, que as delimitações resultam de decisões teórico-metodológicas dos analistas.

Para os propósitos desta pesquisa ressaltamos, ainda, as considerações feitas pela pesquisadora sobre os *abstracts*. Na análise que faz sob o ponto de vista funcional, ela destaca o fato de que os *abstracts* são publicados em obras de divulgação que circulam somente entre os especialistas; eles se destinam à comunicação interna entre um grupo significativo de pessoas. Tanto os produtores como os destinatários são especialistas, e a relação entre os interlocutores pode supor-se relativamente simétrica.

No *abstract*, a perspectiva sobre o tema é teórica, e isto tem consequências futuras no campo aplicado. A leitura do *abstract* implica no acesso à informação central da descoberta por parte do especialista. Assim, do mesmo modo que na descrição do nível funcional, a informação se apresenta como uma valorização positiva e com um explícito grau de certeza.

Quanto a sua estrutura em partes padronizadas, o *abstract* corresponde ao esquema prototípico dos resumos: introdução, métodos, resultados, conclusões ou avaliação, que em termos gerais reproduz a estrutura do artigo científico.

No nível formal-gramatical, é possível observar no *abstract* a presença de máximas comunicativas próprias do estilo científico clássico, como economia e desejo de objetividade, por exemplo. Do ponto de vista léxico-gramatical, vale considerar os traços prototípicos do

estilo científico, como o domínio absoluto da terceira pessoa gramatical, o emprego das formas passivas (formas perifrásticas) e nominalizações. Nos *abstracts*, a terminologia é abundante, correspondendo aproximadamente a um terço do vocabulário.

Do ponto de vista semântico, os *abstracts* compartilham a perspectiva básica sobre o tema, com possíveis derivações aplicadas, e o caráter de serem textos primários. Sob os dois olhares – léxico-gramatical e semântico –, os interlocutores são especialistas da área que interatuam em um meio de comunicação que lhes é próprio – uma revista especializada –, mediante categorias textuais privativas desse círculo. Os traços léxicos e sintáticos destacados são consequências das seleções realizadas nos níveis mais globais.

Concluimos esta subseção, ressaltando que, quanto às consequências teóricas, é desejável que uma tipologia de múltiplas dimensões, como a apresentada por Ciapuscio (1998, 2003), contribua para os estudos do texto, mediante o estabelecimento de critérios para caracterizar textos de especialidade. E, ainda, que esta tipologia seja útil para a Terminologia de base comunicativa, a partir da análise dos fatores contextuais, que influem na variação do termo, tanto denominativa como conceitual, em seu âmbito natural de uso.

Sobre o estudo que Ciapuscio (1998, 2003) desenvolve a respeito da tipologia textual ressaltamos que ele traz contribuições importantes para o campo das Linguagens Documentárias. Neste sentido, destacamos como relevantes para esta pesquisa as seguintes ponderações feitas pela pesquisadora: a dimensão conceitual das comunicações especializadas varia conforme a tipologia textual, e o grau de densidade informativa veiculada se altera conforme os destinatários do texto.

Acentuamos, ainda, o seu posicionamento em relação ao *abstract*, especialmente quando afirma que é um texto com altíssimo grau de densidade informativa. Este posicionamento é essencial para sustentar teoricamente a formação de um *corpus* baseado em *abstracts*, e a partir deles, extrair os termos que devem gerar um vocabulário controlado especializado de uma grande área de conhecimento.

Salientamos, no entanto, que o uso de *abstracts* para a construção de vocabulários controlados especializados em uma área específica do conhecimento como Filosofia platônica e Filosofia aristotélica, não se aplica. Esse entendimento tem origem nas características dos *abstracts*, que remetem à informação central da pesquisa e utilizam a linguagem mais representativa do texto, sem, no entanto, contemplarem a terminologia específica presente no texto, na sua totalidade. E, ainda, pelo próprio fato do *abstract* se caracterizar por ser um texto marcado pela concisão e objetividade.

Finalizamos esta reflexão sobre Estudos de Terminologia, afirmando que o referencial teórico advindo deste domínio mostra-se apropriado para esta pesquisa, na medida em que nos esclarece noções fundamentais para o desenvolvimento de ferramentas – vocabulários controlados especializados – que visam propiciar uma relação mais adequada entre os interlocutores envolvidos nos Sistemas de Recuperação da Informação. Linguagem especializada, texto especializado, unidades terminológicas, unidades fraseológicas especializadas e variação terminológica conceitual e denominativa são algumas das noções discutidas no âmbito terminológico, que devem fundamentar a construção dessas ferramentas.

Destacamos, ainda, que dentre as atividades desenvolvidas pelos Estudos de Terminologia, são relevantes para as atividades documentárias, as seguintes: 1) a identificação de conceitos e das relações entre eles, 2) a análise e modelagem de sistemas de conceitos através de diagramas de conceitos, 3) o estabelecimento de representação dos sistemas de conceitos através de diagramas, 4) a definição de conceitos, 5) a atribuição de designações (predominantemente termos) para os conceitos, em uma ou mais línguas, e 6) o registro e a apresentação de dados terminológicos (terminografia).

Na próxima subseção abordaremos a Linguística de *Corpus* quanto aos seus fundamentos teóricos e metodológicos. Acreditando que ela se apresenta como um recurso a ser utilizado pelos profissionais da Ciência da Informação, dedicaremos a discussão seguinte a este tema. Entendendo-a como uma das aplicações dos recursos informáticos no processamento de dados terminológicos, percebemos que ela se enquadra aos 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.

O relacionamento entre os Estudos de Terminologia e a Informática vem de há muito tempo. Cada vez mais, os dois campos se aproximam e buscam um intercâmbio maior. Os Estudos de Terminologia, inicialmente, ansiavam por recursos operacionais; hoje, eles veem na Informática um campo comum de troca de informações de pesquisas, experiências e aplicações. A Informática, por sua vez, se vale das contribuições dos Estudos de Terminologia, quando se trata de selecionar os termos em uma hierarquia lógico-cognitiva, tal como nas aplicações de Inteligência Artificial e no Processamento da Linguagem Natural (PLN). A Linguística de *Corpus*, por sua vez, é um ramo da Linguística Aplicada que se vale

muito do apoio informatizado para descrever o funcionamento da língua em seus diferentes usos.

2.1.3 Linguística de Corpus

A Linguística 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, podemos fazer referência à sua presença na Antiguidade 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. Tal fato nos 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 se pode ver, a pesquisa linguística 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 linguística, sob os mais variados pontos de vista. A importância do computador para a Linguística de *Corpus* está em viabilizar maior precisão e praticidade aos estudos desenvolvidos a partir de um *corpus*.

A Linguística de *Corpus* descreve padrões de língua, sem um fim obrigatório de gerar uma ferramenta. Por depender de conhecimentos advindos da Informática, estabelece com esta área uma interface perene, pois os profissionais desses dois domínios precisam manter um diálogo constante. Diferencia-se da Linguística Computacional pelo seu uso e finalidade.

A Linguística Computacional é a área de conhecimento que se preocupa “[...] com a compreensão da língua e de técnicas computacionais adequadas para o tratamento da língua escrita e falada, tanto para sua interpretação como sua geração” (VIEIRA, 2004, p.1)³¹. Ela envolve um conjunto de atividades que têm como objetivo viabilizar a comunicação com as máquinas, utilizando as habilidades naturais de comunicação humana.

Ela é também chamada de Processamento de Linguagem Natural (PLN), e é um campo de estudos multidisciplinar que se dedica ao estudo de temas como: Inteligência Artificial, Ciência da Computação, Matemática, Lógica, Estatística, Filosofia, Ciência Cognitiva, Psicolinguística (Psicologia da Linguagem) e outras ciências relacionadas.

³¹ Original obtido na internet.

Com a Linguística Computacional é possível construir sistemas que têm a capacidade de reconhecer e produzir informação apresentada em linguagem natural. Diferencia-se da Linguística de *Corpus* porque usa modelos matemáticos para a construção de uma ferramenta; ela não parte de modelos linguísticos de língua tal como faz a Linguística de *Corpus*.

Conforme Vieira e Lima (2001)³², dentre as aplicações possíveis, citamos: 1) reconhecedores e sintetizadores da fala, 2) corretores ortográficos e gramaticais, 3) tradutores automáticos, 4) geradores de textos e resumo, 5) interfaces em linguagem natural, 6) recuperação de informação, 7) extração de informação, 8) avaliação de sistemas de processamento de linguagem natural, e 9) processamento de linguagem natural e sistemas multi-agentes.

Para os profissionais da computação, os *corpora* são utilizados como fonte de treinamento de programas, desenvolvidos com a finalidade específica de análises lexicais, sintáticas e semânticas. O corretor ortográfico dos processadores de texto é exemplo de um desses programas. Os corretores gramaticais, ainda que não apresentem tanta precisão, caminham nessa mesma direção. O sucesso desses programas – corretores gramaticais – depende de etiquetadores de classe gramatical, os quais são inseridos nos textos dos *corpora*; as etiquetas inseridas indicam, por exemplo, a classe gramatical de cada palavra. Os pesquisadores da área de computação buscam, ainda, alcançar o sonho de produzir tradutores automáticos que realmente funcionem.

A Linguística de *Corpus*, além dos conhecimentos da Informática, utiliza também noções 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 de 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. Enfim, a Linguística de *Corpus* se constitui como um processo interativo entre o homem e a máquina, onde conhecimento científico e tecnologia estão presentes e interagem.

Nesta pesquisa recorreremos aos princípios teóricos e metodológicos da Linguística de *Corpus* para a coleta e análise das unidades terminológicas e unidades fraseológicas especializadas em seu contexto de uso e, assim, para a materialização da proposta de construção de vocabulários controlados especializados, a título de projeto piloto. A coleta dessas unidades será feita a partir de um *corpus* formado por textos de caráter acadêmico de

³² Original obtido na internet.

autoria dos próprios especialistas. A partir da constituição de um *corpus* representativo é possível a análise da terminologia de uma área específica do conhecimento.

Esta subseção se desenvolverá a partir de três outras subseções: Princípios gerais da Linguística de *Corpus*, Características de um *corpus* e Ferramentas para a elaboração de *corpora*. Iniciaremos o item seguinte fazendo algumas considerações sobre a Linguística de *Corpus*, seu conceito e seus pressupostos teóricos.

2.1.3.1 Princípios gerais da Linguística de *Corpus*

A Linguística de *Corpus* possibilita refletir sobre o funcionamento e a natureza da língua pela observação de dados linguísticos autênticos organizados em *corpora*. Para a Linguística de *Corpus*, o *corpus* permite observar o que de fato é falado pelos falantes de uma língua. Ele assume importância primordial como fonte de informação, uma vez que ele registra a linguagem realmente utilizada pelos falantes da língua em situações reais. Além disso, os principais pontos nesse tipo de estudo são a observação do uso real e a consideração de que a língua é constituída por palavras, que se definem uma em função das outras com as quais se combinam. Assim, as colocações e as coligações entre palavras – o que inclui termos, palavras e diferentes unidades da língua – assumem um destaque especial.

Nesta pesquisa a Linguística de *Corpus* é entendida como uma metodologia e uma abordagem teórica na exploração de *corpora* especializados, sendo o *corpus* o ponto central da Linguística de *Corpus* e um dos conceitos mais importantes de uma metodologia entendida sob esse enfoque.

Para Sánchez³³ [1995] um *corpus* linguístico é

[...] un conjunto de datos lingüísticos (pertenecientes al uso oral o escrito de la lengua, o a ambos), sistematizados según determinados criterios, suficientemente extensos en amplitud y profundidad de manera que sean representativos del total del uso lingüístico o de alguno de sus ámbitos y dispuestos de tal modo que puedan ser procesados mediante ordenador con el fin de obtener resultados varios y útiles para la descripción y el análisis.³⁴

Como abordagem teórica, a Linguística de *Corpus* está fundamentada por uma visão empirista e probabilística da linguagem. Empíria em Linguística significa primazia aos dados

³³ Original obtido na internet.

³⁴ Tradução de Berber Sardinha (2004, p. 18): Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus campos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.

provenientes da observação da linguagem. Enquanto enfoque empirista, a Linguística de *Corpus* contrapõe-se a uma visão racionalista da linguagem, segundo a qual, o conhecimento provém de princípios estabelecidos *a priori*. O racionalismo, sob o olhar da Linguística, se fundamenta no estudo da linguagem por meio da introspecção. Seguindo a tradição empirista citamos Halliday, e como o maior representante do racionalismo na Linguística, Chomsky, que percebe a linguagem como possibilidade.

Para Halliday (1991, 1992), que é partidário da posição empirista, a linguagem é vista como um sistema probabilístico, no qual certos traços linguísticos – lexicais, estruturais, pragmáticos e discursivos, por exemplo – ocorrem com maior frequência do que outros. Desse modo, ao ampliar a extensão de um *corpus*, aumenta-se, também, a probabilidade de ocorrência de características linguísticas de frequência menor.

Segundo a teoria sistêmico-funcional de Halliday (1991) e no entendimento da linguagem como um sistema probabilístico, o falante seleciona, dentre um conjunto de escolhas possíveis, apenas uma para cada enunciado. O autor observa que entre o conjunto de escolhas possíveis, umas são mais prováveis do que outras, segundo o contexto em que se apresentam. No entendimento de Halliday (1991), a Linguística descreve a probabilidade dos sistemas linguísticos, em função dos contextos em que os falantes os empregam.

A visão da linguagem como sistema probabilístico, além da possibilidade teórica de ocorrência, pressupõe que uma teoria da linguagem deve incorporar a probabilidade de ocorrência dos traços.

Traço é um

Termo usado na LINGÜÍSTICA e na FONÉTICA para indicar qualquer propriedade típica ou observável da LÍNGUA escrita ou falada. Os traços são classificados em termos de vários NÍVEIS de análise lingüística como os ‘traços FONÉTICOS/FONOLÓGICOS/GRAMÁTICAIS/SINTÁTICOS’, ou em termos de uma dimensão da DESCRIÇÃO, como os ‘traços ACÚSTICOS/ARTICULATÓRIOS/AUDITIVOS’. Em um sentido mais geral, os traços podem ser classificados como lingüísticos (ou ‘intralingüísticos’) em oposição aos ‘não-lingüísticos’ (EXTRALINGÜÍSTICOS ou METALINGÜÍSTICOS’. (CRYSTAL, 2000, p. 256).

Para conhecermos qual a probabilidade de um traço ou estrutura, é preciso recorrer a observação empírica da frequência de uso, realizada por vários usuários e em contextos definidos. Berber Sardinha (2004, p. 30-31), o grande divulgador da Linguística de *Corpus* no Brasil, afirma que

A visão da linguagem como sistema probabilístico pressupõe que, embora muitos traços lingüísticos sejam possíveis teoricamente, não ocorrem com a mesma frequência. Por exemplo, no nível morfossintático, a frequência de

substantivos (no inglês e, com certeza, no português) é maior do que qualquer outra categoria; cerca de 25% das palavras (*tokens*) são substantivos.

Isso significa dizer que, em um *corpus*, a probabilidade de ocorrência de uma palavra ser um substantivo é maior, em seu conjunto, do que a ocorrência de palavras de outras categorias gramaticais; ainda que todas tenham a mesma possibilidade de ocorrência.

Dessas considerações, tiram-se duas conclusões: a importância primordial de um *corpus* como fonte de informação, pois ele registra a linguagem natural realmente utilizada por falantes e escritores da língua em situações reais; e a não-trivialidade da investigação da frequência de ocorrência de traços lingüísticos (lexicais, sintáticos, semânticos, discursivos), pois é pelo conhecimento da frequência atestada que se pode estimar a probabilidade teórica (BERBER SARDINHA, 2004, p. 32).

Ressaltamos que a importância da diferença de frequências entre os traços é a de não serem aleatórias. Se o fossem, o fato das possibilidades estruturais se realizarem com frequências diferentes não seria significativo, ou seja, não acrescentaria informação a respeito da própria estrutura. O que ocorre, na realidade, é que há um mapeamento regular entre a frequência maior ou menor de um traço e um contexto de ocorrência. Em outras palavras, podemos asseverar que há uma correlação entre as características lingüísticas e as situacionais – os contextos de uso –, o que evidencia o fato de que a variação não é aleatória.

Dizer que a variação não é aleatória equivale a afirmar que a linguagem é padronizada. A padronização é evidenciada pela recorrência. Uma colocação, coligação ou estrutura que se repete com uma determinada frequência mostra indícios de ser um padrão lexical ou léxico-gramatical.

A linguagem, por ter como característica a padronização, apresenta regularidade e variação sistemática. Pela característica da regularidade, a linguagem forma padrões estáveis em momentos distintos, isto é, tem frequência comparável em *corpora* distintos, e pela variação sistemática, a linguagem forma padrões que se correlacionam com variedades textuais, genéricas, dialetais, etc.

Sobre os pressupostos teóricos da Linguística de *Corpus*, podemos afirmar, resumidamente, que seu foco é o desempenho lingüístico em detrimento da competência, e a descrição lingüística em lugar dos universais lingüísticos. Acresce a isso o fato dela ter uma visão mais empirista do que racionalista da pesquisa científica.

As teorias de Halliday, ainda que formuladas sob uma visão empirista, na forma de linguística sistêmico-funcional, não se enquadram perfeitamente nos preceitos da Linguística de *Corpus*, embora sirvam de arcabouço maior, no qual ela se pode incluir. A teoria

hallidayana não se pauta pela exigência de um *corpus* ou de um instrumental, como é comumente empregado pelos linguistas de *corpus*.

Os *corpora*, que se caracterizam por serem uma coletânea de textos naturais – porções de linguagem –, são construídos a partir de objetivos específicos com o propósito de serem usados como uma amostra da linguagem. Por serem criados com o fim específico de pesquisa, os *corpora*, em si, são artificiais, ainda que sejam formados a partir de textos naturais, isto é, autênticos.

Por textos naturais compreendemos aqueles que são produzidos pelo homem, excluindo-se, desse modo, qualquer coletânea de textos produzidos por programas de geração de textos. Em razão dos questionamentos relacionados ao conceito de texto, alguns autores preferem usar a expressão porções de linguagem para agrupar as diferentes instâncias do texto, como por exemplo, um artigo científico, um resumo ou um trecho de conversação.

Salientamos que um conjunto de textos em formato eletrônico não é necessariamente um *corpus*, visto ser essencial o fato de um *corpus* ser planejado e concretizado segundo critérios linguísticos de seleção. A seleção de um *corpus* deve ser criteriosa, ou seja, deve refletir a variedade escolhida, o mais fielmente possível.

A princípio, um *corpus* não pode ser formado por sentenças soltas, uma vez que elas não seriam verdadeiras representantes da linguagem. Há o entendimento de que a linguagem natural autêntica não é formada por fragmentos desconexos, tal como são as sentenças soltas.

Em síntese, podemos fazer as seguintes considerações sobre um *corpus*:

1) quanto à origem, os dados devem ser autênticos. Assim sendo, o *corpus* deve ser constituído a partir de textos autênticos, em linguagem natural;

2) em relação ao propósito, ou seja, à finalidade, o *corpus* deve ser um objeto de estudo linguístico, o que significa dizer, que os *corpora* não podem ser criados em linguagem artificial, tal como linguagem de programação de computadores;

3) do ponto de vista da composição, o conteúdo do *corpus* deve ser criteriosamente escolhido, tendo como princípio de seleção de textos, as condições de naturalidade e autenticidade;

4) a respeito da formatação, os dados devem ser legíveis por computador;

5) sob o olhar da representatividade, o *corpus* deve ser representativo de uma variedade linguística, ou mesmo, de um idioma, e a seleção de textos deve ser aleatória;

6) quanto à extensão, o *corpus* deve ser vasto para ser representativo de uma língua ou variedade.

Acrescentamos 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 se dedica exclusivamente aos aspectos intelectuais que orientam os propósitos específicos da pesquisa, tais como: coleta, organização, validação e interpretação dos dados.

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. 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 linguística sob o enfoque da Linguística de *Corpus*. É importante ressaltar, no entanto, que embora um *corpus* seja uma fonte rica e inesgotável para a pesquisa, ele 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.

Quanto às bases teórico-metodológicas da Linguística de *Corpus*, elas são encontradas em J. R. Firth, que é considerado um pioneiro nesta área. O autor criou as bases para uma teoria linguística que viabiliza a descrição das particularidades de uma língua e de seus sistemas. Ele destaca a função comunicativa da linguagem e a importância em estudar a língua, desde seu contexto de uso ou, como o próprio Firth chama, contexto de situação. Embora este termo seja reconhecido como sendo de Firth, na realidade ele deriva do antropólogo Bronislaw Malinowski.

Firth percebe o contexto de situação como parte do aparelho lingüístico, do mesmo modo como são as categorias gramaticais que ele usa, e faz a seguinte afirmação.

My view was, and still is, that ‘context of situation’ is best used as a suitable schematic construct to apply to language events, and that it is a group of related categories at a different level from grammatical categories but rather of the same abstract nature. A context of situation for linguistic work brings into relation the following categories:

A - The relevant features of participants: person, personalities

(1) The verbal action of the participants.

(2) The non- verbal action of the participants.

B - The relevant objects.

C - The effect of the verbal action.³⁵ (FIRTH, 1969, p. 182)

³⁵ Tradução da autora: Minha visão era, e ainda é, que ‘contexto de situação’ é melhor usado como um construto esquemático adequado para aplicar a eventos de linguagem, e que este é um grupo de categorias relacionadas

Mesmo não tendo acesso aos modernos recursos da Informática, o linguista inglês percebeu seu valor para a Linguística. A partir da ideia de que há uma regularidade nos tipos de associações a que se submetem as palavras de uma língua, Firth 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, a frequência e a 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-linguístico da comunicação.

Embora Firth tenha mencionado muitas vezes em seus trabalhos as categorias contextuais, ele nunca chegou a desenvolver essa ideia de modo satisfatório. No entanto, ele teve o mérito de assumir que as categorias contextuais eram suscetíveis de serem descritas linguisticamente.

Fillmore, um dos pioneiros dos estudos linguísticos assistidos pelo computador, fez a seguinte afirmação no famoso *Simpósio Nobel sobre Linguística de Corpus, I*, 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).³⁶

Dentre as vantagens que a Linguística de *Corpus* apresenta estão os 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 Linguística 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

a um nível diferente de categorias gramaticais, mas de preferência da mesma natureza abstrata. Um contexto de situação para o trabalho linguístico estabelece relação entre as seguintes categorias:

A - as características relevantes dos participantes: pessoa, personalidades

(1) A ação verbal dos participantes.

(2) A ação não-verbal dos participantes.

B - Os objetos relevantes.

C - O efeito da ação verbal.

³⁶ Tradução da autora: Tenho duas observações importantes a fazer. 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.

da coleta, armazenamento e análise de dados, e o pesquisador constitui-se como elemento fundamental por determinar quais são os dados que precisam ser coletados, e como devem ser organizados para que seja viável o processo de pesquisa.

Sobre o estatuto da Linguística de *Corpus*, há um debate entre os teóricos da área. Para alguns, ela é uma disciplina, enquanto que para outros ela é uma metodologia. E há, ainda, quem afirme que ela não é nem disciplina, nem metodologia.

Os que afirmam que a Linguística de *Corpus* não é uma disciplina apresentam os seguintes argumentos. A Linguística de *Corpus*, tal como a Psicolinguística e a Sociolinguística, dentre outras, não é uma disciplina porque seu objeto de pesquisa não é delimitado. Ela não se dedica a um assunto definido, mas se ocupa de vários fenômenos comumente enfocados em outras áreas. Dentre esses teóricos, conforme mencionado por Berber Sardinha (2004), citamos McEnery e Wilson, que afirmam que a Linguística de *Corpus* é apenas uma metodologia, e Leech que descreve a Linguística de *Corpus* como uma base metodológica.

Os teóricos que entendem a metodologia como instrumental, alegam que é possível aplicar o instrumental da Linguística de *Corpus* livremente, e preservar a orientação teórica da disciplina original. Sob essa perspectiva, teríamos, por exemplo, a sintaxe ou a fonologia tradicional, e a sintaxe ou a fonologia baseada em *corpus*. Desse modo, os dados, a orientação, os pressupostos teóricos e as implicações dos resultados permaneceriam, sendo que aquilo que mudaria seria o instrumental.

Para os teóricos que percebem a metodologia como um modo típico de aplicar um conjunto de pressupostos de caráter teórico, tal qual John Sinclair, a Linguística de *Corpus* pode ser compreendida como uma metodologia, a partir do entendimento de que ela não se resume, simplesmente, a um conjunto de ferramentas computacionais.

Dentre aqueles que defendem que a Linguística de *Corpus* não é uma metodologia, sustentam seu entendimento no fato de seus praticantes produzirem conhecimentos novos. Para eles, a análise de um *corpus* frequentemente revela fatos a respeito de uma língua, que dificilmente seriam observados sem a análise do *corpus*.

A produção de conhecimento a coloca em condições similares à Linguística aplicada, que tem sido reconhecida por seu caráter essencialmente transdisciplinar. Hoje a Linguística de *Corpus* é entendida como algo maior do que um simples espaço no qual se aplicam os conhecimentos produzidos na Linguística.

Para os que compartilham com a opinião que a Linguística de *Corpus* não é nem disciplina, nem metodologia, como M. Hoey, conforme afirma Sardinha (2004), a Linguística

de *Corpus* é uma maneira de se chegar à linguagem. Sob essa ótica ela não seria apenas um instrumental, mas sim uma abordagem baseada em *corpus*.

Finalizando nossa reflexão sobre os princípios gerais da Linguística de *Corpus*, abordaremos na subseção que segue as características de um *corpus*. Nesse contexto, falaremos de representatividade, extensão, especificidade e adequação.

2.1.3.2 Características de um *corpus*

A representatividade é uma das funções de um *corpus*. Seja de que tipo for, um *corpus* é entendido como representativo da linguagem, de um idioma, ou de uma variedade dele. A característica mais fortemente associada à representatividade é a extensão. Embora a extensão seja um critério fundamental na representatividade de um *corpus*, a definição de critérios mínimos de extensão ainda precisa ser melhor discutida. Ainda que não haja critérios objetivos para a determinação da representatividade, esta deve ser vista sob dois enfoques: o léxico e o semântico.

Quanto ao léxico, é possível diferenciar as palavras entre as de maior frequência e as de menor frequência, sendo que a diferença entre elas é relativa. Certas palavras têm frequência de ocorrência rara; para que a probabilidade de ocorrerem no *corpus* se efetive, é necessário incorporar uma grande quantidade de palavras. De certa forma, isso significa dizer que para ter representatividade, o *corpus* deve ser o maior possível.

Desse modo, quanto maior o número de palavras, maior é a probabilidade de ocorrerem palavras de baixa frequência. Esse entendimento decorre da percepção de que a linguagem é um sistema probabilístico, no qual certos traços são mais frequentes que outros. Como a frequência em si não é suficiente, uma vez que as palavras podem possuir mais de um sentido, entra na discussão da representatividade, a questão do sentido das palavras. Um *corpus* representativo deve conter o maior número possível de sentidos de cada forma.

Sob o ponto de vista semântico, ou seja, do sentido das palavras, é possível distinguir entre os sentidos mais frequentes e menos frequentes dos itens lexicais. Há que ressaltar que até mesmo palavras de alta frequência apresentam sentidos raros, como por exemplo, ‘serviço’ no sentido de saque de tênis. Sob esse aspecto, quanto maior for o *corpus*, maior será a probabilidade de palavras como ‘serviço’, ocorrerem. Assim sendo, representatividade e probabilidade estão fortemente associadas.

Para Biber (2012), a representatividade de um *corpus* não deve ser medida apenas pelo seu tamanho, mas também pela sua variabilidade.

Algumas das primeiras considerações na construção de um *corpus* dizem respeito ao planejamento geral: por exemplo, os tipos de textos que o compõem, o número de textos, a seleção de determinados textos, a seleção de amostras retiradas de textos e o tamanho das amostras de texto.

[...]

Normalmente, os pesquisadores priorizam o tamanho das amostras como o aspecto mais importante para obter representatividade: o número de textos que devem ser incluídos no *corpus* e o número de palavras por amostra. [...] A representatividade se refere ao quanto uma amostra inclui de toda a gama de variabilidade de uma população. No planejamento do *corpus*, a variabilidade pode ser considerada a partir das perspectivas situacionais e linguísticas, pois ambas são importantes na determinação da representatividade. Assim, o planejamento de um *corpus* pode ser avaliação com base no quanto ele abarca dos seguintes itens: (1) a variedade dos tipos de textos em uma língua e (2) a variedade de distribuições linguísticas em uma língua. (BIBER, 2012, p. 11-12).

Um *corpus* é uma amostra de uma população cuja dimensão – a linguagem como um todo – não se conhece. Assim sendo, não é possível estabelecer qual seria o tamanho ideal de amostra, para que seja representativa dessa população. Segundo Berber Sardinha (2004), quanto ao tamanho, um *corpus* pode ser classificado em: pequeno (mais de 80.000 palavras), pequeno-médio (de 80.000 a 250.000 palavras), médio (de 250.000 a 1 milhão de palavras), médio-grande (de 1 milhão a 10 milhões de palavras), e grande (de 10 milhões ou mais).

Por ser a linguagem de caráter probabilístico, há a possibilidade de haver uma relação entre traços que são mais comuns e menos comuns em determinado contexto.

O conhecimento da probabilidade de ocorrência de traços lexicais, estruturais, pragmáticos e discursivos está no cerne da Linguística de *Corpus* e, portanto, o conhecimento acerca da probabilidade de ocorrência da maioria dos traços linguísticos em vários contextos ainda está sendo adquirido (BERBER SARDINHA, 2004, p. 24).

Observando um *corpus* do ponto de vista do léxico, é possível perceber que o léxico de frequência baixa é o mais comum, pois a maioria das palavras de uma língua ocorre poucas vezes. Palavras de baixa frequência têm uma probabilidade baixa de ocorrência, ainda que formem a maior parte do vocabulário de uma língua.

Quanto à extensão, para a constituição de um *corpus* representativo, um *corpus* pode ser definido a partir de três abordagens: impressionística, histórica e estatística. A abordagem impressionística é fundamentada em constatações geralmente feitas por profissionais da área com origem na prática da criação e da exploração de *corpora*. A histórica é calcada na monitoração dos *corpora* usados efetivamente pela comunidade. E a estatística, na aplicação de teorias estatísticas.

Quanto à extensão, ainda, um *corpus* comporta três dimensões, a saber:

1) número de palavras, no sentido de que quanto maior o número de palavras, maior será a possibilidade do *corpus* conter palavras de baixa frequência, as quais compõem a maioria das palavras de uma língua;

2) número de textos, entendendo que um número de textos maior garante uma representatividade melhor quanto ao gênero, registro ou tipo textual;

3) número de gêneros, registros ou tipos textuais.

Uma das formas de alcançar a representatividade de um *corpus* é pela possibilidade de restringir seu conteúdo, tendo como propósito abordagens específicas, como um determinado autor, tema ou características especiais da linguagem, por exemplo. Nesse caso, a representatividade é alcançada pela especificidade do *corpus*. Uma outra forma de se alcançar especificidade é delimitando sua variedade, como tipo de texto, por exemplo. “Uma variedade específica da linguagem demonstra maior padronização e conseqüente menor variação no nível do léxico, da gramática, do discurso. Ou seja, apresenta maior grau de fechamento” (BERBER SARDINHA, 2004, p. 28).

Outro critério fundamental na composição de um *corpus* é a adequação: “Para serem adequados, os *corpora* devem ser afinados com os objetivos de análise” (BERBER SARDINHA, 2004, p. 29). Essa característica atinge principalmente os usuários do *corpus*, embora envolva também seus criadores.

Mesmo entendendo que um *corpus* deve ser representativo, é preciso considerar seus limites. Assim sendo, a questão da pesquisa precisa ser priorizada em relação ao objeto. Desse modo, um *corpus* além de ser representativo precisa estar adequado aos interesses do pesquisador, que necessita de uma questão a investigar, e para a qual precisa de um *corpus* específico.

A pesquisa perde o sentido, se o *corpus* com o qual se está lidando e as perguntas feitas não estiverem adequados aos propósitos da investigação. A adequação do *corpus* tem como consequência o questionamento da validade de *corpora* gerais. *Corpus* como o British National *Corpus* (BNC), representante do inglês britânico falado e escrito, tem uma característica importante, ou seja, a de estar disponível para a comunidade acadêmica e, assim, cumprir seu papel de fontes de dados verificáveis.

O British National *Corpus*, a saber, pode ser usado para a criação de um *subcorpus* especializado. Esse procedimento tem a vantagem de dispensar a coleta de um *corpus* novo e, além disso, de dispor de textos anotados e etiquetados gramaticalmente, o que poupará tempo e recursos. Ressaltamos, no entanto, que *corpora* compilados em pequena escala, por

pesquisadores individuais, para propósitos específicos, tendem a serem mais representativos e produtivos do que os *subcorpora* dos *corpora* gerais.

Apresentado as características de um *corpus*, abordaremos na próxima subseção as ferramentas utilizadas no processamento de *corpora* textuais.

2.1.3.3 Ferramentas para a elaboração de *corpora*

Em Linguística de *Corpus*, ferramentas são instrumentos utilizados para a pesquisa linguística informatizada. Para a exploração de *corpora* há a disposição ferramentas de livre acesso, como: *AntConc*, *CEPRIL*, *LAEL*, *Corpógrafo*, *LácioWeb*, *Simple Concordance Program* e *Caixa de Ferramentas TEXTQUIM*. As ferramentas computacionais são utilizadas para reorganização e extração de informações no *corpus*; elas permitem a observação e a interpretação de dados, viabilizando novos olhares para a análise linguística.

Dentre as ferramentas disponíveis para a análise de textos – processamento de *corpora* textuais –, o *WordSmith Tools* (WS Tools) é uma das mais amigáveis e das mais usadas no mundo todo. É um programa pago, desenvolvido para auxiliar o pesquisador nas tarefas da análise linguística. Ele coloca à disposição do analista uma série de recursos extremamente úteis e poderosos para a análise de vários aspectos da linguagem, como a composição lexical, a temática de textos selecionados, e a organização retórica e composicional de gêneros discursivos.

O *WordSmith Tools* vem sendo usado em pesquisas nas mais diversas áreas, como a tradução, ensino de língua, lexicografia e Estudos de Terminologia, dentre outras. É dedicado a análises lexicais como colocações, agrupamentos e dados estatísticos relacionados a esses e outros itens. É uma ferramenta que se apresenta como sendo de grande utilidade para o desenvolvimento de glossários de áreas de especialidade. Entretanto, uma das limitações para a utilização desta ferramenta é o fato de ela não ser de livre acesso, ou seja, de ser um *software* comercial, o que implica em custo pela sua utilização.

Conforme Maciel (2009), as ferramentas básicas utilizadas no processamento de *corpora* textuais são: o concordanciador, o listador de palavras e o listador de n-gramas.

Concordanciadores são programas que permitem ao usuário a busca por palavras específicas em um *corpus*, oferecendo exaustivas listas para as ocorrências das palavras em contexto. O concordanciador lista as ocorrências no texto de uma determinada palavra com seu contexto imediato. A pesquisa pode ser feita a partir de palavras, ou tendo como elemento de busca uma porção da palavra seguida de asterisco (*). Uma pesquisa com elemento de

busca ‘lingu*’, gerará uma lista de ocorrências com palavras como: língua, linguística, linguagem, interlíngua, etc. acompanhada das palavras que a rodeiam, o que permite ver o seu uso e seu contexto de ocorrência. Assim, o resultado é uma lista de todas as palavras presentes no *corpus* em análise, acompanhadas do contexto em que se apresentam.

O listador de palavras é a lista de frequência de uma palavra dentro do *corpus*. Ele lista individualmente todas as palavras presentes no texto e/ou no *corpus* por ordem alfabética ou de frequência; ele faz a contagem das palavras em um *corpus*.

O gerador de N-Gramas lista as ocorrências de grupos de palavras repetidos ao longo de um texto ou *corpus* com indicação do seu número de frequência, o que permite a análise de um bigrama ou trigrama, por exemplo; a extensão do grupo pode ser previamente determinada. Bigrama é um grupo de palavras composto por duas unidades, como por exemplo, ‘linguística de’, ‘gramática de’, ‘contexto de’. Trigrama é um grupo de palavras composto por três unidades, tal como, ‘contexto de situação’, ‘contexto de enunciado’.

Ainda segundo Maciel (2009), além das ferramentas básicas, há outras ferramentas, tais como: etiquetador, concordanciador alinhado e alinhador de textos.

Os etiquetadores fazem análises automáticas do *corpus* e inserem etiquetas (códigos) de ordem morfosintática, sintática, semântica ou discursiva.

O concordanciador alinhado é a relação das ocorrências de uma palavra em contextos dentro do *corpus*. Ele busca uma palavra em textos paralelos alinhados em pares compostos da língua original e sua tradução. Assim como ocorre com a pesquisa com o concordanciador, a pesquisa com o concordanciador alinhado pode ser feita com o uso do asterisco (*), que pode ser colocado antes ou depois da porção da palavra que se deseja buscar a concordância.

Um alinhador de textos tem como propósito alinhar um conjunto de *corpora* paralelamente.

Feitas essas considerações sobre Linguística de *Corpus*, concluímos a revisão de literatura no âmbito dos Estudos da Linguagem. Daremos continuidade à nossa investigação examinando a Ciência da Informação a partir de duas subseções: 1) Organização do Conhecimento e Representação da Informação, e 2) Sistemas de Recuperação da Informação. Nosso propósito ao revisar esses temas é o de buscar um novo referencial no âmbito da Ciência da Informação.

Visando oferecer uma melhor compreensão da área Ciência da Informação por parte de leitores não familiarizados com os conceitos deste domínio, apresentaremos um apêndice contendo um glossário com os termos mais relevantes mencionados neste texto. (APÊNDICE B – Glossário de termos de Ciência da Informação).

2.2 Ciência da Informação

Início esta seção com o propósito de esclarecer as relações que ocorrem entre os domínios da Biblioteconomia, Documentação e Ciência da Informação, usualmente entendidas sob a ótica da organização da informação.

As primeiras evidências de organização de documentos surgem com as bibliotecas da antiguidade; elas dão origem ao que mais tarde passa a ser concebido como Biblioteconomia. Em sentido estrito, Biblioteconomia pode ser entendida como a área que se preocupa com a organização, gestão e disponibilização de acervos de bibliotecas.

Durante muito tempo, Biblioteconomia e Documentação apresentaram um desenvolvimento similar. Elas surgiram em razão das mesmas necessidades, utilizavam processos e instrumentos comuns e tinham objetivos muito parecidos. A Documentação diferenciava-se da Biblioteconomia porque tinha como propósito fazer uma análise de conteúdo mais profunda. Havia entre os documentalistas uma tentativa de afastar-se dos procedimentos adotados em Biblioteconomia, chegando mesmo a seguir caminhos já trilhados e até mesmo descartados por ela.

A divergência entre bibliotecários e documentalistas era tanta que chegou a se refletir na segmentação das associações destes profissionais. Com o tempo, a cisão entre Biblioteconomia e Documentação tornou-se cada vez mais profunda.

Por mais de quatro séculos, Biblioteconomia e Bibliografia assim como Bibliografia e Documentação foram confundidas em seus princípios teóricos. Foi no século XIX, quando Paul Otlet e La Fontaine sistematizaram e desenvolveram a Documentação, que Biblioteconomia e Documentação começaram a se distinguir uma da outra. As pesquisas contemporâneas devem muito aos documentalistas, uma vez que as práticas documentárias foram as primeiras a contemplar o tratamento da informação.

O termo Documentação foi cunhado por Otlet e La Fontaine, e tinha como propósito significar de forma mais ampla o que antes era denominado de Bibliografia. Por sua importância na área documentária, Otlet vem sendo considerado o precursor e fundador da Documentação e, até mesmo, da Ciência da Informação.

Segundo Ortega³⁷ (2004) a Biblioteconomia deu origem à Bibliografia, que por sua vez fundamentou a Documentação, a qual forneceu insumos à Ciência da Informação. Em sua

³⁷ Original obtido na Internet.

opinião Biblioteconomia, Documentação e Ciência da Informação se relacionam conceitualmente.

A Biblioteconomia é a atividade mais antiga a se preocupar com a organização de documentos e com a preservação das unidades de conhecimento registrado. Com a democratização do acesso à educação e à cultura, essa tendência se altera. A Biblioteconomia não se constitui como uma área científica, ainda que tenha incorporado as novas tecnologias à sua rotina, como o uso de redes cooperativas no âmbito da catalogação com propósitos produtivos e formais; a informação não era o foco.

A Documentação, por sua vez, considerada uma dissidência da Biblioteconomia, é voltada para o tratamento documental. Ela se preocupou com o desenvolvimento de técnicas e princípios relacionados com a organização e recuperação da informação. Ela trabalha com a otimização de processos e técnicas documentárias. A Documentação alcançou significativos graus de sistematização de princípios e modelos, mas entre os anglo-saxões ela foi praticamente substituída pela Ciência da Informação.

A Ciência da Informação tem como objeto de estudo a informação; inicialmente restringia-se à informação técnico-científica, mas com o tempo passou a contemplar também a informação para fins educacionais, sociais e culturais. Ela reflete as alterações que se instauraram no século XX pela comunicação, pela tecnologia eletrônica e pelos fluxos de informação. Conforme Ortega³⁸ (2004) ela

[...] pode ser definida como a área que estuda a teoria, os métodos e as práticas da produção, organização, armazenamento, recuperação, disseminação e promoção do uso da informação, processos esses compreendidos no âmbito dos fluxos comunicacionais [...].

Esta área de conhecimento tem sua história marcada pelas constantes dissidências dos profissionais da área e pela dificuldade de elaborar seus fundamentos teóricos. Ela teve início no final dos anos 50, em consequência do surgimento dos computadores, e ainda hoje busca a consolidação da área, uma vez que permanece em constante processo de discussão. Ela tem suas raízes na cisão da Documentação/Bibliografia e Recuperação da Informação.

Nos Estados Unidos, no final da década de 60, a Ciência da Informação teve um grande desenvolvimento. Inicialmente, por influência norte-americana, era empregada para designar a Biblioteconomia do tipo não tradicional, estabelecendo-se, assim, uma relação entre Ciência da Informação e Biblioteconomia.

³⁸ Original obtido na internet.

Um dos fatores que explica a criação da Ciência da Informação em paralelo à Documentação foi os bancos de dados que ofereciam informações, em oposição à Documentação que oferecia referências de documentos a serem consultados.

Entendida como ciência pós-moderna, a Ciência da Informação se caracteriza pela interdisciplinaridade. Ela estabelece interface com a Biblioteconomia e com a Ciência da Computação, dentre outras.

Segundo Ortega³⁹ (2004),

sendo a Biblioteconomia, a atividade mais antiga de organização de documentos, encontra na Ciência da Informação a possibilidade de construção de referenciais teóricos e de conquista de *status* científico, enquanto esta encontra naquela parte da história e das práticas que compõem aquilo que vem elaborando a partir de diversas disciplinas e aplicações. Já a Documentação, considerada em separado da Biblioteconomia, desenvolveu princípios e técnicas voltadas à organização e recuperação da informação, independente dos suportes e tipos documentais e com base nos contextos de aplicação e tipos de informação. Neste sentido, os princípios documentários permitem à Biblioteconomia maior abstração e adequação na elaboração de seus processos e serviços, e fornecem à Ciência da Informação insumos para uma construção científica sólida, ao conduzir a um foco ou núcleo de referência para a alocação integrada das demais disciplinas e aplicações.

Concluído esses esclarecimentos sobre Biblioteconomia, Documentação e Ciência da Informação, iniciaremos a subseção seguinte abordando o tema da organização do conhecimento e da representação da informação.

2.2.1 Organização do Conhecimento e Representação da Informação

O tema Organização do Conhecimento e Representação da Informação tem sido objeto de inúmeros estudos entre os profissionais da informação. Isso se deve ao número crescente de dados disponibilizados em redes e sistemas de informação, e da dificuldade na busca da informação e recuperação de documentos ali armazenada.

O aumento exponencial dos dados disponíveis em redes e sistemas de informação tem tido uma importância significativa para as técnicas de Organização do Conhecimento e da Representação da Informação. Diversos tipos de estruturas são utilizados com esse propósito. Os cabeçalhos de assunto e os esquemas de classificação ou taxonomias são estruturas que se organizam a partir da classificação e da criação de categorias. As ontologias e os tesouros, de outro modo, se organizam a partir de conceitos e de seus relacionamentos.

³⁹ Original obtido na internet.

Organização do conhecimento segundo Bräscher e Café (2008, p. 8) é o

processo de modelagem do conhecimento que visa a construção de representações do conhecimento. Esse processo tem por base a análise do conceito e de suas características para o estabelecimento da posição que cada conceito ocupa num determinado domínio, bem como das suas relações com os demais conceitos que compõem esse sistema nocional. Como afirma Vickery (2008), organizar o conhecimento é reunir o que conhecemos em uma estrutura sistematicamente organizada.”

Por representação da informação, entendemos o processo que implica em reproduzir o conteúdo dos documentos, visando sua recuperação e circulação por parte de diferentes tipos de usuários. Ela deve estar fundamentada em Linguagens Documentárias, e tem por função permitir que a informação armazenada em Sistemas de Recuperação da Informação (SRIs) ofereça uma linguagem unívoca, para ser compartilhada entre seus interlocutores.

Para Bräscher e Café (2008, p. 92), que entende que a organização da informação (OI) é um processo que envolve a descrição física e de conteúdo dos objetos informacionais, a representação da informação, produto desse processo descritivo, é

[...] entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. Alguns tipos de representação da informação são construídos por meio de linguagens elaboradas, especificamente para os objetivos da OI. Essas linguagens, segundo Svenonius (2000) subdividem-se em linguagens que descrevem a informação e linguagens que descrevem o documento (suporte físico).

A dispersão terminológica – conceitual e denominativa – é uma marca deste subcampo da Ciência da Informação. Conforme mencionado por Lara (2011), encontramos na literatura brasileira o uso simultâneo dos seguintes termos: organização da informação e organização do conhecimento, sendo que ora aparecem associados, ora são usados para expressarem coisas distintas. O termo representação, por sua vez, é frequentemente utilizado com um sentido muito próximo ou, até mesmo, como sinônimo de organização da informação.

Ainda que o Grupo de Trabalho da ANCIB – Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação – adote o termo organização e representação do conhecimento desde 2006 para representar os estudos desenvolvidos sob esse enfoque, optamos nesta pesquisa pelo termo Organização do Conhecimento e Representação da Informação, pelo próprio fato de entendermos de modo diverso, conhecimento e informação⁴⁰.

As reflexões feitas na sequência deste texto se desenvolverão a partir de duas subseções: Linguagens Documentárias e Ontologias.

⁴⁰ Sobre o nosso entendimento em relação à conhecimento e informação, ver subseções 2.2.2.5 Dado, Informação e Conhecimento.

Nossa investigação sobre essas linguagens e os problemas que delas decorrem se desenvolverá sob as seguintes subseções: Princípios teóricos; Sistema nocional; Classificação quanto às formas de representação dos conceitos; Vocabulários controlados; Fenômenos linguísticos, Análise Documentária; Metodologias; Semelhanças e diferenças entre taxonomias, tesouros e ontologias.

Sobre o tema das ontologias apresentaremos algumas considerações sobre os conceitos básicos em que se sustentam, suas características, tipologia e ferramentas para a sua construção. Com tal propósito desenvolveremos duas subseções: 1) Definição, finalidade, características e a relação das ontologias com os mapas conceituais, e 2) ExATOlp – Extrator Automático de Termos para Ontologias em Língua Portuguesa. Nesta subseção, ofereceremos algumas informações sobre este *software*, o qual propiciará a extração automática de termos do *corpus* constituído para a construção do vocabulário controlado especializado oferecido nesta pesquisa, enquanto projeto piloto. Nosso propósito é o de recomendar a Ontologia como uma ferramenta capaz de agilizar a extração de termos de um *corpus*, visando à construção de vocabulários controlados especializados.

2.2.1.1 Linguagens Documentárias

Principiamos esta subseção ressaltando o fato que as Linguagens Documentárias (LDs) se constituem em um instrumento importante nos processos de Organização do Conhecimento e Representação da Informação. Elas precisam ser formuladas com rigor, pois são elas que permitem que o usuário acesse um Sistema de Informação e recupere a informação que deseja. Elas se situam no quadro das representações documentárias, como uma das formas de representação da informação.

Compartilhamos com Lara (2009, p.49), quando afirma que

[...] as representações documentárias são versões sintéticas dos documentos e, embora não o substituam, sinalizem para o seu conteúdo. Do ponto de vista do processo de análise documentária a representação é feita documento a documento; já do ponto de vista do sistema de informações ao qual ele será integrado, a representação não é individualizante, mas remete à classe de documentos que compartilham daquela representação.

As Linguagens Documentárias estão presentes nos Sistemas de Recuperação da Informação (SRIs), nos catálogos eletrônicos, nas ontologias, no sistema *web*, em resumos, índices, bibliografias e inventários. Além de representarem e recuperarem a informação, elas cumprem a função de circular a informação. No ciclo documentário, elas se inserem na etapa de tratamento da informação. Para a Ciência da Informação, ciclo documentário – documento,

coleta, tratamento, difusão e usuários – é a circulação de informações e documentos em um Sistema de Informação.

Para Lara (2011) as Linguagens Documentárias são linguagens intermediárias, uma vez que compartilham simultaneamente da Linguagem Natural e da Linguagem Artificial. No seu entendimento, a Linguagem Documentária é um sistema de significação e de comunicação, cuja função primordial é realizar a intermediação da linguagem.

Na ótica desta pesquisa, enquanto sistema de comunicação ela deve estar fundamentada nas noções de subjetividade e de intersubjetividade. Acreditamos que sob esse referencial, ela terá condições para viabilizar uma relação de interação comunicativa entre os interlocutores de um Sistema de Recuperação da Informação – bibliotecários e usuários, e usuários e Sistema –.

Com o propósito de assegurar, com o mínimo de perda possível, o sentido de um descritor na passagem do processo de indexação para o processo de busca da informação e recuperação de documentos, ela deve estar fundamentada nos princípios teóricos da construção do sentido.

Nesta subseção, e no âmbito da Biblioteconomia, termo e descritor são usados como equivalentes. A Ciência da Informação, através da Documentação, e mais especificamente das Linguagens Documentárias, entende o descritor ou termo como unidade de informação e signo de comunicação. Esse entendimento difere daquele adotado pela Terminologia, que vê o termo em uma perspectiva mais ampla, uma vez que privilegia o termo como unidade de conhecimento, unidade linguística e unidade de comunicação. Dando continuidade à nossa investigação sobre as LDs, faremos uma breve exposição sobre os princípios teóricos em que se sustentam.

2.2.1.1.1 Princípios teóricos

Iniciamos esta explanação, afirmando que as Linguagens Documentárias têm dois propósitos principais: 1) organizar ou categorizar a estrutura conceitual de um campo científico, técnico ou especializado; e 2) viabilizar a circulação da informação em um processo que envolve bibliotecários, SRIs e usuários, tendo como ponto de convergência a linguagem do documento. São as Linguagens Documentárias que viabilizam formas de acesso aos SRIs, e propiciam relações de interação entre bibliotecários e usuários e usuários e Sistemas.

Segundo Lara (2004, p. 232)

A denominação Linguagem Documentária (LD), além de referir-se ao conjunto dos diferentes tipos de instrumentos especializados no tratamento da informação bibliográfica (sistemas de classificação enciclopédico ou facetados e tesouros), designa, de modo mais amplo e completo, a linguagem especialmente construída para organizar e facilitar o acesso e a transferência da informação.

Ainda referenciando a autora, a Linguagem Documentária “[...] deve dispor os seus elementos, uns em relação aos outros, para que, no conjunto obtenha-se um sistema uno e dotado de significação” (LARA, 2004, p. 233). Desse modo ela oferece as condições que possibilitam as operações de representação e de acesso à informação. Assim, além de propor-se como um modo de organização do conhecimento de um domínio, ela propicia a circulação da informação.

Cumprindo a função de expressar a linguagem que identifica o conjunto de conceitos de uma área de conhecimento, uma Linguagem Documentária não tem a finalidade de exibir o conteúdo de um texto específico, mas a de estruturar e representar o campo conceitual ao qual se refere, sendo, por isso, de natureza eminentemente referencial. Essa organização do conhecimento, feita a partir da estruturação de um sistema nocional, aparece sob a forma de vocabulários controlados.

Com o papel de viabilizar a circulação da informação, bibliotecário, Sistema e usuário desempenham papéis diferenciados. O bibliotecário precisa selecionar, adequadamente, os termos de indexação utilizados para a representação da informação expressa em um documento. O usuário necessita formular expressões de busca compatíveis com a linguagem utilizada pelo bibliotecário. O Sistema deve armazenar as informações processadas pelo bibliotecário, e oferecer condições facilitadas de uso, para que o usuário recupere as informações que deseja. É na relação que se estabelece entre bibliotecário, Sistema e usuário, que se caracteriza a circulação da informação. Com o bibliotecário temos a entrada dos dados/informação no Sistema (processo de indexação / documento processado). Com o usuário obtemos a saída dos dados/informação do Sistema (processo de busca da informação e recuperação de documentos / documento localizado).

Tendo como função o tratamento da informação, o principal objetivo das LDs é alcançar eficiência nos processos de indexação e de busca da informação e recuperação de documentos. Ao disponibilizar a informação registrada nos acervos aos diferentes segmentos de usuário, as LDs tornam viável a socialização da informação.

Na tarefa de divulgar e socializar a informação, o bibliotecário enfrenta sérias dificuldades. Isso se explica porque, ao indexar um documento, é preciso assegurar que o sentido pretendido pelo autor do texto se perca o mínimo possível, de forma que as

informações ali contidas possam ser recuperadas pelo usuário, com alto grau de precisão e o menor nível de ruído possível.

O ruído, assim como o silêncio, são fatores que interferem na circulação da informação, e que dificultam o acesso a ela. O ruído ocorre quando o número de documentos recuperados não é pertinente às necessidades de pesquisa do usuário; ele resulta do uso de termos de indexação não apropriados ao documento. O silêncio se apresenta quando os documentos processados em um SRI não são recuperados; ele é consequência da omissão de termos de indexação representativos da temática do documento.

As Linguagens Documentárias tradicionais são linguagens construídas. Elas têm o propósito de representar um conjunto de objetos organizados à maneira de inventário. Sob esse entendimento elas configuram o acervo documental como um conjunto de objeto/livro e não de significado/contexto. Nesse caso, a codificação numérica serve mais para objetivos topográficos do que propriamente intelectuais.

Na tentativa de contemplar a questão da interdisciplinaridade, foram desenvolvidos sistemas de classificação facetada. Nestes sistemas, os assuntos são divididos em categorias ou facetas, isto é, em grupos de classes reunidas por um mesmo princípio de divisão, ou seja, por assuntos que apresentam as mesmas características, e que têm o mesmo tipo de relacionamento entre si e com a classe principal.

Uma Linguagem Documentária se caracteriza por estabelecer relações entre os descritores/termos. Cintra et al. (2002) ressaltam que em uma LD não pode existir duas ou mais palavras para designar um mesmo conceito, nem uma palavra para se referir a mais de um conceito. Considerando que para cada descritor de um tesouro é atribuído um mesmo conceito, de maneira unívoca, entendemos que o sentido de um termo precisa ser fixado pelo Sistema de Informação. Isso significa dizer que um Sistema precisa ser estruturado, considerando como substancialmente relevante a noção de referência, conforme apresentada por Benveniste em sua teoria. Ressaltamos, aqui, o fato de que ainda que a utilização de um descritor seja prescrita por uma Linguagem Documentária, em tesouros diferentes, um mesmo descritor pode designar conceitos diversos.

As LDs estão presas a diversas convenções, como por exemplo, a de que atribuir a uma unidade linguística o estatuto de descritor/termo é convencionar que seu uso está autorizado. Elas utilizam uma série de procedimentos e de instrumentos para operacionalizar a tarefa de indexar documentos. É sob o aspecto do conteúdo temático que reside o grande desafio enfrentado pelo profissional da Biblioteconomia, e que motiva o desenvolvimento desta pesquisa.

[...] 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; FINATTO, 2004, p. 60).

Diferentes denominações, advindas de diferentes orientações teórico-metodológicas, são atribuídas a essas linguagens, dentre elas, as seguintes: 1) linguagens documentárias (J.-C. Gardin, M. Coyaud e J. Chaumier), 2) linguagens de indexação (J. Melton), 3) linguagens descritoras (B. Vickery), 4) codificações documentárias (E. Grolier), 5) linguagens de informação (D. Soergel), 6) vocabulários controlados (F.W. Lancaster), 7) lista de assuntos autorizados (C. Montgomery), 8) linguagens de recuperação da informação, e 9) linguagens de descrição da informação.

Na comparação das Linguagens Documentárias com a linguagem natural, entendemos que ambas se caracterizam por exercerem uma função comunicativa, e se estruturam de modo sistêmico, pois tanto uma unidade linguística como uma unidade documentária não pode ser entendida em separado.

O caráter sistêmico fica garantido com a impossibilidade de se ler uma unidade em separado. De fato, cada unidade só pode ser ‘lida’ na sua relação com as demais unidades componentes do sistema.

Por serem sistemas construídos, as LDs são econômicas. Só que não se trata da aplicação do princípio de economia da LN [linguagem natural] e sim de uma racionalização de escolhas e de procedimentos, que permitam uma utilização eficaz do sistema.

As relações paradigmáticas e sintagmáticas também ocorrem, só que de forma bastante restrita, especialmente nas construções dos sintagmas.

No entanto, nas LDs fica evidente o poder de troca das unidades, numa posição bastante próxima da LN. Cada unidade documentária designa uma realidade dentro do sistema construído, o que torna evidente o valor e a possibilidade de troca, de representação (CINTRA et al., 2002, p.16).

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 extralinguísticas. Sob esse ponto de vista, entendemos que as Linguagens Documentárias precisam ser compreendidas a partir do conceito de estrutura, no sentido concebido por Benveniste em sua teoria, na sua visão sobre linguagem e língua, e conforme apresentada na subseção 2.1.1 Teoria da Enunciação de Benveniste.

Segundo Ortega e Lara (2010, p. 8) “[...] a noção de estrutura, ao pressupor relações mútuas entre os seus elementos, confere maior organicidade aos sistemas documentários e

garante maior qualidade e consistência na recuperação da informação frente às questões dos usuários”. Ainda segundo as autoras citadas, a noção de estrutura precisa ser entendida tanto em relação aos registros de informação, quanto aos sistemas que compõem.

Os registros de informação dos sistemas documentários devem ser produzidos a partir de representações de entidades e de questões de busca (previstas) dos usuários. Para sua funcionalidade, é esperado que os registros apresentem características que garantam aderência entre essas representações e as questões de busca efetivamente realizadas pelos usuários (ORTEGA; LARA, 2010, p. 8).

São consideradas entidades, no contexto dos Sistemas de Informação, tanto os documentos, como os grupos de documentos e/ou partes de documentos que compõem uma coleção.

A Linguagem Documentária, tal qual a linguagem natural, também se caracteriza pela presença de um léxico, de uma rede paradigmática e de uma rede sintagmática. Nas LDs o léxico é representado por uma lista de descritores, entendidos nesta pesquisa como ULRIEs. As relações estabelecidas entre essas unidades, no eixo paradigmático das Linguagens Documentárias, são essenciais e geralmente estáveis; elas correspondem às relações do tipo lógico-semântico. No eixo sintagmático, tais relações são contingentes e válidas no contexto específico onde aparecem. “[...] o eixo sintagmático da maior parte das LDs é muito precário se comparado ao das Línguas Naturais, pois os recursos de combinação entre os termos são muito restritos (nos tesouros, por exemplo, a sintaxe é realizada através de operadores booleanos: and, or, not)” (LARA⁴¹, 2001).

Equivalendo às relações paradigmáticas da linguagem natural, estão as relações associativas de uma LD; elas envolvem o conceito de substituição. As relações sintagmáticas, por sua vez, permitem a combinação de palavras e a representação das informações contidas nos documentos, correspondendo, assim, às unidades de conhecimento destas linguagens.

Por ser um processo complexo, é preciso considerar alguns aspectos na construção de uma Linguagem Documentária, tais como: 1) um sistema padronizado, por si só, não é garantia de sucesso na recuperação da informação; 2) nas comunicações especializadas também é reconhecida a existência da variação terminológica conceitual e denominativa; 3) o apagamento da diversidade denominativa não evita, obrigatoriamente, os ruídos no processo comunicacional; 4) 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.

⁴¹ Original obtido na internet.

Segundo Lara (2004), a complexidade que as Linguagens Documentárias apresentam tem origem na articulação de diferentes códigos: 1) o código da língua, 2) no subcódigo de uma área de especialidade e 3) no subcódigo da Ciência da Informação.

O código da língua remete a um saber prévio, implícito. É no código da língua que a língua se constitui em condição primordial para que a comunicação se efetive. Para que as informações sejam veiculadas, é necessário dominar este código.

O subcódigo de uma área de conhecimento envia a significados específicos. É na linguagem de especialidade e em um contexto determinado que uma palavra adquire um significado preciso.

O subcódigo da Ciência da Informação tem como propósito facilitar o acesso à apropriação da informação. Ele se caracteriza pelo modo como são feitos os recortes nas áreas de especialidade e na linguagem natural, convertendo em unidades de informação, as unidades desses dois códigos – código da língua e subcódigo de uma área de especialidade –. A delimitação de tais unidades, expressas em Linguagem Documentária, depende dos objetivos institucionais e das características dos usuários ao qual o sistema de informação se destina. Ainda que reconheçamos que o Sistema possa ser acessado por um público heterogêneo, é possível identificar certa homogeneidade neste público, em razão da condição de inteligibilidade relacionada às características específicas dos usuários.

Para que cumpra sua razão de existir a LD deve estar apta a ser permanentemente atualizada, de modo a se constituir em um instrumento dinâmico. Com esse propósito, ela 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.

A compreensão que envolve o desenvolvimento de uma LD, 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 buscadas para a construção dessas linguagens, como condição para se obter resultados positivos na busca da informação e recuperação de documentos, 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).

No nosso entendimento, uma Linguagem Documentária, construída com a finalidade de cobrir um domínio específico do conhecimento, tem mais probabilidade de viabilizar a representação de assuntos pela sua especificidade. Nesse direcionamento, a crescente tendência à especialização tem levado a abandonar uma intenção primeira de cobrir o universo do conhecimento em sua totalidade. Da constituição da LD, em relação ao modo como foi concebida em um sistema nocional de um domínio, vai depender o grau de especificidade temática a ser atribuído a um documento. Linguagens Documentárias construídas com o objetivo de abarcar uma extensa área do saber são predominantes na área das Ciências Sociais e Humanidades; essa característica se reflete na construção dos catálogos eletrônicos.

Dos problemas que podem surgir decorrentes de uma LD construída sem propósitos criteriosamente definidos, há a considerar aqueles que se referem: 1) à definição do conjunto de descritores que representarão um campo de conhecimento; 2) à organização dos descritores em uma rede paradigmática para relacioná-las em uma linguagem de especialidade; 3) ao estabelecimento da rede sintagmática para possibilitar uma maior representatividade de novos conceitos; 4) à definição das chaves de acesso ao Sistema, visando à compatibilidade da linguagem do usuário com a do Sistema.

Visando minimizar os problemas enfrentados pelas Linguagens Documentárias, e que afetam a interação que se estabelece entre bibliotecários e usuários, e usuários e Sistemas, apresentaremos na subseção seguinte algumas considerações sobre sistema nocional. É o sistema nocional que determina as relações a serem estabelecidas entre as unidades especializadas de uma área de conhecimento. Estas unidades são utilizadas pelo bibliotecário para descrever o conteúdo dos documentos, com o objetivo de representar a informação indexada para fins de recuperação.

2.2.1.1.2 Sistema nocional

Ao inserirmos nesta pesquisa uma discussão sobre sistema nocional, temos como propósito a compreensão das possibilidades de relacionamentos que se estabelecem entre os descritores/termos no âmbito das relações lógico-semânticas. Iniciamos esta subseção fazendo duas considerações sobre o sistema nocional. Citando Cintra et al. (2002) compartilhamos com a ideia de que o sistema nocional de uma área de conhecimento constitui-se em um

arcabouço fundamental para a construção de uma Linguagem Documentária, na medida em que estabelece os diferentes tipos de relação entre as noções. Entretanto precisamos ter sempre presente, conforme ressalta Barros (2004, p. 108), que um “[...] sistema de conceitos é fruto de uma visão do domínio estudado, podendo estruturar-se diferentemente, de acordo com os critérios empregados”.

Segundo a *Norma ISO 1087* (1990), um sistema nocional ou sistema de conceitos é definido como “Ensemble structuré de notions [...] construit sur la base des relations [...] établies entre ces notions [...] et dans lequel chaque notion [...] est déterminée par sa position dans cet ensemble”⁴². Conceito, por sua vez, é definido como “Unité de pensée constituée par abstraction à partir des propriétés communes à un ensemble d’objets [...]. NOTE – Les notions ne sont pas liées aux langues individuelles. Elles sont cependant influencées par le contexte socioculturel.”⁴³ Ainda citando Cintra et al. (2002) as noções que constituem um sistema nocional são referidas por unidades linguísticas, se constituem nos descritores/termos e são organizadas de forma sistêmica.

A organização nocional de um campo de conhecimento se faz necessária por três razões principais: 1) para a organização do conhecimento de uma área de especialidade, 2) para a compreensão das situações de relacionamento que ocorrem entre os descritores/termos, e 3) para a construção de ferramentas que permitam a eficácia dos processos de indexação e de recuperação da informação. Quando inexistente um sistema nocional, sistematicamente organizado, a Análise Documentária fica sujeita 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, não existe consenso quanto à melhor maneira de estruturá-lo. 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, a organização de um domínio se torna ainda mais difícil.

É inegável que a organização conceitual de um domínio constitui-se em um parâmetro básico de sustentação das Linguagens Documentárias. Entretanto, nesse contexto de conhecimento globalizado, onde o computador permite que os mais complexos sistemas de relacionamentos entre conceitos sejam estruturados, vale refletir sobre os tipos de relação a

⁴² Tradução da autora: Conjunto estruturado de conceitos, construído sobre a base de relações estabelecidas entre esses conceitos, e no qual cada conceito é determinado por sua posição neste conjunto.

⁴³ Tradução da autora: Unidade de pensamento constituída por abstração a partir de propriedades comuns a um conjunto de objetos. NOTA – Os conceitos não são vinculados às línguas específicas. Eles são influenciados pelo contexto sociocultural.

serem estabelecidas. A esse respeito Sager (1990, p. 29) 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.”⁴⁴

Tradicionalmente, entende-se que um sistema nocional é constituído por uma sequência 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 hierárquicas 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.

As relações genéricas são fundamentadas na identidade parcial do conjunto de características envolvidas nas noções superordenadas e subordinadas. As relações específicas, de outro modo, são relações subordinadas que compartilham as mesmas características da noção que lhes é superordenada, mas se diferenciam dela por apresentarem pelo menos uma característica a mais.

Nas relações genéricas, a noção genérica é aquela que é superordenada e que comporta as mesmas características daquelas que lhe são subordinadas. A noção específica, por sua vez, é uma noção subordinada, e se constitui por um processo de separação a partir do conjunto de características comuns. Sendo assim, a noção genérica se estabelece pela soma de características comuns, e a noção específica, por ter as mesmas características da noção genérica, mas também apresentar pelo menos uma característica distinta do grupo de características da noção genérica. Na relação genérica, a superordenação vai das diferenças para as semelhanças, e a subordinação faz o caminho inverso, indo das semelhanças para as diferenças, a partir das primeiras. Em outras palavras, a superordenação vai da espécie para o gênero, enquanto que a subordinação vai do gênero para a espécie. Exemplificando essas relações, tomamos ‘filosofia’ como noção genérica e superordenada em relação à ‘filosofia antiga’, ‘filosofia medieval’, ‘filosofia moderna’ e ‘filosofia contemporânea’. ‘Filosofia antiga’, ‘filosofia medieval’, ‘filosofia moderna’ e ‘filosofia contemporânea’ são noções específicas e subordinadas em relação à ‘filosofia’, mas coordenadas na relação entre si.

A relação partitiva é uma relação hierárquica. Neste tipo de relação, a noção superordenada diz respeito a um objeto considerado como um todo, e as noções subordinadas referem-se a eles – objetos – enquanto partes de um todo. Embora sejam representadas de forma semelhante às relações genéricas, elas não se confundem, porque a característica a ser

⁴⁴ 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.

observada é a da relação entre o todo e as suas partes, sendo que o conceito atribuído à parte depende daquele conferido ao todo. Como exemplo de relações partitivas citamos o corpo humano, enquanto noção referente ao todo e correspondendo à noção superordenada, e braços e pernas, como noções atinentes às partes e, por conseguinte, noções subordinadas.

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. Como exemplo desse tipo de relação, mencionamos regiões montanhosas como noção superordenada, e Andes e Himalaia como noções subordinadas.

As relações não-hierárquicas configuram-se como relacionamentos complexos que não podem ser classificados segundo os critérios de ordem e subordinação lógica. Elas não se submetem a uma hierarquia, mas apresentam entre si contiguidade espacial ou temporal; por tal razão são também chamadas de sequenciais. 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-termos para o conjunto de termos adotados pelo Sistema. Sobre as relações associativas, vale salientar que elas admitem tanto a hierarquia quanto a não-hierarquia.

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 sentido, oferecemos um rigor maior no tratamento da informação e melhores resultados no processo de recuperação. A transformação da unidade de sentido em unidade de representação da informação é a característica primordial de uma Linguagem Documentária.

Finalizando nossas considerações sobre sistema nocional, passaremos a abordar o tema das Linguagens Documentárias sob o ponto de vista da classificação, quanto às formas de apresentação. Essas linguagens podem ser classificadas segundo dois critérios: ordenação dos conceitos e forma de representação. Quanto à ordenação dos conceitos elas podem ser pré ou pós-coordenadas. As linguagens pré-coordenadas são representadas fundamentalmente pelos cabeçalhos de assuntos, e as linguagens pós-coordenadas, pelos tesauros. Nas linguagens pré-coordenadas o indexador determina a ordem de um cabeçalho no momento da

indexação, enquanto que nas linguagens pós-coordenadas é o usuário que combina os assuntos no momento da busca da informação. É quanto à forma de representação dos conceitos que falaremos logo a seguir.

2.2.1.1.3 Classificação quanto às formas de representação dos conceitos

Quanto à forma de representação dos conceitos as Linguagens Documentárias podem se apresentar sob duas formas: ordem sistemática e ordem alfabética. Iniciaremos essa revisão de literatura apresentando a classificação das Linguagens Documentárias sob a ordem sistemática.

2.2.1.1.3.1 Ordem sistemática

A Classificação Decimal de Dewey (CDD), a Classificação Decimal Universal (CDU), a Classificação de Dois Pontos e a Library of Congress (LC) são linguagens que se apresentam sob a forma sistemática.

A Classificação Decimal de Dewey é um sistema de classificação utilizado pelas bibliotecas para organizar livros nas estantes, e foi desenvolvida em 1876 por Melvin Dewey. Ela foi o primeiro sistema de classificação do conhecimento a utilizar números decimais para símbolos de classificação, aplicáveis aos próprios documentos. Este sistema foi inspirado na divisão de conhecimento preconizada por William Torrey Harris (1870). Segundo Piedade (1977), Dewey parece ter sido o primeiro a empregar o princípio da divisão por transferência, expressão usada por Grolier para designar subdivisões do tipo: dividida como.

A Classificação Decimal de Dewey estrutura o universo do conhecimento em dez classes principais. Cada classe é subdividida em outras dez, indo sucessivamente do assunto mais geral ao mais específico. Ela conta com tabelas auxiliares para contemplar diferentes aspectos, tais como: área geográfica, gênero literário, grupos raciais, grupos étnicos, grupos nacionais, línguas e características pessoais. E, ainda, usa números decimais puros para a notação. Notação é o conjunto de símbolos selecionados para representar os termos de classificação, traduzindo em linguagem codificada a temática dos documentos. É usada para indicar a localização dos documentos nas estantes, catálogos e tabelas de classificação. A notação é formada por números decimais e sinais gráficos como o ponto, exclusivamente. Exemplo: 410.3 (dicionário de linguística).

A Classificação Decimal Universal tem uso indicado para bibliotecas especializadas. As edições desenvolvidas oferecem condições privilegiadas para viabilizar a classificação de assuntos especializados. Ela é o segundo sistema de classificação bibliográfica em importância universal. Da mesma forma que a CDD, também divide o conhecimento em dez classes principais, sendo que cada classe é subdividida em outras dez. A notação é formada por números decimais e sinais gráficos, como ponto, dois pontos e apóstrofe, como mostra o exemplo seguinte: 329.11'21 (Partido político com tendências conservadoras e monarquistas: 329.11 e 329.21 – o apóstrofo permite a supressão dos caracteres comuns a dois símbolos não consecutivos).

A CDU foi desenvolvida por Paul Otlet e Henry La Fontaine e é considerada uma expansão da CDD. Graças à utilização de sinais gráficos, ela possibilita uma melhor representação dos assuntos dos documentos, de forma a descrever assuntos compostos ou complexos. A CDU permite a utilização concomitante de duas notações de assuntos principais diferentes, como mostra o seguinte exemplo: 51:52⁴⁵ (Aplicações de Matemática na Astronomia). CDD e CDU são sistemas estruturados linearmente segundo a divisão das ciências pelo princípio gênero/espécie, partindo do universo do conhecimento para assuntos específicos.

A Classificação dos Dois Pontos – Colon Classification – é a teoria desenvolvida por Ranganathan, bibliotecário e professor de matemática. Ele desenvolveu um sistema de classificação com o propósito de estabelecer uma teoria de classificação apropriada para livros, tendo como base o próprio conhecimento. Sua teoria viabiliza a particularização de cada aspecto ou faceta de um assunto, e possibilita a inter-relação dos conceitos, permitindo o acompanhamento da evolução do conhecimento. Ela é utilizada apenas nas bibliotecas indianas, mas tem grande influência sobre aqueles que se dedicam ao estudo de sistemas de classificação. Para a formulação da teoria e desenvolvimento do método, Ranganathan criou uma terminologia própria. Esta teoria é considerada como sendo a produção mais importante da obra deste teórico.

O teórico indiano define cinco categorias ou facetas para expressar o universo do conhecimento: personalidade, matéria (material), energia (ação), espaço e tempo (PMEST). Exemplificando, teríamos: bibliotecário (personalidade), livro (matéria), organização (energia), biblioteca (espaço) e século XX (tempo). Todas as facetas são consideradas como manifestações de uma das cinco categorias fundamentais. O modo como Ranganathan

⁴⁵ Na CDU usam-se dois pontos para documentos que tratam de relacionamentos entre dois ou mais conceitos.

estruturou o conhecimento permite que os assuntos compostos, sinteticamente, surjam a partir de conceitos elementares.

Para Barbosa (1969, p. 166), Ranganathan entende que “[...] analisar um assunto por *facet*s significa que cada aspecto desse assunto pode ser visto como as manifestações de certas *características* ou *facet*s que obedecem a postulados pré-determinados. O sistema torna-se, assim, multidimensional e ilimitado”. Citando Piedade (1977, p. 12-13),

[...] Personalidade, na Medicina, são os órgãos; na Zoologia, os animais; na Botânica, os vegetais; na Religião, as várias religiões; na Linguística e na Literatura, as várias línguas; na Engenharia Civil, os edifícios, as pontes, etc.; na Biblioteconomia, as bibliotecas públicas, as bibliotecas universitárias, etc.

Energia abrange as ações, as reações, as atividades, as operações, os processos, as técnicas, os tratamentos, os problemas, etc. Na Biologia, na Botânica e na Zoologia é representada pela Morfologia, pela Fisiologia, pela Patologia, pela Ecologia, pela Genética e pela Ontologia e sua manifestação. Na Biblioteconomia, serão energia a catalogação, a classificação, a referência, o empréstimo etc.

Matéria consiste em todos os tipos de materiais e substâncias de que são feitas as coisas. Na Engenharia Civil serão os tijolos, as telhas, as pedras etc.; na Biblioteconomia serão os manuscritos, livros, os discos, etc.

Espaço corresponde às divisões geográficas e Tempo, às divisões cronológicas.

Um mesmo conceito pode pertencer a uma categoria num contexto e a outra em contexto diverso.

A análise em *facet*s permite a coordenação de conceitos, significando que um assunto, por mais complexo que seja, pode ser representado pela síntese de mais de uma *facet*a, cada uma indicando conceitos diferentes. Dessa forma, sua teoria tornou possível estabelecer relações de subordinação lógica – relações hierárquicas – com segurança. Na Classificação dos Dois Pontos a notação mistura letras, números e indicadores especiais de *facet*a. Exemplo: T15:3,133 (Ensino elementar através de discos).

Preocupados com a organização do conhecimento, um grupo de pesquisadores, formado por professores, documentalistas e cientistas da informação, constituiu o Classification Research Group, no ano de 1952. Este grupo, o CRG, trouxe contribuições significativas para o desenvolvimento da teoria da classificação. Partindo da premissa de que o universo do conhecimento é o assunto específico a ser classificado, o Grupo estabeleceu sete categorias além das já elencadas por Ranganathan. São elas: objeto, lugar, agente, modo, instrumento, produto e finalidade.

A Library of Congress (LC) é um sistema de classificação criado para atender às necessidades de organização do acervo da Biblioteca do Congresso Americano; não há

previsão para assuntos que não estejam representados na biblioteca. A principal finalidade da LC é a garantia literária. Suas tabelas vão se modificando e se expandindo, segundo as exigências da coleção a que serve; a extensão de cada classe é determinante para uma maior ou menor especificidade de assunto. A LC é utilizada em conjunto com uma lista de cabeçalhos de assuntos, a *Subject Headings of Library of Congress*. É um sistema muito detalhado e bastante enumerativo, podendo ser considerado como uma série de classificações especializadas, mas que recorre à síntese, quando aplica suas inúmeras tabelas auxiliares.

A LC não possui tabelas auxiliares comuns a todas as classes, mas existem muitas tabelas auxiliares destinadas a subdividir determinados grupos de números de cada classe. Há cinco tipos de tabelas auxiliares: a) subdivisões de forma, b) subdivisões geográficas, c) subdivisões cronológicas, d) subdivisões de assuntos específicos e e) subdivisões combinadas. A notação da classificação é mista, constituída de letras, números e sinais gráficos. Exemplo: HF5549.5.J6 (Administração de pessoal).

As Linguagens Documentárias não se estruturam apenas sob a forma sistemática. Elas também podem ser estruturadas sob a forma alfabética. O PRECIS, a lista de cabeçalhos de assuntos e o tesouro são exemplos dessas linguagens. Na próxima subseção abordaremos o PRECIS. A lista de cabeçalhos de assunto e o tesouro serão tratados na exposição que faremos sobre Vocabulários controlados.

2.2.1.1.3.2 Ordem alfabética

O PRECIS (Preserved Context Index System) é um sistema pré-coordenado de indexação alfabética de assunto. Foi desenvolvido, originalmente, pela British Library para a produção automática dos índices da British National Bibliography (BNB). É um sistema fundamentado na análise sintática e semântica da língua, e prevê uma série de operadores de função para caracterizar a posição e o significado dos termos no contexto de um cabeçalho de assunto. É considerado um sistema de indexação flexível e de grande alcance; não se atém a uma terminologia rica e controlada, e nem a um sistema de classificação que se desatualiza em pouco tempo. Este sistema não é constituído de uma lista pré-estabelecida de termos ou expressões, mas de um conjunto de procedimentos de indexação que podem ser aplicados a qualquer área temática e tipo de documento. Ele faz uso de remissivas para relacionar termos que mantêm relações semânticas entre si.

As listas de cabeçalhos de assunto e os tesouros aparecem sob o rótulo de vocabulários controlados. Pela relevância do tema para esta pesquisa, estas ferramentas serão abordadas na subseção seguinte com maior profundidade.

2.2.1.1.4 Vocabulários controlados

Um vocabulário controlado é uma lista de termos autorizados, que se constitui no instrumento de trabalho do indexador, ou seja, do bibliotecário responsável pela análise temática do documento a ser indexado. Esta ferramenta – vocabulário controlado – costuma ser estruturada na forma semântica, e tem como propósito tornar possível o uso de um código, visando à recuperação de um documento sob o enfoque do conteúdo.

Esses vocabulários se preocupam com a organização do conhecimento, uns na forma sistemática e outros na forma alfabética. Tendo como finalidade padronizar os termos usados na indexação, o vocabulário controlado registra a linguagem de uma área de conhecimento sob a forma de um cabeçalho de assunto ou de um descritor.

Nosso interesse nesta pesquisa são as listas de cabeçalhos de assunto e os tesouros. Nestas ferramentas, o arranjo explícito dos termos é alfabético, sendo que nos tesouros dois tipos de relações são identificadas claramente: as hierárquicas e as não-hierárquicas. No tesouro existe uma estrutura hierárquica implícita, incorporada à lista alfabética por meio de remissivas. A lista de cabeçalhos de assunto difere do tesouro porque não distingue claramente as relações hierárquicas das associativas.

Uma lista de cabeçalhos de assunto, conforme Lima (1998), é um vocabulário controlado organizado em ordem alfabética, e que contempla temas preferenciais e não-preferenciais. A unidade básica dessas listas é o cabeçalho de assunto, que é representado por uma palavra ou grupo de palavras que expressam um conceito ou um tópico temático. A lista de cabeçalhos de assuntos nem sempre reflete a organização conceitual de uma área de conhecimento e, muitas vezes, as relações são estabelecidas em função do documento indexado em um determinado momento, não servindo para outros documentos que tratam sobre o mesmo assunto.

A filosofia do cabeçalho de assunto, segundo Cutter (1904, p. 69), é a de priorizar a forma usada pelo usuário na representação da informação indexada. Ele defendia a ideia que “General rules, always applicable, for the choice of names of subjects can no more be given than rules without exception in Grammar. Usage in both cases is the supreme arbiter, – the

usage, in the present case, not of the cataloger but of the public in speaking of subjects⁴⁶". Ele afirmava que os princípios básicos para a elaboração de cabeçalhos de assunto são: o uso e a ordem direta, ou seja, o emprego do termo mais utilizado pela comunidade de usuários, e a seleção do termo mais específico para representar o assunto, evitando termos genéricos. Sob esta ótica, seu pensamento se alinha com o defendido pelos Estudos de Terminologia, mas diferencia-se destes por não ter a linguagem de especialidade como objeto de análise.

A lista de cabeçalhos de assuntos foi desenvolvida para ser utilizada em bibliotecas de assuntos gerais. Ela se estrutura sob a forma alfabética, e representa os assuntos sob a forma de cabeçalhos já estruturados em listas, que arrolam termos de todas as áreas do conhecimento. Ela contempla termos preferenciais e não preferenciais. Segundo Fujita (1988), o conjunto de regras para a construção e arranjo de cabeçalhos de assunto pode ser resumido em dois princípios básicos: princípio específico e princípio sindético. Pelo princípio específico, os assuntos devem dar entrada pelo termo mais específico e não pela classe a que estão subordinados. Pelo princípio sindético, a ligação de assuntos correlacionados é realizada através de uma rede de referências cruzadas.

As regras para a estruturação dos cabeçalhos, que aceitavam somente a linguagem natural (LN) como a única forma de acesso possível, trouxeram alguns problemas para a sua utilização. Os cabeçalhos de assuntos assim constituídos apresentavam mais de uma palavra para um mesmo significado, como também, mais de um significado para uma mesma palavra.

Quanto à forma de apresentação, de modo geral, a lista de cabeçalhos de assuntos traz instruções gerais para uso. Os usuários são remetidos pela referência *veja*, dos termos não utilizados para os termos utilizados, e pela referência *veja também*, dos assuntos mais gerais para os mais específicos, assim como, dos mais específicos para os assuntos mais gerais. A referência *veja também* é utilizada ainda para relações associativas.

“Tesauro é um instrumento de controle terminológico, utilizado em Sistemas de Informação para traduzir a linguagem dos documentos, dos indexadores e dos pesquisadores numa linguagem controlada, usada na indexação e na recuperação de informações” (Bräscher, 2009⁴⁷). O tesauro fundamenta-se em conceitos. Sua metodologia de elaboração é alicerçada na ideia de sistema. Ele é construído sob determinadas regras e é constituído por um conjunto

⁴⁶ Tradução da autora: As regras gerais, sempre aplicáveis, para a escolha dos cabeçalhos de assuntos, não podem mais serem empregadas do modo semelhante às regras da Gramática. O uso, em ambos os casos é o ‘árbitro supremo’; o uso no caso em questão, não é o do catalogador, mas sim o do público falante.

⁴⁷ Original obtido na internet.

de descritores. Ele diferencia-se de uma lista de cabeçalhos de assuntos quanto a sua estrutura e concepção.

O tesouro caracteriza-se por ser uma linguagem de indexação especializada, controlada, estruturada de modo a explicitar as relações estabelecidas *a priori* entre os conceitos; tais relações são reconhecidas como hierárquicas, associativas ou de equivalências. É uma linguagem pós-coordenada, uma vez que a coordenação das expressões de busca a serem recuperadas é feita no momento do processo de busca da informação e recuperação de documentos, e não no momento da indexação do documento. Sob essa concepção ele admite a organização de conceitos em facetas, como por exemplo, o registro do conceito 'Linguística da Enunciação' em campos distintos, uma vez que contém dois conceitos: 'Linguística' e 'Enunciação'.

O descritor representa um conceito e é a unidade descritiva básica que sustenta o processo de Análise Documentária de um documento. A identificação de um conceito obedece a dois princípios: o da monorreferencialidade e o da contextualização. Pelo princípio da monorreferencialidade os termos se referem a uma só realidade, sendo que o conceito mantém com o descritor uma relação unívoca; assim, para cada conceito há apenas um descritor e para cada descritor, apenas um conceito. Pelo princípio da contextualização o contexto é fundamental, pois é ele que especifica o significado de um descritor.

O tesouro vem sendo considerado como o instrumento mais adequado para o tratamento e a recuperação da informação. Ele não só estabelece relações entre os termos utilizados, denominando-os como termos preferidos e termos não-preferidos, como permite uma visão do universo conceitual que está sendo indexado. É uma Linguagem Documentária pós-coordenada, ou seja, é uma linguagem que permite que o usuário combine os termos no momento da pesquisa, o que dá mais flexibilidade ao Sistema, porque torna possível representar mais adequadamente a informação que se deseja obter.

O tesouro estrutura-se segundo um léxico hierarquizado de acordo com as relações semânticas que ocorrem entre os termos. Sua construção está fundamentada em dois princípios referenciais: no conhecimento categorizado em assuntos, e em um *corpus* discursivo do qual são retirados os termos considerados significativos. Desse modo, pretende-se assegurar a organização conceitual de um campo de especialidade e a garantia literária.

As relações documentárias estabelecidas entre os descritores de um tesouro são de três tipos: hierárquicas, associativas ou de equivalência. As relações hierárquicas são simbolizadas por TG/TE (termo geral/termo específico); as relações de equivalência, por UP (usado para); e

as relações associativas, por TR (termo relacionado), TC (termo correlato) ou TA (termo associado).

As relações hierárquicas expressam os graus de superordenação e subordinação entre descritores, indo dos termos mais gerais para os mais específicos. Elas podem ser genéricas, específicas ou partitivas, e constituem a estrutura principal de um tesauro. São também conhecidas como árvores semânticas.

As relações de equivalência são utilizadas para o controle do vocabulário. Em um tesauro atribui-se apenas uma denominação para representar um conceito e ser identificado como descritor; as demais denominações são consideradas como não-descritores; para esses, deve-se fazer uso de remissivas. As relações de equivalência ocorrem para os casos de: 1) nomes comerciais e comuns (ácido acetilsalicílico e aspirina), 2) palavras com grafias diferentes (Rumânia e Rômania), 3) termos atuais e outros de menor uso (países em vias de desenvolvimento e países subdesenvolvidos), 4) regionalismos (mandioca e aipim), e 5) uso popular ou científico (aranhas e aracnídeos), 6) nomes instituídos cientificamente e os que têm origem na linguagem popular (equipamentos de alta fidelidade e equipamentos HI-FI). Conforme a política de indexação estabelecida, podem ocorrer relações de equivalência também entre termos que se relacionam hierarquicamente, como por exemplo: Jardins ver Arquitetura paisagística.

As relações associativas ocorrem entre termos que não são equivalentes nem superordenados ou subordinados uns aos outros, mas que se relacionam por algum outro tipo de associação. Como exemplos de termos associativos, citamos os seguintes: adubo e adubação (relação coisa/aplicação), desemprego e pobreza (relação causa/efeito), veneno e toxicidade (relação coisa/propriedade), cacau e chocolate (relação matéria-prima/produto), veneno e antídoto (relação coisa/seu contra-agente), estética e beleza (relação disciplina/objeto usado), fusão e fusibilidade (relação atividade/propriedade), pesca e pescado (relação atividade/agente ou paciente), apicultura e mel (relação atividade/produto), navio e casco (relação todo/parte), emprego e desemprego (relação entre opostos) e, ainda, ensino e aprendizagem (relação entre atividades complementares).

No tesauro, as relações de um termo com os seus termos associados permitem determinar, ainda que indiretamente, a significação de um descritor. A maioria dos tesauros não traz uma definição do termo, e quando existe ela é expressa em notas de escopo relativas ao seu campo de aplicação. Na prática, os tesauros não têm conseguido normalizar efetivamente a representação do conceito, o que permite a ocorrência de ambiguidades.

Uma das principais vantagens em se adotar um vocabulário controlado – lista de cabeçalho de assunto e tesouro – consiste no fato de ele propiciar as condições necessárias para que se possa melhorar a consistência na representação do conteúdo temático de um documento. Neste sentido, salientamos que o tipo de vocabulário controlado utilizado na indexação influi sobre a consistência. Além da probabilidade de bibliotecários selecionarem diferentes descritores, considerando aqui duas variáveis – pessoa e tempo –, há também a considerar a possibilidade dos autores utilizarem diferentes termos para expressarem suas ideias e, até mesmo, um mesmo termo para expressar ideias diferentes. Se um grupo de indexadores utilizar a mesma ferramenta como suporte para suas atividades de indexação, a tendência é obter uma uniformidade maior no conjunto de descritores selecionados, e em consequência uma maior consistência no processo de indexação. Utilizando vocabulários controlados, o leque de opções para determinação de um descritor representativo de um conceito é mais reduzido, e assim a possibilidade de aumentar a consistência na indexação é maior.

No **vocabulário controlado especializado** que estamos propondo, além do uso de modificadores e qualificadores, estamos sugerindo o uso de paráfrases e o uso do termo na língua original, conforme ficou consagrado na área de um domínio, seguido do termo na forma como é reconhecido na língua do catalogador, como, por exemplo, no Vocabulário Controlado Especializado de ULRIES em Filosofia platônica e Platonismo usaremos Anamnese (Teoria da reminiscência), Dynamis (Força), Eudaimonia (Felicidade), Physis (Natureza).

Paráfrases são termos ou expressões usadas de forma combinada com as unidades lexicais de representação da informação especializada; tratam-se de duas ULRIES que contêm a mesma informação, ainda que utilizem diferentes formas para expressá-las. Elas são usadas entre parênteses e visam oferecer ao usuário uma unidade de informação de maior densidade informativa. Exemplo: Diotima (sacerdotisa de Mantiquéia).

Com o propósito de contextualizar ou especificar o sentido de uma unidade de informação, em alguns casos sugerimos reunir em uma única unidade lexical formas variantes, tal como : Alcibíades Menor (Segundo Alcibíades) e Mito da caverna (Alegoria da caverna)⁴⁸.

Finalizamos esta reflexão observando que os vocabulários controlados são ferramentas que controlam sinônimos, diferenciam homógrafos e agrupam termos afins. Esses fenômenos,

⁴⁸ O diálogo de Platão Alcibíades Menor é também reconhecido na filosofia platônica como Segundo Alcibíades, e o mito platônico Mito da caverna como Alegoria da caverna.

que estão presentes nas Linguagens Documentárias e que são reconhecidos como fenômenos linguísticos, serão o tema da subseção que se segue.

2.2.1.1.5 Fenômenos linguísticos

No controle de vocabulário para tratamento e recuperação da informação, a Linguagem Documentária reconhece alguns fenômenos linguísticos. A utilização da linguagem natural na construção das Linguagens Documentárias tende a gerar um produto que reflete a complexidade própria da linguagem, devido a fenômenos como a polissemia, a ambiguidade, a sinonímia e a hiponímia.

Sobre a polissemia e a ambiguidade, Cintra et al. (2002) afirmam que esses fenômenos devem ser neutralizadas em uma Linguagem Documentária, visando assegurar a monosssemia, ou seja, a presença de termos que apresentam apenas um significado. A polissemia decorre da organização sintático-semântica, e tem a finalidade de distinguir a plurissignificação como fenômeno geral. A ambiguidade ocorre quando há mais de uma interpretação para um termo, em função da plurissignificação.

Pela polissemia uma palavra pode comportar mais de um significado, como por exemplo, na frase que diz: ‘O *cachorro* do meu vizinho é insuportável’, onde *cachorro* pode ser entendido como ser humano ou também como ser animal. Já na frase ‘O menino está sentado perto do *banco*’, a ambiguidade se constitui uma vez que *banco* pode se referir tanto a uma agência bancária quanto ao objeto para sentar.

De forma diversa à da linguagem natural, que se caracteriza pela dinamicidade com que se transforma e evolui a cada momento, a Linguagem Documentária é construída com a finalidade de significar de maneira precisa. Ao contrário do comportamento polissêmico da palavra na linguagem natural, a LD tende à monosssemia e favorece o princípio da univocidade, ainda que admita a presença de variações.

Pelo princípio da univocidade um conceito só pode ser representado por um único termo, na medida em que um termo representa uma ideia. Há um relativo consenso de que este princípio encontra respaldo apenas nas ciências sistemáticas, como a biologia e a zoologia, por exemplo. Nas ciências sociais esse princípio não se aplica. A razão desse entendimento está no fato de que, pelo princípio da variação, é reconhecido que existe mais de uma maneira de denominar um conceito.

A ambiguidade na linguagem natural é resolvida com certa facilidade pelo contexto. Na LD, no entanto, é utilizado o recurso de modificadores e de qualificadores com a

finalidade de contextualizar o sentido de uma unidade de informação. Modificadores, para Bräscher (1999, p. 8), são “[...] os 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.” Exemplificando: Linguagem (aspectos filosóficos), Eutanásia (aspectos éticos), Economia (aspectos políticos). Ainda segundo Bräscher e Carlan (2010, p. 168) “Qualificadores são termos agregados entre parênteses aos descritores para delimitar seu significado e eliminar a homografia. Ex.: tênis (esporte); tênis (calçado).” Ou ainda: valor (filosofia), valor (finanças) e valor (economia).

Em uma Linguagem Documentária, o funcionamento linguístico de um item lexical 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, observamos uma situação de sinonímia, fato linguístico que depende do contexto, muito mais do que de relações de sentido como hiponímia e antonímia.

Sinonímia, outro fenômeno linguístico presente nas Linguagens Documentárias é, segundo a Norma ISO 1087 (1990), a “Relation entre désignations [...] de même langue qui représentent la même notion [...]”⁴⁹. Nas Linguagens Documentárias, o uso de sinonímia permite estabelecer relações de equivalência entre termos preferidos e não-preferidos. Conforme Lara (1999), os sinônimos funcionam como mecanismo de controle da variação de significado e como recurso para o acesso à linguagem do Sistema.

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 hiponímia, identificada por Lyons (1987, v.1, p. 235) como “[...] a relação existente entre um lexema mais específico ou subordinado e um lexema mais geral ou superordenado” também se faz presente nas Linguagens Documentárias. Ela estabelece uma relação do tipo espécie/gênero. Sujeito e objeto, por exemplo, são hipônimos de função gramatical, tal como rosa e tulipa, são hipônimos de flor. Antonímia, por sua vez, é a relação que se estabelece entre duas palavras ou mais, e que apresentam significados contrários.

⁴⁹ Tradução da autora: Relação entre designações de uma mesma língua, representando um mesmo conceito.

Encerrada esta subseção, abordaremos logo a seguir o tema da Análise Documentária, dando ênfase à análise de assunto. No contexto da Ciência da Informação a análise de assuntos é uma das etapas mais importantes do processo de Análise Documentária, e se constitui na primeira etapa do processo.

2.2.1.1.6 Análise Documentária

A Análise Documentária situa-se na etapa de tratamento da informação, visando a sua recuperação sob o enfoque do conteúdo. Intermediada pela Linguagem Documentária, ela possibilita a manipulação de um código comum a um Sistema de Informação e ao seu usuário, de modo a viabilizar o processo de interação entre seus interlocutores. Além de codificar o conteúdo informacional de um documento, ela tem como propósito construir um vocabulário controlado – uma Linguagem Documentária – como uma lista de cabeçalhos de assuntos e um tesouro, por exemplo. Ela é de natureza metodológica e propõe processos de organização do conhecimento, visando à recuperação e à circulação da informação.

Conforme Lara (2011), o termo Análise Documentária tem origem nos trabalhos de Coyaud⁵⁰ e Gardin⁵¹. Este termo foi usado por estes autores para designar operações semânticas que transformam um texto original em uma ou várias palavras-chave, ou, paráfrases, com o propósito de facilitar a representação de conteúdos e a recuperação da informação.

Na literatura brasileira, o termo Análise Documentária foi subscrito formalmente pelo Grupo Temma⁵², que reúne pesquisadores que têm como foco principal de pesquisa a interface entre Documentação e Linguística. Sob a ótica deste Grupo, a Análise Documentária sugere procedimentos para a análise de textos, com o propósito de selecionar conteúdos informativos para serem representados, recuperados e disseminados.

Como meio de diminuir a dispersão frente ao volume da produção, a Análise Documentária persegue a síntese. Buscando a síntese e observando políticas de indexação institucionais, ela procura identificar no texto as unidades informativas do documento.

⁵⁰ Coyaud, M. *Introduction a l'étude des langages documentaires*. Paris: C. Klincksieck, 1996.

⁵¹ Gardin, J.-C. Eléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, v.11, n.5, mai. 1966, p. 171-182. Gardin, J.-C. Análise documentária e análise estrutural em arqueologia. In: CORDIER, S. (ed.). *Lévi-Strauss*. São Paulo: documentos, 1968. Gardin, J.-C. et al. *L'automatisation des recherches documentaires: un modèle général S'YNTOL*. 2. éd. rev. et augm. Paris: Gauthier-Villars, 1968.

⁵² Grupo Temma – Grupo de Pesquisa da Escola de Comunicação e Artes da USP, fundado em 1986.

Na visão de alguns teóricos, o processo de Análise Documentária envolve três procedimentos: análise de assunto, síntese e representação. Na análise de assunto, o conteúdo temático do documento é analisado com o propósito de identificar os assuntos relevantes. Os assuntos são identificados por meio de uma linguagem de indexação, e podem ser expressos em palavras e ou descritores de um vocabulário controlado ou em números de um código de classificação.

Pela importância que a análise de assuntos assume no âmbito da Análise Documentária, na medida em que é considerada como o procedimento básico na recuperação da informação, e em que exerce enorme influência sobre a eficácia dos Sistemas de Recuperação da Informação, este é o assunto que será discutido a seguir.

A análise de assunto é uma tarefa da responsabilidade do bibliotecário indexador e consiste na compreensão e interpretação do conteúdo temático de um documento. É uma das várias denominações atribuídas a esta etapa da Análise Documentária. Ela é reconhecida também como: análise temática, análise conceitual e análise de conteúdo. Alguns teóricos atribuem a ela outras duas denominações: Análise Documentária e processo de indexação. Ela ocorre de forma diferenciada nas diversas áreas do conhecimento, seja pela especificidade própria de cada área, pelo tipo de texto produzido, pela terminologia usada pelos especialistas e, ainda, pela existência ou não de uma taxonomia de assuntos, dentre outras.

Diversos elementos interferem no processo de análise de assunto, dentre eles: 1) a atividade do indexador, 2) a política de indexação adotada pelo Sistema e estabelecida segundo os objetivos da Instituição, 3) o tipo de vocabulário utilizado, e 4) o perfil e necessidades de informação dos usuários.

Indexador é o bibliotecário responsável pelo processo de análise de assuntos, e a atividade desenvolvida por ele é a indexação. Ele é um dos principais responsáveis pelo sucesso ou insucesso de um Sistema de Recuperação da Informação; um conhecimento mínimo da área a ser indexada é indispensável. No exercício da atividade de indexação, ele sofre influência de diferentes ordens, especialmente de fatores subjetivos, linguísticos, cognitivos e lógicos.

A atividade de indexação se caracteriza por ser substancialmente subjetiva e complexa. Ela é afetada pelas tendências ideológicas do indexador, pela sua experiência no exercício da atividade e pelo conhecimento prévio do tema a ser indexado. Os fatores que interferem na análise do assunto são diferentes segundo o tempo de experiência dos indexadores. Toda essa complexidade se manifesta no resultado obtido na indexação de um

documento, fazendo com que diferentes bibliotecários indexem um mesmo tema, sob diferentes formas e a partir de recortes diferenciados.

O indexador descreve o conteúdo temático de um documento empregando um ou vários termos de indexação. O conjunto de termos de indexação deve sintetizar o conteúdo do documento. Os termos atribuídos pelo indexador servem como pontos de acesso para a localização e recuperação da informação.

Quanto maior a representação do conteúdo temático de um documento, mais pontos de acesso são oferecidos ao usuário para a recuperação da informação e localização de um documento. Uma das propriedades mais importantes da representação de conteúdo é a extensão. À medida que se aumenta a extensão da representação, maior é a possibilidade de recuperação de um item de informação.

Quanto à extensão de um registro, sob a ótica da recuperabilidade, podemos optar por uma indexação seletiva ou uma indexação exaustiva. A indexação seletiva apresenta poucos termos e uma indicação muito geral do conteúdo de um documento. A indexação exaustiva, de forma contrária, proporciona uma indicação mais abrangente do conteúdo de um documento e, com isso, oferece um número maior de pontos de acesso.

Para Lancaster (2004, p. 18), a indexação de assuntos envolve duas etapas principais: análise conceitual e tradução. Tradução consiste na “[...] conversão da análise conceitual de um documento num determinado conjunto de termos de indexação”. Esta etapa implica uma decisão sobre quais descritores serão atribuídos ao documento, ou seja, quais os descritores que melhor representam o conteúdo do documento e os interesses dos usuários; na maioria das vezes esses descritores são extraídos de vocabulários controlados.

Sobre as etapas que envolvem a análise de assunto, não há um consenso. Para os propósitos desta pesquisa, nos orientaremos por autores como Lima (2004), que consideram que a análise de assuntos envolve três etapas: 1) leitura do documento, 2) extração de conceitos representativos do conteúdo temático do documento, e 3) representação da atenção⁵³.

A leitura do documento visa determinar o assunto do documento. É um processo que consiste em ler um documento para extrair conceitos que representem seu conteúdo. Não se trata de uma leitura do texto na íntegra, mas de partes significativas do texto, como título, sumário, resumo, capa e contracapa, índice, bibliografia, um rápido exame dos parágrafos iniciais da introdução e de alguns capítulos, assim como de alguns parágrafos da conclusão.

⁵³ O termo atenção refere-se ao assunto de um documento, do que trata um documento.

Conforme Lara (2009), leitura documentária é uma modalidade específica do processo de leitura comum. A leitura para fins documentários se organiza de modo diverso da leitura efetuada usualmente. Na visão de Fujita (2004),

A leitura, apesar da individualidade do ato realizado, é um ato social porque existe um processo de comunicação e de interação entre o leitor e o autor do texto, ambos com objetivos estabelecidos anteriormente dentro do contexto de cada um. Apesar de, aparentemente simples e tão natural, o processo de leitura possui uma complexidade que está subjacente porque depende do processamento humano de informações e da cognição de quem lê, de um texto elaborado por um autor e do contexto de ambos, o que determina os objetivos da leitura.

Segundo Lara (2011), o processo de leitura documentária deve viabilizar o reconhecimento do vocabulário de uma área de especialidade e das formas mais usuais de organização textual. Ainda de acordo com Lara (2011, p. 102) é necessário que o profissional indexador

[...] receba treinamento específico para conhecer estratégias de leitura, para identificar tipologias textuais e para reconhecer a terminologia dos textos. Essas competências, somadas ao conhecimento dos objetivos institucionais e dos perfis de seus usuários, habilita-o a desempenhar a produzir a síntese documentária, meio para construir a informação documentária.

Como bem observa Lara (2009), o leitor-documentalista não se caracteriza como um leitor-modelo, uma vez que não é um leitor que tem domínio do conhecimento veiculado pelo texto e, por outro lado, o autor do texto não prevê o leitor-documentalista como seu leitor.

Na etapa seguinte, a de extração de conceitos, o indexador deve se preocupar em identificar os conceitos que aparecem no documento. Mas é na etapa de representação da atenção que o indexador decide quais aspectos do documento são relevantes aos interesses dos usuários do Sistema. Com esse propósito ele precisa fazer uma seleção dos conceitos que realmente sintetizam o assunto do documento, e representá-los por termos que são extraídos de uma linguagem de indexação. É nesta etapa que são definidos os termos que passam a ser chamados de descritores de assunto, cabeçalhos de assunto, palavras-chave, termos de indexação, ou enunciados.

Atenção é um termo traduzido do inglês *aboutness*. Ainda que tenha sido adotado por Lancaster, como sendo o termo mais adequado para traduzir *aboutness*, Guedes, Martinho e Moraes (2009) observam que não há consenso entre os especialistas da área em relação a esse entendimento.

Para alguns pesquisadores o termo atenção está relacionado com o de relevância. Relevância é entendida como a relação que se estabelece entre um documento e uma necessidade de informação. O tema relevância/pertinência é tratado por vários estudiosos.

Alguns deles tratam atinência como sinônimo de relevância; eles relacionam atinência com as medidas de revocação e de precisão.

Entendendo que o processo de análise conceitual implica no reconhecimento de que um documento tem ou não interesse para determinado grupo de usuários, Lancaster (2004, p. 15) afirma que “‘Análise conceitual’ [...] significa nada mais do que a identificação dos assuntos estudados ou representados num documento”.

Sobre a análise de assuntos, ainda, é importante observarmos que uma mesma publicação pode ser indexada de modo diverso, segundo os interesses de diferentes instituições. Uma biblioteca universitária ou um centro de pesquisa, por exemplo, representam interesses diferentes.

Conforme mencionado por Mai (2000, p. 294),

If one focuses solely on the representation aspect and ignores future users, one might risk representing documents in a way that would be of no use for the users. An indexer who does not pay much attention to the users might choose to represent subjects of documents that are of no interest to the users, or might use a different vocabulary from the users, or might represent the subject on a level that is too broad or too narrow for the users. However, if the indexer pays too much attention to the users of the system, the indexer might represent documents in such a way that the subject representation of the documents only serves the current users and those current information needs.⁵⁴

Ainda referenciando Mai (2001, p. 606), acrescentamos

It would be nearly impossible, of course, for any single person or, in this case, any single indexer, to determine *all* of the ideas and meanings which might be associated with any particular document, since there might always be potential ideas and meanings which different people at different times and places might find in the document. Furthermore, it would be well nigh impossible to predict precisely which of the many possible ideas and meanings that could be associated with the document would be specifically valuable to the users or would have some sort of lasting value for the document. To recognise and accept this fundamental openness is of utmost importance. The indexer must realise from the start that he or she will never discover all the ideas and meanings that could be associated with the document and that, therefore, it is not possible to describe all these ideas and meanings.⁵⁵

⁵⁴ Tradução de Lancaster (2004, p. 13): “Se se focalizar exclusivamente o aspecto da representação, ignorando os usuários futuros, corre-se o risco de representar os documentos de uma forma que não terá qualquer serventia para os usuários. Um indexador que não dê muita atenção aos usuários poderá optar por representar assuntos de documentos que não tenham interesse para eles, ou usar um vocabulário diferente do vocabulário deles, ou representar o assunto em nível que seja muito genérico ou muito específico para eles. No entanto, se o indexador der excessiva atenção aos usuários do sistema, poderá representar os documentos numa forma tal que a representação temática dos documentos somente atenda aos usuários atuais e às necessidades de informação atuais.”

⁵⁵ Tradução de Lancaster (2004, p. 11): “‘Seria quase impossível, naturalmente, para qualquer pessoa ou, neste caso, qualquer indexador, precisar *todas* as idéias e significados que estivessem associados a qualquer documento, posto que sempre haverá idéias e significados potenciais que diferentes pessoas em diferentes

A partir do exposto anteriormente, é possível afirmar que não há um conjunto ‘correto’ de termos de indexação para um mesmo documento. Uma indexação, que concilie os princípios da indexação com os interesses de pesquisa dos usuários, exige que o indexador tenha conhecimento das necessidades de informação da comunidade atendida. Assim, considerando o princípio de indexação orientado para o usuário, e a necessidade de capacitar o bibliotecário indexador para desempenhar sua função, é recomendado que o indexador atue também como bibliotecário de referência.

Finalizando nossas observações sobre a Análise Documentária, passaremos a refletir sobre as metodologias em que se fundamentam as Linguagens Documentárias. Teremos como propósito fazer algumas aproximações entre os diferentes métodos, tomando o tesauro como objeto de análise.

2.2.1.1.7 Metodologias

As metodologias utilizadas para a construção de uma Linguagem Documentária, sob a forma de tesauro, se estruturam fundamentalmente a partir de dois métodos: dedutivo e indutivo. Fujita (1992) observa que tais métodos ganham diferentes denominações entre os teóricos, mas que, de modo geral, resguardam os mesmos conteúdos.

A partir da revisão de literatura⁵⁶, feita sobre as metodologias em que se fundamentam as Linguagens Documentárias, é possível fazer algumas aproximações. Os métodos dedutivo, analítico e *a priori* assemelham-se a partir das seguintes características: a terminologia que formará o vocabulário é reunida antes do momento da indexação, e há uma comissão constituída que tem como responsabilidade a revisão dos termos. Os métodos indutivo, *gestalt* e *a posteriori* são similares nos seguintes aspectos: a terminologia que formará o vocabulário é coletada durante o momento da indexação, e não existe uma comissão constituída; a assistência de técnicos e especialistas ocorre somente quando necessário, a partir de uma solicitação. Na opinião de Guinchat e Menou (1994), na prática é usual a combinação dos dois métodos para a compilação de termos de um vocabulário controlado.

momentos e lugares poderão descobrir nesse documento. Além do que, seria quase impossível prever com exatidão quais das inúmeras idéias e significados que estivessem associados ao documento seriam especificamente úteis para os usuários ou dariam ao documento alguma utilidade duradoura. É da máxima importância reconhecer e aceitar essa indefinição fundamental. O indexador deve compreender, desde o início, que jamais descobrirá todas as idéias e significados que estariam associados ao documento e que, portanto, não é possível descrever todas essas idéias e significados”.

⁵⁶ Para um estudo desenvolvido sobre metodologias em que se fundamentam as linguagens documentarias ver Cervantes (2006).

Ainda sobre a compilação de termos, é preciso destacar que no método dedutivo há uma preocupação com a estrutura conceitual, ou seja, com a categorização do conhecimento. Observamos que a preocupação deste método não está centrada apenas na compilação de termos, uma vez que o domínio ou o subdomínio de uma área é estruturado antes do início da coleta de termos. Entretanto, quando o método de coleta se dá pelo processo indutivo, verifica-se que não existe uma preocupação primeira com a estrutura conceitual ou categorização; a organização do conhecimento em domínio e subdomínio ocorre após a compilação de termos.

Mencionamos, no início desta subseção, que diversos tipos de estrutura são utilizadas na Organização do Conhecimento e Representação da Informação. Com o propósito de esclarecer os conceitos taxonomia, tesouro e ontologia, apresentaremos, na próxima subseção, algumas considerações sobre as semelhanças e diferenças que podem ser observadas nessas ferramentas.

2.2.1.1.8 Semelhanças e diferenças entre taxonomias, tesouros e ontologias

Iniciamos estas ponderações com uma definição de taxonomia.

Taxonomia é, por definição, classificação, sistemática e está sendo conceituada no âmbito da Ciência da Informação como ferramenta de organização intelectual. É empregada em portais institucionais e bibliotecas digitais como um novo mecanismo de consulta, ao lado de ferramentas de busca. Além destas aplicações, a taxonomia é um dos componentes em Ontologias. A organização das informações através do conceito de Taxonomia permite alocar, recuperar e comunicar informações dentro de um sistema de maneira lógica através de navegação (CAMPOS; GOMES⁵⁷, 2008).

Sobre sistemática e taxonomia, vale esclarecer que sistemática é o estudo científico dos organismos em sua diversidade, de sua evolução no tempo e no espaço, enquanto que taxonomia é a parte da sistemática que se ocupa das regras e dos princípios a serem usados para nomear, delimitar e classificar os organismos.

No âmbito da Linguagem Documentária, taxonomia é entendida como uma estrutura que possibilita classificar coisas em grupos hierárquicos, com o propósito de facilitar sua identificação, estudo e localização em um domínio. Uma taxonomia compõe-se de duas partes: estrutura e aplicação. A estrutura apresenta as categorias ou termos e suas relações. A aplicação consiste em uma ferramenta de navegação visando auxiliar a busca.

⁵⁷ Original obtido na internet.

Tradicionalmente a taxonomia é entendida como a ciência ou a técnica de classificação. Inicialmente este conceito era vinculado quase que exclusivamente aos domínios da biologia e da lógica. Hoje vem sendo bastante utilizado no campo das tecnologias da informação e da documentação, no que diz respeito aos conceitos de arquitetura da informação e mineração da informação.

Taxonomias e tesouros se assemelham pelo fato de que ambos são construídos com o propósito de sistematizar o conhecimento, e viabilizar o desenvolvimento da ciência, da pesquisa e da inovação tecnológica. Eles se diferenciam na medida em que as taxonomias são utilizadas, quase que exclusivamente, pelos profissionais de tecnologia da informação, enquanto que os tesouros são desenvolvidos pelos profissionais da informação e da documentação. Elas – as taxonomias – são utilizadas como Linguagens Documentárias, visando dar sustentação aos vocabulários controlados de diferentes domínios.

Tesouro é uma Linguagem Documentária, que se caracteriza pela especificidade e pela complexidade presente no relacionamento entre os termos que propiciam o conhecimento especializado. A partir das inúmeras definições de tesouros revisadas em literatura, é possível afirmar que tesouros são vocabulários controlados formados por descritores/termos relacionados semanticamente. São modelos de representação do conhecimento que atuam como instrumentos de controle terminológico, e que se estruturam hierarquicamente (gênero-espécie e todo-parte) e associativamente (aproximação semântica). São utilizados principalmente para indexar e recuperar informações temáticas.

Os tesouros não são utilizados somente pelos especialistas da informação com fins de indexação. Eles podem ser usados também por usuários da informação durante o processo de busca da informação e recuperação de documentos, visando à localização de documentos. Tesouro é o instrumento que intermedia a relação que se estabelece entre bibliotecários e usuários. Ainda que com propósitos diferenciados, é a linguagem comum entre aquela utilizada pelos indexadores e pelos usuários.

Na opinião de Sales e Café (2008⁵⁸), “Tesouros e ontologias são modelos de representação do conhecimento que, servindo como instrumentos de controle terminológico, auxiliam o processo de indexação e recuperação de informações por assunto.” Esses dois modelos de representação do conhecimento apresentam semelhanças e diferenças no que se refere aos termos, conceitos e objetivos, possibilitando um entendimento mais profundo

⁵⁸ Original obtido na internet.

desses instrumentos, tão necessários para propiciar condições facilitadas para que se efetive a comunicação técnica e científica.

Feito esse esclarecimento sobre taxonomias, tesouros e ontologias, desenvolveremos na continuidade desta pesquisa o tema das ontologias.

2.2.1.2 Ontologias

Ontologia é um modelo de representação do conhecimento que, assim como o tesouro, é utilizada para representar e recuperar informação por meio de estruturas conceituais. São instrumentos desenvolvidos pelos profissionais da engenharia computacional. Elas consistem em especificações formais que tem como propósito descrever estruturas conceituais de domínios específicos. Elas não apresentam sempre a mesma estrutura; sua estrutura pode variar de uma ontologia a outra. Enquanto esquemas conceituais em sistemas de base de dados, elas fornecem uma descrição lógica de dados compartilhados, possibilitando programas de aplicação e interoperabilidade entre as bases de dados.

Na sequência desta pesquisa, abordaremos o tema das ontologias sob diferentes aspectos, dentre outros, os que dizem respeito às suas características e à sua relação com os mapas conceituais.

2.2.1.2.1 Definição, finalidade, características e a relação das ontologias com os mapas conceituais

Há na literatura uma grande variedade de definições para as ontologias. Existe, ainda, uma diversidade muito grande também em relação aos tipos de ontologias, e propostas para a sua construção e para a aplicação em diferentes áreas de conhecimento.

O termo Ontologia tem origem no grego *ontos*, ser, e *logos*, palavra. Ele tem origem em Aristóteles, na palavra categoria, cujo sentido é o de classificar alguma coisa. O filósofo grego enumera dez categorias: substância, quantidade, qualidade, relação, lugar, tempo, estado, hábito, ação e paixão. As categorias apresentadas servem de base para classificar qualquer entidade.

A ideia proposta é a de que o conhecimento das categorias deve resultar em uma maior capacidade de análise e interpretação de elementos e argumentos do discurso. “As categorias possuem três níveis de validade: a) um ontológico, b) um lógico, c) um lingüístico-gramatical.” (REALE, 1995, v.5, p. 42)

No âmbito da organização da informação, Ontologia tem sentido diverso daquele tradicionalmente adotado na filosofia. As definições apresentadas na literatura são inúmeras, e existem contradições entre elas. De modo simples, podemos dizer que para elaborar ontologias, é preciso definir categorias para as coisas de um mesmo domínio.

Segundo Almeida e Bax (2003), uma das definições mais conhecidas para ontologias é a apresentada por Gruber [1993].

An ontology is a specification of a conceptualization.

The word ‘ontology’ seems to generate a lot of controversy in discussions about AI [Artificial Intelligence]. It has a long history in philosophy, in which it refers to the subject of existence. It is also often confused with epistemology, which is about knowledge and knowing.

In the context of knowledge sharing, I use the term ontology to mean a *specification of a conceptualization*. That is, an ontology is a description (like a formal specification of a program) of the concepts and relationships that can exist for an agent or a community of agents. This definition is consistent with the usage of ontology as set-of-concept-definitions, but more general. And it is certainly a different sense of the word than its use in philosophy.⁵⁹

As ontologias são utilizadas para organizar a informação em diversos domínios. Seu objetivo é propiciar um comum acordo na utilização do vocabulário compartilhado, de modo coerente e consistente. Para alguns pesquisadores, as ontologias podem ser consideradas como sinônimo de conceitualização, por entendê-las como possuidoras de uma estrutura semântica intencional, que codifica as regras implícitas, legitimando uma estrutura de uma parte da realidade. Pelas definições encontradas na literatura, e conforme revisão feita por Sales e Café (2008), é possível afirmar que as ontologias proporcionam um vocabulário formal e comum fundamentado em uma estrutura de conceitos específicos de um determinado domínio.

Mesmo não havendo consenso sobre sua definição, as ontologias apresentam características comuns, que estão presentes na grande maioria delas. Por possuírem propriedades distintas, é possível identificar tipos bem definidos.

Segundo Almeida e Bax (2003),

Os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de

⁵⁹ Tradução da autora: Uma Ontologia é uma especificação de uma conceitualização. A palavra Ontologia parece gerar muita controvérsia em discussões sobre AI [inteligência artificial]. Há uma longa história na filosofia, na qual ele se refere ao tema da existência. É confundida igualmente com frequência, com a epistemologia, que trata sobre o conhecimento e o saber.

No contexto do conhecimento que compartilho, eu uso o termo Ontologia para significar uma especificação de uma conceitualização. Isto é, uma Ontologia é uma descrição (como uma especificação formal de um programa) dos conceitos e dos relacionamentos que podem existir para um agente ou uma comunidade dos agentes. Esta definição é consistente com o uso da Ontologia como um conjunto de definições de conceito, mas mais geral. E é certamente um sentido diferente daquele que é usado na filosofia.

um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados) [...].

Ainda de acordo com Almeida e Bax (2003),

Algumas das propostas definem tipos de ontologias relacionando-as à sua função (Mizoguchi, Vanwelkenuysen & Ikeda, 1995), ao grau de formalismo de seu vocabulário (Uschold & Gruninger, 1996), à sua aplicação (Jasper & Uschold, 1999) e à estrutura e conteúdo da conceitualização (Van-Heijst, Schreiber & Wielinga, 1997), (Haav & Lubi, 2001).

Os autores observam que mesmo não havendo consenso, é possível observar que os tipos apresentados guardam semelhanças entre suas funções.

As ontologias diferenciam-se de mapas conceituais por serem formas diferentes de representar o conhecimento. As ontologias são formais e os mapas conceituais são informais. Os mapas conceituais são importantes porque dão sustentação à construção de ontologias, na medida em que descrevem o conhecimento de um determinado domínio. Eles se estruturam a partir de um conjunto de conceitos, e se caracterizam por sua flexibilidade e pelo pouco formalismo. Os conceitos são apresentados sob a forma hierárquica, estando os conceitos genéricos localizados na parte superior da estrutura, e os conceitos específicos na parte inferior da hierarquia.

As ontologias, de outro modo, formalizam o conhecimento sobre determinado domínio; elas se estruturam, estabelecendo um conjunto de relacionamentos entre os conceitos. Elas definem um vocabulário para um domínio.

Nos últimos anos, as ontologias têm sido tema de vários estudos como forma alternativa para a organização da informação; elas são criadas por especialistas e têm sido pensadas visando à organização do conteúdo das fontes de dados. Uma Ontologia define as regras que regulamentam a combinação entre termos e relações em um domínio. Os usuários, quando formulam suas consultas, utilizam conceitos definidos pela Ontologia. Com as ontologias se buscam resultados mais satisfatórios nos processos de recuperação da informação.

Nas palavras de Sales e Café⁶⁰ (2009)

Embora os textos referentes aos tesauros enumerem uma quantidade maior de tipos de relacionamentos entre conceitos, a flexibilidade do relacionamento por meio de axiomas, viabilizada pelo formalismo informático das ontologias, proporciona maior dinamicidade no tangente ao relacionamento conceitual.

⁶⁰ Original obtido na internet.

Encerrada nossas considerações sobre ontologias, apresentaremos na subseção seguinte a ferramenta – ExATOlp – que tem como propósito extrair automaticamente os sintagmas nominais de um *corpus*. Sua relevância para esta pesquisa está em agilizar a coleta de ULRIEs e a organização do conhecimento de um domínio.

2.2.1.2.2 ExATOlp - Extrator Automático de Termos para Ontologias em Língua Portuguesa

A ferramenta ExATOlp – Extrator Automático de Termos para Ontologias em Língua Portuguesa –, desenvolvida por Lopes⁶¹ (2012), tem o propósito de gerar automaticamente ontologias em língua portuguesa à partir de um *corpus*. Para esta pesquisa, a ferramenta *ExATOlp* será utilizada como apoio para a construção de vocabulários controlados especializados. Ela apresenta a vantagem de extrair automaticamente conceitos e noções, e de organizá-los em forma hierárquica. Ela

[...] é uma ferramenta que recebe um *corpus* anotado e extrai automaticamente todos os sintagmas nominais (SN) deste texto. A função primária da ferramenta é extrair termos candidatos a conceitos, auxiliando na construção de ontologias, glossários e outros recursos semânticos. Os sintagmas extraídos são salvos em listas que podem conter tanto os SN na sua forma original no texto, como em sua forma canônica. A ferramenta ainda oferece algumas opções de manipulação usuais para listas de termos como a aplicação de pontos de corte, comparação de listas e cálculo de medidas usuais de precisão e abrangência (LOPES; VIEIRA, 2010).

Segundo Lopes et al. (2009), a arquitetura do *software* é bastante simples. A ideia básica consiste em receber um *corpus* anotado, como um conjunto de arquivos .xml, e sucessivamente executar algumas operações básicas, visando gerar listas de termos significativos e, opcionalmente, algumas medidas numéricas.

As funcionalidades da ferramenta se estendem de tarefas fortemente baseadas em conceitos linguísticos, como a extração de sintagmas nominais, até tarefas essencialmente estatísticas, tal como o cálculo de métricas de avaliação. E, ainda, incluem tarefas como a localização e extração de termos, identificando os textos e as frases onde eles ocorrem.

A ferramenta utilizada para visualização de ontologias é o *Treebolic*, que possui código aberto e é desenvolvido na linguagem Java. *Treebolic* gera uma lista de conceitos e árvores hiperbólicas a partir de um *corpus* para exibição dos resultados. No Apêndice G apresentamos a lista de conceitos gerada pela ferramenta *Treebolic*, a partir do *corpus* de e

⁶¹ Doutora em Ciência da Computação pela Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS –.

sobre Platão. Esse *software* possibilita a geração de árvores hiperbólicas e a criação de *applets* para exibição dos resultados. Uma visualização de árvores hiperbólicas com origem no *corpus* construído para a geração do Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo pode ser vista no Apêndice H.

Concluimos estas ponderações, afirmando que vocabulários controlados especializados e ontologias são ferramentas utilizadas para representar e recuperar informação. Entendendo que as Linguagens Documentárias se constituem em um importante elo entre Sistemas de Recuperação da Informação e usuários, e que tais Sistemas utilizam essas linguagens para viabilizar a organização do conhecimento e representação e circulação da informação, dedicamos a subseção seguinte ao estudo desses Sistemas.

2.2.2 *Sistemas de Recuperação da Informação*

Uma das maiores dificuldades enfrentadas pelo usuário de um Sistema de Recuperação de Informações (SRIs) não é exatamente a recuperação de documentos, mas a recuperação de informações relevantes nos documentos recuperados para seus propósitos de pesquisa. O bibliotecário, como profissional da informação, é responsável pelo gerenciamento de tais Sistemas. Para desempenhar com sucesso este papel, ele precisa aprofundar seu conhecimento sobre o Sistema que ele gerencia, e conhecer os recursos que está a sua disposição para aumentar o grau de satisfação de seus usuários. Sendo assim, uma revisão de literatura sobre os SRIs se faz necessária no contexto desta tese, que tem como pressuposto de que o sucesso ou o fracasso de um Sistema de Recuperação da Informação depende de como os usuários percebem e compreendem a informação que está sendo disponibilizada para uso.

O que pretendemos com o desenvolvimento desta subseção é discutir a relação que ocorre entre SRIs e usuários. Essa discussão implica em uma análise das características dos SRIs e da relação que se estabelece entre os Sistemas e os usuários, visando identificar os aspectos que interferem na eficácia desta relação. Com tal propósito, inserimos uma discussão sobre os modelos de comunicação documentária vistos em revisão de literatura e uma análise desta relação sob parâmetros enunciativos e biblioteconômicos. Tendo um SRI a função de disponibilizar informações, acrescentamos também uma reflexão sobre informação, diferenciando-a de dados e de conhecimento, de modo a esclarecer o que entendemos por esses conceitos. E, complementando, entendendo que dentre os fatores que interferem na eficácia de um SRI está o estabelecimento de uma política de indexação, consideramos importante abordar esse tema nesta pesquisa.

Desse modo, esta investigação está organizada considerando os SRIs sob os seguintes aspectos: Conceituação e características dos SRIs; Eficácia dos SRIs; Comunicação documentária; Processo de comunicação especializada; Modelos de comunicação documentária; Caracterização da relação SRIs e usuários; Dado, Informação e Conhecimento; e Política de indexação.

2.2.2.1 Conceituação e características dos Sistemas de Recuperação da Informação

Os Sistemas de Recuperação da Informação tratam da representação, do armazenamento 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.

Um SRI tem como função principal disponibilizar informações contidas nos documentos indexados, a partir de uma descrição sintética, objetiva e representativa de seu conteúdo formal e temático. Ele integra diferentes processos: seleção, aquisição, indexação e busca e recuperação das informações. Os SRIs informatizados têm como característica o fato de que o auxílio do bibliotecário no processo de busca da informação e recuperação de documentos se torna 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.

Atualmente, os usuários desses Sistemas buscam cada vez mais eficiência, simplicidade e personalização da informação. Ainda que os recursos tecnológicos sejam um dos fatores importantes para a recuperação da informação, ocorre que os usuários enfrentam um desafio ao se defrontarem com as tecnologias disponíveis e com o excesso de informação.

Dentre as vantagens que Cendón (2005, p. 62) apresenta sobre os Sistemas de Recuperação da Informação eletrônica em relação aos não-automatizados, estão

[...] as facilidades que os SRIs automatizados oferecem para busca de informação, [...] maior número de pontos de acesso que os SRIs não-automatizados, podendo-se, muitas vezes, pesquisar palavras-chave que aparecem em qualquer ponto do registro, inclusive no resumo e no texto completo, quando estes estão disponíveis. Além disso permitem realizar pesquisas mais complexas, em que vários conceitos necessitam ser relacionados, pois pode-se combinar grande número de termos de busca com lógica booleana, de maneiras que não seriam possíveis nos SRIs impressos. Permitem também fazer rapidamente, buscas abrangentes, cobrindo vários anos de publicações. Essas e outras facilidades representam uma grande economia de tempo para o usuário, permitindo que uma pesquisa que

poderia tomar muitas horas de trabalho, se realizada manualmente, seja executada bem mais rapidamente, com o uso dos computadores.

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, a saber: das tecnologias utilizadas para a sua implementação, ou do conteúdo informacional a ser inserido. As características específicas do consulente, assim como a sua maneira pessoal de perceber e utilizar a informação e produzir conhecimento, não eram relevantes para a eficácia do Sistema. Até mesmo a maioria daqueles Sistemas considerados amigáveis não tinha como proposta resolver os problemas de informação do usuário, mas apenas oferecer-lhes condições facilitadas de uso.

Nessa nova relação entre os usuários e os SRIs, o processo de interação entre seus interlocutores apresenta sérios problemas. Durante muito tempo os Sistemas de Recuperação da Informação foram considerados como um bom recurso para a recuperação da informação, alcançando um grau de eficiência satisfatório nas pesquisas efetuadas pelos usuários. Ocorre que, antes, 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 da informação e recuperação de documentos, 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 imprescindível é compatibilizar e maximizar os recursos oferecidos por ambos.

Os catálogos eletrônicos, onde são registrados os cabeçalhos de assunto e os descritores selecionados para representarem a informação indexada, são reconhecidos como Sistemas de Recuperação da Informação, por também integrarem diferentes atividades biblioteconômicas como seleção, desenvolvimento de coleções, processamento técnico, busca da informação e recuperação de documentos. O objetivo principal desses catálogos é o de oferecer ao usuário o acesso às informações potencialmente contidas nos documentos, com a finalidade de maximizar o seu 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 dos usuários; mesmo assim, muito pouco se tem investigado sobre como as informações recuperadas são usadas por eles.

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 da informação e recuperação de documentos que o Sistema lhe 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. Nos dias atuais, a sociedade vem se tornando cada vez mais dependente da informação, e os mesmos Sistemas que filtram e permitem a circulação da 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: 1) a necessidade de informação deve levar em consideração a perspectiva da individualidade do usuário, 2) a informação a ser acessada deve estar contextualizada na situação real de onde ela emergiu, e 3) o uso da informação é determinado pelo usuário.

A partir dessas três constatações consideramos 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 podemos esquecer que a busca da informação é um processo dinâmico de tomada de decisão em que o usuário é o agente, o que significa dizer, que 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, com aquela que o usuário faz uso no processo de busca da informação e recuperação de documentos. A esse respeito fazemos duas observações: 1) na maioria das vezes a linguagem empregada pelo bibliotecário durante o processo de indexação não é do conhecimento daquele que utiliza o Sistema; e 2) 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. Com o propósito de discutir a eficácia dos SRIs, desenvolveremos a subseção seguinte.

2.2.2.2 Eficácia dos Sistemas de Recuperação da Informação

A interação entre o usuário e o Sistema é fundamental para que se estabeleça a circulação da informação, e alcance resultados significativos no processo de busca e recuperação das informações. Com esse propósito o bibliotecário precisa conciliar a linguagem criada pelo autor do documento a ser indexado, com 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 sob o ponto de vista das linguagens de indexação e no âmbito do conteúdo de um documento. No primeiro caso a insuficiência do Sistema pode ser verificada pelas seguintes razões: 1) os descritores/termos não serem suficientemente específicos, 2) a hierarquia ser insuficiente, e/ou 3) as 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, ou seja, o grau de precisão e o de revocação que o Sistema oferece.

Lancaster (2004), tal como os autores antes citados, 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 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 tende a perder em precisão, e que quando se ganha em precisão pode haver uma perda na revocação. O que observamos é que muitas das falhas apresentadas no processo de busca da informação e recuperação de documentos provêm das tarefas de indexação temática de um documento.

Em nossa opinião, as falhas que os Sistemas de Recuperação da Informação vêm apresentando são, fundamentalmente, decorrentes da precária interação que ocorre entre bibliotecários e usuários, 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.

O que verificamos na análise desses Sistemas, e conforme mencionado por Guinchat e Menou (1983, p. 287) é que

El acierto y la precisión son cualidades contradictorias. Cuanto más se precisa en la búsqueda, más grande es el riesgo de que no contenga los documentos que se han descrito en términos más generales, pero que pueden referir-se al tema, pues sí el acierto disminuye, el silencio aumenta. Al contrario, cuanto más amplia es la búsqueda o sea cuanto mayor número de documentos relativos al tema ésta pueda hallar, mayor es el riesgo de encontrar al mismo tiempo documentos no pertinentes, y por consiguiente de aumentar la imprecisión y el ruido.⁶²

Segundo Vital e Café (2011, p. 51)

Observando os princípios de especificidade e exaustividade, em relação aos critérios de revocação e precisão de um sistema de recuperação da informação, percebe-se que quanto maior o grau de especificidade, maior será a precisão e menor a revocação na recuperação. Na medida em que se diminui a especificidade, menor será a precisão, havendo uma probabilidade de aumento da revocação. A adoção do princípio de exaustividade atua de forma oposta. Quanto maior a exaustividade, maior a revocação e menor a precisão; quanto menor a exaustividade, maior a precisão e menor a revocação.

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

⁶² Tradução da autora: A revocação e a precisão são medidas de desempenho contraditórias. Quanto mais precisa é a busca, maior é o risco do Sistema não 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 relacionados ao tema forem localizados, maior será o risco de recuperar documentos não pertinentes, aumentando a imprecisão e o ruído.

o processo de indexação é bastante complexo, é praticamente inviável que se proceda a uma indexação exaustiva.

Dentre os autores que questionam e procuram respostas para a insatisfação presente nos processos de busca da informação e recuperação de documentos, mencionamos Araujo (1995, p.1), cuja tese discute essa questão.

Por que vêm falhando os sistemas de informação? E não há dúvidas de que vêm falhando: uma rápida análise da literatura da área evidencia uma série de estudos, projetos etc., visando corrigir, ao menos parcialmente, suas falhas e proporcionar ‘maior satisfação aos usuários’. Observa-se, inclusive, que tais estudos incluem ‘modernas abordagens’ ao planejamento, gestão e, principalmente, *marketing* de seus serviços e produtos, buscando a inserção dos sistemas de informação em uma sociedade pós-industrial, pós-moderna, conhecimento-intensiva.

Araujo (1995), ao tratar das falhas apresentadas pelos SRIs na busca da informação e recuperação de documentos, 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.

Em resposta à pergunta ‘por que vêm falhando os sistemas de informação’ e a consequente insatisfação do usuário, citamos Biscalchin e Boccatto (2011), que em um estudo sobre o uso de Linguagem Documentária em catálogos coletivos de bibliotecas universitárias, apontaram três aspectos principais: 1) a incompatibilidade existente entre a linguagem de busca e a adotada pelo Sistema, 2) a ausência de remissivas, e 3) as inconsistências sintático-semânticas verificadas na linguagem. A pesquisa revelou que a grande maioria dos usuários desconhece a linguagem usada pelo Sistema. Acrescentamos a isso o desconhecimento dos recursos de pesquisa que a base oferece, fundamentada na experiência da autora, trabalhando por mais de 25 anos em bibliotecas universitárias. Lima (1998) observa, ainda, que a presença de várias Linguagens Documentárias em um único Sistema prejudica a circulação da informação e, desse modo, a comunicação entre o Sistema e o usuário.

Kuramoto (1996, p. 185), por sua vez, defende a ideia 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.

Ao discutir os problemas decorrentes do tratamento e recuperação da informação textual, Kuramoto (1996) propõe como abordagem alternativa o uso de sintagmas nominais como termos de indexação.

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 no Sistema e relevantes para os seus propósitos de pesquisa, pois ao localizar textos sobre uma das formas verbalizadas, dificilmente fará uma nova pesquisa com 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: Academia, Academia de Platão e academia platônica com um único sentido.

Com o propósito de melhorar a eficácia dos SRIs defendemos a ideia de que os Estudos de Terminologia oferecem preciosa contribuição, em especial pelo referencial teórico que se fundamenta nas seguintes noções: unidade terminológica, unidade fraseológica especializada, variação terminológica e visão de texto sob o olhar da Terminologia Textual.

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 lexicais de representação da informação especializada, viabilizando uma circulação mais eficiente entre o usuário e os SRIs.

Um planejamento, para o desenvolvimento de um SRI, é 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, com as necessidades de pesquisa do usuário.

Planejar 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 subjetiva quanto à seleção das seções a serem indexados ou pelo uso de uma linguagem de indexação inadequada.

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, e da política de indexação construída para atender os objetivos da Instituição que possui o documento. Um resumo de trabalho apresentado em congressos e simpósios, por exemplo, não deve receber o mesmo tratamento dado a um artigo de periódico, pois são textos que apresentam características textuais bem diversificadas. Desse modo, a exaustividade e a especificidade a ser dada no momento da indexação devem estar relacionadas não apenas com o Sistema a que vai servir, mas também com o documento que está sendo indexado.

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 item indexado vai ser representado por um descritor/termo geral ou específico. Elas são importantes, porque trazem consequências diretas para o processo de busca da informação e recuperação de documentos, pois é sabido que uma informação indexada apenas pelo descritor/termo genérico inviabiliza a recuperação de um documento e de uma temática pela sua especificidade. Como resultado de um procedimento inadequado em relação à coleta e seleção 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, podemos afirmar que o usuário tem se mostrado insatisfeito com os SRIs. Dentre os aspectos que interferem na insatisfação do usuário, é importante abordar as consequências 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 mencionamos não apenas o bibliotecário, mas todos os profissionais que lidam com a informação.

Em relação ao desempenho e 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 a 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 circulação 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 ideia 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 exigências 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: 1) as mudanças que ocorrem no processo de busca da informação e recuperação de documentos, advindas de novas formas de identificar, localizar e obter documentos; 2) a possibilidade de intercâmbio de dados bibliográficos *on-line* entre os diversos Sistemas de Informação; 3) 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 à busca do equilíbrio entre a revocação e a precisão.

Na próxima subseção analisaremos a relação que se estabelece entre SRIs e usuários, sob a perspectiva da comunicação documentária. A investigação que propomos a seguir é a de conceituar comunicação documentária, e a de verificar como os modelos que representam esse tipo de comunicação são usualmente apresentados na literatura da área biblioteconômica. Neste domínio, esses modelos são entendidos como uma representação do modo como se dá a comunicação entre os Sistemas de Recuperação da Informação e os usuários.

2.2.2.3 Comunicação documentária

Comunicação documentária, conforme Lima (1998), é o processo de comunicação entre o acervo de um Sistema de Informação e o seu usuário. Ela se apresenta quando um Sistema codifica o conteúdo informacional dos documentos, objetivando a sua recuperação pelo usuário. A Linguagem Documentária é o código utilizado nesse processo. Desse modo, para que a comunicação documentária se efetive, é necessário que a representação da informação transforme a informação presente em um documento, em uma representação condensada dos itens informacionais identificados no texto.

Ainda segundo Lima (1998), a comunicação documentária é frequentemente prejudicada pela diversidade de representações da mesma informação. Para que a comunicação documentária se efetive com alto grau de satisfação para o usuário, a Linguagem Documentária de um Sistema de Informação é fundamental. Usada como um

código no processo de comunicação documentária, ela viabiliza melhores condições para uma adequada circulação da informação do Sistema ao usuário. De acordo com a autora referenciada, uma comunicação mais eficaz pode ser alcançada com a compatibilização das Linguagens Documentárias com os estudos desenvolvidos em Terminologia.

Os modelos de comunicação documentária, conforme visto em revisão de literatura, têm origem na teoria da informação proposta por Shannon e Weaver (1949). Esta teoria entende que o processo de comunicação se dá a partir de um sistema (fonte), que influencia outro sistema (receptor), transmitindo signos por intermédio de um canal. O canal é o meio que permite a comunicação e que viabiliza a circulação da informação. Esta teoria, ainda que tenha recebido alterações no decorrer do tempo, é a que usualmente fundamenta o processo de comunicação documentária.

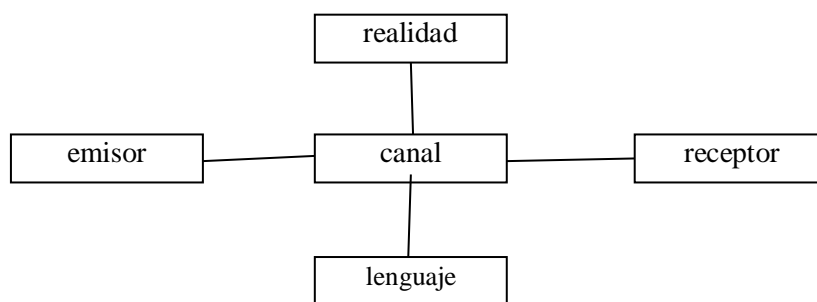
Dando seguimento a esta investigação, trataremos o tema da comunicação especializada a partir do seguinte entendimento: ela tem como propósito o atendimento das necessidades de informação dos especialistas, e ela ocorre em um Sistema de Informação que geralmente é subordinado a uma Instituição. Esse saber, acrescido da necessidade de compreender como se processa a comunicação especializada, é essencial na busca de um modelo de comunicação documentária.

2.2.2.3.1 Processo de comunicação especializada

Segundo Cabré (1993, p. 103-104),

El proceso de comunicación entre especialistas y sobre temas especializados no difiere mucho, en principio, del proceso general de la comunicación. Otras restricciones, que se imponen a los elementos que intervienen en la comunicación especializada, le confieren una especificidad que no tiene la comunicación general.

Todo esquema comunicativo presupone, como ha establecido Jakobson, los cinco elementos siguientes:



El acto de comunicación pone en juego un sexto elemento que relaciona simultáneamente los cinco elementos anteriores: el mensaje o texto, que, restringido al área del lenguaje, puede ser escrito u oral.⁶³

Ainda segundo Cabré (1993), além desses elementos, é preciso considerar que no ato comunicativo que ocorre entre emissor e receptor, estão implícitas duas variáveis: emissor e receptor possuem um conhecimento prévio da realidade que se deseja comunicar; e ambos possuem um conhecimento das normas sociais que regulam o uso da linguagem, das formas linguísticas e das convenções expressivas, como tipos de texto, normas de representação gráfica, formas padronizadas, etc.

E, ainda, os interlocutores participam do ato de comunicação com intenções e expectativas determinadas. As intenções do emissor condicionam o tipo de comunicação a ser estabelecida. Na comunicação especializada, ocorre uma série de restrições que a diferencia da comunicação que se dá entre não-especialistas. Na comunicação especializada, os interlocutores são especialistas em uma temática, e se comunicam sabendo que compartilham um determinado grau de informação sobre o tema que está sendo discutido. O mundo de referência de sua comunicação se restringe à área de sua especialidade. O sistema de comunicação que utilizam inclui a terminologia específica de uma área, e também a linguagem natural, uma vez que a linguagem especializada tem origem na linguagem natural.

Esclarecido o processo de comunicação especializada, conforme exposto por Cabré (1993), passaremos a discutir os modelos de comunicação documentária vistos em revisão de literatura. Em revisão de literatura, constatamos que os modelos de comunicação identificados têm origem em propostas elaboradas por engenheiros atuantes na área de sistemas de comunicação entre máquinas. Durante muito tempo esses modelos foram hegemônicos no campo da Ciência da Informação. Hoje, de acordo com teóricos da informação e bibliotecários, tais modelos tornaram-se obsoletos por excluírem o ser humano do processo, e por desprezarem o contexto e o conteúdo da informação.

⁶³ Tradução da autora: O processo de comunicação entre especialistas e sobre temas especializados não difere muito, em princípio, do processo geral da comunicação. Outras restrições, que se impõem aos elementos que intervêm na comunicação especializada, lhe conferem uma especificidade que não se encontra na comunicação geral. Todo esquema comunicativo pressupõe, como foi estabelecido por Jakobson, cinco elementos: emissor, canal e receptor e, ainda, realidade e linguagem. O ato de comunicação coloca em jogo um sexto elemento, que relaciona simultaneamente os cinco elementos anteriores: a mensagem ou texto que, restrito a área da linguagem, pode ser escrito ou oral.

2.2.2.3.2 Modelos de comunicação documentária

Iniciamos o exame desses modelos, salientando o fato de que o processo de comunicação especializada, conforme entendido por Cabré (1993), acrescenta dois elementos aos da comunicação da língua comum. Além de fonte (emissor), canal e receptor, há a considerar linguagem e realidade.

Conforme Lima (1998, p. 18) “[...] o primeiro modelo de comunicação documentária, ainda que o autor não o designasse assim, foi proposto por Coelho Netto [...].” Neste modelo, a comunicação entre o acervo (fonte), o usuário (receptor) e o bibliotecário (canal) pode ocorrer de modo direto ou indireto, ou seja, com o auxílio do bibliotecário (modo indireto), ou sem (modo direto); nesse caso, o usuário está coletando diretamente da fonte (acervo) aquilo que necessita.

Lima (1998) observa que neste esquema não é mencionada a intersecção entre os códigos da fonte e do receptor, tal como ocorre nos esquemas tradicionais de comunicação e na comunicação especializada. A autora observa, ainda, que a comunicação direta se dá quando o usuário acessa o Sistema sem a intervenção do bibliotecário. Entretanto, se considerar que ao estabelecer comunicação com um Sistema, o usuário está acessando a Linguagem Documentária produzida pelo bibliotecário, esta comunicação não pode ser entendida como uma comunicação direta. Assim, no esquema proposto por Coelho Netto, a rigor há sempre uma situação de comunicação indireta.

Desse modo, a Linguagem Documentária, ao codificar a informação presente em um documento, permite que essa informação seja decodificada pelo usuário; é ela que viabiliza o processo denominado por Lara de comunicação documentária.

Lara (2002) entende que as representações documentárias são utilizadas para transmitir informação. No processo de transferência da informação⁶⁴ é possível identificar: 1) o emissor (instituição que formula o sistema documentário), 2) o canal (veículo pelo qual são veiculadas as mensagens documentárias e que seriam os catálogos eletrônicos), 3) o código intermediário (a Linguagem Documentária usada na indexação e na recuperação da informação e que deve funcionar, simultaneamente, como um sistema de comunicação e de significação), e 4) o destinatário (usuário final do Sistema). Pelo modelo de comunicação

⁶⁴ Utilizaremos a expressão transferência da informação, no sentido que esta expressão é entendida pelos autores citados. Mas usaremos a expressão circulação da informação, sempre que tivermos nos referindo à relação que se estabelece entre Sistemas e usuários.

documentária, assim proposto, a relação entre SRIs e usuários é caracterizada de forma bem simplificada, conforme visualizado no quadro seguinte.

QUADRO 1 – Modelo de comunicação documentária

Elementos da comunicação na teoria da informação	Representação dos elementos da comunicação na comunicação documentária
emissor	instituição que formula o sistema documentário
canal	catálogo eletrônico
código intermediário	Linguagem Documentária
destinatário	usuário final do Sistema

Fonte: a autora

Conforme esse modo de entender

O usuário é o ator que dispara uma situação de comunicação, pois é o ato da recepção que determina o estabelecimento ou não de uma comunicação. [...] para que a comunicação documentária ocorra é necessário que a informação documentária – produto que é transmitido nesse processo – seja portadora de um significado tal que o usuário possa ser capaz de processá-la mentalmente, a fim de obter a informação desejada. A comunicação documentária é então um processo que exige, de um lado, a informação documentária e, de outro, o usuário que se apropria desta informação. Os sistemas documentários lidam com atos de comunicação materialmente explícitos, isto é, o documento, sua representação, a pergunta do usuário, a ação comunicativa do profissional de informação etc. (ORTEGA; LARA, 2010, p. 9).

Ainda segundo as autoras mencionadas, para que o processo de comunicação documentária se efetive, é necessário que ocorra duas situações: a informação documentária – produto que é transmitido neste processo – seja portadora de um significado, e o usuário – ator que dispara uma situação de comunicação – seja capaz de processar a informação documentária, para obter a informação desejada.

Segundo Lima (1998, p. 93),

Da mesma forma que o processo tradicional de comunicação, podemos observar que, na comunicação documentária, o Receptor recebe informações da Fonte através da utilização de um Código comum. Na comunicação documentária a Fonte seria o Acervo, o Receptor o Usuário e o Código a Linguagem Documentária, sendo que a transferência da informação só se efetiva quando o Receptor conhece o código utilizado pela Fonte.

Segundo a autora, ainda, este código – Linguagem Documentária –, para exercer a função de representação da informação deve contemplar as demais linguagens presentes no processo de comunicação documentária, como a linguagem especializada e a linguagem do usuário.

A comunicação documentária será considerada satisfatória na medida em que o usuário, ao buscar uma informação em um Sistema, obtiver respostas positivas adequadas às suas necessidades de pesquisa. A diversidade na representação documentária, caracterizada pelo uso de várias Linguagens Documentárias, impede que esta comunicação se efetive de forma eficaz.

Nesta pesquisa, defendemos a ideia de que a comunicação documentária deve ser vista a partir de parâmetros enunciativos e terminológicos. Sob o ponto de vista enunciativo, desenvolveremos uma proposta no capítulo 3 *Linguagens documentária sob um novo referencial*, que se sustenta na ideia de interação entre interlocutores. E, para que esta interação ocorra de forma satisfatória, recomendaremos que as Linguagens Documentárias estejam fundamentadas em princípios terminológicos.

O léxico temático configura-se, portanto, como um componente lingüístico, não apenas inerente, mas também a serviço de comunicações especializadas, posto que os termos transmitem conteúdos próprios de cada área. Por isso, os termos realizam duas funções essenciais: a de representação e a de transmissão do conhecimento especializado. Ao circunscreverem conteúdos específicos, as terminologias auxiliam também a elidir ambigüidades e jogos polissêmicos, freqüentes no uso do chamado léxico geral da língua, contribuindo para uma desejada precisão conceitual. (KRIEGER, FINATTO, 2004, p. 17).

Finalizamos esta reflexão, salientando que na comunicação especializada a linguagem utilizada é a linguagem especializada, constituída essencialmente por uma terminologia de um domínio específico. Os interlocutores são especialistas de uma área de conhecimento e o mundo de referência de sua comunicação se restringe a um campo específico do saber formalmente conceituado. A terminologia veiculada na comunicação especializada designa conceitos próprios de uma especialidade, caracterizando-se pela monorreferencialidade, uma vez que estabelecem uma única referência com o mundo exterior.

Na subseção que segue, analisaremos a relação que se estabelece entre os SRIs e os usuários, a partir de parâmetros enunciativos e biblioteconômicos.

2.2.2.4 Caracterização da relação SRIs e usuários

Na investigação dedicada às Linguagens Documentárias (2.2.1.1), examinamos essas linguagens sob diferentes aspectos, ressaltando o importante papel que desempenham na organização do conhecimento e recuperação da informação. Vimos a importância de serem desenvolvidas com rigor, para que o usuário possa acessar um Sistema de Informação e recuperar a informação que deseja.

Nesta exposição abordamos o tema dos SRIs, com o propósito de identificar suas características e sua eficácia na relação com os usuários, ou seja, sob a ótica da satisfação do usuário. Observamos que esses Sistemas de Informação se sustentam em uma linguagem especializada. Eles utilizam as Linguagens Documentárias para representar conceitos e expressões representativas de um domínio registrados nos documentos indexados, que são expressos por meio de descritores/termos.

Entendendo que uma análise das Linguagens Documentárias não é suficiente para caracterizar a efetiva interação entre usuários e Sistemas, decidimos proceder a essa análise sob dois enfoques: 1) verificar como se estruturam as Linguagens Documentárias, e 2) identificar os parâmetros necessários para que a interação entre usuários e Sistemas se efetive. Essa dupla análise tem como propósito buscar alternativas, não apenas para alcançar uma interação efetiva entre os usuários e os Sistemas de Recuperação da Informação, mas também para ressaltar a importância do papel que o bibliotecário ocupa nos processos de indexação.

A análise desses Sistemas teve início ainda no período de Mestrado (ARAÚJO, 2006), como aluna do Programa de Pós-Graduação em Letras, na linha de pesquisa Lexicografia e Terminologia: relações textuais. Na época, procurávamos um modelo de comunicação que fosse capaz de contemplar a complexidade que envolve a circulação da informação e a relação entre usuários e Sistemas de Recuperação da Informação.

Dentre as teorias revisadas, em busca de uma alternativa para minorar a insatisfação do usuário na relação com o Sistema, fizemos uma revisão de literatura em diferentes teorias que tratam da questão do sentido – teorias semânticas –. Entretanto, foi na Teoria da Enunciação, em especial no aparelho formal da enunciação proposto por Benveniste, que encontramos as condições essenciais para formular os parâmetros necessários para construir uma representação do processo de comunicação no âmbito documentário, o que, em Documentação é entendido como modelo de comunicação documentária. Esta proposta será apresentada no capítulo 3 – Linguagens Documentárias sob um novo referencial, na seção 3.4 Aparelho formal da enunciação e a relação SRIs e usuários: uma proposta alternativa aos modelos de comunicação documentária.

Assim, com a intenção de apresentar uma proposta caracterizando a relação que se estabelece entre SRIs e usuários, desenvolveremos logo a seguir algumas considerações sobre as características apresentadas por esses Sistemas. Pela identificação dessas características, percebemos que os modelos de comunicação até então apresentados não satisfazem as especificidades que configuram essa relação. Daí a proposição de uma proposta alternativa a esses modelos.

Analisando a relação que se estabelece entre os SRIs e os usuários, observamos que ela apresenta peculiaridades que não são explicadas nos modelos de comunicação documentária vistos em revisão de literatura. O que verificamos é que esses Sistemas apresentam características bem diferenciadas e muito mais complexas daquelas que fundamentam esses modelos. Sob a ótica desta pesquisa eles podem ser assim descritos:

1) são Sistemas em que os locutores – bibliotecário, usuário e SRI – são de natureza heterogênea, e o ato de comunicação que se dá entre eles ocorre em espaço e tempo diferenciados;

2) os códigos utilizados neste tipo de comunicação – linguagem especializada, vocabulário controlado e catálogo eletrônico – são reveladores de diferentes sistemas semânticos;

3) uma mesma unidade de informação – cabeçalho de assunto ou descritor – não recobre a mesma noção de um Sistema a outro;

4) o processo de indexação faz uso de mais de uma linguagem para a representação temática de um documento: a) linguagem especializada (linguagem do documento), e b) vocabulário controlado (linguagem expressa na lista de cabeçalho de assuntos ou tesouro);

5) o processo de busca da informação e recuperação de documentos oferece pesquisas por palavras, em bases diferenciadas quanto a sua estrutura e estratégias de pesquisa;

6) no processo de indexação, bibliotecários e usuários se posicionam como interlocutores, mas no processo de busca da informação e recuperação do documento, usuários e Sistemas são os interlocutores;

7) as marcas da subjetividade e intersubjetividade estão presentes e são relevantes, tanto no processo de indexação como no de busca da informação e recuperação do documento, tanto por parte do bibliotecário, como por parte do usuário;

8) o documento processado é a fonte da informação em ambos os processos.

Visando maximizar o uso da informação e a produção do conhecimento, propomos uma aproximação entre os processos anteriormente mencionados – processo de indexação e processo de busca da informação e recuperação de documentos –. Partindo desse modo de entender, orientamos a pesquisa no sentido de que a relação que se estabelece entre SRIs e usuários não se limita ao processo de busca da informação e recuperação de documentos. A construção de um Sistema, que se proponha a estabelecer uma relação com o usuário mais satisfatória, precisa considerar também aquela que ocorre entre o bibliotecário e o usuário durante o processo de indexação. Mas ainda ficava a pergunta: Que tipo de relação é esta?

Trata-se mesmo de uma relação de comunicação? Encontramos respostas a estes questionamentos na Teoria da Enunciação de Benveniste.

Ao analisarmos os registros bibliográficos de um SRI identificamos uma série de problemas que interferem no sucesso da interação que se estabelece entre os SRIs e os usuários, como por exemplo, a falta de consistência dos catálogos eletrônicos, a falta de especificidade temática na representação do conteúdo de um documento e a inexistência de descritores representativos do texto indexado. Constatamos que a insatisfação do usuário, em sua interação com tais Sistemas, muitas vezes tem origem em procedimentos inadequados de Análise Documentária.

Temos convicção que o sucesso ou o fracasso de um Sistema de Informação depende de como os interlocutores percebem e compreendem a informação que está sendo disponibilizada para uso. Com a automação dos Sistemas de Informação, a representação fiel do conteúdo informacional de um documento é uma exigência, independente de seu suporte. Quando há falhas nesse sentido, o processo de circulação da informação fica seriamente prejudicado.

Entendendo que uma discussão sobre Sistemas de Recuperação da Informação precisa estar fundamentada no esclarecimento dos conceitos de dado, informação e conhecimento, desenvolveremos a subseção seguinte com esse propósito. Esclarecemos que para fins desta pesquisa, interessa discutir apenas a informação que produz conhecimento na e para a comunidade científica.

2.2.2.5 Dado, Informação e Conhecimento

A discussão sobre esses três conceitos – dado, informação e conhecimento – será feita caracterizando-os pelas suas diferenças. Enquanto o dado é puramente sintático, a informação tem conteúdo semântico. A informação é particular, temporária e transitória, mas o conhecimento tem duração perene. O conhecimento, diferentemente da informação, é estruturado, coerente e frequentemente tem caráter universal.

Em nosso entendimento, o dado puro e simples não carrega nenhum tipo de informação; somente o dado formatado de uma maneira inteligível e útil ao usuário representa informação. A informação é criada a partir de dados inseridos em um contexto; é uma abstração informal que representa algo significativo para alguém. Acrescentamos, ainda, que a informação é entendida como um meio de se adquirir conhecimento para o desenvolvimento de uma área do saber, uma vez que todo conhecimento começa por algum tipo de informação.

Assim, por exemplo, teremos dados indexados quando uma obra que trata da filosofia da história é indexada pelos termos Filosofia e História, mas teremos informação, se essa mesma obra for indexada pela expressão Filosofia da História.

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.

O fato do 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 para a 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.

Quando o conhecimento é comunicado, 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).

É importante ressaltar que um conhecimento construído é normalmente retomado nas práticas comunicativas, sendo que a cada retomada há uma nova possibilidade de ressignificação. 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, pois as ideias 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 registradas em um suporte físico, as informações são instituídas em documento, o que permite que o conhecimento não se perca e possa ser compartilhado. Enquanto documento, elas recebem tratamento adequado para propiciar a criação de novos conhecimentos, a sua divulgação e socialização.

Ainda que 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. 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. “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).

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. Se o usuário procurar obras a respeito da ‘Linguística da enunciação’, utilizando como expressão de busca os termos ‘Linguística e Enunciação’, o resultado da pesquisa não obrigatoriamente oferecerá documentos que tratem exclusivamente sobre ‘Linguística da enunciação’, mas sim toda a sorte de documentos que tenham sido indexados com os termos ‘linguística e/ou enunciação’.

Sob esse ponto de vista, a partir 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. O acesso à informação ganha em operacionalidade e velocidade, mas perde em precisão.

Esclarecidos esses três conceitos – dado, informação e conhecimento –, abordaremos na subseção seguinte o tema política de indexação. Ela será direcionada para núcleos de pesquisa acadêmicos, que se caracterizam por possuir um acervo especializado, e por atender prioritariamente usuários com necessidades de pesquisa bem específicas. Esta seção terá início com uma breve exposição sobre o tema.

2.2.2.6 Política de indexação

Uma política de indexação tem por finalidade orientar a tomada de decisões para a otimização dos serviços, a racionalização dos processos e a consistência das operações envolvidas. Ela deve ser construída com base nos seguintes parâmetros: 1) objetivos da organização, 2) tipo de serviços a serem oferecidos, 3) identificação dos usuários, 4) reconhecimento de suas necessidades de informação, e 5) recursos humanos, materiais e financeiros disponíveis.

Dentre os princípios a serem observados em uma política de indexação, citamos:

1) cobertura de assuntos – atendendo a este princípio, torna-se necessário identificar as áreas que necessitam de um tratamento mais aprofundado, e aquelas que receberão um tratamento mais superficial;

2) seleção e aquisição de documentos – cabe aqui decidir sobre a extensão da cobertura do Sistema em relação as áreas de assunto do acervo documental;

3) processo de indexação – uma política de indexação precisa ser estabelecida sob dois aspectos: representação descritiva e representação temática. A representação descritiva trata da descrição dos aspectos formais do documento. Quanto à representação descritiva é preciso definir os instrumentos normativos para a catalogação, como códigos, manuais e *softwares*.

Em relação à indexação temática é preciso estabelecer os níveis de exaustividade e de especificidade requeridas pelo Sistema. Sobre a linguagem de indexação a ser adotada é necessário fazer a opção por uma linguagem livre ou controlada, e uma linguagem que atenda aos princípios de pré-coordenação ou de pós-coordenação. E, ainda, vale uma decisão sobre a capacidade de revocação e de precisão do Sistema. É importante observar que uma mesma publicação pode ser indexada de modo diferente, segundo os interesses de diferentes

instituições. Uma biblioteca universitária ou um centro de pesquisa, por exemplo, representam interesses diferentes;

4) estratégia de busca – a decisão a ser tomada aqui diz respeito à definição de responsabilidade quanto à realização de busca de informações no Sistema, ou seja, se será o bibliotecário ou o usuário que acessará diretamente o Sistema;

5) tempo de resposta do Sistema – a política a ser considerada sob este item é quanto ao tempo permitido para ser gasto no processo de busca da informação e recuperação de documentos, considerando aqui, essencialmente, as variáveis revocação e precisão;

6) formato de saída – uma definição sobre o formato de apresentação dos resultados de busca precisa ser estabelecida em uma política de indexação;

7) avaliação do Sistema – cabe a uma política de indexação, ainda, identificar a forma como o Sistema será avaliado, visando averiguar o grau de satisfação dos usuários, as falhas que o Sistema apresenta, e a forma como as falhas percebidas podem ser corrigidas.

A definição de uma política de indexação, construída em consonância com a Instituição em que está inserido o SRI e com os usuários a que se destina, é indispensável para o alcance dos objetivos de um Sistema de Informação.

Feitas essas considerações, esboçaremos uma política de indexação direcionada aos núcleos de pesquisa acadêmicos. Lembramos, aqui, que a comunidade usuária de um núcleo de pesquisa acadêmico é constituída fundamentalmente por alunos de pós-graduação, professores e pesquisadores, ainda que viabilize o acesso aos alunos de graduação.

Quanto à comunidade usuária de um núcleo de pesquisa, ressaltamos que os alunos de pós-graduação se diferenciam do aluno de graduação por terem necessidades temáticas bem específicas. A maioria dos alunos de graduação centra suas buscas em catálogos eletrônicos, por autor e título. Eles têm como propósito localizar as obras citadas nas bibliografias básicas e complementares recomendadas pelos professores de uma disciplina.

De outro modo, os alunos de pós-graduação buscam não apenas obras de determinados autores, mas também temas específicos. Suas buscas não se limitam às bibliografias recomendadas; eles precisam ir além, seja para elaborar trabalhos acadêmicos, ou para produzir artigos, dissertação ou tese.

Tal qual os alunos de pós-graduação, professores e pesquisadores não se limitam a localizar documentos. Para o desenvolvimento de suas atividades acadêmicas e de pesquisas eles precisam buscar informações específicas para aprofundar seus conhecimentos. A necessidade de informação desses usuários exige um tratamento da informação também diferenciado. Desse modo, a análise temática de um documento em um núcleo de pesquisa

acadêmico precisa ser mais aprofundada e a terminologia usada, mais específica e precisa. É esse procedimento que vai permitir uma maior circulação dos documentos que integram a coleção documental de um núcleo de pesquisa acadêmico e a formação de novos conhecimentos em um domínio.

Quanto à cobertura de assuntos, consideramos que o acervo de um núcleo de pesquisa acadêmico deve se restringir a uma temática específica, como ‘filosofia’ ou ‘relações internacionais’, por exemplo, ainda que ‘relações internacionais’ seja um tema interdisciplinar. Seu acervo deve se diferenciar do acervo de bibliotecas universitárias, por armazenar em sua coleção documental uma temática bem específica, e privilegiar a aquisição de obras na língua original. A seleção e a aquisição de documentos devem estar direcionadas para as obras dos principais teóricos e críticos de um domínio.

Na área da Filosofia, por exemplo, um núcleo de pesquisa acadêmico deve reunir as obras dos principais teóricos e de seus comentadores. Por comentadores entendemos as pessoas que dedicam boa parte da vida a analisar, interpretar e criar novos usos para a filosofia. A leitura dos textos produzidos pelos comentadores é importante para o aprofundamento das ideias dos filósofos.

Levando em consideração que se acredita ser necessário descrever um documento de forma exaustiva, sob o ponto de vista da representação descritiva, visando o atendimento de um usuário especializado, um núcleo de pesquisa acadêmico deverá adotar o nível 3 do Código de Catalogação Anglo-Americano, 2. ed., (AACR2).⁶⁵ São itens obrigatórios na catalogação de suas obras os seguintes: todos os tipos de autoria mencionados na obra, como editor, tradutor, colaborador, autor da introdução, dentre outros; o título original sempre que for identificado na própria obra ou em consulta a outros catálogos eletrônicos; indicação de edição (2.ed., por exemplo), local e ano de publicação; indicação de série e subsérie; número total de páginas ou volumes; ilustração; notas específicas do documento e notas de conteúdo. Deverão ser feitas entradas secundárias para todos os autores e obras mencionados.

A decisão sobre a ferramenta a ser usada visando à otimização do serviço de catalogação ficará a cargo da Instituição a que o núcleo de pesquisa acadêmico está vinculado. O formato MARC 21⁶⁶, por permitir o registro legível por máquina e viabilizar o intercâmbio das informações, tem sido utilizado por inúmeras bibliotecas. Dos *softwares* que utilizam o

⁶⁵ O AACR2 apresenta três níveis de catalogação descritiva. O primeiro nível é o mais simples, oferecendo pontos mínimos de acesso. O segundo nível amplia a descrição do documento, possibilitando sua representação por um número maior de pontos de acesso. O terceiro nível faz uso de todos os elementos constitutivos no código.

⁶⁶ Formato MARC – Machine Readable Cataloging / Catalogação Legível por Computador.

formato MARC 21, mencionamos três possibilidades de aplicação: 1) Pergamum, 2) Aleph e 3) Virtua.

Quanto às tabelas de classificação a serem adotadas, recomendamos a CDD ou a CDU para a classificação de assuntos, e a tabela Cutter-Sanborn, para a classificação de autoria. Em relação ao número de chamada, sugerimos que seja atribuído um código de fácil memorização.

Sob o olhar da representação temática, é preciso buscar equilíbrio entre os níveis de exaustividade e de especificidade, mas sempre se esforçando para alcançar a precisão conceitual. Isso significa dizer que os conceitos e noções representativas de um domínio presentes no texto a ser indexado devem fazer uso de palavras ou expressões que representem, com precisão e especificidade, o que está sendo indexado. Essa representação deve estar fundamentada nas noções advindas dos Estudos de Terminologia, reconhecidas como unidades terminológicas e unidades fraseológicas especializadas. Assim, por exemplo, o assunto 'ética' deve estar representado de forma específica, conforme é tratado na área de conhecimento do núcleo. No âmbito da Filosofia, este tema pode admitir as seguintes representações: 'ética platônica', 'ética aristotélica' e 'ética kantiana', dentre outras. A variação, sempre que houver, deve aparecer no Sistema sob a forma de remissivas.

Se não houver vocabulários controlados especializados para dar suporte ao bibliotecário na seleção e uso de unidades especializadas de um domínio, cabe aos responsáveis pelo núcleo de pesquisa acadêmico – bibliotecários e pesquisadores – unir esforços para a construção destas ferramentas. Neste caso, recomendamos que a coleta das unidades lexicais de representação da informação especializada seja feita a partir de um *corpus* formado por *abstracts*, ou por *abstracts* e textos de caráter acadêmico de autoria dos próprios especialistas. No caso de vocabulários controlados especializados contemplando uma grande área de especialidade, como Filosofia, o uso de *abstracts*⁶⁷ é suficiente para a formação do *corpus*. Mas quando os vocabulários controlados especializados contemplarem áreas específicas de um domínio, como Filosofia platônica ou Filosofia aristotélica, por exemplo, o *corpus* precisa ser formado por *abstracts* e textos completos.

Em atendimento ao princípio estratégia de busca, será definido que cabe ao usuário o acesso à base de dados, e ao bibliotecário dar o suporte que o usuário necessita, uma vez que é ele quem mais conhece os recursos do Sistema. O tempo de resposta do Sistema deve ser

⁶⁷ Os *abstracts* enquadram-se na categoria de textos do âmbito científico-acadêmico e se propõem a estabelecer a comunicação entre os pares; eles correspondem ao nível mais alto de especialização e condensam a informação central da investigação.

medido com base no tempo gasto no momento da recuperação da informação, considerando duas variáveis: a revocação e a precisão. O formato de apresentação dos resultados será definido pelo Sistema ao que o núcleo de pesquisa acadêmico está vinculado.

E, por fim, a avaliação do Sistema deverá ser feita no cotidiano das atividades do núcleo de pesquisa pelo bibliotecário de referência, visando avaliar o nível de satisfação do usuário, e identificar as falhas que o Sistema apresenta. Aqui, é importante considerar a necessidade de integração entre o bibliotecário de referência e o bibliotecário de indexação.

Com a subseção Política de indexação finalizamos o capítulo dedicado à revisão de literatura, que foi constituído por duas seções: Estudos da Linguagem e Ciência da Informação. Nossa revisão de literatura teve início com o tema da Teoria da Enunciação de Benveniste. Na sequência, tomamos para discussão os Estudos de Terminologia e a Linguística de *Corpus*. Nosso propósito, com esses três temas que se inserem no âmbito dos Estudos da Linguagem, era fundamentar nossa discussão, no que diz respeito à eficácia dos Sistemas de Recuperação da Informação e das Linguagens Documentárias.

A Linguística de *Corpus* está presente nesta pesquisa em razão de seus pressupostos teóricos e de sua metodologia para a consolidação da proposta a ser apresentada na construção de vocabulários controlados especializados. Com essa intenção, nossa pesquisa esteve centrada em dois temas que se inserem no âmbito da Ciência da Informação: Organização do Conhecimento e Representação da Informação, Sistemas de Recuperação da Informação.

É fundamentado nesta revisão de literatura que desenvolveremos o capítulo 3 Linguagens Documentárias sob um novo referencial.

3 LINGUAGENS DOCUMENTÁRIAS SOB UM NOVO REFERENCIAL

As questões investigadas nesta pesquisa estão relacionados à eficácia dos Sistemas de Recuperação da Informação (SRIs) e ao funcionamento das Linguagens Documentárias. Buscamos nos Estudos da Linguagem contribuições para a sua resolução, em especial na Teoria da Enunciação de Benveniste e nos Estudos de Terminologia.

Sob o referencial da Teoria da Enunciação, visamos o esclarecimento dos seguintes conceitos e noções: enunciação, linguagem (língua em ação), língua, estrutura, significado, sentido, semiótica, semântica, subjetividade, intersubjetividade, locutor, referência, contexto e categorias de pessoa, espaço e tempo.

Nos Estudos de Terminologia buscamos a compreensão dos seguintes conceitos e noções: unidades terminológicas (UTs), unidades fraseológicas especializadas (UFEs), comunicação especializada, linguagem especializada, textos especializados, variação terminológica e *corpus*.

Foi neste quadro teórico que encontramos sustentação para o desenvolvimento de um referencial enunciativo e terminológico, no âmbito dos Sistemas de Recuperação da Informação e das Linguagens Documentárias.

Este referencial permitiu a formulação de duas propostas, uma fundamentada na Teoria de Enunciação e outra nos Estudos de Terminologia. Com fundamentação no aparelho formal da enunciação proposto por Benveniste, formulamos uma proposta alternativa aos modelos de comunicação documentária concebidos a partir dos conceitos de emissor, canal, código e receptor. Tendo como sustentação teórica os Estudos de Terminologia, propomos a construção de vocabulários controlados especializados sob esse referencial.

O desenvolvimento dessas propostas será apresentado no decorrer deste capítulo, que será construído a partir das seguintes seções: 3.1 As Linguagens Documentárias sob o referencial dos Estudos da Linguagem, 3.2 As Linguagens Documentárias sob o olhar da Teoria da Enunciação de Benveniste, 3.3 A comunicação documentária e os SRIs vistos a partir de parâmetros enunciativos, 3.4 Aparelho formal da enunciação e a relação SRIs e usuários: uma proposta alternativa aos modelos de comunicação documentária, 3.5 Vocabulários controlados especializados sob o referencial dos Estudos de Terminologia, da Linguística da Enunciação e da Linguística de *Corpus*, e 3.6 Vocabulários controlados especializados: uma proposta para a sua elaboração a partir de ULRIEs.

Iniciamos este capítulo abordando as Linguagens Documentárias sob o quadro teórico dos Estudos da Linguagem. Nossa intenção, aqui, é a de reconhecer neste domínio um

referencial que, associado ao dos SRIs e das Linguagens Documentárias, possa contribuir para reduzir os problemas que têm origem na construção do sentido.

3.1 As Linguagens Documentárias sob o referencial dos Estudos da Linguagem

Uma das grandes discussões que se coloca no âmbito das Linguagens Documentárias e da eficácia dos Sistemas de Recuperação da Informação é o da construção do sentido, ou seja o da construção da referência que assegura a contextualização de um descritor, na passagem do processo de indexação para o processo de busca da informação e recuperação de documentos. Ressaltamos que o uso de um descritor contextualizado é uma das grandes dificuldades que estas linguagens apresentam.

Com o propósito de investigar o processo de construção do sentido, desenvolvemos as subseções Teoria da Enunciação de Benveniste (2.1.1) e Sistemas de Recuperação da Informação (2.2.2), inserindo nessas considerações uma discussão sobre os SRIs, comunicação documentária e nosso entendimento sobre dado, informação e conhecimento.

Na perspectiva enunciativa, consideramos conveniente para os propósitos desta investigação a forma como Benveniste (1988a, 1989a) entende tais noções, acrescida da visão de Martins (1990), que introduz o universo da não-pessoa na estrutura da enunciação. O entendimento deste quadro teórico é imprescindível para que possamos estabelecer algumas aproximações com os SRIs e as Linguagens Documentárias, o que é feito neste capítulo.

No âmbito documentário, nossa preocupação se voltou para os modelos de comunicação documentária apresentados em revisão de literatura. Esta revisão foi feita na subseção 2.2.2.3 Comunicação documentária. Da revisão efetuada concluímos serem esses modelos inadequados para explicarem a relação que se estabelece entre os SRIs e os usuários. Da associação dessa leitura com aquela efetuada em Teoria da Enunciação de Benveniste, surgiu a possibilidade de formularmos uma proposta que explicasse essa relação – SRIs e usuários –, e que contemplasse suas peculiaridades. Foi no aparelho formal da enunciação que encontramos os elementos que consideramos adequados para a formulação da proposta que apresentaremos na seção 3.4.

Dando continuidade a essa investigação, focamos nosso olhar na construção de vocabulários controlados especializados. Com este propósito consideramos relevante uma revisão de literatura no quadro teórico da Organização do Conhecimento e Representação da Informação, abordando dentre outros temas as Linguagens Documentárias, os sistemas nocionais, os fenômenos linguísticos, a Análise Documentária, os vocabulários controlados,

tesauros e ontologias. Relembramos que as ontologias serão entendidas nesta pesquisa com a função de agilizar a construção de vocabulários controlados especializados.

Partilhando de outro ponto de vista a respeito da fundamentação teórica das ferramentas documentárias, sejam elas sob o formato de listas de cabeçalhos de assunto ou de tesauros, sustentamos nossa investigação em parâmetros terminológicos e não propriamente lexicográficos.

Optamos, nesta pesquisa, pelas ULRIEs como as unidades dessas ferramentas. Tais unidades representam não apenas conceitos, mas também as expressões representativas de um domínio, sempre que se caracterizarem como unidades terminológicas ou unidades fraseológicas especializadas. Elas se fundamentam no uso e são extraídas das linguagens especializadas. Essas unidades, ainda que se assemelhem aos cabeçalhos de assunto no caso das listas de cabeçalhos de assunto, e aos descritores no caso dos tesauros, se caracterizam por terem uma abrangência maior e não estarem sujeitas a regras gramaticais como singular e plural, por exemplo.

A esse respeito citamos Currás (1995, p. 113) que afirma:

Aceitou-se internacionalmente que os termos que integram um tesouro devem ser substantivos, salvo em alguns casos muito particulares, como quando se trata de certas expressões adverbiais, que sejam absolutamente necessárias, no campo do conhecimento abordado pelo tesouro. [...] Por conseguinte, os termos simples serão → substantivos, → substantivos adjetivados, → verbos substantivados.

Ainda referenciando Currás (1995) selecionamos alguns exemplos sobre o uso de substantivos em tesauros, quanto a sua forma de apresentação, singular ou plural: 1) partes do corpo humano: cabeça, boca, mãos, olhos; 2) partes de uma árvore: troncos, galhos e raízes; 3) substantivos quantificáveis e que respondem a pergunta quantos ou quanto: animais, livros, partidos políticos (quantos) e farinha, mica, neve (quanto).

Esclarecemos que sob o olhar dos Estudos de Terminologia, entendemos tais noções a partir da ideia de unidades terminológicas e de unidades fraseológicas especializadas. Sob esta ótica teríamos cabeça, boca, mão, olho, tronco, galho, raiz, animal, livro, partido político, farinha, mica e neve, uma vez que é sob essa forma que o conceito é representado.

Dando continuação ao desenvolvimento deste capítulo, iniciaremos a próxima seção com uma breve exposição sobre as Linguagens Documentárias sob o olhar da Teoria da Enunciação de Benveniste. Temos como propósito identificar, na teoria deste linguista, conceitos e noções que possam contribuir para formular um novo referencial sobre a relação que se estabelece entre SRIs e usuários, e apresentar uma proposta alternativa aos modelos de

comunicação documentária construídos a partir das noções de emissor, canal, código e receptor.

3.2 As Linguagens Documentárias sob o olhar da Teoria da Enunciação de Benveniste

Com o propósito de investigar a construção do sentido, fizemos uma revisão de literatura em Linguística da Enunciação. Nesta revisão vimos que Benveniste (1988a, 1989a) estuda o funcionamento da língua a partir da oposição forma e sentido, com o propósito de examinar a questão da significação. Ele deixa claro em seu texto *A forma e o sentido na linguagem* que o princípio básico da língua é significar. O linguista propõe para a língua duas espécies e dois domínios, o da forma e o do sentido, e o semiótico e o semântico, respectivamente. Ele identifica na língua duas funções linguísticas: aquela de significar para a semiótica e a de comunicar para a semântica.

Com a noção de semântica, segundo propõe Benveniste, entendemos a língua em emprego e ação. É no domínio da semântica e com a função de comunicar que entenderemos a língua no âmbito das Linguagens Documentárias. Assim, a língua não será entendida como instrumento de comunicação, e sim como possuidora de uma função instrumental. Como vimos em revisão de literatura, opor o homem à linguagem (língua em ação) é colocá-lo em oposição à sua própria natureza.

A distinção entre significar e comunicar é relevante para o entendimento das Linguagens Documentárias. Retomando o que foi visto sobre a teoria de Benveniste, vimos que o linguista esclarece que significação é constitutivo da língua, e não algo acrescentado a ela. O sentido, por sua vez, é inerente ao sistema linguístico e é percebido pelos falantes de uma língua. Sobre o sentido, o linguista acrescenta que cada enunciado e cada termo do enunciado tem um *referendum*. Isso se explica na medida em que aceitamos a ideia de que a linguagem se refere ao mundo dos objetos.

Benveniste (1988a, 1989a) afirma que é próprio da linguagem significar. Para esclarecer seu ponto de vista, ele propõe o seguinte questionamento: isso significa? Chapéu existe? Sim. Chaméu existe? Não. Assim, conclui o linguista, o significado advém do signo e do próprio fato de que a linguagem tem como propriedade significar.

A noção de sentido, de outro modo, está associada à palavra e à frase, e não ao signo. Diferenciando signo de palavra, Benveniste caracteriza o signo como elemento linguístico conceitual, genérico e não circunstancial, e a palavra como a unidade que expressa noções

particulares, específicas e circunstanciais. Segundo Benveniste, a noção de sentido está relacionada com a noção de semântica, enquanto que a de significado, com a de semiótica.

Ainda segundo este teórico é possível afirmar que a semiótica tem no signo a sua unidade, enquanto que a semântica tem sua unidade na palavra. O signo é dotado de significação na comunidade daqueles que utilizam uma língua. A palavra guarda sua própria significação e suas referências; ela é necessária para a codificação do pensamento, da construção de uma ideia. É pela frase que nos comunicamos, explica o linguista, sendo assim, a frase é a expressão semântica por excelência. Isto significa dizer que é na palavra e na frase que uma Linguagem Documentária deve se fundamentar, e não no signo.

Com a frase nos deslocamos para o âmbito do discurso, para aquilo que o locutor quer falar. É com referência à situação de discurso que o sentido se constitui. A noção de referência é fundamental no processo de construção do sentido; ele decorre do próprio ato de enunciação, ou seja, do agenciamento de palavras para a construção de frases. O sentido resulta da atitude do locutor quando coloca a língua em ação; ele é sempre novo a cada enunciação, e é determinado por meio de palavras, na sua relação com o contexto de situação. Desse modo, o sentido é o resultado da ação que as palavras exercem umas sobre as outras.

A posição assumida por Benveniste sobre significação e sentido traz contribuições importantes para as Linguagens Documentárias. Decorrente da sua teoria, inferimos que, nas Linguagens Documentárias, um cabeçalho de assunto ou um descritor adquire sentido e com ele valor de informação a partir da referência a que se referem. Para exemplificar o que queremos demonstrar, citamos o descritor 'valor', que tem sentido diferenciado no quadro da filosofia e no quadro das finanças. É o contexto de situação, ou seja, a sua referência à situação de discurso, que irá determinar o verdadeiro sentido deste descritor.

Ainda considerando a questão do sentido, na qual se insere a construção da referência, trazemos para o âmbito dos SRIs e das Linguagens Documentárias o conceito de estrutura. Em nossa opinião, essa noção é fundamental na construção de SRIs e de Linguagens Documentárias. Entendemos que tal como na língua, tanto os SRIs como os registros bibliográficos devem estruturar-se a partir da noção de estrutura, de modo que cada item de um registro bibliográfico e cada registro de um Sistema seja reconhecido pelo seu todo e pela relação entre as partes.

Com esse referencial, esperamos minimizar os problemas enfrentados pelo usuário na recuperação da informação, causados pela perda de sentido do termo indexado. E, com ele, nos propomos a alcançar o objetivo de viabilizar que um item bibliográfico tenha condições

de recuperar informações e não apenas dados, como ocorre quando os Sistemas desvinculam a informação de seu objeto original, o texto indexado.

Ainda sobre a distinção que Benveniste faz sobre semiótica e semântica, o linguista nos diz que a semiótica se caracteriza como uma propriedade da língua, enquanto a semântica resulta da atividade do locutor que coloca a língua em ação. Sob a ótica da semântica, também, com a frase deixamos o domínio da língua como sistema de signos, e entramos no domínio da língua como possuidora de uma função instrumental, sendo sua expressão o discurso.

Outra contribuição importante, que podemos trazer da teoria benvenistiana para o quadro teórico dos SRIs e das Linguagens Documentárias, é que a ideia é expressa por uma palavra ou frase, e que ela é sempre um ato particular e pronunciado por um locutor. A frase é sempre um acontecimento diferente e seu sentido é dado no instante em que é proferida. Assim sendo, uma frase participa sempre do ‘aqui e agora’, realizando-se, desse modo, sempre no tempo presente. Relacionado ao ‘aqui e agora’ temos duas noções a destacar: o tempo da enunciação e a presença de um locutor.

A afirmação benvenistiana, que diz que a relação do locutor com a língua determina os caracteres linguísticos da enunciação, torna possível analisar aspectos desta relação, com o propósito de determinar empiricamente em que medida o modo como essa relação se estabelece e interfere na instauração da subjetividade e na qualidade discursiva.

O conceito de subjetividade na linguagem é mais uma contribuição da teoria de Benveniste para o âmbito documentário. Há vários textos de Benveniste que abordam este conceito. No texto *Da subjetividade na linguagem* Benveniste faz uma distinção entre categoria de pessoa e não-pessoa, e entre as categorias da língua e categorias do discurso. Em *O aparelho formal da enunciação* o linguista sistematiza sua Teoria da Enunciação. É com base nessa teoria que desenvolveremos uma proposta caracterizando a relação que se estabelece entre SRIs e usuários, e que será desenvolvida na seção 3.4. deste capítulo.

Concluída esta análise fundamentada na Teoria da Enunciação de Benveniste, passaremos a discutir, ainda sob esse referencial, a comunicação documentária e os SRIs. A discussão desse tema é relevante para esta pesquisa, uma vez que os modelos de comunicação apresentados em revisão de literatura, na área documentária, se mostraram insuficientes para explicar a relação que se instaura entre SRIs e usuários.

3.3 A comunicação documentária e os SRIs vistos a partir de parâmetros enunciativos

Percebendo que os modelos de comunicação documentária utilizados para explicar a forma como acontece a comunicação entre SRIs e usuários não satisfazem, pois não contemplam as especificidades que caracterizam a relação que se estabelece entre eles, conforme explicitado na subseção 2.2.2 – Sistemas de Recuperação da Informação, desenvolveremos uma reflexão sobre a comunicação documentária fundamentada em parâmetros enunciativos.

A orientação teórica desenvolvida nesta pesquisa com essa finalidade – explicar a relação que ocorre entre SRIs e usuários – advém do aparelho formal da enunciação formulado por Benveniste (1988a, 1989a), acrescido da reinterpretação de Martins (1990). Benveniste afirma que é a relação do locutor com a língua que determina os caracteres linguísticos da enunciação. Entretanto, como vimos anteriormente, na subseção 2.1.1.9 – Reinterpretação da Teoria da Enunciação de Benveniste, o linguista não contempla em sua proposta a possibilidade de considerar a presença de fatores externos ao enunciado, na relação que se estabelece entre os interlocutores da enunciação. Essa impossibilidade de considerar o conteúdo linguístico da própria enunciação, no enunciado produzido pelos interlocutores, decorre do fato da enunciação ser vista por Benveniste através de suas marcas formais.

Compartilhando da ideia de que a relação que se dá entre os interlocutores não é independente do conteúdo linguístico da enunciação, a reinterpretação de Martins (1990) torna-se relevante aos propósitos desta pesquisa. Sob seu entendimento, o processo de instauração da subjetividade resulta de uma relação tríplice ‘eu-tu-ele’, e não simplesmente de uma relação direta ‘eu-tu’.

Ao inserirmos a noção de subjetividade no âmbito dos SRIs e das Linguagens Documentárias, concebemos o bibliotecário e o usuário como agentes dos processos que conduzem à produção do conhecimento, e não simplesmente como emissor e receptor desses processos. E, ainda, admitindo que a instauração da subjetividade resulta de uma relação tríplice ‘eu-tu-ele’, conforme proposta por Martins (1990), consideramos a noção de sentido como inerente ao universo semântico.

Sob essa concepção, acrescentamos outras duas considerações. Reconhecemos a palavra como unidade semântica possuidora de sentido no âmbito do discurso, e a referência como definidora de sentido no nível semântico, sendo a referência ao mundo parte integrante do ato de enunciação. E, entendendo que a noção de referência está associada à de contexto, partilhamos da ideia de que o sentido de uma palavra é obtido em relação ao contexto de

situação, e que fora do contexto, nem sempre é possível compreender o sentido que resulta da junção das palavras.

Foi pela compreensão dos conceitos e noções advindos da Teoria da Enunciação de Benveniste, já mencionados anteriormente, que se tornou possível perceber a complexidade da relação que se estabelece entre os usuários e o Sistema, e construir uma nova proposta, contemplando os diversos aspectos que caracterizam esta relação.

Analisando esta relação – usuários e Sistema –, que no âmbito biblioteconômico está inserida no processo de comunicação documentária, constatamos não tratar-se de um processo simples, nem se caracterizar como um processo de comunicação como usualmente se tem concebido. Observando-a sob a ótica da enunciação, passamos a reconhecê-la como um processo de comunicação intersubjetiva, resultado de atos de enunciação sempre únicos, produzidos por diferentes interlocutores. Sob essa visão, depreendemos que ocorre aqui uma mudança de paradigma no âmbito biblioteconômico.

Explicitando nosso raciocínio, fazemos as seguintes observações. A complexidade desse processo decorre de três fatos principais: 1) de envolver uma dupla situação de interação, 2) de se manifestar em diferentes espaços e 3) de ocorrer em momentos diferenciados. Para uma melhor visualização do que estamos propondo, apresentamos as informações dispostas no quadro que segue.

QUADRO 2 – Particularidades da relação que se estabelece entre SRIs e usuários

Interlocutores	Linguagem que utilizam durante o ato de enunciação	Momento em que o ato de enunciação se efetiva
bibliotecário e usuário	vocabulário controlado	processo de indexação
usuário e SRI	catálogo eletrônico	processo de busca da informação e recuperação de documentos

Fonte: a autora

Considerando as particularidades dessa relação – SRIs e usuários –, conforme apresentado no Quadro 2, fazemos a seguinte exposição. Segundo nosso entendimento há uma dupla situação de comunicação, uma entre bibliotecário e usuário, e outra entre usuário e Sistema. A situação de interação que ocorre entre bibliotecário e usuário se dá durante o processo de indexação, que tem no vocabulário controlado a origem da informação a ser

registrada em catálogos eletrônicos. A consulta aos catálogos eletrônicos, por sua vez, implica na presença de novos interlocutores, a saber, os usuários e os SRIs. Esta interação se manifestará em outro tempo e em outro espaço. O tempo será o processo de busca da informação e recuperação de documentos, e o espaço é o próprio catálogo eletrônico. Em ambos os processos – processo de indexação e processo de busca da informação e recuperação de documentos – a linguagem especializada está presente. Ela é a fonte de informação e o canal de comunicação que vai permitir que a interação entre bibliotecários e usuários, e usuários e SRIs se efetive de modo apropriado, ainda que não seja possível assegurar que a interação entre estes interlocutores seja exercida de forma plena.

Observamos, ainda, que esse tipo de comunicação se caracteriza por estabelecer relações de intersubjetividade entre interlocutores de natureza heterogênea – bibliotecário, usuário e SRI –. Acresce a isso o fato dos interlocutores não estarem presentes durante o ato de interação, que nesta pesquisa é entendido como um ato de enunciação. Em decorrência do fato de se manifestar em lugares diversos, tem a particularidade de ser constituído por diferentes tipos de códigos – linguagem especializada, vocabulário controlado e catálogo eletrônico –.

Fundamentado nesse referencial, damos início à elaboração de uma proposta alternativa, para aquela que vem sendo apresentada na literatura como modelo de comunicação documentária. Iniciamos nossas considerações a partir de dez proposições. São elas:

- 1) sob esse referencial, a comunicação documentária passa a ser reconhecida como comunicação intersubjetiva efetivada por atos de enunciação;
- 2) nas relações que se estabelecem entre SRIs e usuários são identificados dois processos enunciativos – processo de indexação e processo de busca da informação e recuperação de documentos – com características distintas;
- 3) no processo de indexação, o bibliotecário é o locutor que coloca a língua em ação, mas no processo de busca da informação e recuperação de documentos é o usuário que coloca a língua em ação;
- 4) no âmbito dos SRIs e das Linguagens Documentárias, diversos códigos interferem nas relações que ocorrem entre os interlocutores deste tipo de comunicação: linguagem especializada, vocabulário controlado e catálogo eletrônico;
- 5) o universo da não-pessoa está inserido nos atos de enunciação e é representado pela linguagem especializada, que é o canal de comunicação entre os interlocutores;

6) cabeçalhos de assunto e descritores são entendidos nesta pesquisa como unidades lexicais de representação da informação especializada (ULRIEs);

7) uma ULRIE adquire sentido diferenciado segundo o locutor, o tempo e o espaço próprio de cada ato de enunciação, seja no processo de indexação ou no processo de busca da informação e recuperação de documentos;

8) o sentido atribuído a uma ULRIE participa sempre do ‘aqui e agora’, estando assim, sempre ligado ao tempo presente. Ele é obtido no contexto de situação, ou seja, na relação com as demais ULRIEs de um registro bibliográfico;

9) a referência de um registro bibliográfico é a situação de discurso a que ele se refere. Voltamos a ressaltar que a noção de referência é fundamental na atribuição do sentido a ser dado para cada ULRIE;

10) o registro bibliográfico de um SRI precisa ser construído sob a ótica da noção de estrutura, de modo que uma Linguagem Documentária se constitua em um sistema do qual todas as partes são unidas por uma relação de solidariedade e de dependência.

Com o auxílio da Teoria da Enunciação, e a partir do aparelho formal da enunciação, foi possível identificar a seguinte situação: 1) o tipo de situação comunicativa que se estabelece no processo de comunicação documentária – comunicação intersubjetiva –; 2) as variáveis que envolvem esse tipo de comunicação – categoria de pessoa, espaço e tempo –; 3) os interlocutores que estão envolvidos – bibliotecário, usuário e Sistema –; 4) o espaço e tempo em que se efetiva esta relação – processo de indexação e processo de busca da informação e recuperação de documentos –, ressaltando que esta relação de interação comunicativa ocorre sempre no tempo presente, ainda que em espaço e tempo diferenciado.

Concluimos esta exposição afirmando que, a partir dos pressupostos teóricos enunciativos, entendemos a comunicação documentária como uma comunicação intersubjetiva. Ela se constitui em um ato de enunciação e em um processo de interação comunicativa marcado pela subjetividade e intersubjetividade. Neste tipo de comunicação, as categorias de pessoa, espaço e tempo, tal como entendidas em Benveniste, desempenham papel relevante, sendo que na categoria de pessoa, os atores do processo são o bibliotecário, o usuário e o Sistema.

A proposta alternativa aos modelos de comunicação documentária, fundamentada no aparelho formal da enunciação de Benveniste, será desenvolvida na seção seguinte. Nossa investigação sobre o aparelho formal da enunciação tem a intenção de propor um referencial fundamentado em parâmetros enunciativos, para a relação que se estabelece entre os SRIs e os usuários e que se estende aos bibliotecários, uma vez que ele não pode ser excluído dos

processos que envolvem esta relação. Com a proposta que desenvolveremos a seguir, temos como propósito oferecer uma alternativa aos modelos de comunicação documentária vistos em revisão de literatura e, com ela, dar maior relevância ao papel desempenhado pelo bibliotecário e pelo usuário nesses processos que envolvem a busca da informação e a produção do conhecimento.

3.4 Aparelho formal da enunciação e a relação SRIs e usuários: uma proposta alternativa aos modelos de comunicação documentária

Conforme visto na subseção 2.1.1 desta tese – Teoria da Enunciação de Benveniste –, Benveniste desenvolve o estudo da enunciação a partir do sistema pronominal da língua. As reflexões do linguista sobre a natureza dos pronomes e a sistematização de sua teoria estão na série de artigos intitulada *O homem na língua* (BENVENISTE, 1988a, 1989a). Resumidamente, a descrição que Benveniste faz do processo enunciativo é a seguinte: a categoria de pessoa ‘eu e tu’ se opõe a não-pessoa ‘ele’, que remete a uma situação objetiva.

De acordo com o esquema enunciativo proposto por Benveniste, a relação entre o locutor e a língua seria exterior à relação que se dá entre ‘eu e tu’, mas constitutiva no sentido de ser necessária para a realização do evento. Nesta pesquisa, a relação do sujeito com a língua assumirá um valor constitutivo, não apenas como condição de viabilização do evento enunciativo, mas também como elemento integrante do tipo de enunciado a ser produzido. Isso equivale a dizer “[...] que uma relação *eu-tu* intermediada por uma concepção linguística pautada pela norma interferirá no modo de instauração dessa relação, refletindo-se no enunciado produzido” (GOMES, 2003, p. 63). Essa proposição elaborada por Gomes (2003) tem Martins (1990) como referencial.

Na proposta de Gomes (2003), tal como em Martins (1990), o emprego do sistema pronominal de Benveniste sofre uma reinterpretação no tocante ao estatuto de pessoa. Na teoria de Benveniste, ‘eu e tu’ são elementos constitutivos da cena enunciativa; fora dessa relação de intersubjetividade, a linguagem é pura virtualidade, que se realiza somente no processo enunciativo, ou seja, naquele momento em que o ‘eu’ ao se assumir como locutor, instaura o ‘tu’ neste processo. O ‘ele’, a não-pessoa, tem a função referencial da linguagem, isto é, o ‘ele’ é um substituto de unidades semânticas que refere ao mundo dos objetos, das ideias, das ações. O ‘ele’ é excluído do circuito da enunciação; é não-constitutivo da relação ‘eu e tu’. Em Martins (1990), que insere o universo da não-pessoa na própria estrutura da

enunciação, o 'ele' é o conteúdo semântico desse universo que determinará o caráter de relação interpessoal.

No contexto desta pesquisa, a linguagem é entendida como elemento de interação social, conforme propõe a Teoria da Enunciação de Benveniste. Incluir a noção de interação no âmbito das Linguagens Documentárias significa mobilizar um conceito mais amplo de linguagem nas práticas documentárias. Sob esse enfoque, a relação que se estabelece entre os SRIs e os usuários passa a ser de intersubjetividade, fruto da interação que ocorre entre seus interlocutores. Acresce a esta relação, como interlocutor, o bibliotecário.

A opção por esta teoria justifica-se pelas seguintes razões: 1) ela entende que é pela linguagem que o homem se inter-relaciona, recebe e transmite uma mensagem; 2) é uma teoria que olha para o estudo da língua do ponto de vista do sentido, e que permite desenvolver uma reflexão sobre a construção da referência de um descritor na passagem do processo de indexação para o processo de busca da informação e recuperação de documentos; 3) viabiliza um olhar enunciativo no âmbito dos Sistemas de Recuperação da Informação, permitindo considerar peculiaridades próprias do tipo de relação que se constitui entre os SRIs e os usuários, e que não são visíveis nos modelos de comunicação apresentados na literatura documentária; 4) possibilita inserir o bibliotecário como constitutivo da relação SRIs e usuários, situando bibliotecário e usuário como agentes do processo que conduz à produção do conhecimento; 5) se caracteriza por ser construída a partir de noções como linguagem, subjetividade, intersubjetividade, referência e contexto, que são características identificadas nessa relação e presentes no referencial teórico que fundamenta a proposta que tem origem nesta pesquisa.

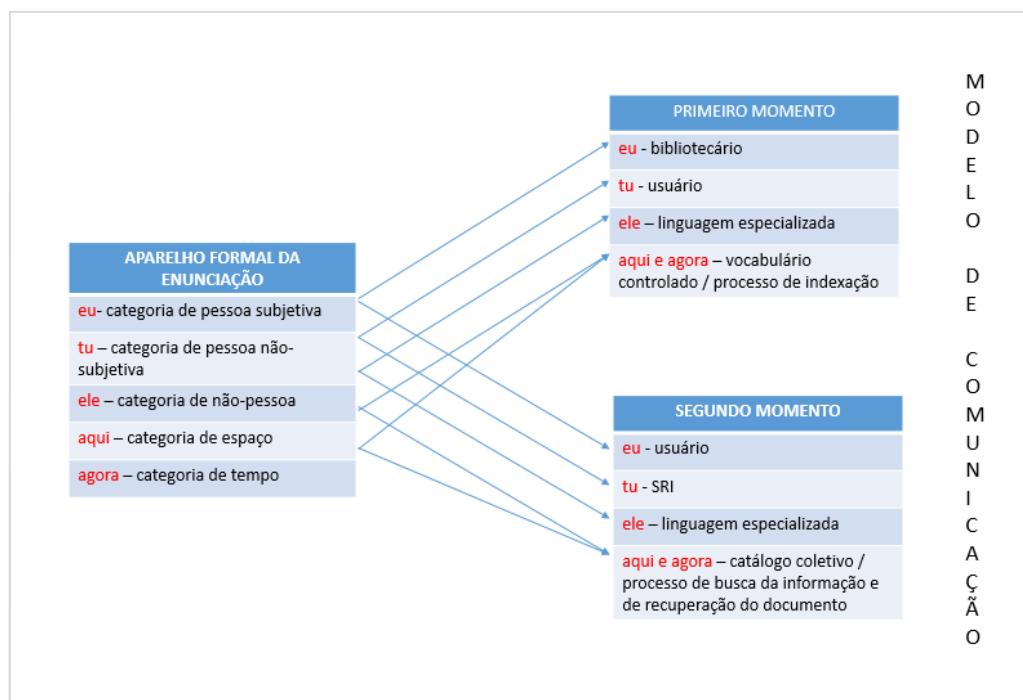
No âmbito desta pesquisa e conforme é entendido por Benveniste, o aparelho formal da enunciação é compreendido como um dispositivo que permite que as línguas possam ser enunciadas. Esse aparelho nada mais é do que a marcação da subjetividade na estrutura da língua. Segundo Flores e Teixeira (2005, p. 37)

[...] cada vez que o locutor se apropria do aparelho formal da enunciação – e por ele se apropria da língua toda – produz um uso novo e como tal irrepitível. Essa irrepitibilidade deve-se ao fato de que jamais tempo, espaço e pessoa – categorias fundamentais em enunciação – podem ser perenizados no uso da língua.

Foi pela compreensão das noções advindas do aparelho formal da enunciação proposto por Benveniste – subjetividade, intersubjetividade, referência, contexto, categorias de pessoa (eu e tu), espaço (aqui) e tempo (agora) – que se tornou possível perceber as peculiaridades e a complexidade da relação que se estabelece entre SRIs e usuários, assim como elaborar e

construir outro referencial sobre esta relação, contemplando os diversos aspectos que a caracterizam. Fundamentado neste referencial apresentamos no quadro seguinte (Quadro 3) uma síntese da proposta que desenvolvemos como alternativa aos modelos de comunicação vistos em revisão de literatura.

QUADRO 3 – Aparelho formal da enunciação e a relação SRIs e usuários



Fonte: a autora

A partir da configuração apresentada fazemos as seguintes considerações. Compreendendo que não se trata de uma situação de comunicação documentária, no modo como vem sendo apresentada, mas de uma comunicação intersubjetiva com origem em atos de enunciação, introduzimos nesta proposta a noção de interlocutores, identificando como interlocutores, não apenas o Sistema e o usuário, mas também o bibliotecário. Ao inserir a noção de subjetividade e intersubjetividade no âmbito desta pesquisa, reconhecemos uma situação de interação entre os diferentes interlocutores envolvidos. Identificamos duas situações de interação, uma entre bibliotecários e usuários, e outra entre usuários e SRIs. Observamos que elas ocorrem em espaço e tempo diferenciados, ou seja, na correlação com o vocabulário controlado (espaço) e processo de indexação (tempo), em um primeiro momento, e na correlação com o catálogo eletrônico (espaço) e processo de busca da informação e recuperação de documentos (tempo), em um segundo momento.

Justificamos nosso raciocínio, lembrando que os termos e expressões que identificam o conteúdo informacional de um documento são localizados nos vocabulários controlados e nos catálogos eletrônicos, durante os processos de indexação e de busca da informação e recuperação de documentos, respectivamente. Esclarecemos que os vocabulários controlados são as ferramentas em que os bibliotecários reconhecem os termos e expressões que identificam o conteúdo informacional de um documento, e que os catálogos eletrônicos são as ferramentas utilizadas pelos usuários para buscar informações e localizar os documentos que contêm as informações que desejam. Nas duas situações – processo de indexação e processo de busca da informação e recuperação de documentos – a linguagem especializada identificada no documento processado é a fonte de informação.

Assim, **sob o referencial do aparelho formal da enunciação de Benveniste, acrescido da reinterpretação de Martins, foi possível formular a seguinte proposta.**

A relação que se estabelece entre os interlocutores da comunicação documentária – bibliotecários, SRIs e usuários – se caracteriza como uma comunicação intersubjetiva, efetivada por atos de enunciação, em que estão presentes os índices de pessoa, espaço e tempo – eu, tu, ele, aqui e agora –, e as marcas de subjetividade, que são substancialmente relevantes, tanto no processo de indexação, quanto no de busca da informação e recuperação de documentos.

Antes de concluir nosso raciocínio, salientamos o fato de que a construção do sentido não é estabelecida *a priori*, e que no âmbito das Linguagens Documentárias ela depende de circunstâncias diferenciadas: 1) de um processo de interação entre os interlocutores – bibliotecário, usuário e Sistema –; 2) da contextualização das informações contidas em um registro, destacando o fato que um mesmo descritor pode adquirir sentido diverso em diferentes registros; 3) dos índices espaciais e temporais que se articulam no processo de indexação, assim como, no de busca da informação e recuperação de documentos; 4) de atos enunciativos sempre únicos implícitos nos processos mencionados.

Desse modo, entendemos que o processo de comunicação intersubjetiva que ocorre no âmbito documentário envolve tanto a relação que se estabelece entre os SRIs e os usuários no processo de busca da informação e recuperação de documentos, como a relação que ocorre entre bibliotecários e usuários durante o processo de indexação. Finalizando nossas considerações, concluímos que os modelos de comunicação documentária apresentados em revisão de literatura a partir de um emissor, canal, código intermediário e destinatário não satisfazem. Mas, se vistos sob a ótica de parâmetros enunciativos, é possível contemplar as diversas peculiaridades que os caracterizam.

Concluída nossa reflexão sobre o modo como se articula a relação SRIs e usuários, fundamentada em parâmetros enunciativos, iniciamos uma nova seção. Nesta, trataremos da geração de vocabulários controlados especializados, tendo como fundamentação teórica os Estudos de Terminologia, a Linguística da Enunciação e a Linguística de *Corpus*.

Reconhecendo os bibliotecários como agentes dos processos de indexação, e entendendo que a interação entre bibliotecários e usuários, e usuários e Sistemas precisa estar calcada em um bom vocabulário controlado especializado, defendemos a ideia de que é necessário qualificar esta ferramenta.

3.5 Vocabulários controlados especializados sob o referencial dos Estudos de Terminologia, da Linguística da Enunciação e da Linguística de *Corpus*

É com a intenção de construir ferramentas mais eficazes, que propomos a geração de vocabulários controlados especializados, buscando o referencial advindo dos Estudos de Terminologia, da Linguística da Enunciação e da Linguística de *Corpus*. Sob o referencial dos Estudos de Terminologia, trazemos para o âmbito das Linguagens Documentárias os seguintes conceitos e noções: linguagem especializada, texto especializado, unidades terminológicas, unidades fraseológicas especializadas e variação terminológica conceitual e denominativa.

Com base nesse referencial as unidades terminológicas e as unidades fraseológicas especializadas passam a ser reconhecidas como unidades de informação de um vocabulário controlado especializado. No âmbito da Ciência da Informação, elas serão identificadas como ULRIEs, e enquanto ULRIEs teremos o registro de uma informação contextualizada, precisa, específica e representativa da linguagem utilizada pelos especialistas. Em conformidade com a perspectiva textual, temos o reconhecimento do texto como unidade de análise e como a unidade que contém termos, e da variação terminológica conceitual e denominativa. Ainda sob a ótica textual, reconhecemos a presença de uma densidade informativa diferenciada em textos especializados. Tal entendimento é importante quando se pensa em construir vocabulários controlados especializados que tenham na construção do sentido, um dos seus referenciais.

Por aceitarmos como válido o entendimento de que bibliotecários e usuários são agentes do processo que conduz à produção do conhecimento, a fundamentação na Linguística da Enunciação se justifica. Sob o referencial enunciativo, consideramos relevantes

na construção de vocabulários controlados especializados, os seguintes conceitos: linguagem, língua, estrutura, subjetividade, intersubjetividade, locutor, referência e contexto.

Com as noções de subjetividade e intersubjetividade fica estabelecido uma relação interpessoal entre bibliotecário e usuário. Ao inserir o bibliotecário na relação que se verifica entre usuários e Sistemas, temos como propósito alcançar melhores resultados nos atos de enunciação que ocorrem nos processos de indexação e de busca da informação e recuperação de documentos. Salientamos ser fundamental a inserção do bibliotecário nessa relação, uma vez que ele interfere significativamente nos atos de enunciações que se efetuam nesses processos.

Ao incluirmos o processo de indexação como constitutivo dessa relação – SRIs e usuários – ressaltamos a importância de qualificar esse processo, e com esse procedimento, obter melhores resultados no processo de busca da informação e recuperação de documentos. Ante esse referencial, pretendemos oferecer informações e não apenas dados ao Sistema e ao usuário. Entendemos que uma ferramenta construída ao abrigo desses princípios, que associa os referenciais enunciativos e terminológicos, tende a oferecer maior satisfação aos usuários, uma vez que acreditamos que os procedimentos de busca da informação e recuperação de documentos tenderão a se efetivar com maior rapidez e precisão.

Reconhecemos, no entanto, que por tratar-se de um processo de comunicação intersubjetiva, que resulta da interação de interlocutores que se caracterizam por não serem de mesma natureza, o resultado nunca será totalmente satisfatório. Mas acreditamos que sob a associação destes referenciais – enunciativos e terminológicos – será possível oferecer novos subsídios para tornar essa comunicação intersubjetiva mais eficaz, viabilizando uma interação mais efetiva entre seus interlocutores.

Sob o referencial da Linguística de *Corpus*, a visão de que a língua é entendida em seu contexto de uso é substancialmente relevante na construção dessa ferramenta. Como vimos anteriormente, a Linguística de *Corpus* tem como objeto de estudo a língua em seu contexto de uso. 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.

Enquanto metodologia, encontramos na Linguística de *Corpus* um recurso que oferece mais agilidade para a geração de vocabulários controlados especializados, a partir da terminologia que veicula em textos especializados, visando registrar, armazenar, buscar informações e recuperar documentos em SRIs.

Recorremos aos princípios da Linguística de *Corpus* para a coleta e análise das unidades lexicais de representação da informação especializada em seu contexto de uso. Essa

coleta é feita a partir de um *corpus* formado por textos de caráter acadêmico de autoria dos próprios especialistas. No caso da construção de vocabulários controlados especializados, mas ainda assim gerais, como um Vocabulário Controlado Especializado de ULRIEs em Filosofia, o texto selecionado é o *abstract*.

Entretanto, no caso de vocabulários controlados especializados, como um Vocabulário Controlado Especializado de ULRIEs em Filosofia antiga (Apêndice I) ou Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo (Apêndice J), a coleta das unidades lexicais do discurso especializado (ULDEs) em *abstracts* não é suficiente, pois apesar de se caracterizarem como textos de altíssimo grau de densidade informativa, estes textos não oferecem a representatividade temática desejada. Nesse caso, há também a necessidade de formação de um *corpus* a partir de textos completos.

Essa constatação tem origem na construção de vocabulários controlados especializados nas subáreas da Filosofia, mas precisa ser comprovada em outras áreas das Ciências Sociais e Humanidades. Fica aqui um questionamento. A inadequação de um *corpus* formado apenas por *abstracts* está relacionada com a área da Filosofia ou com o tipo de ferramenta almejada, vocabulários controlados especializados? Ressaltamos que em experiência anterior, na construção de um vocabulário controlado especializado na área de Estudos da Linguagem, na forma de um projeto piloto, a coleta das ULDEs em *abstracts* satisfaz plenamente.

A proposta ora apresentada está sujeita a discussões, avaliações e reformulações sempre que surgir um novo elemento que possa oferecer melhores resultados. Em seguimento a esta reflexão, apresentaremos nossa proposta sobre a geração de vocabulários controlados especializados, agregando tanto o conhecimento advindo dos estudos enunciativos como dos estudos terminológicos.

3.6 Vocabulários controlados especializados: uma proposta para a sua elaboração a partir de ULRIEs

Nesta seção apresentamos a proposta construída no âmbito desta pesquisa para a geração de vocabulários controlados especializados, que sugere uma base teórica e um formato que julgamos ser inédito. Esta proposta integra o referencial advindo dos Estudos de Terminologia e Linguística de *Corpus*; associado a este conjunto de conceitos e de noções está aquele que tem origem na Linguística da Enunciação e na própria Ciência da Informação.

Quanto ao formato, ressaltamos que esta proposta não considera as relações hierárquicas como relevantes, uma vez que elas estão expressas em outras Linguagens Documentárias, como a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU), por exemplo. Por esta e outras particularidades, explicitadas a seguir, podemos dizer que estamos propondo uma ferramenta diferenciada.

A título de um projeto piloto, a área selecionada para a geração desta ferramenta no contexto das Linguagens Documentárias é a Filosofia. Tal opção deve-se aos seguintes fatores: 1) ter familiaridade com a temática da área, o que é fundamental quando se pensa em construir um vocabulário controlado especializado; 2) não terem sido identificados, em revisão de literatura, vocabulários controlados – listas de cabeçalhos de assuntos e tesouros – que contemplassem as especificidades temáticas desta área; 3) poder dialogar com os especialistas da área selecionada para a construção do vocabulário proposto; 4) ser possível testar a proposta desenvolvida em um SRI disponibilizado via *web*, o SABi/UFRGS.

Quanto à ausência de vocabulários controlados, vale ainda ressaltar que os vocabulários que existem sobre esta temática – Filosofia – estão inseridos em grandes áreas do conhecimento, como as Ciências Sociais e Humanidades, por exemplo. O resultado desta realidade é que a terminologia que veicula nesses vocabulários é muito genérica e, desse modo, oferecem ao bibliotecário uma terminologia pouco representativa da área, o que inviabiliza uma indexação temática que contemple os temas específicos tratados em um documento. A ausência de vocabulários controlados especializados em um domínio, se reflete na terminologia usada em catálogos eletrônicos.

Por integrar o Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho⁶⁸ (NPF), temos a possibilidade de testar o que estamos sugerindo em um Sistema de Automação de Bibliotecas, o SABi/UFRGS⁶⁹. Há a oportunidade, também, de dialogar com

⁶⁸ “O Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho –, criado em 2005, tem por objetivo promover a pesquisa em filosofia, congregando diferentes equipes de trabalho ligadas ao Departamento e ao Pós-Graduação em Filosofia da UFRGS, envolvendo tanto docentes e discentes desta universidade como pesquisadores de outras instituições que atuam como colaboradores. Suas atividades envolvem seminários regulares de pesquisa em nível de graduação e de pós-graduação, apoio a organização de eventos, registro de dados relevantes sobre as atividades de pesquisa desenvolvidas junto ao Departamento e Pós-graduação em Filosofia da UFRGS e organização de acervo bibliográfico e documental. Após a morte do professor Balthazar Barbosa Filho em 2007, seu acervo bibliográfico foi, por generosa decisão de seus herdeiros, doado ao Núcleo, que passou a intitular-se Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho. Somado ao material adquirido graças a recursos de outros projetos de diferentes pesquisadores, trata-se de um acervo especializado de cerca de 8.000 volumes nas mais diferentes áreas da filosofia.” Disponível em: < <http://www.ufrgs.br/ifch/index.php/pesquisa/14/npdf>>. Acesso em: 03/02/2013.

⁶⁹ O SABi/UFRGS, Sistema de Automação de Bibliotecas, foi implantado em 1989 e adota o *software* Aleph 500 para gerenciar as atividades e serviços oferecidos pelas 35 bibliotecas da UFRGS e sua comunidade usuária.

os especialistas da área – pesquisadores e professores do Núcleo –, que são usuários do Sistema, assim como com colegas bibliotecários.

Para os propósitos desta pesquisa propomos que essa ferramenta seja reconhecida por Vocabulário Controlado Especializado de ULRIEs. Relembramos que uma ULRIE é representativa das UTs e das UFEs.

Dentre os pressupostos teóricos norteadores do Vocabulário Controlado Especializado de ULRIEs que estamos propondo, consideramos importante a especificidade temática⁷⁰, a precisão conceitual e a contextualização da informação indexada. Ainda que com a Informática o acesso à informação ganhe em operacionalidade e velocidade, é sabido que há uma perda significativa quanto à precisão da informação recuperada.

Quanto à especificidade almejada na geração de um vocabulário controlado especializado, acreditamos que ela possivelmente só será alcançada na construção de vocabulários controlados de áreas específicas do conhecimento, como um Vocabulário Controlado Especializado de ULRIEs em Filosofia antiga, ou Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo, por exemplo.

O Vocabulário Controlado Especializado de ULRIEs que estamos sugerindo está fundamentado nos seguintes princípios:

1) caracterizar-se como uma ferramenta que se sustenta em parâmetros enunciativos e terminológicos. Sendo assim está alicerçada em conceitos e expressões representativas de um domínio e não em palavras, sentido e não significado, e linguagem especializada e não língua comum;

2) ser capaz de estabelecer uma relação mais satisfatória entre os interlocutores dos processos de indexação e de busca da informação e recuperação do documento pelas seguintes razões:

- a) por reconhecer como interlocutores bibliotecários, usuários e SRIs;
- b) por considerar os aspectos linguísticos, comunicativos e cognitivos da língua;
- c) por entender o descritor sob uma forma mais ampla, ou seja, a partir de conceitos e expressões representativas de um domínio;
- d) por viabilizar o uso de uma unidade de informação de menor variação denominativa e conceitual;

⁷⁰ A especificidade postula que um tópico de assunto do documento deve ser representado pelo termo mais específico que o representa, em oposição à opção por termos genéricos. O aumento da especificidade na indexação costuma aumentar a precisão.

3) servir como ferramenta para o bibliotecário na representação temática de um documento – processo de indexação – e para o usuário na formulação de expressões de busca – processo de busca da informação e recuperação de documentos –;

4) considerar o descritor como uma ULRIE, tendo como propósito expressar conceitos e expressões representativas de um domínio, de modo a propiciar a circulação de informações de uma área de conhecimento;

5) conceber o texto como unidade de análise, de modo a viabilizar a contextualização das ULRIEs;

6) reconhecer a variação tipológica e a presença de uma densidade informativa diferenciada nos textos especializados;

7) coletar as ULDEs a partir de um *corpus* formado por textos especializados de uma área temática;

8) selecionar as ULRIEs a partir da frequência no *corpus* formado por textos especializados de um domínio. Embora a frequência possa se constituir como fator de relevância para a tomada de decisão na identificação das ULRIEs, ela não é fator determinante para seu reconhecimento;

9) identificar as relações conceituais entre as ULRIEs como relações associativas, incluindo sob um mesmo rótulo relações de equivalência, relações hierárquicas e relações interdisciplinares;

10) contextualizar as ULRIEs de um item bibliográfico com o uso de modificadores, de qualificadores, de autores como assunto e, em alguns casos, de variantes e de paráfrases. A contextualização visa construir itens bibliográficos com valor de informação, tal qual ocorre com os dados de catalogação.

Para a construção destas ferramentas serão observadas como procedimentos metodológicos, as etapas seguintes:

1) seleção da área de especialidade (projeto piloto: Filosofia);

2) constituição do *corpus* (*abstracts* para vocabulários controlados especializados em uma grande área de conhecimento, como Filosofia, e *abstracts* e textos completos para vocabulários controlados especializados em domínios específicos do conhecimento, como Filosofia aristotélica e Filosofia platônica, por exemplo);

3) seleção de uma ferramenta computacional de análise de textos (*AntConc*);

4) coleta e análise de dados (*abstracts* e textos completos de teses e dissertações, assim como artigos de periódicos do Portal de Periódicos da Capes);

- 5) definição da arquitetura do vocabulário controlado especializado para o gerenciamento das ULDEs e das ULRIEs;
- 6) seleção do *software* para o registro dos dados e geração desta ferramenta (Excel ou Access);
- 7) categorização das ULDEs em ULRIEs;
- 8) validação da proposta apresentada (SABi/UFRGS).

Apresentaremos, no apêndice I, uma pequena amostra do Vocabulário Controlado Especializado de ULRIEs em Filosofia antiga, e no apêndice J, do Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo, oferecendo uma lista das unidades coletadas e selecionadas para uso. Elas são apresentadas em ordem alfabética, seguidas das informações: ULRIE (unidade lexical de representação da informação especializada), UP (usado para). RA (relação associativa). FR (filósofo relacionado). Em uma única lista aparecem as ULDEs e as ULRIEs, sendo que as ULRIEs são grifadas em destaque.

Com a apresentação desses apêndices, temos o propósito de oferecer uma ideia de como uma determinada subárea do conhecimento é tratada quanto à sua representatividade temática, segundo o referencial teórico sugerido na pesquisa. Essas unidades foram coletadas em textos especializados da área, alguns disponibilizados na *web* e outros escaneados. Elas são representativas das subáreas de conhecimento Filosofia antiga, e Filosofia platônica e Platonismo, propiciando uma terminologia que contempla esta temática, tanto sob o aspecto específico quanto genérico.

Entendemos por unidades lexicais do discurso especializado (ULDEs) aquelas unidades que representam um conteúdo especializado, mas que não foram consideradas como sendo as mais relevantes e precisas para expressarem conceitos e noções de um domínio, como, por exemplo, ‘Alma’, ‘Alma do mundo’ e ‘Alma humana’, que são ULDEs que remetem para a ULRIE ‘Psyché (teoria da alma)’. Dentre as ULDEs estão as formas variantes, como ‘Academia de Atenas’ e ‘Academia de Platão’ em relação à ULRIE ‘Academia platônica’.

Consideramos unidades lexicais de representação da informação especializada (ULRIEs) aquelas unidades que representam um conteúdo especializado, e que foram recomendadas para uso em um Vocabulário Controlado Especializado de ULRIEs. Como ULRIEs, citamos ‘Academia platônica’, ‘Mito da caverna’ e ‘Psyché (teoria da alma)’.

Em consulta aos catálogos eletrônicos – Library of Congress, Bibliothèque Nationale de France, Acervus do Sistema de Bibliotecas da UNICAMP – não identificamos descritores específicos para a representação da temática filosófica. A representação temática das obras

consultadas foi feita sempre por descritores genéricos, conforme podemos observar na sequência desta exposição.

Em pesquisa feita na Library of Congress (LC) verificamos que a obra *Protágoras*⁷¹, um dos diálogos de Platão, não apresenta nenhum assunto para a representação temática do documento indexado. Em consulta a este mesmo catálogo, identificamos este diálogo de Platão na obra – *Protagora, Menone, Fedone* –, que consiste três dos seus diálogos. Para esta obra o assunto utilizado como termo de indexação pela LC é ‘Filosofia grega antiga’. Na Bibliothèque Nationale de France esta obra é verbalizada com o descritor ‘Platon (0427?-0348? av. J.-C.). Protagoras’. Na base Acervus do Sistema de Bibliotecas da UNICAMP esta mesma obra é representada tematicamente sob dois descritores: ‘Protágoras’ e ‘Platonismo’.

Somente no catálogo SABI UFRGS, a obra *Protágoras* processada pela BSCSH/NPF⁷², foi indexada utilizando descritores genéricos e específicos, tais como: ‘Filosofia antiga’, ‘Filosofia grega’, ‘Filosofia platônica’, ‘Diálogo platônico’, ‘Diálogo socrático’, ‘Dialética platônica’, ‘Ética platônica’, ‘Mito platônico’, ‘Mito de Prometeu’, ‘Mito de Epitemeu’, ‘Sofistas’, ‘Teoria das virtudes’ e ‘Teoria do conceito’.

Em uma breve análise dos descritores utilizados pelos diferentes catálogos eletrônicos, para a obra *Protágoras* de Platão, podemos observar o seguinte:

- 1) falta de consistência dos catálogos, uma vez que se constata que nenhuma das verbalizações é utilizada por mais de um catálogo eletrônico;
- 2) falta de especificidade temática na representação do documento indexado; e
- 3) a inexistência de descritores representativos do texto indexado.

Essas três observações, evidenciam duas situações:

- 1) ausência de ferramentas que sustentam as atividades de indexação; e
- 2) ser um processo extremamente subjetivo, o que se constata pela diversidade de descritores utilizados.

Em razão da diversidade de descritores usados para representar uma mesma obra, o usuário enfrenta grande dificuldade na construção de uma expressão de busca na consulta aos catálogos eletrônicos. Acresce, ainda, o fato de que, se o usuário utilizar uma expressão de busca específica, como por exemplo, ‘mito de Prometeu’, ele dificilmente recuperará a obra que contém tal mito.

⁷¹ Trata-se de um diálogo platônico, também reconhecido como diálogo socrático, que trata sobre a virtude e, como tal, da ética. Um dos pontos altos do diálogo é a evocação do mito de Prometeu e do mito de Epitemeu.

⁷² BSCSH/NPF – Biblioteca Setorial de Ciências Sociais e Humanidades / Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho

A indexação diferenciada feita pelo Núcleo é o resultado de um trabalho de pesquisa sobre a terminologia empregada pelos teóricos e especialistas da área filosófica. Finalizamos este capítulo, afirmando que esta pesquisa está viabilizando a geração de Vocabulário Controlado Especializado de ULRIEs em Filosofia antiga e em Filosofia platônica e Platonismo, para uso pelo bibliotecário e pelo usuário, e **formulando a seguinte proposta para a construção de Vocabulários Controlados Especializados:**

Os vocabulários controlados especializados passam a ser reconhecidos como Vocabulários Especializados de ULRIEs. As unidades de informação destes vocabulários passam a ser nomeadas de unidades lexicais de representação da informação especializada (ULRIEs). As ULRIEs passam a ser entendidas como unidades semânticas e ser coletadas e selecionadas a partir de um *corpus* formado por textos especializados de um domínio. O *corpus* será formado por *abstracts* para vocabulários que representam grandes áreas do conhecimento e por *abstracts* e textos completos para vocabulários que representam áreas específicas do conhecimento.

4 PROPOSTA METODOLÓGICA PARA A CONSTRUÇÃO DE VOCABULÁRIOS CONTROLADOS ESPECIALIZADOS

A metodologia em que se sustenta a geração de vocabulários controlados especializados tem como referencial teórico os Estudos da Linguagem e a Ciência da Informação. Com o propósito de dar representatividade a esta ferramenta – Vocabulário Controlado Especializado de ULRIEs – recorreremos à Linguística de *Corpus* enquanto referencial teórico e metodológico. A opção feita por vocabulários controlados especializados, na forma de um Vocabulário Controlado Especializado de ULRIEs, em detrimento de tesouros, justifica-se pelas seguintes razões:

1) pela intenção de propiciar uma ferramenta representativa da linguagem especializada de uma área de conhecimento, contemplando não apenas os conceitos de um domínio, mas também as expressões da área reconhecidas como relevantes;

2) pelo desejo de propor uma ferramenta que se fundamente em relações conceituais associativas, e não em relações hierárquicas como os tesouros;

3) pela vontade de oferecer uma metodologia ágil, eficaz e de menor complexidade na sua construção.

Relações associativas, nesta pesquisa, são entendidas em sentido amplo, ou seja, abarcando todo o tipo de relação existente entre as diferentes ULRIEs. Sobre a hierarquia de conceitos, ressaltamos que de modo algum pretendemos negar a sua importância, mas não há como ignorar a sua complexidade. A esse respeito compartilhamos com Gil (1999, p. 21) quando afirma que “[...] muitos autores vêm procurando definir um sistema de classificação das inúmeras ciências. Nenhum desses sistemas se mostra absolutamente satisfatório”. Relembramos, ainda, que as relações hierárquicas já estão estabelecidas em outras ferramentas usadas no processamento temático de um documento, como nas tabelas de classificação, Dewey e CDU, por exemplo.

Tal posicionamento justifica-se, também, pelo fato de nem mesmo os especialistas serem unânimes em relação à estrutura hierárquica do conhecimento. Isso decorre do fato de que se um mesmo conceito pode ser visto sob diferentes pontos de vista, a estrutura hierárquica do conhecimento será sempre relativa. Desse modo, um mesmo conceito pode ser considerado sob diferentes formas hierárquicas, até mesmo quando se referem a uma mesma

área do saber. Em Filosofia, por exemplo, o conceito *belo*⁷³ pode estar inserido tanto na Ética, como na Estética.

Na área da Filosofia, assim como no domínio das Ciências Sociais e Humanidades, o problema da hierarquia é muito complexo. Essa complexidade é atestada pela falta de ferramentas que ofereçam termos específicos para a representação temática de um domínio. As tabelas de classificação como a CDU e a CDD, por exemplo, ainda hoje são os instrumentos que melhor representam a estrutura hierárquica do conhecimento. E ainda assim, nem mesmo tais tabelas se mostram consistentes, pois a cada nova edição apresentam mudanças estruturais significativas em relação à forma como o conhecimento se organiza.

Fazendo uma análise do assunto Antropologia cultural em três edições das tabelas da Classificação Decimal de Dewey, temos a seguinte situação:

- 1) 15. ed. – 500 Ciências Puras
 - 570 Ciências biológicas
 - 572 Antropologia
 - 572.7 Antropologia cultural e social
- 2) 16. ed. – 500 Ciências Puras
 - 570 Ciências biológicas e antropológicas
 - 572 Antropologia cultural
- 3) 20. ed. – 300 Ciências Sociais
 - 306 Antropologia cultural

Observamos, aqui, uma mudança de concepção em relação à visão que se tem sobre o homem. Enquanto era visto como um ser vivo, ele estava inserido nas Ciências Puras, mas quando passou a ser entendido como um ser social, passou a ser situado nas Ciências Sociais. Esse caso, aqui destacado, demonstra que a evolução da ciência gera mudanças radicais no modo como o conhecimento se estrutura.

A escolha por ULRIEs e não por descritores – unidades de informação de um tesouro – justifica-se pela seguinte razão: a concepção de ULRIEs tem origem nos estudos e práticas da Terminologia. Elas representam as unidades terminológicas (UTs) e as unidades fraseológicas especializadas (UFEs), ou seja, conceitos e expressões representativas de um domínio. Elas diferem dos descritores que representam conceitos. Segundo Dodebei (2002, p. 79) os descritores devem estar expressos

⁷³ Na antiguidade, o *belo* era estudado pelos filósofos no âmbito da Ética, por ser considerado uma virtude. Mas este mesmo conceito pode ser visto também sob a visão da Estética, que é um ramo da filosofia que tem por objeto de estudo a natureza do belo.

[...] por termos, simbolicamente representados por palavras. Desta forma, os conceitos podem estar representados por apenas uma palavra, ou por uma composição de palavras, desde que esta composição designe um único conceito. Por exemplo, 'Ciência da Informação' representa apenas o conceito de um objeto, formado por duas palavras, portanto, é igual a um termo; 'Estudo da Linguagem' representa o conceito de dois objetos (estudo e línguas), devendo constituir dois termos.

Nas seções seguintes serão descritas as etapas metodológicas propostas para a construção de um vocabulário controlado especializado, que considera as ULRIEs como a sua unidade: 4.1) Seleção dos *corpora*, 4.2) Constituição dos *corpora*, 4.3) Seleção das ferramentas, 4.4) Coleta e análise de dados, 4.5) Arquitetura da base de dados, 4.6) Inserção dos dados na base e organização dos relacionamentos, e 4.7) Composição da lista multilíngue das ULRIEs. Iniciaremos este capítulo esboçando nossas ideias sobre a seleção dos *corpora*.

4.1 Seleção dos *corpora*

Os critérios estabelecidos para a seleção dos *corpora* que formarão os vocabulários controlados especializados a título de projetos pilotos, dizem respeito à: 1) temática, 2) tipo de material e 3) idioma. Não houve seleção por datas.

A temática selecionada para a construção destes vocabulários controlados especializados foi a Filosofia, sendo as subáreas Filosofia antiga e Filosofia platônica e Platonismo as eleitas para a aplicação da metodologia que está sendo proposta. As razões pela seleção desta área já foram apresentadas no item 3.6 Vocabulários controlados especializados: uma proposta para a sua elaboração a partir de ULRIEs.

O tipo de material escolhido para a construção dos *corpora* foi: 1) artigos de periódico, 2) teses e dissertações, 3) manuais e glossários específicos da temática filosofia antiga, e 4) textos didáticos.

De acordo com a visão teórico-metodológica adotada, que privilegia a autenticidade da comunicação, foram selecionados para a constituição dos *corpora* textos completos e resumos acadêmicos de teses, dissertações e artigos de periódicos, assim como de fragmentos de textos de manuais e de glossários. As principais razões que motivaram a escolha do resumo acadêmico foram a sua densidade conceitual e o uso de palavras-chave, pois 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.

A importância da inclusão de palavras-chave, nos *corpora* constituídos, se deve ao fato de elas serem as unidades léxicas selecionadas pelo autor como portadoras de maior carga temática no seu trabalho, e de estarem destacadas no final de cada resumo, em sequência à indicação *Palavras-chave:*, obedecendo às normas consensuais de formatação de um texto acadêmico.

Os idiomas eleitos para a constituição dos *corpora* foram: português, francês, inglês e espanhol. Na seção seguinte faremos uma exposição sobre a constituição dos *corpora*.

4.2 Constituição dos *corpora*

Os *corpora* formados para a construção dos vocabulários controlados especializados nos domínios da Filosofia antiga e da Filosofia platônica e Platonismo, ainda que não sejam exaustivos, são representativos destas subáreas de especialidade, o que permite exemplificar de modo concreto todas as etapas da metodologia aqui proposta. Eles são formados por textos completos e resumos de trabalhos disponibilizados no Portal de Periódicos da CAPES, e por textos didáticos extraídos da Internet. No caso do Vocabulário Controlado Especializado de ULRIEs no domínio da Filosofia platônica e Platonismo, o *corpus* foi constituído também pelo uso de textos escaneados produzidos por Reale (1995), Chauí (1994), Paviani (2001) e Zingano (2005), uma vez que são comentadores de Platão e construíram um glossário com os termos da obra do filósofo.

O Portal de Periódicos da Capes⁷⁴

[...] é uma biblioteca virtual que reúne e disponibiliza a instituições de ensino e pesquisa no Brasil o melhor da produção científica internacional. Ele conta atualmente com um acervo de mais de 31 mil periódicos com texto completo, 130 bases referenciais, dez bases dedicadas exclusivamente a patentes, além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual.

Para a construção destes *corpora* textuais foi considerado relevante a sua representatividade, assim como a facilidade de acesso. A representatividade aqui referida não está relacionada com a extensão, mas sim com a população envolvida – comunidade usuária – e com as áreas temáticas – Filosofia antiga e Filosofia platônica e Platonismo –.

A consulta ao Portal de Periódicos da Capes foi feita sob três modalidades: 1) por meio de buscas por assunto, 2) por base de dados e 3) por banco de teses. O *site* que viabiliza a consulta ao Portal é <<http://www.periodicos.capes.gov.br/index.php>>. A maioria dos textos

⁷⁴ Disponibilizado em: <http://www.periodicos.capes.gov.br/index.php?option=com_pinstitutional&mn=69>. Acesso em: junho de 2012.

disponibilizados no Portal tem acesso livre, entretanto, o acesso a alguns deles só foi possível a partir do cadastro no sistema responsável pela publicação.

A busca por assunto foi feita utilizando a expressão de busca ‘filosofia antiga’. Os resultados foram ordenados por relevância. Dentre os resultados oferecidos (245) foram selecionados os realmente relevantes para a constituição dos *corpora*. Para esta seleção foi levado em consideração os propósitos do texto, observando para quem ele é dirigido, se para filósofos ou para usuários e pesquisadores de outras áreas de interesse. A seleção foi feita também a partir da leitura do *abstract*.

Para a composição dos *corpora* os textos selecionados foram copiados no Word, pelo processo copia e cola. Para copiar os arquivos em PDF, foi utilizado o programa *ABBYY FineReader 10*, um conversor de arquivos PDF para outro formato.

Com o propósito de utilizar a ferramenta *ExATOlP*, cujo propósito é extrair automaticamente os sintagmas nominais de um *corpus*, e gerar de modo automático ontologias em língua portuguesa, foi constituído dois *corpora*, um em língua portuguesa e outro em outros idiomas, uma vez que o *ExATOlP* só processa textos em língua portuguesa.

Na busca por base de dados, foi utilizado o termo ‘filosofia’ seguido pela opção ‘busca por área de conhecimento’. Foi selecionada a área do conhecimento ‘Ciências Humanas’ e a subárea de conhecimento ‘Filosofia’. O resultado de busca ofereceu seis bases para consulta. Das seis bases oferecidas, foi escolhida a base *Scielo – Scientific Electronic Library Online* – para a pesquisa, por ser uma biblioteca eletrônica que abrange uma coleção selecionada de periódicos científicos brasileiros, dentre outros.

A opção de busca na base *Scielo* foi a de busca por assuntos. O assunto selecionado foi Ciências Humanas e o resultado de busca foi de 270 periódicos. Destes, foram selecionados os da área da Filosofia. O termo de busca para a pesquisa foi ‘filosofia’, uma vez que a expressão ‘filosofia antiga’ não ofereceu resultados satisfatórios para a constituição dos *corpora*. Para cada título de periódico foi feita uma nova seleção de assunto usando o termo ‘filosofia’, com o propósito de identificar os artigos que efetivamente atendiam às nossas necessidades de pesquisa.

Uma terceira busca para a constituição dos *corpora* foi feita pela opção Banco de Teses. A alternativa de busca foi a de assunto, sendo usada a opção ‘expressão exata’ ‘filosofia antiga’. O resultado desta busca foi de 78 teses e dissertações. Dentre essas, foram selecionadas aquelas que realmente interessavam para a constituição dos *corpora*. Algumas delas foram desprezadas pela impossibilidade de converter o texto no formato PDF, para o formato TXT.

Além do exposto até aqui, foram escaneados alguns textos extraídos de livros, dentre eles, alguns glossários. Para a leitura dos textos escaneados, foi utilizado o programa *ABBYY FineReader 10*, que permite a conversão dos arquivos para o formato DOC e TXT.

Os arquivos que constituem os *corpora* foram salvos no formato DOC ou PDF. Foram transformados em arquivos TXT para que pudessem ser analisados pelo programa de análise de textos selecionado para esta pesquisa, o *AntConc*, versão 3.2.4. Para a identificação dos *corpora* foram criadas tabelas contendo as seguintes informações: 1) identificação do arquivo, 2) tipo de texto, 3) ano de publicação, 4) idioma, 5) número total de palavras (*tokens*), 6) número total de tipos diferentes de palavras (*types*). O número total de palavras e o número total de tipos diferentes de palavras foi obtido pelo uso do *software AntConc 3.2.4*, selecionando a opção *word list* e clicando em *start*. O APÊNDICE C – *Corpus* sobre Filosofia platônica e Platonismo apresenta a identificação dos textos que constituem o *corpus* constituído sobre esta temática.

Quanto ao tamanho e de acordo com Berber Sardinha (2004), conforme pesquisa feita em 10/11/2012, no *AntConc 3.2.4*, na opção *word list*, o *corpus* sobre Filosofia platônica e Platonismo⁷⁵ é classificado como médio-grande, uma vez que conta com 4.616.046 *tokens* (número total de palavras) e 162.817 *types* (tipos de palavras diferentes).

Sobre as características dos *corpora*, é possível dizer que se trata de *corpora* escritos, sincrônicos e contemporâneos. Por ser uma amostragem, são *corpora* estáticos, fechados, de conteúdo especializado, específicos e relevantes quanto à densidade conceitual. Quanto ao idioma, caracterizam-se por ser multilíngue: português, espanhol, francês e inglês. As informações relacionadas à seleção das ferramentas serão apresentadas na próxima seção, que será desenvolvida a partir de duas subseções. São elas: 4.3.1 Ferramentas de análise de textos e 4.3.2 Instrumentos de gerenciamento de dados.

4.3 Seleção das ferramentas

A opção pelas ferramentas computacionais tem origem nas vantagens que os computadores oferecem na investigação da linguagem, tais como: 1) assegurar maior consistência e produtividade na análise dos dados, 2) serem mais eficientes e confiáveis na execução das tarefas, e 3) permitirem uma maior abrangência na quantidade de dados

⁷⁵ O *corpus* sobre Filosofia antiga é mais amplo, pois incorpora o *corpus* sobre Filosofia platônica e Platonismo e é acrescido de outros textos.

coletados para análise. Na sequência serão apresentados os instrumentos usados na análise de textos e gerenciamento de dados.

4.3.1 Ferramenta de análise dos textos

A ferramenta utilizada para análise de textos foi o *software AntConc*, versão 3.2.4, um programa de análise de textos, de livre acesso, 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.

AntConc é um *software* gratuito desenvolvido por Laurence Anthony, professor da Faculty of Science and Engineering, Waseda University, no Japão. Ele é membro do Center for English Language Education in Science and Engineering (CELESE). O *AntConc* pode ler arquivos em diferentes formatos como: TXT, HTML, HTM e XML. Para ler esses arquivos é preciso ativar a opção correspondente nos Global Settings.

Este *software* não necessita de instalação; é um programa executável que pode ser utilizado tão logo tiver sido salvo. É um programa simples que oferece as ferramentas básicas de um programa de análise e de exploração de *corpora*, a saber:

1) *concordance*: gera linhas de concordâncias, ou seja, linhas de um ou mais textos-alvo selecionados pelo usuário (ver APÊNDICE D);

2) *concordance plot*: mostra as concordâncias em um modo gráfico;

3) *file view*: permite a visualização do arquivo original (ver APÊNDICE E);

4) *clusters*: gera uma lista de palavras mostrando as palavras que se agrupam em torno do termo pesquisado. Os *clusters* podem ser ordenados pela frequência, pela parte inicial ou final da palavra e por probabilidade. Esta ferramenta, ainda, permite gerar n-gramas, isto é, grupos de palavras de 1 a x;

5) *collocates*: gera uma lista ordenada de colocações apresentando as palavras que aparecem junto ao termo que se busca;

6) *word list*: gera uma lista de todas as palavras que estão presentes nos arquivos do *corpus*, ordenando-as pela frequência ou pela ordem alfabética;

7) *keyword list*: gera uma lista de palavras-chave.

Desse modo o *AntConc* possibilita a análise de um *corpus* a partir de listas de palavras organizadas por ordem alfabética, por frequência, por palavras-chave e por grupos de palavras, dentre outras possibilidades. Dos seus recursos, os mais utilizados nesta pesquisa foram o *concordance*, o *file view* e o *word list*.

Concordance 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 junto ao seu contexto, visualizadas em linhas.

File view viabiliza acesso ao texto em que o termo está inserido, oferecendo informações relevantes para a categorização das ULDEs em ULRIEs, e para o estabelecimento de relações entre essas unidades.

A consulta ao *word list* permite identificar o tamanho do *corpus*, uma vez que ele oferece informações sobre *tokens* e *types*, ou seja, sobre o número total de palavras e os tipos de palavras.

Salientamos que para a análise dos textos é importante que os arquivos estejam em formato TXT. Neste formato, o texto se apresenta somente com os caracteres do teclado, ou seja, com letras, números e símbolos ortográficos, e sem códigos de formatação específicos como ocorrem no formato DOC, formato padrão para documentos do Word.

4.3.2 Instrumentos de gerenciamento de dados

O *software* utilizado para a constituição da base de dados foi o Excel. Ele é de fácil acesso tanto em relação ao preenchimento dos dados quanto em relação à busca das informações ali armazenadas. Ressaltamos, no entanto, que o Access também satisfaz às necessidades que se impõem para o gerenciamento dos dados, conforme já ficou demonstrado, em experiências anteriores.

O tema da coleta e análise dos dados será abordado na seção seguinte.

4.4 Coleta e análise de dados

Esta seção se desenvolve a partir das seguintes subseções: 4.4.1 Levantamento das unidades lexicais do discurso especializado (ULDEs); 4.4.2 Seleção da terminologia representativa de um domínio; 4.4.3 Produção de concordâncias; 4.4.4 Análise das concordâncias; 4.4.5 Representatividade das ULRIEs; e 4.4.6 Categorização das ULDEs em ULRIEs.

4.4.1 Levantamento das unidades lexicais do discurso especializado (ULDEs)

Em uma primeira etapa foram selecionadas as ULDEs que foram consideradas relevantes nos domínios da Filosofia antiga e da Filosofia platônica e Platonismo, e que poderiam vir a ser categorizadas como ULRIEs no vocabulário controlado especializado a ser construído. Inicialmente as ULDEs foram coletadas durante o processo de indexação das obras que constituem o acervo do Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho; elas foram extraídas do próprio documento que estava sendo indexado. Algumas, no entanto, tiveram origem em textos acadêmicos disponibilizados na *web*, decorrentes de pesquisas feitas no Google Acadêmico.

Em uma etapa posterior, as ULDEs foram submetidas à análise para a sua categorização em ULRIEs, RAs e FRs. Nesta etapa foi essencial a construção dos *corpora*, pois foi pelas informações neles contidas que esta etapa pode ser desenvolvida. Desse modo, a pesquisa nos *corpora* constituídos teve como ponto de partida as unidades já levantadas anteriormente. A consulta aos *corpora* permitiu a seleção de novas ULDEs e sua categorização em ULRIEs, RAs e FRs. Algumas das ULDEs foram descartadas e outras foram categorizadas como variantes. Em muitos casos, foi possível a identificação de ULRIEs em outros idiomas. É intenção desta pesquisa gerar, também, um apêndice aos vocabulários controlados especializados contendo uma lista das ULRIEs expressas nos seguintes idiomas: português, francês, inglês e espanhol, conforme demonstrado no APÊNDICE F.

Esses dados foram registrados no *software* Excel, sob os seguintes rótulos: ULDEs (unidades lexicais do discurso especializado), use (campo fixo), ULRIEs (unidades lexicais de representação da informação especializada), RAs (relações associativas) e FRs (filósofos relacionados), conforme indicado ao final deste capítulo no Quadro 4 – Arquitetura da base de dados terminológica. Para facilitar a análise dessas unidades, elas foram organizadas em ordem alfabética.

A seleção da terminologia dos domínios selecionados para esta pesquisa – Filosofia antiga e Filosofia platônica e Platonismo –, na forma de projetos pilotos, teve origem tanto nos textos especializados que fazem parte do acervo do Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho, como em textos acadêmicos, sejam eles provenientes de consultas feitas ao Google Acadêmico, dos próprios *corpora* constituídos a partir do Portal de Periódicos da Capes e de textos escaneados. Esse será o tema abordado a seguir.

4.4.2 Seleção da terminologia representativa de um domínio

O critério adotado para a seleção da terminologia, representativa de um domínio, teve como princípio a possibilidade de identificar candidatos a termo. Nesta pesquisa, entendemos 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 e/ou 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.

Para a produção de concordâncias, tema da próxima subseção, foi importante o uso do recurso oferecido pelo *AntConc* 3.2.4, o *Concordance*.

4.4.3 Produção de concordâncias

Observando a unidade lexical no contexto foi possível identificar novas ULDEs e categorizá-las em ULRIEs, RAs e FRs, organizando, dessa forma, o conhecimento pelas relações associativas que estas unidades estabelecem entre si.

A opção pelo campo designado como relações associativas, 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 Ciências Sociais e Humanidades essa característica adquire relevância ainda maior, situação que se vê retratada na falta de instrumentos do tipo vocabulários controlados especializados disponibilizados ao bibliotecário.

Dando continuidade a essas considerações, abordaremos o tema da análise das concordâncias na próxima subseção.

4.4.4 Análise das concordâncias

A análise das concordâncias facilitou a identificação das ULDEs e sua configuração ou não como ULRIEs, sejam elas unidades terminológicas ou unidades fraseológicas especializadas. Viabilizou também a identificação das diferentes ocorrências de uso de uma ULRIE no *corpus* analisado, como por exemplo: metafísica *platônica*, metafísica *aristotélica*,

metafísica *ocidental*, *ética* e metafísica e metafísica *da alma*, dentre outras. Isso se tornou possível pelo acesso que se tem ao contexto em que as ULRIEs se apresentam.

Pela análise das concordâncias, várias relações foram evidenciadas entre as ULRIEs, inclusive aquelas que relacionam filósofo às escolas ou sistemas filosóficos, e filósofos e os conceitos que fundamentam as escolas ou sistemas filosóficos a que estão vinculados. Foram identificados, ainda, casos de variação e ocorrências de ULRIEs em outros idiomas, uma vez que os *corpora* são multilíngues. Como exemplo, citamos: estoicismo (língua portuguesa), stoicisme (língua francesa), stoicism (língua inglesa) e estocismo (língua espanhola).

A seleção das unidades terminológicas e das unidades fraseológicas especializadas, em outros idiomas, foi considerada importante por se constituírem em informação relevante 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 linguísticos, terminológicos e do assunto abordado. Para o usuário, a inclusão dessas unidades é pertinente, na medida em que amplia as condições de busca em uma pesquisa às bases de dados internacionais.

Sobre a coleta e a análise dos dados é relevante, ainda, abordar o tema da representatividade das ULRIEs.

4.4.5 Representatividade das ULRIEs

Pela identificação das combinações recorrentes nas concordâncias, percebemos que as unidades terminológicas não eram suficientemente representativas da terminologia de um domínio. Daí a necessidade de considerarmos que as ULRIEs sejam representadas tanto por unidades terminológicas, como por unidades fraseológicas especializadas.

4.4.6 Categorização das ULDEs em ULRIEs

A categorização das ULDEs em ULRIEs – unidades terminológicas (UTs) e unidades fraseológicas especializadas (UFEs) – foi fundamentada no seguinte entendimento: UT é toda a unidade lexical especializada, composta de uma ou mais palavras, que designa um conceito específico de um domínio temático; e UFE é a unidade lexical que representa conceitos e noções de um domínio, e que se caracteriza por ser uma combinação recorrente de palavras semanticamente vinculadas ao conteúdo da comunicação especializada; ela é reconhecida pela sua função, que é a de representar e transmitir conhecimento especializado.

Na categorização das ULRIEs observamos a necessidade de precisar o sentido de algumas unidades terminológicas. Nesses casos, uma ULRIE é constituída por uma unidade terminológica, acrescida de um modificador ou de um qualificador e, às vezes, de uma paráfrase. Na próxima seção desenvolveremos nossa concepção sobre arquitetura da base de dados.

4.5 Arquitetura da base de dados⁷⁶

Para armazenar e gerenciar os dados coletados para a geração dos vocabulários controlados especializados, tendo em vista o registro da terminologia considerada relevante em Filosofia antiga e Filosofia platônica e Platonismo, subáreas da área de especialidade Filosofia, foram criados cinco campos. Para visualizar a configuração da base de dados, ver o QUADRO 4, cujos campos são a seguir especificados.

QUADRO 4 – Arquitetura da base de dados

ULDEs	use	ULRIEs	RAs	FRs

Fonte: a autora

1) ULDEs (unidades lexicais do discurso especializado): são todas as unidades que representam conceitos e noções de um domínio e que podem vir a ser ULRIEs. Elas podem ser compostas de uma ou mais palavras que contém traços semânticos e/ou pragmáticos característicos da área temática, podendo ser a forma lexical plena ou reduzida como sigla, acrônimo, abreviatura e fórmula. Além de configurar-se como possível candidato ao estatuto de ULRIE, uma ULDE configura-se também como possível remissiva, ampliando significativamente o total de pontos de acesso de um vocabulário controlado especializado.

2) use – campo fixo.

3) ULRIEs (unidades lexicais de representação da informação especializada): são as unidades lexicais que representam tanto a unidade terminológica como a unidade fraseológica especializada, e que são recomendadas para uso por serem as mais representativas de um domínio.

⁷⁶ Entendemos por base de dados um conjunto de informações, organizada segundo regras definidas à qual se pode ter acesso para extrair, atualizar, acrescentar ou eliminar informações.

4) RAs (relações associativas): são ULRIEs que mantêm qualquer tipo de relação associativa – hierárquica ou não-hierárquica – com aquelas unidades lexicais registradas no campo ULRIEs.

5) FRs (filósofos relacionados): são ULRIEs que representam o filósofo que está relacionado com a unidade lexical registrada no campo ULRIEs. No APÊNDICE K – Apêndice ao Vocabulário Controlado Especializado de ULRIEs no domínio da Filosofia antiga – Relação de filósofos, apresentaremos uma lista de filósofos da filosofia antiga relacionado com as respectivas ULRIEs.

A inserção dos dados na base e a organização dos relacionamentos serão os aspectos abordados na seção seguinte.

4.6 Inserção dos dados na base e organização dos relacionamentos

A inserção dos dados na base e a organização dos relacionamentos foram feitas em duas etapas: durante o processo de indexação temática dos documentos do Núcleo de Pesquisa e no momento da análise dos *corpora*. Ressaltamos que ambas foram complementadas pela pesquisa no Google Acadêmico, com o propósito de obter maiores esclarecimentos na organização do conhecimento. Salientamos, ainda, que a inserção dos dados teve origem na extração manual e automática de termos.

Foi durante a análise dos *corpora* que ocorreu a categorização das ULDEs em ULRIEs e que as relações entre as ULRIEs foram consolidadas como válidas.

O quadro seguinte apresenta uma amostra de como os dados foram inseridos em seus respectivos campos.

QUADRO 5 – Registro dos dados no Excel

ULDEs ⁷⁷		ULRIEs ⁷⁸	RAs ⁷⁹	FRs ⁸⁰
Abstinência de alimento animal	usar	Vegetarianismo		
Ação de filósofo	usar	Filosofar		
Adiáfora	usar	Adiaforia		
Adiaforia (atitude moral de total indiferença)	usar	Adiaforia	Estoicismo. Ética estóica. Ética grega. Filosofia antiga. Filosofia grega. Helenismo.	Aristo de Quios. Hérilo de Cartago.
<i>Agathon</i>	usar	<i>Agathon</i> (o bem)	Ética aristotélica. Ética grega. Ética platônica. Eudaimonia. Filosofia antiga. Filosofia aristotélica. Filosofia grega. Filosofia platônica. Filosofia socrática. Ideia do bem. Teoria das virtudes.	Aristóteles. Platão. Plotino. Sócrates.
Água	usar	Teoria dos quatro elementos		
Akrasia lúcida	usar	Akrasia (intemperança)		
Indiferentes (adiáfora)	usar	Adiaforia		
Teoria dos quatro elementos (fogo, água, ar e terra)	usar	Teoria dos quatro elementos	Cosmologia grega. Escola pluralista. Filosofia antiga. Filosofia aristotélica. Filosofia da natureza. Filosofia grega. Filosofia platônica. Pré-socráticos.	Aristóteles. Empédocles de Agrigento. Platão.
Vegetarianismo	usar	Vegetarianismo	Ética animal. Filosofia antiga. Filosofia greco-romana. Filosofia grega. Veganismo.	Apolônio de Tiana. Diógenes Laércio. Empédocles de Agrigento. Ovídio. Pitágoras de Samos. Platão. Plotino. Plutarco. Porfírio. Sêneca. Sócrates.

Fonte: a autora

⁷⁷ ULDEs: unidades lexicais do discurso especializado

⁷⁸ ULRIEs: unidades lexicais de representação da informação especializada

⁷⁹ RAs: relações associativas

⁸⁰ FRs: filósofos relacionados

Dando seguimento a nossa exposição, faremos uma análise comparativa do processo de extração de candidatos a termos, sob a forma manual e a automática. Selecionamos para análise o *corpus* constituído para a formação do Vocabulário Controlado de ULRIEs em Filosofia platônica e Platonismo.

4.7 Análise comparativa do processo de extração de candidatos a termos: manual X automático

O processo de extração de candidatos a termos foi feito sob duas formas distintas: extração manual e extração automática. Ressaltamos o fato de que, nas duas situações de extração de candidatos a termos, foi utilizado o *corpus* constituído a partir de publicações do Portal de Periódicos da Capes e de textos escaneados.

De acordo com Berber Sardinha (2004) e conforme mencionado anteriormente – seção 4.2 Constituição do *corpus* –, este *corpus* é considerado de tamanho médio-grande. Entretanto, para os propósitos desta pesquisa, constatamos ser um *corpus* insuficiente, mas ainda assim, produtivo e representativo, conforme veremos nas situações explicitadas a seguir.

A extração manual foi realizada a partir do processo de indexação. As obras indexadas fazem parte do acervo do Núcleo de Pesquisa e Documentação em Filosofia Balthazar Barbosa Filho (NPF), e dizem respeito às obras escritas por Platão – 50 obras – e as escritas por seus comentadores – 96 obras –, totalizando 146 obras. Elas estão disponibilizadas para uso no catálogo SABi/UFRGS.

A extração automática teve origem na anotação do *corpus* pelo *parser* PALAVRAS e no uso da ferramenta ExATOl_p. Ela foi feita a partir da identificação de unigramas (1 token/palavra), bigramas (2 tokens/palavras), trigramas (3 tokens/palavras), quadrigramas (4 tokens/palavras) e n-gramas (n-tokens/palavras). Para fins desta pesquisa e no âmbito da extração automática de candidatos a termos consideramos apenas os unigramas, bigramas, trigramas e quadrigramas.

O processamento manual do *corpus* produzido possibilitou a identificação de 412 ULDEs e 198 ULRIEs. Pelo processo automático, este mesmo *corpus* permitiu a extração de 1.213 candidatos a termos efetivamente confirmados como ULDEs, e 413 unidades de informação aceitas como ULRIEs. Os dois processos identificaram um conjunto comum de 187 ULDEs e 101 ULRIEs. Retomando o que já foi exposto anteriormente, lembramos que as ULDEs servem de pontos de acesso alternativos de representação da terminologia de um

domínio, e correspondem, no nosso entendimento, a um conjunto de remissivas, e as ULRIEs são as unidades aceitas para uso no vocabulário controlado especializado.

O Quadro 6 mostra o percentual de crescimento de ULDEs e ULRIEs na geração de Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo, a partir da comparação dos dois processos de coleta e extração de termos, manual e automático, 194,41% e 108,58%, respectivamente.

QUADRO 6 – % de crescimento de ULDEs e ULRIEs no Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo

PROCESSOS	ULDEs	ULRIEs
Manual	412	198
Automático	1.213	413
% de crescimento	194,41	108,58

Fonte: a autora

Observando as ULDEs e as ULRIEs obtidas pelo processo de extração manual e pelo processo automático, constatamos um conjunto de candidatos a termos presentes no processo de extração manual, mas ausentes no processo de extração automática e vice-versa. Essa divergência tem origem em duas situações diferentes:

1) em procedimentos metodológicos do processo de extração automática, no caso desta pesquisa – do ExATOlp –, que trabalha com diferentes opções de ponto de corte, os quais levam em consideração a frequência relativa ou absoluta de cada termo. Exemplificando os problemas gerados pelas opções de pontos de corte, citamos a seguinte situação: presença de ‘Filosofia antiga’ como bigrama, mas ausência de ‘Filosofia’, como unigrama;

2) no fato da extração manual de termos ter origem no processo de indexação, e não de uma análise mais aprofundada do *corpus*.

O Quadro 7 registra o percentual de ocorrências comuns à coleta manual e à extração automática de termos.

QUADRO 7 – % de ocorrências comuns à coleta manual e à extração automática de ULDEs e ULRIEs no Vocabulário Controlado Especializado de ULRIEs em Filosofia platônica e Platonismo

PROCESSOS	ULDEs	ULRIEs
Manual	412	198
Automático	1.213	413
Ocorrências comuns	187	101
% de ocorrências comuns	45,39	51,01

Fonte: a autora

Conforme citado anteriormente, foram identificados 187 casos de ocorrências comuns de ULDEs e 101 casos de ocorrências comuns de ULRIEs, na comparação dos processos manual e de extração automática de termos, significando que 45,39% das ULDEs e 51,01% das ULRIEs extraídas manualmente se repetiram na extração automática.

Concluindo, dentre as vantagens e desvantagens identificadas para a produção de um vocabulário controlado especializado com o uso dos processos acima mencionados, citamos:

1) a extração automática de termos viabiliza uma maior representatividade de uma área de especialidade, tanto no que se refere à especificidade como à generalidade, por oferecer maior visibilidade de uma área de conhecimento para o pesquisador que se propõe a construir vocabulários controlados especializados. Em consequência, a extração automática de termos oferece um volume significativamente maior de pontos de acesso pelo uso do ExATOlp, tanto no que se refere à ULDEs, como no que diz respeito à ULRIEs;

2) a extração automática de termos apresenta como dificuldade o livre acesso a programas de extração automática de termos, no caso desta pesquisa, ao ExATOlp;

3) a extração manual de termos oferece como vantagem a possibilidade de recuperação de termos importantes que são eliminados pelos processos de extração automática de termos.

Assim, recomendamos o uso dos dois processos para a construção de um vocabulário controlado especializado. É interessante informar que o ExATOlp poderá vir a ser uma ferramenta de livre acesso. Encerrada essa reflexão, desenvolveremos na seção seguinte o modo como concebemos a composição da lista multilíngue das ULRIEs.

4.8 Composição da lista multilíngue das ULRIEs

Inicialmente estava previsto a composição da lista multilíngue das ULRIEs a partir dos *corpora* constituídos. Entretanto, apesar dos *corpora* serem multilíngues, não foi possível a execução desta tarefa apenas consultando os *corpora*. Desse modo, decidimos pelo uso de dois outros instrumentos de pesquisa, o Google Acadêmico e os dicionários especializados na área da Filosofia.

Tomamos esta decisão por considerarmos importante oferecer ao bibliotecário e ao usuário a possibilidade de acessar as ULRIEs de um domínio, não apenas na língua portuguesa, mas também em outros idiomas, viabilizando, assim, a construção de uma ferramenta multilíngue.

Sobre a impossibilidade dos *corpora* de servirem como única fonte de pesquisa para a execução desta tarefa, entendemos que seria necessário não apenas a inserção de novos *abstracts* em outras línguas, como também a inclusão de um número maior de textos completos em outras línguas, o que não foi feito por limitação de tempo para a execução desta pesquisa. Terminaremos este capítulo fazendo algumas considerações.

4.9 Considerações finais sobre a proposta metodológica para a construção de vocabulários controlados especializados

A metodologia proposta nesta tese está direcionada para a área da Filosofia, uma das áreas das Ciências Sociais e Humanidade – CSH –. Pela experiência em indexação de documentos nesta área, atuando em bibliotecas universitárias, acreditamos que a metodologia aqui apresentada possa ser estendida para outras áreas de conhecimento deste domínio – CSH –, tais como Antropologia, Estudos da Linguagem, História, Política e Sociologia, por exemplo. Esperamos, ainda, que outras áreas de conhecimento como as Ciências Biológicas e as Ciências Exatas também possam utilizar essa metodologia. Mas ressaltamos que ela precisa ser testada e possivelmente adaptada às características específicas de cada domínio.

Entretanto, há princípios metodológicos que são válidos para todas as áreas. O bibliotecário que se predispõe a desenvolver um vocabulário controlado especializado deve ter:

- 1) familiaridade com a área de conhecimento a ser indexada;
- 2) experiência nas atividades de indexação;

3) perfil adequado às atividades de pesquisa, ou seja, estar predisposto a enfrentar e vencer os desafios que se apresentam no desenvolvimento da ferramenta que se propôs construir, buscando criar novos conhecimentos, procedimentos e soluções para os diversos problemas com os quais se depara;

4) capacidade de síntese, de trabalhar em equipe e de se adaptar às mudanças;

5) condições de construir o vocabulário controlado especializado, mantendo equilíbrio entre a subjetividade e a objetividade;

6) um espírito crítico e empreendedor, pautado no raciocínio lógico e na autonomia intelectual;

7) fundamentação teórica no campo das Linguagens Documentárias;

Quanto ao conhecimento em Estudos da Linguagem destacamos a necessidade de ter:

1) conhecimentos mínimos em estudos enunciativos e estudos terminológicos;

2) experiência com a construção de *corpus* e com os recursos teóricos e metodológicos que a Linguística de *Corpus* oferece;

3) o entendimento de que a ideia de estrutura é fundamental para a construção de vocabulários controlados especializados;

4) domínio da língua de origem;

5) aptidão para leitura e compreensão de textos em no mínimo mais de dois idiomas estrangeiros.

Em relação aos conhecimentos na área da Informática, destacamos:

1) noções básicas de Informática, mais especialmente as relacionadas ao pacote Office;

2) prática com *softwares* de análise de textos, como o *AntConc*, por exemplo;

3) domínio de *softwares* que permitem a construção de bases de dados, como o Access e o Excel;

4) desenvoltura para pesquisar em bases de dados estrangeiras e nacionais.

Com essas considerações encerraremos este capítulo. Finalizaremos esta pesquisa, apresentando nossas conclusões. Teremos como propósito discutir os objetivos e as hipóteses propostas.

5 CONSIDERAÇÕES FINAIS E PERSPECTIVAS

Finalizamos esta pesquisa entendendo que foi possível alcançar os objetivos propostos, responder positivamente as questões de pesquisa apresentadas e confirmar as hipóteses levantadas.

Tendo como referencial teórico a Teoria da Enunciação de Benveniste, Estudos de Terminologia e Linguística de *Corpus* investigamos a eficácia dos Sistemas de Recuperação da Informação e dos vocabulários controlados especializados, com o propósito de construir um novo referencial no domínio das Ciências da Informação.

Nos conceitos e noções produzidos no âmbito dos Estudos da Linguagem encontramos sustentação para o desenvolvimento de um referencial enunciativo e terminológico na perspectiva das Linguagens Documentárias e dos Sistemas de Recuperação da Informação. Formulamos ainda duas propostas: uma com o fim de caracterizar a relação que se estabelece entre SRIs e usuários, e outra com a intenção de apresentar um referencial teórico e metodológico visando produzir um vocabulário controlado especializado.

Sob o olhar da Linguística da Enunciação e pela compreensão dos conceitos e noções advindos da leitura da obra de Benveniste, aprofundamos o conhecimento das Linguagens Documentárias quanto ao seu funcionamento. É substancialmente relevante para as Linguagens Documentárias a distinção que o linguista faz entre linguagem (língua em ação) e língua, semiótica e semântica, significado e sentido, forma e sentido, e signo e palavra.

A partir da oposição forma e sentido, Benveniste aborda o funcionamento da linguagem com o propósito de examinar a questão da significação. Para o linguista, a significação é constitutivo da língua e não algo acrescentado a ela. A língua é entendida por Benveniste como possuidora de uma função instrumental. No pensamento benvenistiano é impossível falar em língua como instrumento de comunicação, uma vez que a linguagem está na natureza do homem. Isso seria colocar o homem em oposição à sua própria natureza, seria dissociar do homem a propriedade da linguagem. A distinção entre significar e comunicar é pertinente para a compreensão das Linguagens Documentárias.

Sobre significado e sentido Benveniste tem o seguinte entendimento. O significado decorre do signo e da propriedade da linguagem de significar. O sentido, de outro modo, é inerente ao sistema linguístico e é percebido pelos falantes de uma língua. A noção de sentido, de modo diferente da de signo, está associada à da palavra e à da frase. Ao diferenciar signo e palavra, Benveniste caracteriza o signo como elemento linguístico e a palavra como a unidade

que expressa o sentido. Enquanto o signo é dotado de significação, a palavra é possuidora de sentido, guarda sua própria significação e suas referências.

É pela palavra e pela frase que nos comunicamos, e não pelo signo. **Desse modo, é na palavra e na frase que a Linguagem Documentária deve se fundamentar, e não no signo. Mas é no discurso que o sentido se constitui, sendo a referência uma noção fundamental no processo de construção do sentido** e, assim sendo, de considerável importância na geração de vocabulários controlados especializados e de um SRI.

O sentido é obtido pela ação que as palavras exercem umas sobre as outras, e se constitui no momento da enunciação. Sob a ótica desta teoria a enunciação é constituída pelo ato de enunciação, pelas situações em que ele se realiza e pelos instrumentos de sua realização. No modelo de enunciação proposto pelo linguista, um ato de enunciação se realiza sempre no tempo presente, no 'aqui' e no 'agora', e os interlocutores referem e co-referem na atribuição de sentido às palavras. **No âmbito das Linguagens Documentárias, uma ULRIE adquire sentido e valor de informação a partir da referência a que se refere e, desse modo, do contexto de situação.**

Ainda fundamentado nos estudos enunciativos desenvolvemos um referencial com o propósito de viabilizar uma interação mais efetiva entre os Sistemas de Recuperação da Informação e os usuários, um dos objetivos propostos. Com base nesse referencial formulamos um novo modelo de comunicação documentária, sustentado essencialmente pelas noções de subjetividade, intersubjetividade e categorias de pessoa, espaço e tempo. Este modelo integra os processos de indexação e os de busca da informação e recuperação de documentos. A aproximação proposta nesta pesquisa, entre esses dois processos, teve como propósito maximizar o uso da informação e viabilizar a produção do conhecimento. Temos convencimento de que esse objetivo também foi alcançado, pelas razões expostas a seguir.

A noção de subjetividade está presente no processo de indexação. Ela aparece na alta probabilidade de diferentes bibliotecários selecionarem diversos descritores para representar a mesma temática de um documento e, até mesmo, de um único bibliotecário selecionar diferentes descritores para representar tematicamente um mesmo documento, em momentos distintos.

Sobre a atividade de indexação acrescentamos que ela é afetada pelo conhecimento que o bibliotecário tem do tema que é expresso em um documento e, também, pelas tendências ideológicas do indexador e pelas suas experiências. Com a noção de subjetividade inserida no âmbito dos SRIs e das Linguagens Documentárias, e o entendimento que a relação

que se dá entre os interlocutores dos SRIs é de intersubjetividade, bibliotecário e usuário deixam de ser vistos como emissor e receptor, e passam a ser entendidos como interlocutores.

Para explicar a relação que se estabelece entre SRIs e usuários, esta pesquisa buscou o referencial que tem origem em Benveniste (1988a, 1989a), acrescido da reinterpretação de Martins (1990). A reinterpretação de Martins (1990) é essencial aos propósitos desta pesquisa, na medida em que entendemos que a relação que se dá entre os interlocutores de um SRI não é independente do conteúdo linguístico da enunciação.

Trazer para o âmbito das Linguagens Documentárias a noção de interação é importante, porque viabiliza a compreensão de um conceito de linguagem mais amplo, diferenciado. Desse modo, sob a ótica da enunciação, e a partir da noção de interação, **o processo de comunicação documentária passa a ser reconhecido como um processo de comunicação intersubjetiva, marcado fortemente pela subjetividade e intersubjetividade de seus interlocutores.**

Na medida em que passamos a compreender esta relação como inserida em um processo de comunicação intersubjetiva, marcada pela presença de interlocutores de natureza heterogênea – bibliotecários, usuários e SRIs –, aceitamos como válido recomendar a Teoria da Enunciação de Benveniste como referencial teórico apropriado para caracterizar essa relação.

Entendemos, também, que os Estudos de Terminologia oferecem preciosa contribuição para qualificar os SRIs. Com o referencial ali produzido tivemos como propósito selecionar terminologias representativas, precisas e específicas para representar documentos sob sua temática.

Das contribuições advindas dos Estudos de Terminologia destacamos a noção de comunicação especializada. Na comunicação especializada, a linguagem utilizada é a linguagem do especialista. Os interlocutores desse tipo de comunicação são especialistas, e o mundo de referência de sua comunicação diz respeito a um determinado campo do conhecimento.

A noção de comunicação especializada traz para o âmbito das Linguagens Documentárias o uso de uma terminologia padronizada, o que contribui para tornar a comunicação menos ruidosa entre especialistas. Por se caracterizar pela objetividade, pelo uso de termos técnico-científicos e pelo esforço em alcançar a concisão, a precisão e a adequação, ela tende a oferecer uma comunicação mais eficaz entre os interlocutores de um SRI. É de fundamental importância para as Linguagens Documentárias, ainda, o reconhecimento da variação terminológica conceitual e denominativa na comunicação especializada.

Apesar dos SRIs automatizados oferecerem maior facilidade de uso na busca de informação, ainda assim deixam muito a desejar do ponto de vista da satisfação dos usuários. Com a Informática, o acesso à informação ganhou em operacionalidade e velocidade, mas perdeu em precisão; daí a preocupação em construir ferramentas no âmbito das Linguagens Documentárias capazes de controlar os aspectos negativos da Informática, quanto à busca da informação e recuperação de documentos.

Nos Estudos de Terminologia e na Linguística da Enunciação, encontramos um referencial capaz de contribuir para a construção de vocabulários controlados especializados mais eficazes. Ressaltamos que a ausência de vocabulários controlados especializados em um domínio se reflete na terminologia usada em um catálogo eletrônico e, conseqüentemente, na insatisfação dos usuários em pesquisas feitas aos SRIs. É papel do bibliotecário tentar compatibilizar a linguagem de indexação, com a linguagem utilizada pelo autor dos textos e pelo usuário. É da falta de interação entre seus interlocutores, que resulta a insatisfação dos usuários.

Por meio das teorias acima mencionadas foi possível propor uma nova concepção de descritores, passando a entendê-los não apenas como conceitos identificados como unidades terminológicas, mas também como expressões representativas de um domínio e entendidas como unidades fraseológicas especializadas.

Sob o referencial enunciativo e terminológico acreditamos ser possível construir ferramentas mais produtivas. **A especificidade temática, a precisão conceitual e a construção do sentido são os principais pressupostos teóricos dos vocabulários controlados especializados propostos nesta pesquisa.** Trazemos dos estudos enunciativos o referencial que fundamenta a construção do sentido, assim como a visão de linguagem apresentada nesta teoria.

Declaramos como válido o fato de que um Sistema de Recuperação da Informação precisa estar disposto sob o conceito de estrutura, de forma a viabilizar que os dados registrados no Sistema tenham valor de informação. Pelo fato dos SRIs se estruturarem de modo sistêmico, é essencial pensar na noção de estrutura quando se pretende construir Linguagens Documentárias. Sua importância para os SRIs se deve ao fato de conferirem maior organicidade aos sistemas documentários. Ela é importante, ainda, quando se pensa na organização nocional de um campo de conhecimento.

As noções de texto e de contexto, advindas dos Estudos da Linguagem, favorecem a ideia de construir um item bibliográfico de um SRI capaz de propiciar a recuperação de informações. Somos da opinião que uma informação só tem seu sentido garantido quando é

capaz de preservar seu contexto de uso, e que a construção do conhecimento só é possível quando se tem acesso à informação, e não apenas a dados, tal como acontece quando o termo registrado no Sistema não preserva a referência que assegura o sentido da unidade indexada.

Dos estudos terminológicos, ainda, incorporamos uma visão mais ampla e ao mesmo tempo mais precisa de descritores, e a ideia de que um descritor não pode ser desvinculado de seu contexto. Por expressarem o conhecimento especializado, as ULRIEs tendem a oferecer uma comunicação mais eficaz entre os interlocutores de um Sistema de Informação. Elas são as unidades lexicais que representam tanto a unidade terminológica como a unidade fraseológica especializada, e que são recomendadas para uso por serem as mais representativas de um domínio. Sob esse entendimento reconhecemos os sintagmas terminológicos expressando conceitos ou noções de uma área de conhecimento como ULRIEs; ressaltamos que os sintagmas terminológicos ocorrem com uma alta frequência na comunicação especializada.

Formamos a convicção que uma ULRIE é mais adequada que um descritor para representar a informação indexada, na medida em que ela amplia a visão usualmente atribuída a um descritor, que é fundamentalmente baseada na ideia de conceito. **Uma ULRIE se caracteriza por ter como propósito o registro de uma informação contextualizada, precisa, específica e representativa da linguagem especializada.**

Um vocabulário controlado especializado tende a ser mais representativo de um domínio e da linguagem utilizada pelos especialistas, quando construído a partir de conceitos e noções de uma área de conhecimento e da formação de um *corpus* formado por textos de caráter acadêmico de autoria dos próprios especialistas.

Com esse referencial, acrescido daquele advindo da Linguística de *Corpus* e da Ontologia, foi possível sugerir uma nova metodologia para a coleta de ULDEs e seleção de ULRIEs de um vocabulário controlado especializado. Afirmamos que a Linguística de *Corpus* e a Ontologia demonstram ser um recurso adequado para a construção dessas ferramentas, uma vez que agilizam a construção das mesmas. Elas estão em consonância com esta pesquisa, já que permitem examinar o funcionamento da linguagem pela observação de dados linguísticos autênticos, organizados em *corpora*.

A representatividade é uma das características de um *corpus*, estando, assim, em conformidade com a construção de vocabulários controlados especializados conforme propostos nesta pesquisa. Esta representatividade é alcançada pela indexação de documentos por diferentes níveis de especificidade

Há muitos anos atrás, Cutter já priorizava o registro da forma utilizada pelo usuário na representação da informação indexada. Ele recomendava, ainda, o uso do termo mais específico para representar a temática de um documento. Esse seu posicionamento vem ao encontro do que pregam os estudos terminológicos.

O registro da terminologia de um domínio visando a construção de vocabulários controlados especializados foi possível a partir da proposição da arquitetura de uma base de dados. Os *softwares* Excel e Access se apresentam como ferramentas adequadas para tal. A construção de uma base de dados para o registro da terminologia de uma área do conhecimento, estruturada a partir das cinco categorias propostas nesta pesquisa – ULDEs, use, ULRIEs, RAs, FRs – tende a assegurar a contextualização dos dados indexados.

Concluindo que alcançamos os objetivos propostos nesta pesquisa, e que também atingimos os pressupostos teóricos apontados, confirmamos nosso entendimento inicial de que a relação que se estabelece entre os interlocutores de um SRI é de intersubjetividade, e que esta relação deve ser vista de modo integrado entre os processos de indexação e os de busca da informação e recuperação de documentos. E ainda, que o aparelho formal de enunciação formulado por Benveniste (1988a, 1989a), acrescido da reinterpretação de Martins (1990), viabiliza sistematizar essa relação. O uso das noções de linguagem, subjetividade, intersubjetividade, referência, contexto e categorias de pessoa, espaço e tempo tornou possível perceber as peculiaridades presentes nesta relação. Ratificamos nossa ideia de que os interlocutores da relação que se estabelece no processo de comunicação documentária são agentes desse processo, e que são, além dos SRIs e usuários, também os bibliotecários.

O desenvolvimento deste quadro teórico foi fundamental para construir um novo referencial a partir de parâmetros enunciativos e terminológicos, em relação às questões investigadas. **Com o desenvolvimento desta pesquisa foi possível estabelecer uma interface produtiva entre os Estudos da Linguagem e os desenvolvidos em Ciência da Informação**, de forma que as duas áreas se beneficiassem mutuamente.

Como perspectivas para pesquisas futuras, sugerimos aprofundar estudos sobre teorias da enunciação, tipologias textuais e ontologias.

Nesta pesquisa centramos nossa investigação na Teoria de Enunciação de Benveniste. Mas entendemos que outros teóricos merecem ser investigados com o propósito de discutir os modelos de comunicação documentária e os Sistemas de Recuperação da Informação. O desenvolvimento de estudos sobre o comportamento de usuários e de bibliotecários, sob esse referencial, pode contribuir para tornar a relação que se estabelece entre bibliotecários, SRIs e usuários mais satisfatória.

É possível que estudos sobre tipologias textuais sejam úteis para o estabelecimento de políticas de indexação e construção de vocabulários controlados especializados. Dentre as noções abordadas nesses estudos destacamos para exame minucioso, as que se referem a textos especializados e variação conceitual e denominativa. É provável que estudos sobre tipologia textual sejam capazes de oferecer um referencial adequado para fundamentar os procedimentos de determinação do grau de exaustividade e de seletividade com que se deve tratar tematicamente um texto e os conceitos neles produzidos.

As ontologias merecem uma investigação mais aprofundada, especialmente quanto a forma com que podem ser úteis no âmbito da construção de Linguagens Documentárias. Os mecanismos de extração automática de termos que oferecem devem ser aproveitados na construção de vocabulários controlados especializados.

Como perspectivas futuras esperamos, ainda, dar continuidade à construção dos vocabulários controlados desenvolvidos nesta pesquisa em nível de projetos pilotos. Estudos nesse sentido trariam grande contribuição para o desenvolvimento das Linguagens Documentárias e construção de SRIs mais eficazes.

REFERÊNCIAS

- ALMEIDA, Mauricio B.; BAX, Marcello P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, Brasília, v. 32, n. 3, set./dec. 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000300002>. Acesso em: 27/03/2012.
- ARAUJO, 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.
- ARAUJO, Vera Maria Araujo Pigozzi de. *Documentação, Terminologia e Linguística: uma interface produtiva*. 2006. 163 f. Dissertação (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2006.
- ARRUDA, Susana Margaret de ; CHAGAS, Joseane. *Glossário de Biblioteconomia e Ciências afins*. Florianópolis : Cidade Futura, 2002. 229 p.
- BARBOSA, Alice Príncipe. *Teoria e prática dos sistemas de classificação bibliográfica*. Rio de Janeiro: Instituto Brasileiro de Bibliografia e Documentação, 1969. 441 p.
- BARITÉ, Mario et al. Garantia literária: elementos para uma revisão crítica após um Século. *TransInformação*, Campinas, v. 22, n. 2, p. 124, 129, maio/ago., 2010. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/490>>. Acesso em: 27/07/2013.
- BARROS, Lidia Almeida. *Curso básico de Terminologia*. São Paulo: EDUSC, 2004. 285 p.
- BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988a. 387 p.
- BENVENISTE, Émile. Da subjetividade na linguagem. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988b. p. 284-293.
- BENVENISTE, Émile. 'estrutura' em Linguística. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2.ed. Campinas: Pontes, 1988c. p. 97-104.
- BENVENISTE, Émile. Os níveis da análise Linguística. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988d. p. 127-140.
- BENVENISTE, Émile. Estrutura das relações de pessoa no verbo. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988e. p. 247-259.
- BENVENISTE, Émile. A natureza dos pronomes. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988f. p. 277-283.
- BENVENISTE, Émile. Da subjetividade na linguagem. In: BENVENISTE, Émile. *Problemas de Linguística geral I*. 2. ed. Campinas: Pontes, 1988g. p. 284-293.
- BENVENISTE, Émile. *Problemas de Linguística geral II*. Campinas: Pontes, 1989a. 294 p.
- BENVENISTE, Émile. O aparelho formal da enunciação. In: BENVENISTE, Émile. *Problemas de Linguística geral II*. Campinas: Pontes, 1989b. p. 81-90.
- BENVENISTE, Émile. A forma e o sentido na linguagem. In: BENVENISTE, Émile. *Problemas de Linguística geral II*. Campinas: Pontes, 1989c. p. 220-242.
- BENVENISTE, Émile. A linguagem e a experiência humana. In: BENVENISTE, Émile. *Problemas de Linguística geral II*. Campinas: Pontes, 1989d. p. 68-80.

BENVENISTE, Émile. Semiologia da língua. In: BENVENISTE, Émile. *Problemas de Linguística geral II*. Campinas: Pontes, 1989e. p. 43-67.

BERBER SARDINHA, Tony. *Linguística de Corpus*. São Paulo: Barueri, 2004. 410 p.

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.

BEVILACQUA, Cleci Regina. Unidades fraseológicas especializadas: estado da questão em relação a sua definição, denominação e critérios de seleção. *Tradterm*, São Paulo, v. 11, p. 237-253, 2005.

BEVILACQUA, Cleci Regina. *Unidades fraseológicas especializadas: estado de la cuestión y perspectivas*. Barcelona: Universitat Pompeu Fabra, 1999. Trabalho de Pesquisa.

BIBER, Douglas. Representatividade em planejamento de *corpus*. *Cadernos de Tradução*, Porto Alegre, v. 30, p. 11-45, 2012.

BISCALCHIN, Ricardo; BOCCATO, Vera Regina Cesari. Estudo do uso de linguagem documentária em catálogos coletivos de bibliotecas universitárias: uma abordagem qualitativa-sociocognitiva pela perspectiva do usuário. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA, DOCUMENTAÇÃO E CIÊNCIA DA INFORMAÇÃO, 24º, 2011, Maceió. *Anais...* Maceió: CBBB, 2011. Disponível em: <<http://febab.org.br/congressos/index.php/cbbd/xxiv/paper/view/39/508>>. Acesso em: 23/03/2012.

BLACKBURN, Simon. *Dicionário Oxford de Filosofia*. Rio de Janeiro: Zahar, 1997. 437 p.

BLAIS, Esther. Le phraséologisme. Une hypothèse de travail. *Terminologies Nouvelles*, 10, Bélgica, RINT, 1993. p. 50-56.

BRÄSCHER, Marisa. *Curso de elaboração de tesouros*. Brasília, 1999. 23 p. Polígrafo.

BRÄSCHER, Marisa. *Tesouro, Taxonomia e Ontologia: uma evolução?* In: CICLO DE CONFERÊNCIAS SOBRE ORGANIZAÇÃO DA INFORMAÇÃO LEGISLATIVA E JURÍDICA, 2009, Brasília. *Anais...* Brasília: [s.n.], 2009. Disponível em: <<http://projeto.lexml.gov.br/arqs/BRASCHER.pdf>>. Acesso em: 10/10/2011.

BRÄSCHER, Marisa; CAFÉ, Lígia. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9º., 2008, São Paulo. *Anais...* São Paulo: ANCIB, 2008. p. 87-103. Disponível em: < <http://www.pos.eca.usp.br/sites/default/files/enancibdigital.pdf> >. Acesso em: 02/04/2012.

BRÄSCHER, Marisa, CARLAN, Eliana. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, Jaime; BRÄSCHER, Marisa (Org.). *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento – EROIC*. Brasília : IBICT, 2010. Cap.8, p. 147-176. Disponível em: <<http://www2.senado.gov.br/bdsf/bitstream/handle/id/189812/eroic.pdf?sequence=3>>. Acesso em: 26/05/2013.

BRENNAND, Edna Gusmão de Góes ; BRENNAND, Eládio José de Góes. Cognição e redes abertas: a informação interativa como coração dos sistemas inteligentes. *Ciências & Cognição*, Rio de Janeiro, v. 10, p. 54-64, 2007. Disponível em : <http://www.cienciasecognicao.org/pdf/v10/cec_vol10.pdf>. Acesso em: 27/07/2013.

CABRÉ, Maria Teresa. Consecuencias metodológicas de la propuesta teórica (I). In: CABRÉ, Maria Teresa; FELIU, Judit (Org.). *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: IULA, 2001. p. 27-36.

CABRÉ, Maria Teresa. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Antartida/Empuries, 1993. 529 p.

CABRÉ, Maria Teresa. Una nueva teoría de la terminología: de la denominación a la comunicación. In: CORREIA, Margarida (Org.). *Terminología, desenvolvimento e identidade nacional*. Lisboa: Colibri, 2002. p. 41-60.

CABRÉ, Maria Teresa; ESTOPÁ, Rosa. On the units of specialised meaning used in professional communication. *Terminology Science and Research*, Viena, v.14, p. 16-27, 2003. Disponível em: <www.upf.edu/pdi/dtf/rosa.estopa/docums/INSBRUCK.pdf>. Acesso em: 13/09/2010.

CAMPOS, Maria Luiza de Almeida. *Integração de ontologias: o domínio da bioinformática e a problemática da compatibilização terminológica*⁸¹. Niterói: Universidade Federal Fluminense, [2005]. 12 p. [Grupo Temático Ancib: GT 2 : Organização do Conhecimento e Representação da Informação]. Disponível em: <<http://www.uff.br/ppgci/editais/mlconto.pdf>>. Acesso em: 16/09/2013.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Taxonomia e Classificação: o princípio de categorização. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 9, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_01.htm>. Acesso em: 03/04/2012.

CAPES. Portal de Periódicos da Capes. Brasília. Disponível em: <<http://www.periodicos.capes.gov.br/index.php>>. Acesso em: 03/06/2012.

CENDÓN, Beatriz Valadares. Sistemas e redes de informação. In: OLIVEIRA, Marlene de (Coord.). *Ciência da informação e Biblioteconomia: novos conteúdos e espaços de atuação*. Belo Horizonte: Ed. da UFMG, 2005. p. 45-75

CERVANTES, Brígida Maria Nogueira. *Terminologia do processo de Inteligência Competitiva: estudo teórico e metodológico*. Londrina: EDUEL, 2006. 131 p.

CHAUÍ, Marilena. *Introdução à história da filosofia: dos pré-socráticos a Aristóteles*. São Paulo: Brasiliense, 1994. v.1.

CIAPUSCIO, Guiomar Elena. La Terminología desde el punto de vista textual: selección, tratamiento y variación. *Organon*, Porto Alegre, v. 12, n. 26, p. 43-65, 1998.

CIAPUSCIO, Guiomar Elena. *Textos especializados y Terminología*. Barcelona: Institut Universitari de Lingüística Aplicada, 2003. 149 p.

CINTRA, Anna Maria Marques et al. *Para entender as linguagens documentárias*. 2.ed. rev. e ampl. São Paulo: Polis, 2002. 92 p.

CUNHA, Murilo Bastos da ; CAVALCANTI, Cordélia Robalinho de Oliveira. *Dicionário de Biblioteconomia e Arquivologia*. Brasília : Briquet de Lemos / Livros, 2008. 451 p.

⁸¹ Projeto de pesquisa submetido ao CNPq no período de 2005 a 2008.

CURRÁS, Emilia. *Tesouros: linguagens terminológicas*. Brasília: IBICT, 1995. 286 p.

CUTTER, C.A. *Rules for a dictionary catalog*. 4. ed. Washington: Government Printing Office, 1904. Disponível em: <[http://digital.library.unt.edu/ark:/67531/metadc1048/ml/1/?q=rules for a dictionary catalog](http://digital.library.unt.edu/ark:/67531/metadc1048/ml/1/?q=rules%20for%20a%20dictionary%20catalog)>. Acesso em: 17/04/2011.

DEWEY, Melvil. *Dewey Decimal Classification and relative index*. 16. ed. New York: Forest, 1960. 2 v.

DEWEY, Melvil. *Dewey Decimal Classification and relative index*. 20. ed. Albany, New York: Forest, 1989. 4 v.

DEWEY, Melvil. *Sistema de Classificação Decimal: tablas e indice alfabetico auxiliar*. 15. ed. Essex County, New York: Forest, 1955. 1059 p.

DODEBEI, Vera Lúcia Doyle. *Tesouro: linguagem de representação da memória documentária*. Niterói: Intertexto; Rio de Janeiro: Interciência, 2002. 119 p.

FERAUCHE, Thais Maria Yomoto. Teoria geral de Sistemas e Informação. São Paulo : Centro Paula Souza, Faculdade de Tecnologia da Praia Grande, 2006. 56 p. Disponível em: <[http://glauco.net.br/glauconet/si/TGS-%20FATEC.pdf](http://glauco.net.br/glaucenet/si/TGS-%20FATEC.pdf)>. Acesso em: 27/07/2013

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.

FINATTO, Maria José Bocorny. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. *Organon*, v. 25, n. 50, p. 30-45, 2011a. Disponível em: <http://www6.ufrgs.br/textecc/porlexbras/porpopular/arquivos/FINATTO-Organon_FINAL.pdf>. Acesso em: 20/04/2012.

FINATTO, Maria José Bocorny. Estudos sobre linguagens e textos científicos e técnicos: o que é uma terminologia textual? In: BATTISTI, Elisa; COLLISCHONN, Gisela (Org.). *Língua e linguagem: perspectivas de investigação*. Pelotas: EDUCAT, 2011b. p. 153-172.

FINATTO, Maria José Bocorny. Termos, textos e textos *com* termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUIERDO, Aparecida Negri; KRIEGER, Maria da Graça. (Org.). *As ciências do Léxico: Lexicologia, Lexicografia, Terminologia*. Campo Grande: Ed. UFMS, 2004. p. 341-357.

FIRTH, John Rupert. Personality and language in society. In: FIRTH, John Rupert. *Papers in Linguistics: 1934-1951*. London: Oxford University Press, 1969. p. 177-189.

FLORES, Valdir do Nascimento; TEIXEIRA, Marlene. Enunciação, dialogismo, intersubjetividade: um estudo sobre Bakhtin e Benveniste. *Bakhtiniana*, São Paulo, v. 1, n.2, p.143-164, 2009.

FLORES, Valdir do Nascimento; TEIXEIRA, Marlene. *Introdução à Linguística da Enunciação*. São Paulo: Contexto, 2005. 125 p.

FLORES, Valdir do Nascimento et al. *Dicionário de Linguística da Enunciação*. São Paulo: Contexto, 2009. 284 p.

FUJITA, Mariângela Spotti Lopes. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. *DataGramaZero - Revista de Ciência da Informação*, v. 5, n. 4, ago. 2004. Disponível em:< http://www.dgz.org.br/ago04/F_I_art.htm >. Acesso em: 19/02/2013.

- FUJITA, Mariângela Spotti Lopes. *Linguagem documentária em Odontologia: uma aplicação do sistema de indexação PRECIS*. São Paulo, 1992. 3 v. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2006.
- FUJITA, Mariângela Spotti Lopes. *PRECIS na língua portuguesa: teoria e prática de indexação*. Brasília: Ed. da UNB, 1988.
- GAUDIN, François. *Pour une socioterminologie: des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Université de Rouen, 1993. 231 p.
- GIL, Antônio Carlos. *Métodos e técnicas de pesquisa social*. 5. ed. São Paulo: Atlas, 1999. 206 p.
- 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. Disponível em: <<http://www.scielo.br/pdf/ci/v29n1/v29n1a7.pdf>>. Acesso em: 27/07/2013).
- GOMES, Neiva Maria Tebaldi. *Um estudo das relações de (inter)subjetividade presentes na enunciação escrita de professores de língua materna*. 2003. 154 [63] f. Tese (Doutorado em Letras: Estudos da linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.
- GOUADEC, Daniel. Nature et traitement des entités phraséologiques. Terminologie et phraseologie. Acteurs et aménageurs. *Actes du Deuxieme Université d'Automne en Terminologie*. Paris: La Maison du Dictionnaire, 1994. p. 164-193.
- GRUBER, Tom. *What is an Ontology?* [S.l.: s.n., 1993]. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 27/03/2012.
- GUEDES, Emanuel Ferreira; MARTINHO, Noemi; MORAES, João Batista E. de. O 'assunto' na Ciência da Informação: a questão do *aboutness*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 10º, 2009, João Pessoa. *Anais... João Pessoa: ENANCIB, 2009*. Disponível em: <http://dci2.ccsa.ufpb.br:8080/jspui/bitstream/123456789/558/1/GT2%20-%20Poster%20-%20GUEDES,%20Emanuel%20Ferreira_%20MARTINHO,%20Noemi_%20MORAES,%20Jo%C3%A3o%20Batista%20E.%20de%20-%20O%20assunto%20na%20ci%C3%A2ncia%20da%20informa%C3%A7%C3%A3o.pdf>. Acesso em: 27/03/2012.
- GUINCHAT, Claire; MENU, Michel. *Introducción general a las ciencias y técnicas de la información y de la documentación*. Paris: UNESCO, 1983. 477 p.
- GUINCHAT, Claire; MENU, Michel. Linguagens documentárias. In: GUINCHAT, Claire; MENU, Michel. *Introdução geral às ciências e técnicas da informação e documentação*. 2.ed. corr. e aum. Brasília: IBICT, 1994. p. 131-165.
- HALLIDAY, M. A. K. *Corpus studies and probabilistic grammar*. In: AIJMER, Karin; ALTENBERG, Bengt (Org.). *English Corpus Linguistics: studies in honour of Jan Svartvik*. London: Longman, 1991. p. 30-43.
- HALLIDAY, M. A. K. Language as system and language as instance: the *corpus* as a theoretical construct. In: SVARTVIK, J. (Ed.). *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. Berlin, New York: Mouton de Gruyter, 1992. p. 61-78.
- HOFFMANN, Lothar. Conceitos básicos da linguística das linguagens especializadas. *Cadernos de Tradução*, Porto Alegre, n.17, p. 79-90, out./dez. 2004.

- ISO 1087. *Terminology – Vocabulary. Terminologie - Vocabulaire*. Genève: ISO, 1990.
- KRIEGER, Maria da Graça. (Org.). *As ciências do Léxico: Lexicologia, Lexicografia, Terminologia*. Campo Grande: Ed. UFMS, 2004. p. 341-357.
- 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; FINATTO, Maria José Bocorny. *Introdução à Terminologia: teoria & prática*. São Paulo: Contexto, 2004. 223 p.
- KUHN, Tanara Zingano. *Princípios de análise enunciativa de fatos de língua*. Porto Alegre, 2009. 112 f. Dissertação (Mestrado em Estudos da linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul.
- 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 inter-relação lógica*, 2002. 173 f. Tese (Doutorado em Linguística) – Instituto em Letras, Porto Alegre, 2002.
- LALANDE, André. *Vocabulaire technique et critique de la philosophie*. Paris: Presses Universitaires de France, 1997. 1323 p.
- LANCASTER, F.W. *Indexação e resumos: teoria e prática*. 2. ed. rev. atual. Brasília: Briquet de Lemos, 2004. 452 p.
- LARA, Marilda Lopes Ginez de. Conceitos de Organização e Representação do Conhecimento na ótica das reflexões do grupo Tema. *Informação & Informação*, Londrina, v. 16, n. 3, p. 92-121, jan./jun. 2011. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/10391/9285>>. Acesso em: 10/02/2013.
- LARA, Marilda Lopes Ginez de. Linguagem documentária e terminologia. *TransInformação*, Campinas, v. 16, n. 3, p. 231-240, set./dez. 2004. Disponível em: <<http://revistas.puc-campinas.edu.br/transinfo/viewissue.php?id=8>>. Acesso em: 24/02/2012.
- LARA, Marilda Lopes Ginez de. O processo de construção da informação documentária e o processo de conhecimento. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 7, n. 2, p. 127-139, jul./dez. 2002.
- LARA, Marilda Lopes Ginez de. 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: 08/02/2012.
- LARA, Marilda Lopez Ginez de. *Representação e Linguagens Documentárias: bases teórico-metodológicas*. São Paulo, 1999. 208 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1999.
- LIMA, Gercina Ângela Borém. *Protótipo Mapa Hipertextual - MHTX: um modelo para organização hipertextual de documentos acadêmicos por meio do uso de Mapas Conceituais, Análise Facetada e Sistemas Hipertextuais*. Belo Horizonte: Escola de Ciência da Informação da UFMG, 2004. Disponível em: <<http://www.gercinalima.com/mhtx/pages/apresentacao.php>>. Acesso em: 18/02/2013.

- LIMA, Vânia Mara Alves. *Terminologia, comunicação e representação documentária*. 1998. 117 f. Dissertação (Mestrado em Ciências da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1998. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27143/tde-11052004-122839/pt-br.php>>. Acesso em: 20/03/2012.
- LOPES, Lucelene. *Extração Automática de Conceitos a partir de Textos em Língua Portuguesa*. 2012. 156 f. Tese (Doutorado em Ciência da Computação) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2012.
- LOPES, Lucelene et al. ExATOLp – An Automatic Tool for Term Extraction from Portuguese Language Corpora. In: LTC'09 Language and Technology Conference, 4th, 2009, Poznan. *Proceedings...* Poznan: Adam Mickiewicz University, 2009. p. 427-431. Disponível em: <http://www.inf.pucrs.br/~ontolp/exatolp/download/LTC_2009.pdf>. Acesso em: 04/04/2012.
- LOPES, Lucelene; VIEIRA, Renata. Extração automática de sintagmas nominais para construção ontologias. In: ENCONTRO DE LINGÜÍSTICA DE CORPUS - ELC, 9^o, 2010, Porto Alegre. ELC, 2010. Disponível em: <<http://www.inf.pucrs.br/~linatural/Docs/publicacoes/elcLucelene.pdf>>. Acesso em: 03/04/2012.
- 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 linguística: uma introdução*. Rio de Janeiro: LTC, 1987. 322 p.
- MACIEL, Anna Maria Becker. *Terminologia, linguística de Corpus e organização da informação legislativa e jurídica*. Brasília, 2009. Disponível em: <<http://projeto.lexml.gov.br/arqs/MACIEL.pdf>>. Acesso em: 15/04/2012.
- MAI, Jens-Erik. Deconstructing the indexing process. *Advances in Librarianship*, Bingley, v. 23, p. 269-298, 2000. Disponível em: <<http://individual.utoronto.ca/jemai/Papers/2000Deconstructing.pdf>>. Acesso em: 27/02/2012.
- MAI, Jens-Erik. Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, Bingley, v. 57, p. 591-622, 2001. Disponível em: <http://individual.utoronto.ca/jemai/Papers/2001_Semiotics.pdf>. Acesso em: 27/02/2012.
- MARTINS, Eleni. *Enunciação e diálogo*. Campinas: Ed. da UNICAMP, 1990. 196 p.
- MEY, Eliane Serrão Alves ; SILVEIRA, Naira Christofolletti. *Catálogo no plural*. Brasília : Briquet de Lemos/Livros, 2009. 217 p.
- MORENO, Fernanda Passini. *Requisitos funcionais para Registros Bibliográficos – FRBR: um estudo no catálogo da Rede Bibliodata*. 2006. 202 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação – FACE, Universidade de Brasília, Brasília, 2006. Disponível em: <<http://repositorio.unb.br/bitstream/10482/2565/1/DISSERTACAO%20FERNANDA%20MORENO%20-%20UnB.pdf>>. Acesso em: 27/07/2013.
- NASSIF, Mônica Erichsen ; VENÂNCIO, Ludmila Salomão ; HENRIQUE, Luiz Cláudio Junqueira. Sujeito, contexto e tarefa na busca de informação: uma análise sob a ótica da cognição situada. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 8, n.

5, out. 2007. Disponível em: <http://www.dgz.org.br/out07/Art_04.htm>. Acesso em: 16/09/2013.

OLIVEIRA, Djalma de Pinho Rebouças de. *Sistemas, Organização & Métodos: uma abordagem gerencial*. 18. ed. São Paulo : Ed. Atlas, 2009. 480 p.

ORTEGA, Cristina Dotta. Relações históricas entre Biblioteconomia, Documentação e Ciência da Informação. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 5, n. 5, dez. 2004. Disponível em: <http://www.dgz.org.br/out04/Art_03.htm>. Acesso em: 24/03/2013.

ORTEGA, Cristina Dotta; LARA, Marilda Lopes Ginez de. A noção de estrutura e os registros de informação dos sistemas documentários. *TransInformação*, Campinas, v.22, n.1, p. 7-17, jan./abr. 2010. Disponível em: <<http://dci2.ccsa.ufpb.br:8080/jspui/bitstream/123456789/429/1/GT%20%20Txt%201-%20ORTEGA,%20Cristina%20Dotta%3B%20LARA,%20Marilda%20Lopes%20Ginez.pdf>>. Acesso em: 19/05/2011.

PAVIANI, Jayme. *Filosofia e método em Platão*. Porto Alegre: EDIPUCRS, 2001. 256 p.

PEARSON, Jennifer. *Terms in context*. Amsterdam: J. Benjamins, 1998. 242 p.

PIEIDADE, Maria Antonieta Requião. *Introdução à teoria da classificação*. Rio de Janeiro: Interciência, 1977. 185 p.

PINHEIRO, Lena Vânia Ribeiro. Medidas de consistência da indexação: interconsistência. *Ciência da Informação*, Brasília, v.7, n.2, p. 109-14, 1978. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1682/1287>>. Acesso em: 27/07/2013

REALE, Giovanni. *História da filosofia antiga*. São Paulo: Loyola, 1995. v.5.

SAGER, Juan C. *A practical course in terminology processing*. Amsterdam: J. Benjamins, 1990. 254 p.

SALES, Rodrigo de; CAFÉ, Ligia. Semelhanças e diferenças entre tesouros e ontologias. *DataGramaZero: Revista de Ciência da Informação*, Rio de Janeiro, v. 9, n. 4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_02.htm>. Acesso em: 27/03/2012.

SÁNCHEZ, Aquilino (Ed.). *C U M B R E: corpus lingüístico del español contemporáneo. Fundamentos, metodología y aplicaciones de los corpus lingüísticos*. [Alcobendas (Madrid): Sociedad General Española de Librería, 1995.] Disponível em: <www.um.es/lacell/miembros/asp/cumbre.doc>. Acesso em: 15/04/2012.

SANTOS, Gildenir Carolino ; RIBEIRO, Célia Maria. *Acrônimos, siglas e termos técnicos: Arquivística, Biblioteconomia, Documentação, Informática*. Campinas: Ed. Átomo, 2003. 277 p.

SAUSSURE, Ferdinand de. *Cours de linguistique générale*. 3e. éd. Paris: Payot, 1968.

SAUSSURE, Ferdinand de. *Curso de linguística geral*. 4. ed. São Paulo: Cultrix, 1972.

SHANNON, C.E.; WEAVER, W. *The mathematical theory of communication*. Illinois: The University of Illinois, 1949.

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.

TAGNIN, Stella Ester Ortweiler. *O jeito que a gente diz: expressões convencionais e idiomáticas*. São Paulo: DISAL, 2005. 117 p.

VIEIRA, Renata. Linguística Computacional: uma entrevista com Renata Vieira. *Revista Virtual de Estudos da Linguagem: ReVEL*, Novo Hamburgo, v. 2, n. 3, 3 p., ago. 2004.

Disponível em:

<
http://www.revel.inf.br/files/entrevistas/revel_3_entrevista_renata_vieira.pdf >. Acesso em: 19/05/2013.

VIEIRA, Renata; LIMA, Vera Lúcia Strube. Linguística computacional: princípios e aplicações. In : CONGRESSO DA SOCIEDADE BRASILEIRA DA COMPUTAÇÃO, 21º, 2001, Fortaleza. *Anais...* Maceió, SBC, 2001. Disponível em: <http://www.inf.unioeste.br/~jorge/MESTRADOS/LETRAS%20-%20MECANISMOS%20DO%20FUNCIONAMENTO%20DA%20LINGUAGEM%20-%20PROCESSAMENTO%20DA%20LINGUAGEM%20NATURAL/ARTIGOS%20INTERESSANTES/lingu%EDstica%20computacional.pdf>. Acesso em: 08/09/2013.

VIEIRA, Renata; LOPES, Lucelene. Processamento de Linguagem Natural e o Tratamento Computacional de Linguagens Científicas. In: PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO Maria José. (Org.). *Linguagens Especializadas em Corpora: modos de dizer e interfaces de pesquisa*. Porto Alegre: EDIPUCRS, 2010. p. 183-201.

VITAL, Luciane Paula; CAFÉ, Ligia. Proposta para o desenvolvimento de taxonomias em portais corporativos. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 16, n.4, p. 42-54, out./dec. 2011. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362011000400005&lng=pt&nrm=iso&tlng=pt>. Acesso em: 30/05/2011.

ZINGANO, Marco Antonio de Ávila. *Platão e Aristóteles: o fascínio da filosofia*. 2005. São Paulo: Odysseus, 2005. 134 p.