

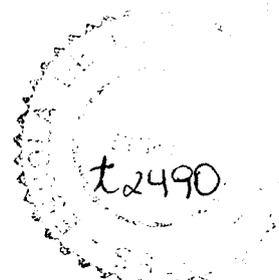
MARILDA LOPES GINEZ DE LARA

A REPRESENTAÇÃO DOCUMENTÁRIA: em jogo a significação

Dissertação apresentada ao Departamento de
Biblioteconomia e Documentação da Escola de
Comunicações e Artes da Universidade de
São Paulo, como exigência parcial para obtenção
do Título de Mestre, sob a orientação da
Profa. Dra. Johanna W. Smit

SÃO PAULO

1993



Mãe
Durval
Rodrigo
Fernão

À mamãe.

Ao Durval, Rodrigo e Fernão.

AGRADECIMENTOS

À Profa. Dra. Johanna W. Smit, pelos ensinamentos e apoio nos vários momentos, pelo carinho e amizade.

À Profa. Dra. Maria de Fátima M. Tálamo, professora e amiga que, com seu dinamismo e vibração me mostrou tantas das várias dimensões da palavra. Em especial, pela oportunidade das inúmeras discussões e pelas muitas sugestões no desenrolar da pesquisa.

À Nair Y. Kobashi, que sempre esteve por perto com sua amizade e carinho e que fez renovar, em mim, a crença na possibilidade do trabalho coletivo. Pelas inúmeras leituras e observações sobre o este trabalho.

À Regina K.O.F. Amaro, à Asa Fujino, ao Waldomiro C.S. Vergueiro, à Maria Helena P. Martins e aos demais colegas de Depto. pelas muitas ajudas, pela amizade e companheirismo.

À Prof. Dra. Maria Aparecida Baccega, pelas observações críticas e sugestões.

À Diana Cohn, pelas contribuições nas nossas leituras metodológicas e pela amizade.

À Sylvia Cioffi, amiga de todas as horas, pelo incentivo e pela paciência e carinho nos momentos de crise.

Ao Marcos Aurélio P. Pessoa, por acreditar no nosso trabalho e pelas muitas discussões teóricas em tempos de prática, no SEADE.

À Cecília de Lara, pelo incentivo e apoio.

À Fundação SEADE e aos amigos de vários tempos que, de diferentes maneiras, contribuíram para que eu pudesse desenvolver este trabalho. Em especial, à Rosa M.G. Beluzzo.

RESUMO

Abordagem semiótica da questão da representação documentária obtida através de instrumentos comutadores, diferenciando-a das demais modalidades de representação da área. Parte-se do pressuposto de que as representações documentárias constituem uma etapa determinada do processo global de Análise Documentária — caracterizada como disciplina de natureza metodológica — e que elas devem funcionar, globalmente, como instrumentos de comunicação. Avaliam-se algumas Linguagens Documentárias tradicionais face a sua capacidade de promover a representação, a normalização de vocabulário e a comunicação de informações. Frente as limitações constatadas das Linguagens Documentárias em exercerem de fato as funções que lhe são associadas, propõe-se as terminologias específicas como seu universo de referência. Obtém-se, assim, linguagens de natureza enciclopédica, suficientemente potentes para promover o tratamento, a recuperação e a circulação de informações documentárias.

ABSTRACT

This paper is a semiotic approach to documentary representation through commuting tools which distinguish these representations from the other in the area. It is presupposed that documentary representations constitute a particular phase in Documentary Analysis' global process — characterized as a methodological subject — and that they must work, globally, as communication tools. Some traditional Documentary Languages are evaluated on the basis of their capacity for promoting representation, vocabulary standardization and communication of information. Due to Documentary Language limitations to exert the functions associated to them, specific terminologies are proposed as their universe of reference. Encyclopaedic languages are obtained, strong enough to promote treating, retrieval and dissemination of documentary informations.

*"Eu vi um homem lá na grimpa do coqueiro, ai-ai,
não era homem, era um coco bem maduro, oi-oi.
Não era coco, era a creca de um macaco, ai-ai,
não era a creca, era o macaco todo inteiro, oi-oi."*

(Cantiga de Espantar Males)

João Guimarães Rosa, epígrafe de "São Marcos", Sagarana.

A representação documentária: em jogo a significação

SUMÁRIO

VOLUME 1

INTRODUÇÃO	1
CAP.1 A PERSPECTIVA DA PESQUISA: repensar o conceito de Representação na Análise Documentária	4
1.1 Análise Documentária e Representação Documentária	4
1.2 O problema e os objetivos da pesquisa	5
1.3 A hipótese de trabalho	8
1.4 Justificativa	9
1.5 As lições da prática	10
CAP. 2 O PERCURSO TEÓRICO E METODOLÓGICO	12
2.1 A questão da representação nos trabalhos sobre Análise Documentária	14
2.1.1 Jean-Claude Gardin	14
2.1.2 A linha de pesquisa em Análise Documentária no Brasil	22
2.2 A contribuição da semiótica segundo Umberto Eco	27
Cap.3 A ATIVIDADE GLOBAL DE ANÁLISE DOCUMENTÁRIA	39

Cap.4	O PROCESSO DE ANÁLISE E SÍNTESE DE TEXTOS	45
4.1	A identificação da informação documentária	45
4.1.1	A leitura documentária	46
4.1.1.1	Características do processo de leitura global	46
4.1.1.2	Características da leitura para fins documentários	49
4.2	O registro da informação documentária: o texto documentário	56
Cap.5	A CARACTERIZAÇÃO DA CONSTRUÇÃO DAS REPRESENTAÇÕES DOCUMENTÁRIAS	62
5.1	Representação documentária, sistemas de significação e informação	62
5.2	As operações de representação	64
5.2.1	O texto documentário como referencial básico	64
5.2.2	A mediação do código documentário	66
5.3	A construção das Linguagens Documentárias	69
5.3.1	A designação dos instrumentos de intermediação documentária	70
5.3.2	A definição e a Linguagem L em formato de enciclopédia como referência para a construção das Linguagens Documentárias	73
5.3.3	A Terminologia como referencial enciclopédico	77
Cap.6	AValiação DE ALGUMAS LINGUAGENS DOCUMENTÁRIAS	85
6.1	Os sistemas de classificação bibliográfica: o Dewey Decimal Classification - CCD e a Classificação Decimal Universal - CDU	85
6.1.1	Identificação	85
6.1.2	Objetivos e princípios de construção	86
6.1.3	Forma de apresentação	87
6.1.3.1	O Dewey Decimal Classification	87

6.1.3.2	A Classificação Decimal Universal	89
6.1.4	Análise de exemplos	91
6.1.5	Avaliação geral	95
6.1.5.1	Estrutura das LDs	95
6.1.5.2	Eficiência para o controle de vocabulário	97
6.1.5.3	Eficiência para a elaboração de índices	98
6.1.5.4	Capacidade para representação da informação	99
6.1.5.5	Eficiência na comunicação documentária	101
6.2	Os tesauros: o Thesaurus POPIN e o MACROTHESAURUS	102
6.2.1	Identificação	102
6.2.2	Objetivos e princípios de construção	102
6.2.3	Forma de apresentação	104
6.2.3.1	O Thesaurus POPIN	104
6.2.3.2	O MACROTHESAURUS	106
6.2.4	Análise de exemplos	108
6.2.5	Avaliação geral	120
6.2.5.1	Estrutura das LDs	120
6.2.5.2	Eficiência para o controle de vocabulário	122
6.2.5.3	Eficiência para a elaboração de índices	124
6.2.5.4	Eficiência para representar a informação	124
6.2.5.5	Eficiência na comunicação documentária	125
	Conclusão	126
	Referências Bibliográficas	130

VOLUME 2

Anexo 1 - Malha de Informações do Setor Saúde	1
Anexo 2 - Exemplos	3

INTRODUÇÃO

Na pesquisa que desenvolvemos procuramos discutir algumas questões relativas à construção da representação documentária, a partir de referencial teórico da área de Análise Documentária - AD. Para inserir tais discussões no âmbito da Comunicação Social buscamos apoio em alguns estudos semióticos que, se não nos forneceram instrumental operacional para o desenvolvimento de modelos de análise em Documentação, nos proporcionaram a oportunidade de identificar algumas características do signo documentário face a outros signos.

No primeiro capítulo procuramos explicitar a perspectiva da análise adotada, a forma de desenvolvimento da pesquisa e as razões que nos levaram a abordar a questão da representação na sua inserção nos processos comunicacionais.

No segundo capítulo apresentamos a síntese das contribuições que consideramos mais significativas para a realização deste trabalho. Num primeiro momento, destacamos a abordagem sobre análise e representação documentárias de Jean-Claude Gardin e de pesquisas brasileiras desenvolvidas a partir desse referencial. Num segundo momento, procuramos explorar as reflexões de Umberto Eco sobre a representação, a partir de um enfoque semiótico.

No terceiro capítulo procuramos refletir sobre o objeto da AD como um todo realizando uma breve comparação com a Análise do Discurso, com o objetivo de identificar, mais claramente, o ponto de partida da análise em direção às representações documentárias.

Em seguida, destacamos a natureza metodológica da AD evidenciando a necessidade de enfocá-la a partir dos processos que ela envolve, e conseqüentemente, os níveis de representação configurados em seus produtos, a cada fase.

Desse modo, procuramos caracterizar o processo de AD como um conjunto de operações de desestruturação e estruturação — do texto original ao documento, e deste à representação — onde se verifica uma intensificação da redução, a caminho da generalização. Inserindo a atividade global de AD na esfera comunicacional, procuramos levantar as inúmeras variáveis intervenientes no processo.

No capítulo 4 procuramos verificar as características das etapas de identificação da informação documentária e seu registro, analisando as especificidades da leitura documentária e a necessidade de explicitação da atividade mental de redução para a otimização da representação via Linguagens Documentárias - LDs.

No capítulo 5 discutimos vários aspectos relativos à representação via intermediação de LDs.

Primeiramente, analisamos a relação entre significação e informação, procurando mostrar que a informação documentária (e a representação correspondente) remete a sistemas de significação consubstanciados nos textos. Sob este aspecto discutimos o porquê da necessidade de utilização de códigos documentários uma vez que as representações sem intermediação mantêm maior fidedignidade de reprodução da significação dos textos originais. Como decorrência, discutimos as características do signo e da semiose “documentárias”.

Em seguida, procuramos enfatizar a necessidade metodológica do registro da redução documentária para a operacionalização da representação via LDs e levantamos as decorrências do uso de instrumentos de intermediação.

Face às várias maneiras de nomear tais instrumentos de intermediação — léxico, nomenclatura, terminologia, linguagem etc. — procuramos discutir a propriedade do uso das denominações que lhe são invariavelmente atribuídas, procurando ressaltar, nesse sentido, que a base da designação encontra-se no próprio conceito de representação utilizado e na função que se deseja que uma LD desempenhe.

Considerando a necessidade de incorporar uma teoria do significado que considere a integração

da semântica e da pragmática, discutimos as explorações teóricas sobre a possibilidade do desenvolvimento de linguagens semióticas, destacando as diferenças entre as potencialidades de uma representação dicionarial e de uma representação enciclopédica.

Num último item do capítulo, sugerimos a Terminologia como referencial enciclopédico para a construção das LDs, tomando como base as contribuições de Le Guern, segundo as quais a condição de referência é exclusiva da palavra no discurso.

O último capítulo foi reservado à parte empírica da pesquisa: a análise das LDs escolhidas — a Classificação Decimal de Dewey, a Classificação Decimal Universal, o Thesaurus POPIN e o MACROTHESAURUS. A análise foi feita através de exemplos e tomou como base os resultados da reflexão teórica. Analisamos aspectos relativos à estruturação formal desses instrumentos e sua eficácia relativamente ao controle do vocabulário, representação e comunicação de informações.

Como conclusão, apresentamos a sistematização das reflexões teóricas e práticas, procurando apontar a necessidade de desenvolvimento de metodologias específicas de análise e representação documentárias, bem como requisitos mínimos, do ponto de vista da AD, para a construção de LDs consistentes.

Esperamos que as discussões realizadas possam auxiliar na sistematização de procedimentos em AD e apontar alguns caminhos a serem explorados.

CAPÍTULO 1

A PERSPECTIVA DA ANÁLISE:

repensar o conceito de Representação na Análise Documentária

1.1 Análise Documentária e representação documentária

O assunto global de nossa pesquisa — a Análise Documentária - AD — define-se como uma atividade específica no interior da Documentação, respondendo pela análise e tratamento da informação com o objetivo de recuperá-la e disseminá-la. Nessa perspectiva, a AD pode ser concebida como atividade essencial para o estabelecimento da comunicação em sistemas documentários — que passaremos a denominar “Comunicação Documentária” — processo que envolve a codificação e a decodificação de conteúdos informacionais, ou seja, o tratamento e a recuperação da informação.

Globalmente, “a Análise Documentária tem por objetivo representar conteúdos de documentos, tendo em vista um fim pragmático: a recuperação da informação” (KOBASHI, 1988, p.19, grifo meu). Entretanto, pode-se identificar no processo de AD dois tipos distintos de representação, como veremos a seguir.

No primeiro tipo, a representação é construída através de um processo de condensação intensiva do texto original. Essa representação gera uma série de produtos documentários (vários tipos de resumo) que apresentam, geralmente, uma relação de contigüidade e semelhança com o texto que lhes deu origem. A operação, nesse caso, vale-se, em princípio, dos mesmos elementos do sistema semiótico utilizado para a elaboração do texto original.

No segundo tipo, a representação é realizada através do uso de um código (1) comutador (2), ou

(1) “Código é um sistema de sinais — ou de signos, ou de símbolos — que, por convenção prévia, se destina a representar e a transmitir a informação entre a fonte de sinais — ou o emissor — e o ponto de destino — ou receptor” (DUBOIS et alii, 1973). “(...) é um sistema de transmutação da forma de uma mensagem em outra forma que permita a transmissão da mensagem” (idem ibidem).

(2) A comutação é uma operação pela qual se “verifica a identidade paradigmática de duas formas da língua” (DUBOIS et alii, 1973). As LDs são tradicionalmente denominadas instrumentos comutadores ou de conversão uma vez que permitem representar a informação presente numa determinada forma lingüística em outra forma, dita documentária.

seja, uma Linguagem Documentária - LD (3), que tem como função a normalização das unidades significantes ou conceituais presentes no texto original, a partir de elementos que constituem, de alguma forma, uma condensação de áreas de assunto. A condensação, nesse caso, é expressa pelos elementos do código de comutação, sendo portanto, exterior ao texto submetido à conversão. Dessa forma, a representação obtida não apresenta necessariamente uma relação de contigüidade e semelhança com o texto original e envolve, pelo menos, dois sistemas semióticos distintos: aquele presente no texto original e aquele estabelecido pelo instrumento comutador ou LD.

Vê-se, portanto, que o processo de AD comporta operações de representação de naturezas diferentes. Nesse sentido, uma distinção clara sobre a concepção de “representação da informação” inerente aos dois processos, pode levar ao aperfeiçoamento das metodologias de AD, contribuindo, dessa forma, para a otimização da “Comunicação Documentária”.

1.2 O problema e os objetivos da pesquisa

Nosso problema de pesquisa é o de verificar a natureza e as características das representações documentárias obtidas através do uso de códigos comutadores.

Para abordar a questão é necessário, num primeiro momento, verificar as diferenças entre as representações realizadas através de condensação de textos originais, daquelas obtidas via utilização de LDs, visando identificar qual é a concepção de representação implícita em cada um dos processos.

Para a consecução desse objetivo, discutiremos:

a) O processo de análise de textos com fins de identificação de conteúdos informacionais que compreendem, tradicionalmente, a análise e síntese documentárias, cujo objetivo primeiro é a representação por condensação;

(3) Linguagem Documentária é um conjunto de termos, providos ou não de regras sintáticas, utilizada para representar conteúdos de documentos científicos, com fins de classificação ou busca retrospectiva de informações (GARDIN et al., 1968, p.26).

- b) O processo de representação desencadeado após a síntese documentária, via LDs;
- c) A capacidade de representação de conteúdos informacionais das LDs de estrutura classificatória e combinatória.

A abordagem desse problema vincula-se a um objetivo geral de pesquisa que é o de realizar uma reflexão teórica sobre o processo global de AD, na perspectiva de sua inserção nos processos comunicacionais. Vincula-se, também, a um objetivo mais específico, qual seja, o de verificar a potencialidade de representação dos códigos de comutação mais utilizados na área da Documentação face a objetivos práticos de recuperação e disseminação da informação. Com isso, pretendemos identificar, também, a relação existente entre o texto original e a representação documentária obtida através das operações de comutação.

De um modo geral, a função de representação das LDs está presente, ou é subentendida, nas definições mais clássicas de LDs.

Coyaud, por exemplo, enfatiza os aspectos comunicacionais das LDs: “As LDs são ferramentas de comunicação mediata entre usuários e autores de documentos” (COYAUD, 1966, p.127) (4), atribuindo às LDs a função de representação quando define a AD como “uma tradução (5) de textos escritos em língua natural para sua representação em uma linguagem documentária” (idem *ibidem*, p.123). Para o autor, a “análise documentária é primeiramente uma operação de reconhecimento das unidades lexicais que representam as noções importantes de um documento” (COYAUD, 1967, p.18).

Para Gardin, representação é um termo empregado para designar o produto da AD resultante de uma operação semântica que transforma um texto original em uma ou várias palavras-chave, ou ainda, paráfrases formuladas em termos diferentes, pressupondo, portanto, a existência de uma metalinguagem autônoma (6) (GARDIN, 1970, p.631-632). Para o autor, tem-se uma ‘represent-

(4) As citações de publicações estrangeiras constituem tradução livre dos textos originais.

(5) Embora a literatura em AD utilize o termo “tradução”, deve-se observar que tal processo documentário é, antes, de transcodificação, uma vez que as LDs constituem códigos substancialmente diferentes das línguas coloquiais. Diferentemente da transmutação, a tradução não altera a forma, uma vez que nela são conservadas as equivalências semânticas e estilísticas da língua-fonte para a língua de chegada (DUBOIS et alii, 1973).

tação' quando os procedimentos de AD envolvem a utilização de “descritores” (7) que, diferentemente dos métodos de tabulação, não pertencem necessariamente à linguagem dos textos estudados (GARDIN, 1973b, p.70). Em trabalhos mais recentes, Gardin utiliza “sistema de representação” para denominar tais “metalinguagens documentárias”: “(...) sistemas de representação nomeados de forma diversa — Linguagens Documentárias, códigos, tesouros — que se utiliza tanto para normalizar as descrições (...) como para interpretar as questões dos usuários (...)” (GARDIN, 1987b, p.13).

Vemos, portanto, que a função de representação das LDs está presente, de forma explícita ou implícita, em trabalhos significativos da área, muito embora a forma de tratamento da questão não tenha enfatizado os aspectos comunicacionais.

Tomando o produto genérico da AD como “representação”, a literatura em geral não aborda a questão do ponto de vista das diferenças entre o produto documentário obtido por condensação e por intermediação de códigos comutadores. Nos trabalhos citados (Gardin e Coyaud), por exemplo, a tônica recai sobre a representação via códigos de conversão, sendo praticamente ignorados produtos como os resumos.

O enfoque da questão da representação sob a perspectiva que adotamos impõe, necessariamente, a discussão sobre o jogo das significações que o processo envolve: se através de procedimentos de condensação, que caracterizam o primeiro tipo de operação de representação, procura-se, via de regra, uma redução que substitua o texto original, poder-se-ia igualmente afirmar que a condensação realizada através da intermediação de um novo sistema semiótico — uma LD — garanta esse status substitutivo?

Essa questão, portanto, é fundamental tanto para a avaliação das LDs enquanto instrumentos de

(6) A designação de metalinguagens dada às LDs é questionável. Para Dubois e outros, uma “metalíngua é uma língua artificial que serve para descrever uma língua natural cujos termos são os da língua objeto de análise, mas que têm uma só acepção e cujas regras de sintaxe são também as da língua analisada” (DUBOIS et alii, 1973). Para Hjelmslev, uma metalíngua se caracteriza pela existência de dois planos — o plano da expressão e o plano do conteúdo — onde o plano do conteúdo é por si mesmo uma linguagem, ou seja, uma semiótica (HJELMSLEV, 1975, p.126). Geralmente, essa metassemiótica (ou metalinguagem) é inteira ou parcialmente idêntica à sua semiótica-objeto (idem *ibidem*).

(7) Descritor: “palavra ou expressão que representa um conceito, ou seja, o termo preferido para a indexação” (IBICT, 1984, p.5)

representação, como para a definição de metodologias específicas de AD.

1.3 Hipótese de trabalho

Nossa hipótese de trabalho é de que o processo de representação em AD é de natureza generalizante, sendo ele intensificado quando do uso de instrumentos comutadores.

Através da condensação do texto original, o produto documentário obtido situa-se entre a generalização e a individualização, expressando a tensão entre esses polos. A representação, neste caso, tem como objetivo evidenciar o que, neste texto, é informação específica, de modo a garantir, sobre a generalidade, a marca do específico, ou seja, ao lado do que é comum, deve destacar o que é particular, individual.

Já a representação via LDs persegue a generalização através de uma normalização crescente da informação do texto original. Para García Gutiérrez, as LDs tradicionais promovem, via de regra, construções desnaturalizadas (8) da informação (GARCÍA GUTIÉRREZ, 1990, p.30, 65, 71 etc.) e como, salvo algumas exceções, não são previstos mecanismos naturalizantes, a recuperação de informações passa a ser apenas uma possibilidade estatística: a partir das representações, pode-se ou não encontrar a informação nos documentos. Acrescente-se ainda que na maior parte das LDs, as noções (quer sejam descritores, símbolos de classificação, cabeçalhos de assunto) e suas inter-relações não são suficientemente definidas e explícitas, sendo que sua utilização pode comprometer o jogo de significações engendrado pelo texto que se quer, através delas, representar. Compromete-se, dessa forma, a recuperação de informações.

Uma vez confirmado o caráter generalizante da representação documentária, deve-se admitir a necessidade de reconsiderar as funções e a estruturação das LDs tradicionais, uma vez que elas representam apenas parcialmente as informações originais, comprometendo, conseqüentemente,

(8) Diferentemente da linguagem coloquial, as LDs são caracterizadas pela quase ausência de articulação. Os mecanismos de naturalização se restringem, na maior parte dos casos, ao uso de conectores lógicos emprestados da álgebra booleana: este e aquele, este ou aquele, este mais aquele, este e não aquele etc.

a recuperação.

1.4 Justificativa

Muito embora a automação tenha contribuído significativamente quanto às possibilidades de busca bibliográfica a partir do aumento de atribuição de descritores utilizados para descrever o conteúdo dos documentos e da utilização de conectores lógicos para a definição de equações para a recuperação, seus resultados ainda estão longe de serem considerados satisfatórios. Não é novidade, por exemplo, que o resultado de uma busca traga invariavelmente um número muito grande de referências, obrigando o usuário a consultar sistematicamente os resumos, como instrumento de decisão para a consulta ou não ao texto original.

Deve-se investigar, portanto, a eficiência dos instrumentos de intermediação (na fase de análise e na de recuperação) para a recuperação da informação, objetivo previsto nas definições clássicas de LDs apontadas acima. Constata-se, pela prática, que através das expressões dos códigos de comutação (descritores, por exemplo), recuperam-se inúmeras referências sobre determinados assuntos, que podem ou não trazer informações pertinentes. Neste sentido, os descritores não recuperam a informação — mas documentos — portanto não a representam satisfatoriamente.

Embora denomine genericamente o produto da AD como “representação”, a literatura não aborda a questão sob o ponto de vista das diferenças relativas ao maior ou menor grau de eficiência da comunicação da informação a partir dos diferentes produtos documentários, quer sejam obtidos por condensação (o resumo, particularmente), quer por intermediação de códigos comutadores.

Deve-se lembrar ainda que, como a maior parte da literatura da área tem enfrentado a questão da representação quase que exclusivamente sob a ótica de sua vinculação com o processo de análise, alguns trabalhos não fazem uma distinção clara entre o processo de representação enquanto atividade e o processo de construção de LDs para a elaboração das representações. Como decorrência, faz-se necessário discutir o conceito de representação implícito nessas diferentes modalidades.

Em resumo, portanto, esta pesquisa proçõe a oportunidade de se discutir a questão qualitativa do problema da representação como elemento de comunicação documentária, bem como de investigar a pertinência do termo para designar o produto documentário obtido através da intermediação de LDs. Por essa razão, não pretendemos, nos limites deste trabalho, dar conta de todos os problemas envolvidos no processo de construção da significação documentária, nem apresentar resultados conclusivos ou delinear metodologias para a construção de instrumentos comutadores.

1.5 As lições da prática

Da nossa atividade prática, buscamos encontrar parâmetros de organização e política de trabalho em documentação mas, principalmente, detectar carências que impediam (ou impedem) a consecução das tarefas de análise e representação de documentos.

Neste sentido, vale salientar a importância do trabalho desenvolvido na Central de Dados e Referências - CDR, da Fundação SEADE - Sistema Estadual de Análise de Dados, onde a necessidade de se dar conta de um amplo espectro de informações sócio-econômicas, cobrindo vários setores de atividade e em distintas formas de apresentação, dirigiu a experimentação de diversos procedimentos metodológicos, nem sempre rigorosos, de tratamento da informação. Na CDR estávamos abertos a uma gama imensa de informações relativas a vários setores sócio-econômicos que, se por um lado nos caracterizava como “desespecializados” fazia dessa não-especialização uma “especial especialidade”, utilizando expressões de um companheiro de equipe (9). Nosso grande debate era imaginar um instrumento para a indexação dos documentos que possibilitasse, por um lado, resolver o problema da especificidade dos conceitos com os quais trabalhávamos, e por outro, respeitar o autor obedecendo o seu “ato de fala”. Julgávamos importante retirar do texto a hierarquia e as articulações utilizadas pelo autor — pois só assim ele se tornaria válido, importante — mas não imaginávamos como isso poderia ser operacionalizado. Por outro lado, só podíamos contar com instrumentos de indexação cujos pontos de vista nos

(9) Marcos Aurélio Pereira Pessoa

pareciam cristalizados. Eles não ofereciam condições para que imprimíssemos nosso enfoque sobre a informação, não abrangiam todos os conceitos com os quais trabalhávamos, e restringiam a possibilidade de combinação, na recuperação, às operações de álgebra booleana.

Na base dessas discussões estava uma forma especial de ver a informação, que se nos apresentou desde o início do SEADE: uma determinada informação não poderia ser vista de forma isolada, fora do contexto situacional a que dizia respeito, e sem considerar seus inter-relacionamentos mútuos. Essa era a base da concepção das Metodologias para a Análise da Informação (SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados, 1978), e particularmente os conceitos subjacentes às “Malhas de Informação” concebidas pela então Equipe de Planejamento daquela instituição, coordenada por Maria Alice Pompéia Gonzaga. Tais malhas (reproduzidas no Anexo 1), apresentam, implicitamente, indicadores daquilo que identificaríamos posteriormente como as “enciclopédias locais” propostas por Umberto Eco (entre outros), guardadas as devidas proporções (ECO, 1984, 1991). As metodologias citadas preconizavam uma nova forma de enxergar a informação, na sua dependência relativamente ao contexto e à situação de colocação, e em função de objetivos determinados. O objetivo implícito naqueles procedimentos de análise, era o de permitir a “democratização da informação”, através da possibilidade de inter-relacionamento de informações de setores sócio-econômicos diversos. Por exemplo, a questão da Mortalidade Infantil poderia ser vista sob o aspecto Demográfico, de Saúde, de Saneamento, de Economia etc., conforme as necessidades informacionais do usuário. Apesar das limitações de seu poder “democratizante” e da ausência de procedimentos rigorosamente delimitados para a operacionalização de análises, a contribuição dessas metodologias se deu no reconhecimento da necessidade de se pensar nas várias leituras possíveis de um documento e nas várias possibilidades de representá-los em função de necessidades de usuários e instituições específicas.

A docência, por outro lado, ao proporcionar a oportunidade de comparar diversos instrumentos utilizados para a representação temática disponíveis na área de Documentação, impôs a necessidade de reflexão sobre aspectos de estruturação e organização formal do conhecimento, sobre a fundamentação dos arranjos conceituais e lexicais, bem como sobre as potencialidades e restrições colocadas face à operacionalização da classificação e da indexação.

CAPÍTULO 2

O PERCURSO TEÓRICO E METODOLÓGICO

O pressuposto básico inerente a esta pesquisa é o de que a análise e a representação de informações é resultado de múltiplas interferências que colocam em jogo a significação. Tal como o afirmou Hjelmslev, “não existem significações reconhecíveis outras que não as significações contextuais. Toda grandeza, e por conseguinte todo signo, se define de modo relativo e não absoluto, isto é, unicamente pelo lugar que ocupa no contexto” (HJELMSLEV, 1975, p.50) (10)

Desde o início, concorrem para o processo de AD variáveis de origens diversas: a língua (tanto a natural como a especializada) e o uso que dela é feito, as articulações e direcionamentos dados pelo texto objeto de análise, as marcas institucionais de vários tipos — a instituição onde se realiza o processo, a formação do analista, o código de intermediação documentária, a ideologia etc., como o sugere CUNHA (CUNHA, 1990, p.70-72).

Sob esse ponto de vista, tanto a informação como a representação documentária são construções, o que implica, entre outras coisas, a necessidade de abordá-las na sua inserção sócio-institucional.

Por outro lado, a natureza metodológica da AD — voltada que é à elucidação dos processos de análise e representação da informação em contextos documentários — direciona a procura de referenciais em outras áreas com objetivos pragmáticos, visando o aperfeiçoamento de seus procedimentos específicos.

Com base nesses pressupostos e para efeito de maior sistematização agrupamos a revisão bibliográfica que realizamos em dois blocos: no primeiro, destacamos os trabalhos diretamente vinculados à questão da AD, concentrando-nos na abordagem de Jean-Claude Gardin, bem como

(10) Significação diferencia-se de significado. “O significado é o conceito ou imagem mental que vem na esteira de um significante, e significação é a efetiva união entre um certo significado e um certo significante” (COELHO NETTO, 1990, p.22-23). “... a questão do significado está no domínio da língua, e a da significação, no da fala. Em outras palavras, a significação de um signo é uma questão individual, localizada no tempo e no espaço, enquanto o significado depende apenas do sistema e, sob este aspecto, está antes e acima do ato individual”(idem *ibidem*).

nos trabalhos brasileiros desenvolvidos a partir desse referencial. No segundo bloco, procuramos explorar as reflexões de Umberto Eco sobre a representação, a partir de um enfoque semiótico.

Quanto aos trabalhos de Jean-Claude Gardin, conferimos maior ênfase àqueles voltados mais especificamente à questão da representação em Documentação. Desta forma, concentramo-nos nas pesquisas do autor desenvolvidas até meados da década de 70, onde a questão documentária é um objetivo claro. Após esse período, altera-se o curso de suas pesquisas e a questão documentária cede lugar, com maior ênfase, às investigações sobre recuperação de raciocínios de área (em Ciências Humanas e particularmente, Arqueologia).

Em relação a Eco, nosso interesse principal prendeu-se aos seguintes fatos: a abordagem conceitual que ele realiza sobre o texto — nosso parâmetro para a seleção de informações — a pesquisa que ele empreende sobre as formas de representação de natureza semiótica e a inserção dessas discussões no âmbito da comunicação. Resumidamente, consideramos os seguintes itens desenvolvidos pelo autor:

- o processo de comunicação e os sistemas de significação;
- a incorporação do elemento pragmático para a análise de texto;
- a noção de enciclopédia para a análise e representação de textos;
- a abordagem do significado;
- a constituição de linguagens enciclopédicas.

A partir das reflexões de Eco, procuramos realizar uma revisão metodológica da atividade de representação em AD, sob o ponto de vista semiótico.

A maior ênfase conferida ao segundo bloco constitui uma tentativa de encontrar elementos para repensar a atividade global de AD exposta no primeiro, conferindo-lhe uma nova perspectiva. Assim, através de Umberto Eco procuramos retomar, ao mesmo tempo, algumas das pesquisas sobre a representação semântica desenvolvidas após 1970, as primeiras delas citadas, entre outros, por Gardin.

2.1 A questão da representação nos trabalhos sobre Análise Documentária

Na maior parte da literatura da área consultada, não se faz uma distinção clara entre a representação obtida através de análise e síntese — representação por condensação — e aquela obtida através do uso de LDs. Isso se deve, em grande parte, ao fato de que a concepção vigente sobre AD identifica a construção de LDs com a própria análise. Por esta razão, julgamos conveniente recuperar as noções de AD e de LDs e localizar, nelas, o enfoque dado à representação.

Por outro lado, é importante verificar quais são os fundamentos subjacentes a essas concepções na literatura da área, e quais as possíveis contribuições de áreas vizinhas como a Lingüística, a Lógica e a Semiótica relativamente à representação textual.

2.1.1 Jean-Claude Gardin

A introdução do termo “Análise Documentária” na literatura da Documentação é em grande parte devida a Jean-Claude Gardin. Nos seus trabalhos iniciais, a AD é identificada com a análise descritiva (que não se deve confundir com representação descritiva, tal como é usada no Brasil para designar as operações de Catalogação). Sua função é a de fornecer uma “representação sistemática de certos fatos que se supõe parcialmente ordenados” (GARDIN, 1968, p.84, 86). Essa representação é de natureza probabilística, uma vez que seus elementos constituem o resultado de escolhas empíricas próprias da AD.

A AD é também concebida como a expressão do conteúdo de textos científicos (GARDIN, 1970, p.630), ou como extração de significado de um texto — o que implica uma referência a algo que é retirado (*taken out*) do texto — e designado por símbolos que não são necessariamente encontrados no texto (uma metalinguagem) (GARDIN, 1973a, p.144). O objetivo da AD é o de isolar o sentido dos textos com a finalidade de permitir pesquisas retrospectivas de informação realizadas sobre o conteúdo ou a significação de textos (GARDIN, 1973b, p.67).

A passagem do texto original à representação constitui, para Gardin, uma “operação semântica,

mesmo que ela não obedeça a uma regra precisa, e que cada organismo, cada analista (...) se limite a ver na ocorrência uma certa regularidade interna, fundada mais sobre a experiência ou hábito do que sobre algum procedimento explícito” (GARDIN, 1970, p.631). Para o autor, esse quadro empírico sofre alterações com a automação, aumentando-se as possibilidades de conferir várias representações a cada documento e exigindo maior formalização dos procedimentos envolvidos nas operações de análise e representação (idem *ibidem*).

Para Gardin, a substituição de textos em LN por representações supõe necessariamente a existência de uma metalinguagem - ML, que é constituída por unidades lexicais (símbolos designando noções ou conceitos) e convenções sintáticas (para a expressão das relações lógicas presentes nos textos em LN). A ML exerce a função de normalização — inexistente com tal caracterização nas LNs — operando semânticamente no tratamento de sinonímias, homotaxias, homografias, polissemias etc. (GARDIN, 1970, p.632).

Observando mais detalhadamente essas MLs, identificam-se três componentes: além do léxico e das relações sintáticas, as relações semânticas ou analíticas, que exprimem as correspondências entre as unidades do léxico. Para Gardin, a diferença entre as relações semânticas e as relações sintáticas (ou lógicas) só tem fins práticos: sob o plano semântico, manifestam-se as relações mais comuns e estabilizadas, enquanto que sob o plano sintático veiculam-se as relações novas, que são objeto do discurso, portanto não estabilizadas. Passa-se, no campo da ciência, progressivamente do plano sintático para o semântico (11): as relações novas, nesse sentido, ganham estabilidade e passam a constituir as relações semânticas (GARDIN, 1973a, p.144-145; 1973b, p.75-76).

As representações obtidas através de uma ML distinguem-se daquelas obtidas por métodos de “tabulação” que, para Gardin, não constituem representações propriamente ditas porque são o resultado de escolhas realizadas diretamente sobre os textos (GARDIN, 1973b, p.68, 70). Ao

(11) “Uma dada relação entre dois ou mais conceitos pode ser expressa tanto analiticamente, através de um dispositivo semântico de algum tipo (por ex., classificação, fatoração, composição etc.) ou sinteticamente, através da geração de uma construção em ML (compare-se, por exemplo, a relação semântica inerente ao termo composto “inseticidas”, ou na estrutura analítica ‘droga - usada em - insetos’, independentemente de sua codificação, com a relação sintática envolvida numa construção em ML como em ‘Drogas (Instrumento), Insetos (Paciente), ou ‘R instrumento/paciente (Droga, Inseto)...’ (GARDIN, 1973a, p.145).

operar diretamente sobre textos em LN sem redução semântica, esbarra-se no problema da língua (idem ibidem), a menos que se postule a existência de uma linguagem semântica universal, e conseqüentemente, características universais, “o que não deixa de ser uma ficção interessante sob o plano especulativo” (GARDIN, 1970, p.633).

A ML utilizada para representação constitui, para Gardin, “(...) o conjunto dos símbolos através dos quais serão expressas as equivalências e as diferenciações introduzidas na fraseologia de um corpus LN dado, para que apareçam as aproximações desejadas entre as palavras, grupos de palavras, proposições etc” (GARDIN, 1970, p.633).

Tal ML é a LD (*information retrieval language*), que é construída no momento da AD: “o jogo próprio da análise documentária engendra, à medida que se desenvolve, a constituição de um sistema de símbolos necessários para nomear, de uma maneira ou outra, os produtos mesmos do jogo, sob o plano lexical e sintático” (GARDIN, 1970, p.634). Essa íntima ligação determina, em larga medida, a forma e o conteúdo dessas análises (idem ibidem, p.636).

Para Gardin, o conteúdo e a estrutura das MLs são, via de regra, inferidas dos textos em LN. Desta forma, MLs diferentes apresentam semelhanças quando relacionadas a um mesmo universo do discurso, o mesmo acontecendo com os resultados obtidos pela AD sobre os mesmos documentos feitas por analistas diferentes (GARDIN, 1973a, p.149).

A concepção de Gardin sobre representação documentária privilegia o jogo de operações lógico-semânticas envolvidas para sua expressão. Para o autor, “(...) toda análise documentária implica o emprego de uma metalinguagem” (GARDIN, 1973b, p.75) e “(...) a definição mesma de análise documentária implica uma referência metalingüística” (idem ibidem). A substituição de textos científicos em LN por paráfrases, “(...) supõe a existência de uma “metalinguagem autônoma” onde são definidos os elementos constitutivos da paráfrase(...)” (GARDIN, 1970, p.632).

Vê-se, portanto, que para Gardin a atividade de AD é uma atividade metalingüística. A “autonomia” dessa linguagem relativamente à LN, configura-se numa LD propriamente dita e justifica seu estatuto distinto das metalinguagens definidas pela Lingüística. Não fosse assim,

como classificá-la enquanto língua artificial cujo plano de conteúdo é uma linguagem semelhante à sua linguagem objeto? (Ver nota 6).

Em relação aos procedimentos para a conversão LN-ML (Linguagem Natural-Metalinguagem), Gardin, contrariamente às pesquisas lingüísticas levadas a efeito à época (anos 60, escola chomskiana), prioriza os aspectos semânticos em detrimento dos sintáticos, passando em seguida, a considerá-los de forma mais integrada (GARDIN, 1973a, p.150). Para o autor, a distinção entre sintaxe e semântica é apenas contingente (12), o que implica que as ferramentas relacionais das LDs devem ser definidas independentemente dessa distinção (GARDIN, 1973b, p.76).

De fato, o uso da gramática chomskiana não apresentou, segundo o autor, resultados práticos para a Documentação. Entretanto, sua rejeição, consubstanciada nas teorias lingüísticas que se seguiram (nos Estados Unidos e na URSS principalmente), revelaram novas convergências de objetivos entre a Documentação e a Lingüística (GARDIN, 1973b, p.73-74).

Gardin destaca, sob esse aspecto, os estudos de Katz e Fodor levados a efeito em 1963, que mostraram a impossibilidade de resolver certas dificuldades sem fazer apelo a dados semânticos. (Enquanto o uso de algoritmos gramaticais implicava resultados ambíguos na interpretação de palavras dos textos, os “marcadores semânticos” combinados com “regras de seleção” permitiam a quebra da ambigüidade) (GARDIN, 1973b, p.78).

O autor ressalta, também, as inúmeras pesquisas que destacam a importância do aparelho semântico na análise Lingüística, como as “árvores semânticas” de Postal, em 1971, o “dicionário *combinatoire-raisonné* da língua russa”, de Mel’Chuk em 1970 e as explorações realizadas por vários teóricos sobre as “redes semânticas” (GARDIN, 1973b, p.79-80).

Para Gardin, tais estudos revelaram muita semelhança com as propostas feitas quinze anos antes para a construção de LDs: para a “(...)passagem de um texto natural à sua representação

(12) “Sobre o plano semântico são consideradas as relações que são sempre válidas em um certo domínio de estudo, ou seja, as relações mais banais, aquelas que, nesse domínio, são tidas como adquiridas — o fato, por exemplo, que o cavalo é um animal ... Sobre o plano sintático, ao contrário, aparecem sobretudo as relações novas, aquelas que fazem o objeto de um discurso porque não são estabelecidas no mesmo grau” (GARDIN, 1973b, p.76)

metalingüística (...)” os documentalistas “(...) construíram redes semânticas *ad hoc*, onde se pode ver uma pré-figuração das redes semânticas da Lingüística atual” (GARDIN, 1973b, p.79). É o caso, por exemplo, do SYNTOL - Syntagmatic Organization Language, desenvolvido pelo autor e sua equipe.

Para o autor, o mesmo pode ser dito em relação ao funcionamento de tais redes semânticas: as LDs permitem que, a partir de uma palavra, encontre-se informação que caracteriza determinado conceito e define sua forma de utilização (GARDIN, 1973b, p.80), como é o caso do dicionário de Mel’Chuk.

A importância das redes semânticas de referência, para Gardin, está na forma de estruturação do enunciado: deixa-se de fazer uso de categorias gramaticais (substantivos, adjetivos, verbos etc.) substituindo-as por categorias cuja natureza é mais lógica do que lingüística, conforme, por exemplo, as categorias sugeridas por Grimes: Estado, Processo, Ação (GARDIN, 1973a, p.164, nota 24). O mesmo acontece com as funções gramaticais (definidas entre sujeito/verbo, verbo/complemento etc.), que cedem lugar às funções lógicas, definidas, por exemplo, a partir dos “casos” de Fillmore (GARDIN, 1973b, p.81-82).

Procedimentos semelhantes aos utilizados pelos lingüistas também são apontados por Gardin no campo da AD, quinze anos antes: no primeiro caso, a utilização de categorias como “Entidade, Processo, Ação” utilizadas pelo SYNTOL; no segundo, a noção de “função” (*role*) e os “operadores de função” para a expressão das relações sintáticas (GARDIN, 1973b, p.81-82). Podemos acrescentar, ainda nesta perspectiva, as categorias PMEST (Personalidade, Matéria, Energia, Espaço e Tempo) utilizadas por Ranganathan em sua Colon Classification na década de 20 e sua re colocação, trinta anos mais tarde, pelo Classification Research Group.

Outro ponto de concordância identificado por Gardin diz respeito à ordem relativa da semântica e da sintaxe no desenvolvimento da análise: segundo Postal, citado pelo autor, “a análise de textos, a análise Lingüística, deve começar a um nível notadamente anterior às estruturas profundas” (POSTAL apud GARDIN, 1973b, p.81), um nível pré-lexical (formacional, cognitivo, para outros autores), o que para Gardin coloca-se no plano metalingüístico remetendo não às palavras,

mas às noções metalingüísticas (idem *ibidem*).

Outras explorações na Lingüística — “estruturas relacionais de base” de Bever, as “grades mentais”, de Wilks, segundo Gardin — podem ser reencontradas sob diversas denominações na AD, sob a forma de quadros lógicos compreendendo as noções de agente, instrumento, objeto, modo, localização no espaço e no tempo, etc. Para ele, entretanto, tais correspondências não se tornaram mais claras com os estudos da Lingüística (GARDIN, 1973b, p.83).

Gardin acrescenta ainda que o reconhecimento da importância das estruturas semânticas de referência e da utilidade das gramáticas funcionais podem ser identificadas na sua concepção de ML para a AD: há “identidade formal entre as relações semânticas definidas sob o primeiro plano e as relações sintáticas sob o segundo” (GARDIN, 1973b, p.83).

O uso de categorias lógicas e de estruturas funcionais, por sua vez, remete à noção de “pressuposição”: para os lingüistas soviéticos, por exemplo, a organização do dicionário articulada com regras de combinação permitiriam aos locutores selecionar a interpretação de interesse; para Katz e Fodor, os marcadores semânticos e as regras de seleção; para Fillmore, os casos direcionam a seleção para a interpretação; para Postal, organizações pré-lexicais determinam a organização dos enunciados etc. Ou seja, embora as categorias utilizadas para a formulação das proposições variem de autor para autor — “estruturas funcionais” (*role structures*), “casos” (*case frames*), “estruturas relacionais de base”, “modelos mentais” (*mental templates*) — todas elas são aplicáveis segundo o mundo particular considerado. Para Gardin, todas essas noções são de caráter metalingüístico (GARDIN, 1973a, p.156-157; 1973b, p.84-85).

Pode-se prever, entretanto, segundo Gardin, conjuntos de funções “universais” para léxicos relativos a determinados “universos” de discurso: Agente, Paciente, Instrumento, Objetivo etc. Exemplo disso na Documentação são, para o autor, os limitadores funcionais de campo (*field-bound roles*) e seus operadores, conforme utilizados pelas Classificações Facetadas (GARDIN, 1973a, p.156-158).

A visão de Gardin, sob este aspecto, contempla a necessidade de incorporação do contexto para trabalhar com a significação na Documentação. As “pressuposições” constituem, na verdade, os

elementos necessários à incorporação do fator contextual à análise das relações de significação.

Entretanto, para Gardin, as “pressuposições” são elementos “impuros” dentro da Lingüística, o que o leva a questionar se eles ainda poderiam ser chamados de mecanismos “lingüísticos” (GARDIN, 1973b, p.84-85). Essa constatação reforça sua posição em relação à pouca contribuição da Lingüística, entendendo que tais mecanismos são derivados da reintrodução da lógica simbólica. Citando Lakoff, Gardin afirma que “as deduções da lógica natural são uma primeira aproximação para a representação formal do processo de pensamento” e “que a gramaticalidade deve ser definida relativamente às suposições sobre contextos situacionais e processos de pensamento” e, conseqüentemente, “o estudo da Lingüística e o estudo da lógica natural devem estar juntos” (LAKOFF, apud GARDIN, 1973a, p.158).

OSYNTOL - Syntagmatic Organization Language, desenvolvido pela equipe de Gardin no início dos anos 60 representa, nesse sentido, uma tentativa de assimilação de referenciais lógicos. Segundo o autor, dois princípios fundamentais nortearam o projeto:

1. A inversão das prioridades que dominavam a teoria Lingüística nos anos 60, com ênfase, na fase inicial do projeto, à aquisição e manipulação de dados semânticos, recorrendo-se à sintaxe apenas para resolver algumas ambigüidades. Num segundo momento, semântica e sintaxe passaram a ser utilizadas mescladamente, sem priorizar uma ou outra.
2. Como o objetivo era construir instrumentos analíticos de natureza razoavelmente geral para dar conta das transformações LN-ML como ocorriam na prática, não se adotou nenhuma posição dogmática para procurar o algoritmo de análise, nem mesmo tentou-se seguir a escola gerativo-transformacional em voga à época. Da mesma maneira, o autor julgou dispensável tomar partido nas discussões sobre as fronteiras entre o conhecimento lingüístico e conhecimento enciclopédico revividas nos anos 60, afirmando que a AD, num contexto particular, já implicava necessariamente algum conhecimento sobre o universo do discurso dessa ciência. Decorrência desses princípios é a proposta do sintagma mínimo ‘ $R_i(x,y)$ ’ como formato relacional único para a representação da estrutura sintática e semântica (GARDIN, 1973a, p.150-151).

Recapitulando. Para Gardin, embora a Lingüística e a Documentação compartilhem os objetivos de “isolar o sentido dos textos” (GARDIN, 1973b, p.67), a primeira não fornece subsídios

significativos para a segunda. Nos métodos de indexação por tabulação, a Lingüística não auxilia na definição de mecanismos de seleção e segmentação dos textos para a geração de índices. Da mesma maneira, a Lingüística não é um recurso para a construção das “metalinguagens” no âmbito da AD, quer sob o aspecto da definição dos descritores — estes são reduzidos a uma mesma base lingüística — quer para a construção das sintaxes que deverão indicar as relações lógicas entre os descritores para a representação de um dado documento (GARDIN, 1973b, p.68-72).

Sob nosso ponto de vista, a discussão colocada por Gardin relativamente à Lingüística deve-se à sua concepção de “Lingüística”, uma vez que, atualmente, agrega-se ao estudo da língua, o estudo da fala, do discurso e das inúmeras relações da língua com o indivíduo em sociedade, fatos que o autor reconhece e integra à sua pesquisa. Trata-se, pois, de um problema de abrangência do conceito empregado.

Após 1970 as pesquisas de Gardin dirigem-se principalmente à análise de textos científicos com o objetivo de resgatar os raciocínios que sustentam as construções nas Ciências Humanas, particularmente na Arqueologia. Esse tipo de análise — que ele denomina “logicista” — tem como objetivo esquematizar, através de um sistema simbólico artificial, os processos de raciocínio (dedutivo ou indutivo) sob a forma de proposições lógicas que vão da descrição à interpretação dos dados.

Mais exatamente, Gardin analisa os textos científicos como “construções”, cuja função é a de “estabelecer o ponto entre as observações empíricas, de um lado (...) e as hipóteses teóricas de outro (...)” (GARDIN, 1987a, p.185). Esse ponto é o raciocínio seguido para realizar a ligação entre as observações empíricas e as hipóteses teóricas (empírico-indutivo ou hipotético-dedutivo). O “cálculo” realizado para a reconstituição dos raciocínios assume a forma de uma seqüência de operações de reescrita, indo de $\{Po\}$ — dados de base ou proposições descritivas, para $\{Pn\}$ — proposições terminais no vértice da construção, intermediadas por $\{Pi\}$ — proposições intermediárias ligadas por operações de inferência empírico-dedutivas ou hipotético-dedutivas, conforme o ponto de partida da análise (idem *ibidem*).

A partir desse momento Gardin se afasta da AD propriamente dita. Para o autor, “a interpretação

de textos específicos numa específica direção ou veio, coloca em jogo operações específicas da LN para a ML que não podem, a priori, ser transferidas de um contexto de aplicação a outro (GARDIN, 1989, p.12). Podemos inferir dessa colocação uma das razões da alteração de rumo de sua pesquisa, que passa, agora, a concentrar-se no universo restrito do discurso arqueológico, buscando identificar as formas de raciocínio presentes na passagem dos dados de base à interpretação nesse universo específico que é sua especialidade.

Resta-nos, enquanto documentalistas, retomar apenas as discussões sobre AD e LDs, uma vez que a recuperação de raciocínios de áreas específicas constitui, segundo nossa opinião, tarefa de especialistas da própria área.

Do ponto de vista lingüístico, a Documentação pode e deve recuperar as reflexões desenvolvidas a partir da década de 70, para verificar a possibilidade de aperfeiçoamento de seus mecanismos de análise e representação textual. A manifestação de paralelismo de procedimentos entre as teorias lingüísticas e a AD identificados por Gardin justificam, por si só, a continuidade da exploração nessa linha. Como seqüência das propostas de integração entre sintaxe e semântica, deve-se pesquisar, portanto, a integração da pragmática, postulada por exemplo, por Umberto Eco.

2.1.2 A linha de pesquisa em Análise Documentária no Brasil

De um modo geral, a construção da representação está intimamente ligada à análise e síntese de textos. É essa a abordagem — com variações para a delimitação das etapas que a compõem — quase que unânime adotada por integrantes da atual linha de pesquisa em AD no Brasil.

Todavia, assim como os trabalhos de Gardin (salvo aqueles sobre o SYNTOL), a questão da representação é enfocada quase que exclusivamente na dependência da análise: via de regra, vincula-se a análise, síntese e representação à construção de LDs.

Analisando a questão SMIT estabelece um paralelo entre as definições de Metalinguagem Científica (Hjelmslev) e as de AD e LD (Coyaud, Gardin) (SMIT, 1973, p.15-18), a partir da noção

de que uma metalinguagem científica “é um sistema construído, onde o plano de conteúdo é constituído por um outro sistema, aquele da significação. Os termos empregados para descrever o sistema de significação constituem um corpo de definições coerente, isto é, eles foram previamente estudados e estruturados em seguida” (SMIT, 1973, p.15-16). A autora discute essa noção (levantando os problemas relativos à relação de homologia ou analogia, validação e conotação), perguntando-se, em seguida, se as LDs podem operar como MLs e se elas cumpririam as tarefas de uma metalinguagem. (idem *ibidem*).

A resposta a essa questão é remetida às leis e características do discurso científico, como “matéria-prima” da Documentação. Segundo a autora, a partir da condensação dos documentos, através do processo de denominação — similar à codificação compreensiva das mensagens discursivas apontadas por Greimas — é lícito identificar o “(...) movimento oscilatório entre a expansão e a condensação, a definição e a denominação” (GREIMAS, 1976, p.101) e o “(...) vai e vem entre a denominação e a definição” da atividade científica e da Documentação. (SMIT, 1973, p.22). Este paralelo lhe permite afirmar que na Documentação “a equivalência entre grupos de lexemas pode existir e ser reconhecida” (idem *ibidem*). A operação denominativa, portanto, é reconhecida como central na Documentação, podendo tomar como ponto de partida as diferentes unidades do discurso, “do tema à frase, passando pelo motivo, o enunciado ou as palavras”, sendo que essa escolha determina o sucesso ou fracasso de uma LD (idem *ibidem*, p.23-24).

Cabe destacar, entretanto, que a autora admite, ao discutir a “(...) ligação que une a linguagem utilizada à sua metalinguagem (...)”, ou o “(...) modelo dessa análise (...)” — tomando a definição de modelo de Kristeva segundo a qual eles seriam “(...) sistemas formais cuja estrutura é isomorfa ou análoga à estrutura de um outro sistema (do estudado) (KRISTEVA, 1969, apud SMIT, p.16) que a ML não valida, em nenhum caso, a análise feita.”

Em trabalho posterior, Smit utiliza um modelo de análise proposto por Gardin, onde a representação é vista como o resultado de duas operações básicas: a “Descrição” e a “Ordenação de Dados”. Da “Descrição”, passa-se à “Interpretação”, com base na identificação de relações lógico-semânticas existentes entre os conceitos, operação que viabiliza a construção progressiva de “metalinguagens” de tradução (GARDIN, 1974). Nessa perspectiva, a autora analisa textos

interpretativos de vários autores sobre o poema “Les Chats”, de Baudelaire, identificando, nos mesmos, Traços Descritivos (unidades portadoras de informação) e Argumentos, que são posteriormente ordenados com base em mecanismos de organização de subconjuntos conceituais, para validar a interpretação (SMIT, 1976).

Com base nesses trabalhos, CUNHA desenvolve um modelo de AD a partir da identificação das etapas de “Análise” (compreendendo leitura e segmentação para a identificação de informações principais) e a “Síntese” (identificação de conceitos/palavra-chave, tradução LN-LD e estruturação do vocabulário) (CUNHA, 1990).

Como o trabalho de SMIT, a pesquisa de CUNHA é centrada na vinculação da representação à fase de análise.

Relativamente à etapa de análise, o modelo de CUNHA foi testado por KOBASHI, que nele identificou problemas relativos à junção de conceitos provenientes de princípios epistemológicos diferentes (KOBASHI, 1988, p.71-73). A etapa de análise desse modelo é novamente explorada por AMARO, incorporando reflexões sobre a teoria do discurso, em particular a noção de macroproposição semântica de Van Dijk. Embora AMARO aponte, em sua conclusão, um esquema básico para a AD composto de duas etapas — uma de “análise e síntese” e outra de “representação” (o que se aproxima da segmentação que adotamos) — sua proposta de trabalho concentrou-se na discussão da primeira (AMARO, 1991).

A questão da representação documentária propriamente dita (entendida como resultado da intermediação de códigos documentários) é abordada por CUNHA na etapa de “síntese”. Nessa etapa, realiza-se a tradução do texto em LN para uma LD. Retomando a formulação de Gardin e outros (GARDIN, LAGRANGE et alii, 1987c), a autora afirma que “no momento da AD existem três linguagens (...) interligadas: a linguagem natural (LN), as linguagens especializadas (LE) e

(13) Sem entrar na discussão sobre divisão ciências exatas/ciências humanas, deve-se lembrar que a medida da maior ou menor sistematização conceitual depende do parâmetro que se adote como referência. É conveniente lembrar que “o vocabulário compartilhado (...) não anula a existência, nem da diversidade teórica ... nem da ambigüidade conceitual, pois o fato de se usar os mesmos conceitos não significa que se lhes esteja atribuindo o mesmo significado” (LOPES, 1990, P.37).

as linguagens documentárias (LD)” (CUNHA, 1991, p.60), que se articulam “de forma diferente nas ciências exatas ou nas ciências humanas” (idem *ibidem*) (13), face o maior grau de sistematização dessas ciências. Essa problemática, segundo a autora, resulta “numa nova atribuição à AD em ciências humanas (...) que é, ao mesmo tempo, hipótese e metodologia de procedimento (...)” (idem *ibidem*, p.61).

Para CUNHA, é tarefa da AD “reconstruir a conotação ou conceptualização feita pelo autor e atribuir-lhe conceitos/palavras-chave tradutoras da informação conceptualizada ou conotada (...)” organizando, em seguida, os termos considerados significativos através da “formulação de categorias operacionais” e da “formação do campo semântico” do texto (CUNHA, 1990, p.112-113).

A partir da identificação de conteúdos a serem traduzidos em “conceitos/palavras-chave” — obtidos a partir de seu modelo de análise — a autora se propõe a desmontar o “(...) bom senso’ bibliotecário na estruturação do vocabulário e na formação do campo semântico”, bem como a seleccionar “(...) os conceitos a serem utilizados no vocabulário e na formação do campo semântico (CUNHA, 1990, p.110-111).

O método desenvolvido por CUNHA compreende basicamente quatro fases. Para a formulação de categorias operacionais — 1a. fase — a autora utiliza as noções de Saussure, distribuindo os conceitos nos eixos paradigmático e sintagmático, tendo como objetivo a construção de “blocos aglutinadores” com a função de “ (...) categorias representativas dos diferentes aspectos conceptuais ou conotativos de formulação do texto” (CUNHA, 1990, p.113). CUNHA utiliza-se dos estudos de “casos” de Fillmore e Pottier, adaptando-os em função dos objetivos de organização de categorias aglutinadoras do vocabulário (idem *ibidem*, p.114-118).

Para a formação do campo semântico — 2a. fase — o ponto de partida adotado são as “ (...) associações mnemônicas e aleatórias”, de caráter mais “intuitivo e etnocêntrico do que científico”, incluindo-se aí aspectos ideológicos (CUNHA, p.119). A autora sugere a incorporação de “procedimentos lógico-lingüísticos sistematizados”, ou seja, a distribuição dos conceitos candidatos a palavras-chave nos eixos paradigmático e sintagmático, considerando suas categorias (definidas na primeira fase) e a identificação daqueles pertencentes ao mesmo nível operacional.

O campo semântico, por-tanto, deverá englobá-los, mantendo seus traços distintivos independentemente das especificidades atribuíveis aos conceitos. A caracterização dos traços distintivos é feita utilizando os atributos de Pottier (Ação, Não-Ação, Descrição, Função, Virtualização), considerando-se também, o conhecimento de área do analista e os usuários possíveis dessa informação, além do uso de “definições standard”, que promoveriam uma “plataforma comum de trabalho” e atenderiam à “neutralidade” aconselhada pelos manuais de construção de vocabulário (CUNHA, 1990, p.122-128).

A 3a. fase —o controle do vocabulário — se inicia a partir da ordenação alfabética dos “conceitos-frase e dos conceitos-termo” levantados do texto, seguida de sua ordenação “em função do campo semântico e das categorias previamente estabelecidas” (CUNHA, 1990, p.129-131).

Por fim, a 4a. fase é reservada à discussão dos conceitos: segundo o “bom senso” (14), são considerados os aspectos relativos à especificidade dos descritores, à resolução de ambigüidades através do uso de notas de alcance, a determinação dos diferentes usos do conceito, lembrando-se que sua fixação sofre interferência do posicionamento ideológico do analista (CUNHA, 1990, p.128-130).

Sinteticamente, essa é a abordagem de CUNHA relativamente à construção do vocabulário. Sua proposta, entretanto, apresenta dois problemas básicos:

- 1) A complexidade do modelo, que vincula a construção e a organização do vocabulário ao processo de representação individual dos documentos;
- 2) O uso sistemático de procedimentos de “bom senso”, ao lado da assimilação de instrumental de teorias semânticas de naturezas diversas.

Revela-se, dessa maneira, a ausência de uma discussão sobre o conceito de representação subjacente. Na falta de procedimentos metodológicos para a elaboração das LDs e sua utilização, o vocabulário se constrói a partir da própria prática de indexar.

(14) O “bom senso” é subjetivo, individual, variando segundo a experiência.

Desse modo, face à complexidade da proposta para a formulação de categorias e formação de campos semânticos a partir da combinação de instrumental lingüístico e bom senso; em razão da própria necessidade apontada pela autora de utilizar o conhecimento de área; e considerando a inviabilidade econômica colocada pelo desencadeamento de uma análise desse tipo a cada texto, o modelo pode ser questionado quanto à sua eficácia e aplicabilidade.

2.2 A contribuição da Semiótica: o conceito de enciclopédia de Umberto Eco

Para Peirce, “representar” significa “estar em lugar de, isto é, estar numa relação com um outro que, para certos propósitos, é considerado por alguma mente como se fosse o outro” (PEIRCE, 1977, p.61).

Na verdade, a comunicação se efetiva no momento da apropriação. Pressupõe, de um lado, o objeto que se quer representar; do outro, um sujeito que deve interpretar essa relação. Embora Peirce não o afirme, podemos dizer, portanto, que a comunicação está implícita nessa concepção de representação (15).

A representação documentária deve ser abordada de forma semelhante. Como produto documentário, ela deve provocar uma interpretação (16) e algum tipo de relação com o texto que lhe deu origem. Uma representação documentária desencadeia, conseqüentemente, uma situação de comunicação.

A abordagem da representação, em Eco, se apóia na concepção de signo de Peirce — “um signo, ou representamen, é aquilo que, sob certo aspecto ou modo, representa algo para alguém” (PEIRCE, 1977, p.46) — e se revela, como decorrência, no seu conceito de texto. Para o autor,

(15) Para COELHO NETTO, a teoria de Peirce é teoria da comunicação, uma vez que ela se apresenta como “uma produção de signos para serem interpretados” (COELHO NETTO, 1990, p.213). Esta afirmação se fortalece quando se considera a forma dialógica de apresentação do pensamento, conforme observação de J. Ransdell citada pelo autor, como também a partir da proposição peirceana da relação triádica entre objeto, signo e interpretante, derivada da noção de relação entre enunciador, enunciação e intérprete (idem *ibidem*).

(16) Pode-se afirmar que, segundo Peirce, a interpretação se dá através da proposição de uma cadeia de interpretantes frente a uma relação Sujeito-Objeto.

“um texto não é um sistema de significação, mas uma realização comunicativa” (ECO, 1984, p.15).

Para Eco, é necessário realizar uma distinção entre sistema de significação e processo de comunicação. Materialmente, o processo de comunicação é aquele onde um emissor envia uma mensagem através de um canal e essa mensagem chega ao destinatário, através da intermediação de um código comum aos dois interlocutores. O código, por sua vez, é um sistema de significação.

Pode-se construir sistemas de significação independentemente da possibilidade de eles serem utilizados em processos de comunicação. Mas não pode existir o processo de comunicação sem que exista um sistema de significação. Existe uma relação de pressuposição, não recíproca, entre comunicação e significação.

Um sistema de significação tem um plano de expressão e um plano de conteúdo (no sentido que Hjelmslev confere a esses conceitos, cf. HJELMSLEV, 1975, p.13-64). A relação entre os dois planos é baseada em “postulados do significado”, cuja validade se dá a partir de juízos emitidos em termos de condições de verdade, dentro de “mundos possíveis”, ou mundos de referência (ECO, 1984, p.12-13).

O “enfrentamento” de um texto, como uma realização comunicativa que se utiliza e atualiza sistemas de significação exige, para Eco, que se considere o elemento pragmático, isto é, as circunstâncias em que o texto foi emitido, a posição do enunciador etc. “(...) o conceito de texto não pode (...) ser reservado apenas a uma análise em termos de sistema de significação, pois se trata já de um problema de comunicação” (idem *ibidem*). E sob esta perspectiva, o texto passa a ser uma mensagem de significação.

Ao não restringir a análise textual aos aspectos semânticos — que dão conta do significado de uma expressão — Eco propõe, portanto, que se incorpore à interpretação, os elementos integrantes de um processo de comunicação: o emissor, o destinatário, o contexto e as circunstâncias de comunicação da mensagem (ECO, 1984, p.16-17).

De forma semelhante, o produto de uma análise de texto em Documentação — a representação documentária — não tem um fim em si mesmo. Esse gênero de representação também constitui uma questão de comunicação. Nesse sentido, o enfoque da atividade global de AD — e da representação que constitui o seu produto — não pode se resumir apenas às operações de análise, síntese e intermediação de códigos: ela deve ser pensada considerando os aspectos relativos ao jogo de significações que envolve quando de sua inserção num processo geral de comunicação.

Como vimos, os documentos científicos constituem a matéria-prima da Documentação. Seu mundo de referência é o da Ciência e os sistemas de significação que a constituem, seu objeto. Ao trabalhar com documentos científicos, a AD opera com textos que utilizam e atualizam os sistemas de significação próprios, específicos da área. Mais à frente, ao elaborar “representações” de documentos, a Documentação se utiliza e cria sistemas de significação que lhe são próprios: suas LDs. A finalidade de todos esses processos é a de gerar produtos documentários para determinados tipos de usuário e em função de objetivos institucionais.

Dito de outra maneira, a Análise e a Representação documentárias devem realizar a interface entre a documentação original e o usuário, valendo-se freqüentemente de sistemas intermediários que constituem novos sistemas de significação: as LDs. O sucesso da Comunicação Documentária está na habilidade da Documentação para levar a efeito o processo como um todo, considerando todas as variáveis intervenientes no seu desenvolvimento, a fim de possibilitar que o destinatário (o usuário, receptor), ao procurar por uma informação, possa entender a mensagem gerada pelo emissor-documentalista.

A despeito dos objetivos diferentes da Análise Textual enfocada por Eco — não é a análise em si que interessa à Documentação, mas a identificação de informações significativas — permanece a necessidade de se considerar os aspectos comunicacionais. A incorporação da pragmática, portanto, para a abordagem das questões documentárias que são nosso objeto, é essencial.

A noção de pragmática em Eco, tem origem na concepção de pragmaticismo de Peirce (que faz questão de distingui-la do pragmatismo para evitar a banalização, cf. PEIRCE, 1977, p.286 e ss.): “uma concepção que deve ser comprovada através de seus efeitos práticos” (PEIRCE, 1977, p.290).

Para COELHO NETTO, o pragmaticismo de Peirce insere o intérprete (sujeito) na prática semiótica — mesmo que Peirce não o diga explicitamente (COELHO NETTO, 1990, p.90). (Peirce fala do homem e das condutas humanas frente às experiências). Afirma o autor que “a significação (...) nunca poderá ser dada por uma cadeia de signos que validem a proposição enunciada mesmo na ausência do Intérprete. Exatamente porque só existe diante desse Intérprete, o significado (...)” de um texto “(...) não pode ser apreendido como se tal texto constituísse uma entidade fechada” (COELHO NETTO, 1990, p.91).

É nesse sentido que a pragmática, para Eco, coloca a necessidade de incorporar, à interpretação textual, contextos e circunstâncias de enunciação.

Por contexto, entendemos a inserção concreta de termos nos textos ou discursos. Para Dubois et alii, “dada uma unidade, ou uma seqüência de unidades A, o contexto é constituído pelas unidades ou seqüências de unidades que precedem ou seguem A e que podem, de uma forma ou outra, fazer pesar sobre A certas coerções (DUBOIS et alii, 1973, p.149). O contexto se refere, portanto, a termos pertencentes a um mesmo campo semiótico. A inserção dos termos num contexto direciona a interpretação textual.

Por outro lado, as circunstâncias de enunciação dizem respeito aos elementos de outros códigos semióticos que devem estar presentes no momento de interpretação textual. Referem-se, por exemplo, às condições de produção do discurso (no sentido de Pêcheux, cf. PÊCHEUX, 1969) e ao contexto situacional ou social de inserção do texto (no caso da AD, as variáveis que intervêm no seu processo). A configuração das circunstâncias de enunciação condicionam a interpretação a ser dada a um texto.

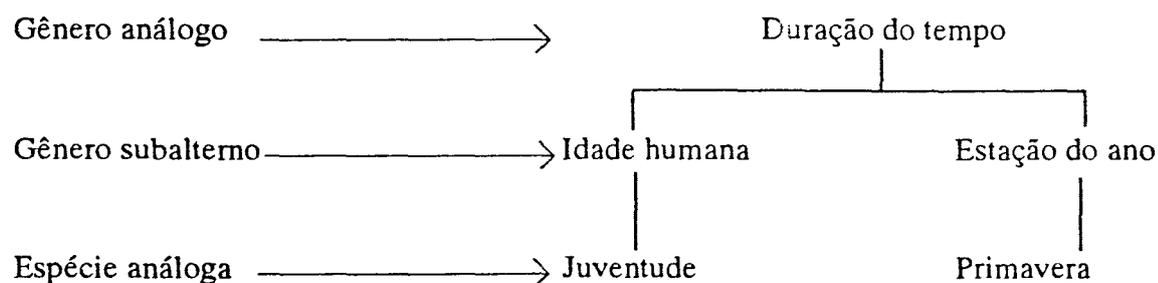
A AD opera a partir do texto. Supõe-se então que, da mesma maneira, devem ser considerados, para isso, contextos e circunstâncias de enunciação. Se, entretanto, o “enfrentamento” do texto nos processos de AD é realizado com objetivos diferentes — não é a análise em si que interessa, mas a identificação de informações significativas — sua inserção num contexto comunicacional permanece.

O problema da comunicação — incorporado aqui à discussão da representação documentária — remete, por sua vez, à noção de enciclopédia.

A necessidade de um modelo enciclopédico de análise foi imperativo para a interpretação da “metáfora”. Já no séc. XVII, Emanuele Tesauro afirmava que a construção de uma metáfora era um “exercício sobre a cultura”. Reportando-se ao autor, Eco observa que “(...) é preciso que se faça uma imensa colheita do já-dito, do intertextual, e em seguida organizar um índice de categorias com o uso de fichas e tabelas (...)” que configuraria um “modelo de universo semântico organizado” (ECO, 1984, p.78).

O índice de Tesauro tomava as categorias aristotélicas — substância, quantidade, qualidade etc. — reunindo, sob cada categoria, todas as coisas subjacentes. A demonstração da construção da metáfora da primavera como juventude da vida, por exemplo, é realizada a partir da consulta ao índice de categorias, buscando “um gênero análogo que é “duração do tempo”, um gênero subalterno que é “idade humana vs. estação do ano”, espécie análoga “juventude, “primavera” (ECO, 1984, p.78-79). Após todos esses percursos no interior da rede enciclopédica, Eco reproduz o seguinte esquema:

QUADRO 1



(ECO, 1984, p.79)

Eco destaca que a análise da parte construída de seu índice mostra que a noção subjacente à sua construção não é o inventário de propriedades e similaridade das coisas, mas predicados, ou

“propriedades culturais (...) de interpretantes” (ECO, 1984, p.79). Essa constatação leva-o a investigar a possibilidade de construir um modelo semântico enciclopédico feito através de interpretantes, a partir da constatação de que o uso de modelos semânticos dicionariais só permite trabalhar com “um pacote restrito de propriedades dicionariais”, que são “marcas fixas”. Os modelos enciclopédicos permitiriam trabalhar com propriedades que variam segundo diversas seleções contextuais.

O autor reconhece, entretanto, que a enciclopédia é um postulado semiótico, e que deve permanecer como um postulado dada a impossibilidade de se descrever todas as interpretações possíveis de um texto. A enciclopédia é, na verdade, uma “hipótese reguladora” com base na qual (...) o destinatário decide construir uma porção de enciclopédia concreta que lhe permita reconhecer (...) uma série de competências semânticas” (ECO, 1991, p.113).

O desenvolvimento da proposta de Eco relativamente à possibilidade de construção de um modelo semântico enciclopédico está assentado em sua concepção de signo. Eco sustenta que “todo signo é um texto virtual — ou uma virtualidade de textos — e que o texto é a expansão do signo, inscrito num universo sígnico em forma de enciclopédia” (PAIS, in ECO, 1984, ix). Essa posição articula as propostas de Peirce e os modelos de Hjelmslev e representa, portanto, a confluência de reflexões epistemológicas sobre as ciências da significação e da linguagem (idem *ibidem*).

Ao adotarmos esse referencial para pensarmos as questões colocadas pela análise e representação documentária, buscamos referendar a maneira como concebemos os processos envolvidos na construção e representação da informação documentária. O processo global de AD se dá na estrita dependência de inúmeras variáveis intervenientes (a estrutura textual, a instituição, o analista etc.), e dessa forma, não há como pensá-lo restringindo-nos a uma abordagem exclusivamente semântica e/ou sintática.

Por outro lado, uma vez que a questão da AD remete à abordagem do significado, tanto em relação à identificação do que se considera informação, no sentido documentário, como à sua representação num contexto documentário mais amplo, buscamos em Eco o referencial relativo à questão da significação, cujo embasamento remete a Peirce.

Em comentários à obra de Lady Welby (*What is meaning?* 1903), Peirce discute o problema da significação, através de um exemplo. Para ele, a possibilidade de interpretação de uma sentença como “Napoleão é uma criatura letárgica” depende do que ele denomina experiência colateral, ou seja: se a pessoa que interpreta essa sentença nunca ouviu falar em Napoleão, a sentença não significará, para ela, mais do que uma pessoa ou coisa que era chamada Napoleão e que era letárgico. Neste caso, a palavra Napoleão não pode determinar a mente dessa pessoa a identificar Napoleão com uma série de atributos. Não se estabeleceu nessa pessoa, um “hábito” pela qual essa palavra possa trazer à tona uma variedade de atributos de Napoleão, o homem. Peirce afirma que “Quase tudo isso é verdadeiro em relação a qualquer signo. Nessa sentença, além do Objeto “Napoleão”, temos um outro “Objeto Parcial” que é a Letargia: e a sentença não pode veicular seu significado a menos que a experiência colateral tenha ensinado a seu intérprete o que é Letargia, ou o que ‘letargia’ significa nessa sentença” (PEIRCE, 1977, p.160-161). Para Peirce, experiência ou observação colateral é “o pré-requisito para se obter qualquer idéia significada pelo signo (...). Não se trata de uma familiaridade com o sistema de signos, mas uma “prévia familiaridade com aquilo que o signo denota” (PEIRCE, 1977, p.161).

A experiência colateral no processo documentário é evidenciada, por exemplo, nas diferenças de desempenho profissional de analistas com especialização nas suas áreas de trabalho. O resultado de operações como a indexação, por exemplo, podem ficar comprometidas, a não ser que o “hábito” já tenha se responsabilizado pelo fornecimento de um quadro de referência mínimo para o reconhecimento de determinadas remessas (tipologia textual, referencial institucional, sistemas de significação das ciências etc.).

Portanto, a questão da significação não se resolve apenas do ponto de vista de sua decodificação semântica e, a não ser que a concepção de semântica inclua o elemento pragmático — caso de algumas orientações da semântica textual (Eco, Rastier) — deve-se pesquisar, para a delimitação do significado de uma palavra, além de sua inserção contextual, elementos de outras naturezas (que no nosso caso deverão incluir informações sobre a situação de inserção do ato documentário, pertinentes).

Uma vez que concebemos a AD como uma série de processos que se orientam a partir do texto

e de uma série de informações de caráter extra-textual, a abordagem da significação a partir da inclusão da pragmática é necessária.

Exploramos, portanto, a teoria da significação adotada por Eco para embasar epistemologicamente nossa abordagem, e a noção de enciclopédia semântica decorrente como instrumental teórico para analisar a questão da análise e representação documentárias.

Devemos ressaltar que o paralelo que estabelecemos entre a interpretação da metáfora e a construção da representação documentária (que supõe uma interpretação) deve-se à função da “relação de interação entre as palavras” (DUCROT & TODOROV, 1988, p.253) pressuposta pela metáfora e presente no processo de representação (onde se realiza um jogo complexo e delicado de substituição). A metáfora é uma figura lingüística através da qual se emprega “(...) uma palavra num sentido que se assemelha e no entanto difere de seu sentido habitual” (DUCROT & TODOROV, 1988, p.254). É o “(...) surgimento numa determinada cadeia significante de um significante vindo de outra cadeia, transpondo este significante, a barra (‘resistente’) do algoritmo para perturbar, com sua ‘disrupção’, o significado da primeira cadeia, onde ela produz um efeito de sentido (...)” (DUCROT & TODOROV, 1988, p.316). No processo documentário, a substituição considera núcleos semânticos similares e a combinação e a interação entre palavras institui novos significados.

Nesse sentido, a utilização do conceito de enciclopédia e a investigação sobre a possibilidade de construção de um modelo semântico enciclopédico, pode ser utilizado tanto para compreender o processo de AD como para investigar os requisitos necessários à construção de LDs.

Destacam-se, para a investigação desses problemas, as contribuições de ECO relativas à abordagem do significado que o modelo adota (que interessa particularmente na fase de análise) e aquelas que dizem respeito às formas de organização das propriedades dos termos (noção útil para a discussão dos instrumentos de comutação documentária).

Para Eco, “a organização de propriedades é do tipo empírico”, o que dificulta a adoção de definições de dicionário que implicam, necessariamente a aceitação das relações de entailment ou

implicação (ECO, 1984, p.50). Ou seja, não há como representar o conhecimento empírico a partir de uma árvore de Porfírio; “(...) não existe uma única árvore de Porfírio” (ECO, 1984, p.50-53), e uma definição dicionararial não é potente para dar conta do significado. As árvores desse tipo são uma construção, o que implica a necessidade de considerar o contexto onde elas se inserem (Idem *ibidem*).

De fato, a implicação lógica (ou implicação lógica) é base dos principais sistemas de classificação bibliográfica, onde prevalecem as relações aristotélicas do tipo gênero/espécie, todo/parte. Prova de que tais relações não dão conta do mundo empírico é que esses sistemas incorporam — às vezes caoticamente — os recursos de enumeração (segundo o princípio de indução de Bacon).

Deve-se destacar, também, a proposta da semântica componencial que permite descrever hierarquias de propriedades que variam conforme as diferenças colocadas em questão, considerando o contexto e as circunstâncias onde o texto se insere. As propriedades de um termo são atribuições dadas institucionalmente, revelando seu caráter cultural.

A noção de enciclopédia semântica, portanto, pode embasar os procedimentos de AD. Propomos a explorá-la mais particularmente face à representação documentária via intermediação de códigos normalizadores. Interessa-nos investigar em que medida a concepção subjacente à organização de universos semânticos, a partir de propriedades dadas culturalmente, pode conferir maior consistência à organização de inventários para representação documentária.

Para efeito deste trabalho, realizaremos, também, uma aproximação da noção de propriedades dicionariais — que considera uma série restrita de postulados de significado e remetem apenas ao conhecimento lingüístico (ECO, 1984, p.43) — à noção de predicado livre postulada por Le Guern, que se refere às palavras na língua: possuem um significado mas não uma referência, exprimindo, dessa forma, apenas um conjunto de propriedades (LE GUERN, 1989). À noção de propriedades enciclopédicas, aproximaremos a de predicado relacionado de Le Guern, referindo-se à palavra no discurso, tendo portanto, um referente e designando um objeto concreto. A determinação da porção da enciclopédia a ser acionada para a compreensão de uma palavra remete a determinados contextos e circunstâncias.

Diremos, com Eco, que a noção de enciclopédia será adotada como uma “hipótese reguladora” que deverá permitir a recuperação de determinadas competências semânticas (ECO, 1991, p.113), ou seja, ela encaminhará o analista — e em seguida o usuário da informação — aos interpretantes que definem as significações num texto e numa dada área de conhecimento. Dito de outra maneira, a hipótese reguladora propõe um quadro referencial teórico e conceitual determinado, capaz de se constituir no parâmetro regulador para a interpretação.

Para Eco, a noção de enciclopédia semântica coloca em evidência que a reconstrução para o registro de propriedades é dada com base em categorias textuais, a partir da “narcotização” e “ênfase” dos semas a serem considerados ou deixados de lado.

Na operacionalização da análise e síntese de textos científicos, o quadro referencial e teórico — a enciclopédia — e a estrutura textual — base para a definição do documento — orientam e determinam a construção da informação documentária. São esses parâmetros que determinam a “narcotização” e a “ênfase” no processo documentário.

Na verdade, “(...) todos os semas são enciclopédicos; os semas dicionariais são uma espécie de ato sucessivo metalingüístico que praticamos, quando construímos ad hoc a árvore de Porfírio que nos interessa (ECO, 1984, p.93-94). Uma boa representação enciclopédica é pontencialmente infinita, “ (...) já é um texto in nuce, um texto incoativo, um texto pronto para expandir-se (...)”. É um “ (...) mecanismo que prescreve quais propriedades na representação enciclopédica dos termos que o compõem devem ser narcotizadas e quais enfatizadas” de modo a se poderem dar amálgamas e, assim, estabelecer níveis de sentido ou isotopias no interior do texto” (idem ibidem, p.96).

Antes do texto, os termos não têm sentido ou têm todos os sentidos virtuais: a atribuição de sentido é dada no texto. Estes contêm dispositivos que orientam o intérprete a realizar operações “sobre a reserva enciclopédica enquanto repertório de textos virtuais” (Idem ibidem).

A identificação de indicadores para a elaboração da síntese documentária é direcionada pelo texto, mas dependente de informações extra-textuais, o que faz ressaltar também, o caráter seletivo das operações de identificação da informação documentária, em função do contexto situacional

onde se insere o processo.

Além disso, como a representação documentária acentua o processo de generalização (perdem-se os vínculos com o documento original), a construção de LDs deve prever o uso dos predicados relacionados. Assim, abre-se a possibilidade da identificação de propriedades de objetos concretos bem como da remissão à porção da enciclopédia pertinente para sua interpretação. Dito de outra maneira, as linguagens de representação textual devem orientar-se no sentido de possibilitar a recuperação de propriedades que evoquem “interpretantes” os quais, por sua vez, remetem a determinada organização de significados, e não só a possíveis organizações lexicais.

Uma linguagem assim construída possibilita a geração de conhecimento a partir da reutilização da informação — nova leitura da informação já produzida — através do engendramento de novos tipos de relações. As propriedades dicionariais — que remetem apenas ao conhecimento lingüístico e não incluem o empírico — só permitem a reprodução da informação já postulada e cristalizada. Esta constitui, portanto, a contribuição central de Eco que adotaremos para a discussão da representação documentária, que retomaremos em capítulo específico.

A questão da representação, por sua vez, supõe que o sentido dos textos remete a contextos e circunstâncias determinados e que as LDs devem viabilizar a conversão dos textos em LN para uma linguagem normalizada. Para isso, os procedimentos de conversão devem garantir que a representação seja feita a partir de “marcas” do contexto situacional presentes nos textos. O parâmetro dessas “marcas é parte da enciclopédia (o quadro referencial) que regula a interpretação.

Em resumo, é possível pensar, ao menos teoricamente, um instrumento comutador que permita conferir às expressões de representação documentária o caráter de “signos documentários”. Uma LD desse tipo seria, portanto, de natureza semiológica. Nesse instrumento, a representação do significado deveria remeter ao mundo material.

As LDs tradicionais, entretanto, nem sempre realizam essa remissão, uma vez que se baseiam em termos preferenciais ou são utilizadas como nomenclaturas, possibilitando apenas o estabelecimento de equivalências: a ausência de definições implica, não raras vezes, que uma determinada

palavra encontrada num texto seja convertida por um descritor similar encontrado na LD. A representação resultante não é de significação, baseando-se exclusivamente na equivalência lexical. Do ponto de vista metodológico, tais LDs revelam uma visão estritamente lingüística da LD. Mesmo admitindo processos de inferência de significado dadas através do arranjo gráfico das expressões no instrumento comutador (por exemplo, TG - Termo Geral, TE - Termo Específico, TR - Termo Relacionado etc), da ausência de explicitação podem derivar procedimentos baseados na identificação lexical.

Dadas as amplas possibilidades virtuais contidas em cada palavra integrada ao léxico documentário, a significação só será possível a partir da identificação da porção da enciclopédia necessária à recuperação de determinadas competências interpretativas.

É nesse sentido, também, que analisamos a proposta de uma linguagem “L” em formato de enciclopédia. A postulação de uma linguagem desse tipo é feita com base numa teoria do significado que integra semântica e pragmática. Por exemplo, para dar conta do que Gardin chama de caráter não-lingüístico do problema da representação do conhecimento (embora acreditemos que, antes de tudo, este seja um problema relacionado aos limites de atuação e à concepção de Lingüística), recorreremos a Eco, que afirma, com Bar-Hillel, que a pragmática “é o estudo da dependência essencial em comunicação, em Linguagem Natural, do falante e do ouvinte, do contexto lingüístico e do contexto extralingüístico” e da “disponibilidade dos conhecimentos fundamentais, da prontidão em obter esse conhecimento fundamental e da boa vontade dos participantes do ato de comunicação” (BAR-HILLEL, 1968, apud ECO, 1991, p.75).

Uma linguagem “L” deve conter, entre as próprias regras de significação, instruções pragmáticas orientadas (ECO, 1991, p.76-77). Ela pode fornecer o significado contextual — que vai muito além dos significados lexicais — se fornecer: “a) significados lexicais em forma de instrução para a inserção contextual e b) roteiros” (idem *ibidem*, p.78-79).

CAPÍTULO 3

A ATIVIDADE GLOBAL DE ANÁLISE DOCUMENTÁRIA

Os textos veiculam informações de diversas naturezas. Para a Documentação, tradicionalmente, eles constituem os documentos, que são a matéria prima da AD. O objetivo genérico da AD é extrair, desses documentos, a informação documentária propriamente dita, para, em seguida, representá-la através de códigos próprios, de modo a recuperá-la convenientemente. O objetivo último das operações de AD é o de realizar a disseminação da informação documentária.

A AD é, portanto, uma modalidade de Análise de Textos, e para entender sua especificidade é preciso compará-la com a Análise dos Discursos.

Durante muito tempo, confundiram-se as noções de texto, discurso e documento e as análises supostamente documentárias realizaram-se a partir de aparatos rigorosos que, apesar disso, não contemplavam o discurso como manifestação da estrutura documentária de produção e representação da informação. Desse modo, quanto mais a AD se aproximava das diferentes modalidades de Análise do Discurso, mais evidente se tornava a precariedade das metodologias para as análises específicas de tratamento da informação.

Para a Análise do Discurso, o discurso é um objeto concreto e se propõe invariavelmente como expressão da subjetividade, isto é, apropriação da língua por um “Eu” definido (17). Analisá-lo implica, portanto, extrapolar os aspectos que alguns teóricos consideram puramente lingüísticos do texto.

Normalmente, a Análise do Discurso propõe-se como um empreendimento multidisciplinar que busca, na pesquisa das condições de produção — jogo de imagens entre os interlocutores, tempo

(17) Na linguagem se dá a possibilidade da subjetividade, já que ela “contém todas as formas lingüísticas apropriadas para a sua expressão, e o discurso provoca a emergência da subjetividade, porque ele consiste de instâncias discretas. A linguagem propõe (...) formas “vazias” que cada locutor, no exercício do discurso, se apropria e que ele liga à sua “pessoa”, definindo, ao mesmo tempo, ele mesmo como eu e um parceiro como tu. A instância do discurso é assim constitutiva de todas as coordenadas que definem o sujeito (...)” (BENVENISTE, 1966, p.263).

e espaço da enunciação, expressão da intencionalidade etc. — o seu modo de articulação específico. Assim sendo, cada discurso é único, e a análise tem como objeto discursos individuais ou corpus bastante restritos. Procura-se, portanto, através da Análise do Discurso, o sentido, como forma de apropriação individual de um sistema de significação, que é a língua.

O texto, por sua vez, é um objeto material tomado fora do contexto de sua produção, passível, portanto, de ser analisado em si mesmo. Define-se “pela sua autonomia e por seu fechamento (...) constitui um sistema que não se deve identificar com o sistema lingüístico mas em relação a ele: relação ao mesmo tempo de contigüidade e semelhança” (DUCROT & TODOROV, 1988, p.268).

No texto se manifestam as estruturas informacionais, ou seja, os elementos que respondem por sua coesão e progressão. Assim, o ponto de partida da AD é o texto, já que, ao operar com procedimentos de desestruturação, estruturação e representação documentária, o documentalista persegue uma estrutura bastante semelhante. Além disso, elementos presentes nas condições de produção — que remetem, portanto, ao discurso — são codificados segundo procedimentos documentários diferenciados: lugar, tempo e autor de uma obra, por exemplo, definem a representação descritiva e não temática de um texto. Mais uma vez se evidencia o fato de que as variáveis do discurso não se confundem com as variáveis temáticas e que essas últimas só podem ser isoladas a partir da estrutura informacional que é própria do texto. Enquanto no discurso se expressa o sentido — apropriação individual da significação articulada pelo sistema (BENVENISTE, 1974, p.262, v.2) — no texto manifesta-se a significação como veículo de informação.

Sob o ponto de vista da AD, as unidades selecionadas do documento, denominadas informação, constituem formas e não substâncias, no sentido hjelmsleviano. A estrutura do documento apresenta-se como articulação de formas (18), o que torna possível a identificação e a análise da

(18) O conteúdo de um pensamento “ (...) recebe forma quando ele é enunciado e somente assim. Ele recebe forma de língua e na língua, que é a mola de toda a expressão possível (...)” (BENVENISTE, 1966, p.64). A língua “é organizada como agenciamento de ‘signos’ distintos e distintivos, suscetíveis eles mesmos de se decompor em unidades inferiores ou de se agrupar em unidades complexas. Essa grade estrutural, que enforma as estruturas menores e de vários níveis, dá sua forma ao conteúdo do pensamento. Para chegar a ser transmissível esse conteúdo deve ser distribuído entre morfemas de certas classes, agenciados numa certa ordem etc. Em resumo, esse conteúdo deve passar pela língua e dela emprestar seus quadros (...) A forma lingüística é, então, não somente condição de transmissão, mas primeiramente, condição de realização do pensamento (idem *ibidem*).

informação, qualquer que seja o suporte físico em que se manifesta. (Um documento científico, por exemplo, desenvolve-se, via de regra, através da apresentação de objetivos, metodologia, análise e conclusões).

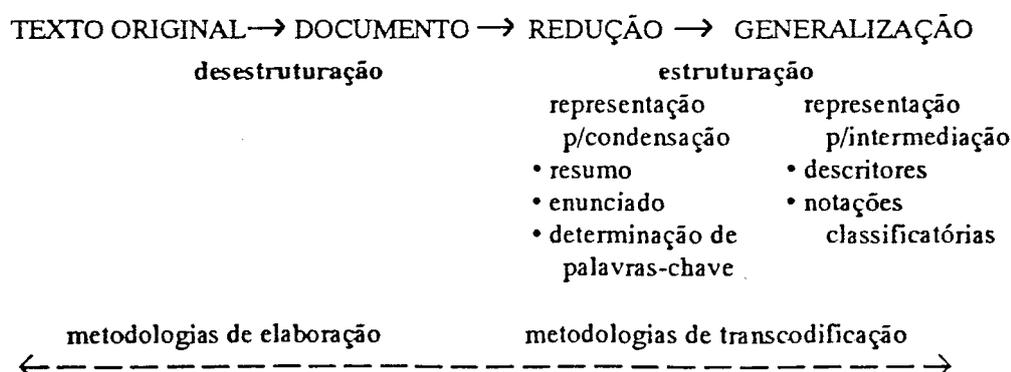
Uma vez identificada, a estrutura informacional do texto é submetida a várias operações documentárias — análise e condensação (resumos, enunciados, palavras-chave), representação, utilizando-se como instrumento comutador a LD — que a transforma em mensagens de natureza diferente daquela que a originou. Por isso se afirma que a AD tem como objetivo recuperar e disseminar informação e não textos, discursos ou documentos — suportes materiais da informação. É preciso entender, também, que não se recupera qualquer tipo de informação (a expressiva, por exemplo) mas apenas informação documentária. Esta se apresenta como produto específico obtido através de procedimentos documentários a partir da informação original — presente no texto — operada por instrumentos documentários.

É necessário ainda observar que os procedimentos documentários caracterizam-se por operar uma redução crescente: do texto para o resumo, do resumo para o enunciado, do enunciado para a unidade de tradução via código documentário. A atividade de AD caracteriza-se, portanto, como uma sucessão de processos de transformação do texto original, observando-se, a cada etapa, graus crescentes de generalização. Neste sentido, alguns procedimentos documentários podem ser considerados como uma modalidade específica de tradução.

A natureza metodológica da AD deve contemplar, conseqüentemente, a diferenciação entre as etapas e seus produtos, evidenciando procedimentos e níveis de representação da informação.

A partir do exposto, pode-se esquematizar o processo global de AD do seguinte modo:

QUADRO 2 - PROCESSO DE ANÁLISE DOCUMENTÁRIA



Deve-se considerar, ainda, que o processo global de AD é condicionado por determinadas variáveis, endógenas e exógenas, a saber:

- a) a estrutura do texto original;
- b) a informação bruta presente neste mesmo texto;
- c) o estado de sistematização metodológica e terminológica da área em questão;
- d) a instituição na qual o processo se desenvolve;
- e) o usuário da informação documentária;
- f) o estoque de conhecimento anterior do documentalista e sua formação ideológica.

A estrutura textual reflete o arranjo utilizado pelo autor para a apresentação das informações. Tais arranjos apresentam-se como “esquemas formais de organização de textos” a partir dos quais pode-se identificar “traços básicos” que os caracterizam. O reconhecimento da tipologia dessas organizações pode facilitar o trabalho documentário: a identificação de superestruturas textuais (19) pode conduzir, com maior eficácia, à identificação das partes mais significativas de um texto (KOBASHI, 1989, p.33). O reconhecimento da estrutura do texto científico, por exemplo, permite identificar, com maior clareza, as informações mais significativas do documento, para a divulgação da informação documentária.

As informações brutas são aquelas encontradas nas partes mais significativas do texto e constituem os elementos referenciais que os integram. São as informações de base da mensagem (VANOYE, 1991).

O estado de sistematização das terminologias de área dificulta ou facilita o processo de AD. Os discursos das ciências onde a necessidade de comunicação obrigou a busca de precisão terminológica facilitam, conseqüentemente, o trabalho documentário.

(19) As superestruturas constituem as formas convencionais do texto. São estruturas hierárquicas de categorias, como por exemplo, introdução, desenvolvimento, conclusão. Segundo Van Dijk, “muitos tipos de discurso parecem exibir uma estrutura esquemática convencional e, conseqüentemente, variável de acordo com a cultura, uma forma global que organiza a macroproposição (o conteúdo global do texto). Assim sendo, atribui-se usualmente às estórias um esquema narrativo, que consiste em uma estrutura hierárquica de categorias convencionais tais como situação, complicação e resolução” (VAN DIJK, 1992, p.30). O autor denomina esses esquemas de superestruturas do texto. “Uma superestrutura fornece a sintaxe completa para o significado global (...)” (idem *ibidem*).

A instituição onde o processo de AD se desenvolve, ou seja, o contexto geral em que se insere, estabelece o usuário, determinando seu perfil. Em seguida, direciona e condiciona a seleção de informações determinando retenções ou descartes, e dá formato aos produtos documentários segundo objetivos específicos.

Finalmente, a formação e o quadro de referência do documentalista, incluindo sua ideologia, determinam todos os processos de construção do texto documentário e sua representação.

A interferência desses fatores no processo global de AD concorre para a construção de um produto não isento e neutro.

QUADRO 3 - VARIÁVEIS ENDÓGENAS E EXÓGENAS PRESENTES NO PROCESSO DE ANÁLISE DOCUMENTÁRIA

VARIÁVEIS	ANÁLISE E SÍNTESE	CONVERSÃO
Estrutura textual	████████████████████	
Informações brutas	████████████████████	
Instituição	██	
Usuário	██	
Perfil do documentalista	██	
Sistematização da área	██	

Nos capítulos subseqüentes procuraremos observar o desenvolvimento dos processos de AD com maior detalhe, objetivando distinguir os diferentes níveis de representação da informação e as peculiaridades da representação via LDs.

É necessário ressaltar, entretanto, que este trabalho volta-se à AD enquanto área de atividade, cujo objeto são os textos técnico-científicos. Nos textos representativos dessa área, as relações de

significação pressupõem uma univocidade de significação, ou seja, um corpo conceitual proposto por consenso (ou discussão, quando da quebra de paradigmas), conhecimento, produção, rearranjos, segundo correntes institucionais de pensamento e posição: ou seja, um referente e um contexto. Desse modo, pretende-se tomar o signo monossêmico como base para a transmissão da informação documentária.

CAPÍTULO 4

A ANÁLISE E SÍNTESE DE TEXTOS COMO PROCESSOS DE ANÁLISE DOCUMENTÁRIA

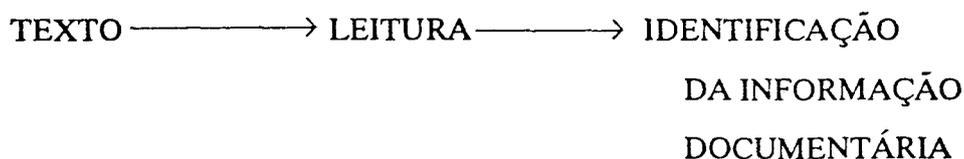
Neste capítulo, procuraremos explicitar as características relativas ao processo de análise e síntese documentárias levadas a efeito a partir da leitura de um texto selecionado, considerando os condicionamentos particulares a que tais processos estão sujeitos.

Para melhor abordar a questão, a segmentação que adotamos privilegia dois momentos: a identificação da informação documentária e o registro dessa mesma informação. Desse modo, pretendemos observar a seqüencialidade do processo e focar os indicativos individualizantes inerentes a cada etapa, supostamente necessários a uma sistematização posterior.

4.1 A identificação da informação documentária:

O processo de identificação da informação documentária envolve operações de análise e síntese documentárias. O objetivo dessa fase é o de retirar do texto original os principais indicadores de conteúdo para construir o “texto documentário” — síntese concreta — a partir de operações de leitura e seleção de informações. O texto documentário nada mais é do que o produto da representação por condensação, ou o “resumo”. Pode-se submeter esse produto, em seguida, à intermediação de um código normalizador.

QUADRO 4



4.1.1 A leitura documentária

A leitura documentária é uma modalidade específica de leitura. Para caracterizá-la, é necessário analisar o processo global de leitura.

4.1.1.1 Características do processo de leitura global

Na perspectiva de Umberto Eco, um texto é “um jogo de estratégia” que fornece uma série de instruções diretivas para que se delineie uma imagem de “Leitor Modelo” (ECO, 1984, p.99). Entretanto, ele é também, simultaneamente, uma “máquina preguiçosa”, que não realiza todo o trabalho que deveria, deixando ao leitor a tarefa de preencher os buracos dos não-ditos (Idem *ibidem*, p.97). A esse leitor, portanto, compete uma parte importante na construção do próprio texto.

Para Eco, além de ser um aparato de comunicação, um texto questiona os sistemas de significações preexistentes a ele, freqüentemente os renova, às vezes os destrói” (ECO, 1991, p.31). A manifestação textual, nesse sentido, pode alterar e reconstruir “funções sígnicas pré-existentes” (idem *ibidem*).

O processo de leitura é, a seu modo, um processo de comunicação leitor-texto e um momento para a tomada de decisões: o leitor negocia com o texto, a fim de preencher os “não-ditos” — o que significa também, descartar alguns “ditos” — e, assim, estabelecer uma interpretação particular do texto em questão. Compete, portanto, a esse leitor-modelo genérico, construir o processo de interpretação, munido de uma enciclopédia particular, que lhe permite executar a tarefa de preencher os vazios deixados pelo autor, bem como descartar alguns ditos. É a enciclopédia, portanto, a responsável pela efetivação do processo de comunicação. Caso o leitor não disponha de uma enciclopédia apropriada para a leitura de um texto, ele não disporá das chaves para a interpretação textual. A comunicação será, nesse sentido, incompleta.

Para Grice, a forma ideal de comunicação supõe uma espécie de cooperação entre os interlocutores.

O “princípio de cooperação” reporta-se a regras dispostas sob quatro categorias: quantidade (relativa ao volume de informação que deve ser fornecida pelo locutor ao seu ouvinte); qualidade (relativa à veracidade das informações); relação (que diz respeito ao propósito da informação) e modalidade (que se refere à forma pela qual se apresenta o dito, incluindo clareza, ausência de obscuridade, ausência de ambigüidade, brevidade e método). A eficiência da comunicação, portanto, é relativa à observação ou não dos interlocutores às máximas conversacionais (GRICE, apud TÁLAMO, 1982, p.54 e ss.).

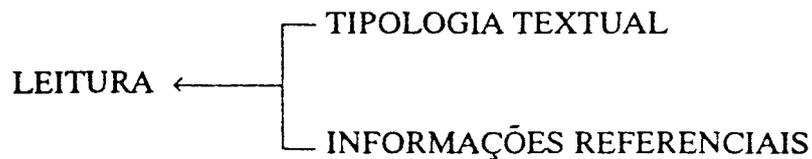
Pode ocorrer (e usualmente acontece), entretanto, que algumas máximas não sejam obedecidas, supondo-se a existência de informações implícitas na conversação. Essa implicitação ou implicação conversacional se dá sempre que houver infração de um princípio de cooperação (no caso da omissão da informação, a implicação não é conversacional já que não se pode determiná-la), ou na suposição de que, mesmo desrespeitando uma das máximas, o locutor “dá a entender”, ou faz “subentender” alguma informação em sua mensagem. Nestes dois últimos casos têm-se, respectivamente, uma intenção particular de comunicar (a implicação conversacional é pública e determinável) e uma intenção que, embora não seja explícita no nível conversacional, o é ao nível da comunicação implícita (TÁLAMO, 1982, p.54-60).

Por outro lado, a interpretação da mensagem, na perspectiva de Vanoye, é dada a partir da identificação de elementos referenciais integrantes dos textos (para nós, técnico-científicos) (VANOYE, 1991, p.68). Como mensagens escritas, os textos caracterizam-se por exercer a função referencial. Devem portanto, apresentar um conjunto de elementos — palavras — estruturados segundo um princípio de organização, que responde pelo ponto de vista ou enfoque através do qual o autor propõe e situa o tema de sua comunicação. Os elementos referenciais (ou as palavras referenciais) “definem as informações de base da mensagem — informações enumeráveis, sintetizáveis, ‘brutas’ ou ‘objetivas’”, enquanto que os “outros elementos definem o trabalho de organização dessas informações em função do propósito almejado pelo autor da mensagem” (idem *ibidem*, p.68). Desse modo, a identificação dos termos (palavras) referenciais permite a reconstituição da informação bruta veiculada pelo texto (VANOYE, 1991, p.68-69).

A leitura, na perspectiva comunicacional, envolve, portanto, tanto o reconhecimento da tipologia

textual, como a identificação dos elementos referenciais para uma interpretação apropriada.

QUADRO 5



De fato, para Verón, a questão da determinação da significação não se reduz à investigação do conteúdo de uma mensagem. A significação surge de uma comparação, pois nenhum texto se esgota em si mesmo. “Não se analisa jamais um texto: analisam-se pelo menos dois, quer se trate de um segundo texto escolhido explicitamente para a comparação, quer se trate de um texto implícito virtual, introduzido pelo analista, muitas vezes sem que ele o saiba” (VERÓN, 1980, p.14). Nesta perspectiva, a ideologia se manifesta como “um sistema de relações entre o textual e o extratextual” (idem ibidem, p.14). Como bem o assinala Edward Lopes, no prefácio à edição brasileira de “A produção do sentido”, de E. Verón, “a pertinência de uma leitura, enquanto prática produtora de um discurso, depende das suas condições de produção, do relacionamento, portanto, do texto com fatores extratextuais que deixaram vestígios no próprio discurso (...)” (idem ibidem, p.5).

Já para Fêcheux, o discurso deve ser considerado como fenômeno da língua em funcionamento. Para ele, a lingüística de Saussure, ao definir seu objeto — a língua — exclui a fala, condicionando a abordagem da língua a partir dos seus elementos invariantes. Entretanto, ao defini-la como instituição social, Saussure abre a possibilidade de expandir o campo da investigação dos fenômenos lingüísticos. É possível, a partir da incorporação dos elementos da teoria sociológica, considerar nos estudos lingüísticos as variantes, ou seja, a língua em funcionamento. Na verdade, a fala, embora não constitua um fenômeno puramente lingüístico, deve ser considerada como um

mecanismo pertencente ao funcionamento da língua. Tal mecanismo pode ser definido face às “condições de produção”: “os fenômenos lingüísticos de dimensão superior à frase podem (...) ser concebidos como um funcionamento, não integralmente lingüístico (...), que só pode ser definido em referência ao mecanismo de localização dos protagonistas e do objeto do discurso (...)” (PÊCHEUX, 1969, p.16). Inclui-se aqui o referente, o contexto, a situação na qual se insere o discurso (PÊCHEUX, 1969, p.9-23).

A partir dos conceitos expostos, é possível promover uma aproximação entre as noções de “estratégia de negociação leitor-texto”, “princípio de cooperação”, “identificação de elementos referenciais a partir de informações brutas” e “condições de produção”, com o objetivo de imprimir uma interpretação globalizante das variáveis intervenientes no processo de leitura. Os “ditos” e “não-ditos” de Eco, por exemplo, poderão ser identificados ou preenchidos a partir dos resultados advindos da obediência ou não às máximas conversacionais de Grice, incluindo as determinações “dadas a entender” ou “dadas como subentendidas”, que constituem instruções existentes implicitamente no texto e que solicitam o acionamento de enciclopédias dos interlocutores. Tais instruções, presentes na estrutura organizacional do texto, auxiliam, por outro lado, a identificação dos “elementos referenciais”, de Vanoye. As condições de produção do texto (Pêcheux), por sua vez, remetem às enciclopédias, ou quadros de referência do autor-emissor e do receptor-leitor, incluindo aí os referenciais ideológicos.

4.1.1.2 Características da leitura para fins documentários

Para a Documentação, os textos são considerados documentos (como o são as imagens registradas em suportes físicos específicos), a partir, justamente, da perspectiva referencial que assumimos (cf. Vanoye). A forma documental é o registro necessário para que determinados conhecimentos ou reflexões tenham garantidas sua inserção no sistema de comunicação social.

As operações de leitura documentária propriamente dita objetivam a identificação e extração de referenciais dos textos originais, para sua transformação em texto documentário. Adotando a perspectiva de Vanoye, diremos que os principais referenciais para essa construção são as

“informações brutas” contextualizadas por determinada organização textual, informações essas que funcionam, em Documentação, como os principais indicadores para a construção do texto documentário propriamente dito.

O leitor-documentalista, entretanto, não pode ser classificado como um leitor-modelo, conforme a conceituação de Eco. O leitor especialista (o leitor-modelo) procede à leitura a partir de uma enciclopédia que lhe permite avaliar, criticamente, os procedimentos metodológicos de construção do texto, bem como à questionar a validade de suas proposições enquanto contribuição à ciência. Pode, portanto, verificar a pertinência e a adequação do problema investigado, a validade das hipóteses levantadas, a adequação das conclusões obtidas e proceder a vinculação desse texto a um universo de pesquisas mais amplo. É um leitor que tem condições de fazer inferências a partir do texto.

Já o leitor-documentalista tem objetivos distintos. A ele não compete a crítica da ciência e a avaliação das contribuições veiculadas pelos textos (mesmo porque ele não dispõe, em geral, de enciclopédia que o autorize a proceder dessa forma). O objetivo do leitor-documentalista é o de localizar proposições veiculadas pelo texto. Sua estratégia de leitura deve voltar-se à identificação e extração de informações. A identificação de informações novas é, nesse sentido, problemática, uma vez que a porção da enciclopédia necessária para que tal tarefa se concretize só é acessível àqueles que tenham especialização na área ou que possuam um tipo de “especialização” proveniente da experiência cotidiana com o material bibliográfico sobre os assuntos em questão.

A partir dessas operações de identificação, o analista deve construir substitutos compactos de documentos de tipologia variada, para diversas finalidades: resumos, para uso direto, ou textos documentários para a operacionalização da tradução via código documentário.

A leitura do documentalista caracteriza-se, portanto, pela presença de operações seletivas. A análise dessa leitura particular deve considerar as variáveis que lhe são específicas, além daquelas que são características da leitura de um leitor comum (não profissional da informação) — instruções textuais, condições de produção, circunstâncias de enunciação, momento e lugar de enunciação etc.

O leitor-documentalista defronta-se com restrições impostas pela inserção do seu trabalho num espaço institucional, devendo desenvolver a leitura voltada à produção de construções documentárias pertinentes aos objetivos da entidade onde se encontra inserido e, ao mesmo tempo, adequadas às necessidades específicas de seus usuários. O contexto institucional funciona, portanto, como o grande parâmetro para a seleção de informações documentárias. É a partir da instituição, sua organização e objetivos, que se determinam os paradigmas de pertinência de informação. Ao se proceder a leitura, determinados objetivos são perseguidos quando da identificação e do descarte de informações presentes nos documentos.

Da mesma forma, a interferência de variáveis de ordem individual, ou seja, a ideologia do documentalista, sua visão de mundo e sua formação pessoal e profissional, sua enciclopédia particular — que pode ou não incluir o conhecimento de área — determinam a retenção ou descarte de informações.

Por sua vez, o usuário que, além de usuário da informação é um elemento integrado de algum modo à instituição e aos seus objetivos, também condiciona a seleção. É ele o beneficiário do processo de tratamento e recuperação da informação, e é em sua função que as ações documentárias são dirigidas.

Assim sendo, tornam-se pertinentes aquelas informações que tenham espaço garantido no contexto situacional onde se inserem: “(...) daquilo que permanece semanticamente incluso ou implícito, o leitor só explicita o que lhe serve. Ao fazer isto, ele magnifica algumas propriedades, ao passo que mantém outras sob narcose” (ECO, 1979, p.69). Desta forma, são “iluminadas” (selecionadas) aquelas informações brutas que sejam pertinentes ao documentalista e à sua instituição e que possam garantir a relação usuário/instituição e conteúdo selecionado.

A instituição condiciona, ainda, o nível de leitura que se fará de um texto: áreas especializadas — que produzem terminologias bastante específicas e diferenciadas — requerem leituras mais acuradas (muito embora apresentem, provavelmente, textos mais estruturados, com terminologia mais sistematizada que facilitam a tarefa); áreas mais abrangentes, leituras mais superficiais e generalizantes. O primeiro usuário é mais facilmente identificável — e com necessidades de

informação definidas — que o segundo usuário.

As operações de seleção de informação documentária refletem, por outro lado, a política de gerenciamento da informação, a qual implica a definição de universos de interesse específicos e a idealização de seus produtos, motivo pelo qual acaba por determinar os níveis de profundidade e especificidade de leitura e análise dos documentos.

A leitura documentária pressupõe, também, regras documentárias que lhe são específicas, que dizem respeito à normalização de seus produtos para uso documentário. A influência de todas essas variáveis reflete-se, conseqüentemente, na redução do texto original e construção do texto documentário.

Segundo Ana Maria Cintra, para a leitura documentária concorrem vários fatores: a qualidade do texto (nível de estruturação do texto), o conhecimento prévio e as estratégias de leitura. Excetuando-se o primeiro — identificado em função de parâmetros institucionais — os fatores restantes são diretamente dependentes da habilidade do profissional documentalista (CINTRA, 1987, p.28).

O conhecimento prévio se refere ao estoque de conhecimento armazenado na memória do leitor, que se configura como uma espécie de “quadro de referência, formado por uma rede multidimensional de unidades conceituais a partir da qual o input visual é avaliado” (CINTRA, 1987, p.30). Nesse estoque de conhecimento se incluem referências lingüísticas (estoque de palavras — vocabulário — frases, formas de organização textual, tipologias textuais) e entidades lingüísticas e conceituais.

As estratégias de leitura dizem respeito à forma pela qual o leitor realiza esse processo: a partir de estratégias cognitivas (processamento automático e inconsciente) e das estratégias metacognitivas (quando o leitor tem consciência do que está lendo). Os dois processos podem estar presentes no ato de leitura, sendo que a medida de legibilidade de um texto pode ser feita a partir do uso equilibrado dessas duas estratégias (CINTRA, 1987, p.32-33).

As estratégias cognitivas correspondem às atividades realizadas automaticamente, em função de esquemas prévios armazenados na memória. Elas “comandam a compreensão ortográfica, sintática e semântica do texto” (CINTRA, 1987, p.34). Já as estratégias metacognitivas referem-se ao “domínio consciente de algumas etapas, que permitem “identificar ‘fatias’ ou segmentos significativos do texto, estabelecer relações de sentido e referência entre certos segmentos, manter a coerência entre as macro-proposições semânticas do texto (...)” (idem *ibidem*).

Reportando-se a Grice, a autora afirma que “todo texto é produzido para determinados receptores e que a sua eficácia depende, em boa parte, da capacidade do autor em estabelecer com seus leitores uma relação cooperativa” (CINTRA, 1987, p.28). Ressalta, entretanto, que esse princípio é rompido pela leitura documentária, uma vez que o documentalista não foi previsto como leitor.

De fato, o leitor documentalista não pode ser considerado um leitor apto a interpretar quaisquer textos. Não sendo especialista, ele não tem uma enciclopédia (ou um quadro de referência) que lhe permita relacionar-se com o texto de forma a identificar integralmente sua organização e compreender todas as informações colocadas no nível implícito. Não tem, portanto, condições de identificar todos os ditos e, portanto, de preencher os não-ditos. O nível de interpretação de determinados textos por parte de documentalistas, desse modo, fica condicionado, embora não totalmente, à sua maior ou menor familiaridade com o vocabulário utilizado na área. Esse dado acrescenta mais um fator à complexidade do processo de leitura.

Cintra ressalta, ainda, que a leitura para fins documentários não é “neutra”, uma vez que, “sendo a linguagem intrinsecamente comprometida com o cultural e o ideológico, tanto o processo de produção do texto, quanto o de recepção não se isentam destes componentes. E é esse comprometimento que garante o caráter interativo da leitura” (CINTRA, 1987, p.29)

A questão da neutralidade também é apontada por Cunha, para quem a crença na “leitura única e absoluta”, por parte do bibliotecário ou documentalista, é um preconceito: o leitor documentalista, como indivíduo inserido no processo de produção, consumo e tradução de informações, superpõe sua visão ideológica à linguagem e ideologia dos textos que lê e analisa. Para a autora, é necessário considerar o sistema onde se insere a disseminação dos produtos e informações documentárias e,

conseqüentemente, considerar a “variável produtor-tradutor-consumidor da mesma”, sabendo-se que lhe é inerente, uma “linguagem/ideologia refletora do pensamento e da situação em que se encontram inseridos” (CUNHA, 1987, p.38-39).

Portanto, além da impossibilidade de uma leitura neutra dada pela inserção do documentalista como indivíduo na sociedade, deve-se considerar também que o seu papel é condicionado pelos imperativos ditados pela instituição à qual se vincula. A leitura que ele procede, portanto, tem, ao menos implicitamente, parâmetros institucionais, o que o remete a selecionar determinados “ditos” como pertinentes a objetivos institucionais. Isso não quer dizer, entretanto, que o documentalista não possa ter uma visão crítica dos objetivos institucionais. Ele pode, nesse sentido, imprimir critérios de seleção que dependem de parâmetros ideológicos particulares.

A partir dessas observações e de modo genérico, poderemos caracterizar o leitor-documentalista, conforme segue:

1. O perfil do leitor documentalista é particular: não é o leitor-modelo genérico postulado por Eco: ele não dispõe, necessariamente, da enciclopédia para interpretar as instruções textuais;
2. O documentalista não é previsto pelo autor como leitor (o princípio cooperativo não prevê o documentalista como interlocutor), ou seja, a enciclopédia do documentalista não é necessariamente a mesma da do especialista. A leitura do documentalista deve se desenvolver, portanto, com o auxílio de regras adicionais que permitam a “identificação” daqueles referenciais que normalmente possibilitam ao leitor previsto pelo texto, proceder à sua interpretação. (A identificação não implica necessariamente, a interpretação, mas a localização de ocorrências típicas e atípicas num dado universo discursivo. Experiências colaterais distintas, mas convergentes, determinam o estágio de leitura). As regras adicionais podem se constituir de informações sobre ocorrências típicas: quadro de tipologias textuais, organização dos textos e discurso dentro dessas tipologias, terminologia de área e quadros de sistematização conceitual. Se tais informações não permitem o estabelecimento do princípio cooperativo, não substituem as enciclopédias especializadas. Portanto funcionam, em todo caso, como indicadores para o reconhecimento de referenciais standard, ou seja, ocorrências tipológicas já identificadas e registradas pelo quadro de referências.

3. O leitor-documentalista pode estar capacitado a realizar adequadamente a leitura, a partir do uso equilibrado das estratégias cognitivas e metacognitivas. Essas estratégias, entretanto, também dependem de seu conhecimento prévio, e de seu quadro de referência básico. (Parte de sua enciclopédia deve ser acionada).

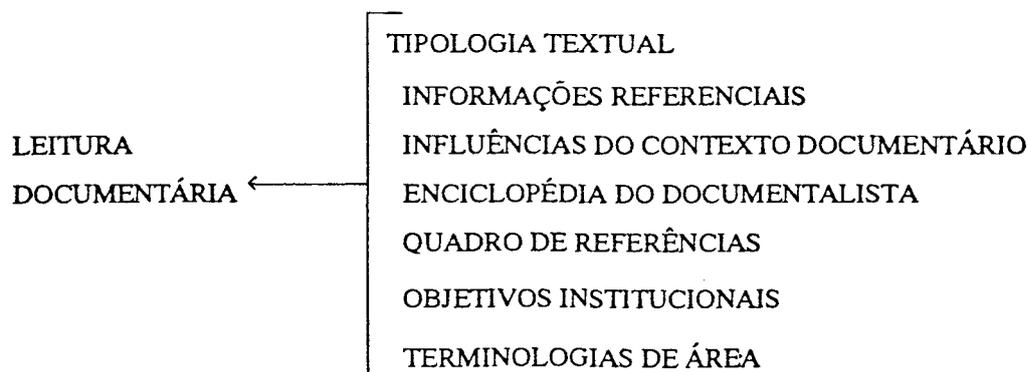
4. A interpretação dos textos supõe necessariamente a intervenção de componentes ideológicos: o leitor documentalista não é imune a essas intervenções.

5. As condições de produção condicionam a leitura: deve-se pensar os discursos em relação ao lugar de seus protagonistas e ao seu objeto. Parafraseando Pêcheux, as construções documentárias (os textos documentários) se fazem a partir de normas de análise que incluem a norma da instituição na qual estão inseridas, de forma a definir a posição de um conteúdo particular em relação a essa mesma norma. O documentalista, portanto, aciona (ou deve acionar) os dados de sua enciclopédia relativos aos imperativos institucionais e ao perfil de seus usuários.

A estratégia adequada para realizar a leitura documentária, portanto, deve ser traçada de forma a fornecer, ao documentalista — além de regras documentárias propriamente ditas — referenciais indiretos para a melhor identificação das informações brutas que permitam a reconstituição e interpretação de informações veiculadas pelo texto. Tais referenciais são:

- a) a posse de um quadro de referência que permita a identificação tipológica do texto;
- b) o conhecimento aprofundado dos objetivos institucionais e dos perfis de usuário;
- c) as terminologias de área, que constituem referenciais de enciclopédias necessárias à interpretação textual e discursiva.

QUADRO 6



As soluções encontradas pela prática recente têm colocado a especialização do documentalista como uma alternativa para a realização de uma leitura e uma redução adequadas. Observa-se, também que, em algumas bibliografias, os próprios autores têm elaborado resumo de seus trabalhos.

A primeira alternativa — a especialização do documentalista — nos parece inviável, se considerarmos a necessidade de conciliar a tarefa documentária — e seu volume — com a de atualização profissional. A segunda — o próprio autor elaborar o resumo de seus trabalhos — pressupõe predisposição por parte desses mesmos autores, além da necessidade de treinamento específico para a construção de textos documentários adequados. Ambas, portanto, esbarram em restrições de custo.

4.2 O registro da informação documentária: o texto documentário

Como vimos, a síntese documentária é o produto da leitura e seleção de informações brutas presentes nos documentos, elaborada com fins documentários. Constitui o resultado da identificação da informação documentária e, como tal, é uma construção que se reporta aos itens considerados mais significativos selecionados do texto original, em função de objetivos institucionais.

A síntese documentária pode se apresentar como uma síntese concreta — a partir do registro da hierarquização de conceitos e informações presentes no texto, enunciados, resumos — ou como uma síntese mental (20) — no caso de traduções diretamente feitas a partir do texto original, como se faz usualmente.

O registro de palavras significativas, unitermos e palavras-chave representa um produto documentário bastante rudimentar, cuja característica principal é a simples enumeração: lista de palavras retiradas do texto, lista de idéias ou itens mais relevantes do texto. O uso de LN, nestes casos, é um procedimento comum e a operação realizada, na maior parte das vezes, caracteriza-

(20) A síntese mental constitui um processo pelo qual não são explicitadas as etapas de seleção e organização das informações, muito embora elas estejam presentes.

se como uma extração de palavras, feita à medida que estas vão surgindo nos textos submetidos à análise. O uso de unitermos e palavras-chave constitui, em alguns sistemas documentários, o recurso para a recuperação de documentos e informação, apresentando, todavia, problemas inerentes ao uso de uma linguagem não uniforme e normalizada.

A hierarquização de conceitos constitui um produto documentário um pouco mais elaborado. Sua construção supõe a organização dos itens de registro segundo um esquema de subordinações sucessivas, utilizando tanto termos da LN como da LE. Para construir hierarquias, o documentalista usa seu estoque de conhecimentos agregado ao quadro referencial da área e a estrutura textual, visando à organização das idéias principais e dos conceitos presentes no documento.

Os produtos documentários apresentados acima não constituem frases, sendo, antes, um registro de palavras — isoladas ou organizadas — presentes ou não no documento original.

O enunciado é um produto documentário que pressupõe uma abstração maior. É, geralmente, constituído por uma frase sintética, através da qual se procura registrar a idéia principal do texto. Sua construção prevê o uso da LN — para a organização da frase — e da LE. O uso de palavras presentes no texto original pode constituir um recurso para a montagem da frase. Sua construção se faz, segundo Van Dijk, por meio da inferência, supondo o conhecimento que um sujeito-falante tenha das regras subjacentes às relações entre as frases, ou seja, sua competência textual (VAN DIJK, 1977, p.203).

Já o texto documentário é uma síntese concreta, fruto de uma sistematização metodológica, construído com a finalidade de exercer uma função documentária. A designação “texto documentário” pretende contemplar um resumo documentário, que se reporta ao texto original, pois mantém com ele uma relação de contigüidade e semelhança. O texto em questão é uma modalidade de representação por condensação, realizada com fins documentários, marcada pela presença e organização de informações brutas do texto original.

O texto documentário é um elemento importante e indispensável para a realização consistente da tradução (operação documentária de conversão de textos em LN e/ou LE para uma linguagem

normalizada e controlada, a LD, denominada indexação).

O resumo é o único produto para o qual a Documentação, até o momento, prescreveu regras de apresentação. A literatura em documentação prevê diversos tipos de resumo (cf. MACEDO e TÔLAMO, 1978). Dentro dessa tipologia, o resumo científico é um “texto documentário” no sentido que adotamos, uma vez que constitui uma síntese que mantém respeitadas as estruturas informacionais significativas do discurso científico: a hipótese, o modo de tratamento do problema, argumentos, solução do problema e conclusão (idem *ibidem*, p.65). Para as autoras, um resumo de texto científico, “não se apresenta como simples enumeração dos itens desse texto, mas deve expressar concisamente a relação com o objetivo do autor e as discussões que conduzirão à conclusão(...) Constitui uma mensagem completa, inteligível por si mesma” (idem *ibidem*, p.71).

A diferença básica que caracteriza os diversos tipos de produtos documentários obtidos a partir de condensação, portanto, é sua função: fornecer indicadores sobre os temas ou assuntos tratados no documento, de forma não necessariamente normalizada (unitermos, palavras-chave, registro de palavras significativas, registro da hierarquização de conceitos) e constituir substitutos documentários, como instrumentos para a tomada de decisão sobre a ida ou não ao texto original (textos documentários ou resumos científicos).

Do ponto de vista da A.D., ambos podem ser tomados como produtos intermediários, se se visa à representação por intermediação de LDs: nos primeiros, há uma remissão a referenciais de assuntos tratados nos textos, e não aos textos eles mesmos; nos segundos, a referência é mais precisa e particularizante. Se os resumos (textos documentários) podem ser considerados como o produto documentário ideal para proceder a tradução e representação, imperativos de ordem econômica inviabilizam esta possibilidade: o tempo e o volume de informação e trabalho exigidos na elaboração de resumos não justifica que eles sejam feitos exclusivamente como produto intermediário.

Do ponto de vista do usuário, a função das palavras-chave e unitermos (que aparecem, geralmente, acompanhando resumos de publicações) é a de fornecer, de forma extremamente sintética, indicativos sobre o conteúdo dos documentos. Não dispensam a leitura do resumo — e do texto — e

funcionam apenas como informação adicional.

Já os resumos constituem uma representação condensada do documento original, configurando-se como um produto acabado, pronto para ser consumido. Funcionam, sob esse aspecto, como instrumento de decisão sobre a consulta ao texto original.

Salvo alguns casos, a Documentação não dispõe de regras explícitas para a construção dos registros de síntese obtida a partir da leitura e seleção documentárias. Dispõe-se, entretanto, de instruções, na sua maioria de caráter normativo, para a classificação e indexação de documentos ou para a elaboração de resumos (supondo que as operações de leitura e seleção já tenham sido efetuadas).

Em relação às regras para classificação de documentos (que pode ser vista como um processo de registro de síntese a ser traduzida, em seguida, por um código de classificação), a maior parte dos textos se limita a destacar a importância da “identificação do assunto principal”, como uma operação a ser feita a partir da “leitura técnica do documento”, encaminhando o classificador às partes físicas do documento onde ele provavelmente encontrará “pistas” para a localização do assunto principal (informações de capa, sumário, orelha da obra, introdução etc). Contribuição mais significativa a esse respeito encontramos na obra de Shera e Egan, onde se sugere um método de análise para a classificação de documentos, a partir das características do material a ser analisado, e “das categorias de características que ele apresenta” (SHERA, e EGAN, 1969, p.139). Os autores sugerem a aplicação de uma “fórmula geral que desvie a atenção do classificador da consideração do livro como um exemplo de unidade de propósito do autor, focalizando-a em direção aos elementos constitutivos que passam a ser de utilidade presumível para qualquer leitor” (Idem *ibidem*). Ou seja, não são considerados os discursos em si, mas as informações que eles veiculam e/ou seu assunto genérico. O contexto considerado é a unidade bibliográfica.

Essa proposição tem como base a sintaxe gramatical e a possibilidade de abstração da estrutura da sentença, buscando detectar o padrão comum de relações abstratas que podem ser substituídas por “qualquer termo de acordo com cada problema ou situação particular” (SHERA & EGAN, 1969, p.139). Sinteticamente, essa fórmula de análise consiste em determinar: o agente (ou assunto

enquanto campo de estudo ou disciplina acadêmica), o ato, o instrumental, o objeto, o tempo, o espaço e o produto resultante. Reproduzimos abaixo, a lista de elementos e sua conceituação.

A aplicação de método similar foi utilizado por CUNHA, para iniciar o processo de estruturação de vocabulário. A partir de experiência sugerida por CINTRA (CINTRA, 1983, p.5-22) e pelo PRECIS (AUSTIN, 1974) com base nos estudos de casos (CUNHA, 1990, p.114-119), CUNHA retoma os textos de Fillmore e Pottier sobre a “especulação acerca da existência e possibilidade de determinação de casos numa língua”, e sobre a “identificação dos casos conceptuais” (idem *ibidem*, p.114). A adequação realizada para uso documentário é bastante próxima àquela sugerida anteriormente por Shera e Egan, prevendo: Instrumento, Agente, Objeto, Modo, Lugar, Produto e Finalidade (idem *ibidem*, p.118).

Em relação às regras para elaboração de resumos, verificamos que a maior parte delas se refere à sua apresentação formal, não discutindo os procedimentos para a sua gênese, exceção feita ao trabalho de MACEDO e TÁLAMO, mencionado anteriormente.

Esse fato é confirmado, por exemplo, a partir de uma pesquisa sobre regras para resumo, feita por Kuhlen na Universidade de Constance, na Alemanha, a partir da tabulação de questionários de cinquenta e uma organizações engajadas em serviços de resumos. A atividade de elaboração de resumos é vista como uma arte baseada na experiência e no conhecimento pessoal, sendo que sua qualidade depende, em grande parte, da homogeneidade do campo de assunto coberto ou do conhecimento que o resumidor tem das necessidades de seus usuários. Em relação às regras para sua produção (que geralmente se reportam às regras universalmente aceitas - INIS-71, DIN 1426, ISO 5966 etc), verificou-se que o aspecto mais importante — sua produção intelectual — não é contemplado, remetendo, invariavelmente, ao conhecimento do resumidor, o que é extremamente vago. Um extrato das referidas regras reproduzidas pelo artigo pode ilustrar o fato: “tanto quanto possível, aquilo que seja essencial”; “pequeno, preciso, significativo”; “deve apresentar os fatos e descobertas mais importantes de um estudo científico, de uma maneira breve e clara”; “o texto deve ser informativo e concreto”; “uma clara condensação dos argumentos e descobertas essenciais constantes no original. Deve indicar o tópico, a metodologia, e particularmente os resultados e conclusões obtidas, com ênfase nos aspectos quantitativos” etc (KUHLEN, 1983).

Vê-se, portanto, que, com raras exceções, a Documentação não dispõe de instruções seguras para a construção do texto documentário. Não é nosso objetivo, entretanto, nos limites desse trabalho, aprofundarmo-nos na questão. As reflexões que fizemos tiveram como objetivo principal levantar algumas observações sobre a necessidade de contar com indicadores para a construção do texto documentário, gênero de representação que mais se aproxima do documento original. Por outro lado, um texto documentário bem construído pode se constituir numa base consistente para a tradução e representação documentárias.

CAPÍTULO 5

A CARACTERIZAÇÃO DA CONSTRUÇÃO DAS REPRESENTAÇÕES DOCUMENTÁRIAS

5.1 Representação documentária, sistemas de significação e informação

As representações de caráter individualizante pressupõe um vínculo estreito com o documento original (como é o caso dos resumos). Para Eco, “os momentos de representação são mais facilmente aqueles em que um código nasce do que aqueles em que um código preexistente é observado” (ECO, 1991, p.282).

A representação documentária, entretanto, supõe, como vimos, a intermediação de um instrumento comutador — uma LD. O código intermediário pressupõe uma normalização e pode ser visto como um “recurso destinado a aumentar o rendimento informativo de uma mensagem” (COELHO NETTO, 1980, p.140). Tradicionalmente, procura-se no texto a invariante documentária registrada pelo código. Por esses motivos, as representações documentárias são de caráter generalizante. Por sua vez, o código intermediário determina o nível de informação a ser veiculado.

Temos, portanto, uma dicotomia: se por um lado a representação sem a intermediação de códigos documentários tem a vantagem de substituir, de forma reduzida, o texto original, por outro, a necessidade comunicacional impõe o uso de códigos de intermediação, prevalecendo, neste caso, a generalização.

O código de intermediação funciona, nesse sentido, como um elemento para assegurar um rendimento informativo modal. A normalização obtida através do código documentário, embora tenha, por um lado, um caráter preditivo impedindo a representação da subjetividade expressa nos textos originais, garante, por outro, a circulação de informações, revelando, dessa forma, seu caráter dinâmico.

Neste caso, a normalização proposta pelo código documentário exerce uma função social, otimizando a comunicação e a disseminação de informações.

Deve-se lembrar, entretanto, que, em sistemas de Documentação mais especializados, é desejável alcançar algo mais que a informação modal, ou seja, através das representações, pretende-se reportar às relações de significação presentes nos textos. Mas a representação documentária obtida através da intermediação de códigos documentários, na maior parte dos casos, remete antes a áreas de assunto estabilizadas, previstas e normalizadas pelo código comutador, e não às relações de significação consubstanciadas nos textos.

A conversão realizada para se chegar à representação documentária supõe uma operação do tipo “tradução”, de uma LN ou Especializada - LE, para uma LD normalizada. Parte-se do resultado de um processo de condensação de textos particulares, em LN e/ou LE (objeto do processo de análise e síntese documentárias), para se chegar à representação documentária normalizada. O instrumento de comutação — a LD escolhida — tem, portanto, uma função normalizadora.

Como já o afirmamos anteriormente, vários sistemas semióticos integram e concorrem para a configuração desses produtos: a língua, enquanto sistema social, a instituição, a ideologia etc., impondo determinados recortes no continuum da realidade. (A correspondência entre LD e ideologia é enfatizada por CUNHA, cf. CUNHA, 1990, p.72). Cada um desses elementos remete, por sua vez, a determinados sistemas de significação: a representação fica, desse modo, como o resultado de múltiplas interferências. O significado de uma representação, portanto, não vai estar nas palavras que a compõem, mas nas estruturas de significação a que remete. Para interpretar uma representação documentária deve-se, conseqüentemente, reportar-se a sistemas de significação determinados.

A informação documentária — objeto da representação documentária — remete a sistemas de significação que só se consubstanciam nos textos. Em decorrência, a representação documentária deve se remeter a esses sistemas se quiser transmitir informação. Como a representação documentária, conforme afirmamos anteriormente, não consegue operar com as informações particulares dos textos e tenha, ao mesmo tempo, que exercer a função comunicativa, chega-se a

um impasse. Esse conflito é imposto pela ausência de uma articulação que assegure que as representações mantenham uma relação de contigüidade e semelhança com o texto original, desde a fase de tratamento até a recuperação.

Segundo a abordagem proposta, portanto, a incorporação de uma gramática — que não se restrinja a uma sintaxe (enquanto conjunto de regras que permite a combinação lógica dos termos, ou a sintagmatização) — para a geração das representações documentárias, é essencial. Acrescente-se ainda, a necessidade de uma discussão sobre a possibilidade de se pensar um signo documentário que, no sentido semiótico, deveria ser instrucional, ou seja, deveria permitir a recuperação não apenas das semelhanças e equivalências entre palavras, mas estabelecer a cadeia de relações possíveis desencadeadas a partir de determinada palavra, numa dada área de conhecimento, num determinado contexto e circunstância.

Interpretar um signo, no sentido peirceano, “significa definir a porção de conteúdo veiculada em suas relações com outras porções derivadas da segmentação global do conteúdo; e definir uma porção através do emprego de outras porções, veiculadas por outras expressões” (ECO, 1991, p.60). Na conversão documentária, portanto, coloca-se em jogo a significação.

5.2 As operações de representação

5.2.1 O texto documentário como referencial básico para realizar a operação de representação documentária

A principal marca dos produtos documentários convencionais é seu caráter generalizante. E é esse aspecto que confere à Documentação a possibilidade de construir e constituir sistemas. A generalização, entretanto, conforme já afirmamos, implica a perda de indicativos individuais dos textos, ou seja, a perda de autoria. Nesse sentido, a construção de textos documentários dá início ao processo de "desautenticação" de textos e discursos.

O texto documentário — produto da síntese ou condensação — representa o primeiro passo em

direção à generalização. Conceitos, idéias e fatos presentes no texto original são colocados em classes categoriais, visando a facilitar seu reconhecimento posterior, a partir de operações de comparação (e diferenciação). Portanto, os textos documentários desencadeiam o primeiro movimento de perda dos referenciais concretos do texto.

A intermediação de códigos acaba por acentuar a generalização, passando a desvincular as representações documentárias dos referenciais presentes nos textos.

Por outro lado, mesmo na presença de vocabulários controlados, a atribuição empírica de palavras-chave, descritores etc. para a construção das representações através da indexação, baseada que é na experiência ou hábito (cf. GARDIN, 1970, p.631), reduz a possibilidade de consistência dos produtos documentários deste nível.

É neste sentido que o texto documentário pode constituir um referencial importante para a realização das operações de intermediação. A síntese mental e a conversão simultâneas não deixam explícitas as operações de natureza lógico-semânticas realizadas, fazendo com que o aprendizado para a indexação fique condicionado a procedimentos baseados no ensaio e erro e dependentes da performance de cada analista.

Sendo o texto documentário o registro articulado das informações referenciais presentes no documento original (conforme expusemos no cap. anterior), fica simplificada a identificação dos elementos que deverão ser objeto de tradução via código documentário. Desse modo, procura-se no texto documentário — síntese que apresenta uma relação de contigüidade e semelhança com o texto original — os elementos referenciais a serem representados.

Na indexação realizada diretamente a partir do texto original, a perda de referenciais concretos é minimizada quando o profissional envolvido com a tarefa é um especialista de área (em condições, portanto, de reconhecer a arquitetura conceitual do campo em questão e identificar e interpretar ocorrências típicas da linguagem técnico-científica), ou quando a experiência faz, do documentalista, um especialista.

Do ponto de vista metodológico, entretanto, deve-se prever a possibilidade de que o processo se desenvolva na ausência das condições acima relacionadas.

Embora se transfira à tarefa de análise o problema de identificação de informações a serem representadas, esta é a maneira de assegurar uma tradução fidedigna. Resta, entretanto, assegurar a existência de um código documentário consistente para realizar a comutação.

5.2.2 A mediação do código documentário

Os processos de AD empreendidos diretamente a partir de textos em LN ou LE, sem a intermediação de códigos documentários — caso dos processos de extração simples, como os índices e concordâncias, sistemas KWIC e KWOC, ou das experiências hoje realizadas em sistemas de Inteligência Artificial — serão excluídos deste trabalho em razão, ou da precariedade dos resultados e rejeição dos métodos essencialmente lexicais (1o. caso), ou do desconhecimento (e uma certa temeridade) dos procedimentos que embasam as operações sobre os próprios textos (2o. caso). Consideraremos, portanto, os aspectos relacionados à construção de representações documentárias em função de códigos de conversão documentários.

Os códigos de conversão (instrumentos de comutação, intermediação etc.) constituem uma modalidade de metalinguagem elaborada a partir de sistemas de significação de áreas constituídas (organizadas). Por sua vez, constituem sistemas de significação (LDs) com função específica: normalizar os conceitos de área, controlar seu uso e viabilizar a interface documentação-usuário.

Visando à comunicação, uma LD utiliza um sistema de significação. Na terminologia de Hjelmslev, uma LD apresenta um plano de expressão e um plano de conteúdo (HJELMSLEV, 1975, p.53 e ss.), organizados em função de um determinado paradigma: determinados postulados de significado (ECO, 1984, p.12) são considerados para sua montagem e estruturação.

Dito de outra maneira, o arranjo das significações numa LD representa uma perspectiva de organização do conhecimento. De forma explícita ou implícita, as definições das expressões e suas

inter-relações internas são relativas a uma determinada visão de mundo. O uso de uma LD, portanto, imprime aos produtos resultantes da intermediação a perspectiva selecionada. Por exemplo, a Classificação Decimal de Dewey (19a.ed.) separa a Antropologia Física (572) da Antropologia Cultural (306), enquanto que o Tesouro de Sociologia, de Jean Viet os considera como termos equivalentes (cf. VALE, 1987, p.13-14). Revelam-se, desse modo, concepções diferentes sobre o mesmo assunto, face aos postulados de significado tomados como base para realizar a relação forma de expressão/forma de conteúdo. Nesse sentido, as Linguagens Documentárias constituem a expressão de um “recorte” no conteúdo, condicionando, conseqüentemente, os procedimentos de interpretação e comunicação.

Por outro lado, “toda linguagem formalizada (...) não é mais exata ou exclusivamente uma linguagem, sobretudo se dela for retida sua função de comunicação” (GRANGER, 1974, p.139). Para Granger, as linguagens formalizadas são difíceis de memorizar e o que elas ganham em rigor, perdem em eficácia, sendo que na realidade, sua função comunicativa é apenas virtual. Para o autor, tal fato se deve, primeiro, à ausência de uma “segunda” articulação: o sentido de suas expressões “é diretamente embreado (21) no sistema dos sintagmas que corresponde ao primeiro nível de articulação das línguas naturais” e em segundo lugar, porque “as línguas formais ignoram os símbolos de ‘embreagem’ numa experiência vivida”, remetendo apenas a uma combinação de regras simbólicas que constitui seu ‘objeto’ no sentido de Peirce (GRANGER, 1974, p.140-141).

Nesse sentido, as LDs têm prejudicada sua função comunicativa. Para GARCÍA GUTIÉRREZ, “o processo documental se inscreve na comunicação humana através de documentos. Análise e linguagem utilizam o adjetivo “documental” como matização do suporte material dos discursos e textos observados. Mas estes são produzidos em LN e portanto, deve caminhar-se para a “naturalização” de todos os elementos e normas da Lingüística Documental” (GARCÍA GUTIÉRREZ, 1990, p.26-27), se se quiser manter a função comunicativa. “Trabalhamos com vocabulários, normas e sistemas fechados, frente ao caráter aberto da linguagem natural” (idem ibidem, p.31). Deve-se lembrar, todavia, que em função da necessidade de delimitação rigorosa do significado não é viável uma volta à LN como solução.

(21) Embreado, ou seja, articulado.

Quanto ao segundo aspecto levantado por Granger (a perda da 'embreagem' na experiência vivida), deve-se lembrar que numa LD, via de regra, opera-se o isolamento das expressões de conteúdo relativamente aos textos onde elas podem ocorrer. A ausência de um corpo sistemático de definições que se reportem a contextos determinados, ou o uso indevido das expressões de conteúdo de uma LD, conseqüentemente, compromete sua função comunicativa.

O predomínio da utilização da equivalência lexical (quando este procedimento só deveria ser utilizado para resolver os problemas de sinonímia lingüística ou funcionar como um meio de entrada no sistema), transforma uma LD em nomenclatura e, conseqüentemente, torna inúteis os produtos documentários enquanto veículo de informação, uma vez que, sob esta forma, uma dada palavra encontrada num texto pode ser "representada", automaticamente, por palavra semelhante encontrada na LD, alterando-se os vínculos de significação.

Muitas vezes, a ausência de definições nas LDs é a responsável pelo uso indiscriminado da equivalência lexical. Este é um aspecto reconhecido por Gardin quando ele afirma que "não importa que termos sejam admitidos como equivalentes entre si: só importa sua definição, não sua designação", o que nos permite enfatizar a importância de sua explicitação (GARDIN, 1970, p.632).

Um exemplo pode ilustrar este fato. A nota de escopo do Thesaurus Popin recomenda o uso do descritor "planejamento familiar" para representar o "controle intencional da fecundidade por parte de indivíduos que decidem ter filhos somente se e quando querem" (VIET, 1986). Indexar um documento sobre planejamento familiar produzido pela BENFAM sob esse descritor, significa esquecer que as ações desenvolvidas por esse órgão não coincidem com a definição atribuída ao descritor, uma vez que, nesse caso, não se delega a decisão sobre o controle da fecundidade aos indivíduos, mas antes, a eles, lhe é imposta.

Os procedimentos baseados nas relações de equivalência são bastante comuns no cotidiano do documentalista, tanto pela falta de uma sólida formação profissional, como pela precariedade dos instrumentos de conversão utilizados para a representação documentária. O uso indevido pode ser saneado a partir de definição de metodologias que não sejam baseadas apenas na performance, ou

bom senso. O rigor metodológico vai impor, necessariamente, a consideração do contexto onde as palavras se inserem, para realizar a representação.

Pode-se, também, abordar a questão sob um outro aspecto. Para CINTRA, os termos que compõem uma LD são baseados em regras “de jure” (CINTRA, 1983), ou seja, estabelecidos aprioristicamente de maneira impositiva. Ao realizar uma correlação palavra no texto — palavra na LD, realiza-se uma implicação nem sempre verdadeira. Uma implicação falsa — sem referente material — não informa nada.

Exemplo disso é o termo “Bancos de Desenvolvimento”, descritor do Macrothesaurus das Nações Unidas, onde a “inclusão do Banco Mundial entre os Bancos de Desenvolvimento pressupõe a aceitação de uma definição de desenvolvimento que permita identificar as ações desenvolvidas por esse órgão com aquelas que se encontram nos limites do conceito de desenvolvimento, tal como é caracterizado pela instituição em questão (OCDE e Nações Unidas) (TÁLAMO e outros, 1990, p.6-7). Incluir, por exemplo, o BADESP, Banco de Desenvolvimento do Estado de São Paulo sob o descritor “Bancos de Desenvolvimento” constituir-se-ia uma implicação falsa, uma vez que são bancos de naturezas diferentes.

Vê-se, dessa forma, que a intermediação do código intensifica outras intermediações (intermediação do indivíduo-profissional, vícios profissionais, língua etc.). Múltiplas determinações respondem pela configuração do produto documentário.

5.3 A construção das Linguagens Documentárias

As reflexões que realizamos sobre a construção das LDs reportam-se a uma concepção de instrumento de mediação que permita, entre outras coisas, o exercício da função comunicativa. Por outro lado, consideramos que a tarefa de construção de LDs é, essencialmente, interdisciplinar. Essas premissas impõem, conseqüentemente, um limite ao nosso trabalho: sob a ótica da Documentação, restringiremo-nos a levantar alguns aspectos teóricos relativos a essa construção, abstando-nos de propor procedimentos rigorosos de operacionalização.

5.3.1 A designação dos instrumentos de intermediação documentária

Neste trabalho utilizamos a expressão “Linguagem Documentária” para nos referirmos aos instrumentos de comutação documentária, a despeito das restrições que apresentam para exercer a função comunicativa. Adotamos essa denominação em razão do levantamento de problemas que possam auxiliar na sua configuração enquanto “termo”, obedecendo, desta forma, a um outro critério: o da difusão da expressão que reflete, por sua vez, uma necessidade.

A denominação desses instrumentos varia dentro da literatura: léxico documentário, vocabulário, nomenclatura, terminologia, e, mais freqüentemente, LD. Confere-se a esses instrumentos, quase que invariavelmente, o status de metalinguagem. Essa variedade de designações, mais sua classificação como metalinguagem, sugere a necessidade de realizar algumas distinções.

Os *léxicos* designam o conjunto das unidades que formam a língua de uma comunidade, atividade humana etc (DUBOIS et alii, 1988). Há instrumentos de intermediação documentária que apresentam esta característica, na medida em que são apenas uma listagem de termos sem articulação. (Para Gardin, uma metalinguagem documentária possui um léxico, por um lado, e uma sintaxe, de outro, a qual permite realizar as articulações. O que ele caracteriza como um “léxico” — por exemplo, no SYNTOL — não é um léxico no sentido da definição que adotamos) (GARDIN, 1973a; 1973b).

Os *vocabulários* se referem ao conjunto das ocorrências que integram um determinado corpus discursivo, como uma lista de unidades da fala (DUBOIS et alii, c1973). (Exemplo de vocabulário encontramos no trabalho de CUNHA, relativamente às ocorrências registradas nos discursos sobre política colonial de Adriano Moreira) (CUNHA, 1990).

Uma *nomenclatura*, por sua vez, supõe biunivocidade da relação significado-significante (DUBOIS et alii, ibidem).

De um modo geral, as *terminologias* referem-se ao conjunto de termos de uma área, definidos rigorosamente para designar as noções que lhe são úteis (idem ibidem). (Exemplo de terminologia,

o Glossário de Termos em Educação: BRASIL, Ministério da Educação e Cultura, 1980).

A definição do termo *linguagem* indica a capacidade específica da espécie humana de se comunicar por meio de um sistema de signos vocais (ou língua), (...) e supõe a existência de uma função simbólica (...) (idem ibidem); uma das suas características é a de ser articulada (BORBA, 1971).

Uma *metalinguagem*, no sentido hjelmsleviano, ou metasemiótica, é uma semiótica cujo plano do conteúdo é, ele próprio, uma semiótica, ou seja, uma semiótica cujo conteúdo é uma semiótica (HJELMSLEV, 1975, p.126). E uma metalinguagem científica (metasemiótica científica), é uma metalinguagem cuja semiótica-objeto é uma semiótica científica (idem ibidem).

Nesse sentido, uma LD deveria funcionar como uma metalinguagem, já que o seu conteúdo pretende ser uma semiótica: um sistema de signos (e de significação) que se reporta a outro sistema de signos (e de significação).

Para que os instrumentos de representação atinjam o status de linguagem, entretanto, devem ser articulados e cumprir uma função de comunicação. Já que raras vezes tais instrumentos cumprem essa função, designá-los como “Linguagens” revela a utilização de um termo impróprio. (O SYNTOL, desenvolvido por Gardin e sua equipe, pretende realizar esta articulação, mas ela é limitada em função da visão estritamente lógica das relações sintáticas).

Para funcionar como metalinguagem, a linguagem deve realizar uma função referencial, possibilitar uma conversão, exercendo, dessa maneira, a função comunicativa. Mas, ao mesmo tempo, uma LD deve controlar o vocabulário (inclusive para possibilitar a comunicação) e constituir um instrumento de natureza semiótica que permita desencadear cadeias de relações — a semiose — a partir de uma palavra (ou conjunto de palavras) sem se restringir a equivalências, semelhanças e correlações simples (exercer a função de signo e possibilitar a semiose). Esse conjunto de atributos, portanto, deveria ser considerado para a construção de LDs consistentes.

Para GARCÍA GUTIÉRREZ, vários fatores contribuem para que as LDs deixem de se caracterizar como metalinguagens cujo referente é a descrição discursiva. Dentre eles, destaca: a inserção das

LDs em processos comunicativos para regular formas de entrada e saída de enunciados; a existência de uma estrutura relacional e sintática e as diferentes técnicas para sua criação e operacionalização. Por outro lado, “a partir do isolamento terminológico que precede a construção de uma LD, as relações que conformam sua estrutura não se realizam em função do campo discursivo representado senão mediante mecanismos de expansão e oposição gerados em um sistema distinto” (GARCÍA GUTIÉRREZ, 1990, P.71-72).

Para Granger, os interpretantes que definem as significações numa língua científica constituem “(...) remessas a próprias experiências já estruturadas no simbolismo, ou pelo menos, suscitando o problema imediato de sua estruturação” (GRANGER, 1974, p.152). Uma expressão metalingüística que remete ao sistema das ciências é um significante, que “(...) remete a outros objetos e estas correlações elevam-no necessariamente ao nível de conceito” (idem ibidem).

Assim sendo, uma LD deve, antes, se remeter a esse simbolismo, e não valer-se de sua incidência texto a texto. Isto porque no sistema de significação das ciências, ou mais precisamente, na terminologia dessas ciências, as denominações são significantes conceituais e não variantes textuais.

A questão da designação dos instrumentos de comutação documentária não é, portanto, uma simples questão de denominação. Ela deve ser analisada a partir do conceito de representação que embasa sua concepção, bem como da delimitação clara de suas funções.

Nesse sentido, a vinculação estrita do processo de construção das LDs ao processo de análise de textos com fins documentários deve ser vista com reservas. Ela representa uma opção de desenvolvimento puramente empírico de construção de LDs e acrescenta à atividade de análise, a função de escolha das denominações e definições que deveriam ser remetidas ao sistema de significação das ciências.

5.3.2 A definição e a linguagem L em formato de enciclopédia como referência para a construção das Linguagens Documentárias

Para Eco, uma linguagem L em formato de enciclopédia é aquela que “de algum modo contenha entre as próprias regras de significação instruções pragmaticamente orientadas” (ECO, 1991, p.76). Deve-se entender “L não como um sucinto dicionário, mas como um *complexo de competências semânticas enciclopédicas*” (idem *ibidem*, p.77).

A noção de enciclopédia semântica, portanto, enfatiza a necessidade de considerar a inserção comunicacional dos enunciados, ou seja, a impossibilidade de realizar representações no âmbito tradicionalmente considerado lingüístico.

O verbete “Enciclopédia” elaborado por Diderot, na “*Encyclopédie*”, diz que “Esta palavra significa *encadeamento de conhecimentos*; é composta pela preposição “*en*”, em, e pelos substantivos “*ciclos*”, círculo, e “*paidéia*”, conhecimento (...)” (DIDEROT, 1989, apud COELHO & CHIARETTI, 1989).

Embora Eco não se refira diretamente ao conceito de enciclopédia registrado pela *Encyclopédie*, é a partir de uma definição desse gênero que se desenvolve a noção de “enciclopédica semântica”. A base da construção de uma linguagem enciclopédica, para Eco, está assentada na discussão sobre “definição”, numa abordagem que incorpora uma teoria do significado que prevê a integração da semântica e da pragmática, inserindo, desta forma, a questão da representação num contexto de comunicação.

A noção de “definição” de Eco é idêntica às definições clássicas no que se refere aos modos de apresentação: ela não deve ser circular e deve se referir à essência (ECO, 1991, p.80, p.96; HEGENBERG, 1974, p.27 apud TÁLAMO, 1987, p.89; ISO 704, 1987), prevendo a necessidade de absoluta reciprocidade entre *definiens* e *definiendum* (ARISTÓTELES, apud ECO, 1991, p.96).

Ao privilegiar a questão da diferença para o estabelecimento da definição, a concepção de Eco atende também às recomendações já sistematizadas pela literatura, ou seja, o privilégio do enfoque

“intensional” sobre o extensional (ISO 704, 1987; DESCLÉS, 1987, p.59) para a elaboração de definições. Nesse sentido, gêneros e espécies constituem parâmetros extensionais — valores de verdade que definem classes — que constituem referências para a construção de árvores a partir de diferenças que, por sua vez, vão fixar o regime intensional. (ECO, 1991, p.105). A intensão refere-se à totalidade de características do conceito, enquanto que a extensão à totalidade de espécies ou objetos que tem todas as características do conceito (ISO 704, 1987, p.2).

Ainda segundo Eco, a definição é também uma operação sobre a linguagem (é sempre um discurso) (ECO, idem ibidem, p.97). Não se pode, portanto, entender a palavra “definição” como sinônimo de uma coisa ou de uma idéia, confundindo mundo exterior e o universo da linguagem: a linguagem é sempre uma forma de organização (TÁLAMO, 1987, p.88, 90). Através da linguagem, a definição remete a postulados de significado que são estabelecidos em bases pragmáticas (ECO, idem ibidem, p.88). A pertinência de uma definição, portanto, remete à realidade extralingüística, e “(...) resolve--se no âmbito das formulações de objetivos” (TÁLAMO, 1987, p.94) ou seja, a definição varia na dependência do ponto de vista privilegiado e dos postulados de significado determinados pelo enfoque escolhido.

Entretanto, ao privilegiar a questão da diferença, ECO vê o gênero como irrelevante: o gênero é apenas um nome que rotula grupos de diferenças (ECO, idem ibidem, p.105, como “(...) fantasmas verbais que cobrem a verdadeira natureza da árvore e do universo que ela representa, *um universo de puras diferenças*” (idem ibidem, p.106). Se, por um lado, essa constatação permite confirmar a ascendência da enciclopédia sobre o dicionário — “a árvore de gêneros e espécies, de qualquer modo que seja construída, dissolve-se numa poeira de diferenças (...) não hierarquizável de *qualia*” (idem ibidem p.109) e, conseqüentemente, o dicionário “dissolve-se necessariamente (...) numa galáxia potencialmente desordenada e ilimitada de elementos de conhecimento do mundo” (idem ibidem) dando lugar à enciclopédia —, por outro, perde-se a referência necessária — o gênero — para a organização concreta de uma definição.

A diferença supõe necessariamente uma semelhança, e o gênero próximo é um elemento essencial para o estabelecimento da conjunção em função da qual a diferença permitirá realizar a disjunção. O privilégio da diferença constitui, na verdade, um argumento para destacar que as possibilidades

de organização do mundo, através da linguagem, não podem ser observadas apenas sob a forma de uma “árvore”, mas como um “rizoma”, onde cada ponto pode e deve ser unido a qualquer outro ponto, ou seja, como uma “rede de árvores que se abrem em toda direção (...) O rizoma não tem centro” (DELEUZE e GUATARI, apud ECO *ibidem* p. 115).

A enciclopédia, no entanto, é um postulado semiótico (ECO, p.113) e a possibilidade concreta de seu uso só se dá através de “enciclopédias locais ou parciais”, conforme, por exemplo, em Quine, Fillmore, Petõfi etc. (*idem ibidem*, p.116-131). Dito de outra forma, a enciclopédia supõe o dicionário, como forma de representar localmente a enciclopédia (*idem ibidem*, p.131).

Vê-se, portanto, que a importância da contribuição de Umberto Eco se dá na necessidade de incorporação da pragmática (e portanto, das visões de mundo e do elemento comunicacional). No nível prático, entretanto, ela não fornece instrumental claro para operacionalização.

Nos estudos de classificação, a importância da fixação das diferenças para a organização de conceitos é identificada, por exemplo, a partir das classificações facetadas — cuja origem pode ser computada à Colon Classification, desenvolvida por Ranganathan — introduzidas pelo Classification Research Group - CRG, na década de 50, na Inglaterra.

A concepção suposta pelas classificações facetadas representam uma alteração significativa nas formas de organização do conhecimento (para uso bibliográfico) pressupostas pelos Sistemas de Classificação tradicionais, como a Classificação Decimal de Dewey - CDD e a Classificação Decimal Universal - CDU. Se nestas últimas, disciplinas convencionais determinavam a organização dos assuntos através da enumeração, nas classificações desenvolvidas pelos integrantes do CRG a organização por facetas privilegiava o uso.

Ao incorporar a noção de “faceta” (flexibilizando o âmbito da utilização do termo em relação à proposta de Ranganathan), o CRG demonstra, empiricamente, a necessidade de considerar as “diferenças” como determinantes do enfoque a ser dado à organização do conhecimento. Dentro das facetas, entretanto, prevalecem princípios empíricos e nem sempre consistentes de organização. Exemplo disso é a Classificação das Ciências do Solo, de Vickery, que agrupa os aspectos

sob os quais podem ser analisados os assuntos a partir de categorias — tipo, estado, propriedade, reação, operação etc. — prevendo uma sintaxe para combiná-los, conforme os assuntos se apresentem nos documentos (VICKERY, 1980). Mais recentemente, a noção de faceta é identificada com “categoria semântica” (VICKERY, 1986).

Do mesmo modo, os tesouros incorporam a noção de faceta para a organização de seus termos, estendendo, entretanto, a extensão do conceito de faceta para o momento da recuperação, quando se abandona a árvore hierárquica que sustenta sua organização, para aproximar termos que representam noções que a experiência empírica aproxima nos documentos.

A incorporação da experiência empírica da organização dos significados para a construção de instrumentos de intermediação documentária é também considerada na organização do SYNTOL, que prevê a constituição ‘por aprendizagem’ de novos agrupamentos lexicais” (GRANGER, 1974, p.215), em função do material bibliográfico tratado.

Embora as LDs relacionadas acima considerem a importância da flexibilidade face à representação de informações do mundo empírico, elas não se constituem — por motivos diversos e diferentes entre si — instrumentos que assimilem totalmente a informação pragmática e a importância da definição como mecanismo de controle de vocabulário e como elemento de comunicação. (Estes instrumentos serão analisados mais detalhadamente no cap. 6).

Em relação aos desenvolvimentos empreendidos pela Inteligência Artificial — que utilizam redes semânticas situadas em espaços multidimensionais — a questão da importância da definição também não é totalmente observada. Em seu trabalho sobre redes semânticas, Desclés identifica vários modelos de representação por traços (assentados na idéia de que uma rede é composta de traços semânticos, onde ‘traço’ equivale à ‘representação mental de propriedade’, que se reportam às análises em semas (Pottier) e às análises componenciais (Lyons, Rastier). Procurando evidenciar a pertinência teórica dessas redes, o autor analisa como elas se aproximam dos modelos ‘a traços’ da psicologia cognitiva, das análises sêmicas lingüísticas e das análises lógicas. Destaca-se dessa discussão a questão da natureza dos conceitos das redes: elas são de natureza intensional, desde que o conceito de intensão se refira a uma função com um valor dentro de um conjunto de

valores de verdade, e que o conceito de extensão, diga respeito à classe de objetos que se subordinam ao conceito (DESCLÉS, 1987, p.59-60).

A concepção de Desclés coincide com a de Eco (ECO, 1984, p.10; 1991, p.67). A determinação do nó principal que organiza os termos em hierarquia, define e delimita sua aplicação.

A enciclopédia como conceito, por outro lado, evidencia a necessidade de remissão à especialidade de área. É o que trataremos a seguir.

5.3.3 A Terminologia como referencial enciclopédico

Para construir representações documentárias (no seu sentido tradicional) trabalhamos, normalmente, com dois subconjuntos de produção lingüística: de um lado palavras ou sintagmas — que constituem as LDs ou os instrumentos de comutação — e de outro, textos, nos quais as palavras se atualizam, ganhando uma referência e um sentido.

Para falar dos sentidos das palavras podemos recorrer, por exemplo, aos dicionários. Nestes últimos os sentidos são, via de regra, “codificados” pela língua (codificação lingüística). Alguns dicionários podem incluir, também, outras significações devidas a uma codificação cultural (por exemplo, “cão, associado à fidelidade” (DUCROT & TODOROV, 1988, p.235).

Nos textos, todavia, as palavras ganham uma referência — e a informação só se consubstancializa nos textos em função dessa referência — e a consulta ao dicionário só irá fornecer elementos relativos à codificação lingüística (e algumas vezes, cultural). As modalidades de enunciação textual não são identificáveis apenas a partir dos significados registrados pelo dicionário porque, entre outras coisas, o dicionário não vai fornecer a referência, que é inerente àquele texto. “O problema da *referência* está estreitamente ligado à enunciação; como já havia notado Peirce, para que um signo possa denotar, deve passar pelo intermediário de um ‘índice’” (DUCROT & TODOROV, 1988, p.290). Do mesmo modo, só a referência permite colocar a questão da verdade: “em si mesmo, o enunciado não é verdadeiro nem falso, torna-se verdadeiro ou falso unicamente

no decorrer de uma enunciação particular” (idem *ibidem*).

Na operação de conversão documentária, portanto, pode-se comprometer a representação da informação, dada a ausência da referência ao contexto situacional, uma vez que as informações presentes nos textos passam a ser “representadas” por expressões (na maior parte das vezes, palavras) retiradas do instrumento de comutação.

Como o objetivo da representação documentária é o de promover uma situação de comunicação documentária, além da necessidade da articulação ela deve dispor de mecanismos que possibilitem a restituição dessas referências.

Nesse sentido, a utilização da noção de enciclopédia semântica como “hipótese reguladora”, pode viabilizar a restituição de referências, se não dos textos propriamente ditos, pelo menos das significações conforme conceptualizadas pelas Ciências.

Uma LD com características instrucionais pode prever a reconstituição de enciclopédias parciais (ou locais) de forma a possibilitar, além da representação, a identificação do formato de competência necessário para interpretar as unidades do sistema, e dessa forma, não só recuperar informações como realizar novas inferências e novas associações.

Um sistema instrucional enciclopédico deve estabelecer regras de implicação a partir do uso de expressões que funcionem como “interpretantes” (enquanto signos que definem as significações nos sistemas das ciências), permitindo e disciplinando o funcionamento da semiosis. Esse sistema pode permitir, pelo menos teoricamente, o reconhecimento das informações nos textos (ou nos textos documentários) e a mudança de código — a tradução documentária — minimizando as perdas de informação. Não se trata, pois, apenas de instaurar a substituição — equivalência — mas interpretação, onde uma proposição (um termo), como um argumento, pode permitir prever futuras ocorrências contextuais de outros signos. Deve-se lembrar que o critério de interpretância de Peirce “(...) não só retraduz o ‘objeto imediato’ ou o conteúdo do signo, mas amplia sua compreensão” (PEIRCE, apud ECO, 1991, p. 60).

Como, entretanto, as LDs tradicionais têm muitas características de léxicos — não têm referência precisa — as palavras que as integram nem sempre têm significado, ou podem remeter a vários significados relativos a enciclopédias não especificadas.

É nesse sentido, portanto, que a Terminologia pode fornecer elementos para a restituição de competências semânticas pertinentes, aumentando, dessa forma, a possibilidade de recuperação da informação e o desenvolvimento da semiosis.

A partir dos instrumentos convencionais a semiose fica sob uma camisa-de-força (LDs como nomenclatura) ou é “disparada” aleatoriamente (LDs como léxico). Como interessa a uma LD que ela permita desenvolver a semiosis, a nomenclatura é descartada. Não é função de uma LD estabelecer relações de denominação.

Vejamos, portanto, o que seria conceber uma LD a partir de palavras do léxico.

Para estabelecer as diferenças entre a palavra no léxico e a palavra no discurso, e para avaliar a capacidade do primeiro em recuperar o segundo, recorreremos à Le Guern.

Para Le Guern, a principal diferença entre as palavras na língua — que constituem o léxico — e as palavras no discurso — as palavras em funcionamento — reside na questão referencial. Na língua, as palavras não estão em relação imediata com as coisas, com o concreto; exprimem apenas um conjunto de propriedades, têm um significado, mas não uma referência; dizem respeito a propriedades, não a substâncias; a qualidades e não a objetos. Segundo a tipologia peirceana, as palavras na língua são “legisignos simbólicos remáticos” (PEIRCE, apud LE GUERN, 1989, p.340). E embora possam ser classificadas pelos lexicógrafos como substantivos, só são predicados: o signo de uma propriedade é um predicado. As palavras na língua não são, portanto, termos (LE GUERN, 1989, p.340-341).

Já no discurso, as palavras têm uma dimensão referencial: juntando léxico e sintaxe, as palavras no discurso designam objetos. (A unidade mínima que tem a possibilidade de designar um objeto

é um sintagma nominal). O emprego das palavras no discurso através do encerramento do predicado por um quantificador transforma esse predicado num termo (LE GUERN, p.340-341).

No léxico, as palavras significam a despeito de se tomar um universo qualquer que seja. São predicados livres. Nele os elementos não têm extensão, só compreensão, ou seja, não existem em função de alguma escala de valor (LE GUERN, 1989, p.341).

No discurso, à medida em que existe um referencial — um universo constituído de objetos determinados — as palavras passam a ter uma extensão, assumindo valores relativamente a esse universo particular. A construção dos sintagmas nominais pressupõe a tomada de palavras do léxico e sua inserção em determinados universos de discurso, ou seja, a associação de um predicado livre, a uma classe de objetos tomados dentro de um universo particular. O predicado, então, deixa de ser livre, para transformar-se num predicado relacionado (LE GUERN, 1989, p.341-342).

O predicado relacionado é, portanto, um predicado ao qual se confere um valor dentro de um sistema relacional de significados, ou seja, é um predicado que passa a ter uma extensão.

Mais uma vez, a distinção entre palavra no léxico e palavra no discurso leva-nos a concluir que a capacidade das LDs em recuperar os termos substantivos dos textos está diretamente ligada à questão da referência. E a referência só vai ser dada a partir da palavra no texto, não no léxico.

A Terminologia, entretanto, ao trabalhar com etiquetas ou expressões de classes de objetos tomados num universo determinado, constitui predicados relacionados, ou seja, termos referenciais.

Na terminologia de uma determinada área, uma palavra designa um determinado objeto, na medida em que o insere numa classe particular dentro da respectiva área. Essa mesma palavra, num léxico, exprimiria apenas um conjunto de propriedades, independentemente de qualquer objeto que seja e de qualquer universo que seja. (As palavras no léxico, significam, a despeito dos possíveis referentes).

A Terminologia trabalha com traços de substância, onde a escolha dos termos é estabelecida em

função de pontos de vista próprios de cada disciplina. Constitui, dessa forma, um exemplo de enciclopédia local ou parcial.

Isso posto, podemos inferir que a capacidade de um instrumento de intermediação documentária em recuperar as informações presentes nos textos, está condicionada à sua capacidade de referir. Para que essa referência exista concretamente, a construção das LDs deve ser feita com base na terminologia de área, em campos bastante delimitados.

É relativamente a esse aspecto que se colocam as críticas à maioria das LDs. Na ausência de parâmetros teóricos para sua elaboração, as LDs utilizam “(...) palavras da língua, e não do discurso”, o que “exclui o sentido contextual da composição do léxico, impedindo, portanto, que seja assegurada a invariante da informação. E um dos meios para atingir essa invariante é a normalização terminológica, objeto específico da terminologia” (TÁLAMO, 1991, p.2-3).

Por outro lado, uma vez que a Terminologia remete a sistemas de significação de área, ela constitui base legal para a construção das LDs. Restituindo sentidos — que só são possíveis nos discursos — ela pode tornar a LD interpretável no sentido peirceano, possibilitando não apenas a restituição de informações dos documentos do sistema, mas o engendramento de novas relações; não só correlações lexicais, mas instruções para o desencadeamento dos processos de inferência contextual. Os recursos oferecidos pela Terminologia podem imprimir um nível de conotação desenvolvido às LDs, sedimentando uma certa porção de enciclopédia de área.

Nesse contexto, cabe vincular a definição — que nos léxicos só é dada através de propriedades — à delimitação de conceitos em uma determinada área particular. Utilizando referencial da Terminologia, as LDs serão constituídas de termos analisáveis obtidos a partir de valores constituídos dentro de determinados universos, através de operações de “delimitação”. E se a Terminologia se refere a determinada área, tem-se assegurada uma referência para os léxicos documentários (os valores daquela lhe são transferidos), consubstanciada na seleção de termos, a partir de determinados universos discursivos dados.

Segundo Rastier, o conceito de valor rompe com a concepção do léxico como nomenclatura. Para

ele, “uma palavra não pode ser definida isoladamente, em relação àquilo que designa. Ela o deve ser relativamente a outras palavras” (RASTIER, 1987, p.104). As palavras não podem ser definidas fora de seu contexto, porque a língua não é nomenclatura. É o método onomasiológico que vai orientar as denominações: do significado ou conceito para o significante, e não o contrário (idem *ibidem*).

A diferença entre o significado lexical e o significado textual também é abordada por Eco. Para ele, da mesma forma que para Dubois e Rastier, o significado lexical não pode ser dado de forma independente dos contextos, porque essa hipótese implicaria tomar o significado textual como função do lexical. Citando Grice, Eco fala do significado convencional e do situacional. Em alguns casos, o significado convencional, numa dada linguagem L, não é suficiente para instaurar um processo efetivo de comunicação, ou seja, o destinatário pode ficar em dúvida ou interpretar erroneamente determinado enunciado se não agregar ao seu processo de interpretação, informações contextuais. (“Um contexto é uma classe de ocorrências de cadeias ou grupos de expressões (pertencentes a um ou vários sistemas semióticos ao mesmo tempo); define-se ao contrário como co-texto a ocorrência atual e específica de um membro dessa classe”) (ECO, 1991, p.73).

Dessa forma, a interpretação de um significado lexical, num corpus textual, não remete somente à identificação de determinadas propriedades correspondentes a essas expressões, mas também a uma série de instruções para que ela possa ser interpretada segundo um contexto pertinente. O significado convencional, todavia, exclui a possibilidade de interpretar aquilo que se “pretendia dizer” (excluindo-se o caso da mentira), ou seja, pode-se pretender dizer algo mais do que comporta um significado convencional. Para Grice esse é o significado situacional, que, na interpretação de Eco, exige a incorporação de elementos pragmáticos (ECO, 1991, p.73).

Numa representação semântica de L, os significados convencionais são função dos significados lexicais, o que não inclui a possibilidade de representação (para posterior interpretação), de significados situacionais. A representação semântica, nesse sentido, é exclusivamente do tipo dicionarial. E é para que a possibilidade de incorporação dos significados contextuais se dê que Eco trabalha com a hipótese de postular uma língua L que contenha, de alguma forma, “entre as próprias regras de significação instruções pragmaticamente orientadas” (ECO, 1991, p.75). A

possibilidade de constituição de uma tal língua L incorpora, portanto, as possibilidades de interpretação não exclusivamente semânticas — incluindo as significações convencionais e as situacionais, de Grice (ECO, 1991, p. 75-77).

A partir dessas premissas, poderemos afirmar que se uma LD quiser funcionar como uma linguagem instrucional — e se quiser referir — ela não pode ser construída a partir de palavras do léxico, devendo, portanto, buscar referência na Terminologia. A Terminologia, ao referir-se a objetos concretos — ela sim se reporta aos discursos — refere-se às palavras em funcionamento vinculadas a determinados contextos, o que lhe possibilita exercer a função referencial de designação.

Como na maior parte dos casos as LDs não explicitam suas definições com base na Terminologia — suas unidades são muitas vezes construídas à base de palavras preferenciais — as conversões são realizadas utilizando simultaneamente palavras que têm referência no discurso e outras que só constituem palavras do léxico, misturando, dessa forma, predicados livres e predicados relacionados. Dito de outra forma, algumas palavras das LDs são termos — e se reportam à terminologia de determinada área (e, conseqüentemente, aos discursos que lhe deram origem); outras, não constituem termos, podendo desencadear interpretações variadas: “os interpretantes não previstos de início passam a ser articulados quase que livremente, convivendo com hierarquias estabelecidas” (TÁLAMO et alii, p.9).

Pode-se citar como exemplo, “Reforma agrária”, do Thesaurus Popin. Uma vez que essa expressão não é acompanhada de definição e nem de nota de alcance — e nesse sentido, não constitui um termo — ela constitui uma unidade preferencial. Pode-se, portanto, com esse descritor, indexar documentos cuja concepção sobre reforma agrária sejam conflitantes.

Resta, portanto, encontrar mecanismos concretos para a operacionalização de LDs como instrumentos semióticos. A utilização dos casos de Fillmore, pode ser retomada, uma vez que os casos remetem à área de conhecimento, constituindo modos de organização da estrutura profunda. Dizem respeito, portanto, à competência e constituem acessório metodológico para a construção das LDs. Outra sugestão é analisar pormenorizadamente as contribuições da TESWEST de Petöfi (teoria

da estrutura do texto e da estrutura do mundo) (PETÖFI apud ECO, 1991, p.124-126), bem como as apontadas por Desclès e Rastier, desde que representem exemplos de aplicações cuja base é a semântica intensional.

CAPÍTULO 6

AVALIAÇÃO DAS PRINCIPAIS LINGUAGENS DOCUMENTÁRIAS

Neste capítulo, analisaremos, através de exemplos, algumas LDs consideradas significativas, ou em função do seu uso intensivo, ou em função dos princípios que elas supõem. Para efeito da análise, agruparemos as LDs em dois grupos: os sistemas de classificação e os tesouros.

Incluem-se no primeiro grupo a Dewey Decimal Classification, conhecida pela sigla CDD, em português, e a Classificação Decimal Universal - CDU.

No segundo grupo procuraremos analisar algumas características dos Tesouros, tomados como o exemplo mais significativo de LDs de estrutura combinatória. Escolhemos, para o desenvolvimento da análise, trechos do Thesaurus POPIN - Thesaurus Multilingüe de População e do Macrothesaurus para el procesamiento de la información relativa al desarrollo económico y social, das Nações Unidas.

Para facilitar a consulta, agrupamos a reprodução dos exemplos no Anexo 2, v.2 deste trabalho.

6.1 Sistemas de classificação bibliográfica: a Dewey Decimal Classification - CDD e a Classificação Decimal Universal - CDU

6.1.1 Identificação

A Dewey Decimal Classification - CDD é considerada a “primeira classificação bibliográfica no sentido moderno” (FOSKETT, 1973, p.199). Foi inicialmente desenvolvida por Melvil Dewey, em 1876, sendo sucessivamente revisada e ampliada.

Analisamos, no presente trabalho, exemplos da 20a.ed., publicada em 1989, sob a coordenação de J. Comaromi, J. Beall, W.E. Matthews Jr. e G.R. New. A área temática abrangida é geral, ou

seja, o sistema pretende abarcar o conhecimento enciclopédico. Não se dispõe, até o momento, de uma tradução do sistema.

Sistema similar desenvolvido a partir da 5a. edição da CDD (FOSKETT, 1973, p.199) é a Classificação Decimal Universal - CDU, originalmente preparada por Paul Otlet e Henri de La Fontaine em 1895.

A edição analisada é a 2a. ed. média em língua portuguesa publicada pelo IBICT, em 1987. Como a CDD, a área temática pretende abarcar o conhecimento em geral.

6.1.2 Objetivos e princípios de construção

Sistemas datados do final do século XIX, as classificações bibliográficas mais conhecidas — a CDD e a CDU — constituem exemplos de Linguagens Documentárias ainda utilizados intensivamente pelas bibliotecas e centros de documentação brasileiros.

Tais sistemas se apresentam como uma lista estruturada de assuntos que se apresentam como referência para a classificação de documentos, fornecendo elementos para o arranjo de acervos, organização de inventários, catálogos e bibliografias, bem como sua recuperação. A representação dos assuntos é realizada a partir de “notações” ou símbolos compostos por números, letras e sinais.

A CDD, que serviu como base para a construção da CDU, tem como referencial básico a divisão do conhecimento postulada por Bacon para a delimitação das classes. Foskett afirma, entretanto, que a influência de Bacon no desenvolvimento desses sistemas de classificação, mais do que a repartição do conhecimento em três grandes classes — História, Poesia, Filosofia, que derivaram dos processos mentais de Memória, Imaginação e Razão — deve-se ao fato da assimilação de experiências originadas da observação (FOSKETT, 1967, p.97). “A influência que o filósofo exerceu na classificação das ciências não se nota tanto na larga distribuição das classes como na exaustiva compilação e denominação dos fatos — o método da enumeração” (idem *ibidem*).

Ambos os sistemas dividem o conhecimento em dez classes principais, correspondentes à disciplinas, à época, delimitadas. Essa divisão se reporta à adotada por Willian Torrey Harris, para a biblioteca Pública de St. Louis, EUA, em 1870, baseada na inversão de Bacon. (PIEDADE, 1977, p.65).

As classes da CDD e da CDU são subdivididas hierarquicamente utilizando a base decimal, segundo a proposta de André Marie Ampère (1775-1836) (CDU, 1987, p.viii).

A lógica formal aristotélica fundamenta a estruturação de tais sistemas a partir da aplicação sucessiva de princípios de dedução. Semelhante aplicação sucessiva de uma qualidade, atributo ou característica de divisão, dá origem a uma cadeia conceitual hierárquica (gênero/espécie, todo/parte), de forma que cada conceito ou termo deve subordinar-se necessariamente ao seu imediatamente superior, bem como a todos os termos superordenados que lhe deram origem, até o topo da cadeia, que dá nome à classe.

Segundo Foskett, a formação de rubricas individuais é feita “por meio de um processo de pensamento que, a cada divisão, se confina a uma diferenciação que parece inequívoca: a inclusão ou não de uma qualidade postulada” (FOSKETT, 1967, p.96). A aplicação dos ‘predicáveis de Aristóteles’ (gênero, espécie, diferença, propriedade e acidente) é ilustrada pela Árvore de Porfírio, particularmente em relação aos três primeiros predicáveis (FOSKETT, idem ibidem).

6.1.3 Forma de apresentação:

6.1.3.1 DEWEY DECIMAL CLASSIFICATION

A Dewey Decimal Classification, na sua 20a. edição, é apresentada em 4 volumes, conforme segue:

- v.1 Tabelas auxiliares
- v.2 Esquema: classes 000 a 500
- v.3 Esquema: classes 600 a 900
- v.4 Índice relativo

Os “esquemas” constituem o corpo principal da tabela, incluindo as dez classes de assunto (disciplinas) e suas subdivisões sucessivas, que constituem as subclasses e seus detalhamentos.

Pela reprodução abaixo pode-se visualizar a primeira e segunda subdivisões.

Second Summary* *The Hundred Divisions*

000	Generalities	500	Natural sciences & mathematics
010	Bibliography	510	Mathematics
020	Library & information sciences	520	Astronomy & allied sciences
030	General encyclopedic works	530	Physics
040		540	Chemistry & allied sciences
050	General serials & their indexes	550	Earth sciences
060	General organizations & museology	560	Paleontology Paleozoology
070	News media, journalism, publishing	570	Life sciences
080	General collections	580	Botanical sciences
090	Manuscripts & rare books	590	Zoological sciences
100	Philosophy & psychology	600	Technology (Applied sciences)
110	Metaphysics	610	Medical sciences Medicine
120	Epistemology, causation, humankind	620	Engineering & allied operations
130	Paranormal phenomena	630	Agriculture
140	Specific philosophical schools	640	Home economics & family living
150	Psychology	650	Management & auxiliary services
160	Logic	660	Chemical engineering
170	Ethics (Moral philosophy)	670	Manufacturing
180	Ancient, medieval, Oriental philosophy	680	Manufacture for specific uses
190	Modern Western philosophy	690	Buildings
200	Religion	700	The arts
210	Natural theology	710	Civic & landscape art
220	Bible	720	Architecture
230	Christian theology	730	Plastic arts Sculpture
240	Christian moral & devotional theology	740	Drawing & decorative arts
250	Christian orders & local church	750	Painting & paintings
260	Christian social theology	760	Graphic arts Printmaking & prints
270	Christian church history	770	Photography & photographs
280	Christian denominations & sects	780	Music
290	Other & comparative religions	790	Recreational & performing arts
300	Social sciences	800	Literature & rhetoric
310	General statistics	810	American literature in English
320	Political science	820	English & Old English literatures
330	Economics	830	Literatures of Germanic languages
340	Law	840	Literatures of Romance languages
350	Public administration	850	Italian, Romanian, Rhaeto-Romanic
360	Social services; association	860	Spanish & Portuguese literatures
370	Education	870	Italic literatures Latin
380	Commerce, communications, transport	880	Hellenic literatures Classical Greek
390	Customs, etiquette, folklore	890	Literatures of other languages
400	Language	900	Geography & history
410	Linguistics	910	Geography & travel
420	English & Old English	920	Biography, genealogy, insignia
430	Germanic languages German	930	History of ancient world
440	Romance languages French	940	General history of Europe
450	Italian, Romanian, Rhaeto-Romanic	950	General history of Asia Far East
460	Spanish & Portuguese languages	960	General history of Africa
470	Italic languages Latin	970	General history of North America
480	Hellenic languages Classical Greek	980	General history of South America
490	Other languages	990	General history of other areas

*Consult schedules for complete and exact headings

As tabelas auxiliares (v.1) relacionam os complementos que podem ser agregados à notação do assunto principal para a especificação da forma de apresentação do documento ou assunto (conforme o caso), área geográfica a que se refere, gênero literário, línguas individuais, grupos raciais, étnicos e nacionais, idiomas e características pessoais. Sua aplicabilidade varia, conforme a tabela em questão: algumas podem ser utilizadas em todo o sistema (salvo instruções contrárias), como é o caso da tab. 1 e 2; as restantes só podem ser utilizadas vinculadas à classes específicas, a partir de instruções.

tab. 1 - Standard subdivisions;

tab. 2 - Geographic areas, historical periods, Persons;

tab. 3 - Subdivisions for Individual Literatures, for
specific literary forms;

tab. 4 - Subdivisions of Individual Languages;

tab. 5 - Racial, Ethnic, National Groups;

tab. 6 - Languages;

tab. 7 - Groups of Persons.

A notação do sistema é dita “pura”, pois caracteriza-se pelo uso exclusivo de números para a composição dos símbolos de classificação.

Deve-se observar que a 20a. edição é a primeira a apresentar um manual de uso, que integra o v.4., juntamente com o índice do sistema.

6.1.3.2 CLASSIFICAÇÃO DECIMAL UNIVERSAL

A edição analisada da Classificação Decimal Universal, em português, é apresentada em 2 volumes:

v.1 Tabelas auxiliares e principais;

v.2 Índice remissivo.

Diferentemente da CDD, a CDU não possui sumários que orientem sua consulta. No início do volume, após a introdução e as instruções para utilização, são relacionadas as tabelas auxiliares e seus respectivos sinais.

- + Adição;
- / Extensão;
- [] Subagrupamento;
- :: Ordenação;
- = Auxiliares comuns de língua;
- (...0) Auxiliares comuns de forma;
- (...) Auxiliares comuns de lugar;
- (=...) Auxiliares comuns de raças e nacionalidades;
- "..." Auxiliares comuns de tempo
- A/Z Especificação de assunto por meio de notações que não pertencem à CDU;
- .00 Auxiliares comuns de ponto de vista;
- .03 Auxiliares comuns de materiais;
- .05 Auxiliares comuns de pessoas e características pessoais;

Além dessas tabelas auxiliares, há subdivisões especiais especificadas por .0, -1/9 ou *, conforme o caso, distribuídas no interior das tabelas auxiliares e principais.

A aplicabilidade das tabelas auxiliares (exceto as especiais) é válida para todo o sistema, desde que façam sentido. Contrariamente às tabelas auxiliares da CDD, algumas tabelas auxiliares da CDU podem ser utilizadas independentemente, sem acompanhar a notação de um assunto principal.

Em seguida são relacionadas as tabelas principais, semelhantes, quanto à sua apresentação básica, à CDD, exceto pela ausência dos dois zeros que sucedem o número de cada classe principal e pela inexistência da classe 4 (400 na CDD, reservada à Lingüística e Filologia) integrada à classe 8, Literatura.

A notação da CDU é mista, caracterizada pelo uso de números, letras e sinais para a composição

dos símbolos de classificação.

Na ausência de um sumário mais detalhado, reproduzimos abaixo a relação das classes principais:

- 0 Generalidades. Ciência e conhecimento.
 - 1 Filosofia. Psicologia.
 - 2 Religião. Teologia.
 - 3 Ciências sociais.
 - 5 Matemática e ciências naturais.
 - 6 Ciências aplicadas. Medicina. Tecnologia.
 - 7 Arte. Belas-artes. Recreação. Diversões. Esportes.
 - 8 Linguagem. Lingüística. Literatura.
 - 9 Geografia. Biografia. História.
- (CLASSIFICAÇÃO DECIMAL UNIVERSAL, 1987, p.vi)

6.1.4 Análise de exemplos

Exemplo 1

Via de regra, a estruturação hierárquica é obedecida, utilizando-se a inclusão como critério para a formação das árvores conceituais do sistema de classificação, mantendo-se, conseqüentemente, o princípio da subordinação.

Nas subdivisões .1 a .8 do exemplo acima, dominam os sistemas socialistas, enquanto "modo" de realização de organização de sociedades. O princípio é quebrado, entretanto, no item .9, que diz respeito ao "produto" concreto de um determinado sistema: tipos de comunidades socialistas.

O "0" antecedendo o numeral em .02 (normalmente reservado às subdivisões de forma, na CDD), representa um artifício utilizado pelo sistema para a enumeração de aspectos adicionais de um determinado assunto, quando a subdivisão na base 10 não é suficiente para dar conta de toda a enumeração necessária.

Revela-se, por outro lado, neste exemplo, a orientação ideológica do sistema, que inclui, sob o número 335 "Socialismo", escolas utópicas.

Exemplo 2

Observa-se, neste caso, que a junção de vários assuntos no item 613.7 implica uma série não-coordenada logicamente. Dessa maneira, as relações de cada termo com a classe superordenada variam segundo a faceta privilegiada. A hierarquia é, nesse sentido, descaracterizada. “Cultura física”, por exemplo, não se relaciona a “Sono”, ao mesmo tempo que “Feriados” pode ser abordado como “Lazer”, “Recreação” ou “Sono”, conforme os aspectos passíveis de serem analisados. Neste último caso ocorre uma remessa anárquica à definição suposta no termo superordenado.

Exemplo 3

Neste caso, misturam-se “modo” (processos) e “produto”: 637.11 “Ordenha”, é um processo. Já os tipos de leite relacionados em .12, .13 e .14 são produtos.

Observa-se também, que em relação à 637, a série 637.1, .2 e .3 não é coordenada, uma vez que “Manteiga” e “Queijos” são derivados de leite, devendo, segundo um esquema lógico, constar como subdivisão de 637.1.

Explorando a mesma classe pode-se observar que em 637.3 “Queijos e sua fabricação” (que já representa um amálgama de produto e processo), .33 trata dos “Processos de fabricação do queijo, tendo como assuntos subordinados .334.2 “Coalho, coagulação”.

Os procedimentos apresentados acima podem comprometer a elaboração dos índices para catálogos sistemáticos (22), uma vez que a sua montagem exige a recuperação da cadeia conceitual do termo (do mais específico — número mais longo — ao mais genérico). Se a hierarquia está mal construída, o fato se refletirá no índice correspondente.

(22) Catálogos formados por um índice alfabético e um índice numérico. Cada entrada no índice alfabético remete a um o número de classificação correspondente à palavra ou frase usada como cabeçalho de assunto. O índice numérico, por sua vez, é organizado na seqüência prescrita pelo sistema de classificação utilizado.

Exemplo 4

Evidencia-se, neste caso, a junção de gêneros de natureza diferente. Sob a notação 656.1 a .5 são discriminados os tipos de transporte terrestre; sob .6 e .7, os outros tipos de transporte: por água e por ar, respectivamente. A hierarquia deveria agrupar, sob .1, todas as subdivisões de transporte terrestre. Não o fazendo, ficam incorretamente coordenados, por exemplo, “Transporte ferroviário” e “Transporte por água”.

Já sob a subdivisão .8 são relacionados os serviços de transporte postal, fato que introduz novo gênero na cadeia de transporte: de .1 a .7 são relacionados os tipos de transporte segundo o meio e a forma de locomoção; sob .8, segundo a função.

Deve-se lembrar que o último item pode ser relacionado aos anteriores: transporte postal pode ser feito por via terrestre, aérea ou aquática (a CDU permite que tais relacionamentos sejam feitos através do uso do sinal próprio), mas isso não resolve o problema do viés conceitual existente na hierarquia construída.

Segundo os princípios da Lógica, o esquema deveria organizar, hierarquicamente, sob o termo “Transportes”, “Transporte Terrestre”, “Transporte por água” e “Transporte por ar”; sob o termo “Transporte Terrestre” deveriam estar subordinados os vários tipos. Ao agrupar sob o número 356.1/.5 os vários tipos de transporte terrestre, a CDU realiza uma falsa coordenação, ou seja, permite estabelecer uma equivalência, por exemplo, entre “Transporte ferroviário” e “Transporte por água”, quando o correto seria “Transporte terrestre” e “Transporte por água”.

Exemplo 5

A CDD, neste exemplo, separa “Life Sciences” de “Botanical Sciences” e “Zoological Sciences”, o que representa uma inconsistência do ponto de vista lógico e conceitual. Como recurso para analisar, em Botânica ou Zoologia, aspectos relacionados à Ciências da vida, a CDD prevê mecanismos que permitem a utilização de subdivisões de 570 nas classes 580 e 590 (síntese notacional).

Exemplo 6

Este exemplo revela a concepção subjacente à CDD, discutível atualmente, de separar Antropologia Cultural (no âmbito das Ciências Sociais) da Antropologia Física (no âmbito das Ciências Naturais). Este (e muitos outros exemplos) remetem à concepções cristalizadas, o que dificulta e compromete a classificação de documentos recentes.

Exemplo 7

Evidencia-se, neste exemplo, a enumeração simples, que relaciona, numa mesma hierarquia, termos de categoria diferentes: “instrumento” e “modo”, principalmente. Observe-se, entre outros, sob esse aspecto, 303.623 “Estratégia da entrevista” como um termo que pode ser coordenado com .624 e .625 “Entrevistador” e “Entrevistado”, respectivamente.

Exemplo 8

Transparece, no exemplo acima, o privilégio que a CDD dá às religiões cristãs (para as quais foram reservadas as subdivisões 200, 220 a 280) face às outras religiões, agrupadas sob a notação 290.

Exemplo 9

A subdivisão 025 de Biblioteconomia na CDU revela uma opção pelo enfoque tradicional das atividades biblioteconômicas que são relacionadas de forma estanque, como setores administrativos de uma biblioteca. Observe-se que o título da subdivisão é “Departamentos Administrativos das Bibliotecas”, relacionando, entre outros, os processos de Seleção, Aquisição, Catalogação, Classificação, Indexação, Referência etc.

As hierarquias não são logicamente construídas, uma vez que amalgamam aspectos das categorias “modo” e “instrumento”, como por exemplo em 025.4, Classificação e Indexação. Linguagens de indexação e recuperação etc.

Exemplo 10

Sob a notação 150.19, a CDD oferece um exemplo de hierarquia lógica, agrupando nas subdivisões os vários sistemas e escolas da Psicologia. Nota-se que a discriminação das tipologias não apresenta, em geral, grandes problemas em todo o sistema.

Deve-se observar, entretanto, que a classe “Psychology” foi incluída na classe Filosofia (100), revelando, simultaneamente, uma vinculação hoje carente de sentido e problemas derivados da rigidez esquemática da divisão do conhecimento em 10 classes ou disciplinas. Uma das decorrências do fato é a existência de notações excessivamente longas para designar quaisquer aspectos da disciplina.

6.1.5 Avaliação geral

6.1.5.1 Estrutura das LDs

Pela análise dos exemplos pode-se concluir que domina na CDD e na CDU, o uso de mais de uma característica de divisão para a construção das hierarquias, implicando, conseqüentemente, na coordenação de termos de natureza diferente e na possibilidade de remessa anárquica à definição suposta pelo termo superordenado.

Esse problema foi identificado por Coyaud, para quem a CDU “se contenta em amalgamar diversas relações em uma só relação de ‘inclusão’ ” (COYAUD, 1966, p.36). Sob esse mesmo aspecto, o autor também aponta a heterogeneidade da hierarquia dos conceitos, que coloca, sob o mesmo plano, noções de níveis de generalidade diversos (idem ibidem).

O uso de facetas de natureza diferente para a enumeração dos assuntos, por seu lado, tem como conseqüência o amálgama de traços diversos (processos e produtos, ações, entidades etc.), na mesma hierarquia, resultando no comprometimento das cadeias conceituais do sistema.

Conforme Coyaud, Kervegant propôs que se agregasse à CDU um sistema claro de relações (Pertinência, Processos, Dependência, Orientação, Comparação), que, entretanto, não foi aceito pelo Comitê responsável pela sua edição (COYAUD, 1966, p.37).

A CDD e a CDU não são sistemas flexíveis, uma vez que se utilizam de grades rígidas para a análise dos assuntos, fato que resulta na compartimentação das disciplinas e dificulta, por isso, a análise dos aspectos onde domina a interdisciplinaridade. A grade rígida é um problema comum aos dois sistemas de classificação, fruto da organização do conhecimento em 10 disciplinas delimitadas originalmente.

Como decorrência, a inclusão de novas disciplinas é feita de forma discutível, através de subordinações nem sempre lógicas e coerentes. Na verdade, ambos os sistemas apresentam sua marca teórica e ideológica, consubstanciada nas classes rigidamente organizadas e na manutenção de pontos de vista cristalizados.

Sob o ponto de vista da organização e desenvolvimento dos sistemas pode-se observar que a noção de indução baconiana —segundo a qual deve-se tomar como base a observação metódica dos fatos da experiência para a formação das séries e ordens (BACON, 1973, p.20) — combinada ao princípio de que “o conhecimento e o domínio do homem sobre o meio que o cerca aumentam na medida em que cresce, também, a sua observação da ‘Ordem natural’” (FOSKETT, 1967, p.89) foi assimilada como simples processo de enumeração (ver exemplo 1, 2, 3, 4 etc.).

Deve-se acrescentar que, ao compartilhar a concepção de progresso dependente da observação da “ordem natural” do universo físico, tais sistemas de classificação assimilam a concepção do conhecimento condicionado pelo tempo cronológico. As noções derivadas a partir da sucessão de observações da ordem natural, por seu lado, tem como implicação a crença em um movimento infinito do conhecimento em direção à verdade absoluta. Pode-se afirmar, daí, que os conceitos decorrentes se aperfeiçoam no tempo, e mesmo o aforismo baconiano que descarta a noção de progresso nas ciências pela “superposição ou enxerto do novo sobre o velho” (BACON, 1979, af.31, p.19) não é suficiente para derrubar a crença na depuração das noções através do tempo.

Como decorrência, tais sistemas de classificação mantêm a estrutura inicial de divisão do conhecimento agregando, nas sucessivas edições, novas enumerações dentro de cada classe, sem rever o seu conjunto face às novas configurações conceituais.

Por outro lado, o uso simultâneo de mais de uma categoria de divisão (mais de uma disjunção) para a construção das hierarquias e o uso simultâneo de relações lógicas (do tipo gênero/espécie) e relações associativas de outra natureza, impedem que se afirme que o sistema é baseado na lógica aristotélica, como o faz a maior parte da literatura na área. De fato, verifica-se o uso simultâneo de princípios de implicação estrita e de associação livre não previstos pela lógica aristotélica, demonstrando que domina, na organização desses sistemas de classificação, a base empírica do registro.

Tais ocorrências apontam para dois tipos de problema, de natureza diferente: por um lado, o uso inadequado da lógica formal para a construção de hierarquias; por outro, a própria ineficácia de um sistema que só comporta relações de implicação estrita, uma vez que o registro do conhecimento empírico não se reduz a relações de implicação. Nesse sentido, o limite imposto pelas dez classes e pela subdivisão decimal faz com que relações de natureza lógica passem a conviver com relações de natureza semântica.

6.1.5.2 Eficiência para o controle de vocabulário

Uma LD, como afirmamos anteriormente, tem como uma de suas funções, o controle do significado do vocabulário. Os elementos integrantes de um sistema de classificação deveriam, portanto, assegurar a possibilidade de remissão unívoca aos conceitos representados. Entretanto não é isso que ocorre com os sistemas de classificação analisados, uma vez que as expressões documentárias são constituídas de palavras simples ou frases, retiradas da LN, podendo induzir, conseqüentemente, a interpretações diversas.

Nos sistemas analisados não há definição formal das expressões: a definição é, via de regra, deduzida a partir de sua inserção lógica na cadeia conceitual da classe. A definição (pressuposta)

fica, portanto, limitada às relações de inclusão. A indicação do âmbito da aplicabilidade das expressões é limitada a alguns casos. Por outro lado, há um comprometimento da definição (pressuposta) em razão do uso aleatório de outras relações associativas na mesma cadeia hierárquica.

6.1.5.3 Eficiência para a elaboração de índices

Uma LD deve fornecer elementos para a construção de índices remissivos. A função de uma expressão documentária é a de possibilitar a representação de informações presentes nos documentos ou, pelo menos, como prescrevem os sistemas de classificação, tornar viável a reunião de referências de documentos sobre assuntos similares.

Dado o uso intensivo de expressões retiradas da LN e considerada a estruturação dessas expressões nos sistemas analisados, a única forma de índice coerente possível é o sistemático (índice que recompõe a cadeia conceitual a partir da notação mais específica, para a mais geral, como já foi observado). O índice sistemático vincula os termos de entrada às definições (pressupostas) dos termos superordenados, sucessivamente até o termo da classe mais geral.

Verifica-se, entretanto, que a ausência de rigor na construção das cadeias conceituais dos sistemas de classificação analisados (uso simultâneo de relações de inclusão e de associação livre, ausência de elos na cadeia, presença de elos falsos etc.), acaba por gerar problemas na construção dos índices, que passam a refletir os defeitos dos sistemas.

Por outro lado, na ausência de termos normalizados nos sistemas de classificação, a construção do índice sistemático é feita através da escolha empírica das expressões que constituirão as entradas no catálogo, supondo quais delas poderão ocorrer ao usuário na busca da informação. Predomina, portanto, o uso de vocabulário livre para a recomposição das cadeias e o estabelecimento de relações de associação aleatórias.

Observa-se que a prática bibliotecária, dada a inconsistência do vocabulário dos dois sistemas,

obriga o uso de outros vocabulários sistemáticos para a construção do índice. Abandona-se o sistema de relações prescrito pelos sistemas em prol do uso de “Listas de Cabeçalhos de Assunto”, construídas externamente à CDD e à CDU. (Por exemplo, a Library of Congress Subject Headings, que é, por sua vez, traduzida aleatoriamente pelas várias Bibliotecas). Constroem-se, geralmente, a partir das Listas de Cabeçalhos de Assunto, índices alfabéticos que compõem um catálogo dicionário, abandonando, dessa maneira, qualquer vinculação com o sistema de classificação inicialmente adotado.

Acresça-se, ainda, que o produto documentário (índice) obtido apresenta uma articulação pobre, ou seja, as expressões são desnaturalizadas, podendo remeter a interpretantes diversos ou condicionando a interpretação às associações previstas pelo analista que elaborou o índice.

A articulação passível de ser visualizada no catálogo sistemático, por sua vez, acarreta a reprodução do parâmetro teórico e ideológico dos sistemas, uma vez obedecidas suas hierarquias. O uso de grades rígidas para interpretação das mensagens informativas dos textos implica a perpetuação de paradigmas e cristalização do conhecimento já configurado por esses sistemas de classificação.

6.1.5.4 Capacidade para representar informação

Os sistemas de classificação não foram construídos para representar a informação, mas sim assuntos. Esse limite desautoriza uma análise que extrapole os objetivos do sistema. Na prática, entretanto, tais instrumentos são utilizados na tentativa de abarcar também esta função, o que nos permite levantar algumas observações.

Nos sistemas de classificação, a potencialidade de representação da informação depende de sua capacidade de gerar novas relações entre os termos: além das relações analíticas geralmente previstas por tais instrumentos, eles deveriam contar com mecanismos que permitissem a síntese entre os elementos já existentes. Observa-se, entretanto, que predomina nesses sistemas, relações de subordinação (inclusão) que caracterizam as relações hierárquicas, amalgamadas a outros tipos de relações de associação de natureza não especificada.

A síntese (combinação) entre assuntos de classes distintas é restrita, na CDD, às indicações da tabela, sendo mais ampla na CDU, dada a possibilidade de uso de sinais específicos (+, /, :) (23). Ambos os sistemas prevêm síntese para a complementação da notação do assunto principal, consubstanciada a partir das tabelas auxiliares (forma de apresentação do documento, do assunto, agregação geográfica, língua etc.).

Nos dois sistemas, uma mesma noção pode estar ligada a mais de uma classe, sendo seu significado deduzido segundo sua inserção contextual na classe.

Por outro lado, de um modo global, a não determinação do domínio (sistemas enciclopédicos) dificulta a representação de noções específicas e a ausência de um corpo de definições compromete a univocidade da representação de conceitos.

Tanto em relação à CDD como à CDU pode-se afirmar que representa-se o código (quando da reprodução da cadeia conceitual no caso dos catálogos sistemáticos) ou remete-se a uma noção extremamente genérica, através do uso de uma Lista de Cabeçalhos de Assunto. Em ambos os casos, trata-se de “reprodução”, não de “representação”.

A transcodificação operada a partir dos sistemas se faz, via de regra apenas, pela forma de expressão. Assim, se um texto apresenta a palavra “Planejamento familiar”, a classificação se faz pela equivalência lexical, uma vez que os sistemas relacionam “assuntos” e não prevêm a possibilidade de abordar as variações conceituais expressas pelos termos nos textos.

A possibilidade de classificação é, nesse sentido, extremamente generalizante e o nível de leitura exigido para realizar a operação de classificação é superficial: compromete-se a representação da especificidade e perde-se totalmente os vínculos com a informação presente no documento original.

(23) + (adição): “(...) liga dois ou mais números separados (não consecutivos), para indicar um assunto composto para o qual não existe um número simples (...)” (CLASSIFICAÇÃO DECIMAL UNIVERSAL, 1987, p.1).
 / extensão consecutiva): “(...) liga o primeiro e o último número de uma série de números consecutivos, para indicar um assunto amplo ou uma série de conceitos (...)” (idem ibidem).
 : (relação simples): “(...) indica uma relação geral, coordenada e recíproca, na qual A:B e B:A têm o mesmo valor (...)” (idem ibidem).

6.1.5.5 Eficiência na comunicação documentária

Via de regra, o usuário do sistema de informação que utiliza a CDD ou a CDU não tem acesso aos códigos de classificação, dada a sua arquitetura sofisticada, e mesmo supondo essa possibilidade, ele contaria com um instrumento pouco eficiente como sistema de significação.

Os produtos documentários obtidos a partir desses sistemas caracterizam-se pela existência de articulação pobre e de expressões desnaturalizadas.

A ausência de “ajuste” para a especificação de novas relações de significado implica, por outro lado, a inexistência de um vocabulário dinâmico.

O usuário (inclusive o usuário documentalista) não conta com um sistema de referência explícito que funcione como mecanismo de transcodificação: o sistema de significação se deduz dele mesmo e a interpretação fica praticamente livre e indeterminada.

A ausência de um sistema de referência explícito e a ausência de definições implica a impossibilidade de univocidade conceitual, o que induz à subjetividade e à ambiguidade, comprometendo, conseqüentemente, a circulação de informações.

Do ponto de vista institucional, deve-se considerar que os objetivos de classificação diferem de uma instituição para outra e os sistemas de classificação, cujos objetivos não são claros, são utilizados para atender necessidades múltiplas e variáveis. Sendo assim, a ampla aplicabilidade do sistema acaba por comprometer o atendimento a demandas específicas, impedindo, desse modo, a otimização da comunicação de objetivos.

6.2 OS TESAUROS: o Thesaurus POPIN e o MACROTHESAURUS

6.2.1 Identificação

O Thesaurus POPIN - Thesaurus Multilingüe de População, foi traduzido a partir da segunda edição do Population Multilingual Thesaurus, publicado pela primeira vez em 1979 pelo Comitê para a Cooperação Internacional em Pesquisa Nacional em Demografia (CICRED), sob a coordenação de Jean Viet. A área temática nuclear é relativa à “População” sob o enfoque demográfico.

A tradução e adaptação foi feita em 1986 por Leticia Borges Costa, da Fundação SEADE e inclui descritores utilizados pelo DOCPOP - Sistema de Documentação em População no Brasil e pelo DOCPAL - Documentación sobre Población de America Latina y Caribe.

O Macrothesaurus for Information Processing in the Field of Economic and Social Development foi também coordenado por Jean Viet. A edição analisada é trilingüe (inglês/francês/espanhol) e foi publicada em 1985 pela United Nations. A área temática desse tesouro é a de “Informação sócio-econômica para o planejamento”.

6.2.2 Objetivos e princípios de construção

A origem dos tesauros se reporta à Colon Classification - CC, de Ranganathan e às experiências posteriores desenvolvidas pelo Classification Research Group - CRG (24), referentes à estruturação

(24) A CC, desenvolvida pelo indiano Ranganathan, em 1933, é um sistema de classificação enciclopédico que divide o conhecimento em classes básicas de assunto. A análise dos assuntos presentes nos documentos é realizada a partir da manifestação das categorias (ou facetas) Personalidade, Matéria, Energia, Espaço e Tempo. (Ex.: Fabricação (E) de luvas (P) de lã (M) em Bradford (S) no séc.19 (T)). A importância da CC não está no sistema em si, mas na forma de abordagem dos assuntos: um mesmo assunto pode representar a manifestação de uma determinada categoria numa classe, e de outra categoria numa outra classe. (Ex.: Livro, em Biblioteconomia é Matéria; em Bibliografia é Personalidade). Uma mesma categoria pode ocorrer mais de uma vez numa mesma classe básica. O sistema, entretanto, condiciona a leitura dos assuntos às disciplinas pré-fixadas. (Para detalhes, ver MANIEZ, 1976; FOSKETT, 1973). Já as experiências do CRG procuram flexibilizar a noção de faceta a partir da maior delimitação de domínios e subdomínios específicos. Exemplo significativo é a Classificação das Ciências do Solo, de Vickery, onde estes são observados a partir de sua Tipologia, Estrutura, Constituintes, Propriedades, Processos, Operações, Técnicas de laboratório (VICKERY, 1980).

das áreas do conhecimento a partir da noção de “faceta”, ou seja, do privilégio de determinados pontos de vista sobre o arranjo dos domínios e sub-domínios específicos.

Entre as finalidades destacadas pelo IBICT, um tesouro é um instrumento utilizado para:

- a) “controlar os termos usados na indexação(…)” através de uma “linguagem mais controlada(…)”
- b) “assegurar ... uma prática consistente entre diferentes indexadores (...)”
- c) “limitar o número de termos necessários atribuídos aos documentos (...) que devem representar, tão especificamente quanto possível, os conceitos expostos pelo autor, sem que haja necessidade de incluir termos de conotação mais ampla e demais termos associados, se tais termos estiverem implicados nos quadros normais de referência;”
- d) “servir como auxiliar de busca na estratégia de recuperação, inclusive em sistemas de texto livre.” (IBICT, 1984, p.1-2).

A fonte de referência para a construção dos tesouros é a estrutura teórico-conceitual de domínios específicos, sendo que os instrumentos incluem termos do domínio nuclear — área de especialização propriamente dita — e de domínios periféricos — áreas complementares.

No caso do Thesaurus POPIN, não há referência explícita aos princípios de classificação adotados, embora fique clara a preocupação de compatibilizar os vários instrumentos existentes na área de população, considerando, também, a proximidade temática com o Macrothesaurus (VIET, 1986, p.5).

A classificação adotada pelo Macrothesaurus reflete, simultaneamente, a estrutura geral do desenvolvimento sócio-econômico e os vários aspectos de especialização concernentes às atividades das agências de desenvolvimento (VIET, 1985, p.xi).

Nos dois tesouros analisados, os domínios são estruturados hierarquicamente, como nos sistemas de classificação anteriormente analisados. Da mesma forma, a hierarquia (com o predomínio das relações gênero/espécie, todo parte) é utilizada para a organização dos termos (denominados “descritores”), sendo que adicionalmente são apresentadas outras relações de associação entre os termos (relações de equivalência e associações de natureza não especificada).

Segundo o IBICT, em relação à sua estrutura, “o tesouro patenteia as relações vigentes entre os termos — sinonímicas, hierárquicas e outras — que, no seu conjunto, constituem a linguagem de indexação” (IBICT, 1984, p.2).

Para indicar o gênero de relações entre os termos, os tesouros analisados utilizam os símbolos e abreviaturas especificados a seguir:

Relações hierárquicas:

TT Top Term (Termo genérico mais abrangente; em português, TGM); no nosso caso, o TT é usado apenas pelo Macrothesaurus;

BT Broader Term (Termo genérico; em português, TG);

NT Narrower Term (Termo específico; em português, TE);

Relações de equivalência:

UF Used for (Usado para; em português, UP);

Use (Use);

SN Scope Note (Nota de escopo; em português, NS)

Relações associativas não-hierárquicas:

TR Related Term (Termo relacionado; em português, TR).

OBS.: nos dois tesouros analisados um mesmo termo pode estar ligado a mais de uma hierarquia, ao mesmo tempo.

6.2.3 Forma de apresentação

6.2.3.1 Thesaurus POPIN

Inicialmente, o POPIN apresenta a “Estrutura Agregada do Thesaurus”, que representa a estrutura geral de classificação adotada e a especificação das áreas e sub-áreas temáticas

acompanhadas de notação numérica de facetas.

- 01.00.00**
DEMOGRAFIA. PESQUISA. METODOLOGIA
- 01.01.00** MATEMÁTICA. CIÊNCIAS NATURAIS. CIÊNCIAS SOCIAIS
- 01.02.00** CENTROS DE PESQUISA
- 01.03.00** PESQUISA
- 01.04.00** METODOLOGIA
- 01.05.00** CENSOS. LEVANTAMENTOS
- 01.06.00** REGISTROS
- 01.07.00** MÉTODOS DE ANÁLISE
- 01.08.00** MODELOS
- 01.09.00** ANÁLISE ESTATÍSTICA
- 01.10.00** ESTIMATIVAS. PREVISÕES. PROJEÇÕES
- 02.00.00**
POPULAÇÃO: IDADE. SEXO. GRUPOS ÉTNICOS
- 02.01.00** POPULAÇÃO
- 02.02.00** IDADE
- 02.03.00** SEXO
- 02.04.00** GRUPOS ÉTNICOS
- 02.05.00** NACIONALIDADE
- 03.00.00**
FAMÍLIA. DOMICÍLIOS
- 03.01.00** DOMICÍLIOS
- 03.02.00** FAMÍLIA
- 04.00.00**
DISTRIBUIÇÃO DA POPULAÇÃO
- 04.01.00** HABITAT
- 04.02.00** MEIO AMBIENTE. RECURSOS NATURAIS. ALIMENTAÇÃO
- 04.03.00** ZONAS URBANAS
- 04.04.00** ZONAS RURAIS
- 04.05.00** RESIDÊNCIA
- 04.06.00** HABITAÇÃO
- 05.00.00**
ORGANIZAÇÃO SOCIAL
- 05.01.00** SOCIEDADE. SISTEMAS SOCIAIS
- 05.02.00** GRUPOS
- 05.03.00** ESTRUTURA SOCIAL
- 05.04.00** CONFLITOS. MUDANÇA SOCIAL. PROBLEMAS SOCIAIS
- 05.05.00** CONTROLE SOCIAL
- 05.06.00** RELIGIÃO
- 06.00.00**
CULTURA. EDUCAÇÃO. INFORMAÇÃO
- 06.01.00** CULTURA
- 06.02.00** EDUCAÇÃO
- 06.03.00** COMUNICAÇÃO
- 06.04.00** INFORMAÇÃO. DOCUMENTAÇÃO
- 07.00.00**
FORÇA DE TRABALHO. EMPREGO
- 07.01.00** MÃO-DE-OBRA. VIDA ATIVA
- 07.02.00** MERCADO DE TRABALHO
- 07.03.00** OCUPAÇÕES
- 07.04.00** TRABALHO. CONDIÇÕES DE TRABALHO
- 08.00.00**
ECONOMIA
- 08.01.00** CIÊNCIAS ECONÔMICAS
- 08.02.00** CONDIÇÕES ECONÔMICAS
- 08.03.00** PRODUÇÃO
- 08.04.00** FINANCIAMENTO. CUSTOS
- 08.05.00** MERCADO. CONSUMO
- 08.06.00** RENDA. PREVIDÊNCIA SOCIAL
- 08.07.00** SISTEMAS ECONÔMICOS
- 08.08.00** FINANÇAS PÚBLICAS
- 08.09.00** RELAÇÕES ECONÔMICAS. COMÉRCIO INTERNACIONAL

09.00.00
ADMINISTRAÇÃO. LEGISLAÇÃO. POLÍTICAS GOVERNAMENTAIS

- 09.01.00** SISTEMAS POLÍTICOS. VIDA POLÍTICA
- 09.02.00** ADMINISTRAÇÃO. GOVERNO
- 09.03.00** DIREITO. LEGISLAÇÃO
- 09.04.00** DIREITOS HUMANOS. LEGITIMIDADE
- 09.05.00** POLÍTICAS GOVERNAMENTAIS
- 09.06.00** RELAÇÕES INTERNACIONAIS
- 09.07.00** ORGANIZAÇÕES INTERNACIONAIS

10.00.00
FATORES PSICOLÓGICOS

- 10.01.00** PSICOLOGIA
- 10.02.00** PERSONALIDADE. MOTIVAÇÃO
- 10.03.00** ATITUDES. IDEOLOGIAS
- 10.04.00** PRECONCEITO RACIAL
- 10.05.00** PAPEIS MASCULINOS. PAPEIS FEMININOS
- 10.06.00** SEXUALIDADE

11.00.00
VIDA. SAÚDE

- 11.01.00** BIOLOGIA
- 11.02.00** MEDICINA. SERVIÇOS DE SAÚDE

12.00.00
MORBIDADE. MORTALIDADE

- 12.01.00** DOENÇAS
- 12.02.00** DOENTES
- 12.03.00** DURAÇÃO DA VIDA
- 12.04.00** MORTALIDADE

13.00.00
NUPCIALIDADE

- 13.01.00** NUPCIALIDADE
- 13.02.00** ESTADO CIVIL
- 13.03.00** ESCOLHA MARITAL. CERIMÔNIA MATRIMONIAL
- 13.04.00** TIPOS DE UNIÃO
- 13.05.00** UNIÃO MARITAL
- 13.06.00** SEPARAÇÃO. DIVÓRCIO

14.00.00
FECONDIDADE. PLANEJAMENTO FAMILIAR

- 14.01.00** FECUNDIDADE
- 14.02.00** PLANEJAMENTO FAMILIAR
- 14.03.00** ANTICONCEPÇÃO
- 14.04.00** ABORTO

15.00.00
MIGRAÇÃO

- 15.01.00** MIGRAÇÃO. SALDO MIGRATÓRIO
- 15.02.00** FLUXOS MIGRATÓRIOS
- 15.03.00** MIGRAÇÃO SAZONAL. MIGRAÇÃO DO TRABALHO
- 15.04.00** MIGRAÇÃO FORÇADA. MIGRAÇÃO CONTROLADA
- 15.05.00** TURISMO. VIAGENS

16.00.00
DINÂMICA DEMOGRÁFICA. TAXAS DE REPRODUÇÃO

- 16.01.00** DINÂMICA DEMOGRÁFICA
- 16.02.00** DESPOVOAMENTO. CRESCIMENTO DEMOGRÁFICO
- 16.03.00** TAMANHO DA POPULAÇÃO

17.00.00
TEORIA DE POPULAÇÃO

18.00.00
PAÍSES E REGIÕES

- 18.01.00** ÁFRICA
- 18.02.00** AMÉRICA
- 18.03.00** ÁSIA
- 18.04.00** EUROPA
- 18.05.00** ÁSIA OCIDENTAL
- 18.06.00** OCEANIA

O corpo principal desse tesouro — “Apresentação Temática dos Descritores” — é constituído por grupos de facetas, sendo que os termos no interior de cada faceta são relacionados em ordem alfabética. Sob cada descritor são apresentados os termos (descritores e não-descritores) com os quais ele mantém algum tipo de relacionamento.

Para facilitar a consulta, integra o Thesaurus POPIN um “Índice Permutado” que relaciona descritores e não-descritores em ordem alfabética a partir de todas as suas palavras significativas (ou seja, excluindo-se artigos e preposições).

6.2.3.2 MACROTHESAURUS

Diferentemente do Thesaurus POPIN, o corpo principal do Macrothesaurus é apresentado em ordem alfabética — o “Alphabetical Thesaurus”. Nessa listagem os descritores são acompanhados dos respectivos números de faceta e dos descritores com os quais ele se relaciona.

Em seguida, a “Subject category fields” apresenta a árvore classificatória básica do tesouro — reproduzido em seguida — que é apresentada mais detalhadamente no “Descriptor Group Display”, segundo grupos temáticos.

01. INTERNATIONAL COOPERATION. INTERNATIONAL RELATIONS.

- 01.01 INTERNATIONAL COOPERATION.
- 01.02 INTERNATIONAL RELATIONS.
- 01.03 INTERNATIONAL ORGANIZATIONS.
- 01.04 COUNTRIES AND REGIONS.

02. ECONOMIC POLICY. SOCIAL POLICY. PLANNING.

- 02.01 ECONOMIC POLICY. PLANNING.
- 02.02 SOCIAL POLICY.
- 02.03 SOCIAL SECURITY.
- 02.04 SOCIAL PROBLEMS.
- 02.05 SOCIAL SERVICES.

03. ECONOMIC CONDITIONS. ECONOMIC RESEARCH. ECONOMIC SYSTEMS.

- 03.01 ECONOMIC RESEARCH. ECONOMICS.
- 03.02 ECONOMIC CONDITIONS.
- 03.03 ECONOMIC SYSTEMS.

04. INSTITUTIONAL FRAMEWORK.

- 04.01 LAW. LEGISLATION.
- 04.02 HUMAN RIGHTS.
- 04.03 GOVERNMENT. PUBLIC ADMINISTRATION.
- 04.04 POLITICS.

05. CULTURE. SOCIETY.

- 05.01 SOCIAL SCIENCES.
- 05.02 CULTURE.
- 05.03 SOCIETY.
- 05.04 ETHICS. RELIGION.
- 05.05 ART.
- 05.06 LANGUAGES.
- 05.07 COMMUNICATION.

06. EDUCATION. TRAINING.

- 06.01 EDUCATIONAL SCIENCES.
- 06.02 EDUCATIONAL DEVELOPMENT. EDUCATIONAL POLICY.
- 06.03 EDUCATIONAL SYSTEMS.
- 06.04 EDUCATIONAL INSTITUTIONS.
- 06.05 CURRICULUM. TEACHING. LEARNING.
- 06.06 STUDENTS. TEACHING PERSONNEL.

07. AGRICULTURE.

- 07.01 AGRICULTURAL ECONOMICS.
- 07.02 LAND ECONOMICS.
- 07.03 AGRICULTURAL ENTERPRISES.
- 07.04 AGRICULTURAL EQUIPMENT.
- 07.05 AGRICULTURAL PRODUCTION.
- 07.06 AGRICULTURAL RESEARCH.
- 07.07 PLANT PRODUCTION.
- 07.08 FORESTS.
- 07.09 ANIMAL PRODUCTION.
- 07.10 FISHERY.

08. INDUSTRY.

- 08.01 INDUSTRIAL ECONOMICS.
- 08.02 INDUSTRIAL ENTERPRISES.
- 08.03 INDUSTRIAL ENGINEERING. INDUSTRIAL EQUIPMENT.
- 08.04 INDUSTRIAL PRODUCTION. INDUSTRIAL PRODUCTS.
- 08.05 INDUSTRIAL RESEARCH.
- 08.06 FOOD INDUSTRY.
- 08.07 WOODWORKING INDUSTRY. PULP AND PAPER INDUSTRY.
- 08.08 TEXTILE INDUSTRY. LEATHER INDUSTRY.
- 08.09 RUBBER INDUSTRY.
- 08.10 CONSTRUCTION INDUSTRY. CERAMICS INDUSTRY. GLASS INDUSTRY.
- 08.11 ENERGY.
- 08.12 CHEMICAL INDUSTRY.
- 08.13 MINING.
- 08.14 METALWORKING INDUSTRY.
- 08.15 ELECTRONICS. ELECTRICAL EQUIPMENT.
- 08.16 COMMUNICATION INDUSTRY.

09. TRADE.

- 09.01 DEMAND. MARKET. CONSUMPTION.
- 09.02 PRICES.
- 09.03 MARKETING.
- 09.04 DOMESTIC TRADE.
- 09.06 INTERNATIONAL TRADE.

10. TRANSPORT.

- 10.01 TRANSPORT ECONOMICS.
- 10.02 GOODS. PASSENGERS.
- 10.03 TRANSPORT INFRASTRUCTURE.
- 10.04 MEANS OF TRANSPORT.
- 10.05 MODES OF TRANSPORT.
- 10.06 LOADING. PACKAGING.
- 10.07 INTERNATIONAL TRANSPORT. URBAN TRANSPORT.
- 10.08 TRAFFIC.
- 10.09 FREIGHT.

11. PUBLIC FINANCE. BANKING. INTERNATIONAL MONETARY RELATIONS.

- 11.01 PUBLIC FINANCE. TAXATION.
- 11.02 CURRENCIES. FINANCING.
- 11.03 INTERNATIONAL MONETARY SYSTEM.

12. MANAGEMENT. PRODUCTIVITY.

- 12.01 ENTERPRISES.
- 12.02 ECONOMIC CONCENTRATION.
- 12.03 ENTREPRENEURS.
- 12.04 MANAGEMENT.
- 12.05 EQUIPMENT.
- 12.06 TECHNOLOGY.
- 12.07 PRODUCTION. PRODUCTIVITY.
- 12.08 PRODUCTS. PRODUCT DEVELOPMENT.
- 12.09 COST ACCOUNTING. PROFIT.

13. LABOUR.

- 13.01 HUMAN RESOURCES.
- 13.02 EMPLOYMENT SERVICES. OCCUPATIONAL QUALIFICATIONS. PERSONNEL MANAGEMENT.
- 13.03 WORKING CONDITIONS.
- 13.04 OCCUPATIONAL SAFETY.
- 13.05 DISMISSAL. LABOUR MOBILITY.
- 13.06 LABOUR RELATIONS.
- 13.07 WAGES. WAGE INCENTIVES.
- 13.08 LEISURE.
- 13.09 OCCUPATIONS.

14. DEMOGRAPHY. POPULATION.

- 14.01 POPULATION DYNAMICS.
- 14.02 AGE GROUPS.
- 14.03 ETHNIC GROUPS.
- 14.04 HABITAT. RURAL AREAS. URBAN AREAS.
- 14.05 FERTILITY. FAMILY PLANNING.
- 14.06 MORTALITY.
- 14.07 MIGRATION.

15. BIOLOGY. FOOD. HEALTH.

- 15.01 BIOLOGY. PARASITOLOGY. BIOCHEMISTRY.
- 15.02 ANATOMY. GENETICS. PHYSIOLOGY.
- 15.03 FOOD. NUTRITION.
- 15.04 MEDICINE. DISEASES.
- 15.05 PHARMACOLOGY. TOXICOLOGY.

16. ENVIRONMENT. NATURAL RESOURCES.

- 16.01 ECOLOGY.
- 16.02 NATURAL RESOURCES.
- 16.03 DISASTERS. POLLUTION.
- 16.04 POLLUTION CONTROL. ENVIRONMENTAL ENGINEERING.
- 16.05 RESOURCES CONSERVATION.

17. EARTH SCIENCES. SPACE SCIENCES.

- 17.01 METEOROLOGY.
- 17.02 CLIMATE.
- 17.03 GEOGRAPHY.
- 17.04 GEOPHYSICS. GEOLOGY. SOIL SCIENCES.
- 17.05 HYDROLOGY. WATER.
- 17.06 OCEANOGRAPHY.
- 17.07 SPACE SCIENCES.

18. SCIENCE RESEARCH. METHODOLOGY.

- 18.01 RESEARCH. SCIENCE.
- 18.02 ORGANIZATION OF RESEARCH.
- 18.03 RESEARCH METHODS. THEORY.
- 18.04 DATA COLLECTING.
- 18.05 EXPERIMENTS.
- 18.06 MEASUREMENT.
- 18.07 MAPPING.
- 18.08 MATHEMATICS. STATISTICAL ANALYSIS.
- 18.09 COMPARISON. EVALUATION.
- 18.10 FORECASTS. TIME FACTOR.

19. INFORMATION. DOCUMENTATION.

- 19.01 INFORMATION.
- 19.02 DOCUMENTS.
- 19.03 TERMINOLOGY.
- 19.04 CONFERENCES.

Já o “Hierarchical Display” apresenta as cadeias de descritores do tesouro indo dos termos genéricos (TTs) aos mais específicos (NTs).

Segue-se o “KWOC Index” (key-word-out-of-context) que relaciona as palavras significativas utilizadas para a composição dos descritores do tesouro, agrupando, sob cada uma, a relação dos descritores onde essa palavra ocorre.

6.2.4 Análise dos exemplos

Exemplos 11 e 12

O Thesaurus POPIN agrupa, sob cada descritor — quando for o caso — as facetas pelas quais os termos podem ser analisados. É o caso de “Geografia da População”, que pode ser abordado tanto do ponto de vista da demografia (indicado pelo BT demografia 01.01.00), como sob a ótica da geografia (BT 04.01.01) e de Paróquias, analisável segundo a divisão territorial (04.01.04) e as instituições religiosas (05.06.04).

Embora a classificação por facetas não seja privilégio dos tesouros, a apresentação utilizada por tais instrumentos é mais organizada e sistemática do que a dos índices dos sistemas tradicionais de classificação, uma vez que registra, sob cada descritor, todas as facetas pelas quais ele pode ser analisado. (Na verdade, a facetação só muito recentemente foi introduzida nos sistemas de classificação. Os índices que deveriam indicar os múltiplos significados de um termo, entretanto, são bastante precários).

Exemplo 13

Neste exemplo realiza-se uma aproximação de conceitos do tipo “deslocamento genérico”, com base, provavelmente, na frequência dos termos na literatura-objeto. O resultado de procedimentos do gênero é a intensificação da generalização de conceitos e a conseqüente perda de especificidade. Deve-se ainda ressaltar que este exemplo é exclusivo da adaptação em português (VIET, 1986,

Exemplo 14

Neste caso estabelece-se uma relação de equivalência entre os termos “Cortiços” e “Favelas”. Contrariamente ao exemplo anterior, entretanto, não se pode afirmar tratar-se de um deslocamento genérico, uma vez que os traços semânticos dos dois descritores são de natureza diferente.

“Cortiço”, no Brasil, refere-se à “habitação coletiva das classes pobres; casa de cômodos...” (NOVO DICIONÁRIO AURÉLIO, sd). Já “Favela” designa o “conjunto de habitações populares toscamente construídas e desprovidas de recursos higiênicos” (idem *ibidem*).

A origem do cortiço está na deterioração do uso de uma unidade de moradia. Sua imagem se reporta a um condomínio ou a uma pensão. Já uma favela se reporta a uma imagem de um bairro urbano. O cortiço, ou “casa de cômodos” constitui um imóvel único, geralmente uma antiga habitação unifamiliar, caracterizada por um padrão médio de construção. Já uma favela é geralmente constituída de unidades independentes, como múltiplos imóveis de baixo padrão de construção.

Via de regra, no cortiço domina a locação, o que nem sempre ocorre na favela, onde a posse dos barracos é geralmente individual.

Quanto ao uso do espaço, o cortiço caracteriza-se pela subdivisão do imóvel em unidades familiares, enquanto que a favela por unidades geralmente autônomas.

O cortiço dispõe, em grande parte dos casos, da infra-estrutura básica oferecida pelo município (água, luz, telefone, coleta de lixo etc.), muito embora isso não se estenda, necessariamente, às unidades de moradia individuais. A favela — salvo aquelas “urbanizadas” não tem acesso a esses benefícios.

Sem estender a análise, poder-se ia pesquisar as diferenças entre o perfil sócio-econômico dos moradores de favelas e cortiços, incluindo suas expectativas em relação à vida nas cidades. Os traços diferentes levantados, entretanto, são suficientes para constatar a inconsistência da equivalência estabelecida entre os dois termos. É redundante afirmar que tal equivalência

corroborar a perda de especificidade na representação dos conceitos.

Exemplo 15

A oposição realizada, no exemplo, é feita em função de uma definição (implícita) operacional, para fins estatísticos: deve-se dar conta, na análise dos dados sobre o estado civil da população, das diversas situações que envolvem o vínculo matrimonial estabelecido por lei, por um lado, bem como daquelas onde esse vínculo está relaxado ou desfeito, por outro.

Contrariamente à oposição global mais comum em linguagem coloquial — casados/não casados — a grande oposição da qual parte o tesouro POPIN é entre “alguma vez casados” e “não casados”.

De fato, no referido tesouro, “Casados” subordina-se a “alguma vez casados” e relaciona, como TEs, os indivíduos que mantêm um vínculo matrimonial sob o ponto de vista estrito da lei. Isso explica a inclusão de “separados” entre os “casados”.

Se “separados” pode representar, em LN, um não casado (podendo participar como termo intermediário entre casado e não casado, para dar conta das transformações observadas na vida cotidiana relativas às relações entre indivíduos), o tesouro escolheu a situação institucional (a lei) para realizar a contabilização estatística das informações relativas à situação de estado civil da população.

Essa seria uma opção clara não fosse a inclusão de “separados” na nota de escopo relacionada à “Não Casados”. Embora sob este último termo não seja incluído o termo “Separados” como NT, a nota de escopo introduz ambigüidade, uma vez que utiliza, ao mesmo tempo, elementos da LN e da LE. Por outro lado, a opção pelo vínculo legal também não se sustenta se inferirmos a definição de “Separados” a partir do subgrupo “União Marital”: NTs “Separação de fato” e “Separação legal” se referem tanto aos aspectos institucionais quanto aos relativos ao costume. Assim, na nota de escopo relativa à “Não casado” utiliza-se o critério relativo ao “costume”, enquanto a relacionada à “Casados” adota-se o critério legal.

Por outro lado, a ausência de definição para “Alguma vez casados”, torna ambíguo o campo de aplicação do descritor. Para compreendê-lo deve-se tomar as notas de escopo relacionadas à “Casados” (“casados atualmente”) e à “Não casados” (“pessoas que não estão atualmente casadas...”), devendo-se entender que “casados” reflete uma situação de vínculo matrimonial estritamente institucional (supondo um “deslize” na inclusão de “Separados” na nota de escopo de “Não Casados”).

Entretanto, ao incluir no campo semântico de “Estado Civil” (13.02.00) “Coesposa”, “Coesposo”, “Companheiro”, “Concubina”, o tesouro volta a introduzir ambigüidade, uma vez que estes são termos onde prevalecem vínculos não institucionais, ou seja, vínculos de união de natureza consensual.

O mesmo pode ser observado em relação a “Cônjuges”, NT de “Casados”, pressupondo uma relação de matrimônio legal. Na nota de escopo existente sob o descritor “União Consensual”, os cônjuges são indivíduos que têm uma união sexual estável e “coabitam, sem ter nenhuma cerimônia formal de casamento”.

Conclui-se, portanto, que o termo “Cônjuges” remete ora a um conceito, ora a outro, e que é essa remessa desordenada (baseada na ausência de uma definição explícita) que faz aparecer, no campo semântico de Estado Civil, ao mesmo tempo, “Homens casados”, “Mulheres Casadas”, e os termos “Coesposa(o)”, “Companheiro”, “Concubina”, que representam indivíduos que mantêm vínculos não formais de união. Uma vez admitida a oposição “Alguma vez casados”, “Não Casados” (vínculo estritamente institucional), “Coesposa(o)”, “Companheiro”, “Concubina” não poderiam fazer parte do campo semântico de Estado Civil.

Além disso, uma vez que não se explicita a definição de “Estado civil” e considerando-se a equivalência estabelecida entre este e o termo “Estado conjugal”, poderia-se propor novas associações. Vejamos: “União conjugal” equivale a “União marital”, cuja nota de escopo se refere à “união sexual entre um homem e uma mulher de natureza estável e duradoura”. Disso pode-se inferir que uma possível situação de união sexual estável e duradoura constitui um dos traços de “Estado civil”.

Da mesma forma, “União marital” implica coabitação, já que “União temporária” não implica coabitação, nem padrão regular de relações sexuais. Deduz-se daí, novo traço para Estado Civil: situação possível de união sexual transitória, sem coabitação, nem padrão regular de relações sexuais.

Vê-se, dessa maneira, que na ausência de uma definição clara e precisa para Estado Civil, os significados dos termos não são unívocos, remetendo simultaneamente à terminologia de área e à linguagem coloquial.

Outra inconsistência diz respeito aos problemas causados pelo uso indevido do “deslocamento genérico” (25). “Desquitados”, embora apareça na nota de escopo de “Não Casados”, não é relacionado como NT deste último: é remetido a “Divorciados” supondo uma relação de equivalência semântica. Mas “Divorciados” se refere a indivíduos cujo vínculo matrimonial foi legalmente dissolvido, que não é o caso de “Desquitados”, que poderia ser incluído em “Casados”, uma vez que o vínculo legal ainda existe, tendo sido apenas relaxado legalmente.

Pode-se lembrar ainda que “Divorciados” — e por que não Divorciadas? — assim como “Viúvos(as)”, poderia ser incluído em “Alguma vez casados”, já que esta é uma situação concreta dos indivíduos acima discriminados. Mas o tesouro interdita esse uso restringindo o âmbito de aplicação do termo à “Casados”, relacionado como seu NT; no entanto, “Desquitados” não é um termo semanticamente equivalente a Divorciados, uma vez que, perante a lei brasileira, os indivíduos desquitados ainda são casados: com o desquite, ao contrário, do divórcio, o vínculo matrimonial não está dissolvido mas apenas relaxado.

Pode-se levantar, ainda, outros casos resultantes da ausência de definição de “Estado Civil”. Se esta é a denominação dada à subcategoria, como explicar a inclusão de “Celibato”, “Celibato definitivo”, “Educação pré-nupcial”, “Mães solteiras”, “Pais solteiros”, “População casadoura”, “População não casadoura”, além dos acima mencionados “Coesposa(o)”, Companheiro, Concubina? Teoricamente, sob “Estado Civil”, deveriam figurar termos da mesma natureza semântica.

(25) Caso de sinonímia, em tesouros, caracterizada pelo uso de equivalência entre o nome de uma classe e os nomes de seus membros, ficando o termo mais geral como descritor (IBICT, 1984, p.24)

Exemplo 16

Neste exemplo, embora a definição de migração não esteja explícita, todos os descritores estão em concordância com o conceito “estrito” de migração adotado pelas Nações Unidas: “translado de uma zona definitiva de migração a outra (ou traslado de uma distância mínima especificada) que se realiza ou durante um intervalo de migração determinado e que implicou uma mudança de residência” (NAÇÕES UNIDAS, 1972, apud PATARRA & CUNHA, 1987, p.33).

A tipologia das migrações especificada de 15.01.02 a 15.04.02 (mencionada sob migração, como NTs), entretanto, dá conta dos novos tipos de movimentos populacionais decorrentes de alterações estruturais e conjunturais ocorridas nos últimos anos (cf. PATARRA & CUNHA, 1987, p.34), exceção feita à “migração pendular”.

Para o CELADE, o conceito de movimento migratório é limitado ao deslocamento “com alteração de residência dos indivíduos de um lugar de origem, ou lugar de partida, a um lugar de destino, ou lugar de chegada, e implica atravessar os limites de uma divisão geográfica” (DICCIONARIO DEMOGRAFICO MULTILINGÜE, 1985), devendo-se estabelecer uma “diferença entre migração e deslocamento temporário, mediante critérios baseados na duração da ausência do lugar de origem e na duração da presença no lugar de destino” (idem *ibidem*). Embora considere a regularidade cíclica e a importância econômica e social dos deslocamentos de trabalhadores ou movimento pendular, este não é considerado como um tipo de migração (idem *ibidem*).

Já para PATARRA & CUNHA, deve-se considerar o movimento pendular correspondente aos “deslocamentos temporários de indivíduos de seus lugares de residência habitual para frentes ou locais permanentes de trabalho” como um tipo de migração: a “Migração pendular” (idem *ibidem*). Conforme os autores do referido artigo, “este tipo de movimento ... deixaria bastante constrangido, ao ser interpelado sobre o tema, aquele que apregoasse uma definição rígida do que poderia ser um movimento migratório”. Para os autores, “as várias referências em geral recomendadas (sejam elas espaciais, temporais, etc.), apesar de extremamente úteis e importantes, deveriam ser apenas instrumentos adequados aos objetivos explícitos dos estudos; ou seja, a definição depende depende da proposição clara do problema” (idem *ibidem*, grifo meu).

No POPIN, não há menção à migração pendular, mas pode-se localizar o descritor “Deslocamento residência-trabalho” usado para identificar “Movimento pendular”. Ao lado da ausência de definição ou âmbito de validade de aplicação do descritor, o uso de equivalência entre “Deslocamento residência-trabalho”/“Movimento pendular” pode ser visto como um reflexo de um momento de transformação conceitual caracterizado pela ausência de uma unidade terminológica — o termo — para precisar e representar o novo conceito. O descritor escolhido é o resultado de uma apropriação da LN, e não da LE, podendo, portanto, induzir à ambigüidade, ocasionando remessas não unívocas. Nada impede, por exemplo, que se atribua o referido descritor a deslocamentos residência-trabalho que não caracterizam um movimento migratório.

Os termos coordenados relativos aos tipos de migração são adequadamente obtidos a partir da aplicação de características de divisão semelhantes (modos de realização da ação migração), a exceção de 15.03.01 onde “primeira migração” e “última migração” introduzem elementos de naturezas diferentes na série (resultados da ação de migração).

Exemplo 17

Analisaremos aqui, o descritor “Communication”, do Macrothesaurus. Inicialmente, reproduzimos sua apresentação na pare alfabética do tesouro (ver A).

Para definir os termos sob “Communication” o Macrothesaurus escolheu três facetas: a Comunicação sob o ponto de vista da Cultura e Sociedade (notação 05), da Indústria (notação 08) e da Informação e Documentação (notação 19) (ver B).

Essa escolha se reflete na organização hierárquica adotada para agrupar os termos sob Comunicação que, por sua vez, privilegia seis entradas para encabeçar a organização dos termos: “Development communication” (05.07.01), “Information transfer” (19.01.07), “Mass communication” (05.07.01), “Postal Services” (08.16.00) e “Telecommunications” (05.07.03) (ver C). A sustentação dessa escolha se reporta ao fato de que todos eles têm BTs e TTs

coincidentes e todos têm como BT e como TT, “Communication” (26).

Já os descritores subordinados a cada um dos relacionados acima representam, na cadeia hierárquica, a introdução de novas facetas para a enumeração dos assuntos.

A cadeia formada sob a faceta “Information Transfer” (19.01.07) — portanto, “Information. Documentation” — reúne descritores da mesma classe: 19.01.07 e 19.02.07. Já sob Mass Communication (05.07.01) os descritores são relativos a outras facetas: 09.03.04 (“Advertising” e “Public Relations”), sob “Marketing” (grupo 09) “Trade”; 04.04.02 “Propaganda”, sob “Politics” (grupo 04) “Institutional Framework”.

Vejamos, em seguida, como se organizam alguns dos subgrupos obtidos da hierarquia sob Communication relacionada acima, na parte do tesauro relativa aos Grupos de Descritores.

Analisando o subgrupo temático relativo à Comunicação (05.07) (ver D), no interior do grupo (ou faceta) Cultura e Sociedade (05), verifica-se que o único traço comum que une os descritores é o fato de que todos podem se relacionar, de uma maneira ou outra, à Comunicação.

Sob esse tópico os descritores são reunidos em três grupos, onde são enumeradas: ações em comunicação em geral, principalmente (05.07.01); público (05.07.02) e meios de comunicação (05.07.03) propriamente ditos.

O princípio de subdivisão adotado para a obtenção dos descritores não é claro tanto no primeiro como no segundo grupo, nos quais parece dominar a associação livre. Observa-se, nesse caso, a existência de descritores de natureza diferente (obtidos, portanto, por processos de divisão distintos) como por exemplo “Communication Planning” — como modo de operacionalizar sistemas de comunicação — e “Communication Systems” — instrumento concreto de comunicação — ocorrendo o mesmo com “Communication Policy” — modo — e “Confidentiality” —

(26) Na reprodução da cadeia hierárquica (Hierarchical display), acrescentamos, ao lado dos descritores sob “Communication”, as notações correspondentes, com o objetivo de visualizar as facetas ou agrupamentos realizados sob o termo. Pesquisamos, depois disso, cada um dos descritores que encabeçam as facetas para verificar seus BTs e TTs.

finalidade de uma ação de comunicação.

No terceiro grupo pode ser identificada uma tipologia dos meios de comunicação, reunindo instrumentos específicos (“Books”, “Films”, “Radio”, “Television” etc.) ao lado daqueles que os englobam — portanto, mais genéricos — (“Media”, “Mass Media”). A cadeia hierárquica especificada sob “Media”, por exemplo, comprova o fato.

Fatos semelhantes ocorrem quando analisamos a faceta “Indústria” (08) de “Communications”, subgrupo temático 08.16, “Communication Industry” (Ver E). São relacionadas, no mesmo nível, as indústrias ligadas à comunicação (os agentes do desenvolvimento sócio-econômico), os produtos e os modos para sua obtenção.

Sob o subgrupo temático 19.01.07, na faceta “Information. Documentation” (notação 19), são enumerados descritores de níveis de generalidade diferentes, como “Data transmission” e “Transnational data flow”, ou “Information Dissemination” e “SDI” (Selective Dissemination of Information) (Ver F).

A análise dos elementos sob cada faceta analisada demonstra que, via de regra, o arranjo dos grupos temáticos de descritores é feito através de enumeração, buscando organizar, com base no “bom senso”, termos que pertencem a campos conceituais semelhantes: em alguns casos adota-se a forma hierárquica para o registro dos descritores; em outros, não se especificam os critérios de organização.

Deve-se considerar, ainda, que, embora o Macrothesaurus contenha algumas notas de escopo para dar conta da abrangência dos descritores, inexistem definições precisas dos seus elementos.

Exemplo 18

A análise comparativa dos descritores idênticos, nos dois tesouros, apresenta semelhanças quanto à organização da cadeia hierárquica: em ambos, por exemplo, o descritor “Imigração” é subordinado

à Migração Internacional. O maior número de NTs registrado pelo Thesaurus POPIN deve-se, obviamente, à maior especificidade desse instrumento relativamente ao assunto abordado.

Já em relação aos RTs, os tesauros realizam associações diferentes: além do descritor comum, “Imigrantes”, o Macrothesaurus registra “Immigration Law” e “Population Growth”, enfatizando, portanto, os aspectos legislativos e aqueles da dinâmica populacional relacionados aos fenômenos de movimento da população segundo os lugares de origem dos migrantes, enquanto que o Thesaurus POPIN registra apenas “Política de Imigração”.

A especificidade do domínio dispensa ao POPIN a inclusão de “Population Growth” (ou “Crescimento demográfico”, tradução correspondente), consequência imediata da Imigração. Por outro lado, “Política de Imigração” revela o enfoque demográfico do POPIN, contrariamente ao enfoque legislativo adotado pelo Macrothesaurus para abordar o problema. De qualquer forma, deve-se observar que, embora não fique explícito o gênero de associação realizada, o registro dos TRs é determinado, em grande medida, pela faceta privilegiada na elaboração do domínio.

Exemplo 19

Ambos os tesauros subordinam o descritor “Planejamento Econômico” (09.05.02, no POPIN; 02.01.02, no Macrothesaurus) a Planejamento. O Macrothesaurus relaciona inúmeros NTs, contrariamente ao Thesaurus POPIN que não registra nem um descritor.

Os TRs de “Planejamento Econômico” no Thesaurus POPIN também são registrados pelo Macrothesaurus que, entretanto, oferece uma lista mais extensa.

Note-se que o descritor “Economic Indicators”, cujo correspondente no POPIN é “Indicadores econômicos” (08.01.00) (VIET, 1985, exemplo p.77) não se acha associado à “Planejamento Econômico” neste último. No POPIN, o descritor “Planejamento econômico” se insere no agrupamento temático, “Políticas Governamentais” (09.05.00), enquanto que no Macrothesaurus ele está sob o agrupamento “Economic Research. Economics” (03.01).

Vê-se, desse modo, que as associações não-hierárquicas registradas pelos dois tesouros diferem entre si em função dos aspectos que cada um privilegia. Fica difícil, entretanto, identificar a natureza dos relacionamentos associativos registrados. A classificação proposta pelo IBICT, embora ressalte a ausência de base teórica para sua determinação precisa, oferece alguns elementos para sua identificação (IBICT, 1984, p.31).

Segundo essa classificação teríamos:

Imigração/Imigrantes = campo de estudo/fenômeno estudado; operação/agente; ação e resultado da ação; dependência causal;

Imigração/Política de imigração = fenômeno estudado/campo de estudo (inversão da classificação proposta); operação/instrumento; resultado da ação/ação (inversão); dependência causal;

Immigration/Immigration law = fenômeno estudado/campo de estudo (inversão); operação/instrumento; resultado da ação/ação (inversão); dependência causal;

Immigration/Population growth = campo de estudo/fenômeno estudado; agente/processo (inversão); ação/resultado da ação; dependência causal;

Planejamento econômico/Desenvolvimento econômico = campo de estudo/fenômeno estudado; agente/processo (inversão); ação/resultado da ação; dependência causal;

Planejamento econômico/Economia planejada = ação/resultado da ação; dependência causal;

Planejamento econômico/Política econômica = campo de estudo/fenômeno estudado; instrumento/operação (inversão); dependência causal;

Economic Planning/Economic administration = campo de estudo/objeto estudado; instrumento/operação (inversão); dependência causal;

Economic Planning/Economic Indicators = campo de estudo/objeto estudado; operação/instrumento; dependência causal;

Economic Planning/ILPES = processo/agente.

O exercício acima demonstra que a classificação proposta não é potente para especificar a natureza das relações de associação: um mesmo descritor, em ambos os tesouros, pode ser analisado segundo diferentes enfoques e a associação continua livre, isto é, em determinado nível todos os descritores se relacionam. Não se propõe, de fato, quais as relações de associação são potentes para a representação e a recuperação de informações.

Nas operações de indexação (e recuperação) os termos relacionados são utilizados, antes, para auxiliar a definição do campo semântico do descritor em destaque, do que para a recuperação. Adicionalmente, tais termos podem conduzir o usuário (incluindo o indexador) ao descritor que melhor se adequa à representação desejada.

Na verdade, as relações gênero/espécie e de equivalência são muito importantes para a recuperação da informação sendo que os outros tipos de relações de associação contribuem para aumentar a redundância, propondo o acréscimo necessário para aperfeiçoar a recuperação. O excesso de redundância, entretanto, pode descaracterizar conceitualmente os descritores.

A indeterminação da natureza da associação estabelecida deve ocorrer, também, em função da ausência de determinação clara do enfoque a ser privilegiado por cada tesouro. As categorias utilizadas para sua classificação são universais e sua aplicação é subjetiva. Contrariamente às associações lógicas, onde se determina a direção de leitura das relações realizadas (gênero/espécie, todo/parte), as associações não-hierárquicas são potencialmente realizáveis em todas as direções, dependendo do ponto de partida selecionado para a análise.

6.2.5 Avaliação geral

6.2.5.1 Estrutura das LDs

Conforme vimos, os tesouros analisados estruturam seus termos a partir de hierarquias básicas onde domina a relação de inclusão, ao lado de relações de equivalência e de associações de naturezas diversas.

Pode-se observar, entretanto, pelos exemplos analisados, que embora as regras para a construção de tesouros prescrevam a obediência a princípios lógicos (27) para a construção das hierarquias, elas nem sempre são observadas, prevalecendo o uso de relações associativas de vários tipos para a enumeração dos descritores.

Conforme já dissemos anteriormente, através das relações lógicas não se esgota o registro dos fatos relativos ao conhecimento do mundo empírico, o que explica a utilização de outros tipos de relações para a enumeração dos descritores.

Deve-se lembrar, todavia, que a natureza das relações estabelecidas entre os descritores deve ser suficientemente clara, de modo a assegurar a consistência do instrumento de comutação, ou seja, deve-se assegurar um conjunto de procedimentos rigorosos para a organização e registro dos descritores.

Esse rigor nem sempre é observado, como pudemos verificar através de alguns dos exemplos analisados: as associações (hierárquicas e não-hierárquicas) evidenciam o registro aleatório do conhecimento relativo aos domínios, prevalecendo o uso do “bom senso” para a enumeração das unidades que se pretendem significativas para a representação de conceitos.

Tal fato se comprova, por exemplo, relativamente ao Macrothesaurus, quando se verifica que o

(27) Um relacionamento lógico-hierárquico “exprime os graus ou os níveis de super-ordenação e de subordinação entre os conceitos (...)” (IBICT, 1984, p.25). Deve-se assegurar que os termos colocados em relação hierárquica representem “o mesmo tipo básico de conceito, por exemplo, de que ambos denotam operações, ou coisas ou propriedades” (idem *ibidem*).

sistema de classificação — cuja síntese é representada pelos “Subject Category Fields” e pela expansão configurada nos “Descriptor Groups” — é caracterizado pela ausência de princípios explícitos. Conforme pode-se verificar na introdução daquela obra, ela foi derivada do vocabulário da área, tendo sido “gradualmente moldada pelas relações semânticas” (VIET, 1985, p.xi). Não é elucidada, entretanto, a forma de registro dessas relações.

O “bom senso” também caracteriza a estruturação do Thesaurus POPIN, muito embora a maior delimitação de domínio e a existência de maior sistematização conceitual e terminológica do campo da Demografia acabe por conferir maior grau de rigor na estruturação dos termos daquele instrumento. As relações associativas não-hierárquicas, explicitadas através de TRs, são mais livres e devem depender, em grande medida, da garantia literária, ou seja, das associações registradas e mais freqüentes na literatura de área. Dito de outro modo, as associações entre os descritores de cada tesauro analisado diferem entre si em razão das diferentes experiências práticas envolvidas em cada domínio.

Não há uma formalização única e rígida das relações associativas dadas a partir de uma palavra porque elas dependem das ordens de arranjo funcionais determinadas pelos objetivos e interesses dos domínios em questão. Isso explica as diferenças verificadas na listagem de descritores associados à “Imigração” (exemplo 18) e “Planejamento econômico” (exemplo 19), em tesouros com enfoques e objetivos distintos.

Percebe-se, pelo exposto acima, que as associações estabelecidas pelos tesouros são de natureza lógica e semântica, e que as associações lógicas não são suficientes para dar conta do conhecimento registrado. As relações semânticas entre descritores, por seu lado, dependem do domínio e da forma pela qual ele é observado, organizando-se diferentemente conforme o ponto de vista privilegiado.

Deve-se lembrar, ainda, que a noção de campo semântico de Granger afirma a necessidade de considerar que o léxico comporta várias organizações que correspondem a vários pontos de vista do uso lingüístico (GRANGER, 1974, p.200). “Não é o “léxico da habitação tomado em si mesmo e abstratamente que é necessário analisar, mas as organizações desse léxico tais como aparecem

do ponto de vista da construção, do ponto de vista da ocupação, do ponto de vista da estética arquitetural etc...” (idem *ibidem*).

Comparados aos sistemas de classificação, entretanto, os tesauros são instrumentos mais flexíveis para a indexação de informações, uma vez que a grade classificatória, embora definida de início, é mais abrangente. Acrescente-se ainda que, na utilização dos tesauros, abandona-se a hierarquia básica em prol do uso centrado nos descritores, considerando suas relações verticais e horizontais (subordinação lógica e não-lógica), sem se prender excessivamente à hierarquia classificatória. Isso garante, conseqüentemente, maior flexibilidade de uso.

6.2.5.2 Eficiência para o controle de vocabulário

Como já o dissemos, as LDs devem funcionar como instrumentos para o controle do significado do vocabulário. Entretanto, na ausência de definições explícitas, o significado dos descritores dos tesauros analisados deve ser buscado apenas a partir da posição relacional de seus elementos.

Na relação entre os Termos Específicos e seus superordenados, efetiva-se a conjunção, determinada pela existência de traços constantes (mesma característica de divisão). A conjunção, portanto, delimita o domínio e define toda uma classe de termos que estarão, por esse motivo, associados. Traços específicos definem a disjunção, de modo a individualizar os termos no interior da classe (TÁLAMO, 1987, p.91).

Nos tesauros analisados, entretanto, o uso de um traço constante para a organização da classe não é um princípio permanente, e nem sempre os traços específicos são claramente identificáveis. Não raras vezes, o conjunto dos descritores remete a vários interpretantes ao mesmo tempo, cujo resultado é a ausência de especificidade.

Desse modo, as relações entre os descritores não fornecem elementos suficientes para determinar definições, remetendo, invariavelmente, à traços possivelmente existentes em sistemas de significação externos aos instrumentos analisados.

Assim sendo, pode-se afirmar que os tesouros, genericamente, não funcionam como dicionários e não são também sistemas de significação propriamente ditos, uma vez que a compreensão de seus elementos não se dá no interior do próprio código, dependendo invariavelmente do conhecimento registrado externamente. Como esse recurso ao conhecimento exterior também não é explicitado, têm-se prejudicada a função de significação, o que torna os tesouros instrumentos inadequados para exercer, entre outras, a função de controle de vocabulário.

Acrescente-se ainda que a eficiência de um tesouro enquanto instrumento de controle de vocabulário impõe a necessidade de uma delimitação precisa do vocabulário-objeto, considerando, inclusive, sua evolução e atualização.

Pelos exemplos analisados verifica-se que no tesouro POPIN utiliza-se tanto o vocabulário comum, quanto o especializado. Por exemplo: na linguagem comum, “casado” inclui tanto o casado “legalmente”, quanto o casado “pelo costume”. A evolução da linguagem comum altera o significado original de “casado” baseado na relação legal entre sexos opostos. Ou seja, a palavra “casado” perde alguns traços mais específicos passando a denominar relações estáveis, mas não necessariamente “legais”.

Na linguagem comum as palavras perdem ou ganham traços, e o “ajuste” do significado é muito rápido. Já nas LDs, como não existe esse mecanismo de ajuste, o significado pode ficar “cristalizado”, determinando rigidamente os conceitos a que os termos se referem e causando problemas quando de seu uso na indexação de informações que se refiram a conceitos novos. A incorporação de novos significados ao instrumento é, portanto, inviável.

Quanto ao Macrothesaurus, deve-se observar que seu domínio é caracteristicamente interdisciplinar, formado pela intersecção dos diversos setores que compõem a questão do planejamento sócio-econômico. Essa mesma interdisciplinaridade da área se consubstancia no vocabulário: a apropriação dos termos é realizada a partir dos diversos setores que compõem a área. Por isso, a generalização conceitual é constante, já que não se consegue determinar com rigor a especificidade dos termos.

Esse vocabulário é, conseqüentemente, reflexo da maior ou menor sistematização conceitual e

terminológica dos domínios que compõem a área em questão, oscilando entre o uso de termos tomados de empréstimo da LN e da terminologia especializada.

6.2.5.3 Eficiência para a elaboração de índices

Face aos sistemas de classificação, os tesouros permitem a elaboração de índices mais consistentes dada a especificação de domínios e a forma de apresentação de seus descritores.

A despeito dos problemas do uso da LN em alguns descritores, os tesouros operam uma normalização mais rigorosa para a apresentação dos seus termos, quer seja do ponto de vista lingüístico (tratamento de homografias, sinonímia, definição de gênero, número, grafia, empréstimos lingüísticos, tratamento de termos compostos etc.), ou gráfico.

Ao utilizar um tesouro não é necessário — como ocorre nos sistemas de classificação — recorrer a outras listas para definir as entradas do índice.

Relativamente ao texto, entretanto, o índice obtido pelo tesouro é “desnaturalizado” e a álgebra booleana (e, ou, e não etc.) utilizada como mecanismo de recuperação da informação, não responde pela “naturalização”.

6.2.5.4 Eficiência para representar a informação

A ausência de definições nos tesouros também prejudica a função de representação e conseqüentemente, a explicitação dos objetivos específicos a que se propõe.

Ao representar uma informação via tesouro, a ausência da definição pode implicar, no momento da indexação, o uso exclusivo da equivalência lexical: as operações de indexação assim realizadas alteram a característica de representação, uma vez que não se representam as informações, mas as etiquetas previstas pelo código de conversão.

Esse fato fica evidente no exemplo retirado do Thesaurus POPIN relativo ao Estado Civil da população: não se dispondo de definição, têm-se prejudicada a organização e representação dos conceitos já no próprio instrumento e, conseqüentemente, nas operações de representação (indexação) da informação presente nos textos.

O mesmo ocorre com o Macrothesaurus, onde a definição é deduzida unicamente da posição relacional dos descritores (como na maior parte dos tesauros), e remete aos interpretantes previstos apenas pelo instrumento. Condiciona-se a representação, desse modo, aos registros efetuados pelo instrumento. Veja-se, por exemplo, o caso de “Comunicações” na faceta Cultura e Sociedade, cuja suposta definição é restrita à enumeração de ações de comunicação, público e meios de comunicação.

Em ambos os tesauros analisados corre-se o risco de induzir o usuário (incluindo-se aqui o documentalista), a interpretantes ao mesmo tempo restritivos e diversos, dada a ausência de remessa precisa a um sistema de conceitos.

6.2.5.5 Eficiência na comunicação documentária

Para que exista comunicação é necessário que o sistema de significação utilizado seja claro e explícito. A ausência de definição nos tesauros — e conseqüentemente, a ausência de delimitação precisa do sistema conceitual a que se refere — mais a artificialidade do sistema, acabam por dificultar a comunicação.

Comparados aos sistemas de classificação, entretanto, deve-se destacar a superioridade dos tesauros, uma vez que tais instrumentos definem, com maior precisão, pontos de vista específicos para a abordagem dos conceitos.

Deve-se pesquisar, portanto, a viabilidade de tesauros acoplados a glossários terminológicos, como forma de dinamizar a comunicação.

CONCLUSÃO

Pudemos confirmar, ao longo do trabalho, que as LDs analisadas não são construídas com o rigor necessário para dar conta do significado. Vários critérios são utilizados, de forma empírica, inexistindo uma preocupação com a palavra como fonte de sentido.

As LDs se propõem como metalinguagens mas não dispõem de procedimentos de elaboração capazes de caracterizá-la como tal. De fato, uma metalinguagem é utilizada para “descrever” textos da LN, função distinta daquela de uma LD, que se propõe a representar e possibilitar a recuperação da informação documentária.

Vale lembrar, em relação às características de uma metalinguagem que, via de regra, predomina, na descrição, a expansão textual, contrariamente ao que acontece com a representação para fins documentários, onde se busca a condensação ou síntese.

Acrescente-se a isso que uma metalinguagem é constituída de elementos semióticos idênticos utilizados pela linguagem do texto a ser descrito, que também não é o caso das LDs, já que estas envolvem o uso de, pelo menos, dois sistemas semióticos diferentes, como já o dissemos anteriormente.

Na ausência de instrumental para a normalização semântica das expressões documentárias as LDs passam a integrar, na prática, as denominadas “anomalias” da LN a que são chamadas a “sanar”. Desse modo, a função normalizadora de uma LD ou é inexistente ou, quando existe, apresenta-se de forma precária.

Uma LD só seria uma metalinguagem se ela se consubstanciasse, de fato, como elemento da própria análise, ou seja, se ela funcionasse como instrumento de conhecimento.

Na verdade, no desenrolar do trabalho pudemos verificar que as funções associadas às LDs — seja pela literatura, ou pela expectativa pessoal e profissional — transcendiam às funções que elas

concretamente poderiam desempenhar. As LDs analisadas não se caracterizam por definições claras quanto aos seus objetivos, não têm uma definição clara quanto ao seu objeto de análise, não revelam o conceito de representação utilizado, nem a forma de operacionalização necessária à análise. Ao buscarmos, portanto, verificar se as LDs analisadas “cumpriam sua função”, deparamo-nos com a ausência de rigor sobre o próprio conceito funcional das LDs.

Isso nos levou a perceber que, embora de maneira implícita, tais LDs foram construídas na suposição da existência de um signo monossêmico, onde significado e significante pudessem estar numa relação sígnica fixa e estável, quase imutável. A monovalência do signo, entretanto, só se justificaria se uma LD fosse de natureza taxionômica, onde a cada signo deveria corresponder um único significado.

Compreende-se, desse modo, porque os aspectos gramaticais das unidades lexicais que compõem uma LD constituem a única orientação clara nos manuais de construção de tesouros.

Tal fato revela a inconsistência do princípio implicitamente adotado: o signo monossêmico é uma “ficção” (parafrazeando Gardin) e o usuário sempre entra num sistema documentário munido de tantas possibilidades interpretativas quantas lhe forem asseguradas pela sua experiência colateral.

Conclui-se, desse modo, que não há como trabalhar com a hipótese de um signo “cristalizado”, já que, sua condição de signo, remete necessariamente à virtualidade interpretativa.

Portanto, é preciso rever qual é a normalização necessária para trabalhar com a plurivalência do signo, caso o objetivo primordial da representação documentária seja efetivamente a transmissão de informação e sua circulação efetiva.

A normalização, como vimos, é um elemento fundamental para que as representações documentárias sejam integradas a um sistema de comunicação documentária. Tal sistema tem sua existência vinculada à comunicação e à circulação de informações, não tendo sentido discuti-lo na ausência deste contexto.

Entretanto, não é isso o que ocorre. Na verdade, a questão da comunicação raramente se coloca no âmbito da Documentação. Não fosse assim ter-se ia procurado resolver o problema da monossemia, pelo menos, através do recurso ao “dicionário”.

Sem entrar nos problemas da incorporação da noção de dicionário, deve-se destacar a arbitrariedade que normalmente caracteriza esse tipo de recurso: as definições vão sempre variar segundo a decisão do autor, que imprimirá este ou aquele enfoque para a sua construção. Desse modo, a alternativa “dicionário” leva, em última análise, à necessidade de definição de noções primitivas para a arquitetura das definições. E ficou evidente que sua utilização (como por exemplo, os universais semânticos, os “casos”, as categorias etc.), só se desenvolve adequadamente uma vez definida sua inserção contextual. Sem esse parâmetro, tudo pode: “reforma agrária” é paciente, é instrumento, é finalidade etc.

A comunicação, geralmente impensável na concepção tradicional de LDs só se torna possibilidade com a introdução de universos de referência.

Nesse sentido, dizer que uma LD é uma metalinguagem, não resolve a questão. Por outro lado, buscar a “naturalização” da LD também não soluciona o problema, uma vez que ela não é um instrumento para “descrever” a linguagem-objeto, mas para representá-la documentariamente. As LDs são desnaturalizadas porque são construídas buscando a normalização. Metalinguagem e linguagem desnaturalizadas são essencialmente diferentes.

A normalização nas LDs tradicionais como instrumento de controle de vocabulário não ultrapassa à normalização gramatical. Não se determinam formas únicas de interpretação, mas apenas formas únicas de apresentação. A variação é, assim, “camuflada”, porque os conceitos não são estabilizados e não podem sê-lo em linguagens que ignoram a normalização semântica.

O conceito tem componentes retirados de conceitos já existentes e continuamente reconfigurados na evolução da ciência. Não é possível cristalizar a relação significado/significante, e se a LD persegue a normalização para assegurar a univocidade conceitual, ela parte de um princípio inconsistente. Um instrumento para a representação da informação deve ser edificado prevendo

relações entre conceitos e não só relações entre significantes.

A proposta da enciclopédia — nos limites do dicionário local e viabilizada pelas terminologias de área — consegue, sob nosso ponto de vista, controlar melhor a variação, porque acompanha a produção terminológica de área e não cerceia, de todo, a possibilidade interpretativa, ao fornecer referentes concretos à interpretação.

A perspectiva que apresentamos considera, portanto, as características específicas da “apropriação documentária”, ou do “discurso” documentário (se é permitido extrapolar momentaneamente o conceito de discurso) e do jogo de significações que ele envolve. Essa condição possibilita o engendramento do signo documentário: um signo que, a partir da referência terminológica, pode orientar a semiose, sem cerceá-la ou dispará-la aleatoriamente.

O “discurso” documentário deve se caracterizar pela apropriação das informações dos discursos individuais, de forma a evidenciar a “notícia” ou o “assunto”, conforme o caso e o objetivo do sistema, mas não dados isolados. Dessa forma, a representação documentária não é uma expressão notacional, um termo, ou um descritor, mas um conjunto articulado que, embora não substitua, em hipótese alguma, o texto original (a generalização é necessária à constituição de sistemas documentários), deve funcionar como signo, remetendo aos sistemas de significação pertinentes.

Nos limites da “escala industrial” de redução do conteúdo dos textos, o signo documentário deve “desencadear uma situação de comunicação” e, ao mesmo tempo, funcionar como elemento de conhecimento.

BIBLIOGRAFIA E REFERÊNCIAS BIBLIOGRÁFICAS

- AMARO, R.K.O.F. (1991). *Contribuição da análise do discurso para a análise documentária: o caso da documentação jornalística*. São Paulo, ECA-USP. (Dissertação de Mestrado)
- BACON, F. (1979). *Novum organum ou Verdadeiras indicações acerca da interpretação da natureza*. Trad. de José Aluysio Reis de Andrade. 2.ed. São Paulo, Abril Cultural. p.1-231.
- BARANOW, V.G. (1979). Aspectos lingüísticos de linguagens de indexação. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA. Anais. Rio de Janeiro, IBICT, v.1, pp.295-310.
- BEGHTOL, C. (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *J. of Documentation*, v.42, n.2, pp.84-113, jun.
- BENVENISTE, É. (1966). Catégories de pensée et catégories de langue. In: _____. *Problèmes de linguistique générale*, I. Paris, Gallimard. cap. 6.
- BENVENISTE, É. (1966). De la subjectivité dans le langage. In: _____. *Problèmes de linguistique générale*, I. Paris, Gallimard. cap. 21.
- BORBA, F. da S. (1971). *Pequeno vocabulário da lingüística moderna*. São Paulo, Ed. Nacional/EDUSP.
- BRASIL. Ministério da Educação e Cultura. Serviço de Estatística da Educação. *Glossário de termos utilizados na estatística educacional*, I. Rio de Janeiro, Fundação MUDES, 1980. 144p.
- CABRE, M.T. (1991). Terminologie ou terminologies? specialité linguistique ou domaine interdisciplinaire? *Meta*, v.36, n.1, 55-63.
- CDU - CLASSIFICAÇÃO DECIMAL UNIVERSAL. (1987) Ed. média em língua portuguesa. Brasília, IBICT. 2 v. (Publicação FID, n.665)
- CHAUMIER, J.(1978). *Les langages documentaires: le traitement linguistique d'information documentaire*. Paris, Entreprise Moderne d'Édition.
- CINTRA, A.M.M. (1983). Elementos de lingüística para estudos de indexação. *Ci. da Inf.*, Brasília, v.12, n.1, pp.5-22.
- CINTRA, A.M.M. (1987). Estratégias de leitura em documentação. In: SMIT, J.W. (org.). *Análise documentária: a análise da síntese*. Brasília, IBICT, p.28-35.
- COELHO NETTO, J.T. (1990). *Semiótica, informação e comunicação*. 3.ed. São Paulo, Perspectiva.
- COYAUD, M. (1966). *Introduction a l'étude des langages documentaires*. Paris, C. Klincksieck.
- COYAUD, M. & SIOT-DECAUVILLE, N. (1967). *L'analyse automatique des documents*. Paris, Mouton.
- CUNHA, I.M.R.F. (1987). *Análise documentária*. In: SMIT, J., org. *Análise documentária: a análise da síntese*. Brasília, IBICT. pp.37-60.
- CUNHA, I.M.R.F. (1987). Estruturação do vocabulário. In: SMIT, J., org. *Análise documentária: a análise da síntese*. Brasília, IBICT. pp.61-85.
- CUNHA, I.M.R.F., org. (1989). *Análise documentária: considerações teóricas e experimentação*. São Paulo, FEBAB.
- CUNHA, I.M.R.F.C. (1990). *Do mito à análise documentária*. São Paulo, EDUSP. (Teses, 11)
- DASCAL, M. (1978). Las sémiologies contemporaines. In: — *La sémiologie de Leibniz*. Paris, Aubier Montaigne. cap. 2.
- DESCLÈS, J.-P. (1987). Réseaux sémantiques: la nature logique et linguistique des relateurs. *Langages*, n.87, pp.55-78.
- DEWEY DECIMAL CLASSIFICATION devised by Melvil Dewey. (1989) 20.ed. Ed. by J. P. Comaromi et alii. Albany, N.Y., Forest Press. 4v.
- DICCIONARIO DEMOGRAFICO MULTILINGÜE. 2. ed. Lieje, Unión Internacional para el Estudio Científico de la Poblacion; Centro Latinoamericano de Demografia; Ed. Ordina, 1985.
- DIDEROT, D. (1989). Enciclopédia, s.f. (Filosofia). Trad. de Marcelo Coelho e Marco Chiaretti. *Folha de São Paulo*, 9 dez. 1989, p.G-1. (Letras)
- DUBOIS, J. & C. (1971). *Introduction à la lexicologie: le dictionnaire*. Paris, Larousse. cap. 6-11.
- DUBOIS, J. et alii (1988). *Dicionário de lingüística*. Trad. de Frederico Pessoa de Barros e outros. São Paulo, Cultrix.
- DUCROT, O. e TODOROV, T. (1988) *Dicionário enciclopédico das ciências da linguagem*. Trad. de Alice Kyoko Miyashiro e outros. 2.ed. rev. ampl. São Paulo, Perspectiva.
- ECO, U. (1980). Peirce et la sémantique contemporaine. *Langages*, n.58, pp.75-91, juin.
- ECO, U. (1984). *Conceito de texto*. Trad. de Carla de Queiroz. São Paulo, TA Queiroz; EDUSP.

- ECO, U. (1986). *Lector in fabula: a cooperação interpretativa nos textos narrativos*. Trad. de Atílio Cancian. São Paulo, Perspectiva. (Estudos, 89)
- ECO, U. (1991). *Semiótica e filosofia da linguagem*. Trad. de Mariarosaria Fabris e José Luiz Fiorin. São Paulo, Ática. (Fundamentos, 64)
- FERRATER MORA, J. (1971). *Diccionario de filosofia*. Buenos Aires, Ed. Sudamericana.
- FOSKETT, A.C. (1973). *A abordagem temática da informação*. Brasília; São Paulo, UnB; Polígono.
- FOSKETT, D.J. (1967). Classificação. In: ASHWORTH, W. *Manual de bibliotecas especializadas e de serviços informativos*. Lisboa, Fund. Calouste Gulbenkian. p.89-158.
- GARCÍA GUTIÉRREZ, A. (1990). *Estructura lingüística de la documentación: teoría y método*. Universidad de Murcia.
- GARCÍA GUTIÉRREZ, A. & LUCAS FERNÁNDEZ, R. (1987). *Documentación automatizada en los medios informativos*. Madri, Paraninfo.
- GARDIN, J.-C. (1968). Análise documentária e análise estrutural em arqueologia; trad. de J. T. Coelho Netto. In: CORDIER, S., ed. *Lévi-Strauss*. São Paulo, Documentos. pp.81-86.
- GARDIN, J.-C. (1970). Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J. et MARANDA, P. (orgs.). *Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème. anniversaire*. (Tirage à part). Paris, Mouton.p.628-657.
- GARDIN, J.-C. (1973a). Document analysis and linguistic theory. *J. of Documentation*, v.29, n.2, pp.137-168, June.
- GARDIN, J.-C. (1973b). Linguistique et documentation. *Bollettino d'Informazioni*, v.13, n.2/3, pp.67-85.
- GARDIN, J.-C. (1974). *Les analyses de discours*. Paris, Delachaux et Niestlé.
- GARDIN, J.-C. (1987). La logique, naturelle au autre, dans les constructions des sciences humaines. *Revue européenne des sciences sociales*, v.25, n.77, pp.179-195.
- GARDIN, J.-C. (1987). Principes généraux. In: —(org.). *Systemes experts et sciences humaines*. Paris, Eyrolles, pp.3-55.
- GARDIN, J.-C. (1989). Artificial Intelligence and the future of semiotics: an archaeological perspective. *Semiotica*, v.77, n.1/3, pp.5-26.
- GARDIN, J.-C. (1991). Introduction. In: —. *Le calcul et la raison: essais sur la formalisation du discours savante*. Paris, École des Hautes Études en Sciences Sociales. pp. 17-37.
- GARDIN, J.-C. et alii. (1968). *L'automatisation des recherches documentaires: un modèle général "Le SYNTOL"*. ed. revue et augmentée. Paris, Gauthier-Villars.
- GARDIN, J.-C. et alii. (1987). *La logique du plausible: essais d'epistémologie pratique en sciences humaines*. Paris, Ed. de la Maison de l'Homme.
- GOLDMANN, L. (1984). *As ciências humanas e a filosofia*. São Paulo, DIFEL.
- GRANGER, G.G. (1974). *Filosofia do estilo*. Trad. de Scarlett Zebetto Marton. São Paulo, Perspectiva; EDUSP. (Estudos, 29)
- GREIMAS, A.-J. (1976). *Semântica estrutural: pesquisa de método*. Trad. de Haqira Osakabe e Izidoro Blikstein. 2.ed. São Paulo, Cultrix/EDUSP.
- HJELMSLEV, L. (1975). *Prolegômenos a uma teoria da linguagem*. Trad. José Teixeira Coelho Netto. São Paulo, Perspectiva. (Estudos, 43)
- IBICT (1984). *Diretrizes para elaboração de tesouros monolíngües; projeto coord. por Hagar Espanha Gomes*. Brasília, IBICT.
- ISO (1987). *Principles and methods of terminology*. ISO 704, 1.ed.
- KOBASHI, N.Y. (1988). *Política científica e tecnológica: da análise do discurso à análise documentária*. São Paulo, ECA-USP. (Dissertação de Mestrado)
- KOBASHI, N.Y. (1989). Análise documentária: considerações sobre um modelo lógico-semântico. In: CUNHA, I.M.R.F., org. *Análise documentária: considerações teóricas e experimentação*. São Paulo, FEBAB. p.45-57.
- KOBASHI, N.Y. (1989). Análise documentária: tipologias discursivas. In: CUNHA, I.M.R.F., org. *Análise documentária: considerações teóricas e experimentação*. São Paulo, FEBAB. p.31-44.
- KUHLEN, R. (1983). Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting. In: *INTERNATIONAL RESEARCH FORUM IN INFORMATION SCIENCE*, 5, Heildelberg, Germany. (Representation exchange of knowledge as a basis of information process). p.87-109.
- LE GUERN, M. (1989). Sur les relations entre terminologie et lexique. *Meta*, v.34, n.3, p.340-343.
- LOPES, E. (1987). *Fundamentos da lingüística contemporânea*. São Paulo, Cultrix.

- LOPES, M.I.V. (1990). *Pesquisa em comunicação: formulação de um modelo metodológico*. São Paulo, Loyola.
- MACEDO, N.D. & TÁLAMO, M.F.G. (MOREIRA) (1978). Resumos: subsídios para sua elaboração. *R. Bras. Bibliotec. Doc.*, São Paulo, v.11, n.1/2, pp.65-72, jan./jun.
- MANIEZ, J. (1976). *Le role de la syntaxe dans les systèmes de recherche documentaire*. Dijon, Université de France-Comté. (Thèse pour le doctorat de 3ème. cycle). 2v.
- NOVO DICIONÁRIO AURÉLIO/Novo dicionário da língua portuguesa (s.d.), org. por Aurélio Buarque de Holanda Ferreira. 1a.ed. Rio de Janeiro, Nova Fronteira.
- PATARRA, N.L. & CUNHA, J.M.P. da (1987). Migração: um tema complexo. *São Paulo em Perspectiva*, São Paulo, v.1, n.2, p.32-35.
- PEIRCE, C.S. (1977). *Semiótica*. Trad. de José Teixeira Coelho Netto. São Paulo, Perspectiva. (Estudos, 46)
- PÊCHEUX, M. (1969). *Analyse automatique du discours*. Paris, Dunod.
- PIEDEDE, M.A.R. (1977). *Introdução à teoria da classificação*. Rio de Janeiro, Interciência.
- RANGAHATHAN, S.R. (1963). *Colon Classification*. 6.ed. reprinted with amendments. Bombay, Asia Publ. House.
- RASTIER, F. (1987). Présentation. *Langages*, n.87, pp.5-19. (N.Esp. Sémantique et intelligence artificielle).
- RASTIER, F. (1987a). Répresentation du contenu lexical et formalismes de l'intelligence artificielle. *Langages*, n.87, pp.79-102. (N.Esp. Sémantique et intelligence artificielle).
- REY, A. (1977). *Le lexique, images et modèles: du dictionnaire à la lexicologie*. Paris, Armand Colin. cap.1 e 4.
- SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados. (1978). Documento básico: Sistema Estadual de Análise de Dados Estatísticos. 2.ed. rev. e atualizada. São Paulo, ago. 20p.
- SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados. Unidade de Planejamento de Informação (1978). *Uma metodologia para sistemas setoriais de informação: Governo do Estado de São Paulo*. São Paulo, abr. 24p. (Metodologia, 1)
- SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados. Unidade de Planejamento de Informação (1978). *Metodologia para sistemas setoriais de informação: Governo do Estado de São Paulo - Setor Saúde*. São Paulo, out. 34p. (Metodologias, 2)
- SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados. Unidade de Planejamento de Informação (1978). *Metodologia para sistemas de informação nas Secretarias de Estado: Governo do Estado de São Paulo*. São Paulo, abr. 22p. (Metodologia, 3)
- SÃO PAULO (Estado) Secretaria de Economia e Planejamento. Coordenadoria de Análise de Dados. Unidade de Planejamento de Informação (1978). *Metodologia para capacitação de recursos humanos em sistemas de informação: Governo do Estado de São Paulo*. São Paulo, dez. 25p. (Metodologia, 4)
- SHERA, J.H. & EGAN, M.E. (1969). *Catálogo sistemático: princípios básicos e utilização*. Trad. de Maria Neile Teles Landau. Brasília, UnB.
- SMIT, J.W. (1973). *Les langages documentaires comme metalangages du discours scientifique*. Paris, École Pratique des Hautes Études. (Mémoire)
- SMIT, J.W. (1974). Análise semântica e análise documentária. *Rev. Bras. de Semiótica*, n.1, p.168-177.
- SMIT, J.W. (NATALI) (1976). *De l'analyse documentaire à l'analyse poétique: a propos des critiques des "Chats" de Baudelaire*. Paris, École Pratique des Hautes Études en Sciences Sociales. (Thèse de 3e. Cycle)
- SMIT, J.W. (NATALI) (1978). Documentação e lingüística: inter-relação e campos de pesquisa. *Rev. Bras. Bibliotecon. Doc.*, v.11, n.1/2, jan./jun.
- TÁLAMO, M.F.G.M. (1982). As máximas conversacionais de Grice. In: ———. *Comunicação e funcionamento da linguagem*. São Paulo, ECA-USP. (Dissertação de Mestrado). pp.54-63.
- TÁLAMO, M.F.G.M. (1987). A definição semântica para a elaboração de glossários. In: SMIT, J., org. *Análise documentária: a análise da síntese*. Brasília, IBICT. p.87-98.
- TÁLAMO, M.F.G.M. (1991). *Terminologia, linguagens documentárias e representação do conhecimento*. Trab. apresentado ao Congresso Nacional de Informação - SUCESU, 23-27 set.1991.
- TÁLAMO et alii. (1990). *Instrumentos de controle terminológico: limites e funções*. Trab. apresentado ao II Simpósio Latinoamericano de Terminologia. Brasília, 9-4 set, 1990.
- UNISIST (1978). *Princípios de indexação*. R. Esc. Bibliotec. UFMG, Belo Horizonte, v.10, n.1, pp.83-94, mar.
- VALE, E.A. (1987). Linguagens de indexação. In: SMIT, J., org. *Análise documentária: a análise da síntese*. Brasília, IBICT, p.11-26.
- VAN DIJK, T.A. (1977). Gramáticas textuais e estruturas narrativas. In: CHABROL, C., apres. *Semiótica narrativa e textual*. Trad. de Leyla Perrone Moisés e outros. São Paulo, Cultrix; EDUSP, pp.196-229.
- VAN DIJK, T.A. (1992). *Cognição, discurso e interação*. São Paulo, Contexto.

- VANOYE, F. (1991). **Usos da linguagem: problemas e técnicas na produção oral e escrita.** Trad. e adaptação. de Clarisse Madureira Sabóia e outros. 8.ed. São Paulo, Martins Fontes.
- VERÓN, E. (1980). **A produção do sentido.** Trad. de Alceu Dias Lima e outros. São Paulo, Cultrix.
- VIET, J. (1985). **Macrothesaurus para el procesamiento de la información relativa al desarrollo económico y social.** 3. ed. New York, Naciones Unidas.
- VIET, J. (1986). **Thesaurus POPIN: thesaurus multilingüe de população.** Trad. e adapt. de Letícia B. Costa. São Paulo, SEADE.
- VICKERY, B.C. (1963). **La classification à facettes: guide pour la construction et l'utilisation de schémas spéciaux.** Paris, Gauthier-Villars.
- VICKERY, B.C. (1980). **Classificação e indexação nas ciências.** Trad. de Maria Christina Girão Pirolla. Rio de Janeiro, BNG/Brasilart.
- VICKERY, B.C. (1986). Knowledge representation: a brief review. *J. of Documentation*, v.42, n.3, p.145-159.