

O processo de construção da informação documentária e o processo de conhecimento*

Marilda Lopes Ginez de Lara**

*Tal como ocorre no processo de conhecimento (e de aprendizagem), a construção da informação recorre a recortes de conteúdo, à analogia e à generalização. Discute-se essas semelhanças a partir da análise do texto *Marco Polo e o unicórnio*, de Umberto Eco, procurando mostrar que a informação documentária não existe 'a priori', mas é resultado da segmentação de conteúdos feita a partir de hipóteses de organização. Esse processo ganha referências concretas a partir do uso da terminologia, permitindo a formulação de linguagens documentárias mais consistentes.*

Palavras-chave: Análise documentária; Linguagem documentária; Informação documentária; Terminologia; Lingüística; Semiótica

Recebido em 16.04.2002 - Aceito em 16.08.2002

127

Introdução

Preteu-se, neste artigo, estabelecer uma analogia entre o processo de conhecimento¹ e o processo de representação e construção da informação documentária², tendo como ponto de partida o texto *Marco Polo e o unicórnio*, do livro *Kant e o ornitorrinco*, de Umberto Eco (ECO, 1998), sobre as experiências do explorador ao tentar classificar e nomear um animal desconhecido.

Na concepção de Eco, o texto é uma máquina preguiçosa que solicita a participação do leitor na compreensão dos ditos e no preenchimento dos não ditos e, conseqüentemente, toda leitura é parcial e dependente da experiência anterior do leitor, sua motivação e objetivos de leitura, mais seu estoque prévio de conhecimento. Na leitura que se fará objetiva-se reter alguns aspectos que julga-se pertinentes para discutir os processos de classificação, representação, construção e uso da informação documentária.

* Texto baseado no artigo O Unicórnio (o Rinoceronte, o Ornitorrinco...), a Análise Documentária e a Linguagem Documentária, publicado na revista eletrônica *DataGramaZero - Revista de Ciência da Informação*, v.2, n.6, dez. 20001.

** Profª. do CBD-ECA-USP, Doutora em Ciência da Comunicação - larama@usp.br

¹ Entendemos por processo de conhecimento aquele que se relaciona à apreensão de um objeto (objeto de conhecimento) e sua representação. Muitas são as interpretações de apreender e representar, porém há relativa concordância sobre o fato de que o conhecimento, além de uma atividade intelectual, é uma atividade intelectualmente motivada. Para a filosofia contemporânea (pragmatista), o processo de conhecimento é necessariamente ligado à experiência (FERRATER MORA, J. *Dicionário de filosofia*. São Paulo: Martins Fontes, 1998).

² O qualificativo documentária pretende distinguir entre a acepção geral de informação (notícia, dado, indagação, sinal etc.) e o conceito de informação como resultado do processo de análise documentária: produto documentário que designa um conteúdo, que se apresenta sob determinada forma (resumos, descritores) e que visa o estabelecimento de uma relação comunicativa particular.

A analogia na construção das representações

O processo de conhecimento se realiza fundamentalmente através da analogia. Frente a um fenômeno desconhecido *"reagimos por aproximação, procuramos aquele recorte de conteúdo, já presente na nossa enciclopédia, que bem ou mal parece prestar contas do novo fato"* (ECO, 1998, p.55). Ao se defrontar, em Java, com animais nunca vistos (hoje, rinocerontes), Marco Polo procura identificá-los (e nomeá-los) com base no seu estoque anterior de conhecimentos: lançando mão de características de animais que já conhece e a partir das descrições disponíveis em sua cultura, tenta estabelecer relações e, por aproximação, classifica o novo a partir de referências existentes. Se de imediato identifica-os com os unicórnios - seu corpo, as quatro patas, um chifre acima do nariz etc., - rapidamente verifica que a classificação feita não é inteiramente própria, uma vez que, nessa comparação, constata a existência de diferenças relativamente ao modelo anterior.

Os animais que se lhe apresentam agora

"não são brancos e ágeis, mas têm 'pêlo de búfalos e pés de elefantes', o chifre é negro e desgracioso, a língua espinhosa, a cabeça semelhante àquela de um javali: É uma besta muito feia de se ver. Não é, como se diz por aqui, que ela se deixe pegar como uma donzela, mas acontece ao contrário" (MILIONE 143, *apud* ECO, op. cit.).

Tais características não condizem exatamente àquelas do unicórnio presentes em sua enciclopédia, motivo pelo qual prossegue na sua tentativa de classificação. Neste momento, deve tomar uma decisão: ou segmenta novamente o conteúdo sugerindo um novo animal, ou modifica a intenção dos unicórnios, alterando sua descrição anterior para acomodar, na classe, os novos animais.

Se Marco Polo optasse por modificar a intensão dos unicórnios para poder dar conta dos animais que agora conhecia, acrescentaria características ao conceito original. Esse acréscimo resultaria, entretanto, na perda de especificidade dos unicórnios. A intensão do conceito seria um amálgama indiferenciado.

Processo similar experimenta-se ao se construir representações de conteúdo na área da ciência da informação. Opera-se sempre por analogia e generalização, procurando reunir os conceitos a partir de seus traços comuns, numa tentativa de organizar a informação e de garantir sua recuperação posterior. Ao classificar (e indexar) segmenta-se o conteúdo a partir de referências que já possuímos, formando agrupamentos em função de suas propriedades comuns, ou mais exatamente, das características que se julga pertinentes para os nossos propósitos, já que nunca se esgota as propriedades das coisas e dos fenômenos. Frente ao que se desconhece, ou não se conhece suficientemente, encontra-se dificuldades para realizar segmentações e formar grupos.

Para que as coisas fiquem mais claras, retoma-se alguns conceitos acima verificando seu significado sob a ótica da terminologia³, antes de prosseguir na análise:

- *objeto*: tudo o que pode ser percebido ou concebido;
- *conceito*: unidade de conhecimento criada pela combinação de características únicas;

³ Terminologia: ciência que estuda a estrutura, a formação, o desenvolvimento, o uso e a gestão de terminologias em diferentes domínios (ISO 1087-2000).

A enciclopédia é o conjunto de *"todas as interpretações, concebíveis como a biblioteca das bibliotecas, onde uma biblioteca é também um arquivo de toda a informação não verbal de algum modo registrada, das pinturas rupestres às cinematecas"* (ECO, 1991, p.113). Mas ela é um postulado semiótico, ou seja, não é passível de descrição na sua totalidade, já que as possibilidades interpretativas são infinitas, às vezes contraditórias e implicam sistematicamente novas segmentações no *continuum* da realidade. A enciclopédia é uma espécie de competência global constituída de diversas enciclopédias parciais que são ativadas, à medida do necessário e individualmente, no processo interpretativo. Enquanto conjunto não ordenado de marcas pode ser organizado parcialmente, sob forma de dicionário,

"toda vez que queremos circunscrever a área de consenso dentro da qual um discurso se move (idem, p.133). É o conhecimento enciclopédico, e não apenas o lingüístico, que nos permite interpretar metaforicamente um enunciado do tipo "O cão do meu vizinho uivou a noite toda".

Retomando nossa análise, verifica-se que as informações presentes na enciclopédia de Marco Polo não permitiam dar conta adequadamente da descrição dos novos animais, porque as características do unicórnio ali retidas não eram suficientemente extensas. Ao tentar corrigir a descrição inicial dos unicórnios, Marco Polo *"modifica a intensão⁴ deixando a extensão sem juízo"*, ou seja, altera a definição original incluindo nela novas características, mas incorre na descaracterização da classe que permitiria acomodá-los. A modificação da intensão do termo unicórnio pelo acréscimo (ou eliminação) de características, interfere na extensão do termo, do que resulta que nem os unicórnios, nem os novos animais, podem pertencer, de direito, à nova classe resultante, pois os traços não são específicos nem de um, nem de outro e não servem mais para descrevê-los num mesmo conjunto. Frente a essa situação resta a alternativa de uma nova segmentação do conteúdo: Marco Polo poderia acrescentar um novo animal ao universo dos seres vivos.

Caso mais difícil de resolver teria enfrentado Marco Polo se tivesse chegado à Austrália e tivesse visto um ornitorrinco. *"O ornitorrinco é um estranho animal, que parece concebido para desafiar qualquer classificação..."* tendo em média uns cinquenta centímetros, cerca de dois quilos, o corpo chato coberto por uma pelugem marrom-escuro, sem pescoço, e com uma cauda de castor, bico de pato, de cor azulada por cima e rosa ou matizada por baixo, sem pavilhões auditivos, as quatro patas terminam com cinco dedos espalmados, mas com garras; fica bastante tempo debaixo d'água (e ali come) para ser considerado um peixe ou um anfíbio, a fêmea põe ovos, mas amamenta os próprios filhotes, mesmo se não se vê nenhuma teta (além disso, não se vê no macho sequer os testículos, que são internos) (*idem*, p.55-56). Um animal assim, tão estranho, representaria uma dificuldade de classificação: mamífero, anfíbio ou peixe? Nesse caso, as características do unicórnio, mesmo um pouco modificadas, não se prestariam à descrição e o estranho animal fugiria aos vários modelos.

Fica claro a partir deste texto que os problemas relacionados ao conhecimento derivam, por raciocínio hipotético, do conhecimento de fatos externos e dos acontecimentos precedentes (PEIRCE *apud* ECO, 1991, p.56). Eco sugere que se Peirce analisasse o caso talvez tivesse afirmado que Marco Polo, antes de propor um

⁴ Na tradução publicada no Brasil, pela Record, o termo intensão está grafado com "ç", erroneamente. Trata-se de dois conceitos distintos.

a) Ao considerar a informação como um dado, a classificação só pode ser reprodução. Essa crença está fundada na preocupação com a busca da essência ou, dito de outro modo, da informação objetiva. Tal procedimento remeteria ao *polimento do espelho*, já que a informação estaria dada aprioristicamente, requerendo um esforço para resgatá-la do documento;

b) A ativação daquilo que é considerado *conteúdo do documento* é feita a partir de uma combinatória universal - uma grade de classificação única válida para quaisquer contextos. Os sistemas bibliográficos universais reproduziriam a realidade e permitiriam, por esse motivo, que se representasse fielmente os conteúdos dos textos.

Já a partir de uma perspectiva pragmática, a informação é vista sob outro ângulo, para o que corrobora o texto de Umberto Eco. A informação nunca é um dado, ao contrário, é sempre uma construção, porque:

a) As generalizações a que procedemos são um produto cultural, fruto de nossa experiência com a realidade;

b) Enquanto produtos culturais, as classificações (e generalizações) partem necessariamente de hipóteses; podem existir, concomitantemente, distintas hipóteses, que privilegiam determinadas características em detrimento de outras. Sob esse aspecto, uma classificação não pode ser avaliada como certa ou errada, mas mais, ou menos, adequada para determinados propósitos, para o que se selecionam determinados aspectos e não outros. Dito de outra forma, não se trata de polir o espelho para encontrar a informação, mas de construí-la em função de objetivos. Sob essa perspectiva, compreende-se que qualquer proposta de classificação universal é datada, isto é, parcial, momentânea e centrada num ponto fixo de enunciação. Isso explica porque as classificações bibliográficas de natureza enciclopédica⁴ e quase que exclusivamente hierárquicas oferecem dificuldade para dar conta de distintos aspectos não contemplados de início. Tais propostas classificatórias têm uma articulação dura entre o plano do conteúdo e o plano de expressão e funcionam exclusivamente a partir de operações de encaixe.

c) A informação envolve sempre uma relação de pertença ao que nos rodeia (GADAMER, *apud* NUNES, 1998, p.10). Para haver informação é preciso que ela faça sentido para alguém, ou seja, que exista algum nível de compartilhamento a partir do qual se estabeleça um vínculo para a interpretação e o sentido (BAITELLO, 1994).

No âmbito documentário, portanto, representar conteúdos para constituir informação significa recortar, segmentar. Como qualquer segmentação, ela não é neutra, mas tributária de hipóteses de organização. Além disso, conscientes ou não, desde que usamos a língua estamos utilizando convenções: a língua é o primeiro sistema modelizante a partir do qual todos os outros sistemas derivam. Também não é necessariamente unívoca a interpretação do usuário, sujeito que também tem suas convicções e está apto a formular hipóteses, tanto quanto o produtor do sistema de representação. Com efeito, para que se possa falar de transmissão de informação documentária é necessário formular hipóteses (e explicitá-las) razoavelmente compartilhadas pela comunidade de usuários do sistema que se propõe.

⁴ O entendimento de *enciclopédicas*, neste contexto, refere-se à pretensão universalizante de abarcar, num só sistema, todo o universo do conhecimento. A enciclopédica, entretanto, é um postulado semiótico: o que se consegue registrar é sempre parcial frente à infinita possibilidade de interpretação enciclopédica.

Uma reformulação da noção de conteúdo implica verificar que, sob a ótica do enunciador, a constituição de sistemas de informação para transferência é uma operação carregada de intencionalidade. A representação não tem um fim em si mesma, ao contrário, é uma construção que persegue objetivos específicos. Num determinado contexto uma vaca é um tipo de mamífero; noutra, um bicho sagrado etc. Contrariamente à visão tradicional que distingue as informações pelas características (documentos de bibliotecas, de museus etc.), são suas funções informativas que devem ser privilegiadas (SMIT, 1988, p.4-5) para a constituição de sistemas. Necessidades pragmáticas determinam os recortes do conteúdo, de forma a procurar viabilizar um fluxo informativo efetivo: um mesmo universo documental pode ser organizado e representado de diversas maneiras, pautando-se nos distintos propósitos a que se destinam. Sob a ótica do enunciatário, a informação depende de vínculos: a informação existe quando faz sentido para o usuário.

Hipóteses de organização dos instrumentos de representação da informação

Ao proceder à classificação dos animais, a hipótese de organização utilizada por Marco Polo era tributária das segmentações propostas pela cultura de sua época. Também as linguagens da biblioteconomia e da documentação geralmente partem de hipóteses, pontos de vista que estão na base das formas de organização dos domínios visados. Tais formulações variam histórica e funcionalmente, conforme os pressupostos considerados e os objetivos da classificação.

a) Os sistemas de classificação bibliográfica universal, como o *Dewey Decimal Classification* e a Classificação Decimal Universal, têm sua origem na proposta baconiana de divisão do conhecimento que foi considerada, durante muito tempo, um parâmetro estável para organizar o universo bibliográfico. Com o crescimento do conhecimento registrado e, conseqüentemente, das segmentações propostas para analisar a realidade, esses sistemas foram submetidos a inúmeras modificações para dar conta do novo. As sucessivas edições afastam-se, progressivamente, da hipótese inicial, mas sofrem os efeitos da ausência de alteração radical nas formas de segmentação do novo universo documental. Passam a conviver, nesses sistemas, segmentações de origem distinta, do que decorrem amálgamas de características provenientes de distintas segmentações, na mesma hierarquia. É por esse motivo que tais sistemas têm sido questionados quanto à eficácia para cumprir as funções de organização da informação.

b) Listas de cabeçalhos de assunto, como a *Library of Congress Subject Headings – LCSH-*, pautam-se na garantia literária (presença dos termos na literatura) para inventariar sugestões de entradas de assunto, porém esse procedimento não caracteriza efetivamente uma proposta de segmentação assente num princípio organizador, razão pela qual elas não formam um conjunto uno. Sua apresentação sob a forma de tesouro, nas edições contemporâneas, também não tem a contrapartida relativa à estruturação de seu conteúdo, já que as relações entre os seus termos não são formuladas rigorosamente quer em relação ao controle da sinonímia, quer na distinção entre relações hierárquicas e associativas (WELLISCH, 1991, p.388). De forma semelhante podem ser caracterizados os inventários construídos empiricamente pelas bibliotecas a partir dessas listas, cujos problemas são agravados na ausência de



parâmetros rigorosos de tradução. A garantia literária, tal como é utilizada nas listas de cabeçalhos de assunto, é aleatória e não chega a configurar uma hipótese de organização, do que decorrem dificuldades para o engendramento de mecanismos inferenciais.

c) A recuperação de textos feita com base na frequência e na ocorrência de termos também não caracteriza uma segmentação do tipo classificatório, já que o recorte é pulverizado para as palavras retidas no processo de extração. O parâmetro classificatório, baseado no relacionamento de características (semelhança, diferença, associações, equivalência) é, neste caso, substituído por um parâmetro estatístico. A identificação da informação resulta altamente arbitrária e carente de vínculos necessários à instauração de situações de comunicação.

d) Os tesouros representam uma das formas mais consistentes de apresentar uma proposta de organização de um domínio, já que são formulados segundo princípios lógico-semânticos através dos quais é possível constituir um todo significativo. As diretrizes de elaboração de tesouros prescrevem a identificação de relações de natureza hierárquica, associativa e de equivalência entre os termos, relações estas que provêem o arranjo necessário a uma proposta de organização do conhecimento. Todavia, as normas não dizem como encontrar as referências para sedimentar o processo de identificação e de relacionamento entre os termos, razão pela qual há muitos tesouros que o são apenas no nome, pois não formulam claramente os vértices a partir dos quais devem ser organizados os termos. Para que os tesouros sejam estruturas significantes, seus termos devem remeter a conceitos de um domínio e a delimitação de conceitos e termos não é uma operação simples, pois remete a princípios teórico-metodológicos específicos.

e) Os tesouros terminológicos são instrumentos que usam simultaneamente princípios de organização de tesouros e de organização de sistemas de conceitos. A terminologia teórica e concreta permite fundamentar a delimitação de domínios e a operação de seleção dos termos que lhe são próprios. No âmbito dos domínios, os termos correspondem a conceitos delimitáveis a partir de redes de relacionamento que moldam o universo conceitual visado.

O simples uso dos termos de uma área (tal como o processo de garantia literária utilizado na LCSH) não constitui um trabalho terminológico e não garante, por consequência, que o sistema conceitual do domínio do saber ou da área de atividade (que se sustenta por uma hipótese) esteja representado. Tal procedimento resulta da visão ingênua que associa o significado dos termos de uma linguagem documentária⁵ com algo que está fora dela, como se o simples agrupamento de palavras típicas fosse suficiente para caracterizar uma linguagem (TÁLAMO, 1997). Se uma linguagem é um sistema (ou estrutura), a determinação do sentido de cada termo se estabelece na relação mútua entre os termos do mesmo conjunto. Uma linguagem pressupõe a noção de arranjo, e esse arranjo se desenha relativamente a uma proposta de organização.

⁵ Linguagem documentária é uma linguagem construída para a indexação e recuperação da informação.

A terminologia como fonte para a identificação da hipótese de organização de um domínio

As segmentações (ou recortes de conteúdo) não podem ser operações aleatórias se visam a construir linguagens documentárias que permitam instaurar situações de comunicação. A busca de uma hipótese para a operacionalização desse recorte e, em seguida, sua organização para constituir uma estrutura significativa, remete necessariamente a objetivos funcionais e a uma delimitação clara do domínio visado. É nesse sentido que a terminologia subsidia os trabalhos da ciência da informação.

A terminologia trabalha a partir do conceito. O conceito, como já vimos, é uma unidade de conhecimento formada por uma combinação de características. A formação do conceito é um processo de abstração que tem como base a seleção de propriedades (características) de um conjunto de um ou mais objetos. O conceito, conseqüentemente, não tem vida independente, mas é o resultado de uma proposição feita dentro de um domínio do saber. Cada conceito é determinado pela posição que ocupa no conjunto estruturado de conceitos, denominado sistema conceitual.

A unidade básica da terminologia é o termo, designação verbal do conceito dentro de um domínio específico. A designação é uma representação do conceito por um signo que o denota (um símbolo, um nome ou um termo). Para elaborar um sistema de conceitos, enquanto conjunto estruturado segundo as relações que os unem, o trabalho terminológico parte de definições. A definição é a representação de um conceito por um enunciado descritivo que permite diferenciá-lo de conceitos relacionados dentro de um domínio de especialidade.

O domínio é um campo especializado do conhecimento (áreas do saber ou de atividades) expresso por uma língua de especialidade (LSP⁶, ou língua utilizada pelo domínio e caracterizada pelo uso de meios de expressão lingüística particulares). Enquanto subconjunto do sistema lingüístico, as línguas de especialidade refletem o modo mais ou menos normalizado de se comunicar num dado domínio do saber, o que seria impossível na língua geral onde os sentidos são variáveis.

As definições tomadas como ponto de partida podem ser intensionais ou extensionais. A definição intensional (ou por compreensão) descreve a intensão do conceito indicando o conceito superordenado e as características distintivas do conceito em foco. Dito de outro modo, faz menção ao conceito genérico mais próximo, já definido ou supostamente conhecido, e às características distintivas que delimitam o conceito a ser definido. Já a definição extensional (ou definição por extensão) descreve o conceito enumerando todos os conceitos subordinados segundo um critério de subdivisão⁷.

O trabalho terminográfico é realizado sobre um conjunto de documentos que são selecionados como representativos do universo do domínio enfocado: dicionários e glossários já existentes, textos produzidos na área dos quais são criteriosamente coletados termos-candidatos. Normalmente, a coleta é realizada utilizando-se, como suporte, fichas terminológicas, onde se transcrevem literalmente as definições formais presentes na documentação, ou os contextos onde aparecem os termos-candidatos. As

⁶ LSP: Language for special purposes.

⁷ Lembre-se aqui as tentativas de classificação feitas por Marco Polo para identificar o 'rinoceronte', primeiramente, a partir de uma definição intensional. A impossibilidade de generalização motivou-o a modificar a definição original, opção abandonada em seguida pelo prejuízo que causou à definição extensional.

fichas terminológicas são, em seguida, analisadas, comparando-se as características presentes nas definições formais ou nas ocorrências de uso dos termos. Esse procedimento fundamenta a seleção dos termos considerados mais apropriados para representar os conceitos, possibilita a identificação de variantes e apoia a construção ou reconstrução de definições. Embora nem todos os trabalhos terminológicos cheguem efetivamente a apresentar mapas conceituais dos domínios (muitas vezes o trabalho se encerra na elaboração de um dicionário ou de um glossário), é a partir do processo definicional que são criadas as condições para se estabelecer redes de relações entre os termos⁸.

Os dicionários terminológicos constituem, assim, uma ferramenta imprescindível para a construção da linguagem documentária, já que permitem conferir referência aos descritores. Se antes a biblioteconomia e a documentação trabalhavam empiricamente a partir de referências da classificação filosófica (ênfatisando apenas a segmentação, mas não sua expressão lingüística) e, em seguida, para dar conta das formas significantes, a partir de palavras (através dos processos de extração baseados em freqüência ou ocorrência) ou da seleção empírica de unidades significativa (sem a definição daquilo que poderia caracterizar uma unidade significativa), a partir da Terminologia ela passa a contar com instrumentos que trabalham com o termo, unidade que representa o conceito dentro de um domínio ou área de atividade.

A palavra é uma unidade do léxico de uma língua; a representação do léxico constitui o dicionário da língua. O empreendimento de organizar sistematicamente o léxico de uma língua é muito difícil, senão inviável, pelo seu tamanho e pela dificuldade em dar conta das várias possibilidades de sentido⁹. Já o termo é uma unidade da terminologia. Seu significado é fixado no interior de um domínio dado, ou seja, corresponde à palavra contextualizada no uso (no discurso) de determinado campo do saber (LE GUERN, 1989). O significado de um termo é delimitado pelas relações que estabelece com os outros termos do mesmo domínio (fixado no processo definicional, para facilitar a comunicação).

O uso do termo (e de mapas conceituais de domínios) permite à ciência da informação transformar uma atividade anteriormente empírica, de categorização, num trabalho teórico e metodologicamente sedimentado em referências concretas dos domínios de especialidade. O uso da terminologia teórica e das terminologias concretas cria condições para formular propostas de organização das linguagens documentárias à medida que provê fundamentos para a seleção e estruturação dos termos.

A linguagem documentária não é terminologia: objetivos do sistema de informação e referências de uso para a formulação das hipóteses de organização

Embora o uso da terminologia seja crucial para referendar o processo de construção de linguagens documentárias - e com isso possibilitar melhor representação e recuperação da informação -, ela não garante por si só o sucesso da comunicação

⁸ Alguns trabalhos terminográficos partem, inicialmente, de um mapa conceitual construído mais ou menos empiricamente, que é feito à medida que o trabalho de coleta e organização se desenvolve.

⁹ Os tesouros de língua, como o *Thesaurus de Roget*, são exemplos de léxicos da língua organizados por associação de idéias e representam uma possibilidade de organização, entre outras.

documentária¹⁰. Isso significa dizer que as hipóteses de organização de tais linguagens são mais complexas do que poderia parecer inicialmente, uma vez que a atividade de informar requer, além dos parâmetros de um sistema conceitual, a observação das necessidades de uso e recepção da informação.

A linguagem documentária, portanto, não é a simples reprodução da terminologia de determinada área. Num tesouro terminológico, que dentre as linguagens documentárias se destaca por apresentar uma boa solução para estruturação do eixo paradigmático¹¹, as unidades são, antes, preferenciais. Os descritores são termos¹² de um tesouro que devem ser empregados para representar uma noção (ou conceito) presente num texto, ou para estabelecer uma equação de busca; os não-descritores, termos que não devem ser empregados, mas incorporados à linguagem como referência de acesso a um descritor a ser utilizado em seu lugar.

O termo preferencial é escolhido como descritor entre um conjunto de termos equivalentes (ISO 5127/6-1983-E/F). A definição apontada na norma não esclarece que critérios devem ser utilizados para a preferência, porém pode-se considerar que eles remetam simultaneamente à garantia literária (como o conceito é designado com maior frequência na literatura) e à garantia do uso (como os usuários procuram por informação¹³). Embora não existam orientações precisas para preferir ou preterir um termo, é lícito supor que o critério pragmático (uso pela literatura e pela comunidade de usuários) deve ser perseguido, para o que deveria existir metodologia. A existência de um critério pragmático, embora venha sendo tratado de maneira empírica (exceto pelo fato de que a terminologia também se refere no compartilhamento de linguagens), é expressão evidente de que os sistemas documentários, construídos para transmitir informação, não podem descolar-se das referências de uso. Convém investigar, dentre outras possibilidades, como a socioterminologia poderia contribuir para definir procedimentos de identificação e incorporação de tais referências à organização das linguagens documentárias.

A linguagem documentária também não é terminologia porque ambas se propõem tarefas distintas. Se a terminologia visa estudar a estrutura, a formação, o

¹⁰ Utilizou-se o termo comunicação documentária para identificar uma situação particular de comunicação. As representações documentárias não têm um fim em si mesmas, ao contrário, são utilizadas para transmitir informação. No processo de transferência da informação pode-se identificar o emissor (instituição que formula o sistema documentário), o canal (veículo através do qual são veiculadas as mensagens documentárias), o código intermediário (a linguagem documentária usada na indexação e na recuperação que deve funcionar, simultaneamente, como um sistema de comunicação e de significação) e o destinatário (usuário final do sistema). Os 'signos documentários' podem ser identificados com as representações construídas com a intermediação dos códigos; as mensagens documentárias, as representações feitas a partir da seleção de signos documentários ou de combinação deles nas equações de busca (LARA, 1993). Hutchins sugere, inclusive, que nos sistemas de indexação também ocorre a semiose. Para o autor, o signo (s) é a seqüência de formas físicas da linguagem documentária (LD): os intérpretes (i) são os indexadores e os usuários dos índices. Os designata (d) são as relações entre a forma física dos signos da LD e seus referentes (r), o conteúdo informacional dos documentos; os efeitos (e) são as reações dos usuários (ex.: procurar o documento ou não, de acordo com seu julgamento sobre sua relevância); e os contextos (c) são a organização física e o ambiente geral do índice e do centro de informação etc. (HUTCHINS, 1975).

¹¹ Uma linguagem documentária, para Gardin, é constituída de um léxico, uma rede paradigmática e uma rede sintagmática (CINTRA et al., 2002). O eixo paradigmático das LDs é representado pela estruturação vertical (conjunções e disjunções) e horizontal (associações não-hierárquicas) do conjunto de termos do léxico. Já o eixo sintagmático é muito precário se comparado ao das Línguas Naturais, pois os recursos de combinação entre os termos são muito restritos (nos tesouros, a sintaxe é realizada através de operadores booleanos: *and*, *or*, *not*). O investimento nas linguagens documentárias sintagmáticas, como foi o caso do SYNTOL - *Syntagmatic Organization Language* (Gardin), do *Semantic Code ou WRU* (Kent, Perry e Melton) e da Análise Codificada (Pagès), todos entre as décadas de 60-70, precisaria ser retomado.

¹² Termo, neste contexto, é uma designação genérica e não se confunde com o termo da Terminologia.

¹³ As principais referências de uso são consideradas através das relações de equivalência, muito embora não sejam as únicas. A função das relações de equivalência é facilitar o acesso do usuário ao sistema. Medida de economia em sistemas que ocupavam muita memória em disco até bem pouco tempo, interessa, atualmente, apenas como medida de controle: a proposta de uma entrada única já não é imperativa em face aos recursos hipertextuais, porém permanece a necessidade de dar conta da variabilidade das formas de busca.

desenvolvimento, o uso e a gestão de terminologias em diferentes domínios, a ciência da informação, quando constrói linguagens documentárias, tem como objetivo assegurar a organização e a transferência da informação. Uma vez que a informação é uma construção, a linguagem documentária, embora se aproprie das terminologias e dos sistemas conceituais por ela mapeados, se propõe como modo de organização da informação, para o que articula temas, subtemas e termos orientados para dar conta de conjuntos documentários e para transferir informação a determinados públicos segundo objetivos específicos. Com efeito, segmenta, a seu modo, o universo enfocado segundo perspectivas pragmáticas. A partir dessa segmentação, que se apresenta como proposta organizadora, a linguagem documentária configura-se como instrumento facilitador da comunicação em contextos documentários específicos.

Portanto, se para o senso comum as referências são genéricas e não necessariamente localizadas, podendo remeter a enciclopédias variadas, para a transferência da informação é necessário conjugar critérios terminológicos e pragmáticos que permitam, por um lado, identificar a porção da enciclopédia a ser ativada (a partir da qual seja possível interpretar um termo) e, por outro, determinar os vínculos de comunicação que funcionem efetivamente em contextos informacionais. Se Marco Polo estivesse interessado, ao invés de classificar os novos animais, em saber para que servem, certamente teria selecionado características diferentes.

138

Comentários finais

Os processos de tratamento e recuperação de informações apresentam características similares às que ocorrem no processo de conhecimento em geral. Não é sem razão que as metodologias de organização da informação fazem uso de referenciais da lógica, da semiótica, da lingüística, da semântica, da pragmática e das ciências cognitivas. Procuram-se, a partir desses referenciais, operacionalizar com maior rigor os procedimentos de identificação e registro das conjunções (superordenação), das disjunções (subordinação), das associações não-lineares e das equivalências entre os termos, à semelhança do que ocorre, grande parte das vezes de forma inconsciente, em qualquer processo de conhecimento.

O que particulariza a atividade documentária é sua função: organizar para transferir, transferir para viabilizar a apropriação da informação. Nessa perspectiva, compreender como se desenvolve o processo interpretativo e identificar quais são as condições mínimas para que ele se desenvolva com eficácia em contextos informacionais pode significar a diferença entre simplesmente estocar e transmitir informação para o uso efetivo.

A exploração do texto de Eco permite verificar que os problemas enfrentados para a representação (ou melhor, construção) da informação não são simples. À semelhança do que ocorre na relação do homem com a realidade, a atividade de análise e representação documentária se caracteriza por segmentações ou recortes de conteúdo, visando imprimir certa organização a um *continuum* indiferenciado. Na vida cotidiana, as distintas formas de organização são condição para a vida em sociedade; em contextos informacionais, escolhemos modos de organização específicos para alcançar objetivos específicos. Nos dois contextos é necessário existir compartilhamento, o que se torna possível a partir da presença de vínculos de

significação. Nos sistemas informacionais, porém, onde a busca de graus de univocidade de significação é premente, faz-se necessário lançar mão de referenciais que permitam que esses vínculos tornem possível o estabelecimento de contextos de interpretação específicos. Sob esse aspecto, já não basta recorrer indiferenciadamente às enciclopédias disponíveis em nossa cultura: é preciso ir além, determinando qual dessas enciclopédias permite referendar o processo interpretativo, ao qual deve seguir-se a possibilidade de apropriação e uso da informação. No contexto documentário, entendido como o conjunto de fatores externos que influenciam o processo de análise documentária (HUTCHINS, 1975), portanto, a classificação do rinoceronte ou do ornitorrinco deveria ser feita segundo um critério pragmático, determinando-se, a partir da função do sistema informacional, quais os aspectos a serem retidos em prol da criação de boas condições de recepção da informação.

The process of constructing documentary information and the knowledge process

The process of constructing documentary information as well as the knowledge process use content segmentation, analogy and generalization. Starting with the discussion of Umberto Eco's text 'Marco Polo e o unicórnio', we intend to show that documentary information is not an a priori fact, but it is related to contents segmentation carried out from organization hypothesis. These operations must be supported by terminology to give more consistency to documentary languages.

Key-words: *Documentary analysis; Documentary language; Documentary information; Terminology; Linguistics; Semiotics*

Referências

- BAITELLO JR., N. Comunicação, mídia e cultura. *São Paulo em Perspectiva*, v.12, n.4, p.11-16, out./dez. 1998.
- CINTRA, A.M.M.; TÁLAMO, M.F.G.M.; LARA, M.L.G. de; KOBASHI, N.Y. *Para entender as linguagens documentárias*. 2.ed. rev. e ampl. São Paulo: APB/Polis, 2002..
- ECO, U. Marco Polo e o unicórnio. In: _____. *Kant e o ornitorrinco*. Rio de Janeiro: Record, 1988.
- ECO, U. *Semiótica e filosofia da linguagem*. Trad. de Mariarosaria Fabris e José Luiz Fiorin. São Paulo : Ática, 1991.
- FERRATER MORA, J. *Dicionário de filosofia*. São Paulo : Martins Fontes, 1998.
- HUTCHINS, W.J. *Languages of indexing and classification: a linguistic study of structures and functions*. Southgate House, Stevenage, Herts :Peter Peregrinus, 1975.
- ISO 704. *Principles and methods of terminology*. 2.ed. [s.l.:s.n] 2000.
- ISO 1087-1. *Terminology work - principles and methods*. Part 1: theory and application; Travaux terminologiques - vocabulaire. Partie 1: théorie et application. [s.l: s.n], 2000-E/F.
- ISO 5127/6. *Documentation and information - vocabulary - part 6: Documentary languages; Documentation et information - vocabulaire - partie 6: Langages documentaires*. [s.l: s.n], 1983-E/F.
- LARA, M. L. G. de. *A representação documentária: em jogo a significação*. 1993. Dissertação (Mestrado) - ECA-USP, São Paulo.
- LARA, M. L. G. de. *Representação e linguagens documentárias: bases teórico-metodológicas*. 1999. Tese (Doutorado) - ECA-USP, São Paulo.
- LE GUERN, M. Sur les relations entre terminologie et lexique. *Meta*, v.34, n.3, p.340-343, 1989.
- LARA, M.L.G. de. O unicórnio (o rinoceronte, o ornitorrinco ...), a análise documentária e a linguagem documentária. *DataGramaZero: revista de ciência da informação*, v.2, n.6, dez. 2001. Disponível em (www.dgz.org.br)
- LOPES, E. *Fundamentos da lingüística contemporânea*. São Paulo: Cultrix, 1987.
- NUNES, B. (1998). Nós somos um diálogo. *Folha de S. Paulo*, São Paulo, 8 ago. 1998. Cad. Especial, Jornal de Resenhas, p.10 (Caderno Discurso Editorial/USP/Unesp).
- RORTY, Richard. *A filosofia e o espelho da natureza*. Lisboa: Dom Quixote, 1988.
- SMIT, J. W. A perspectiva interdisciplinar da informação no contexto da Ciência da Informação. In: CONGRESSO BRASILEIRO DE ARQUIVOLOGIA, 12., 1998, João Pessoa. *Anais...* João Pessoa, 1998.
- TÁLAMO, M.F.G.M. *Linguagem documentária*. São Paulo: APB, 1997. (Ensaios APB, n.45).
- WELLISCH, H. H. Thesauri. In: _____. *Indexing from A to Z*. New York: The H. W. Wilson, 1991.

Perspect. cienc. inf., Belo Horizonte, v. 7, n. 2, p. 127-139, jul./dez. 2002