

Lingüística textual e elaboração de informações documentárias:

Algumas reflexões¹

Nair Yumiko Kobashi²

1 Introdução

Um sistema de recuperação de informação, de fato uma forma de expressão da memória coletiva, é construído para responder, de maneira duradoura, a necessidades específicas de informação. A seleção dos objetos informacionais para integrar tais dispositivos deve ser sistemática, fundamentada simultaneamente no conhecimento dos organismos que os produzem e nos interesses dos usuários. O tratamento, por sua vez, deve ser regido por parâmetros claros, sistematizados em uma política explícita e em procedimentos metódicos. A difusão, finalmente, deve pautar-se por um princípio básico: os produtos e serviços devem ser planejados e executados de modo a responder às especificidades da demanda. Assim, cada etapa da cadeia documentária é refratária à improvisação.

Infere-se, a partir daí, que o tratamento documentário adequado dos objetos informacionais, aqui considerados de forma ampla com quaisquer tipos de textos, é fundamental para constituir sistemas de armazenamento e recuperação de informação. Portanto, entre a produção de informação documentária e sua recuperação existe uma relação de pressuposição não recíproca: a recuperação não é factível sem uma ação anterior de tratamento.

No tratamento da informação deve-se estabelecer distinção entre dois aspectos inerentes aos objetos informacionais: seu suporte material e seu

¹ Este texto é uma síntese da tese: *A elaboração de informações documentárias, em busca de uma metodologia*, defendida em 1994. Algumas idéias novas, surgidas após a defesa, foram aqui incorporadas.

² Professora doutora livre-docente no curso de Biblioteconomia e Documentação da Universidade de São Paulo (USP/SP). E-mail: <nykobash@usp.br>.

conteúdo. Cada aspecto supõe processos de tratamento diferenciados. Será realizada, a seguir, uma breve caracterização dos dois aspectos do tratamento, enfatizando que este texto limita-se à abordagem do tratamento de conteúdo.

O tratamento do suporte material – a catalogação – visa descrever, de forma normalizada, os aspectos físicos do documento (nome da obra e do autor, local e ano de publicação, editora, entre outros). No contexto da difusão eletrônica de informações, denominam-se metadados as estruturas de organização desses tipos de dados.

O tratamento do conteúdo, por sua vez, visa elaborar representações condensadas daquilo que é dito em um dado texto. As representações documentárias típicas são o resumo e o índice. A primeira – o resumo – responde pela função de condensação da informação, individualizando os textos, e a segunda – a indexação – generaliza o conteúdo, de modo a integrar os textos em classes de assuntos.

2 Análise documentária: teoria e metodologia de produção de informação documentária

A análise documentária é, no interior da Ciência da Informação, a área que responde pelo tratamento do conteúdo de textos; é o “conjunto de procedimentos utilizados para exprimir o conteúdo dos documentos científicos sob formas destinadas a facilitar a sua localização ou consulta” (GARDIN, 1987, p. 48-49). Essas operações não são neutras porque as coerções imanentes aos sistemas documentários afetam cada operação realizada, desde aquelas relacionadas à seleção de documentos até as que se referem aos produtos a serem gerados (KOBASHI, 1994).

Desse modo, a análise documentária não se reduz à proposição de um conjunto de regras perenes, utilizáveis em todas as circunstâncias. Para avançar teoricamente e praticamente, ela critica de forma permanente seus pressupostos, procedimentos e instrumentos; ao mesmo tempo, com base na reflexão e no trabalho empírico, elabora novas hipóteses de trabalho que possam contribuir para aperfeiçoar processos e sistemas. A análise documentária apresenta, portanto, três aspectos indissociáveis, pertencentes a diferentes esferas:

- a esfera teórica, na qual se determina o próprio objeto da análise documentária, suas funções, seus métodos e procedimentos metodológicos;
- a esfera da produção, em que são engendradas as regras de geração das diversas modalidades de informação documentária;
- a esfera pragmática, na qual se identificam as condições de aderência entre usuários e sistema documentário.

Do ponto de vista metodológico, a análise documentária define-se como uma operação com textos. Primeiramente, eles são desestruturados, isto é, analisados e sintetizados, operação fundada na distinção entre informação essencial e acessória. A seleção é uma tarefa complexa, na medida em que implica atribuir valor às informações do texto. Em seguida, as informações selecionadas são estruturadas, isto é, submetidas a processos de combinação, de modo a convertê-los em novos textos, ditos resumos, ou em símbolos de uma linguagem específica, dita linguagem documentária. Em ambos os casos, os produtos são obtidos por meio de operações sucessivas de análise e síntese, devendo os produtos resultantes manter, necessariamente, relações de semelhança e contigüidade com o objeto submetido à análise.

No caso do resumo, tais relações poderão ser garantidas caso sejam mantidos os valores referenciais e a estrutura do texto original (LE ROUX, 1990). Uma cadeia de termos de indexação é, por outro lado, uma representação na qual subsiste apenas o tópico (assunto) do texto, representado por meio da palavra, de um sintagma ou uma unidade polilexical mais complexa. Nesse caso, "o produto da análise é uma representação do texto formulada em uma linguagem que não se confunde com a linguagem do texto, mesmo que os termos tenham aparentemente a mesma forma" (GARDIN, 1987, p. 49).

A informação documentária, assim produzida, tem, por definição, função de relevância. Ela deve, de um lado, promover a identificação de itens informacionais que respondam de modo pertinente a determinada questão do usuário e, de outro, embasar a tomada de decisão sobre a consulta ou não do documento original. Para cumprir adequadamente tais funções, a informação documentária deve ser elaborada por meio de metodologias que garantam a equivalência de sentido entre o texto-fonte e sua representação. É esse o objetivo central da análise documentária, que tem procurado responder às tarefas conceituais, metodológicas e pragmáticas, aqui enunciadas, pela apropriação e adaptação de conceitos e métodos das áreas que, como ela, lidam com a linguagem e os textos.

2.1 Caracterização do objeto e das operações documentárias

A elaboração de informações documentárias compreende um conjunto de operações esquematizáveis em três fases: a) ler um texto; b) selecionar dele o conteúdo informacional considerado pertinente para usos determinados; e, finalmente, c) representá-lo, ou seja, dar forma documentária às informações selecionadas.

No senso comum, a elaboração de resumos é vista como mero ato técnico de reduzir textos; a indexação, como o ato de estabelecer equivalências entre

palavras do texto e palavras de uma linguagem documentária. No entanto, para serem eficazes, essas operações requerem uma visão clara dos parâmetros específicos no interior dos quais se instaura a situação comunicativa, sendo os mais importantes os objetivos institucionais, os tipos de usuários e suas demandas informacionais.

Os parâmetros situacionais são a base sobre a qual são formuladas as políticas que orientam o conjunto de operações a serem realizadas. O ato documentário requer, por fim, metodologias que permitam construir os objetos que darão materialidade à comunicação entre usuário e sistema de informação. Portanto, uma análise mais acurada dos processos documentários indicará que suas operações centrais – ler, selecionar e representar – apresentam problemas cuja complexidade não poderá ser explicada segundo visões ingênuas.

A análise documentária, em sua dimensão operatória, manipula e transforma textos em dois tipos básicos de representações: o resumo e o índice. No resumo, o texto-base é desestruturado de modo a permitir que certos fragmentos, organizados sob a forma de um novo texto, reconstruam potencialmente o sentido do original. Na indexação, procura-se obter um grau ainda maior de compactação do texto-base: do processo de análise resultam palavras ou sintagmas indicadores do conteúdo.

Se o resumo mantém, de fato, relação de contigüidade e semelhança com o texto que lhe dá origem, porque procura reter a estrutura informacional do original, o mesmo não pode ser afirmado em relação à indexação. Neste último caso, instaura-se uma nova ordem de sentido, mediada por um instrumento comutador – a linguagem documentária –, que generaliza as informações do texto, integrando-as às “classes” previstas no referido instrumento. Nessa medida, as representações documentárias mediadas por linguagens documentárias “são de caráter generalizante”, porque a indexação não particulariza as informações do texto (LARA, 1993).

Já se afirmou que a palavra é neutra porque ela assume todos e nenhum significado. Ao discutir as palavras e seu significado, Ullmann (1967, p. 133) observa que:

o caráter *genérico* de nossas palavras é uma das principais fontes de sua imprecisão. Com exceção dos nomes próprios e de um pequeno número de nomes comuns que se referem a objetos únicos, as palavras denotam, não entidades singulares, mas classes de coisas ou de acontecimentos ligados por algum elemento comum.

A natureza genérica das palavras é um fenômeno geral da linguagem, “um elemento de abstração” (ULLMANN, 1967, p. 134). Desse modo, as unidades da linguagem documentária, embora sejam fixadas semanticamente, pelo fato

de designarem “classes”, e não informações particulares, tendem a veicular informação com certo grau de imprecisão, à semelhança das palavras da linguagem natural. Por essa razão, no âmbito pragmático da recuperação da informação, que opera com a palavra, a indexação e o resumo atuam de forma complementar. A primeira etapa de busca de informação é feita por meio dos termos de uma linguagem documentária. Em geral, obtêm-se respostas com um número significativo de documentos. Dessa forma, a busca requer operações sucessivas de filtragem, até que se obtenha um conjunto manipulável de documentos. Para verificar a pertinência dos itens informacionais recuperados, é útil recorrer aos resumos.

O processo global de análise documentária defronta-se, portanto, com dois objetos, ao menos: a) o texto, como sua unidade de análise; e b) a linguagem documentária, como instrumento comutador. Sobre isso, algumas questões: Como abordar tais objetos? Qual é a teoria necessária para discutir criticamente a análise documentária? Quais são os procedimentos necessários para analisar os objetos informacionais? Que tipos de técnicas poderiam otimizar a fabricação de representações de objetos informacionais?

Granger (1989) propõe a conduta global a ser adotada ao afirmar que a radical descontinuidade dos paradigmas é um mito, porque um modelo sempre se estabelece “a partir de um momento anterior do pensamento”, pela “ressonância do passado no presente” (GRANGER, op. cit., p. 25). Para o autor, “mesmo quando se pretende fazer tábula rasa dos sistemas que o precederam, este não pode instituir-se sem levá-lo em conta, mais ou menos explicitamente e de modo mais ou menos rigoroso” (*Idem*, p. 25). Granger indica, igualmente, a necessidade de delimitar os fatos específicos que se deseja observar ao afirmar que:

Não se poderia, com certeza, exigir das ciências uma definição universal do fato: o conhecimento científico, por sua natureza, repousa numa determinação específica e, por assim dizer, regional do fato e, em constante evolução, porque a cada momento de sua história, cada ramo do pensamento científico delimita a classe de fatos que se quer explicar com os meios materiais e conceituais dos quais se pode dispor (GRANGER, op. cit., p. 13).

Diante dos múltiplos marcos teóricos e metodológicos a serem examinados, duas tentações precisam ser evitadas: a) amalgamar perspectivas irreconciliáveis entre si; b) estabelecer equivalências indevidas entre conceitos ou mesmo assimilá-los de forma precipitada ao campo específico de trabalho. Assim, serão definidos primeiramente alguns termos para garantir a clareza das questões a serem abordadas. O primeiro deles é o conceito de representação de informação.

2.2 A representação da informação

A palavra representação evoca inúmeras noções. Para evitar equívocos, já que ela é utilizada em contextos variados, será determinado, neste trabalho, seu sentido específico.

Na Ciência da Informação, o termo “representação” é um conceito pré-teórico, associado, de um lado, à descrição dos aspectos que identificam materialmente os objetos informacionais (catalogação) e, de outro, ao processo e ao produto da condensação de conteúdos de textos, ou seja, à indexação e à elaboração de resumos (processo) e aos próprios índices e resumos (produtos).

Ao descrever os aspectos materiais dos objetos, a identificação ocorre por apreensão instantânea. Na análise documentária, a representação resulta de procedimentos inferenciais mais complexos. Gardin, por exemplo, assimila a operação de construir essas representações aos processos realizados no interior das ciências que analisam e interpretam textos:

A passagem do texto original para esse gênero de “representação” – empregaremos doravante este termo para designar o produto da análise documentária – é sem dúvida uma operação semântica, mesmo que ela não obedeça, na maioria das vezes, a nenhuma espécie de regra precisa, e que cada organismo de documentação, e mesmo cada analista, se limite a buscar no documento a ocorrência de uma certa regularidade interna, fundada muito mais na experiência ou no hábito, do que em algum tipo de procedimento explícito (GARDIN, 1974, p. 81-82).

A informação documentária, ao ser abordada como algo resultante de operações de natureza semântica, permite que ela seja observada como uma forma de representação simbólica. Ela é, contudo, uma representação particular porque, sob suas diversas formas de manifestação, deve responder de modo inequívoco às exigências da recuperação da informação, mantendo, com o texto original, relação de similaridade. Embora a informação documentária seja obtida, de um lado, pela neutralização do poder expressivo do texto e, de outro, moldada de acordo com formas previamente determinadas, prevalece a idéia de algo que, apesar de ser formalmente diferente do original, portanto “representação”, é equivalente a ele, do ponto de vista do conteúdo informacional. Nesse sentido, a informação documentária é representação construída a partir de um objeto efetivamente presente, que o substitui para certas finalidades.

Aceitando a idéia de que o texto é uma forma de representação simbólica, deve-se aceitar igualmente que a análise documentária, ao operar com textos dos vários domínios da atividade humana, opera com objetos que trazem as

marcas das várias funções da consciência: de representação (mito), da função expressiva (linguagem) e da função significativa (ciência) (CASSIRER, 2001).

Além disso, se for aceito o princípio segundo o qual o homem não entra em contato direto com as coisas, mas se apropria da realidade por meio da mediação de estruturas conceptuais e lógicas, parece adequado afirmar que a geração dos produtos da análise documentária – as representações documentárias – não escapa, também, a essa mediação.

2.3 A significação nos processos documentários

O reconhecimento da natureza semântica da análise documentária aproximou-a das Ciências da Linguagem. Nesse diálogo, a análise documentária tem se apropriado dos conceitos e metodologias propostos pela Lingüística, pela Análise do Discurso, pela Lingüística Textual e, mais recentemente, pela Terminologia.

A semântica do código permite discutir as relações de significação entre as unidades constitutivas das linguagens documentárias; a semântica textual apresenta, por sua vez, conceitos fundamentais para realizar as apropriações parafrásticas que caracterizam as operações de condensação. As dificuldades da tarefa são previsíveis, pois se trata de aproximar ciências conceituais, como a Lingüística, de uma área na qual os modelos devem ser operacionalizáveis.

A linguagem documentária, código comutador – conhecido também por outras denominações: linguagem de indexação, sistema de classificação, linguagem de informação, listas de cabeçalhos de assuntos –, via de regra composto por um conjunto limitado de termos, estabelece interfaces entre as formas de entrada e de busca a serem utilizadas pelo indexador ou pelo usuário em um sistema de informação. As unidades constitutivas da linguagem documentária são selecionadas de diversas fontes: linguagens de especialidade, linguagem de uso corrente e terminologias de área.

Esse fato revela questões semânticas complexas para a estruturação das linguagens documentárias, porque elas têm como ponto de partida sistemas de significação em que as relações entre significado e significante são de natureza diversa. Na linguagem natural, por exemplo, o fenômeno da plurissignificação é a regra, enquanto, nas linguagens de especialidade, predomina a busca de estabilização semântica dos termos. As linguagens documentárias operam, por sua vez, em campos do conhecimento especializado para descrever e tornar recuperáveis os textos produzidos. Desse modo, a operação de construção de linguagens documentárias supõe, necessariamente, os instrumentos terminológicos como fontes de referência (TÁLAMO et al., 1992, 1992a, 1992b; CINTRA et al., 1993).

As terminologias não se confundem, portanto, com as linguagens documentárias, antes, instauram-se como o universo referencial destas últimas. Nesse quadro, as linguagens documentárias operam como classificações intermediárias³ construídas na intersecção de três paradigmas designacionais, ao menos: a linguagem de especialidade (núcleo de idéias); as classificações científicas e suas definições (terminologias de áreas); e a linguagem natural (TÁLAMO et al., 1992b; CINTRA et al., op. cit.).

2.4 Análise de textos: algumas abordagens

Ao definir a análise documentária como uma disciplina que se preocupa com o tratamento do conteúdo de textos, torna-se inevitável sua comparação com outros campos que também operam com o texto. Embora todas elas observem um mesmo objeto empírico, denominado discurso ou texto, as noções recobertas pelos termos utilizados variam de área para área, sendo igualmente diversos os objetivos, modos e instrumentos utilizados para analisá-lo, como será exposto a seguir.

Os estudos do discurso e do texto remontam aos anos 1950, período em que Harris (1962) realizava os trabalhos pioneiros da área. Seu desenvolvimento mais amplo, contudo, ocorreu a partir dos anos 1980 (VAN DIJK, 1992). Até então, as pesquisas lingüísticas observavam, de forma dominante, a palavra e a frase.

Aqueles que estudam o discurso reconhecem que sua abordagem não se esgota dentro de uma única disciplina. Textos e discursos são objetos pluridisciplinares cuja apreensão impõe a formulação de um quadro teórico para o qual concorrem o lingüístico, o literário e o filosófico (BAKHTIN, 1992), além do sociológico (PÊCHEUX, 1969) ou mesmo o psicológico (KINTSCH & VAN DIJK, 1978).

Na Análise do Discurso de Pêcheux, os discursos, por serem articulados no interior de condições de produção determinadas, são objetos que exibem certos traços estáveis de manifestação da ideologia. Para o autor, a Análise do Discurso é, de fato, a teoria que procura “apreender como no lingüístico inscrevem-se as condições sócio-históricas de produção” (BRANDÃO, 1991, p. 83). Dito de outro modo, o modelo proposto por Pêcheux inscreve-se no conjunto de trabalhos interdisciplinares sobre o discurso e as relações de confronto entre as classes sociais. Essa perspectiva revela uma característica central da Análise do Discurso: operar em dois níveis, de modo a:

3 Dentro de uma perspectiva semelhante, Gardin discute as relações entre a linguagem natural, a linguagem especializada e a linguagem documentária, caracterizando esta última como uma linguagem intermediária (GARDIN, 1987, p. 52).

Integrar, por meio da enunciação, uma abordagem interna do texto, indispensável para que se reconheçam os mecanismos e regras de engendramento do discurso, com a análise externa do contexto sócio-histórico, em que o texto se insere e de que, em última instância, cobra sentido (BARROS, 1988, p. 5).

Também Adam (1989), embora sob uma ótica distinta, afirma não ser possível analisar o discurso de um ponto de vista puramente lingüístico:

A lingüística e a pragmática textuais devem, portanto, definir um campo de pesquisa específico dentro do domínio mais vasto do discurso porque outras disciplinas (como a história, a sociologia, a psico-sociologia, a psicanálise, etc.) têm, provavelmente, mais condições para descrevê-lo (ADAM, 1989, p. 191).

Tem centralidade, nos modelos citados, a idéia de que o discurso deve ser abordado em sua dimensão social. No entanto, Adam (op. cit.) distingue discurso de texto: este último não é um objeto empírico; ao contrário, é um objeto teórico depurado dos traços que caracterizam o discurso: "O texto é um objeto abstrato, resultante da subtração do contexto sobre o objeto empírico (Discurso)" (ADAM, op. cit., p. 191). Um modo esquemático de distinguir a Análise do Discurso da Teoria do Texto, ou Lingüística do Texto, pode partir da descrição das tarefas que a cada uma delas se atribui. No caso da Lingüística do Texto, elas podem ser resumidas em dois pontos: a) determinar o que faz um texto ser um texto; b) diferenciar as várias espécies de textos (FÁVERO & KOCH, 1988, p. 14). Já a Análise do Discurso, segundo Diana Barros (1988), deve considerar:

Três questões básicas para a concepção de discurso e sua análise: a) a relação do discurso com a enunciação e as condições de produção e de recepção; b) o discurso como o lugar, ao mesmo tempo do social e do individual; c) a articulação entre narrativa e discurso, isto é, o discurso constituído sobre estruturas narrativas que o sustentam (BARROS, 1988, p. 3).

Por mais sedutores que sejam as propostas da Análise do Discurso, deve-se admitir que as diversas tentativas de incorporar seus procedimentos à Análise Documentária, realizadas até o momento, não foram bem-sucedidas. Duas razões, ao menos, explicam os percalços:

- 1) Os dois níveis de análise (externo e interno), por serem destinados a desvendar a “relação do falante com o processo de produção de frases (enunciação) ou a relação do discurso com o grupo social a que ele se destina (sociolinguística)” (DUBOIS et al., 1973, p. 50-51), são de difícil operacionalização no âmbito da Análise Documentária, cujo objetivo fundamental é condensar textos e representá-los documentariamente, para que circulem como informação.
- 2) Em segundo lugar, é preciso considerar que a noção de informação relevante de um texto varia de acordo com os objetivos perseguidos. Os procedimentos para selecioná-la são regidos, necessariamente, por essa conceitualização inicial. Desse modo, seria contraditório supor que os mecanismos utilizados no âmbito da Análise do Discurso possam ser transferidos diretamente para a Análise Documentária.

Considerando a possibilidade de isolar do objeto empírico “discurso” o objeto teórico “texto”, tal como propõe Adam (1989), serão apresentados, a seguir, os conceitos essenciais relacionados ao objeto “texto”, discutindo, ao mesmo tempo, sua operacionalidade para os fins documentários.

O texto, em uma primeira aproximação, define-se por sua “autonomia e seu fechamento” (DUCROT & TODOROV, 1972, p. 375). O quadro conceitual, no interior do qual se discute o texto, pode ser, *grosso modo*, classificado em duas vertentes: de um lado, as de natureza estritamente linguística, centradas em fenômenos frásticos; e, de outro, a pragmática textual, que procura dar conta da dimensão global da textualidade, definindo o texto como um todo, e não como a reunião de proposições. É essa segunda vertente que será discutida agora.

Segundo Ducrot (1972, p. 87), todo texto apresenta simultaneamente:

- a) a condição de progressão, já que a ausência de informação nova implica paráfrases perpétuas;
- b) a condição de coesão e coerência, expressa pela reiteração e repetição de certos conteúdos que reaparecem regularmente ao longo do texto.

Em perspectiva semelhante, Adam (1989, p. 209) afirma que: “A textualidade pode ser definida como um equilíbrio delicado entre continuidade-repetição, de um lado, e progressão de informação, de outro”.

A coerência e a coesão, ao se referirem à permanência de categorias semânticas no texto, podem ser assimiladas à noção de isotopia,⁴ proposta

4 “O termo isotopia designa, segundo A. J. Greimas (1976), a propriedade característica de uma unidade semântica que permite apreender um discurso como um todo de significação” (DUBOIS et al., 1973, p. 355). A isotopia equivale a um plano homogêneo de significação que o leitor deve construir para compreender o texto como um todo (LOSANO et al., 1989, p. 30).

por Greimas. É importante observar que os conceitos de coesão, coerência e isotopia se articulam à idéia de sentido global do texto, ou tópico do discurso. Segundo Adam (1989, p. 194-195), “[...] o tópico ou tema do texto instaura-se como uma relação não linear de coesão-coerência”. Ou, ainda, segundo a formulação de Eco (1979, p. 97), “[...] a isotopia pode ser vista como um fenômeno de coesão semântico-referencial, enquanto o tópico do discurso é um fenômeno pragmático relacionado com a coerência e a pertinência”. Essas noções caracterizam o texto como uma unidade de sentido que apresenta redundâncias, reiteraões e repetições. Infere-se, daí, a possibilidade de condensar o texto, suprimindo as redundâncias, sem que se descaracterize sua informação central.

O conceito de tipologia textual está intimamente associado à idéia de que todo texto apresenta uma sintaxe que organiza suas várias partes. É essa forma global de organização do texto que Van Dijk & Kintsch (1983) chamam superestrutura, a qual é um elemento considerado fundamental para a apreensão do significado do texto, porque o domínio desses esquemas globais permite ao leitor monitorar sua leitura, de modo a integrar as várias informações àquilo que é central.

A superestrutura designa uma estrutura convencional. Enquanto paradigma de organização textual, o esquema, ou superestrutura, fornece a base para a interpretação do texto. No modelo em discussão, a compreensão é descrita como um processo que supõe a transferência das unidades semânticas identificadas no texto para um esqueleto conceitual derivado do esquema de conhecimento. O texto-base resultante é, em suma, o produto do casamento entre conhecimento do esquema (superestrutura) e informação textual (VAN DIJK & KINTSCH, 1983, p. 48).

Para os mesmos autores, a compreensão do discurso é de natureza estratégica. As estratégias são, por sua vez, parte do conhecimento geral e, como tal, aprendidas e reaprendidas, antes de serem automatizadas. Nessa perspectiva, novos tipos de discursos e formas de conhecimento requerem o desenvolvimento de novas estratégias (VAN DIJK & KINTSCH, 1983, p. 11). Desse modo, o bom leitor é capaz de explorar de forma mais direta o texto, não se preocupando com a decodificação *bottom-up*. Em suma, bons leitores utilizam melhor os processos *top-down*, nos quais a superestrutura assume papel preponderante (*Idem*, p. 23).

O modelo de Van Dijk permite hipotetizar um esquema de análise documental fundamentado no reconhecimento da superestrutura textual, a qual poderá ser o elemento capaz de imprimir agilidade e consistência à coleta e à organização de informações para a elaboração de resumos e índices. Afirma-se, portanto, que a elaboração de resumos e de índices requer paradigmas de coleta de dados motivados, de um lado, pelo que são as representações documentárias e, de outro, pelo que são o texto e sua estrutura.

3 Condensação de textos

A condensação de textos, sem a presença de um esquema que parametrize a coleta de dados, resulta, via de regra, em produtos que não exibem as propriedades necessárias para as finalidades da recuperação da informação. A ausência de regras e procedimentos metódicos pode causar seleção incompleta ou equivocada de dados, motivada pela utilização de estratégias de leitura inadequadas para os objetivos documentários. Os resumos e índices elaborados sem rigor metodológico são, dessa forma, “o produto de um impulso, não de uma intenção” (BARTHES, 1971, p. 20).

A metodologia aqui proposta pretende ser uma ferramenta de trabalho apta a promover a produção de informações documentárias pertinentes, por contemplar, simultaneamente, a coleta de dados textuais, sua organização e sua combinação por meio de procedimentos objetivados e sistemáticos.

3.1 Indexação e elaboração de resumos: semelhanças e diferenças operatórias

A indexação e a elaboração de resumos, como visto, são operações que compreendem: a) análise e compreensão do texto; b) seleção das informações principais; e c) representação das informações segundo padrões estabelecidos por políticas de informação.

Os índices e os resumos são representações que, embora possam ser reunidas sob a denominação genérica de informação documentária, apresentam características distintas: a) o **resumo** representa o original sob a forma de um novo **texto condensado**; e b) a **indexação** caracteriza o **tema** do texto por meio de palavras ou sintagmas de uma linguagem documentária.

A linguagem documentária, por ser um instrumento constituído por léxico reduzido e sintaxe precária, não permite a produção de mensagens com a mesma riqueza informacional dos resumos. Desse modo, para a indexação, interessa analisar o texto para determinar sua macroestrutura,⁵ que será, em seguida, representada por palavras ou expressões sintéticas, os descritores.

No caso do resumo, a compreensão e a seleção assumem caráter mais complexo em relação à indexação, ao menos por duas razões:

- 1) a mensagem veiculada pelo resumo é, em princípio, mais completa do que a mensagem veiculada pelo descritor; enquanto o resumo mantém

⁵ A macroestrutura é definida como uma rede de proposições associadas hierarquicamente (VAN DIJK & KINTSCH, 1983).

- relação de semelhança (no plano da mensagem) e contigüidade (no plano da organização textual) com o texto de partida, o índice opera basicamente no eixo da semelhança (mensagem);
- 2) a elaboração de resumos implica responder aos requisitos básicos da textualidade por meio de operações comprometidas com a construção formal de uma mensagem dotada de coesão e coerência, enquanto o índice implica elaborar representações por meio de descritores justapostos.

Em síntese, tanto a elaboração de resumos quanto a indexação requerem a compreensão global de um texto. O grau de redução a que se deve submetê-los para obter um ou outro tipo de representação documentária sugere que, no caso da indexação, pode-se operar com a noção de tema⁶ e de estrutura temática; no caso do resumo, a noção de superestrutura poderá nortear a operação de seleção de informações textuais.

3.2 Tema e estrutura temática

A análise e a compreensão do texto, para fins documentários, requerem estratégias adequadas de leitura. A leitura documentária é orientada para uma missão. Não é leitura de fruição, sendo, antes, atividade comprometida com a intenção precisa de comunicação: elaborar produtos que permitam recuperar informação. Os produtos documentários destinam-se ao menos a: auxiliar a tomada de decisões sobre a leitura ou não do documento original; dar ao leitor uma quantidade suficiente de informações, a fim de que ele, em certas circunstâncias, não tenha necessidade de ler integralmente o original. Nessa perspectiva, a leitura documentária é, fundamentalmente, uma operação de hierarquização e seleção das informações presentes no texto. A seleção das informações é factível, apenas, com base na distinção entre informação essencial e informação acessória.

Na leitura documentária, a compreensão global de um texto requer a identificação de seu tema. Os temas são as idéias principais discutidas ou tratadas em um documento. Definem-se, portanto, como o foco principal de um trabalho, e não como um aspecto periférico. A abordagem de um tema implica, normalmente, a referência a conhecimentos anteriores ou requer a introdução de outros conceitos; algumas idéias podem, por outro lado, ser usadas para ilustrar certos fatos. Tais itens são dados que cumprem no texto várias funções, tais como auxiliar a compreender o texto ou dar credibilidade a

⁶ "O tema (ou tópico) de um ato de enunciação é aquilo sobre o qual o locutor fala, é o objeto do discurso [...]" (DUCROT & TODOROV, 1972, p. 345).

uma afirmação, mas não são temas centrais. O “tema” é o elemento em torno do qual se estrutura a mensagem, é seu núcleo informativo.

A identificação do tema central resulta de um processo de condensação semântica. García Gutiérrez & Lucas (1987, p. 47-49) propõem a utilização das perguntas conceituais de Lasswell (Who, What, When, Where, Why) como método de abordagem analítica do texto de atualidade, para fins de indexação. Lasswell (1971) observou que um ato de comunicação pode ser descrito por meio de respostas fornecidas a cinco perguntas, a saber: quem, diz o quê, em que canal, para quem, com que efeito? Essa formulação de Lasswell, expressa pelos “5 W”, determina a estruturação da informação jornalística.

Considerando, portanto, a indexação como uma operação que implica: a) análise e compreensão do texto; b) seleção das informações principais; e c) representação das informações na linguagem utilizada no sistema, as duas primeiras operações (análise e seleção de informações) poderão ser realizadas por meio dos mecanismos de resposta às categorias da estrutura temática.

3.3 Superestrutura e classificação de textos

Há diversas tentativas de classificar os textos segundo tipologias. Os textos podem ser classificados, entre outros enfoques, pela estrutura interna (descritivo, narrativo, dissertativo) ou pela finalidade a que se propõem (técnico, científico, didático, jornalístico, jurídico, político, etc.).

Os primeiros estudos sistemáticos sobre os “gêneros” textuais e as “partes do discurso” são atribuídos a Aristóteles. Suas idéias, apresentadas no âmbito da retórica, fundamentam ainda hoje as modernas teorias discursivas relacionadas à persuasão (CITELLI, 1989; OKASABE, 1979).

A retórica, para Aristóteles, não se reduz a um conjunto de normas ou à oratória, mas ao esforço de “descobrir especulativamente, em qualquer dado, o persuasivo” (ARISTÓTELES *apud* OKASABE, 1979, p. 142). É nessa perspectiva que Aristóteles afirma: “Não há senão duas partes no discurso, pois é necessário dizer qual é o tema e demonstrá-lo [...] Dessas duas partes, uma é a proposição; a outra, a confirmação. Como se tratasse, de um lado, do problema e, do outro, da demonstração” (*Idem*, p. 156).

Os textos argumentativos não escapam a essa caracterização. Contudo, a esse esquema genérico – problema/demonstração – podem ser agregados outros constituintes, de modo a caracterizar variantes de um tipo.

O texto técnico é, via de regra, de natureza argumentativa, elaborado com o intuito de expor metodicamente os resultados da observação de um fenômeno. Para Gardin (1987, p. 4):

a expressão concreta de um raciocínio científico é o texto científico, no qual o autor expõe as operações do espírito que o conduziram da observação de certos fatos empíricos ao enunciado de proposições denominadas de forma diversa: teses, hipóteses, interpretações, comentários, conclusões, explicações, etc....

É no interior da atividade de conhecer, portanto, que se define o texto científico canônico: uma unidade de comunicação do saber, dotada de certos elementos estruturais. No âmbito das pesquisas experimentais, esses elementos estruturais são o tema, o problema, a hipótese, a metodologia, os resultados e a conclusão (ASTI-VERA, 1979; BUNGE, 1973).

Texto Tipo 1

| | |
|-------------|------------------|
| Problema | Indagação |
| Hipótese | Conjetura |
| Metodologia | Observação |
| Resultado | Interpretação |
| Conclusão | Comentário final |

O texto argumentativo pode apresentar outros constituintes: a tese, que se caracteriza como apresentação de um ponto de vista; os argumentos, que são as provas apresentadas para qualificar positivamente o ponto de vista e, simultaneamente, desqualificar um ponto de vista contrário; a conclusão, por fim, que confirma o ponto de vista inicialmente apresentado. O quadro apresentado a seguir sintetiza a superestrutura desse tipo de texto:

Texto Tipo 2

| | |
|------------|----------------|
| Tese | Ponto de vista |
| Argumentos | Provas |
| Conclusão | Confirmação |

O trabalho pioneiro sobre o texto expositivo, segundo Ducrot & Todorov (1972, p. 36), foi feito por Becker, em 1966. Este último caracteriza o texto expositivo como aquele que apresenta as categorias problema-solução como seus constituintes básicos. Na realidade, o texto expositivo (tipo 3, neste trabalho), amplamente utilizado na comunicação do conhecimento técnico-científico, pode ser visto como uma estrutura constituída de três categorias: a) apresentação do problema; b) causas e conseqüências do problema; c) solução do problema, que pode ser esquematizado como a seguir:

Texto Tipo 3

| | |
|----------|----------|
| Problema | Questão |
| Causas | Razões |
| Solução | Resposta |

Neste trabalho, a proposta não é inventariar exaustivamente as variantes textuais. Apresenta-se uma hipótese de trabalho – a possibilidade de utilizar a superestrutura textual como paradigma para a coleta de dados para fins documentários. Com esse objetivo, foram apresentadas, anteriormente, três variantes do texto técnico-científico e, a seguir, a estrutura do texto narrativo (Tipo 4) e do descritivo (Tipo 5), classificação fundamentada na estruturação interna.

Texto Tipo 4

| | |
|---------|------------------|
| Quem | Actantes/Atores |
| O quê | Acontecimento |
| Quando | Aspecto temporal |
| Onde | Aspecto espacial |
| Como | Modo |
| Por quê | Razão |

Texto Tipo 5

| | |
|-----------------|----------------------|
| Referente | Objeto do mundo |
| Características | Predicados/Atributos |

4 Condensação para indexação e resumos

As atividades documentárias são processos nos quais se identificam objetivos e operações. Os objetivos referem-se à natureza dos processos, enquanto as operações, aos procedimentos realizados. Nessa perspectiva, a indexação tem como **objetivo** caracterizar o conteúdo de um documento por meio de uma linguagem documentária; suas **operações** básicas são: identificar o tema do texto; elaborar um enunciado temático e representar o enunciado temático. Para identificar o tema, lê-se o texto tendo como parâmetros as categorias da estrutura temática, expostas no item 3.2. Deve-se observar ainda que, na indexação, há uma operação final de conversão do enunciado temático para a linguagem documentária adotada no sistema, observando, igualmente, a política de indexação.

Tal como na operação de indexação, a elaboração de resumos supõe a seleção de informações textuais, realizada com base em sua hierarquização.

Segundo Van Dijk & Kintsch (1983), a superestrutura é um elemento fundamental para a compreensão dos textos porque: a) ele tem caráter convencional, sendo conhecido e reconhecido por uma comunidade lingüística; b) a superestrutura configura-se como um esquema abstrato que estabelece a ordem global de um texto e se compõe de uma série de categorias, cujas possibilidades de combinação fundamentam-se em regras convencionais. O domínio das superestruturas permite um processamento *top-down* coerente, porque a leitura é feita com base em hipóteses, e não pela construção do sentido por meio de processamento *bottom-up* (VAN DIJK, 1992).

Os passos da elaboração de resumos documentários podem, tal como na indexação, ser observados como princípios e operações. O objetivo é produzir representações parafrásticas condensadas de um texto. As operações básicas são: identificar o tipo de texto (superestrutura) e selecionar, nos constituintes, as informações consideradas relevantes para a elaboração de diferentes tipos de resumos – resumo indicativo ou resumo informativo –, propostas em normas de documentação. O resumo informativo supõe a condensação de cada constituinte da superestrutura e sua articulação, de modo a, em certas circunstâncias, substituir o texto de partida; o resumo indicativo, como o próprio nome sugere, é uma caracterização sumária do conteúdo. Para a elaboração de resumos, sugere-se a seleção de constituintes segundo o tipo de resumo pretendido.

| Superestrutura | Resumo informativo | Resumo indicativo |
|----------------|--------------------|-------------------|
| Problema | | |
| Hipótese | | |
| Metodologia | | |
| Resultados | | |
| Conclusões | | |

| Superestrutura | Resumo informativo | Resumo indicativo |
|----------------|--------------------|-------------------|
| Tese | | |
| Argumentos | | |
| Conclusões | | |

| Superestrutura | Resumo informativo | Resumo indicativo |
|----------------|--------------------|-------------------|
| Problema | | |
| Causas | | |
| Solução | | |

| Superestrutura | Resumo informativo | Resumo indicativo |
|-----------------|--------------------|-------------------|
| Objeto do mundo | | |
| Predicados | | |

| Superestrutura | Resumo informativo | Resumo indicativo |
|----------------|--------------------|-------------------|
| Quem | | |
| O quê | | |
| Quando | | |
| Onde | | |
| Como | | |
| Por quê | | |

5 Considerações finais

A **informação documentária**, para cumprir adequadamente suas funções, deve manter relações de contigüidade e semelhança com o texto-base. Essas relações são rompidas quando a seleção de dados é incompleta e equivocada. Dessa forma, sugere-se que a adoção de parâmetros que guiem metodicamente o trabalho documentário, em suas várias etapas, desde a seleção de dados até sua combinação final, possa ser feita com base nas contribuições da Linguística Textual e nas teorias do processamento cognitivo de textos. Afirma-se, portanto, que as atividades documentárias, em particular a indexação e a elaboração de resumos, podem ser monitoradas, deixando de lado a concepção de que a análise de textos é pura subjetividade, não se prestando à formulação de parâmetros. No entanto, como em todo jogo de linguagem, a produção de informações documentárias será mais efetiva se realizada observando regras.

A metodologia proposta procura integrar os conceitos de textualidade, superestrutura textual e estrutura temática, tornando operatórias as duas últimas para as finalidades da indexação e elaboração de resumos. Esses conceitos poderão, portanto, objetivar a identificação e hierarquizar o conteúdo informacional de textos.

A utilização da superestrutura e da estrutura temática nos processos documentários apóia-se no pressuposto de que o texto não se oferece à compreensão de forma direta e imediata. Desse modo, é necessário dispor de conceitos operatórios que facilitem a apreensão do texto e a identificação e seleção de informação segundo o produto a ser elaborado. A noção de superestrutura, no entanto, não deve ser utilizada de forma ingênua ou dogmática, já que os textos concretos não são puros. É possível, porém, operar com a noção de dominância para classificá-los no interior de tipos. A complexidade do real não pode paralisar e impedir a busca de modelos para compreender, cada vez melhor, os fenômenos e objetos do mundo.

Considerando que a forma de exposição do conhecimento tende a seguir padrões estabelecidos por cada comunidade de discurso, sugere-se que, em cada área, procure-se identificar os tipos de textos produzidos para aprimorar as operações documentárias.

A tipologização de textos não é uma tarefa propriamente documentária; ela é uma atividade específica da Linguística do Texto. Cabe à análise documentária apropriar-se das classificações propostas, testando-as e adaptando-as às finalidades documentárias. As atividades documentárias são feitas para promover a identificação dos produtos da atividade intelectual, sendo a condensação uma operação que reduz textos para viabilizar seu armazenamento e circulação como informação. Desse modo, a indexação e os resumos jamais poderão substituir a riqueza dos textos. Deve-se, no entanto, procurar modos de garantir a identificação adequada do conteúdo informacional dos textos, objetivando os processos de elaboração dos produtos documentários.

O fazer documentário se realiza sob coerções pragmáticas. Essa compreensão implica avaliar sistematicamente os métodos disponíveis, propondo novas formas de trabalho que substituam aqueles métodos que se tornam obsoletos ou são pouco rigorosos. A presente proposta de elaboração de informação documentária, inclusive, deve ser submetida a testes e críticas para ser validada. Ela certamente apresenta lacunas. São bem-vindas as críticas.

Referências bibliográficas

- ADAM, J. M. Pour une pragmatique linguistique et textuelle. In: REICHLER, C. *L'interprétation de textes*. Paris: Minuit, 1989. p. 183-219.
- ASTI-VERA, A. *Metodologia da pesquisa científica*. Porto Alegre: Globo, 1979.
- BAKHTIN, M. *Marxismo e filosofia da linguagem*. São Paulo: Hucitec, 1992.
- BARROS, D. L. P. *Teoria do discurso: fundamentos semióticos*. São Paulo: Atual, 1988.
- BARTHES, R. *O grau zero da escritura*. São Paulo: Cultrix, 1971.
- BRANDÃO, H. N. N. *Introdução à análise do discurso*. Campinas: Ed. UNICAMP, 1991.
- BUNGE, M. *La investigación científica*. Barcelona: Ariel, 1973.
- CASSIRER, E. *A filosofia das formas simbólicas*. São Paulo: Martins Fontes, 2001.
- CINTRA, A. M. M. et al. Do termo ao descritor: estudo exploratório. *Revista Comunicações e Artes*, São Paulo, 1993.
- CITELLI, A. O. *Linguagem e persuasão*. São Paulo: Ática, 1989.
- DUBOIS, J. et al. *Dicionário de linguística*. São Paulo: Cultrix, 1973.
- DUROT, O. *Dire et ne pas dire*. Paris: Hermann, 1972.
- DUROT, O.; TODOROV, T. *Dictionnaire encyclopédique des sciences du langage*. Paris: Seuil, 1972.
- ECO, U. *Conceito de texto*. São Paulo: T. A. Queirós, 1979.
- FÁVERO, L. L.; KOCH, I. G. V. *Linguística textual: introdução*. São Paulo: Cortez, 1988.
- FELBER, H. *Manuel de terminologie*. Paris: UNESCO/Inforterm, 1984.
- GARCÍA GUTIÉRREZ, A.; LUCAS, R. *Documentación automatizada de los medios informativos*. Madrid: Paraninfo, 1987.
- GARDIN, J. C. *Les analyses de discours*. Neuchâtel: Delachaux et Niestlé, 1974.

- GARDIN, J. C. L'analyse logiciste. In: GARDIN, J. C. et al. *Systèmes experts et sciences humaines*. Paris: Eyrolles, 1987. p.17-26.
- GRANGER, G. G. *Por um conhecimento filosófico*. São Paulo: Contexto, 1989.
- GREIMAS, A. J. *Semântica estrutural: pesquisa de método*. São Paulo: Cultrix, 1976.
- HARRIS, Z. S. *Discourse analysis reprints*. Haia: Mouton, 1962.
- KINTSCH, W.; VAN DIJK, T. Toward a model of text comprehension and production. *Psychological Review*, v. 85, n. 5, p. 363-394, 1978.
- KOBASHI, N. Y. *A elaboração de informações documentárias, em busca de uma metodologia*. 1994. Tese (Doutorado) – ECA-USP, São Paulo.
- LARA, M. L. G. *A representação documentária: em jogo a significação*. 1993. Dissertação (Mestrado) – ECA-USP, São Paulo.
- LASSWELL, H. D. A estrutura e a função da comunicação na sociedade. In: COHN, G. *Comunicação e indústria cultural*. São Paulo: Nacional/EDUSP, 1971.
- LE ROUX, D. Automatisation de l'activité resumante: essai de typologie. *Colloque sur le résumé de texte*. Pont à Mousson, 1990 (separata).
- LOSANO, J. et al. *Análisis del discurso: hacia una semiótica de la interacción textual*. Madrid: Cátedra, 1989.
- OKASABE, H. *Argumentação e discurso político*. São Paulo: Kairós, 1979.
- PÊCHEUX, M. *Analyse automatique du discours*. Paris: Dunod, 1969.
- TÁLAMO, M. F. G. M. et al. A contribuição da terminologia para a elaboração de tesau-
ros. *Ciência da Informação*, v. 21, n. 3, p. 197-200, 1992.
- TÁLAMO, M. F. G. M. et al. Instrumentos de controle terminológico: limites e funções. In: SIMPÓSIO LATINOAMERICANO DE TERMINOLOGIA, 2. Brasília: IBICT; Paris: União La-
tina. *Anais...* 1992a.
- TÁLAMO, M. F. G. M. et al. A interface análise documentária, lingüística documental
e terminologia. In: SIMPÓSIO IBEROAMERICANO DE TERMINOLOGIA, 3. *Anais...* Espanha,
1992b.
- ULLMANN, S. *Semântica*. Madrid: Aguillar, 1967.
- VAN DIJK, T. *Cognição, discurso e interação*. São Paulo: Contexto, 1992.
- VAN DIJK, T.; KINTSCH, W. *Strategies of discourse comprehension*. Orlando: Academic
Press, 1983.