

**Eduardo Wense Dias  
Madalena Martins Lopes Naves**

**Análise de Assunto  
Teoria e Prática**

**Estudos Avançados em Ciência da Informação - V. 3**



**Ficha Técnica**

Revisão: O autor  
Projeto gráfico: Cláudia Gomes  
Capa: Hugo Oliveira

Dados Internacionais de Catalogação na Publicação (CIP)

---

D541a Dias, Eduardo Wense  
Análise de assunto: teoria e prática / Eduardo Wense  
Dias e Madalena Martins Lopes Naves. — Brasília :  
Thesaurus, 2007  
p. 116 (Estudos Avançados em Ciência da Informação, 3)

1. Ciência da Informação 2. Biblioteconomia 3. Cabeçalho  
de Assunto 4. Indexação 1. Naves, Madalena Martins Lopes  
II. Título

CDU 025.43  
CDD 015

---

ISBN 978-857062-620-2



**Presidente Editorial**  
**Antonio Miranda - UnB**

**Secretária Executiva**  
**Iza Antunes Araújo**

**Comissão Editorial**  
**Kátia Carvalho - UFB**  
**Neusa Dias de Macedo - USP**  
**Elmira Simeão - UnB**

Todos os direitos em língua portuguesa, no Brasil, reservados de acordo com a lei. Nenhuma parte deste livro pode ser reproduzida ou transmitida de qualquer forma ou por qualquer meio, incluindo fotocópia, gravação ou informação computadorizada, sem permissão por escrito dos autores. Está é uma publicação THESAURUS EDITORA DL BRASÍLIA LTDA, SIG Quadra 8, lote 2356 — CEP 70610-480 — Brasília, DF fone (61) 3344-3738 — Fax: (61) 3344-2353 [www.thesaurus.com.br](http://www.thesaurus.com.br), e-mail: [editor@thesaurus.com.br](mailto:editor@thesaurus.com.br) — Contato com os autores: [edias@eci.ufmg.br](mailto:edias@eci.ufmg.br)

Composto impresso no Brasil  
*Printed in Brazil*

## Como referenciar os capítulos do livro?

DIAS, Eduardo Wense; NAVES, Madalena Lopes. Título do capítulo. In: \_\_\_\_\_. **Análise de assunto:** teoria e prática. Brasília: Thesaurus, 2007. Cap. Número do capítulo, p. Página inicial-Página final.

**OBS.: preencha todos os campos (basta dar um clique em cima de cada um) com os dados necessários e copie depois todo o modelo de referência acima e cole no local desejado.**

*Páginas iniciais e finais de cada capítulo no livro original impresso de onde se extraiu o texto*

Capitulo	Título	Página Inicial	Página Final
1	Tratamento temático da informação	13	26
2	O processo de tratamento temático	27	34
3	Leitura do texto pelo indexador	35	50
4	A prática da leitura técnica	51	62
5	Extração de conceitos	63	72
6	Determinação da atinência	73	78
7	O caráter interdisciplinar da análise de assunto	79	92
8	A pesquisa em análise de assunto	93	106
	Referências	107	114

## Sumário

Apresentação .....	5
1 Tratamento temático da informação .....	7
1.1 Os processos.....	10
1.2 Os instrumentos .....	12
1.3 Os produtos .....	13
2 O processo de tratamento temático .....	15
2.1 Consistência e relevância na indexação .....	17
3 A leitura do texto pelo indexador .....	20
3.1 O texto.....	20
3.2 A leitura.....	25
4 A prática da leitura técnica .....	29
5 Extração de conceitos.....	35
5.1 Conceito .....	35
5.2. Assunto .....	37
5.3 Contexto .....	39
6 Determinação da atinência.....	41
6.1 Atinência .....	41
6.2 Significado .....	42
7 O caráter interdisciplinar da análise de assunto .....	44
7.1 Aspectos lingüísticos .....	44
7.2 Aspectos cognitivos e lógicos .....	47
8 A pesquisa em análise de assunto .....	53
8.1 Os temas .....	53
8.2 Estratégias metodológicos .....	55
8.3 A pesquisa no Brasil .....	57
8.4 Tendências .....	59
Referências .....	61

## Apresentação

O tratamento da informação<sup>1</sup> tem sido considerado uma das funções de grande importância no desempenho de bibliotecas e sistemas de recuperação da informação. Pode ser dividido em dois grandes tipos, de acordo com a natureza da atividade que desenvolve: o tratamento descritivo, que lida com dados físicos extraídos do documento, e o tratamento temático, que procura descrever o conteúdo (o “assunto”) do documento.

A primeira etapa do tratamento temático é a análise de assunto, considerada por muitos como a etapa intelectual por excelência do trabalho do indexador, catalogador de assunto ou classificador. A análise de assunto é o processo de ler um documento para extrair conceitos que traduzam a essência de seu conteúdo. Essa tarefa está sujeita à interferência de diversos fatores ligados à pessoa daquele profissional, como nível de conhecimento prévio do assunto de que trata o documento, formação e experiência, subjetividade, além de fatores lingüísticos, cognitivos e lógicos.

Desde logo, deve-se alertar o leitor para os problemas de variação terminológica nessa área, inclusive em língua estrangeira. Neste livro, utiliza-se principalmente a expressão análise de assunto, mas, em língua portuguesa, vai-se encontrar ainda expressões como análise conceitual, análise temática, análise documentária e análise de informação, entre outras. Em inglês, há a expressão *subject analysis* que, em muitos casos é utilizada nesse sentido de análise de assunto que aqui empregamos; mas a expressão é muito utilizada também para significar *tratamento temático da informação*, ou seja, a área maior em que se insere a análise de assunto.

Outro alerta a ser feito é que há algumas outras atividades de tratamento temático da informação, e mesmo fora da área de tratamento, que às vezes são designadas de análise de assunto, ou mesmo entendidas como fazendo ou devendo fazer parte da especialidade. Típico do primeiro caso é o trabalho que se realiza no âmbito da construção de linguagens de indexação visando conhecer e definir a área de assunto que a linguagem vai cobrir. Fora da área de tratamento, é o caso da tarefa de analisar uma consulta feita por um usuário a um sistema de recuperação de informação. Essa análise visa definir a estratégia de busca a ser utilizada. Não se descarta o entendimento de que essas atividades têm grandes similaridades e afinidades com a tarefa de *análise de assunto*, como definida para este texto, mas não se vai nele discutir questões relativas a essas outras esferas.

É interessante a observação de que o termo análise de assunto apareça na literatura com diversos sentidos. Há uma confusão com relação ao real significado do termo, verificando-se que essa situação acarreta dificuldades generalizadas, tanto para os indexadores, quanto para os professores que ministram disciplinas na área, no momento de transmitir aos alunos conhecimentos sobre esse tema. Conseqüentemente, também o usuário pode ser vítima dessa situação, no momento de buscar a informação de que necessita.

Gardin (1973), de acordo com a abordagem francesa, utiliza o termo análise documentária, entendida, segundo ele, como a extração do significado de documentos. No caso em estudo, documentos escritos. A influência francesa é observada nos trabalhos de Kobashi (1994), para quem a elaboração de informações documentárias de natureza textual supõe a transformação de um objeto (documento) em outros objetos que possam representá-lo. Trata-se de uma

---

<sup>1</sup> Também conhecido por organização da informação e representação da informação.

modalidade específica de análise de textos, em que esses últimos são desestruturados, sintetizados e transformados em novos textos (resumos documentários e termos de indexação), com base na distinção entre informação essencial e acessória.

A análise documentária é definida como um conjunto de procedimentos efetuados com o fim de expressar o conteúdo de documentos sob formas destinadas a facilitar a recuperação da informação. Essa passagem de um texto original para um tipo de representação é uma operação semântica, mesmo que não obedeça a nenhuma regra precisa e variem função de cada organismo e do analista. Este seleciona as palavras-chave, normalmente de forma intuitiva, em função de sua ocorrência do seu interesse para a instituição.

Como se vê, no estudo do processo de análise de assunto, é necessário estar atento à confusão conceitual que gira em torno do tema, tendo-se verificado que tanto ocorre que diferentes termos são atribuídos ao mesmo processo, quanto o fato de que o mesmo termo seja dado a diferentes processos.

Muitos professores e estudiosos do tratamento da informação questionam a necessidade de se estudar o processo de análise de assunto, tendo em mente o desenvolvimento cada vez mais veloz das novas tecnologias da informação. Segundo eles, esse desenvolvimento tornaria desnecessário o estudo de um tema que estaria caminhando para a obsolescência. No entanto, sabe-se que o processo intelectual de definir o assunto de um documento pode ser considerado o limite da máquina, pois esta não pode abstrair, perceber, compreender e interpretar o que está num texto, sendo esse um trabalho que só pode ser desenvolvido por um ser humano. Acredita-se que à máquina, cabe a tarefa de indexar automaticamente grande quantidade de documentos, sem que se espere qualidade e precisão nos resultados obtidos e, conseqüentemente, no trabalho de recuperação das informações. Ao indexador humano caberia a indexação de coleções pequenas e especializadas, em que se pode adotar grande especificidade e esperar uma alta precisão na busca de informações.

O processo de análise de assunto, do ponto de vista do indexador é iniciado com a fase de leitura do texto. Para isso, é necessário que se conheçam tipos e estruturas de textos para iniciar-se a sua leitura com fins específicos. Após essa leitura, passa-se à fase da extração de conceitos que possam representar o conteúdo temático do texto, para se chegar ao momento da fase de representação da atinência, em que são definidos os termos em linguagem natural. Depois de traduzidos para uma linguagem de indexação, esses termos passam a ser chamados de descritores de assunto, cabeçalhos de assunto, palavras-chave, termos de indexação ou enunciados. Todas as fases do processo sofrem interferência de fatores lingüísticos, cognitivos e lógicos, o que confere ao processo de análise de assunto um caráter interdisciplinar.

Eduardo Wense Dias  
Madalena Martins Lopes Naves

## **1 Tratamento temático da informação**

Desde os primórdios da humanidade que o homem vem exercitando suas faculdades de agir sobre o meio ambiente e de transformá-lo. A partir de determinado momento de sua história, desenvolveu também a capacidade de elaborar registros documentais dessas e de outras ações, o que certamente foi motivado pelo reconhecimento da importância que esses registros poderiam ter, principalmente como meio de transmissão das informações neles contidas para outras pessoas, outros lugares e outras épocas. Desde então, a quantidade, a variedade e a complexidade de tais registros só faz aumentar, o que é estimulado não apenas pelo simples crescimento da população, mas também pelo desenvolvimento de novas tecnologias e métodos de elaboração e produção desses registros.

Esse aumento da quantidade de informação produzida no mundo e, por conseguinte, disponível para os interessados, tem se constituído num dos fatores principais de dificuldade para o trabalho de identificação e acesso a essas informações. Em muitas situações de necessidade de informação, o interessado estará às voltas com os problemas de saber o que existe de publicado que possa ser de seu interesse e, em seguida, escolher, no conjunto das informações identificadas, aquela ou aquelas que possam realmente atender às suas necessidades. Por isso, foi-se desenvolvendo também, paulatinamente, a necessidade de produzir informação sobre informação, ou seja, informações cujo objetivo é documentar os registros existentes nas várias áreas do conhecimento, suas características, e onde esses registros podem ser encontrados, entre outros detalhes.

Fala-se em identificar e acessar informações como tarefas cruciais, mas na verdade a identificação, que é uma tarefa por si só árdua e complexa, serve principalmente de etapa preliminar para o que, então, poder-se-ia caracterizar como a etapa-fim desse processo: o acesso efetivo à informação/documentos. Em outras palavras, para aqueles que precisam de informação o que interessa, em última análise, é ter acesso à informação desejada. Por isso, pode-se dizer que o acesso à informação/documentos é a questão básica de interesse da ciência da informação.

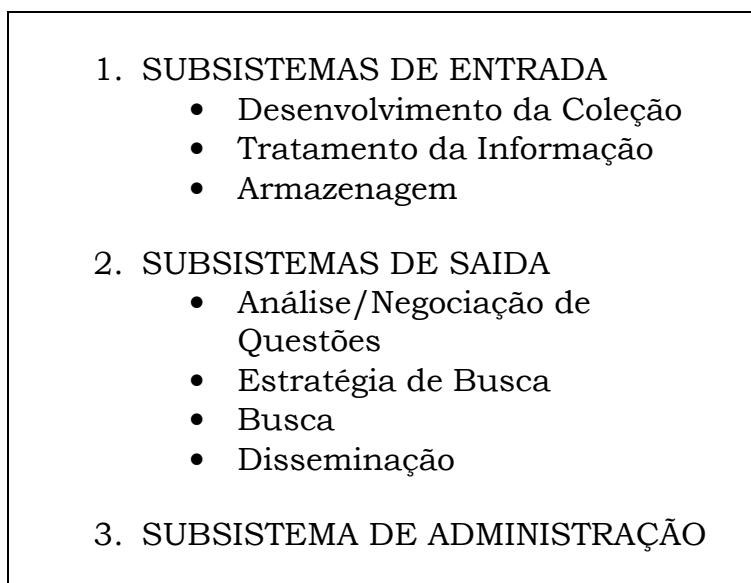
O acesso à informação tem vários aspectos — financeiro, físico, intelectual, social, entre outros — todos eles representando, eventualmente, barreiras para o usuário, dependendo das circunstâncias. Por isso, surgiu a necessidade de se criarem meios, instrumentos e instituições que pudessem facilitar esse acesso. Entre as instituições que têm se revelado mais eficazes nesse sentido destacam-se as bibliotecas e os sistemas de recuperação da informação. As bibliotecas, em suas várias conformações (nacionais, públicas, acadêmicas, especializadas), põem à disposição de seus usuários as informações de que necessitam. É um acesso físico com um aspecto econômico porque a grande maioria dos usuários de uma biblioteca não teria condições financeiras de adquirir e manter uma biblioteca semelhante às que são mantidas por um país, estado ou município; ou por uma empresa ou organização; ou por uma escola ou universidade. Não apenas custa caro adquirir os materiais que contêm informação, ou seja, livros, periódicos, vídeos, entre outros tipos de materiais usualmente encontrados em bibliotecas. Estas, para cumprir de forma condizente suas funções, precisam de várias outras coisas. Precisam, por exemplo, de profissionais capazes de

administrá-las e que precisam conhecer bem todos os seus recursos, para poder orientar os usuários e ter com estes um importante diálogo.

As características acima mencionadas dizem respeito não apenas à biblioteca tradicional<sup>2</sup>, mas continuam válidas para as bibliotecas eletrônicas (virtuais digitais etc.). A percepção é de que a necessidade dos elementos essenciais da biblioteca tradicional — a seleção das informações, seu tratamento, os diversos tipos de profissionais necessários — foi potencializada, na biblioteca virtual, talvez na mesma proporção em que as informações disponíveis cresceram nestas últimas (DIAS, 2001).

O sistema de recuperação da informação, por outro lado, visa otimizar o acesso ao conteúdo das informações, sejam estas as existentes numa determinada biblioteca, sejam quaisquer outras informações, inclusive aquelas que podem não estar em nenhuma biblioteca. Diz-se otimizar porque uma biblioteca já é organizada de forma a facilitar o acesso a seu conteúdo, mas o catálogo, que é o principal sistema de recuperação da informação de uma biblioteca, incrementa esse acesso.

Tanto as bibliotecas quanto os sistemas de recuperação da informação são constituídos de vários subsistemas, onde se destacam dois grandes grupos: subsistemas de entrada e subsistemas de saída. Uma síntese da variada terminologia utilizada na literatura poderia resumir esses subsistemas no seguinte esquema:



**Fig. 1 — Os subsistemas de uma biblioteca ou sistema de recuperação de informação**

O tratamento da informação é, por conseguinte, um dos subsistemas de entrada, encarregado de descrever aspectos físicos e de conteúdo dos documentos trazidos à biblioteca ou ao sistema de recuperação da informação pelo subsistema de desenvolvimento da coleção. Sua designação na prática e na literatura, entretanto, pode variar, sendo comum se referir a esse trabalho como:

**Catálogo e Classificação.** Esta foi e continua sendo a terminologia usual para descrever esse tipo de trabalho quando realizado nas bibliotecas tradicionais. A catalogação visa criar representações dos documentos, conhecidos

<sup>2</sup> Bibliotecas que possuem não apenas documentos como também vários espaços: para guardar todos ou alguns desses documentos; para receber e acomodar os leitores/usuários; e para desenvolver atividades e serviços que variam de acordo com o tipo de biblioteca, O conceito contrapõe-se ao de bibliotecas virtuais ou digitais, que existem apenas num ambiente eletrônico, como uma rede de computadores, por exemplo.



como fichas de catalogação, ou fichas catográficas, que descrevem tanto os aspectos físicos, objetivos do documento (autor, título etc.), quanto os aspectos de conteúdo, ou seja, o assunto ou assuntos de que trata. À atividade de descrição dos aspectos físicos costuma se dar o nome de catalogação descritiva, e à do conteúdo, catalogação por assunto.

A classificação visa também identificar o conteúdo dos documentos, mas com duas características muito próprias: a) o objetivo é o de determinar um lugar para o documento numa coleção organizada por assunto; e b) utiliza um sistema de classificação bibliográfica para representar os assuntos.

**Indexação.** É a terminologia mais usada para designar o trabalho de organização da informação quando realizado nos chamados serviços de indexação e resumo. Esses serviços têm por finalidade organizar informações referentes, principalmente, a artigos de periódicos. Essas informações não são, normalmente, organizadas nas bibliotecas. Estas costumam se limitar à organização dos periódicos considerados em seu todo. Os produtos principais daqueles serviços são os índices/abstracts, que tanto podem estar disponíveis em forma impressa como na forma de bases de dados.

**Metadados.** Este é o termo que vem sendo usado nas bibliotecas digitais para designar o trabalho de descrição física dos recursos eletrônicos. Desse ponto de vista, cumpre, no contexto digital, função semelhante à da catalogação nas bibliotecas tradicionais. Mas o conceito de metadados embute também um objetivo de normalização, necessidade que já era enfatizada no âmbito das publicações impressas há bastante tempo.

**Ontologias.** É o termo que tem sido usado nos contextos digitais para designar o trabalho de organização dos recursos eletrônicos com base em seus conteúdos, de forma a possibilitar a posterior recuperação dessas informações. Atividade, por conseguinte, com objetivos bastante semelhantes ao trabalho de classificação, indexação e catalogação por assunto.

Mas as atividades de tratamento da informação não se restringem ao trabalho desenvolvido nas bibliotecas e nos sistemas de recuperação da informação. Instrumentos importantes nesse trabalho, como os códigos de catalogação, os sistemas de classificação bibliográfica, as listas de cabeçalhos de assunto, os *thesauri* e as normas de documentação raramente são elaborados no próprio âmbito dessas instituições. Pelo menos não como parte das funções clássicas neles existentes. Ao contrário, são geralmente desenvolvidos e mantidos por organismos ou instituições muitas vezes criadas especialmente com essa finalidade.

Podemos, portanto, sintetizar o conceito de tratamento da informação da seguinte forma: expressão que engloba todas as disciplinas, técnicas, métodos e processos relativos a: a) descrição física e temática dos documentos numa biblioteca ou sistema de recuperação de informação; b) desenvolvimento de instrumentos (códigos, linguagens, normas, padrões) a serem utilizados nessas descrições; e c) concepção/implantação de estruturas físicas ou bases de dados destinadas ao armazenamento dos documentos e de seus simulacros (fichas, registros eletrônicos, etc.).

Compreende as disciplinas de classificação, catalogação, indexação, bem como especialidades delas derivadas, ou terminologias novas nelas aplicadas, tais como metadados, e ontologias, entre outras.

Essa caracterização do tratamento da informação foi levando, progressivamente, à identificação de duas grandes especialidades nessa área: tratamento descritivo e tratamento temático. O tratamento descritivo cuida

daqueles aspectos mais objetivos capazes de bem identificar, extrinsecamente, um documento: o autor, o título, a editora, e elementos similares. O tratamento temático, ao contrário, tem uma forte carga subjetiva pois, como o nome indica, visa caracterizar o documento do ponto de vista do seu conteúdo. É o que muitos vão chamar de assunto do documento, só que esse assunto vai depender muito de quem faz a leitura. No caso das bibliotecas/SRIs, essa leitura é feita por um profissional de informação, que tem sido designado de classificador, catalogador, indexador, além de outras eventuais denominações, em consonância com a diversidade terminológica notada acima. É esse profissional da informação que irá examinar o documento e tomar a decisão, muitas vezes difícil, de identificar o assunto ou assuntos de que trata o mesmo. O tratamento descritivo e o tratamento temático são também conhecidos como representação descritiva e representação temática, respectivamente.

Há uma tendência a se considerar as atividades de abordagem temática da informação (classificação, catalogação por assunto, indexação temática) como uma única atividade. Nesse sentido, essa variedade terminológica seria confusa e incongruente. É verdade que as semelhanças são muito grandes, de vez que em todas as três especialidades o que se faz, basicamente, é o mesmo, ou seja:

- Identificar o assunto ou assuntos de um documento; e
- Traduzir esse assunto(s) para uma linguagem de indexação.

Há, entretanto algumas diferenças que talvez possam justificar a variedade terminológica. Uma diferença básica entre classificar um documento para colocação nas estantes de uma biblioteca (ou, mais genericamente, para situá-lo numa coleção) e catalogá-lo ou indexá-lo por assunto é que, no primeiro caso, o classificador deve optar por um único lugar para o documento na coleção, o que não acontece nos outros casos, pois o catalogador de assunto e o indexador podem decidir que um determinado documento tenha várias entradas de assunto, num catálogo ou num índice. O catalogador ou indexador tem como limite apenas a política de indexação do sistema, em que princípios de recuperação da informação têm evidentemente um papel fundamental, mas considerações de ordem econômica também podem operar.

## **1.1 Os processos**

Há, por conseguinte, dois grandes processos básicos de tratamento da informação: a descrição física e a descrição temática. No âmbito desses macro processos vão-se encontrar grupos ou conjuntos de atividades, a que também se pode chamar de processos, como a análise de assunto, objeto deste livro e que é um processo específico no âmbito da descrição temática.

O processo de descrição física compreende, em primeiro lugar, um exame do documento com o objetivo de identificar certos elementos nele constantes e que vão servir para identificá-lo. Se o documento traz de forma clara, todos os elementos necessários a esse tipo de identificação, a tarefa é relativamente fácil, de modo geral. Entretanto, há casos bastante complexos que podem demandar extensas pesquisas para se estabelecerem as informações precisas. Por exemplo, pode não estar claro quem é o autor do livro, situação fácil de ocorrer quando o autor tem um nome muito comum. Nesse caso, é provável que haja outros autores com aquele mesmo nome, constituindo-se um problema identificar entre todos eles o verdadeiro autor do documento que se tem em mãos. É muito comum, igualmente, a ausência de dados importantes, tais como o local onde foi

publicado o documento e/ou a data de publicação. Com o tempo, passou-se a chamar a esse exame do documento para fins de sua descrição física de *leitura técnica do documento*, expressão que vai ser encontrada na literatura e na prática profissional<sup>3</sup>.

As etapas que se seguem, na descrição física, são as de decidir que dados serão aproveitados e que forma deverão ter. Todas essas decisões são geralmente tornadas com base num código de catalogação, instrumento essencial a uma descrição física eficiente e consistente, no caso de sistemas mais complexos.

Na descrição temática, repete-se (ou aproveita-se) o processo de exame, aqui também chamado de leitura técnica, apenas que, agora, com um objetivo diferente: o de identificar o assunto ou assuntos do documento. Esse exame constitui a fase inicial da chamada análise de assunto, primeira grande etapa da descrição temática. As etapas seguintes vão incluir processos como o de extração e seleção de conceitos, até que se determine afinal o assunto do documento.

A propósito, o profissional encarregado de fazer a análise de assunto pode ser designado por nomes diversos, dependendo principalmente do tipo de documento com que lida e do lugar onde desenvolve seu trabalho. Conforme foi dito acima, inicialmente, quando o tipo de material existente eram apenas os livros, é claro que a análise de assunto se limitava a esse tipo de material. Essa análise era feita nas bibliotecas. Mas não desde sempre e, mesmo hoje em dia, não em todas as bibliotecas, apenas naquelas onde havia a decisão de organizar os livros por assunto e também organizar por assunto as representações, ou simulacros (fichas de catálogos e referências bibliográficas) desses livros. Essas representações foram, por muito tempo, conhecidas como fichas de catalogação e, depois de elaboradas, havia ainda a necessidade de organizá-las para que pudessem ser devidamente pesquisadas. A forma padrão de organização ficou sendo o catálogo que, então, definia-se como um conjunto de fichas de catalogação organizadas de acordo com determinados princípios. O profissional que elaborava as fichas e organizava as em catálogos passou a ser conhecido como catalogador. Por conseguinte, o catalogador é, historicamente, o primeiro nome que é utilizado para designar a pessoa que faz o trabalho de análise de assunto. Como a catalogação consiste tanto de dados descritivos (autor, título etc.) como de dados temáticos (os cabeçalhos que expressam o assunto ou assuntos do livro), com o tempo surge o termo *catalogação por assunto* (*subject cataloging*, em inglês), utilizado quando se queria especificar apenas esta parte do trabalho de catalogação. Logicamente, quando também se queria especificar o profissional que faz esse trabalho, passou-se a usar a expressão *catalogador de assunto* (*subject cataloger*, em inglês). Embora não seja muito comum na prática, a distinção entre catalogador descritivo e catalogador de assunto é bastante utilizada na literatura técnica especializada bem como em cursos de formação de catalogadores.

Se a análise de assunto é feita com o objetivo de classificar um documento, ou seja, encontrar um número para ele num sistema de classificação bibliográfica, então denomina-se o profissional que desempenha essa tarefa de classificador. Se ao invés, o trabalho é realizado num serviço bibliográfico de elaboração de índices/abstracts, o nome usado para o profissional vai ser, provavelmente, indexador. E, em sintonia com as novas terminologias que começam a aparecer na área, fruto da emergência das bibliotecas virtuais e digitais, é provável que brevemente surjam termos como ontólogos e semelhantes,

---

<sup>3</sup> A prática dessa leitura técnica será tratada com detalhes no capítulo 4.

para designar os que trabalham com ontologias e com as demais novas especialidades.

## 1.2 Os instrumentos

Os principais instrumentos necessários ao tratamento da informação são, no que diz respeito ao tratamento descritivo, os códigos de catalogação e os formatos de metadados, e no que diz respeito ao tratamento temático, as linguagens de indexação. Uma variedade de normas de documentação são aplicáveis tanto a um tipo de tratamento quanto ao outro. Esses instrumentos serão abordados, por conseguinte, dentro de cada tipo de tratamento.

O tratamento descritivo de coleções tais como as de uma biblioteca, especialmente bibliotecas de médio e grande porte, exige a utilização de um *código de catalogação*. Este tem, por finalidade, estabelecer regras a serem seguidas na descrição de cada documento. Essas regras cumprem primordialmente uma função intelectual, a da correta identificação da obra que se manifesta na forma de um determinado documento. Por exemplo: Dom Quixote, o romance famoso, é de autoria de Miguel de Cervantes, escritor espanhol. Mas se, ao invés do original, tem-se uma versão condensada desse romance, o autor continua sendo Cervantes? Ou seria a pessoa que fez a condensação? É o tipo de problema que um código de catalogação procura resolver. Entretanto, o código cumpre também uma função econômica, pois a existência de regras agiliza o trabalho de catalogação que, dependendo do tipo de documento, pode até ser feito por técnicos de nível médio, ao invés de exigir um profissional de nível superior. A existência de regras é também um dos pressupostos do trabalho cooperativo, que também tem um impacto positivo muito grande na economia desses sistemas.

Devido à complexidade e variedade dos registros, os códigos tornaram-se instrumentos igualmente complexos e sofisticados. Necessitam de revisões periódicas e de toda uma infra-estrutura para a realização dessas revisões. Por conseguinte, a criação e manutenção de um código de catalogação é uma tarefa dispendiosa. Com o tempo, tem-se provado não haver viabilidade econômica para a existência de vários códigos simultaneamente. A cena é dominada, desde algum tempo, pelo Código de Catalogação Anglo-Americano — 2 edição - CCAA2 (CÓDIGO, 1983-1985), traduzido em várias línguas e utilizado em bibliotecas do mundo inteiro.

Os *formatos de metadados* cumprem, na descrição de documentos eletrônicos (um sue da Internet, por exemplo), funções semelhantes às desempenhadas pelos códigos de catalogação em relação aos documentos impressos. Certas características dos documentos eletrônicos exigem formas descritivas não contempladas nos códigos, que foram planejados basicamente para descrever documentos armazenados em suportes convencionais (papel, principalmente). Vários formatos de metadados têm sido propostos mas, também à semelhança do que vem acontecendo com os códigos de catalogação, um formato parece começar a impor-se, o chamado Dublin Core (DUBLIN, 1995-2006).

Outra importante aplicação de tratamento descritivo existente é na elaboração de bibliografias, queira-se com isto significar obras vastamente complexas, como os chamados índices/*abstracts*, ou as singelas bibliografias necessárias num trabalho escolar ou num artigo de periódico. Em ambos os casos, faz-se necessário descrever documentos por meio de representações ou

simulacros, neste caso chamados *referências bibliográficas*, ou citações. Para esse tipo de tratamento, faz-se também necessária a utilização de regras. Normas para a padronização das referências bibliográficas têm sido elaboradas por organismos nacionais e internacionais de normalização, tais como a International Standards Organization (ISO) e, no Brasil, a Associação Brasileira de Normas Técnicas (ABNT).

No tratamento temático da informação, a necessidade é de *linguagens de indexação*, instrumentos que vão fornecer os termos padronizados para representar o assunto ou assuntos identificados nos documentos analisados. Existem dois tipos principais de linguagens de indexação: linguagens alfabéticas e linguagens simbólicas. Os principais tipos de linguagens alfabéticas que existem são as listas de cabeçalhos de assunto e os tesouros. Os cabeçalhos são para uso principalmente de bibliotecas públicas e outros tipos de bibliotecas cuja abrangência de assunto seja mais geral. Os tesouros, por outro lado, são voltados para coleções especializadas. Quanto às linguagens simbólicas, estas se limitam praticamente aos sistemas de classificação bibliográfica, que utilizam símbolos como letras e números para representar os assuntos dos documentos. Atualmente, o mais popular desses sistemas é a Classificação Decimal de Dewey, usada por bibliotecas do mundo inteiro, inclusive bibliotecas virtuais (DEWEY, 2003).

Além das linguagens de indexação, um outro importante instrumento, geralmente usado em conjunto com um sistema de classificação, são as tabelas de notação. Essas tabelas têm por finalidade auxiliar o trabalho de classificação em sua tarefa de prover uma localização única e fixa para cada documento numa coleção, no que se refere ao tratamento temático, deve-se mencionar instrumentos como as normas para elaboração de resumos, desenvolvidas também sob a responsabilidade das instituições de normalização acima citadas. Deve-se mencionar, finalmente, os manuais de indexação e de catalogação por assunto. Os primeiros, usados nos serviços de indexação e resumo e em bibliotecas especializadas/técnicas ou centros de documentação; e os últimos, em todos os tipos de bibliotecas.

### **1.3 Os produtos**

Os principais produtos do trabalho de tratamento da informação são:

- registros bibliográficos;
- registros catalográficos;
- resumos;
- metadados (registros bibliográficos ou catalográficos de documentos eletrônicos);
- pontos de acesso de catálogos;
- pontos de acesso de bibliografias;
- arranjo sistemático de coleções de documentos.

Os registros bibliográficos e catalográficos têm muitos elementos em comum, pois de ambos podem constar alguns, ou mesmo todos, os elementos de descrição física e temática acima mencionados. São, por conseguinte, representações dos documentos a que se referem, pois retratam o essencial dos mesmos por meio de um número reduzido de informações. A denominação clássica para o registro bibliográfico é referência bibliográfica, ao passo que o registro catalográfico foi conhecido durante muito tempo como ficha catalográfica,

ou ficha de catalogação. A aplicação dos computadores nesse tipo de trabalho logo se revelou especialmente útil e, com eles, chegou também muita terminologia nova, caso do termo registro, que está sendo aqui utilizado, e que era desconhecido antes da chegada dos computadores. Se esses registros bibliográficos e catalográficos se referem a documentos eletrônicos, o nome que vem sendo usado na prática profissional para designá-los é metadados, conforme foi visto.

A diferença entre registros bibliográficos e catalográficos dizia respeito principalmente à existência física, ou não, de um determinado documento numa biblioteca ou centro de informação. Assim, se a biblioteca possuía um documento e fez sua catalogação, haveria um registro catalográfico desse documento que poderia ser utilizado sob várias circunstâncias para representá-lo. Por exemplo, mesmo se o documento estivesse emprestado, ainda assim o leitor nele interessado poderia encontrar informações sobre o mesmo no registro catalográfico (ou ficha de catalogação), que seria encontrado no catálogo (hoje, mais comumente, esse registro estaria armazenado num computador). Se a biblioteca não possuía um determinado documento, então não poderia fazer um registro catalográfico do mesmo, apenas possuir um registro bibliográfico, provavelmente elaborado externamente à biblioteca e obtido em alguma fonte de informação, principalmente bibliografias importantes, como as bibliografias nacionais.

Resumos, como se observou acima, são muito úteis para o próprio profissional da informação, pois podem ajudá-lo no trabalho de identificar o assunto de um documento. Mas nem todo documento vem acompanhado de um resumo, sendo muitas vezes necessário que sejam elaborados, o que é feito em muitas bibliotecas/SRIs, para que o usuário possa melhor avaliar a pertinência de um determinado item.

Os pontos de acesso são aqueles elementos da descrição que poderão ser utilizados na pesquisa no catálogo. O processo de escolha desses pontos de acesso, nos sistemas manuais, tendia a ser altamente seletivo, principalmente pelo custo que representava cada ponto a mais adicionado ao catálogo. Assim, aqueles elementos da descrição que tivessem pouca probabilidade de servirem como chave de busca eram geralmente desprezados. Por exemplo: o número de páginas de um livro. Com os sistemas automatizados, há uma tendência a se ampliar essa escolha, ou a se ser menos seletivo. Mas escolher um determinado elemento como chave de busca continua tendo um custo, pois na maioria dos sistemas significa a necessidade de criar o índice apropriado que vai permitir essa busca.

O arranjo sistemático de uma coleção significa organizá-la de acordo com algum tipo de classificação por assunto. Os arranjos mais sofisticados e mais eficientes são geralmente conseguidos com o uso de um sistema de classificação bibliográfica, como o de Dewey, acima mencionado. É em tais sistemas que têm se baseado a organização física dos documentos nas principais bibliotecas do mundo, desde o século XIX. O arranjo sistemático permite, entre outras coisas, a função do browsing, que continua importante nos sistemas automatizados, como se pode observar em vários tipos de websites (NAVES, 1998).

## 2 O processo de tratamento temático

No contexto do tratamento da informação, o termo *indexação* possui dois sentidos: um, mais amplo, quando se refere à atividade de criar índices, seja de autor, título, assunto, tanto de publicações (livros, periódicos), quanto de catálogos ou banco de dados, em bibliotecas ou centros de informação. O outro sentido, mais restrito, se refere apenas à indexação ou catalogação de assuntos das informações contidas em documentos.

Essa distinção se torna necessária face ao uso do termo na literatura da área, ora no primeiro sentido, ora no segundo, sem que seja feita a diferenciação, o que torna o estudo do tema um pouco confuso. Um exemplo disso ocorre com os artigos do periódico em língua inglesa *The Indexer*, que traz grandes contribuições para a literatura de indexação, mas não deixa muito clara essa distinção, fazendo referência, apenas raramente, entre indexação *back-of-book* e indexação acadêmica.

Essa distinção é feita por Farrow afirmando que a indexação *back-of-book* permite ao leitor localizar informação sobre um tópico dentro do livro; a tarefa do indexador é ler o texto, distinguir entre informação relevante e periférica e empregar juntos o processamento *top-down* (conceitual) e *bottom-up*, presentes na leitura fluente. Por sua vez, a indexação acadêmica fornece um termo útil estabelecido pela indexação praticada em bases de dados de resumos e em catálogos de bibliotecas, usando predominantemente a abordagem *top-down*. A segunda é considerada menos exaustiva que a primeira (FARROW, 1995).

Este livro concentra-se nos aspectos do segundo sentido acima citado e, neste caso, a indexação acadêmica compreende duas etapas distintas: a extração de conceitos que possam representar o assunto de um documento e a tradução destes para termos de instrumentos de indexação, que são as chamadas linguagens de indexação ou linguagens documentárias. Fugmann (1985) afirma que a indexação é o processo de: (a) discernir a essência de um documento e (b) representar essa essência com um grau suficiente de predicabilidade e fidelidade, isto é, num modo de expressão em linguagem de indexação.

A primeira etapa ou operação envolve representações implícita ou explícita do documento por uma frase de indexação, já a outra etapa envolve a tradução dos termos da frase de indexação para o léxico do vocabulário controlado de indexação, com a devida consideração pela semântica e sintaxe da linguagem de indexação (FROHMANN, 1990).

Outros, como Chu & O'Brien (1993) já consideram quatro estágios no processo, a saber:

- Análise de assunto do texto;
- Expressão do conteúdo do assunto nas palavras dos indexadores (linguagem natural);
- Tradução para um vocabulário de indexação;
- Expressão do assunto em termos do índice.

Para as autoras citadas, de acordo com a experiência de catalogadores de assunto e indexadores, os três primeiros passos podem acontecer quase simultaneamente, o que não significa que seja uma atividade simples. Certamente, se indexadores abordam um texto inteiramente com a intenção de decidir a questão de assunto em sistema de vocabulário, eles podem perder

algumas nuances que poderiam acrescentar os subseqüentes termos do índice (CHU; O'BRIEN, 1993).

Estudos de indexação se debruçam principalmente em questões das linguagens utilizadas no sistema de recuperação. Segundo Frohmann (1990), muita pesquisa focaliza apenas a segunda operação, enquanto sobre a primeira, considerada como uma opção intelectual junto à indexação, pouco é feito. Acrescenta um pensamento de Forkett, quando diz que a *operação chave da indexação, que é a decisão sobre o que o documento é, ainda é menos discutida e a menos reduzível a regras.*

Essa idéia parece antiga, pois ponto de vista semelhante já havia sido mostrado por Hutchins (1978) quando afirmava que a literatura de indexação e classificação contém pouca discussão do processo de indexação e classificação; encontra-se grande quantidade de literatura sobre a construção de linguagens de indexação e sistemas de classificação, sobre os princípios de classificação, sobre a correta formulação de entradas de índice e sistemas de informação. Mas pouco é encontrado sobre como indexadores decidem qual é o assunto de um documento, como eles decidem *sobre o que ele trata.*

O número baixo de pesquisas que aprofundam no tema é um elemento importante que tornam desafiadores os esforços na busca de dados que possam clarear alguns aspectos e, conseqüentemente, auxiliar no desempenho dessa atividade no tratamento da informação, Farrow é otimista a esse respeito e afirma que, em anos recentes, tem aumentado o interesse pela questão de como o texto é analisado para a indexação, sendo considerada uma área que requer mais pesquisa. Os processos cognitivos subjacentes à indexação (sumarização e produção de texto) foram tema de pesquisas na década de 70 e início da década de 80, principalmente no campo da psicolingüística, tendo exercido influência o modelo proposto por Kintsch e Van Dijk. Nesse modelo, o texto é primeiro convertido numa lista de proposições, usando a técnica de análise proposicional. As proposições são seletivamente suprimidas pela aplicação de macro-regras, modelando a forma pela qual a mente relembra o sentido geral do texto, embora ela não possa lembrar cada detalhe. O artigo de Farrow (1995) analisa a extensão pela qual a análise proposicional e as macro-regras auxiliam no processo de indexação.

Problemas no processo de indexação de assuntos são apontados por Bernier (1965). Ele acredita que há dificuldade em descrevê-lo e avaliá-lo, afirmando que até o conceito de assunto é difícil de definir, ensinar e aplicar, especialmente quando os assuntos são complexos ou implícitos.

Especialistas em recuperação de informação são os primeiros a declarar que a indicação de termos apropriados capazes de representar o conteúdo de itens de uma coleção são, ao mesmo tempo, a mais importante e a mais difícil de todas as operações normalmente usadas no processamento de informações contidas em documentos (SALTON; MCGILL, 1983).

A indexação é uma operação delicada, por vezes complicada, que ainda não obteve grandes suportes teóricos. É entendida como processo básico na recuperação da informação. Campos ressalta o *caráter integrativo* das duas etapas da indexação e admite que, apesar de distintas, deva se admitir que se trata de uma única operação em dois momentos. Afirma ainda que não há exagero em acentuar a *unicidade* das duas operações, sobretudo tendo em vista que a teoria da indexação deveria não apenas indicar o modelo de elaboração de uma linguagem de indexação, mas deveria também apresentar instrumentos para interpretação do documento. Este segundo aspecto não tem sido suficientemente



posto em relevo, mesmo porque as dificuldades que normalmente se encontram, são maiores do que, à primeira vista, se poderia imaginar (CAMPOS, 1987).

Antes de considerar indexação como uma rotina de situação cognitiva ou como compreensão de texto, David e outros autores acreditam que deve-se considerá-la como uma atividade de solução de problema: determinar os temas contidos em um documento e produzir uma lista de descritores obtidos de um tesouro o qual poderia ser mais ou menos adequado. Como qualquer problema, assuntos começam num estado de conhecimento inicial e se movem pelo espaço do problema (que é a representação, isto é, os possíveis estados do problema, as operações cognitivas ou físicas), até um estado final de solução (DAVID *et al.*, 1995).

A questão da subjetividade presente no processo de indexação é lembrada por vários especialistas da área, como Strehl (1998), quando ressalta que a atividade de indexação é um processo subjetivo. Cita Pinheiro: *A indexação envolve julgamento e, conseqüentemente, oscila muito no seu nível de concordância e apresenta discrepâncias*. Pode-se afirmar que, o que coloca em risco a coerência da indexação é que diferentes indexadores, ou até um mesmo indexador em momentos distintos, percebem de forma diferente:

- *O conteúdo real do documento;*
- *A parte desse conteúdo que será suscetível de responder realmente às necessidades (inevitavelmente futuras) dos usuários*
- *Os conceitos importantes que devem ser conservados para representar este conteúdo;*
- *Os descritores definidos para representar esses conceitos (STREHL, 1998).*

Para Pinheiro, não existem critérios objetivos de indexação e o termos podem ser considerados mais pertinentes, mais relevantes, mais informativos, etc.

Destaca-se, no processo de indexação, a necessidade de se estabelecer uma *política de indexação*, imprescindível na orientação da atividade do indexador. Contendo uma política bem definida, tendo em vista o perfil de seus usuários, o sistema de recuperação de informação apresenta maiores chances de eficácia no alcance de seus objetivos. Dentre os elementos que compõem essa política, podem ser citados os níveis de exaustividade e especificidade, capacidade de revocação e precisão do sistema, estratégia de busca, tempo de resposta do sistema forma de saída e avaliação do sistema.

## **2.1 Consistência e relevância na indexação**

Dois conceitos sempre presentes na literatura sobre indexação são os de consistência e relevância. É importante abordar sobre ambos, mesmo que superficialmente, para o entendimento do processo como um todo. No entanto, um estudo em profundidade exigiria dados sobre a satisfação dos usuários do sistema, o que foge ao escopo deste livro.

Uma opinião de Chu & O'Brien (1993) sintetiza bem o primeiro conceito: *a consistência é crítica em todos os estágios do processo de indexação*.

Um dos problemas que têm sido observados nos índices de assuntos, e que colocam em xeque o trabalho desenvolvido por indexadores, é o da variedade de termos que são definidos para representar o conteúdo de um mesmo documento num mesmo sistema, o que levanta a questão da *consistência* na indexação.

Quando é feita uma comparação entre os termos de indexação definidos por dois ou mais indexadores, para o mesmo documento, certamente são detectadas diferenças de julgamento de quais termos seriam os mais adequados. Conforme Rolling, essa diferença de julgamento introduz incerteza em qualquer sistema de recuperação baseado em indexação humana. Esse autor afirma que:

*Desde que a consistência de indexação manifesta-se na similaridade dos termos de indexação atribuídos a um dado documento por diferentes indexadores, e desde que a seleção de termos de indexação por um indexador reflita seu julgamento quanto à informação contida no documento e sua representação, a consistência da indexação é essencialmente a medida de similaridade de reação de diferentes seres humanos processando a mesma informação. Assim, mais precisamente, a consistência de indexação é definida como o grau de concordância na representação do conteúdo de informação (essencial) do documento por certas listas de termos de indexação selecionados individualmente e independentemente por cada indexador (ROLLING, 1981;p.69).*

Também é usado o conceito de inconsistência entre indexadores, o que significa que diferentes indexadores determinam diferentes termos para o mesmo documento. O nível de inconsistência varia caso a caso, mas observa-se que em nenhum estudo sobre a questão, a inconsistência é completamente ausente ou insignificante (BLALR 1990).

Uma das razões para a inconsistência pode ser resultante do momento da expressão de conceitos, que é feita de diferentes maneiras e em diferentes níveis de especificidade, o que ocorre durante o processo de análise de assunto. Svenonius (1981) concorda com esse ponto de vista, e lembra que a aparente inconsistência mostrada em estudos dos anos 60 do século passado resultaram do fato de que os conceitos selecionados para indexação foram verbalizados diferentemente, por diferentes indexadores, em diferentes níveis de especificidade.

Tonta (1991) cita Zunde & Dexter (1969)<sup>4</sup>, segundo os quais, essencialmente, consistência na indexação é vista como a medida de similaridade da reação de diferentes seres humanos processando a mesma informação. Consistência da indexação num grupo de indexadores é definida como o grau de concordância na representação da informação essencial do conteúdo do documento por certos grupos de termos de indexação, selecionados individualmente e independentemente, por cada indexador do grupo.

Estudos de consistência na indexação têm mostrado, ainda conforme Tonta (1991), que os valores da consistência variam muito entre indexadores. Os valores da consistência dependem do número de fatores sob os quais a indexação foi executada. Outra variável que afeta a consistência na indexação é a medida de consistência usada na avaliação, e existem diferentes métodos e fórmulas para calcular esses valores de consistência.

Critérios de pertinência, exaustividade e especificidade auxiliam os indexadores a avaliar o trabalho de indexação, mas esses são freqüentemente difíceis de serem operacionalizados. A medida de consistência entre indexadores dá uma indicação da confiabilidade na indexação, e uma demonstração parcial de validade do processo (DAVID et al., 1995).

Outro conceito sempre presente nos estudos de indexação é o de relevância. No contexto da recuperação da informação, dá-se o nome de relevância ao

---

<sup>4</sup> ZUNDE, Pranas; DEXTER, Margaret E. Indexing consistency and quality. *American Documentation*, v. 20 n. 3, p. 259-267, 1969.

juízo feito pelo indivíduo ao se confrontar com o resultado de sua busca em um sistema de recuperação da informação.

Relevância seria a conexão existente entre situações de usuários e a informação contida nos documentos. Seria a avaliação da satisfação do usuário com relação à representação de documentos.

Segundo Hiliman (1964), o termo é aplicado para definir a relação entre duas coisas, por exemplo, a questão e a resposta. Ele mostra que quatro fatores entram na definição do termo: a questão, a resposta, o grau e um *corpus*. A interpretação de conceitos depende do campo de conhecimento (*corpus*) para o qual estão sendo aplicados.

O termo tem sido usado na literatura de pesquisa experimental para se referir à relação entre um documento e uma questão, onde essa questão é a representação de uma necessidade de informação. Alternativamente, o termo pertinência, ou outro aspecto da relevância subjetiva, tem sido usado para se referir à relação entre documento e a necessidade de seu uso (HARTER, 1992).

Verifica-se que o conceito de relevância está diretamente ligado à necessidade de informação do usuário. Desta forma, a medida de relevância está ligada à sua satisfação, o que exige, para o estudo dessa questão, um estudo de usuário e o estabelecimento de critérios de avaliação.

### 3 A leitura do texto pelo indexador

Para a análise de assunto que se realiza em textos escritos é preciso que seja feita uma leitura que possibilite a extração de conceitos que sintetizem o conteúdo desses textos. Sabe-se da importância do texto, de estruturas e de tipos de textos, bem como a forma específica da leitura que deve ser feita pelo indexador.

#### 3.1 O texto

Considera-se texto como o meio, o veículo que permite a comunicação de idéias entre o sujeito que cria e dissemina informação (emissor, no caso, o autor) e o sujeito que necessita e adquire informação (receptor, no caso, o leitor). É o objeto que permite a transmissão das informações contidas em documentos, sendo também visto como uma coleção de símbolos, os quais são intencionalmente estruturados pelo emissor para mudar a estrutura de imagem do receptor. Existem textos orais mas, no caso em estudo, a atenção fica restrita a textos escritos, maioria esmagadora na produção científica de documentos publicados.

A palavra texto é, muitas vezes, confundida com documento, discurso, informação, dado, conhecimento ou até mesmo com literatura, que é adotado por Kaiser (1911)<sup>5</sup>, citado por Svenonius (1981). Nas escolas francesas e anglo-saxônicas, muitas vezes os termos texto e discurso são usados como sinônimos. Van Dijk afirma que texto é um conceito abstrato que se manifesta ou realiza em discursos concretos (KOBASHI, 1994).

O texto é considerado a unidade da análise do discurso e o que caracteriza a relação entre discurso e texto, para Orlandi (1987), é o seguinte: eles se equivalem, mas em níveis conceituais diferentes, ou seja, o discurso é tomado como conceito teórico e metodológico, e o texto, em contrapartida, como o conceito analítico correspondente, sendo o texto considerado uma unidade complexa de significação, tendo em vista as condições de sua realização. Pode ter qualquer extensão, desde uma simples palavra, até um conjunto de frases.

A esse respeito, Ginez de Lara (1993) considera que, para a *análise do discurso*, o discurso é um objeto concreto e se propõe invariavelmente como expressão da subjetividade. Analisá-lo implica extrapolar os aspectos que alguns teóricos consideram puramente lingüísticos do texto. Cada discurso é único e a análise do discurso tem como objeto discursos individuais ou um *corpus* bastante restrito, Essa é uma visão estruturalista do discurso, tendo-se uma visão mais atual em Santos:

A análise do discurso tem sido, nos últimos anos, uma das áreas de estudo da linguagem que mais tem recebido atenção de profissionais de diversos segmentos, devido ao seu caráter interdisciplinar. Consideram o processo comunicativo inerente às formas de expressão, observando, por exemplo, a intenção do emissor, sua interação com o receptor, a adequação de mensagem, enfim, toda a 'mise-en-scène' do contrato comunicativo (SANTOS, 1996; p. 3).

O texto é um objeto material tomado fora do contexto de sua produção, passível, portanto, de ser analisado em si mesmo. Enquanto no discurso se

---

<sup>5</sup> KAISER, J. *Systematic indexing*. London: J. Gibson, 1911. Paragraph 20.

expressa o sentido-apropriado individual da significação articulada pelos sistemas, no texto manifesta-se a significação como veículo de informação (COSTA VALA, 1991).

Tomado num sentido mais científico, o termo texto não corresponde ao seu uso comum. Refere-se a *grupo de unidades lingüísticas ligadas num conglomerado total de intenção comunicativa* (PETÕFI; GARCIA BERRIO, 1978<sup>6</sup> apud PINTO MOLINA, 1995). A dificuldade de uma descrição científica do texto deriva do fato de que ele é um sinal *open ended*, tendo um significado dinâmico e aberto, sobre o que o documentalista deveria pensar em grande profundidade. O estudo de macro símbolos demanda a integração interdisciplinar de variáveis semióticas da linguagem: sintaxe, semântica e pragmática. Segundo Renkema (1993), essas três áreas fazem parte, então, da Semiótica, que é o estudo de sinais, sendo a sintaxe relação entre sinais e o sistema de sinal; a semântica relação entre sinais e os objetos a que eles se referem; e a pragmática relação entre sinais e as pessoas que os usam, lidando, portanto, com questões sobre como os sinais funcionam.

Conforme Pinto Molina (1995), o texto é o ponto de partida necessário para operações analítico-documentárias. A autora considera a perspectiva de contemplar, no texto, a presença dos elementos conteúdo e forma, como partes essenciais do mesmo. No seu estudo de análise documentária, Ginez de Lara afirma que os textos veiculam informações de diversas naturezas e, para a documentação, tradicionalmente, eles constituem os documentos, que são a matéria prima da análise documentária.

Para facilitar seu trabalho, é necessário que o indexador conheça estruturas e tipos de textos. O sentido geral do texto é baseado na seguinte trilogia estrutural: *microestrutura* (estrutura superficial, que corresponde à realidade física do texto e seus símbolos de significação, as palavras), *macroestrutura* (concebida como um tópico representativo hierárquico e coerente da unidade textual, envolvendo mínima estrutura da representação textual sintático-semântica), e a *superestrutura* (estrutura retórico-esquemática, um tipo de esquema de produção convencional para o qual o texto é adaptado, podendo ser considerado como transição entre estruturas de superfície e de profundidade). Para Cintra (1983), os constituintes básicos de um determinado tipo de texto é que definem a sua superestrutura. Ela afirma que pesquisas têm demonstrado que leitores com conhecimento prévio específico sobre superestruturas textuais executam a tarefa de ler de forma mais fácil que leitores que não vêem essa superestrutura e, por isso, são obrigados a um maior apoio na leitura palavra por palavra, dificultando a integração das informações no texto como um todo. O leitor que domina as superestruturas textuais capta com mais facilidade as idéias centrais do texto, pois tem como parâmetro a identificação dos constituintes básicos.

Segundo Van Dijk, citado por Kobashi (1994), a superestrutura é considerada um elemento fundamental para a apreensão do significado do texto. Ela permite ao leitor monitorar a leitura, de modo a integrar as várias informações textuais. A superestrutura é, por definição, uma estrutura convencional, uma organização paradigmática, e o processo de compreensão supõe a transferência das unidades semânticas identificadas no texto para esse esqueleto conceitual.

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 podem-se identificar traços

---

<sup>6</sup> PETÕFI, J., GARCIA BERRIO, A. *Lingüística dei texto y crítica literária*. Madrid: Alberto Corazón, 1978

básicos que os caracterizam. O reconhecimento da tipologia dessas organizações pode conduzir, com maior eficácia, à identificação das partes mais significativas de um texto (GINEZ DE LARA, 1993).

Sabe-se, então, que todo texto se constrói segundo alguns princípios tipológicos. Existe grande diversidade de tipos de textos, e alguns tipos podem ser citados, como o de E. Werlich, que desenvolve um modelo tipológico textual que leva em conta dois tipos de critérios fundamentais: os dados do contexto extralingüístico, basicamente social, e as estruturas das orações que formam a chamada base textual. Seu modelo compreende os seguintes tipos: descritiva, narrativa, expositiva (sintética e analítica), argumentativa e instrutiva. Outro modelo é o de Grosse, que apresenta uma tipologia baseada no conceito de função textual, assinalando as seguintes funções: normativa, de contato, de indicação de grupo, poética, de auto-manifestação, de ordem/exigência, de transferência de informação, e das funções que atuam simultaneamente. A cada uma dessas funções corresponde um tipo de texto (BÉRNARDEZ, 1982).

A tipologia mais comumente utilizada divide os textos em narrativos e informativos. Os textos narrativos se caracterizam pela marcação cronológica (no uso dos diversos tempos para sinalizar diversos momentos narrativos, havendo referências a diversos momentos no tempo real da história, uma vez que o momento em que se dá a ação é importante para o desenrolar da mesma) e pela causalidade (o porquê do fato, sua motivação, são importantes também para desenvolver a estória) Para os textos informativos, a classificação mais conhecida, segundo Giasson (1993), é a de Meyer, e compreende: descrição (dá informações de um sujeito e especifica alguns de seus atributos e características) enumeração (o texto apresenta uma lista de elementos ligados entre si por um ponto comum), comparação (texto que serve para comparar objetos, pessoas ou acontecimentos entre si, tendo em conta suas semelhanças e diferenças), causa-efeito (é possível identificar no texto uma relação causal entre as idéias) e problema-solução (pergunta/resposta o problema antecede à solução).

Dentre os textos informativos, pode-se reconhecer o texto científico e, nesse tipo de texto, o conteúdo é quase inteiramente determinado pelo autor; geralmente, relatórios de pesquisa são altamente informativos e os autores os constroem numa estrutura convencional com introdução, metodologia, resultados e discussão, o chamado modelo clássico. Para Tibbo (1992), esse não é, entretanto, o formato tipicamente usado no campo das humanidades, que raramente incluem os tipos de seções mencionados, faltando-lhe qualquer tipo de padronização.

Pinto Molina (1995) faz também uma diferenciação entre textos produzidos no campo das ciências naturais daqueles produzidos pelas ciências sociais e humanidades, afirmando que há diferenças entre eles, acima de tudo quanto à metodologia e à importância dada à produção textual como um canal de expressão.

Uma vez identificadas a estrutura e a tipologia informacional do texto, esse é submetido a várias operações documentárias, tais como análise e condensação (resumos, enunciados, palavras-chave), representação, utilizando-se como instrumento comutador a linguagem documentária, que transforma o texto em mensagens de natureza diferente daquela que o originou. Por isso se afirma que a análise documentária tem, como objetivo recuperar e disseminar informação, e não, propriamente, textos, discursos ou documentos — suportes materiais da informação. Ginez de Lara afirma

(...) 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 (GINEZ DE LARA, 1993; p. 41).

O texto é visto muito mais do que como a simples soma das frases e das palavras que o compõem, e, para Koch, a diferença entre frase e texto não é meramente de ordem quantitativa, e sim qualitativa. Passou-se então a pesquisar o que faz com que um texto seja um texto, isto é, quais os elementos ou fatores são responsáveis pela textualidade (KOCH, 1991). Beaugrande & Dressler (1983)<sup>7</sup>, citados por Costa Val (1991), apontam sete fatores responsáveis pela textualidade: a coerência e a coesão, que se relacionam com o material conceitual e lingüístico do texto, e a *intencionalidade*, a *aceitabilidade*, a *situacionalidade*, a *informatividade* e a *intertextualidade*, que têm a ver com os fatores pragmáticos envolvidos no processo sociocomunicativo.

Embora muitos autores desconsiderem a distinção entre coesão e coerência, hoje em dia já se tornou praticamente um consenso de que se trata de noções diferentes.

O conceito de *coesão* textual é apresentado por Halliday & Hasan (1973)<sup>8</sup>, citados por Koch (1991), como um conceito semântico, que se refere às relações de sentido existentes no interior do texto e que o definem como um texto. Segundo eles, “a coesão ocorre quando a interpretação de algum elemento no discurso é dependente da de outro.” Nenhum elemento pode ser decodificado a não ser por recurso ao outro.

A coesão é uma relação semântica entre um elemento do texto e algum outro elemento crucial para sua interpretação e, por estabelecer relações de sentido, diz respeito ao conjunto de recursos semânticos por meio dos quais uma sentença se liga com a que veio antes, aos recursos semânticos mobilizados com o propósito de criar textos. A coesão manifesta-se no nível microtextual, referindo-se aos modos como os componentes do universo textual, isto é, as palavras que ouvimos ou vemos, estão ligados entre si dentro de uma seqüência.

A coerência é vista como responsável pela continuidade dos sentidos no texto, não se apresentando como mero traço do texto, mas como o resultado de uma complexa rede de fatores de ordem lingüística, cognitiva e interacional. Manifesta-se, em grande parte, macrotextualmente, referindo-se aos modos como os componentes do universo textual, isto é, os conceitos e as relações subjacentes ao texto de superfície, se unem numa configuração, de maneira reciprocamente acessível e relevante. Assim, a coerência é o resultado de processos cognitivos operantes entre os usuários.

Para Koch e Travaglia (1990), a coerência é que faz com que o texto faça sentido para os usuários, devendo ser entendida como um princípio de interpretabilidade, ligada à inteligibilidade do texto, numa situação de comunicação, e à capacidade que o receptor tem para calcular o sentido desse texto. Esse sentido deve ser do todo, pois... *a coerência é global*.

Fazendo uma relação entre coesão e coerência, Fávero afirma que os fatores de coesão são os que dão conta da estruturação da seqüência superficial do texto, e os de coerência os que dão conta do processamento cognitivo do texto e permitem uma análise mais profunda do mesmo. Diante disso, pode-se afirmar

---

<sup>7</sup> BEAUGRANDE, R; DRESSLER, W. *Introducion to text linguistics*. London: Longman, 1983.

<sup>8</sup> HALLIDAY, M.A.K.; HASAN R. *Cohesion in spoken and written English*. Londres: Longman, 1973.

que, para a Análise de assunto de textos, interessa mais o estudo da coerência textual, já que se lida com a macroestrutura do texto e a estruturação do sentido. Em resumo, tem-se que

(...) todo falante de uma língua tem a capacidade de distinguir um texto coerente de um aglomerado incoerente de enunciados e esta competência é lingüística. (...) Qualquer falante é também capaz de para frasear um texto, de resumi-lo, de atribuir-lhe um título, de produzir um texto a partir de um título dado e de distinguir um texto segundo os vários tipos de texto. Todas essas habilidades explicitam competência textual (FÁVEFO, 1993; p. 6).

Entre os fatores pragmáticos de textualização, ainda segundo Beaugrande e Dressier (1983)<sup>9</sup>, citados por Costa Vai (1991), a *intencionalidade* e a *aceitabilidade* se referem ao ato da comunicação. A intencionalidade concerne ao empenho do produtor em construir um discurso coerente, coeso e capaz de satisfazer os objetivos que tem em mente, numa determinada situação comunicativa. Diz respeito ao valor ilocutório do discurso. A *aceitabilidade* concerne à expectativa do receptor de que o conjunto de ocorrências com que se defronta seja um texto coerente, coeso, útil e relevante. O terceiro fator é a *situacionalidade*, que diz respeito aos elementos responsáveis pela pertinência e relevância do texto quanto ao contexto em que ocorre. É a adequação do texto à situação sociocomunicativa. O quarto fator é a *informatividade*, que diz respeito à medida na qual as ocorrências de um texto são esperadas ou não, conhecidas ou não, no plano conceitual e no formal. E o último componente definido pelos autores acima é a intertextualidade, que concerne aos fatores que fazem a utilização de um texto dependente do conhecimento de outro(s) texto(s).

Beghtol (1986) discute a relação existente entre o texto e os outros textos derivados desse, que também tem sido chamada de intertextualidade e que, segundo a autora... é o princípio por meio do qual a textualidade de qualquer texto surge da interação com outros textos. Para Koch e Travaglia (1990) a intertextualidade pode ser de forma ou de conteúdo: (a) de forma: ocorre quando o produtor de um texto repete expressões, enunciados ou trechos de outros textos, ou então o estilo de determinado autor ou de determinados tipos de discurso; (b) de conteúdo: pode-se dizer que, quanto ao conteúdo, a intertextualidade é uma constante, pois os textos de uma mesma época, de uma mesma área do conhecimento, de uma mesma cultura dialogam uns com os outros.

A existência de várias formas de intertextualidade é inquestionável e tem sido consistentemente reconhecida nas práticas tradicionais de catalogação e classificação. Existem relações entre um trabalho e seus vários *metatextos* derivativos, como sumário, resumo e sua expressão numa linguagem documentária. O texto original é chamado *texto primário*, um sumário ou resumo, *texto secundário*, e a expressão do texto primário numa linguagem documentária, *texto terciário*. São considerados textos independentes.

Um novo texto surge na atualidade, o chamado texto contemporâneo, que Levy (1996) considera um texto dinâmico, correndo em redes, fluido, *desterritorializado*, *mergulhado no meio oceânico do ciberespaço*. É o *hipertexto*, uma matriz de textos potenciais que se encontra num suporte digital e, por isso, permite novos tipos de leituras, nas quais os leitores podem criar, através de elos de ligação, seus próprios textos. modificando e acrescentando novos textos, imagens. O texto não é definido, mas compõe-se de dados atualizados em tempo

---

<sup>9</sup> BEAUGRANDE, R; DRESSLER, W. *Introducion to text linguistics*. London: Longman, 1983.



real. Como se vê, a nova concepção de texto deixa confuso o indexador; responsável pela organização e tratamento das informações que nele estão contidas.

Para ter uma competência textual é preciso que, além de conhecer o texto que tem em mãos para análise sob todos os aspectos até aqui abordados, o indexador faça dele uma leitura adequada, e sabe-se que um texto pode gerar muitas leituras, interessando mais, neste estudo, a leitura para fins documentários.

### 3.2 A leitura

A leitura é um tema muito explorado na literatura e existem diversos estudos sobre o tema em várias áreas do conhecimento, abordados sob diferentes enfoques. Para que os conteúdos dos textos sejam assimilados e compreendidos, é preciso que as informações sejam processadas na mente do leitor. No estudo ligado às questões da recepção da informação, a leitura aparece como um termo que *não se apresenta como um conceito preciso e rigoroso* (BIRMAN, 1994).

Ao se considerar a leitura como uma prática de dar sentido perceptivo e intelectual a um texto, está sendo inserido, nesse contexto, o sujeito e toda a sua capacidade subjetiva de interpretar. Como consequência disso, o estudo da leitura toma-se interdisciplinar, por conectá-la às ciências cognitivas e à semântica, principalmente.

Barthes e Compagnon (1987)<sup>10</sup>, citados por Birman (1994), sugerem que a leitura pode ser uma técnica, uma prática social, uma gestualidade, uma forma de sabedoria, um método e uma atividade voluntária. No aprofundamento de uma concepção de leitura, uma questão é levantada: a relação do sujeito com o texto.

Pode-se verificar que o conceito de leitura sofre transformações ao longo da história ocidental, e considera-se que a constituição do leitor como intérprete seja uma concepção de leitura bastante tardia, tendo-se fundado na passagem de uma concepção semiológica do signo para uma concepção hermenêutica.

Na Semiologia, procura-se estabelecer a origem e a referência da representação, enquanto na Hermenêutica a interpretação não se refere mais a uma suposta origem e a um objeto absolutos, mas remete a uma cadeia infinita de interpretações (BIRMAN, 1994).

Os avanços progressivos da leitura são abordados numa cronologia feita por Chartier, desde a leitura silenciosa na Idade Média à entrada no mundo da leitura extensiva do fim do século XVIII. Numa de suas reflexões, Chartier coloca a leitura como articulação de três séries de transformações: as tecnológicas (revoluções nas técnicas de reprodução de textos), as formais (mutações nas formas específicas do livro) e as culturais (mudanças em larga escala, tanto das competências quanto dos modos de leitura).

Esse autor cita um texto de Michel de Certeau, que contrasta o escrito — conservador, fixo, durável — com as leituras — sempre na ordem do efêmero. Para Chartier (1994), a leitura não está, ainda, inscrita no texto. O sentido lhe é imposto posteriormente e a interpretação é feita por seus leitores. Em outras palavras, o texto em si não tem significação, a não ser através de seus leitores. Cita ainda Paul Ricoeur, que analisa a maneira como se dá o encontro entre o *mundo do texto* e o *mundo do leitor*. Esse processo de atualização de textos exige considerar que as suas significações sejam dependentes das formas pelas quais eles são recebidos e apropriados por seus leitores. Na análise semântica do texto,

---

<sup>10</sup> BARTHES, A; COMPAGNON, A. Leitura, In: *Enciclopédia Einaudi*. Lisboa: Imprensa Nacional, 1987. v. 11.

estão presentes a crítica estruturalista, bem como as teorias literárias que reconstróem a recepção das obras (CHARTIER, 1994).

A leitura pressupõe uma prática de gestos, espaços e hábitos. As várias maneiras de ler distinguem comunidades de leitores e tradições de leitura. Nesse contexto, a dimensão social do processo de leitura pode ser identificada no abismo que surge entre alfabetizados e analfabetos, intensificando as desigualdades sociais. No entanto, pode-se afirmar que *o escrito está instalado no coração da cultura do analfabeto*, ou seja, estando presente nos rituais, nos espaços públicos, nos espaços de trabalho, o escrito se torna acessível mesmo àqueles incapazes de ler, ou que têm apenas uma compreensão rudimentar, graças à palavra que o decifra ou à imagem que o desdobra (CHARTIER, 1994).

Outros fatores sociais também interferem nesse processo, como, por exemplo, as características dos leitores (sexo, geração, adesões religiosas, grupos comunitários, tradições educativas e corporativas), pois influenciam na forma como ocorre a recepção da informação nos mais diversos níveis.

Na concepção moderna de leitura, é ressaltado o papel do leitor como produtor do sentido, numa dinâmica de forças que perpassa a relação do sujeito com o texto. Para Widdowson (1979)<sup>11</sup>, citado por Dell'Isola (1999), ler é um processo que envolve a combinação entre a informação textual e a informação que o leitor traz para o texto. Há uma espécie de diálogo entre o texto e o leitor.

*Um texto só existe se houver um leitor para lhe dar um significado.* Uma reflexão sobre essa afirmativa põe em evidência os dois componentes principais, atuantes no processo de leitura: o texto e o leitor, numa alternância de papéis, ora de ativos, ora de passivos (CHARTIER, 1994).

Ao tomar a leitura como um processo interativo leitor/texto, três fatores básicos são apontados por Kato (1985a) como suporte para a legibilidade: a qualidade do texto, o conhecimento prévio do leitor e o tipo de estratégias que o texto exige para ser lido. O caráter interativo da leitura pode ser garantido pelo comprometimento dos aspectos cultural e ideológico da linguagem tanto na produção do texto quanto na sua recepção (CINTRA, 1983).

Atualmente parece ser consenso entre os especialistas em leitura que o processamento do ato de ler se dá interativamente, dependendo dessa interação a não compreensão ou a compreensão de um texto. Essas duas possibilidades, entretanto, não podem ser inteiramente atribuídas ao texto, pois a legibilidade envolve outros elementos, além da boa formação de sentenças, da coesão e da coerência textuais, ocorrendo que, mesmo apresentando todos esses elementos, o texto pode não ser compreendido (ORLANDI, 1987).

A recepção da informação obtida do texto através da leitura dá-se de forma diferenciada de leitor para leitor, pela maneira de cada um atribuir sentido a seu conteúdo. Canclini (1995) fala em *estética da percepção*, relativa à experiência que os leitores absorvem das obras, *uma relação pura e imediata entre os sinais' emitidos pelo texto e o 'horizonte de expectativa' do público para o qual é dirigido.*

A leitura é considerada por Pinto Molina (1995) como a única maneira de se ter acesso ao conteúdo de um documento, sendo um processo simultâneo e não simplesmente simétrico à escrita. Há uma natureza interativa que depende tanto do texto quanto do leitor, consistindo em urna série de procedimentos coordenados que incluem operações perceptivas, lingüísticas e cognitivas.

Podem-se distinguir dois grupos de estratégias no processo de leitura: as cognitivas, que são comportamentos automáticos e inconscientes, e as metacognitivas, que supõem comportamentos não automáticos, em que o leitor

---

<sup>11</sup> WIDDOWSON, H. G. *Explorations in applied linguistics*. Oxford: Oxford University Press, 1979.

tem consciência de como está lendo. Essa distinção é feita por Kato (1985a) e ela acredita que o texto legível é aquele que exige uma aplicação equilibrada das duas estratégias. A autora dá uma abordagem psicolinguística ao processo de leitura. Na sua opinião,

(...) as pesquisas em leitura, principalmente na área da psicologia e da psicolinguística, são unânimes em afirmar que, na leitura proficiente, as palavras são lidas não letra por letra, mas como um todo não analisado, isto é, por reconhecimento instantâneo e não por processamento analítico-sintético (KATO, 1985a; p. 25).

Para a autora citada, a leitura eficiente é produto de três processos distintos. Num primeiro processo, cada bloco (segmentos maiores que as palavras) seria analisado em categorias correspondentes a palavras, através do conhecimento de regras léxico-sintáticas, atuando estas como unidades mínimas. O segundo processo seria o entendimento por respostas instantâneas do bloco, por poder ele ser extraído, não de um léxico mental, mas de um glossário mental. O terceiro processo seria o reconhecimento dos sintagmas, feito através da leitura de um sintagma recorrente no texto, que pode ser considerado como um tópico ou como um sub-tópico.

Ainda seguindo o raciocínio de Kato (1985a), tem-se que a leitura de palavras e blocos pode se dar de três maneiras:

- Através da resposta instantânea diante do estímulo, devido à existência do item no acervo de palavras e blocos do armazém da memória permanente, em sua forma e conteúdo;
- Através da resposta instantânea ao estímulo desse item no estado de consciência do leitor, ou memória a médio prazo, em sua forma e/ou conteúdo;
- Através da análise e síntese dos componentes do bloco, reconhecidos por um dos três processos aqui hipotetizados.

O processo de leitura depende da competência comunicativa do leitor, competência essa que sofre influência de vários fatores e, dentre esses fatores, destacam-se : (a) a ação da memória que, incessantemente, relaciona o não conhecido ao conhecido e (b) a participação da razão e suas atividades complementares de indução e dedução, análise e síntese (KATO, 1985b).

Cavalcanti vê a leitura como um processo em dois estágios: redução e mudança. A redução refere-se à tradução, que implica numa simplificação conceitual das próprias idéias do autor. Esse estágio é caracterizado pela inter-relação entre o conhecimento prévio e acumulado do leitor e sua atribuição de relevância às partes do texto. O leitor não passa por esse estágio sem realizar elaborações que levam à simplificação conceitual. A mudança refere-se à utilização da informação, isto é, à avaliação da informação processada na leitura de textos acadêmicos, ou seja, na leitura analítica. Esse estágio é relacionado ao efeito que o texto possa ter nas estruturas de conhecimento e nos sistemas de valores do leitor, que são as reações do leitor face ao texto. Para a autora, *uma vez que o texto potencialmente representa uma tentativa de mudar a visão de mundo do leitor, é possível identificar reações diferentes em diferentes leitores interagindo com um mesmo texto* (CAVALCANTI, 1989).

Segundo essa autora, podem surgir problemas durante todo um processo de leitura, como os apontados a seguir:

- Problemas Contingentes: são ligados, por exemplo, ao significado de uma palavra desconhecida;
- Problemas Táticos: são relacionados à organização do texto, por exemplo, a confusão entre os planos principal e secundário do discurso, devido a falha do leitor ou do escritor;
- Problemas Modais: devem ser resolvidos pelo próprio leitor. Sua solução, às vezes, depende da compreensão de um conceito ou de uma teoria;
- Problemas Ontológicos: são provenientes das crenças ideológicas do leitor, são construídos pelo leitor e geralmente resultam em comentários críticos sobre o texto.

A interferência desses vários fatores toma a leitura um ato subjetivo e individual, posto que o sentido dado ao texto lido sempre vai variar de leitor para leitor. Diante disso, o que poderia ser afirmado com relação à leitura para fins documentários? Existe algum modelo a ser seguido? E com relação ao leitor/indexador?

Sabe-se que um documento, inserido num Sistema de Recuperação da Informação, antes de ser lido pelo leitor, usuário final do sistema, é lido por um leitor técnico, o indexador, aquele que faz a leitura para fins documentários. Esse tipo de leitura, conhecido como leitura documentária ou leitura do indexador, tem certas características, não sendo realizada para lazer ou aprendizagem, nem é prazerosa, muito pelo contrário, O alto grau de incerteza, ansiedade e responsabilidade contido na atividade já mostra que a mesma traz pouca satisfação. É um tipo de leitura bem racional e rápido, em que o leitor técnico não tem chances de aproveitar a leitura, já que seu propósito é o de extrair o conteúdo informativo do texto, tendo em vista a sua posterior recuperação por um leitor interessado.

Outro aspecto que merece ser ressaltado é que o autor do texto, ao escrevê-lo, tem em mente um determinado leitor alvo para o qual direciona suas idéias; suas intenções não são dirigidas para o leitor/indexador e não lhe interessa se esse vai ter capacidade para interpretar as informações que aquele texto está veiculando. Cintra (1983) concorda com esse ponto de vista ao afirmar que, na leitura para fins documentários, é rompido o princípio de cooperação autor/leitor, já que o autor não previu o documentalista como leitor, e que esse fator deve ser acrescido à complexidade natural do processo.

As respostas àquelas questões propostas na página anterior já podem ser previstas. Obviamente que, considerando o fator subjetividade, tudo o que for dito pode ser correto ou não, dependendo da interpretação que as pessoas fazem a esse respeito. São várias as tentativas de se estabelecerem alguns critérios e de sistematizar o processo, mas não há um consenso quanto à forma mais adequada de se fazer essa leitura, visando à extração e ao posterior tratamento das informações contidas no texto.

## 4 A prática da leitura técnica

Conforme se pode observar dos capítulos anteriores, os conceitos de texto e de leitura são muito ricos e podem suscitar reflexões bem profundas, e o conhecimento de sua natureza e de suas peculiaridades só pode contribuir para que o analista de assunto exerça de forma melhor a sua tarefa. Entretanto, no dia-a-dia da grande maioria das bibliotecas/SRIs, o tempo que pode ser dedicado ao estudo e à reflexão sobre o próprio ato de análise de assunto é provavelmente pequeno, porque à espera do indexador estará certamente, quase sempre, um grande volume de documentos para serem lidos e analisados. Por conseguinte, é importante que, a par da reflexão e do estudo da questão, se ofereça também ao indexador metodologias e processos mais práticos de levar a cabo sua tarefa.

No exame do documento, as práticas foram inicialmente desenvolvidas em bibliotecas e principalmente voltadas para livros, que era o tipo de documento predominante. Daí nasceu a expressão *leitura técnica do documento*, significando uma forma de leitura do conteúdo do documento que fosse apropriada para a realização das demais tarefas da análise de assunto: identificação dos conceitos; seleção dos conceitos; e expressão do(s) assunto(s) do documento na forma de uma frase, ou frases, de indexação.

A leitura documentária poderia ser feita nos moldes de qualquer outro tipo de leitura, de vez que seu objetivo é idêntico ao de qualquer outro tipo de leitura, ou seja, a apreensão do conteúdo pelo leitor. Entretanto, fosse feita assim, iria exigir um tempo muito maior do profissional de informação. Isso não seria problema se o procedimento fosse uma exigência para que o processo de análise de assunto resultasse eficaz, ou mais eficaz. Entretanto, o que se foi descobrindo com o tempo é que, para fins do tratamento da informação, essa leitura normal, completa, não só é impraticável, por falta de tempo, mas também desnecessária.

Impraticável porque a quantidade de documentos incorporados a qualquer biblioteca/SRI é sempre tão grande que, para fazer uma leitura completa de cada um desses documentos, seria necessário um quadro de recursos humanos muito além do que a maioria dos recursos financeiros desses sistemas poderia suportar. Desnecessária porque foi-se identificando certas partes do documento cujo exame provou-se suficiente para que se atingisse os objetivos desse trabalho, que é o de determinar o assunto ou assuntos de um documento. Essas partes são: o título, a introdução, o resumo, os títulos dos capítulos e das seções, entre outras.

A leitura técnica tornou-se, assim, a estratégia clássica de leitura para análise de assunto. O que é a leitura técnica? Segundo Lancaster, “um misto de ler e passar os olhos pelo texto” (LANCASTER, 1993). Mais especificamente, consiste numa leitura direcionada para certas partes do documento onde vai-se encontrar elementos especialmente importantes para a identificação do assunto ou assuntos do documento. As partes mais comumente citadas nos manuais de catalogação e indexação são as seguintes:

- Título
- Subtítulo
- Sumário
- Resumo
- Introdução
- Prefácio
- Apresentação
- Títulos dos capítulos

- Bibliografia.

Vamos analisar, a seguir, cada um desses elementos e mostrar como podem ser explorados na prática da análise de assunto.

**Título.** A importância do título deriva do fato de que sua função precípua é a de sintetizar o conteúdo do documento. Essa é uma função que é efetivamente cumprida em relação a vários tipos de textos e de documentos:

**BIOLOGIA PARA ALUNOS DO ENSINO MÉDIO E VESTIBULANDOS**

www.tiraduvidabiologia.hpg.com.br<sup>12</sup>

**A BIOLOGIA NOS VESTIBULARES**

SPINGA, Renato Mascarenhas. *A biologia nos vestibulares*. São Paulo: Navegar, 2001. 160 p.

**HISTÓRIA DA BIOLOGIA COMPARADA**

LLORENTE, Jorge; PAPAVERO, Nelson. *História da biologia comparada*. São Paulo: HOLOS, 2000. 168 p.

Examinando-se os documentos a que se referem esses títulos, pode-se observar que cada um desses títulos consegue transmitir uma idéia muito boa do que seja o conteúdo do respectivo documento. É claro que dentro das limitações que o elemento título apresenta, ou seja, não se deve esquecer que o título representa uma condensação dramática de um conteúdo que, muitas vezes, pode estar contido ou desenvolvido em centenas de páginas.

Entretanto, nem sempre o título reflete bem o assunto, e isso usualmente ocorre com títulos muito genéricos:

**AÇÚCAR**

FREYRE, Gilberto. *Açúcar*. São Paulo: Companhia das Letras, 1997.

Para o leitor que tiver de se valer apenas do título para ter uma idéia da obra acima, o mesmo não será, provavelmente, de muita utilidade. O açúcar, como assunto, pode ser tratado sob vários pontos de vista. Em conseqüência, o único entendimento perfeitamente adequado seria o de que essa obra trata do assunto 'açúcar' das várias perspectivas possíveis, tais como o ponto de vista técnico (produção do alimento), econômico, social, entre outros. Embora o leitor que tenha alguma familiaridade com a obra de Gilberto Freyre pudesse deduzir o possível enfoque no tratamento do tema, com base no conhecimento que tivesse desse autor e de sua obra, ainda assim as possibilidades de abrangência da temática continuariam grandes.

Os problemas de se fazer análise de assunto com base no título, entretanto, começam com os títulos que simplesmente nada dizem sobre o assunto do documento. Por exemplo, estes dois:

- *A dança dos mestres Wu Li*
- *O ensino através da televisão*

---

<sup>12</sup> Os documentos eletrônicos disponíveis na web, pela própria natureza do suporte e da rede, estão sujeitos a grande instabilidade quanto ao seu conteúdo, podendo inclusive simplesmente desaparecerem, da noite para o dia. Os exemplos citados neste texto não dependem de consulta aos próprios documentos para entendimento do que se quer demonstrar, mas caso o leitor depare com algum documento eletrônico que não estiver mais disponível na rede, ainda assim poderá tentar encontrá-lo no Internet Archives. Como o próprio nome indica, o IA procura arquivar todos os documentos eletrônicos que em algum momento estiveram disponíveis na rede. Endereço do Internet Archives: <http://www.archive.org>

É praticamente impossível antecipar, a partir desses títulos, que o primeiro trata da física moderna, ou que o segundo trata especificamente do ensino de ciências.

Mas o verdadeiro problema com os títulos que nada dizem sobre o assunto do documento, do ponto de vista do tratamento da informação, é quando os indexadores se baseiam nele para fazer a indexação. Muitos exemplos reais comprovam isso, como o do livro *A ceia dos cardeais*, classificado como livro de culinária, numa determinada biblioteca. Tal classificação foge completamente do conteúdo do documento, pois trata-se de uma peça teatral do autor português Júlio Dinis e que nada tem a ver com culinária. Por conseguinte, uma classificação apropriada deveria ser literatura portuguesa ou teatro português, dependendo da necessidade da biblioteca/SRI de classificá-lo de forma mais, ou menos, específica.

Títulos de documentos ou de capítulos que contenham a conjunção ‘e’ merecem atenção especial, porque o ‘e’ pode ter dois sentidos diferentes:

- a) expressar algum tipo de relação entre partes de um mesmo assunto. Neste caso, estaríamos lidando com um assunto com- posto;
- b) expressar dois assuntos diferentes tratados num mesmo documento.

No primeiro caso temos um item que deve ser tratado como uma entidade singular; no segundo caso, o item pode ser dividido e tratado como se, na verdade, estivéssemos lidando com dois documentos diferentes.

Outro problema é que, mesmo quando o título é informativo, ele se refere mais aos fenômenos estudados do que à forma de conhecimento (LANGRIDGE, 1989).

**Subtítulo.** Embora nem sempre mereça grande atenção, o subtítulo pode conter informação de muita valia para o analista de assunto. Voltemos a alguns dos exemplos acima dados e observemos como os subtítulos acrescentam informação importante aos títulos dos documentos:

**AÇÚCAR; uma sociologia do doce, com receita de bolos e doces do Nordeste.**

FREYRE, Gilberto. Açúcar: uma sociologia do doce, com receita de bolos e doces do Nordeste. São Paulo: Companhia das Letras, 1997.

**A BIOLOGIA NOS VESTIBULARES; resumo teórico e vestibulares resolvidos.**

SPINGA, Renato Mascarenhas. *A biologia nos vestibulares; resumo teórico e vestibulares resolvidos*. São Paulo: Navegar, 2001. 160p.

**HISTÓRIA DA BIOLOGIA COMPARADA; desde o gênese até o fim do império romano do Ocidente.**

LLORENTE, Jorge; PAPAVERO, Nelson. *História da biologia comparada; desde o gênese até o fim do império romano do Ocidente*. São Paulo: HOLOS, 2000. 168 p.

Como se pode observar, os subtítulos acrescentam informação importante aos títulos dos documentos. Dessa forma, colaboram para que se tome mais fácil a tarefa de identificar o conteúdo dos mesmos.

**Sumário.** Depois do título e subtítulo, pode-se passar ao sumário, ou aos próprios títulos dos capítulos, na ausência de um sumário. O sumário tem como objetivo mostrar a seqüência de desenvolvimento do assunto no livro. Por isso, o

sumário tanto vai complementar informações já encontradas em outros dos elementos aqui tratados, como trazer informações sobre esse desenvolvimento, que podem servir para uma melhor caracterização de como o assunto é tratado no documento.

O sumário, de acordo com as normas de documentação mais utilizadas, deve vir logo no início do documento. Na prática, pode aparecer no final do mesmo. Pode também aparecer com outros nomes, sendo muito comum utilizar-se a designação de índice. Isso costuma induzir a alguma confusão no leitor menos familiarizado com essas diversas partes do documento, porque o índice é também uma parte importante em muitos tipos de livros. No entanto, tem uma função completamente diferente da do sumário.

**Resumo.** Outra parte importante do texto ou documento, quando existente, é o resumo, porque, assim como o título e o subtítulo, é também uma tentativa de apresentar uma síntese do texto. O diferencial é que pode e costuma ser mais informativo do que o título e o subtítulo.

Observe-se abaixo o resumo referente ao livro de Gilberto Freyre acima citado:

**AÇÚCAR; uma sociologia do doce, com receita de bolos e doces do Nordeste.**

FREYRE, Gilberto. *Açúcar*. uma sociologia do doce, com receita de bolos e doces do Nordeste. São Paulo: Companhia das Letras, 1997.

RESUMO: Em '*Açúcar*', o sociólogo pernambucano Gilberto Freyre reúne preciosas receitas de bolos e doces guardadas por tradicionais famílias nordestinas. Enfocando o açúcar como o elemento responsável pela liga entre diversos paladares e culturas, o autor demonstra a existência de uma arte do doce no país criada à sombra da escravidão.

Com efeito, o resumo traz algumas informações adicionais que não constavam nem do título nem do subtítulo, e que definem melhor o escopo da obra. A partir desses dados, o analista de assunto passa a ter uma concepção melhor do modo como o assunto *açúcar* é tratado no livro.

Um aspecto para o qual o indexador deve atentar, quando for utilizar o resumo para fins de análise de assunto, é o da autoria do resumo, que tanto pode ser do próprio autor do texto quanto de terceiros. Embora muitos possam se entusiasmar com o fato de que o próprio autor fez o resumo, e essa é uma tarefa para a qual ele certamente possui algumas das melhores credenciais, é bom observar que a elaboração de resumos é, por si só, uma técnica. Como tal, exige conhecimento não apenas da matéria substantiva de que trata o texto, mas também de técnicas documentais, que podem fazer com que um resumo seja mais ou menos eficiente nos seus objetivos. Evidentemente, não se pode esperar de todos os autores de textos que além de serem especialistas na matéria de que estão tratando, sejam também especialistas em técnicas documentais.

**Introdução.** A introdução, ou mesmo alguns dos parágrafos iniciais do documento, é também uma das partes destacadas como propícias a fornecer indicações sobre o assunto do texto. Com efeito, num texto bem estruturado e de acordo com os melhores princípios da comunicação escrita, a introdução deve começar falando do objetivo do documento, e esse tem que apontar necessariamente para o cerne do conteúdo deste.



**Conclusão.** Assim como ocorre com a introdução, também na conclusão constante de um texto bem escrito vai-se encontrar referência ao objetivo do documento e, por conseguinte, ao seu assunto.

**Série.** A série, ou coleção, a que eventualmente um documento pertença, é de modo geral um indicador ou indicador adicional do seu assunto. Observe-se, por exemplo, os seguintes documentos, todos pertencentes a diversos tipos de coleções:

**A TRANSIÇÃO PROLONGADA: O PERÍODO PÓS-CONSTITUCIONAL.**

FERNANDES, Florestan. *A transição prolongada: o período pós-constitucional*. São Paulo: Cortez, 1990. 239 p. (série: Biblioteca da educação; série economia política).

**AMÉRICA LATINA**

BESSONE, Tânia Maria Tavares; QUEIROZ, Tereza Aline P. (Coord.) *América Latina: imagens, imaginação e imaginário*. Rio de Janeiro: São Paulo: Colação, [s.d.j], 755p. (Série: América 500 anos).

**CURSO DE REDAÇÃO**

ABREU, Antonio Suarez. *Curso de redação*. São Paulo: Ática, 1989. 144p. (Série: Básica universitária).

**HÁ QUEM DIGA QUE O AMOR É UM RIACHO**

PADRE Zezinho. *Há quem diga que o amor é um riacho*. São Paulo: Paulinas, 1992. 45p. (Série: Adolescência e juventude).

Com base nessas informações do índice, pode-se observar, por exemplo, que duas categorias de obras de referência — os dicionários e as enciclopédias — que normalmente costumam merecer atenção igual em obras dessa natureza, aparentemente receberam atenção bem diferente, pelo menos a julgar pelo número de páginas a elas dedicadas: oito páginas para os dicionários e apenas duas páginas para as enciclopédias. Evidentemente, nem sempre um dado puramente quantitativo como esse pode bastar para se chegar a conclusões quanto ao tratamento dado a um assunto num determinado documento, mas em muitos casos pode ser um indicativo inicial ou complementar. Um tal dado, combinado com outros, toma a análise ainda mais segura.

**Bibliografia.** As listagens bibliográficas também constituem uma outra parte importante para o trabalho de análise de assunto. Essas listagens se resumem quase que exclusivamente à Bibliografia Consultada, mas podem eventualmente incluir outros tipos de listagens, como a Bibliografia Recomendada, por exemplo. A clara natureza temática da relação entre a bibliografia consultada e o próprio documento há muito que foi estabelecida na ciência da informação, tendo resultado num importante tipo de instrumento de recuperação de informações, que é o índice de citações. O pressuposto básico é que o assunto dominante, ou predominante, da bibliografia é, ou deveria ser, o mesmo assunto que é tratado no documento. Se assim for, então um exame da bibliografia e a identificação do assunto dela, ou do assunto nela predominante, representará seguramente um bom indicador no trabalho de identificação do assunto do próprio documento.

Embora o método da leitura técnica tenha sido desenvolvido para aplicação a documentos escritos e apoiados no suporte papel, pode muito bem ser aplicado a outros tipos de informação (visual, sonora) e a outros suportes, inclusive o suporte eletrônico, que se toma cada vez mais importante. Assim, é perfeitamente

possível se fazer uma leitura técnica de um site da Internet, por exemplo. Isso é óbvio se pensarmos que muitos desses recursos eletrônicos são apenas reproduções literais de documentos escritos. Entretanto, podemos também aplicar a leitura técnica àqueles recursos eletrônicos mais sofisticados, porque do ponto de vista documental as diferenças que já foram possíveis de serem identificadas entre documentos em papel e documentos eletrônicos não têm, praticamente, nenhum impacto do ponto de vista temático. A única característica mais saliente dos documentos eletrônicos que poderia ter esse impacto é a sua volatilidade, ou seja, a facilidade com que podem ser alterados ou mesmo simplesmente desaparecer do espaço virtual da noite para o dia, conforme se tem observado com frequência.

A fase seguinte da análise de assunto é a extração de conceitos que representem o conteúdo de itens, tema a ser abordado no próximo capítulo.

## 5 Extração de conceitos

Para definir em termos adequados o assunto de um texto é necessário que, primeiro, se extraíam os conceitos que nele estão contidos. Se, para fazer uma análise conceitual, devem-se extrair conceitos, pergunta-se: o que é um conceito? Como identificá-lo? Qual a sua importância no processo de análise de assunto? Para responder a essas questões, são feitas, a seguir, algumas considerações sobre conceito, assunto e contexto, todos eles constantes no processo em estudo.

### 5.1 Conceito

A base de todo o campo da ciência é o seu corpo conceitual, constituído e desenvolvido com muita reflexão e análise crítica por parte de seus pesquisadores. Sua representação vem a formar o conjunto de termos relativo a determinado campo, conjunto esse denominado terminologia, utilizada, assim, para representar esse corpo conceitual.

É importante a clareza no entendimento do que possa significar o termo conceito, que muitas vezes é empregado com imprecisão.

Conceitos são unidades do conhecimento identificadas através de enunciados verdadeiros sobre um item de referência e representados por uma forma verbal (termo ou palavra) (MEDEIROS, 1986).

Em síntese, conceito pode ser definido, também, como um conjunto de características, que são os elementos dos conceitos e traduzem os atributos das coisas designadas. A característica mais geral é chamada categoria, que é o conceito na sua mais ampla extensão.

Os conceitos são essenciais à vida dos indivíduos, pois eles simplificam sua percepção do ambiente e permitem a identificação dos objetos que se encontram no seu ambiente e o acréscimo de novos elementos aos esquemas individuais de cada um. Definir um conceito é uma operação verbal e lógica bem clara, na qual se usa uma série de idéias logicamente subordinadas para chegar a uma conclusão geral.

Para Oliveira (1997), conceito é visto como uma imagem subjetiva do mundo objetivo e revela aspectos essenciais, universais do objeto, abstraindo-se dos aspectos secundários. Para formar o conceito de algo, faz-se uso de vários processos mentais, como análise, síntese, abstração e generalização.

- Análise: operação mental que consiste em separar em partes, decompor, fragmentar um todo (objeto ou fenômeno) em seus elementos constituintes;
- Síntese: operação mental que, ao contrário da análise, consiste em recompor um todo a partir de seus elementos constituintes, a fim de compreendê-lo em sua totalidade;
- Abstração: operação mental que consiste em isolar ou separar, para considerá-lo à parte, um elemento ou parte de um todo que não é separável na realidade, a fim de distinguir o particular (acidental) do geral (essencial);
- Generalização: operação mental que consiste em estender a toda uma classe de objetos ou fenômenos os elementos essenciais, gerais, universais, constatados num certo número de objetos ou fenômenos da mesma classe.

Conforme Oliveira, dessa maneira é que se formam os conceitos e quanto mais abstrato for o conceito mais geral ele será. Todos os quatro acima são processos mentais inseparáveis.

A construção de conceitos pode ser afetada por alguns fatores, como:

- Percepção: esta é que inicia o trabalho mental, é a matéria prima para o pensamento;
- Emoção e Atitude — toda emoção inclui pelo menos três aspectos inter-relacionados: sentimentos, alterações orgânicas e impulsos para a ação. O termo atitude designa tipos de predisposição para a ação, como opiniões, preconceitos e nível de abstração;
- Linguagem — é um sistema de símbolos verbais, palavras, elaborado e utilizado por uma comunidade humana para exprimir e comunicar sentimentos e pensamentos.

Nas bases da Lógica, Dahlberg afirma que os filósofos alemães Immanuel Kant e Gottlieb Frege inferiram, já, a geração do conceito pela predicação. Para a autora,

(...) um conceito é criado pela predicação sobre um objeto de concernência, o chamado 'Referente', é originado nos escritos desses filósofos. Qualquer predicação sobre esse referente produz uma característica do conceito desse referente. A soma total das características dessas predicções possíveis irão compor a soma total das características de um conceito e assim determinar os conteúdos de um conceito (DAHLBERG, 1987; p.65).

Os itens que entram na elaboração final completa do conceito são representados pelas relações entre o objeto de referência, o próprio conceito e sua expressão lingüística, incluindo: (a) a referência de um item (da realidade); (b) as afirmativas sobre o item de referência, produzindo os elementos ou características da unidade do conceito e a necessária verificabilidade (ou controlabilidade) por outras dessas afirmativas; e (c) a designação por um termo, representando a síntese dos elementos do conhecimento. Esses itens são derivados do chamado *Triângulo do Conceito*, versão do *Triângulo Semântico*, criado por Ogden & Richards em 1936.

Farradane (1980), em seu estudo do escopo da ciência da informação, faz uma distinção entre conceitos e palavras individuais, e acha que conceitos individuais são elementos de pensamentos e palavras individuais são somente rótulos para os conceitos tratados na mente. Conceitos são conectados na mente por regras de pensamento, mas essas regras não devem ser confundidas com gramática e sintaxe, ou as regras da linguagem, que variam de uma linguagem para outra.

Já Fugmann (1985) distingue os conceitos como individuais ou gerais. Para ele, os individuais são representados por nomes de coisas individuais, em linguagem simples e sucinta, freqüentemente por uma simples linha de caracteres alfanuméricos, somente por expressões lexicais. E os gerais são representados por nomes de classes de coisas e podem ser expressos em uma multiplicidade de expressões lexicais e não lexicais.

## 5.2. Assunto

A extração de conceitos tem como produto um assunto, que representa o conteúdo informacional de um texto. Parece uma coisa óbvia explicar o que é assunto. No entanto, para muitos, esse termo é considerado ambíguo. É um conceito impreciso e difícil de definir e ensinar. A essência do tema e sobre o que o autor escreveu são outras formas de designar assunto. Uma pessoa pode, usualmente, selecionar um assunto de um item que ela compreende, pode parafraseá-lo e registrá-lo. Mas não pode estar apto a dizer como (por qual processo) ela seleciona e parafraseia o assunto. Essa idéia é de Bernier (1965), e, para ele, as pessoas podem fazer as coisas sem estar aptas a dizer precisamente como. A respeito da questão, esse autor afirma que *identificação de assunto é difícil de ensinar e aplicar, especialmente quando os assuntos são complexos ou implícitos. A pessoa sem conhecimento no campo indexado acha impossível a identificação consciente de assuntos* (BERNIER, 1965).

Metcalf (1977) acredita que há conflito e confusão no uso da palavra, tanto para quem compila quanto para quem consulta catálogos e índices de assunto. Há ambigüidade no seu uso geral e específico, bem como no seu uso por objetos de informação, geral e especial, concreto e abstrato, real e imaginário. O autor conclui que *assunto* não é um termo satisfatório na recuperação da informação, por causa dessa ambigüidade. No entanto, apesar da complexidade e da falta de consenso entre os estudiosos, esse é o termo mais utilizado na literatura.

Segundo Todd (1992), nesse campo há uma considerável confusão terminológica. Ele cita autores como Cutter (que define assunto como tema ou tópico, podendo ou não estar no título do documento), Kaiser (que toma *assuntos como coisas em geral*, reais ou imaginárias, e as condições para designá-las, que são chamadas *concretos e processo*), Ranganathan (que fala sobre o pensamento contido no documento), Coates (que identifica assunto como abstração da idéia global corporificada no conteúdo de uma unidade literária dada), e Vickery (que se refere ao tema a partir do qual livros, parte de livros, artigos ou parte de artigos são escritos). Todd (1992) acrescenta que, na literatura mais recente, aparece o uso do termo *aboutness* (neste livro, traduzido por atinência), como sinônimo do termo *subject* (assunto) de um documento.

Talvez essa confusão terminológica existente em torno do termo *assunto* possa ser responsável pela complexidade que envolve a representação dos assuntos dos documentos. Sobre essa questão, merece destaque o artigo de Hjørland (1992), que mostra o termo *assunto* como uma idéia, ou num sentido objetivo (platônico), ou num sentido mais subjetivo. Considera o conceito de *assunto* como subjetivo-idealístico e como objetivo-idealístico. O primeiro, o autor acredita que seja a expressão das percepções ou visões de um ou mais indivíduos; conceitos e assuntos são aquilo que é subjetivamente compreendido ou entendido por eles. Portanto, a chave do conceito de *assunto* está no estudo da mente de algumas pessoas. Já com relação ao conceito objetivo-idealístico, o autor pensa que esse não considera o assunto como subjetivo e, sim, tende a enfatizar aspectos da análise teórica e os torna absolutos e, para ele, as idéias existem independentemente da consciência humana. Em síntese, é a seguinte a sua teoria sobre o que assuntos são: as potencialidades de documentos para o avanço do conhecimento.

A noção de *assunto* de um documento é indeterminada, pois há casos em que é impossível, em princípio, decidir qual, de duas diferentes e igualmente precisas descrições, é a descrição do assunto, ou se o documento tem dois assuntos, ao invés de um. De duas descrições, que não são descrições de uma

mesma coisa, pode ser impossível dizer qual precisamente, descreve o assunto. Quanto mais vaga e geral nossa representação de assunto de um documento, menos ela é aberta a questões, e quanto mais exatos e precisos tentamos ser, mais provável é que várias descrições igualmente exatas, de diferentes coisas, serão formuladas, dentre as quais não se pode escolher uma, exceto agindo arbitrariamente. Isso significa que há sérias dificuldades na escolha de um assunto que seja considerado o principal de um documento, deixando-se outros, que ele inclua, em plano secundário.

Hjørland (1992) afirma que temos uma inclinação a dizer que o que é verdade sobre as coisas deve ser verdade dos escritos sobre as coisas, mas que se deve resistir a isso, bem como à afirmação de que deve haver um assunto definido para um documento, e que existirão algumas descrições do assunto que são absolutamente precisas e exatas, sendo todas as outras imprecisas ou inexatas. Na sua opinião, deve-se evitar dizer 'a' descrição 'do' assunto. Ele oferece um novo conceito de assunto, como a totalidade dos potenciais epistemológicos de documentos.

A atividade de identificar a(s) idéia(s) principal(ais) do texto exige a capacidade de compreensão de seu conteúdo, o que está ligado a processos cognitivos. Para Monday (1996), *o domínio dessa atitude é feito à base de uma leitura crítica e de uma facilidade de síntese*. A questão levantada pela autora é quanto à idéia principal de um texto e de como se pode definir essa noção. Como um leitor analisa o conteúdo de um texto para extrair a mensagem principal? Ela afirma que os tipos de textos são ignorados e que é importante distinguirem-se gêneros literários. Pode-se dizer que cada gênero possui sua própria definição de idéia principal.

Giasson (1993) faz uma distinção entre assunto e idéia principal. Este último conceito encontra-se expresso em diversos vocábulos tais como mensagem do autor, visão de conjunto, elementos importantes, ponto de vista principal, idéia central do texto, havendo, a respeito, diversidade de concepções. A avaliação do que é importante num texto pode variar de um leitor para outro, sendo consideradas duas categorias de informação importante: a *textualmente importante* (a informação é importante porque o autor a apresenta como tal) e a *contextualmente importante* (a informação pode ser importante porque o leitor a considera como tal, devido à sua intenção de leitura). Essa informação importante pode variar segundo os tipos de texto: nos textos narrativos, a idéia principal tem a ver com os acontecimentos e a sua interpretação, enquanto nos textos informativos o que é importante pode ser um conceito, uma generalização, uma regra. Há uma certa confusão entre assunto e idéia principal. O primeiro pode ser descoberto quando se pergunta de que trata um artigo, e o segundo quando se pergunta sobre qual é a coisa mais importante que o autor nos quer dizer, no texto, esperando-se, aí, obter, como resposta, uma idéia principal.

Questões relevantes são levantadas por Albrechtsen (1993), como: De que estamos falando quando falamos sobre assuntos de livros e outros documentos? Existem diferentes concepções de assuntos e, daí, de análise de assuntos; então, são essas concepções interconectadas com métodos aplicados pela indexação? Em seu artigo, a autora apresenta um modelo alternativo para discutir análise de assunto e indexação, com a intenção de tentar colocar a indexação num contexto social mais amplo, além dos métodos de evolução mecânica, e apontar novos desafios para indexadores.

Alguns autores como Blair, Hjørland, Weinberg e Soërgel apontam novas direções para a indexação, restabelecendo o conceito de *assunto* numa parte

principal da prática e da teoria da indexação. A função principal que a indexação deveria ter é de busca por conhecimento. Recomenda-se que o indexador não focalize exclusivamente o conteúdo de documentos, mas tente antecipar o impacto e o valor de um documento para seu uso potencial (BLAIR, 1990).

Discutindo, em seus estudos, toda a problemática da questão de assunto, Pinto Molina (1995) prefere utilizar o termo *conteúdo*, afirmando que, em espanhol, é difícil estabelecer uma definição clara para *conteúdo* e *matéria*, apontando similar fato no inglês, quando se refere a palavras como *content*, *aboutness* e *subject*. Também no português, enfrenta-se essa problemática, pois análise de conteúdo é também o termo usado para um procedimento metodológico nas pesquisas qualitativas, adotado nas ciências humanas para analisar as comunicações, ou melhor, o conteúdo de mensagens. A principal diferença entre os dois tipos de análises é que a análise documentária tem como objetivo a representação condensada da informação, para consulta e armazenagem, ao passo que a análise de conteúdo tem como objetivo a manipulação de mensagens.

Um fator agravante do problema de nomenclatura adequada para o tratamento da informação é que, devido ao atraso nas pesquisas e à escassez de produção científica nesse campo específico, recorre-se a termos importados que não possuem equivalente no português. Verifica-se também, nesse sentido, a falta de entrosamento entre grupos de pesquisa, que sofrem influências diversas, ora da documentação, da escola francesa, ora da ciência da informação, da escola americana.

O número de termos a serem definidos para representar um documento depende da complexidade do assunto. Pela interdisciplinaridade, ou pela multidisciplinaridade, ou pela transdisciplinaridade, cada vez maior entre os campos do conhecimento, raros são os documentos que tratam de assuntos simples, de uma só classe de uma área, havendo, por isso, uma tendência ao aumento da complexidade dos mesmos. As novidades que se observam em alguns campos de assunto podem, ainda, trazer dificuldades para o indexador que não possui treinamento na área.

Conforme Hagler, citado por Stone (1993), identificação de assunto é algo confuso e freqüentemente indeterminado. Na área de controle bibliográfico, quase tudo parece já poder ser programado no computador. No entanto, essa área ainda deve permanecer como dependente do domínio do julgamento humano. Hagler acrescenta que relatórios de experimentos em indexação e classificação executadas pela máquina indicaram que a intervenção humana é necessária para se atingirem resultados aceitáveis.

Concorda-se com essas idéias e defende-se a necessidade da presença do indexador humano nesse tipo de atividade. Com todos os avanços tecnológicos que vêm sendo realizados na atualidade, acredita-se que a atividade intelectual de definir o assunto de um documento não possa ser feita eficazmente pela máquina.

Outro aspecto a ser considerado é que não se pode definir um assunto sem que se leve em conta o contexto em que está inserido. Os dois conceitos — assunto e contexto — estão sempre juntos nas teorias tradicionais de indexação.

### **5.3 Contexto**

Na determinação do assunto, é preciso que se verifique o contexto no qual o documento é produzido e para o qual ele existe, em determinado momento.

Descrições contextuais típicas, para Blair (1990) são autor(es) e data de publicação, dentre outras. São aspectos observados na leitura técnica que se faz inicialmente, ao analisar um documento. Informação contextual pode também incluir informação sobre o ambiente de recuperação no qual o documento existe (por exemplo, um documento, num sistema de recuperação computadorizado, poderia ter informação contextual indicando quantas vezes ele foi recuperado, a data em que ele foi incluído na base de dados, a data da última vez em que foi solicitado, etc.).

A seguinte questão é levantada por Metcalfe (1977): *Quando um assunto não é um assunto?* E a compara com a questão: *Quando a porta não é uma porta?* E a resposta deve ser: *Quando ela está sozinha, não inserida numa parede.* Refletindo-se sobre isso, pode-se observar realmente a importância do contexto para o significado de um termo, para definir um determinado assunto.

Schustack, Ehrlich e Rayner (1987) citam estudos que mostram que o tempo requerido para responder a uma palavra incluída num texto contínuo está fortemente influenciado pelo contexto que precede essa palavra. O estudo desses autores levanta a questão da falta de clareza, em pesquisas anteriores, quanto à influência dos fatores contextuais que primeiramente identificam a palavra, integram-na na representação mental interna da passagem, ou realizam ambos os processos: são testados diferentes tipos de fatores contextuais, que podem ser levados a produzir seus efeitos facilitadores via diferentes mecanismos. Eles fazem ainda a distinção entre os aspectos *local* ou *lexical* do contexto, e o mais *global*, ou aspecto estrutural do contexto.

No processo de comunicação da informação, que vai desde o emissor (no ponto de vista documentário, o autor) até o receptor (o usuário), a ação de diferentes tipos de contextos retira e altera o significado das mensagens. Essa opinião é defendida por Pinto Molina (1995) e, segundo ela, aspectos situacionais ou contextuais correspondem à perspectiva pragmática que é estabelecida como a dimensão prevalente na pesquisa do texto, de forma que uma teoria textual implica uma teoria do contexto.

Sob o ponto de vista documentário, a autora citada faz uma distinção entre contexto de produção e contexto de reprodução. O núcleo textual é cercado por várias membranas, ou *peles contextuais*, as quais representam os contextos científicos (estabelece a área característica do conhecimento e sub-linguagem), lingüístico (impõe uma característica *straight jacket* — revestimento contínuo — ao qual todos os símbolos usados no texto devem visar) e documentário (sem dúvida, um dos mais importantes e que define a categoria de usuário e sua demanda possível).

Após essas considerações sobre conceito, assunto e contexto, pode-se afirmar que esses são aspectos interdependentes, que ocorrem durante todo o processo de análise de assunto, até o momento de se afirmar sobre o que trata um documento, que é a chamada *Atinência*, tema a ser tratado no próximo capítulo.



## 6 Determinação da atinência

Terminada a fase de extração dos conceitos do texto para análise de assunto, é necessário que se faça uma seleção daqueles que realmente sintetizem o assunto do documento, partindo-se, assim, para a terceira fase do processo de Análise de assunto, que é a determinação da chamada atinência, que envolve ainda o estudo do significado.

### 6.1 Atinência

A atinência é termo traduzido do inglês *aboutness*, e, como tal, carrega toda a problemática da utilização de termos estrangeiros. Para Baranov (1983), *aboutness* pode ser traduzido como “concernência” (do verbo *concernir*; adjetivo *concernente*). Segundo o autor, esse é um termo mais adequado do que o esdrúxulo anglicismo “sobrecidade”. Já Medeiros (1986), em seu estudo sobre terminologia, sugere que, em português, o conceito deva ser representado por “temacidade” (substantivo ligado ao termo *temático*). Como pode ser observado, não há um consenso entre os especialistas da área com relação ao termo mais adequado para traduzir *aboutness*, tendo sido, neste texto, adotado o termo atinência, usado na tradução do livro de Lancaster (1993), por ser o mais utilizado na escassa literatura nacional sobre o tema.

Ainda em 1969, o *Annual Review of Information Science and Technology* (ARIST) publicou uma revisão de literatura de Fairthorne (1969)<sup>13</sup> em que o assunto é abordado (BEGHTOL, 1986). *Aboutness* é visto como uma propriedade de alguns tipos de discurso. O termo vinha sendo estudado há longo tempo por lógicos e lingüistas, pois representa um aspecto do significado. Essa revisão expõe julgamentos sobre *aboutness*, considerados importantes em recuperação de informações.

Um modelo que guia esse processo é proposto por Swift, Winn e Bramer (1978), do *Information Research Group*. Para eles, não é clara a noção de *about* e sugerem o procedimento abaixo:

- Indexadores rotulam e agrupam documentos de acordo com o conteúdo desses;
- Pesquisadores formulam suas questões de acordo com o conteúdo dos documentos;
- Questões relativas à pesquisa são cotejadas com o que os documentos do sistema oferecem. Dessa forma, o indexador objetiva possibilitar ao pesquisador delinear o material que ele busca.

É importante ressaltar que esse modelo é baseado na suposição de que o conteúdo do documento é uma função do próprio documento e também impõe a crença de que o *sobre o que trata o documento* é a base da busca. O artigo discute em profundidade a questão das dificuldades dos diferentes usos do termo *about*, quando na forma de *documentos sobre*, principalmente no campo das ciências sociais. Essas dificuldades se dão por motivos como diferenças de interpretação, existência de várias maneiras de se pensar sobre os tópicos, sobre os argumentos dos autores ao se basearem em pontos de vista diversos, e tornam essa noção de indexação de documentos interpretada como *sobre o que eles tratam* difícil de

---

<sup>13</sup> FAIRTHORNE, Robert A. Content analysis, specification, and control. *Annual Review of Information Science and Technology*. v. 4, p.73-109, 1969.

sustentar. Segundo os autores, é evidente que nessa área não há uma maneira simples de caracterizar documentos e, por esse motivo, propuseram uma *abordagem multi-modal*.

A *abordagem multi-modal* assume uma visão diferente do processo de busca do modelo anterior. A lógica do argumento sugere que a busca em ciências sociais deve ser, necessariamente, um processo *openended*. O procedimento geral é que o indexador tome uma cadeia de pontos de vista claramente definidos, em círculo (por exemplo, orientação teórica, método de pesquisa) e gere uma lista de palavras-chave para cada documento, cada frase tomando como seu próprio foco um aspecto diferente do mesmo, mas referindo-se ao todo. As frases associadas a cada dimensão do sistema são apresentadas como séries separadas (índices separados), cada um deles organizado de acordo com sua própria lista de cabeçalhos, escolhidos pela relevância do ponto de vista empregado.

O modelo acima citado se assemelha à idéia de Cesarino e Pinto (1980), de que há dois momentos em que se faz análise de assunto: o momento em que se analisa o documento e o momento em que se analisam as questões do usuário. Este é um modelo que dá maior ênfase à forma como o usuário vai fazer sua busca de informação.

Uma distinção proposta por Fairthone é mostrada, no artigo de Beghtol, entre atinência extensional e intensional. Extensional, nos termos do primeiro, é o assunto inerente ao documento e intensional é a razão ou propósito pelo qual ele foi adquirido pela biblioteca ou requerido por um usuário. Esses dois tipos de atinência podem ser distinguidos mais acentuadamente contrastando-se atinência (extensional) com significado (intensional) (BEGHTOL, 1986). Beghtol cita ainda Boyce (1982)<sup>14</sup>, que faz uma distinção usando os termos *topicality* (atinência) e *informativeness* (significado). Os termos de Boyce parecem sugerir que buscas bem sucedidas automaticamente fornecem nova informação, quando elas podem, de fatos confirmar informação previamente conhecida do investigador.

## 6.2 Significado

Beghtol (1986) afirma que um texto tem uma atinência relativamente permanente, mas um número variado de significados. Há, portanto, uma forte relação entre a atinência de documento e seu(s) significado(s) identificado(s) pelos indivíduos. Isso varia de acordo com o uso que a pessoa pode encontrar da atinência do documento numa certa época, e o mesmo documento pode ter diferentes significados para o mesmo leitor em diferentes épocas, mas o documento, este, imutável, possui uma atinência fundamental.

Nesta fase da determinação da atinência para representar os conceitos extraídos do texto, inicia-se um processo lingüístico e o problema de descrever documentos para recuperação é, principalmente, o problema de como a linguagem é usada. Assim, qualquer teoria de indexação ou representação de documentos pressupõe uma teoria da linguagem e do significado (BLAIR, 1990).

Na aplicação de considerações sobre palavras e seus significados, Tinker (1966) sugere a expansão da definição de relevância. Afirma que estamos cientes de como uma dada palavra pode ter um certo número de significados, e como o significado específico pode ser aplicado às palavras; isso depende do campo de conhecimento no qual está sendo usado. Propõe a seguinte definição para

---

<sup>14</sup> BOYCE, B. Beyond topicality: a two stage view of relevance and the retrieval process. *Information Processing and Management*, v.18, p.105.109, 1982.

significado: *pode ser definido como a relevância de uma palavra para o conceito que ela rotula*. Acrescenta que a consistência da indexação serve como uma medida da precisão do significado. Na designação de significado estão presentes fatores subjetivos, daí ser difícil o consenso de várias pessoas com referência a um assunto a ser indexado (TINKER, 1966; p. 97).

Considera-se significado como a representação, na linguagem, de um significante, correspondendo o primeiro ao conceito ou à noção, e o segundo à forma. O processo da significação é aquele em que um significante e um significado são correlatos, e o produto é o sinal. A significação impõe uma relação transitória entre expressão (significante) e conteúdo (significado).

O conteúdo é, ao mesmo tempo, conhecimento, informação e significado. Pinto Molina (1994) cita Hirsch (1967)<sup>15</sup>, para quem o *significado é aquilo que é apresentado num texto; é aquilo que representa uma intenção do autor concretizada pelo uso de uma seqüência particular de sinais. É o que estes sinais representam*. Significância, por outro lado, é a relação entre o significado e uma pessoa, ou uma concepção, ou uma situação. O conteúdo (no caso, o conteúdo documentário) é mais próximo ao significado.

Véron (1980)<sup>16</sup>, citado por Ginez de Lara (1993), afirma que *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.

No campo da lingüística, é a semântica a responsável pelo estudo do significado. Como disciplina lingüística, sempre se debateu, de um lado, com as dificuldades de precisão metodológica, de outro, com a identificação, especificamente, do significado, que é ligada a fatores lógicos, antropológicos, sociológicos e psicológicos. A questão do significado, segundo Cintra (1983), é considerada o problema de contorno mais complexo para as linguagens documentárias, no trato com termos da linguagem natural. A redução dos significados, isto é, *a operação de corte num universo praticamente ilimitado* se vale de alguns conceitos correntes em linguagem natural, como polissemia (nome dado à pluralidade de sentidos de uma mesma forma), homonímia (correspondente à igualdade entre significantes diferentes), sinonímia (coincidência de significado entre diversas palavras, sendo a utilização do sinônimo uma grande dificuldade das linguagens de indexação) e antonímia (decorrente das significações contrárias de dois vocábulos, ditos autônomos).

Apesar da complexidade do estudo do significado e da sua relação direta com a atinência, acredita-se que ambos estejam relacionados exatamente ao momento em que o indexador diz: *O texto X trata de tal assunto*. Pode-se afirmar que este é um momento muito importante pois, após uma complexa atividade mental, ele finalmente sente-se apto a definir termos, ainda na linguagem natural, os quais são chamados por Frohmann (1990) de *frases de indexação*. Esses termos devem representar o assunto do texto em análise, para só depois, numa segunda etapa, serem traduzidos para termos de uma linguagem de indexação.

---

<sup>15</sup> HIRSCH, E. D. *Validity in interpretation*. New Haven; London: Yale University Press, 1967.

<sup>16</sup> VÉRON, E. *A produção do sentido*. São Paulo: Cultrix, 1980.

## **7 O caráter interdisciplinar da análise de assunto**

Acredita-se que a análise de assunto seja uma atividade complexa por sofrer interferência de fatores lingüísticos, cognitivos e lógicos, levando o indexador a enfrentar obstáculos ao desenvolver a atividade.

Diante das quatro áreas, ciência da informação, lingüística, ciência cognitiva e lógica, está claro que a maior interessada na interseção é mesmo a ciência da Informação, disciplina nova, ainda sem um corpo teórico consolidado que possa subsidiar qualquer tentativa de resolver problemas que, sabe-se, não são poucos. Haveria interesse, das três outras áreas, em despendar esforços no desenvolvimento dessas pesquisas interdisciplinares? Questiona-se, ainda, qual seria a receptividade de lingüistas, psicólogos e filósofos em trabalhar com profissionais da informação?

Tenta-se, com os aspectos abordados a seguir, dar um suporte teórico, ainda que superficial, para essas discussões, procurando-se, sempre que possível, direcionar as abordagens para o âmbito da análise de assunto ou, pelo menos, para a indexação como um todo.

### **7.1 Aspectos lingüísticos**

Apesar das dificuldades de interseção entre lingüística e indexação existirem desde a primeira vez que alguém indexou um documento, ou seja, traduziu o conteúdo desse documento em palavras que tornassem possível sua recuperação, só recentemente os pesquisadores das duas áreas se preocuparam com o assunto e passaram a adotar procedimentos científicos para os métodos envolvidos na indexação.

Para Navarro (1988), durante muito tempo trabalhou-se na base do bom senso e *somente a partir da necessidade de sistematizar as relações entre a linguagem natural e as linguagens documentárias, principalmente visando-se procedimentos automatizados, a interseção foi percebida e trabalhada a nível teórico.*

Quando analisa um texto com o objetivo de representar seu conteúdo, o indexador lê a linguagem do autor, que é a linguagem natural. Através de processos mentais, tenta extrair os conceitos contidos no texto, para definir os mais adequados e expressar, com suas palavras (também em linguagem natural), o resultado da sua análise. Numa etapa posterior, ele vai traduzir as palavras para os termos de uma linguagem artificial, chamada linguagem de indexação. Essas são, então, as etapas do processo de indexação e pode-se observar que, em todos os momentos, fala-se em linguagens.

Segundo Blair, *o processo de representar documentos para recuperação é fundamentalmente um processo lingüístico, e o problema de descrever documentos para recuperação é, primeiro e principalmente, um problema de como a linguagem é usada.* Assim, qualquer teoria de indexação ou representação de documento pressupõe uma teoria da linguagem e do significado (BLAIR, 1990; p.122).

A lingüística é definida como a ciência que estuda as línguas naturais. A importância da lingüística para a análise documentária não se restringe apenas às suas propostas de análises e, dado que a análise documentária trabalha com a língua e elabora linguagens de documentação, é necessário fazer um recorte teórico da lingüística no que ela apresenta de instrumental, não só na fase de

procedimentos, como na fase de elaboração de instrumentos e produtos a serem obtidos (CUNHA, 1987).

A análise documentária recebe colaboração da lingüística, tanto no reconhecimento de textos, como na estruturação das linguagens documentárias. No entanto, Cunha (1987) afirma que não existem estudos interdisciplinares com objetivos definidos que envolvam a lingüística e a documentação, e diz que a primeira recorta da segunda as propostas que lhe interessam no seu fazer pragmático. Mas não é claro o recorte que se deve fazer para atender ao interesse documentário, apesar de a área ser reconhecida como de apoio aos trabalhos de documentação, de modo especial, aos procedimentos de indexação.

A tentativa de localizar, dentro do âmbito da lingüística geral, aspectos auxiliares à indexação, é feita também no estudo de Cintra, e a autora identifica vários fenômenos lingüísticos que intervêm na conversão da linguagem natural para a linguagem documentária.

Diferentes disciplinas lingüísticas se ocupam do estudo de uma língua, como a fonologia, a sintaxe e a semântica. Para estudos de indexação, importam de modo especial a *semântica*, disciplina que se ocupa do sentido ou da significação dos elementos; a *sintaxe*, que se ocupa das relações que se estabelecem a partir da organização sintagmática dos elementos e, de certo modo, a *morfologia*, que, sintetizando parcialmente aspectos da *semântica* e da *sintaxe*, se encarrega da identificação das partes da palavra e de suas condições de ocorrência (CINTRA, 1983).

A relação entre lingüística e documentação se evidencia na análise do texto, na análise de conteúdo dos documentos (análise documentária). A vizinhança entre lingüística e a documentação, apesar de evidente a partir de um certo nível de abstração, é menos explorada do que se poderia esperar. Smit (1987) acha que, salvo algumas exceções, os trabalhos de síntese ainda não foram efetuados; aponta algumas razões para essa contradição: (1) as pesquisas lingüísticas desenvolveram-se primordialmente, até agora, no campo da sintaxe, cujo interesse é importante mas não suficiente numa perspectiva documentária e (2) a prática da análise documentária baseou-se muito tempo em operações implícitas, circunscritas à 'cultura' individual do indexador (indexação manual). Com a explosão de informações científicas, fez-se sentir a necessidade de uma documentação 'industrial' e automatizada, surgindo o imperativo da explicitação das operações de indexação anteriormente implícitas. A lingüística tornou-se fundamental para uma documentação automatizada.

O estudo das linguagens documentárias ultrapassa o âmbito deste livro, cujo objetivo é a primeira etapa da indexação, que é a extração do conteúdo dos documentos. Diante disso, percebe-se que, dentre as disciplinas da lingüística, a que mais interessa, no caso, é a semântica, por abordar a questão do significado.

Para Artandi *a compreensão do significado vai além do contexto lingüístico fornecido pelo próprio texto, desde que não é um conceito lingüístico, mas sim um conceito psicológico. O significado de mensagens varia com as características dos indivíduos humanos que o constroem e o interpretam* (ARTANDI, 1976; p.239).

Sendo indexação, a condensação (elaboração de resumos) e a busca atividades que envolvem a extração do significado do texto, a mecanização desses processos implicaria em que o computador possa compreender o significado do texto. O estudo do significado contribui para a pesquisa em processamento automático de texto. Com a evolução dessa automatização de textos, feita principalmente em linguagem natural, é que foram iniciadas as pesquisas interdisciplinares, pois a utilização dessa linguagem traz sempre muita

ambigüidade e inconsistência. A contribuição da lingüística tem sido mais observada e desenvolvida, portanto, principalmente nos estudos da indexação automática.

Há considerável consenso entre pesquisadores em ciência da informação de que muito conhecimento necessitado para que se vá além do processamento automático de texto, em geral, e indexação mecânica, em particular, deve vir de estudos lingüísticos, de estudos que ajudam a compreender melhor a organização e a operação da linguagem. Por causa da considerável orientação sintática da teoria lingüística, entretanto, as contribuições da lingüística têm sido um tanto limitadas (ARTANDI, 1976; p. 238).

Para Baranov (1983), algumas das dificuldades surgidas no interrelacionamento da ciência da informação com a lingüística, além do mútuo desconhecimento, se devem a expectativas que a lingüística não pode e talvez não poderá preencher. Ele cita a publicação da Federação Internacional de Documentação (FID), da década de 70 do século passado, como o trabalho mais abrangente sobre lingüística e ciência da informação; *uma das conclusões a que chegaram os autores diz respeito às dificuldades de se casarem as técnicas lingüísticas com os objetivos da recuperação informacional*. Dentro da ciência da informação, o campo mais sensível às perspectivas interdisciplinares é, na opinião do autor, a indexação automática. Vêm sendo realizadas pesquisas que buscam soluções desenvolvidas através do modelo lingüístico.

Dando um enfoque didático à questão dos fenômenos lingüísticos relevantes para a ciência da informação, Baranov acha que, além de introduzir os conceitos gerais vigentes na lingüística, deve-se incluir, em um curso de lingüística para cientistas da informação, noções de sintaxe e de semântica, bem como noções de psicolingüística e de sociolingüística, sempre com vistas às eventuais interfaces com a ciência da Informação.

Baranov cita um trabalho de Basilio<sup>17</sup>, que se dedica à morfologia, mais especializada nas potencialidades de análise das estruturas lexicais com vistas à ciência da informação, mostrando como o conhecimento de padrões morfológicos na formação de palavras pode ser relevante para a indexação. A moderna abordagem morfológica, utilizada no trabalho, cita importantes aplicações, tanto na indexação automática, quanto na manual, e abre uma perspectiva de cooperação entre lingüistas e cientistas da informação.

Um dos fatores de distorção considerados no trabalho de indexação é a limitação lingüística, encontrada principalmente em indexação na área de humanidades, mais do que na técnica. Conforme opinião de Bell (1991), trata-se da questão da terminologia disponível para descrever a vida humana e suas relações: as limitações de idéias que podem ser expressas, particularmente, a lentidão com que a linguagem acompanha as mudanças que ocorrem na sociedade.

As bases lingüísticas e filosóficas para a organização do conhecimento estão sempre presentes nos trabalhos de Dahlberg. Ela afirma que, junto aos estudos das bases filosóficas, quando se considera a lógica, a teoria da ciência, a epistemologia, a ontologia, a fenomenologia, a aletologia e a metafísica, nenhum conteúdo dessas disciplinas pode ser comunicado sem um portador, que seria a linguagem (palavras e termos). Quando se lida com a geração e a produção de coisas novas, essas precisam ser nomeadas, tendo-se de criar uma forma verbal para os conceitos, quando então surge a linguagem. E, para cada campo de

---

<sup>17</sup> BASÍLIO, M. deP. Interface Lingüística e ciência da informação: potencialidades na análise de estruturas lexicais. In: REUNIAO BRASILEIRA DE CIENCIA DA INFORMACAO; IBICT. **Anais...** Rio de Janeiro: IBICT 1978. vi, p233-246.

especialidade, surgem os termos que estabelecem a terminologia da área (DAHLBERG, 1992).

Smit (1987) relata que as preocupações acerca da explicitação do processo de análise documentária (e as questões lingüísticas aí envolvidas) datam da década de 70 do século passado. Para ela, as discussões surgidas pela automação da problemática da análise documentária não se resumia a questões de sintaxe e semântica, uma vez que, ao ultrapassar o limite da frase e passando a se preocupar com o texto, um outro componente intervinha igualmente: a Lógica, que será abordada a seguir.

## **7.2 Aspectos cognitivos e lógicos**

Acredita-se que, junto aos aspectos lingüísticos, os aspectos cognitivos e os lógicos formam o tripé de sustentação do processo de análise de assunto. Eles interferem e interagem simultaneamente, não havendo possibilidade de delimitar em que ponto começa ou acaba a influência de cada um dos três.

Para entender como ocorrem os processos cognitivos, é importante conhecer alguns conceitos básicos, como ciência cognitiva, psicologia cognitiva, e os processos cognitivos propriamente ditos.

Segundo Daniels, a ciência cognitiva consiste *em disciplinas como psicologia cognitiva, lingüística, inteligência artificial, filosofia, educação e, idealmente, ciência da informação*. Essa conceituação, por si só, já mostra a interdisciplinaridade também presente nesse campo de estudo, o que torna quase impossível o estudo de uma disciplina separada- mente (DANIELS, 1986; p.272).

A psicologia cognitiva é uma ciência empírica baseada na realidade do homem como um ser comunicativo. É uma área relativamente nova, que lida com os processos e estruturas mentais implicadas na aquisição, processamento e uso de conhecimento ou informação, entre os quais podem ser diferenciados os processos mentais básicos (memória e atenção), as representações mentais (imaginação, formulação de proposições e estabelecimento de categorias) e os processos mentais complexos (compreensão, raciocínio e solução de problema) (PINTO MOLINA, 1994; p.124).

*Os processos cognitivos são atividades mentais como pensamento, imaginação, lembrança e solução de problemas* (ALLEN, 1991; p.13). Como outras atividades humanas, essas são desempenhadas diferentemente por indivíduos que têm diferentes níveis de habilidade em raciocínio lógico, memória visual ou experiência em vocabulário, o que pode afetar o desempenho da recuperação da informação.

Para analisar os processos cognitivos da geração e compreensão dos sistemas de linguagem (textos ou discursos), Beghtol (1986) considera que os textos lingüísticos tomam emprestado, da psicologia Cognitiva, os conceitos de dois tipos de processamento mental da informação: *top-down* e *bottom-up*, já citados no cap. 2 deste livro.

Pesquisas empíricas mostram que esses dois tipos de processamento mental parecem ocorrer, contínua e simultaneamente, durante a compreensão do texto pelo leitor.

A compreensão de um texto implica na identificação dos conceitos essenciais nele contidos, e essa compreensão necessita da interação de processos intelectuais complexos, que envolvem a memória do indivíduo e a ativação das estruturas cognitivas.

A representação condensada de um texto é uma atividade intelectual e implica na identificação das palavras e na compreensão do léxico utilizado, além da elaboração das representações cognitivas percebidas, sua colocação na memória e sua integração aos conhecimentos anteriores. Há um consenso de que, para ler e compreender um texto, para conservar e recuperar na memória as informações, o leitor deve possuir conhecimentos prévios.

A essência da abordagem cognitiva, conforme Belkin (1990), é a idéia da percepção humana, da cognição e das estruturas do conhecimento. Cita De Mey (1977)<sup>18</sup> que no estudo da cognição, afirma que qualquer processamento de informação, se perceptivo ou simbólico, é mediado por um sistema de categorias ou conceitos os quais, para o mecanismo do processamento de informação, são um modelo de seu mundo.

Para a ciência da informação, ter uma visão cognitiva tem tipicamente significado considerar seu escopo como sendo concernente com o sistema de comunicação humana, no qual textos têm um papel chave, bem como a interação do indivíduo com esses textos.

O interesse da psicologia para a documentação tem seus limites e há uma lacuna entre ciência cognitiva e ciência da informação. Esse é o ponto de vista de Garcia Marco e Esteban Navarro (1993). A cognição humana é também um processo social que requer especialização de funções e comunicação. Somente no sentido sociológico a informação documentária pode ser falada como sendo conhecimento, como parte da herança cultural da humanidade. Os autores afirmam que *como os processos de informação social são de fato mediados pelos processos psicológicos, parece claro que as descobertas da psicologia cognitiva são intensivamente relevantes para ciência da informação.*

Eles acreditam, ainda, que a psicologia cognitiva é interessante para cientistas da informação por duas razões principais: (a) ciência da informação e psicologia cognitiva são, ambas, ciências cognitivas num sentido amplo; ambas são interessadas na maneira como informação produz conhecimento, como informação é processada e como é realizada a melhor adaptação da realidade; e (b) processos psicológicos mediam o ciclo da informação. Isso acontece, principalmente, em todos os tipos de atividades de interface entre homens, e entre homens e máquinas.

Conforme Alien (1991), desde 1977, pesquisadores em ciência da informação têm adotado, em seus estudos, o foco dos sistemas conceituais de usuários da informação e, como resultado, registra-se o aumento do número de projetos de pesquisa que são identificados como parte da perspectiva cognitiva. A pesquisa cognitiva em ciência da informação deriva de metodologias de estruturas explanatórias desenvolvidas em ciência cognitiva.

Outro trabalho que enfoca as investigações em aspectos cognitivos da recuperação da informação é o de Ingwersen (1982). O objetivo da pesquisa foi o de obter informação detalhada sobre processos de transferência de informação em bibliotecas públicas: interação do usuário com a organização de documentos, processo de negociação entre usuário e bibliotecário e procedimentos de busca do bibliotecário (com padrões de busca e processos de pensamento). Apesar de os experimentos envolverem somente bibliotecas públicas e usuários em geral, o autor reconhece serem os resultados úteis e relevantes para outros setores da informação e documento. No caso do processo de análise de assunto, foram extraídas idéias das considerações e abordagem teórica da referida pesquisa.

---

<sup>18</sup> DE MEY, M. The cognitive viewpoint: its development and its scope. In: *International Workshop on the cognitive viewpoint*. Ghent: University 01 Ghent, 1977.



O autor acima cita De Mey, que afirma: *qualquer processo de informação, seja perceptual ou simbólico, é mediado por um sistema de categorias ou conceitos os quais — para o mecanismo do processamento de informação — são um modelo de seu mundo*, seja o mecanismo humano ou mecânico. Esse modelo de mundo é determinado por experiências individual, social/coletiva, educativa, de treinamento, etc. Para o autor, a tarefa da recuperação da informação é trazer estruturas cognitivas de autores, analistas de sistemas e indexadores, de acordo com a necessidade real. Mecanismos cognitivos coletivos, freqüentemente descritos em teorias de paradigmas, também influenciam a estrutura dos sistemas de classificação e indexação e, assim, podem ter implicações para a relação de tópicos e conceitos tratados no corpo da literatura e necessidades de informação.

A compreensão do texto é um ramo do processamento de informação humano. Farrow (1995) cita Todd que escreveu, em 1970, sobre a necessidade urgente da profissão em desenvolver um conhecimento base teórico para indexação. Segundo Todd, *a emergência, na metade dos anos 70, da psicologia cognitiva como uma ciência distinta, levou ao desenvolvimento de um corpo de pesquisa que pode ser aplicado para auxiliar nosso entendimento do processo de indexação* (TODD, 1992; p.243).

Segundo Luria (1994), o problema central da psicologia da compreensão de texto é a assimilação, não somente dos significados superficiais e de suas combinações, mas, também, da assimilação do sistema interno, profundo, de subtextos ou sentido. Precisamente, este aspecto é o mais complexo e importante problema da psicologia da compreensão de textos. Esse autor afirma que todos esses problemas constituem o conteúdo fundamental de uma nova ciência, que se encontra nos limites entre a psicologia e a lingüística, e que recebeu o nome de psicolingüística.

Os processos de produção e compreensão são similares no trabalho de indexadores, classificadores, elaboradores de resumos (*abstractors*), de condensações e sinopses: o processo comum entre eles é a análise de conteúdo e o propósito, comum, de representação do documento. Há, entretanto, segundo Farrow (1995), algumas diferenças quanto aos objetivos a serem alcançados: para o *classificador*, o objetivo é a representação da atinência do documento, traduzida para a linguagem do sistema de classificação; para o *elaborador de resumos*, o propósito é representar em prosa contínua os argumentos principais do documento; para o *indexador*, há dois tipos: (a) para o indexador de livros, a tarefa é ler o texto, distinguindo entre informações relevantes e periféricas, e empregando uma mistura do processamento *top-down* (conceitual) e *bottom-up* (perceptivo) obtidos na leitura fluente normal; e (b) para o indexador acadêmico, a indexação é menos exaustiva e usa, predominantemente, a abordagem *top down*.

Farrow (1995) acredita que as teorias de Van Dijk e Kintsch têm uma atração imediata para indexadores: indexação é a redução do texto e se tem, aqui, uma explanação teórica desse processo. O discurso de Van Dijk oferece uma descrição teórica da operação cognitiva da compreensão de assunto de um texto.

Com o objetivo de entender o baixo nível de consistência entre indexadores, David e outros autores consideraram necessário estudar os processos subjacentes à atividade de indexação e, dada que esta atividade é intelectual, é preciso apropriar-se de teorias e métodos da psicologia cognitiva, e da teoria de solução de problemas, em particular. Antes de ver indexação como uma situação cognitiva de rotina ou uma compreensão de texto, eles a consideraram como uma atividade de solução de problema (DAVID *et al.*, 1995).

Além dos aspectos cognitivos, o uso da linguagem cria vários problemas do ponto de vista lógico, como, por exemplo, a homonímia (várias coisas com o mesmo nome), a sinonímia (vários nomes para a mesma coisa), e recorre-se à lógica para resolver esses tipos de obstáculos.

A lógica é um ramo da filosofia que permite uma análise do pensamento. Consiste numa operação mental que possibilita, através do raciocínio, o surgimento de novas proposições através de proposições já existentes.

Nos processos de análise textual, dois tipos de raciocínio, ou de argumentos lógicos, são usados: *dedutivo* (silogismo aristotélico) e *indutivo* (inferência probabilística). O dedutivo ocorre quando não se aumenta o conteúdo factual das proposições básicas (premissas). A dedução não é ampliativa; sua conclusão, ao contrário, é compulsória (não comporta duas saídas) e, como tal, está contida inteira nas premissas. Se essas estiverem certas, a conclusão estará igualmente certa. O risco está em que uma das premissas seja falsa. O indutivo ocorre quando se aumenta o conteúdo factual das proposições que possibilitam o argumento. Isto significa que, com a indução, obtemos uma conclusão que diz mais do que as proposições iniciais. A indução é ampliativa.

A lógica formal, assim chamada pelo fato de que lida exclusivamente com a forma (significante) dos enunciados, é a variação da lógica geral e, como o nome indica, limita-se a um sistema de símbolos, relacionando o exercício do raciocínio lógico a um cálculo algébrico.

A importância da lógica formal para a análise de assunto vem do fato de que símbolos lógicos, ao contrário dos lingüísticos, têm um significado perfeitamente exato. Uma das atividades do cientista é o uso de raciocínios; como estes para se explicitarem empregam a linguagem, são ditos discursivos. Diante de uma série de fatos, o cientista muitas vezes acredita poder obter conclusões através de argumentos em que usa palavras e expressões como ora, portanto, por conseguinte, etc. Quando entra nesse processo, está raciocinando sobre os fatos que analisou e, desta forma, procurando tirar conclusões (inferindo) a partir deles.

Conforme Freire-Maia, toda a lógica clássica se baseia sob dois princípios: o princípio da identidade (no caso, *duas noções iguais a uma terceira são iguais entre si*) e o princípio *dicci de omni, dici de nullo* [ser dito de todos (os casos implicados), ser dito de nenhum]. Esse segundo princípio significa que o que se diz de todos (por exemplo, todo Homem é mortal) corresponde a que o inverso pode ser dito de nenhum (nenhum homem é imortal) (FREIRE-MAIA, 1995; p39).

Dentro da filosofia, a lógica é considerada uma ciência normativa. Santaella (1992) faz um estudo da classificação das ciências feita por Peirce. Para entender sua concepção do caráter normativo da lógica, é feita uma diferenciação entre lógica *utens* e lógica *docens*. Segundo o autor, Peirce achava que qualquer pessoa, na vida comum, tem um instinto para o raciocínio ou um hábito para o raciocínio, com os quais forma sua opinião relativa a muitos assuntos de grande importância, conhecida como lógica implícita do homem comum: é a lógica *utens*. Por outro lado, quando o homem se defronta com fatos surpreendentes, não usuais, que exigem capacidade inventiva, generalização, teoria, essa primeira lógica citada não é suficiente. Nesse momento, são exigidos estudos dos processos de raciocínio e a investigação de métodos, que dêem mais confiança e apressem o avanço do conhecimento para os resultados desejados. Essa é a lógica *docens*.

O conselho de Peirce é *apoiarmos nossa conduta tanto quanto possível no instinto, mas quando, diante de uma situação complicada, tivermos de raciocinar, que o façamos com uma lógica severamente científica* (SANTAELLA, 1992; p.124).

Diante da análise da concepção de lógica de Peirce, torna-se evidente que o esforço necessário para a análise de assunto está relacionado à lógica *docens*, pela exigência da capacidade inventiva, generalização e teoria, por parte do indexador, muito além da lógica do bom senso do homem comum.

Avançando nos estudos da lógica, Peirce fez, por volta de 1900, uma diferenciação entre dois sentidos desta área: (a) no sentido mais estreito, é a ciência das condições necessárias para se atingir a verdade e (b) no sentido mais amplo, é a ciência das leis necessárias do pensamento, ou melhor (o pensamento sempre ocorrendo por meio de signos) é *semiótica geral*, que trata da verdade, das condições gerais dos signos. das leis de evolução do pensamento, que coincidem com o estudo das condições necessárias para a transmissão de significado de uma mente à outra, e de um estado mental a outro.

Dando continuidade à análise feita por Santaella, tem-se que a lógica, no sentido mais amplo, é considerada semiótica, uma ciência que lida não só com símbolos, mas também com todas as espécies de representações. Fazem parte do estudo da semiótica todas e quaisquer formas de linguagem, todos os signos de todas as coisas.

A semiótica geral, como um todo, é dividida em três partes: (a) gramática especulativa, que tem por função estudar a fisiologia dos signos de todos os tipos; (b) lógica crítica, ou lógica propriamente dita, primariamente desenvolvida como uma teoria unificada da abdução, dedução e indução; (c) metodêutica ou retórica especulativa, que está fundada na crítica, estudando as condições gerais da relação dos símbolos e outros símbolos com seus interpretantes.

Não se pretende entrar no debate filosófico e trabalhar as diferentes abordagens a respeito da lógica. A opinião de Cunha (1987) é de que a lógica, enquanto ciência, encontra-se bem distante das preocupações dos bibliotecários/analistas da documentação, mesmo quando se utiliza automação. Do mesmo modo que a análise documentária utiliza a lingüística, enquanto instrumento teórico para a sua atividade pragmática, a lógica interessa à análise documentária enquanto parâmetros de procedimentos, mais como aparatos formais que como metodologia ou epistemologia da ciência do discurso. Para a autora, deve-se considerar tanto a lógica geral quanto a lógica formal. Na primeira, propõe-se examinar no texto os processos de aquisição dos conhecimentos científicos, assim como a construção dos princípios, das hipóteses, leis gerais e teorias. Na segunda, o interesse é determinar como se elaboram os raciocínios do ponto de vista formal.

A lógica formal é a ciência que aumenta a possibilidade de aplicação com respeito à análise de assunto, desde que ela possibilita operar logicamente com diferentes unidades de significação.

Verifica-se que a lingüística, a psicologia e a lógica agem separadamente, mas em colaboração mútua. É o caso do conhecimento humano materializado em textos científicos e o *mundo* dos fatos lógicos (conceitos universais) arranjados numa estrutura lógica. Conclui-se, assim, a interferência direta dessas áreas durante o processo de extração do assunto de documentos.

\* \* \*

Com o presente capítulo, dá-se por encerrada a descrição do processo de análise de assunto. O capítulo final, a seguir, traça um panorama da pesquisa sobre análise de assunto, com destaque especial para a realidade brasileira.

## **8 A pesquisa em análise de assunto**

Conforme o leitor pode observar, muitas das informações apresentadas nos capítulos anteriores deste livro resultaram, de forma direta ou indireta, de pesquisas realizadas na área do tratamento da informação e, mais especificamente, na área de análise de assunto. Por isso, na formação dos profissionais que vão atuar como analistas de assunto, é importante não apenas dotá-los das informações necessárias no trabalho prático do dia-a-dia, mas também sensibilizá-los para a importância da atividade de pesquisa. É no bojo dessa atividade que serão geradas as informações que poderão trazer avanços para o trabalho de análise de assunto, seja em seu aspecto de técnica utilizada nas bibliotecas e nos sistemas de recuperação da informação, seja como uma área de estudo, ensino e pesquisa. Por conseguinte, se não se faz pesquisa numa determinada área ou tema, ou se está usando conhecimento defasado ou conhecimento produzido por meios não científicos.

De um modo geral, os alunos se sentem mais atraídos pela prática de uma profissão, rejeitando ou repelindo as questões teóricas. Isso seria, inclusive, especialmente verdadeiro em relação aos bibliotecários, conforme observou Butler (1971). Entretanto, como é o entendimento de vários especialistas, a pesquisa científica rigorosa visa principalmente o desenvolvimento da teoria de um determinado campo do conhecimento. Assim, devemos concluir que o trabalho de motivar os alunos e profissionais para a pesquisa será uma tarefa ainda mais difícil entre os alunos e profissionais de biblioteconomia e ciência da informação. Este capítulo, que procura traçar um panorama geral da pesquisa em análise de assunto, com destaque especial para aquela que tem sido feita no Brasil, tem como um de seus principais objetivos sensibilizar os leitores para a importância da pesquisa em análise de assunto.

Para isso, vamos discutir os temas que vêm sendo enfocados nessa pesquisa, assim como aqueles que poderiam ou deveriam estar sendo igualmente contemplados. Vamos falar também das questões metodológicas, tão importantes quando se trata de pesquisa, porque as estratégias utilizadas do ponto de vista metodológico muitas vezes podem significar a diferença nos resultados obtidos. Finalmente, vamos falar das tendências da pesquisa na área.

### **8.1 Os temas**

Os temas abordados na pesquisa em análise de assunto estão vinculados às principais atividades realizadas na prática dessa etapa do tratamento temático da informação. Podem também se voltar para o processo como um todo, limitando-se ou não a um determinado contexto de informação (geral ou especializado) ou ainda a determinadas categorias de materiais. Assim, por exemplo, é natural que se investigue a existência de peculiaridades do processo como um todo numa biblioteca pública, em contraste com o que ocorre, digamos, numa biblioteca especializada. Ou então que se parta do pressuposto de que a análise de assunto de material áudio-visual implique em problemas que a análise de assunto do material textual talvez não enfrente.

Se for tomado o processo como um todo, independentemente de fatores contextuais ou de suportes de informação, geralmente pode-se encontrar os temas localizados em três grandes áreas;

1. Leitura e compreensão do texto

2. Identificação dos conceitos
3. Seleção dos conceitos

A leitura do texto, em análise de assunto, como foi dito anteriormente, é realizada tradicionalmente por meio da chamada leitura técnica. Entretanto, essa é uma estratégia que foi desenvolvida de forma intuitiva pelos bibliotecários, assim como tantas outras técnicas, métodos e processos utilizados nas bibliotecas antes do desenvolvimento da pesquisa científica no campo da biblioteconomia/ciência da informação. Por conseguinte, é de se esperar que outras alternativas possam ser pensadas e sugeridas. Hjørland (1997; p. 58), por exemplo, comentando sobre os quatro métodos propostos por Wilson (1968) para análise de assunto, relaciona esses quatro métodos com quatro diferentes abordagens teórico-filosóficas:

- a. O primeiro tem uma relação com a teoria da hermenêutica clássica, que analisa o texto principalmente por meio do estudo das intenções, personalidade e biografia do autor;
- b. O segundo tem ligações com as modernas abordagens de natureza psicológica, cognitiva e de orientação para o usuário;
- c. O terceiro é moldado a partir das abordagens de análise de textos de natureza positivista, bibliométrica e estatística; e
- d. O quarto tem base nas abordagens de análise de documentos de natureza lingüística.

E conclui HJØRLAND: “Métodos para a determinação do assunto de um documento (ou algum item lingüístico oral ou escrito, assim como objetos informacionais) são, portanto, intimamente conectados com teorias do significado e da interpretação.” Isso demonstra possibilidades muito amplas de investigação de problemas de leitura na análise de assunto. Fujita (1999) lembra algumas questões específicas que podem ser suscitadas nas pesquisas sobre a leitura do indexador:

- Como é o processo de leitura do indexador?
- A leitura do indexador é igual à do leitor comum?
- O indexador faz a leitura de todos os documentos de maneira idêntica, ou muda os procedimentos de acordo com o assunto?
- A leitura para a indexação de livros é igual à leitura de artigos de periódicos?

De qualquer forma, está claro que a leitura — que é preocupação de várias áreas — tem um papel central nesta etapa da análise de assunto. Isso significa que questões de leitura identificadas como importantes em outros campos do conhecimento — em psicologia, em comunicação, em letras — podem muito bem ser candidatas a serem estudadas do ponto de vista da ciência da informação.

Um exemplo de pesquisa de compreensão textual é a de Farrow (1995), que buscou conhecer o modelo situacional do indexador, ou seja, compreender como o indexador procede na tarefa de leitura em sua atividade profissional.

Como observamos acima, a análise de assunto tanto pode ser estudada de forma geral, sem referência a uma área específica do conhecimento, como limitada a uma determinada área. No que diz respeito a estudos restritos a determinadas áreas, exemplos são os estudos de Guimarães (1994; 1997; 2000), que investiga os problemas na área do direito do trabalho.

A identificação de conceitos é uma etapa que, embora logicamente bem próxima da atividade de leitura, tem, entretanto, características próprias. A habilidade principal exigida nesta atividade é a de síntese. Com efeito, a análise de assunto caminha de um texto que pode ter centenas de páginas rumo a uma frase de indexação ou enunciado temático que pode ser expresso por apenas uma palavra. Assim, um livro como *Introdução à administração*, de Peter Drucker, tratando, como trata, de todos os aspectos da administração, só pode ser descrito, tematicamente, como tratando de... administração. Ou seja, o livro que em sua edição brasileira de 1991, pela editora Pioneira, tem 713 páginas, tem todo esse vasto conteúdo textual reduzido a uma única palavra, para efeitos de descrição temática e posterior recuperação desse documento.

Entretanto, esse é um exemplo extremo de enunciado temático, o que se reduz a apenas uma palavra. Em muitos outros casos vai acontecer do conteúdo ser expresso por mais de uma palavra, ou por frases. O trabalho de identificação de conceitos tem uma proximidade grande com um outro trabalho de redução de conteúdos documentais, que é o trabalho de elaboração de resumos e de sumarização. O primeiro, da própria esfera da ciência da informação e o outro, muito comum nas áreas de educação e de letras. Por conseguinte, vale aqui a observação feita anteriormente, de que nessas áreas externas à ciência da informação em que se estuda a condensação de textos e de outros tipos de informação, os estudiosos de análise de assunto podem se inspirar quanto a temas importantes para investigação.

Finalmente, a etapa em que os conceitos identificados são selecionados para efetiva representação no sistema de informação exige do analista de assunto ainda outros tipos de conhecimentos. Os dois mais importantes são o conhecimento dos usuários e o conhecimento da coleção. Como esses são conhecimentos necessários em diversas outras áreas das unidades de informação, certamente que um aspecto importante a ser investigado diz respeito ao que especificamente interessa ao analista de assunto nesses dois tópicos.

## **8.2 Estratégias metodológicas**

Há uma tendência a se entender metodologia apenas no seu aspecto técnico de condução de uma pesquisa, ou seja, em todas as etapas de sua realização, da formulação do problema até o seu final, com a divulgação dos resultados. Com uma ênfase especial na etapa de coleta de dados: definição de estratégias, elaboração de instrumentos etc. Entretanto, o termo é utilizado também no sentido de concepção genérica, teórica, filosófica, ideológica do que seja pesquisa. Por conseguinte, quando alguém desenvolve uma pesquisa é inevitável que na concepção do respectivo projeto e na sua implementação estejam embutidas idéias e crenças de natureza teórica, filosófica e ideológica que constituem a visão de mundo do pesquisador.

As metodologias utilizadas na pesquisa em análise de assunto enfrentam, de modo geral, os mesmos problemas das demais áreas da ciência da informação. Há um consenso de que esta pertence ao domínio das ciências sociais. Com isso, começam as dificuldades, já que as ciências sociais vêm utilizando os modelos metodológicos das ciências exatas, levantando questionamentos quanto à propriedade dessa utilização. Entretanto, o que se tem observado na prática é certa atitude de pragmatismo, ou seja, continuar a fazer pesquisa apesar dos mencionados questionamentos. Inclusive porque já existe hoje um corpus de

resultados científicos nas ciências humanas e sociais que não é desprezível, em termos do seu impacto.

Uma das principais estratégias metodológicas que têm sido utilizadas é a de investigar o trabalho realizado por indexadores. Exemplo de pesquisa nessa linha é a de Bertrand e Cellier (1995), que investigaram a habilidade do indexador com diferentes níveis de experiência. Eles concluíram que indexadores com mais prática utilizam estratégias diferenciadas, tomando a leitura mais consistente. Descobriram também que a perícia no uso da linguagem de indexação e a familiaridade no domínio do assunto influenciam igualmente o trabalho dos indexadores experientes, se comparados com os iniciantes.

Fujita e Silva (2000) estudaram o trabalho de indexadores da Sub- Rede Nacional de Informação em Ciências da Saúde. Inicialmente, as pesquisadoras distribuíram um formulário de identificação onde os pesquisados deveriam fornecer informações sobre sua formação profissional e relatar os procedimentos adotados na leitura documentária, inclusive dificuldades enfrentadas. Em seguida, dos nove indexadores, quatro foram observados enquanto trabalhavam, utilizando-se do chamado protocolo verbal, ou pensar alto (*think aloud*). Trata-se de uma técnica de coleta de dados em estudos de observação, que se destaca por incluir dados gerados a partir de manifestações espontâneas do sujeito, expressadas enquanto realiza seu trabalho. Outras pesquisas recentes, também em análise de assunto e que também utilizaram a técnica do *think aloud* são as de Naves (2000) e Neves (2004).

Um outro tipo de estratégia privilegia a visão do usuário. Neste caso, parte-se do pressuposto de que a análise de assunto visa, em última instância, fazer uma leitura em nome do usuário, sendo importante, por conseguinte, procurar conhecer os usuários dessa perspectiva. O estudo de Dias, Naves e Moura (2001) parte da perspectiva do usuário, visto em seu contexto de uso e busca da informação. O pressuposto é o de que as características e os comportamentos associados com essas atividades podem ter grande importância para a análise de assunto. Estudos de usuários têm sido feitos em grande quantidade, mas nem sempre de uma perspectiva que possa redundar em resultados proveitosos para os serviços ou setores específicos de uma biblioteca ou sistema de recuperação de informação.

O estudo de Dias (2004) teve por objetivo verificar a percepção de usuários sobre o conteúdo de textos com que estão familiarizados. O delineamento do estudo incluía a previsão de que, passado algum tempo, se voltasse aos respondentes com a mesma questão original. Os resultados mostram certas coincidências e/ou aproximações entre os termos utilizados nos dois momentos da pesquisa, bem como alguma discrepância.

Um estudo que procura combinar as duas perspectivas, a do usuário e a do profissional da informação, é o de Saarti (2000), cujo objetivo era verificar como o conteúdo de obras de ficção é sumarizado. Para isso, o autor do estudo pesquisou tanto usuários quanto bibliotecários de uma biblioteca pública na Finlândia. A estratégia metodológica de se combinar grupos ou segmentos sob análise, ou diferentes métodos de coletas de dados, é muito conhecida e muito utilizada em várias áreas do conhecimento. Baseia-se na lógica de que estratégias isoladas sempre têm deficiências que podem ser compensadas pelo uso de outras estratégias, paralelamente.

Uma última estratégia que pode ser mencionada, por fim, é a de se investigarem os autores dos textos indexados. Aqui, como se pode perceber, a lógica é a de que ninguém melhor que os próprios autores para saber do que



trata um determinado documento. É claro que alguns vão discordar da base de sustentação desse tipo de abordagem, principalmente aqueles que entendem que um texto, de certa forma, se torna independente do seu autor, depois que se torna público. Desse ponto de vista, os leitores podem dar a um texto significados que o próprio autor não previu.

Exemplo de estudo nessa linha é o de Bramrn e Bruil (1992), que investigaram autores indexados no *Chemical Abstracts* online. O estudo utilizou 270 artigos de periódicos originários de 54 periódicos da área de química. Os autores eram solicitados a opinar sobre os descritores utilizados pelos indexadores.

### 8.3 A pesquisa no Brasil

Como se sabe, a maior parte da pesquisa que se faz no Brasil, em qualquer área do conhecimento, é produzida nas universidades, principalmente as públicas, isso é especialmente verdadeiro no caso da ciência da informação, por não ter esta muita visibilidade como área de grande impacto social ou econômico que pudesse justificar o interesse do setor privado, ou mesmo do investimento público de fomento. Como a análise de assunto é uma especialidade da ciência da informação, é natural que a pesquisa nessa área seja desenvolvida nas escolas ou departamentos de biblioteconomia e/ou ciência da informação. Embora nada impeça que essa pesquisa possa se desenvolver em escolas ou departamentos que oferecem apenas cursos de graduação em biblioteconomia e/ou ciência da informação, entretanto, o ambiente mais propício e mais favorável para o desenvolvimento de pesquisas é o ambiente da pós-graduação, onde a pesquisa tem papel fundamental.

No Brasil, atualmente são oito os programas<sup>19</sup> de pós-graduação em ciência da informação<sup>20</sup>, vinculados às seguintes universidades:

Pontifícia Universidade Católica de Campinas (PUCAMP); Universidade de Brasília (UnB); Universidade de São Paulo (USP); Universidade Estadual Paulista, campus de Marília (UNESP/Marília); Universidade Federal da Bahia (UFBA); Universidade Federal de Minas Gerais (UFMG); Universidade Federal de Santa Catarina (UFSC); e Universidade Federal Fluminense (UFF), esta em convênio com o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). A esse conjunto deve-se acrescentar, como *locus* significativo de formação e pesquisa em ciência da informação, a linha de pesquisa em ciência da informação existente no programa de pós-graduação em comunicação da Universidade Federal do Rio Grande do Sul (UFRGS).

Entretanto, isso não quer dizer que em todos esses programas haja pesquisa em análise de assunto. Deve-se fazer uma distinção entre iniciativas individuais de pesquisa, e a pesquisa que se faz em tomo de uma *linha de pesquisa*, conceito fundamental da pós-graduação. Assim, embora possa haver pesquisas individuais, isoladas, em várias universidades, o leitor interessado na pesquisa numa determinada especialidade da ciência da informação deve, primeiro, procurar identificar aqueles programas de pós-graduação onde a especialidade seja objeto de uma linha de pesquisa. Isso significa que num tal

---

<sup>19</sup> Dá-se o nome de programa à atividade de pós-graduação realizada numa determinada universidade, podendo consistir de cursos de especialização, mestrado e doutorado, além de outras modalidades menos comuns (aperfeiçoamento, pós-doutorado etc.).

<sup>20</sup> Ciência da informação é hoje terminologia oficial tanto do CNPq quanto da CAPES para designar uma subárea do conhecimento localizada dentro das ciências sociais aplicadas. Nada impede, entretanto, que um determinado curso use o termo biblioteconomia ou qualquer outro dos termos existentes no campo do conhecimento (arquivologia, por exemplo) para designar a sub-área.

lugar a pesquisa sobre aquela especialidade terá um volume e uma qualidade provavelmente muito maiores que a possível de se conseguir com iniciativas isoladas.

As linhas de pesquisa de um determinado programa de pós-graduação podem sofrer mudanças, de tempos em tempos, pois dependem basicamente de haver professores e pesquisadores da instituição interessados no tema. Atualmente, os programas que têm linha de pesquisa onde é dado destaque à análise de assunto são os da UFMG, UNESP/ Marília e USP.

Na pós-graduação da USP, o tópico vem tendo destaque desde a criação da linha de pesquisa Análise Documentária, cuja atuação resultou em várias publicações, teses e dissertações. Pode-se citar, como exemplo, os trabalhos de Smit (1987) e de Kobashi (1994). No Programa de Pós-Graduação em Ciência da Informação da UFMG, a análise de assunto é destaque na linha de pesquisa Organização e Uso da Informação, que incorpora a antiga linha Tratamento da Informação e Bibliometria. Dentre as publicações, teses e dissertações produzidas na UFMG, pode-se citar: Naves (2000), Dias (2004), Neves (2004) e Silva (2004). No Programa da UNESP/Marília, de mais recente implantação, os estudos de análise de assunto estão compreendidos dentro da linha de pesquisa Tratamento da Informação, com várias pesquisas e dissertações. Entre estas últimas, podem ser citadas as de Fagundes (2001), Souza (2002) e Silva (2004); entre as pesquisas, os trabalhos já mencionados de Fujita (1999) e de Guimarães (1994).

Até a criação da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação e Biblioteconomia (ANCIB), a eventual pesquisa em análise de assunto deveria aparecer de forma dispersa nos congressos e reuniões científicas da área, assim como nos periódicos, principalmente aqueles que durante as décadas de 70, 80 e 90 do século passado veicularam a maioria da produção intelectual do campo de biblioteconomia/ciência da informação: a) *Ciência da Informação*, publicado pelo Instituto Brasileiro de Informação em Ciência e Tecnologia IBICT; b) *Revista da Escola de Biblioteconomia da UFMG*, publicada pela Escola de Biblioteconomia da UFMG; tanto o periódico quanto a Escola tiveram seus nomes alterados, na década de 90: a revista passou a chamar-se *Perspectivas em Ciência da Informação*, e a instituição, Escola de Ciência da Informação da UFMG; c) *Revista de Biblioteconomia de Brasília*, publicada pela Associação dos Bibliotecários do Distrito Federal com a colaboração do Departamento de Biblioteconomia da UnB; e d) *Revista Brasileira de Biblioteconomia e Documentação*, publicada pela Federação Brasileira de Associações de Bibliotecários (FEBAB). As duas últimas estão com sua publicação praticamente interrompida, de tempos em tempos acenando com um retorno. Entretanto, se for feita uma pesquisa nesses periódicos e também nos anais dos eventos científicos do mesmo período, talvez nada se encontre de pesquisas que, a rigor, se pudesse dizer que tratam de análise de assunto. Com a criação da ANCIB, passou a realizar-se, de dois em dois anos, por sua iniciativa, um Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB). Os seis Encontros até agora realizados ocorreram nos locais e datas a seguir especificados: 1) Belo Horizonte, 1994; 2) Valinhos, São Paulo, novembro de 1995; 3) Rio de Janeiro, setembro de 1997; 4) Brasília, novembro de 2000; 5) Belo Horizonte, novembro de 2003; 6) Florianópolis, novembro de 2005.

No bojo desses encontros e com o apoio e respaldo da ANCIB foram criados Grupos de Trabalho especializados nos vários subcampos da ciência da informação:

- Informação Tecnológica e Administração de Serviços

- Representação do Conhecimento/Indexação/Teoria da Classificação
- Novas Tecnologias/Redes de Informação/Educação à Distância
- Informação e Sociedade/Ação Cultural
- Produção Científica/Literatura Cinzenta
- Formação Profissional e Mercado de Trabalho

É logicamente no grupo de Representação do Conhecimento/Indexação/Teoria da Classificação que iremos encontrar a pouca pesquisa existente sobre análise de assunto na área.

#### **8.4 Tendências**

Langridge (1989) afirma que o desempenho de um sistema de recuperação de informação depende principalmente da qualidade da análise de assunto nele realizada. Se essa afirmação for aceita como um pressuposto correto, tem-se que concordar que as perspectivas da pesquisa na área são as mais auspiciosas possíveis. Se a pesquisa nessa área ainda não atraiu a atenção que merece seria, por conseguinte, fruto do não reconhecimento da importância que tem a atividade no contexto das bibliotecas e dos sistemas de recuperação de informação.

O surgimento dos documentos eletrônicos, e a facilidade de sua reprodução e disponibilização pública, levou ao desenvolvimento natural de bibliotecas digitais ou virtuais, destinadas a fazer, em relação a esses documentos, o mesmo que faziam as bibliotecas tradicionais com os documentos impressos: reunir, tratar, disseminar os documentos. A quantidade desses documentos eletrônicos, entretanto, é enorme, se comparada aos padrões com que se estava acostumado no universo dos documentos impressos. Assim, a maior biblioteca do mundo, a Biblioteca do Congresso dos Estados Unidos (Library of Congress), tinha, pouco antes da explosão da Internet, cerca de 12 milhões de títulos de documentos. Essa coleção, considerada enorme se comparada às coleções das demais bibliotecas então existentes no mundo, mesmo outras também consideradas grandes bibliotecas, soa ínfima em termos do número de páginas da Internet, ou seja, documentos eletrônicos, disponíveis atualmente.

Por isso, a função de tratamento da informação, que já se revelava complexa e financeiramente onerosa nas bibliotecas tradicionais parece quase impraticável no ambiente de rede. É essa constatação que tem levado muitos a descrever da viabilidade de se praticar, na biblioteca digital, o tratamento tal como se fazia nas bibliotecas tradicionais. Há entretanto um dado que parece sempre ser esquecido nessa discussão: o de que qualquer biblioteca ou sistema de recuperação da informação tem como pressuposto uma seleção de informações. Em outras palavras, nenhuma biblioteca ou sistema de recuperação da informação da era pré-Internet incorporava a suas coleções “tudo” o que havia de publicado em sua respectiva área de interesse. E esse procedimento não era determinado apenas por questões de custo. Antes de tudo, se baseava no conceito de que era importante, para qualquer biblioteca, não apenas possuir documentos que tivessem um potencial de atendimento às necessidades de informação de seu público, mas também poupar esse público de documentos que não tivessem esse potencial. Aqueles familiarizados com as cinco leis da biblioteconomia do bibliotecário indiano Ranganathan irão recordar aqui, imediatamente, da lei que diz: “poupe o tempo do leitor”.

No ambiente de rede, ao contrário, ainda não se percebe a preocupação em fazer uma seleção de informações, mas simplesmente indexar o maior número

possível de sites. Tanto que muitos dos motores de busca usam o número de sites indexados como um grande diferencial. Com isso, o princípio de poupar o tempo do usuário é desprezado, como podem certamente comprovar a maioria dos usuários desses sites de busca em relação à maioria das consultas que fazem. Acredita-se que aplicado o processo de seleção na construção de bibliotecas digitais ou sistemas de recuperação da informação, o número de documentos em suas coleções seria reduzido de forma drástica. E quanto menor o número de documentos numa biblioteca/SRI, mais viável se torna o tratamento da informação nesses sistemas.

O tratamento da informação como um problema para as bibliotecas/SRIs já havia sido constatado muito antes da explosão digital. É o que demonstra toda a pesquisa em busca de métodos automatizados na área que veio sendo desenvolvida desde a década de 50 do século passado. Entretanto, nada se fará de produtivo nessa área sem se conhecer os processos humanos de tratamento da informação, que deverão constituir necessariamente a base para o desenvolvimento dos respectivos processos automatizados. Entre esses processos, como se procurou demonstrar neste livro, destaca-se como especialmente desafiador do ponto de vista intelectual o de análise de assunto, cerne e base de todo o tratamento temático da informação. Repetindo o pensamento de Langridge (1989) com que se iniciou este capítulo: nenhum sistema de recuperação de informação — e, logicamente, nenhuma biblioteca — pode ser melhor que a análise de assunto que lhe dá sustentação.

Assim, pode-se concluir que a tendência da pesquisa em análise de assunto é no sentido de identificar os padrões de processamento dos indexadores de forma a não apenas auxiliar no aperfeiçoamento desses padrões, como também servir de insumo à cada vez mais necessária automatização do processo.

## Referências

- ALBRECHTSEN, Hanne. Subject analysis and indexing: from automated indexing to domain analysis. *The Indexer*, v. 18, n. 4, p. 219-224, Oct. 1993.
- ALLEN, B. Cognitive research in information science: implications for design. *Annual Review of Information Science and Technology*, v. 26, p. 3-37, 1991.
- ARTANDI, Susan. Machine indexing: linguistic and semiotic implications. *Journal of the American Society for Information Science*, v. 27, n. 4, p. 235-239, July/Aug. 1976.
- BARANOV, Ulf Gregor. Perspectivas na contribuição da lingüística e de áreas afins à ciência da informação. *Ciência da Informação*, v. 12, n. 1, p. 23-35, 1983.
- BEGHTOL, Clare. Bibliographic classification theory and text linguistics: aboutness, intertextuality, and the cognitive act of classifying documents. *Journal of Documentation*, v. 42, n. 2, p. 84-113, June 1986.
- BELKIN, Nicholas J. The cognitive viewpoint in information science. *Journal of Information Science*, v. 16, p.11-15, 1990.
- BELL, Hazel K. Bias in indexing and loaded language. *The Indexer*, v. 17, n. 3, p. 173-177, Apr. 1991.
- BÉRNARDEZ, Enrique. *Introducción a la lingüística del texto*. Madrid: Espasa-Calpe, 1982. 324p.
- BERNIER, Charles L. Indexing process evaluation. *American Documentation*, v. 16, n. 4, p. 323-328, Oct. 1965.
- BERTRAND, A.; CELLIER, Jean Marie. Psychological approach to indexing: effects of the operator's expertise upon indexing behavior. *Journal of Information Science*, v. 21, n. 6, p. 459-472, 1995.
- BIRMAN, bel. Leitura crítica: questões sobre recepção. In: SIMPÓSIO NACIONAL DE LEITURA, Rio de Janeiro, 1994. *Leitura, saber e cidadania*. Rio de Janeiro: Biblioteca Nacional, 1994. p. 103-112.
- BLAIR, David C. *Language and representation in information retrieval*. Amsterdam; New York: Elsevier Science, 1990.
- BRAAM, R. R.; BRUIL, J. Quality of indexing information: authors' view on indexing of their articles in Chemical Abstracts online CA-file. *Journal of Information Science*, v. 18, n. 5, p. 399-408, 1992.
- BUTLER, Pierce. *Introdução à ciência da biblioteconomia*. Rio de Janeiro: Lidador, 1971.
- CAMPELLO, Bernadete Santos et al. (Org.). *Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: Ed. UFMG, 2000.
- CAMPOS, Astério Tavares. A indexação. *Revista de Biblioteconomia de Brasília*, v. 15, n. 1, p. 69-72, jan./jun. 1987.
- CANCLINI, Néstor Garcia. Consumidores do séc. XXI, cidadãos do séc. XVIII. In: \_\_. *Consumidores e cidadãos: conflitos multiculturais da globalização*. Rio de Janeiro: UFRJ, 1995. p. 13-47.
- CAVALCANTI, Marilda do Couto. *Interação leitor-texto: aspectos de interpretação pragmática*. Campinas: Ed. UNICAMP, 1989.
- CESARINO, Maria Augusta da Nóbrega; PINTO, Maria Cristina Melio Ferreira. Análise de assunto. *Revista de Biblioteconomia de Brasília*, v. 8, n. 1, p. 32-43, jan./jun. 1980.

CHARTIER, Roger. Comunidade de leitores. In: \_\_\_\_\_. *A ordem dos livros: leitores, autores e bibliotecas na Europa entre os séculos XIV e XVIII*. Brasília: Ed. UnB, 1994. p. 11-31.

CHU, Clara M.; O'BRIEN, Ann. Subject analysis: the critical first stage in indexing. *Journal of Information Science*, v. 19, p. 434-454, 1993.

CINTRA, Anna Maria Marques. Elementos de lingüística para estudos de indexação. *Ciência da Informação*, v. 12, n. 1, p. 5-22, 1983.

CÓDIGO de catalogação anglo-americano; preparado por the American Library Association... [et al.]; coordenado por Michael Gorman e Paul W. Winkler. 2. ed. São Paulo: FEBAB, 1983-1985.

COSTA VAL, Maria da Graça. Texto e textualidade. In: \_\_\_\_\_. *Redação e textualidade*. São Paulo: Martins Fontes, 1991. cap. 1, p. 3-16.

CUNHA, Isabel M. R. Ferin. Análise documentária. In: SMIT, Johanna W. (Coord.). *Análise documentária*. 2. ed. Brasília: IBICT, 1987. p. 38-60.

DAHLBERG, Ingetraut. Classification and "The Tree of Cognition" (Editorial). *International Classification*, v. 14, n.3, p. 125-126, 1987.

DAHLBERG, Ingetraut. Knowledge organization and terminology: philosophical and linguistic bases. *International Classification*, v. 19, n. 2, p. 65-71, 1992.

DANIELS, P. J. Cognitive models in information retrieval: an evaluative review. *Journal of Documentation*, v. 42, n. 4, p. 272-305, Dec. 1986.

DAVID, Claire et al. Indexing as problem solving: a cognitive approach to consistency. In: ANNUAL MEETING OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE, 58, 1995, Washington, DE. *Proceedings...* Washington, DE, 1995. p. 49-55.

DELL'ISOLA, Regina Lúcia Peret. *O contexto e a compreensão lexical na leitura em português - língua estrangeira*. 1999. 370f. Tese (Doutorado em Letras lingüística aplicada) - Faculdade de Letras, Universidade Federal de Minas Gerais Belo Horizonte, 1999.

DEWEY, Melvil. *Dewey decimal classification and relative index*. 22. ed., edited by Joan S. Mitcheil et al. Dublin, Ohio: OCLC Online Computer Library Center 2003.

DIAS, E. W. Contexto digital e tratamento da informação. *DataGramaZero*, Rio de Janeiro, v. 2, n. 5, out. 2001. Publicação eletrônica disponível em: <<http://www.dgz.org.br>>. Acessado em: 03 mar 2006.

DIAS, E. W. Análise de assunto: percepção do usuário quanto ao conteúdo de documentos. *Perspectivas em Ciência da Informação*. Belo Horizonte, v. 9, n. 2, p. 146-157, jul./dez. 2004.

DIAS, E. W., NAVES, M. M. L., MOURA, M. A. O usuário-pesquisador e a análise de assunto. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 6, n. 2, p.205 -221, jul./dez. 2001.

DUBLIN Core Metadata Initiative. Dublin, OH: OCLC, 1995-2006. Disponível em: <http://dublincore.org/> Acessado em: 07 março 2006.

FAGUNDES, Silvana Aparecida *Leitura em análise documentária de artigos de jornais*. 2001. 261f. Dissertação (Mestrado em ciência da informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Manha.

FARRADANE, J. Knowledge, information, and information science. *Journal of Information Science*, v. 2, p. 75-80, 1980.

FARROW, John E All in the mind: concept analysis in indexing. *The Indexer*, v. 19, n. 4, p. 243-247, Oct. 1995.

- FÁVERO, Leonor Lopes. *Coesão e coerência textuais*. 2 ed. São Paulo: Ática, 1993.
- FREIRE-MAIA, Newton. *A ciência por dentro*. 3. ed. Petrópolis: Vozes, 1995. 262p.
- FROHMANN, B. Rules of indexing: a critique of mentalism in information retrieval theory. *Journal of Documentation*, v. 46, n. 2, p.81 -101, June 1990.
- FUGMANN, Robert. The complementarity of natural and indexing languages. In: CHAN, L. M.; RICHMOND, P.; SVENONIUS, E. (Org.). *Theory of subject analysis*. Littleton, CO: Libraries Unlimited, 1985. p. 390-402.
- FUJITA, M. S. L. A leitura do indexador: estudo de observação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 4, n. 1, p. 101-116, jan./jun. 1999.
- FUJITA, M. S. L.; SILVA, M. R. A leitura documentária na prática de indexação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 4, 2000, Brasília. *Trabalhos apresentados*. Brasília: ANCIB, 2000.
- GARCIA MARCO, Francisco Javier; ESTEBAN NAVARRO, Miguel Angel. On some contributions of the cognitive sciences and epistemology to a theory of classification. *Knowledge Organization*, v. 20, n. 3, p. 126-132, 1993.
- GARDIN, Jean-Claude. Document analysis and linguistic theory. *Journal of Documentation*, v. 29, n. 2, p. 137-168, June 1973.
- GIASSON, Joce Lyne. *A compreensão na leitura*. Lisboa: Asa, 1993. 320p.
- GINEZ DE LARA, Marilda Lopes. *A representação documentária: em jogo a significação*. 1993. 133f. Dissertação (Mestrado em biblioteconomia e documentação) — Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.
- GUIMARÃES, J. A. C. *Análise documentária em jurisprudência: elementos para uma metodologia de indexação de acórdãos trabalhistas brasileiros*. 1994. Tese (Doutorado em ciências da comunicação) — Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1994.
- \_\_\_\_\_. *Condensação documentária em legislação e jurisprudência: elementos instrumentais para a elaboração de ementas*. 2000. Tese (Livre Docência) - Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, 2000.
- \_\_\_\_\_. *A diplomática como aparato metodológico no processo de análise documentária no âmbito jurídico*. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 3, 1997, Rio de Janeiro. *Anais...* Rio de Janeiro: ANCIB, 1997. p. 29.
- HARTER, Stephen P. Psychological relevance and information science. *Journal of the American Society for Information Science*, v. 43, n. 9, p. 602-615, Oct. 1992.
- HILLMAN, D. J. The notion of relevance. *American Documentation*, v. 15, n. 1, p. 26-34, 1964.
- HJORLAND, Birger. The concept of “subject” in information science. *Journal of Documentation*, v. 48, n. 2, p. 172-200, June 1992.
- HJORLAND, Birger. *Information-seeking and subject representation; an activate theoretical approach to information science*. Westport, Connecticut; London: Greenwood, 1997.
- HUTCFIINS, W. J. The concept of “aboutness” in subject indexing. *Aslib Proceedings*, v. 30, n. 5, p. 172-181, May 1978.
- INGWERSEN, Peter. Search procedures in the library - analysed from the cognitive point of view, *Journal of Documentation*, v. 38, n. 3, p. 165-191, Sept. 1982.

- KATO, Mary. *O aprendizado da leitura*. São Paulo: Martins Fontes, 1985a. 120p
- \_\_\_\_\_. Uma visão interativa da legibilidade. *Ilha do Desterro*, v. 13, p. 57-66, 1985b
- KOBASHI, Nair Yumiko. *A elaboração de informações documentárias: em busca de uma metodologia*. São Paulo: ECA/USP, 1994. 195f. Tese (Doutorado em ciências da comunicação) — Escola de Comunicação e Artes, Universidade de São Paulo. São Paulo, 1994.
- KOCH, Ingedore Grunfeld Viliça. *A coesão textual*. 4.ed. São Paulo: Contexto, 1991. 75p.
- KOCH, Ingedore Grunfeld Viliça; TRAVAGLIA, Luiz Carlos. *A coerência textual*. São Paulo: Contexto, 1990. 94-p.
- LANCASTER, E W. *Indexação e resumos: teoria e prática*. Brasília: Briquet de Lemos/Livros, 1993.
- LANGRIDGE, D. *Subject analysis*. London: Bowker-Saur, 1989.
- LÉVY, Pierre. *A virtualização do texto*. In: \_\_\_\_\_. *O que é virtual*. São Paulo: Ed. 34, 1996. cap. 3, p. 35-50.
- LURIA, A. R. *Desenvolvimento cognitivo*. 2. ed. São Paulo: Ícone, 1994.
- MEDEIROS, Mansa Brisher Basílio. Terminologia brasileira em ciência da informação. *Ciência da Informação*, v. 15, n. 2, p. 135-142, jul./dez. 1986.
- METCALFE, John. When is a subject not a subject? In: RAWSKI, C. H. *Towara a theory of librarianship*. Metüchen, N.J.: Sacrecrow, 1977. p. 305-339.
- MONDAY, Isabelle. Les processus cognitifs et la redacción de résumés. *Documentation et Bibliothèques*, v. 42, n. 2, p. 55-62, avr./juin 1996.
- NAVARRO, Sandrelei. Interface entre lingüística e indexação: revisão de literatura. *Revista Brasileira de Biblioteconomia e Documentação*, v. 21, n. 1/2, p. 46-62, jan./jun. 1988.
- NAVES, M. M. L. Aspectos conceituais do browsing na recuperação da informação. *Ciência da Informação*, Brasília, v. 27, n. 3, p. 279-283, 1998.
- \_\_\_\_\_. *Fatores interferentes no processo de análise de assunto*. 2000. Tese (Doutorado em ciência da informação) — Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2000.
- NEVES, D. A. de B. *Aspectos metacognitivos na leitura do indexador; um estudo exploratório*. 2004. 165 f. Tese (Doutorado em ciência da informação) — Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.
- OLIVEIRA, Silvio Luiz de. *Tratado de metodologia científica*. São Paulo: Pioneira, 1997.
- ORLANDI, Eni Pulcinelli. *A linguagem e seu funcionamento: as formas do discurso*. 2. ed. rev. aum. Campinas: Pontes, 1987. 276p.
- PINTO MOLINA, María. Documentary abstracting: toward a methodological model. *Journal of the American Society for Information Science*, v. 46, n. 3, p. 225-234, 1995.
- \_\_\_\_\_. Interdisciplinary approaches to the concept and practice of Written Documentary Content Analysis (WTDEA). *Journal of Documentation*, v. 50, n. 2, p. 111-133, June 1994.
- RENKEMA, Jan. *Discourse studies: an introductory text book*. Amsterdam; Philadelphia: J. Benjamins, 1993.
- ROLLING, L. Indexing consistency, quality and efficiency. *Information Processing & Management*, v. 17, p. 69-76, 1981.



SAARTI, Jarmo. Taxonomy of novel abstracts based on empirical findings. *Knowledge Organization*, v. 27, n. 4, p. 213-220, 2000.

SALTON, G.; MCGILL, M. J. *Introduction to modern information retrieval*. New York: McGraw-Hill, 1983.

SANTAELLA, L. A cartografia das ciências. In: \_\_\_\_\_. *A assinatura das coisas: Peirce e a literatura*. Rio de Janeiro: Imago, 1992. cap. 5, p. 101-140.

SANTOS, Leonor Werneck dos. *Discurso, coesão e argumentação*. Rio de Janeiro: Oficina do Autor, 1996. 110p.

SCHUSTACK, Minam W.; EHRLICH, Susan E; RAYNER, Keith. Local and global sources of contextual facilitation in reading. *Journal of Memory and Language*, v.26, p. 322-340, 1987.

SILVA, Maria dos R. *A localização do tema no artigo acadêmico para indexação*. 2004. Dissertação (Mestrado em ciência da informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Manha, 2004.

SILVA, Marilée Patta e. *Análise de assunto auxiliada por um sistema de informação, em textos da computação*. 2004. Dissertação (Mestrado em ciência da informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

SMIT, Johanna W. (Coord). *Análise documentária: a análise da síntese*. Brasília: IBICT, 1987. p. 5-10.

SOUZA, Vânia Regina Alves. *Leitura em língua estrangeira (inglês) para elaboração de resumos documentários*. 2004. Dissertação (Mestrado em ciência da informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Manha, 2004.

STONE, Alva T. That elusive concept of "aboutness": the years work in subject analysis, 1992. *Library Resources and Technical Services*, v. 37, p. 277-298, July 1993.

STREHL, Letícia. Avaliação da consistência da indexação realizada em uma biblioteca universitária de artes. *Ciência da Informação*, v. 27, n. 3, p. 329-335, set./dez. 1998.

SVENONIUS, Elaine. Directions for research in indexing, classification and cataloging. *Library Resources & Technical Services*, v. 25, n. 1, p. 88-103, 1981.

SWIFT, D.F.; WINN, V.; BRAMER, D. Aboutness' as a strategy for retrieval in the social sciences. *Aslib Proceedings*, v. 30, n. 5, p. 182-187, May 1978.

TIBBO, Helen R. Abstracting across the disciplines: a content analysis of abstracts from the natural sciences, the social sciences, and the humanities with implications for abstracting standards and online information retrieval. *Library and Information Science Research*, v. 14, p. 31-56, 1992.

TINKER, John F. Imprecision in meaning measured by inconsistency of indexing. *American Documentation*, v. 17, n. 2, p. 96-102, Apr. 1966.

TODD, Ross T. Academic indexing: what's it all about? *The Indexer*, v. 18, n. 2, p. 101-104, Oct. 1992.

TONTA, Yasar. A study of indexing consistency between Library of Congress and British Library Catalogers. *Library Resources & Technical Services*, v. 35, n. 2, p. 177-185, Apr. 1991.

WILSON, P. Subjects and the sense of position. In: CHAN, L. M.; RICHMOND, P.; SVENONIUS, E. (Org.) *Theory of subject analysis; a sourcebook*. Littleton, Co Libraries Unlimited, 1985. p. 306-325.