

ARTIGO***ANÁLISE DOCUMENTÁRIA E REPRESENTAÇÃO DA INFORMAÇÃO****Nair Yumiko Kobashi**Professora, USP / ECA***Resumo**

Apresentação de aspectos da Análise Documentária, campo de pesquisa relacionado aos processos de organização e representação de informações documentárias. São caracterizados os seus referenciais teóricos, conceituais e metodológicos mais importantes e a natureza de seu diálogo com as diversas disciplinas que têm o texto como objeto (Análise de Conteúdo, Análise Logicista, Análise do Discurso, Linguística Textual, Tratamento de Linguagem Natural). Nessa perspectiva, ressalta-se a importância dos conceitos de representação, significação e texto, bem como as operações realizadas com este último, no âmbito da Análise Documentária.

Palavras-chave

Análise Documentária
Organização - Informação
Representação - Informação

1 Introdução

A preocupação teórica com a organização e a representação de informações, com fins documentários, é fato relativamente recente se levarmos em conta as práticas relacionadas a esses processos, que são executadas, pelo menos, desde o segundo milênio A.C. É suficiente lembrar os documentos mesopotâmicos da época das cidades-estado, que já apresentavam uma forma de tratamento documentário: as obras cunhadas em tábuas de argila eram protegidas por uma espécie de envelopes, sobre os quais se transcreviam informações que cumpriam função semelhante à dos modernos resumos (fornecer informação concisa sobre o documento original). Desse modo, era possível conhecer o conteúdo das tábuas sem que fosse necessário quebrar o seu invólucro protetor. (WITTY, F. J. 1973) Da mesma forma, os papiros da Biblioteca de Alexandria eram condensados de

modo a facilitar as pesquisas dos estudiosos da época. O resumo de Agamêmnon, de Ésquilo, transcrito abaixo, ilustra o notável trabalho documentário ancestral a que nos referimos:

"Agamêmnon, ao partir para Tróia, prometeu a Clitemnestra que, se saqueasse Tróia, ele a avisaria, no mesmo dia, através de um sinal luminoso. Por isso, Clitemnestra contrata um observador que a informa assim que vê o sinal. Ela, então, convoca a assembléia dos anciãos - que formam o coro - para consultá-los a respeito do sinal. Ao saberem do fato, uma parte deles entoou uma canção de triunfo. Logo depois, Taltíbios (o arauto) aparece e descreve em detalhes os eventos da viagem. Em seguida, Agamêmnon vem em uma carruagem seguida por uma outra, na qual está Cassandra, seu espólio. No momento em que ele entra no palácio com Clitemnestra, Cassandra, antes de entrar, faz profecias sobre sua própria morte, a de Agamêmnon e sobre o matricídio de Orestes; a seguir, adentra precipitadamente e, como alguém pronto para morrer, deixa cair as suas insígnias. Esta parte da peça é admirável, porque provoca temor e compaixão. Como é peculiar em Ésquilo, a morte de Agamêmnon dá-se nos bastidores; nada é dito sobre a morte de Cassandra, até que seu cadáver seja exposto. Egisto e Clitemnestra revelam, cada qual, suas próprias razões para o assassinato: o dela, é o sacrifício de Ifigênia; o dele, os infortúnios causados por Atreu a seu pai. A peça foi encenada durante o governo de Filaco, no segundo ano da octagésima Olimpíada (459/8 A.C.). Ésquilo ganhou o primeiro prêmio com Agamêmnon, Coéforas (Libation bearers), Eumênides e sua peça satírica Proteu. Xenocles Afidnaios dirigiu o Coro." (WITTY, F.J. 1973. p.195)

A Documentação, enquanto sistema público de circulação de informações, nasceu no século XVII, com a edição do "Le Journal des Sçavans". Publicado em Paris, a partir de 1665, pela Academia Francesa de Ciências, esse periódico apresentava os resumos dos trabalhos científicos, filosóficos e artísticos produzidos no período. (BORKO, H., BERNIER, C. 1975)

Publicações semelhantes cresceram em número e abrangência de cobertura durante os séculos XVIII e XIX, até chegarem aos formatos atuais, as bases de dados, em suas inúmeras formas legíveis por computador: fitas magnéticas, disquetes, CD-ROMs, etc.

Apesar dessa longa história, as preocupações teóricas da Documentação se manifestam de modo sistemático apenas a partir da década de 50, estimula-

das, em parte, pelos experimentos automáticos de indexação, elaboração de resumos e recuperação de informações.

Um dos marcos mais importantes desse percurso encontra-se, sem sombra de dúvida, na proposta de indexação e elaboração automática de resumos, de Luhn (1958 e 1959). Sua contribuição foi, durante quase duas décadas, o paradigma norteador de inúmeros projetos que visavam modificar radicalmente os métodos tradicionais de armazenamento, tratamento e recuperação de informações.

As experiências realizadas apontaram, porém, os limites do modelo acima referido. Os critérios estatísticos (ocorrência/frequência) de tratamento de textos, propostos por Luhn, mostraram ser eficientes para identificar formas significantes, mas pouco úteis para determinar significados e, portanto, para reconhecer mensagens e condensá-las. A frustração decorrente desses experimentos pode ser sintetizada nesta afirmação de Borko:

"Após alguns meses de trabalho, cheguei à relutante conclusão de que não há meio razoável de se fazer indexação automática com base somente no princípio da exclusão." (BORKO, H. *apud* ARTANDI, S. 1970)

1.1 A interface Lingüística/Análise Documentária

Os limites do modelo de Luhn levaram os pesquisadores da área a buscar novas hipóteses de trabalho. Passa-se, por exemplo, a atribuir valor aos aspectos comunicacionais dos sistemas documentários e aos instrumentos que possam promover níveis crescentes de interação entre sistema e usuário, tais como os tesouros. Inicialmente utilizados para servir ao indexador para selecionar o termo mais adequado para descrever o conteúdo informacional dos documentos, os tesouros passam a ser propostos, igualmente, como ferramentas valiosas para auxiliar o usuário a explorar adequadamente os sistemas documentários. (ARTANDI, S. 1970)

Nesse mesmo período, observa-se a aproximação entre a Análise Documentária e a Lingüística Aplicada, fato justificado inicialmente pelas semelhanças existentes entre os processos documentários e a tradução automática, ressaltadas, evidentemente, as

especificidades de cada uma delas. Destaca-se, nessa vertente, o arqueólogo francês Jean-Claude Gardin. (GARDIN, J. C. 1969, 1991)

Devem ser a ele creditadas a própria denominação do campo que hoje se conhece por Análise Documentária, a criação das bases fundantes de uma teoria específica sobre os processos documentários e a terminologia adotada pela Ciência da Informação de orientação francesa.

Os aspectos documentários de maior interesse para o referido pesquisador são os processos relacionados à passagem do texto original para a sua representação através de linguagens documentárias. Suas reflexões levam em consideração o fato de ser a Documentação um domínio no qual o número de textos analisados é sempre vasto (atinge hoje alguns milhões), o que lhe permite caracterizar a Análise Documentária como o ramo industrial da análise de textos.

Uma de suas contribuições mais notáveis dá-se precisamente na proposição de um modelo formal de linguagem documentária. (CROSS, R. C., GARDIN, J. C., LÉVY, F. 1964) Segundo Gardin, toda linguagem documentária é constituída por um conjunto de termos (**o léxico**), por relações entre as unidades lexicais, determinadas *a priori* (**o eixo paradigmático**) e por uma sintaxe que articula os encadeamentos entre os termos da linguagem (**o eixo sintagmático**).

Outros pesquisadores compartilham, com Gardin, a idéia da necessidade de se promover a aproximação Lingüística/Documentação. Edmundson (1969), por exemplo, afirma que os métodos de elaboração automática de resumos devem levar em conta as características sintáticas e semânticas da linguagem e do texto, já que as formas de tratamento baseadas em evidências estatísticas mostraram ser pouco eficientes para realizar adequadamente tais tarefas.

Essa mesma posição é reforçada por Montgomery (1972), que analisa a aproximação Lingüística/Ciência da Informação do ponto de vista dos benefícios mútuos. Segundo a autora, a Lingüística contribuiria para resolver os problemas de representação dos itens documentários e das perguntas dos usuários ao sistema, por serem tais processos muito semelhantes àqueles presentes na tradução automática. A Lingüística Aplicada, por sua vez, poderia beneficiar-se da Ciência da Informação enquanto campo privilegiado de experimentação de teorias e modelos.

Seguindo a mesma linha de raciocínio, o trabalho de Sparck Jones e Kay (1973), procura aprofundar a inter-relação das duas áreas, indicando as potenciais contribuições que o primeiro campo poderia dar ao segundo. Nesse trabalho, os autores abordam as questões sintáticas e semânticas da recuperação da informação e das linguagens documentárias. No entanto, as observações apresentadas por Sparck Jones e Kay são otimistas apenas nas questões relacionadas à sintaxe e consideram que as tentativas de aplicação das teorias semânticas à Documentação ainda estão longe de apresentar resultados conclusivos.

A interface Ciência da Informação/Lingüística, inaugurada em fins dos anos 60, continua sendo enriquecida com novas abordagens. Recentemente, García Gutiérrez, pesquisador espanhol com inúmeros trabalhos sobre as linguagens documentárias, caracteriza a Teoria Lingüística como o

"... fundamento dos métodos e processos documentários (...) no qual irrompem o discurso, o conhecimento, a análise, a tradução, a redução e a representação." (GARCÍA GUTIÉRREZ, A. 1990. p. 11)

Os modelos lingüísticos têm fornecido subsídios para os trabalhos de diversos pesquisadores brasileiros envolvidos com testes ou proposição de modelos que dêem conta do processo global de Análise Documentária. (SMIT, J. 1976, 1987; BARANOW, U. G. 1983; CINTRA, A. M. M. 1983, 1987; CINTRA, A. M. M. et al. 1993; CUNHA, I. M. R. F. 1990; KOBASHI, N. Y. 1989, 1989a; TÁLAMO, M. F. G. M. 1987; TÁLAMO M. F. G. M. et al. 1992, 1992a, 1992b; LARA, M. L. G. 1993) Concentrada durante um certo período nos processos de análise e condensação de textos, a mais recente produção teórica dos referidos pesquisadores aponta para a interface Terminologia/Análise Documentária como um campo com possibilidades de imprimir rigor às práticas empíricas de construção de vocabulários para fins documentários.

1.2 A interface Análise Documentária/ Inteligência Artificial

A Ciência da Informação, assim como outras disciplinas, tem procurado se beneficiar dos progressos obtidos pela Inteligência Artificial (IA).

Esse paradigma, utilizado inicialmente na concepção de Sistemas Especialistas para explorar bases de dados ou para orientar a catalogação, incorpora-se de forma crescente às pesquisas de indexação e de produção automática de resumos.

Essas metas, perseguidas há ao menos 40 anos, desde os trabalhos pioneiros de Luhn, ganharam um novo alento, em face dos resultados positivos alcançados por alguns sistemas desenvolvidos pela IA. Nas pesquisas desenvolvidas no âmbito da Ciência da Informação, são referências obrigatórias os projetos de compreensão automática de textos de Schank e Abelson (1977), da Universidade de Yale, e o de Lehnert (1984), da Universidade de Amherst. Nesses sistemas, a atração maior reside nos módulos de geração de resumos, utilizados como mecanismos de validação da compreensão de textos de natureza diversa: textos jornalísticos, mensagens de telex, narrativas.

O modelo de Schank e Abelson inspirou, por exemplo, o projeto TOPIC (Text Oriented Procedures for Information Management and Condensation of Expository Texts), desenvolvido na Universidade de Konstanz (Alemanha).

O TOPIC é um sistema que, além de efetuar as operações básicas dos sistemas convencionais de recuperação de informação, é dotado de um módulo que condensa textos de aproximadamente 2000 a 4000 palavras (entre 10 a 12 laudas), em língua alemã, no domínio das tecnologias de informação. Sua característica mais notável é a de elaborar produtos, dentre os quais resumos, adaptados aos diferentes tipos de demandas. O módulo de indexação automática, por outro lado, dispõe de um mecanismo que converte a informação do texto para os termos de um vocabulário controlado. (HAHN, V., REIMER, V. 1984)

Um outro modo de responder aos objetivos acima formulados - imprimir qualidade e pertinência aos produtos documentários, renovando as técnicas de tratamento - é através da construção de bancos de dados em domínios cada vez mais especializados, em contraposição aos modelos enciclopédicos das décadas de 60 e 70.

Essa tendência está presente nos trabalhos do INIST (Institut de L'Information Scientifique et Technique, da França), o qual vem procurando construir novos tipos de produtos documentários. Embora as fontes de informações sejam convencionais - as bases de dados Pascal e Francis - esses produtos

apresentam maior valor agregado, já que os resultados das buscas documentárias são apresentados sob a forma de "mapas conceituais" elaborados a partir do tratamento lingüístico e estatístico dos termos presentes nos produtos da recuperação. As estruturas conceituais criadas *ad hoc* pretendem refletir, de modo pertinente, a dinâmica da criação científica contemporânea, por excelência interdisciplinar. Desse modo, as referidas estruturas podem ser utilizadas como parâmetros para as operações de indexação de novos documentos e para refinar as estratégias de busca do sistema. Além disso, a edição final do documento, apresentado na forma de hipertexto, otimiza a interface homem/máquina, dada a qualidade de visualização do produto final. (DUCLOY, J. et al. 1991)

Na mesma perspectiva, enquadra-se a proposta de base de dados inteligente de Juan Barceló (1991), desenvolvida junto ao grupo de pesquisa de Jean-Claude Gardin, no CNRS/ Centre de Recherches Archéologiques.

O autor citado parte do princípio de que a solução de um dado problema científico depende muito mais de conhecimentos estruturados do que de um conjunto de dados empíricos apresentados em forma bruta, ou mesmo de um conjunto de referências apenas justapostas. Desse modo, sua proposta, mais do que uma base de dados relacional clássica, instaura-se como um sistema de produção de novos dados, a partir da articulação de conhecimentos já disponíveis em um dado domínio.

O sistema por ele idealizado é composto de:

- um conjunto de conceitos que descrevem conhecimentos relevantes relacionados a um problema;
- uma representação "ativa" desses conceitos, o que permite ligar conceitos e perguntas do usuário;
- um conjunto de regras que gerencia a representação conceptual;
- um conjunto de operadores lógicos que articula a linguagem de representação;
- três tipos de meta-conhecimentos:
 - a) conhecimentos sobre o problema a ser resolvido;
 - b) conhecimentos sobre a estrutura do sistema;
 - c) conhecimentos sobre a estratégia para resolver o problema.

Observa-se que a base de dados inteligente proposta por Barceló é uma modalidade de Sistema Especialista que adota a técnica do hiperdocumento na sua interface com o usuário.

2 A organização e a representação de informações documentárias

As pesquisas relacionadas à organização e representação de informações documentárias devem levar em conta certas restrições pragmáticas:

- o fato de a Documentação operar sempre com grandes quantidades de textos, mesmo sendo restritos os domínios cobertos;
- o fato de o valor do produto documentário estar, na maioria dos casos, relacionado à atualidade da informação.

Estes aspectos exigem a adoção de princípios e de mecanismos que garantam resultados simultaneamente rápidos e potentes, aptos a produzir informações documentárias que observem os parâmetros de qualidade e de pertinência exigidos pela demanda. Não é uma tarefa trivial, como veremos a seguir.

A elaboração de informações documentárias, independentemente do *corpus* documental ou do sistema documentário no interior do qual se realiza, compreende um conjunto de operações, esquematizáveis em três fases:

- ler um texto;
- selecionar, no mesmo, o conteúdo informacional considerado pertinente para usos determinados;
- representar, ou seja, dar forma às informações selecionadas, de modo a torná-las documentariamente manipuláveis.

Tradicionalmente, concebe-se a elaboração de resumos como meros atos técnicos de reduzir textos, ou a indexação como o ato de estabelecer equivalências entre palavras do texto e palavras de uma Linguagem Documentária.

No entanto, todo ato documentário é um ato de comunicação que tem a finalidade de promover a circulação da informação. Desse modo, para ser eficaz, o ato documentário requer, inicialmente, uma visão

clara dos parâmetros específicos no interior dos quais instaura-se a situação comunicativa (objetivos institucionais, tipos de usuários e de demandas, estrutura organizacional, etc.). Os parâmetros situacionais constituem, portanto, a base sobre a qual se formulam as políticas que determinam e controlam o conjunto das operações a serem praticadas.

O ato documentário requer, por fim, metodologias que permitam construir os objetos que darão materialidade à Comunicação Documentária¹. Portanto, uma análise mais acurada dos processos documentários indicará que suas operações centrais - ler, selecionar e representar - apresentam problemas cuja complexidade não poderá ser explicada por enfoques ingênuos.

A Análise Documentária, em sua dimensão operatória, manipula e transforma textos em dois tipos básicos de representações: o resumo e o índice.

No resumo, o texto-base é desestruturado de modo a permitir que certos fragmentos, organizados sob a forma de um novo texto, reconstruam potencialmente o sentido do original. (GARCÍA GUTIÉRREZ, A. 1990)

Na indexação, procura-se obter um grau ainda maior de compactação do texto-base: do processo de segmentação resultam fragmentos que procuram caracterizá-lo por meio de palavras ou de sintagmas.

Se o resumo mantém, de fato, uma relação de contigüidade e de semelhança com o texto que lhe dá origem, já que procura reter a sua estrutura informacional, o mesmo não pode ser afirmado em relação ao produto da indexação. Neste último caso, instaura-se uma nova ordem de sentido, mediada por um instrumento comutador - a linguagem documentária - que generaliza as informações do texto, integrando-o a "classes" já previstas no referido instrumento.

A natureza da linguagem documentária e das unidades que a constituem e, por extensão, a natureza dessa modalidade de representação documentária já foram amplamente discutidas por Gardin (1974), e retomadas por Marilda Lara (1993).

Segundo a referida autora, as representações documentárias mediadas por linguagens documentárias "são de caráter generalizante", porque a indexação não opera com as informações particulares do texto. Segundo suas palavras:

¹ Segundo Lara, a "Comunicação Documentária [é um] processo que envolve a codificação e a decodificação de conteúdos informacionais, ou seja, o tratamento e a recuperação da informação." (LARA, M. L. G. 1993. p. 4)

"O código de intermediação funciona, nesse sentido, como um elemento para assegurar um rendimento informativo modal. A normalização obtida através do código documentário, embora tenha, por um lado, um caráter preditivo, impedindo a representação da subjetividade expressa nos textos originais, garante, por outro lado, a circulação de informações, revelando, dessa forma, seu caráter dinâmico." (LARA, M. L. G. 1993. p. 62)

Os termos da linguagem documentária são, portanto, unidades semânticas orientadas para a univocidade, já que o seu significado é delimitado por operações de fixação semântica. As informações documentárias submetidas a essas linguagens veiculam, desse modo, conteúdos prefixados.

Já se afirmou que a palavra é reversível porque ela assume todos e nenhum significado. Ao discutir as palavras e o seu significado, Ullmann, observa que:

"o caráter *genérico* de nossas palavras é uma das principais fontes de sua imprecisão. Com exceção dos nomes próprios e de um pequeno número de nomes comuns que se referem a objetos únicos, as palavras denotam não entidades singulares, mas classes de coisas ou de acontecimentos ligados por algum elemento comum." (ULLMANN, S. 1967. p. 133)

A natureza genérica das palavras é um fenômeno geral da linguagem, um **elemento de abstração da linguagem**. (ULLMANN, S. 1967. p. 134) Desse modo, as unidades da linguagem documentária, embora sejam fixadas semanticamente, pelo fato de designarem "classes", e não informações particulares, elas tendem a veicular informação que apresenta certo grau de imprecisão, à semelhança das palavras da linguagem natural.

Por essa razão, no âmbito pragmático da recuperação da informação, que opera com a palavra, a indexação e o resumo atuam de forma complementar. Normalmente, a primeira etapa de busca de informação faz-se por meio dos termos de uma linguagem de indexação. Em face dos resultados da pesquisa bibliográfica, que em geral fornecem um número significativo de documentos, identificados por um determinado termo ou conjunto de termos da linguagem utilizada, é necessário recorrer aos resumos. Estes últimos determinam a pertinência dos itens informacionais recuperados. É nesse contexto que se manifesta de forma evidente a necessidade de se garantir a correspondência entre informações textuais e representa-

ções documentárias, sejam eles resumos ou índices.

O processo global de Análise Documentária defronta-se, portanto, com pelo menos dois objetos:

- o Texto, como a sua unidade de análise;
- a Linguagem Documentária, como instrumento comutador.

Como abordar tais objetos? Qual é o *corpus* teórico necessário para discutir a Análise Documentária? Quais são os procedimentos necessários para se observar o documento? Que tipos de técnicas poderiam otimizar o processo de fabricação de representações documentárias?

Granger nos propõe a conduta global a ser adotada, ao afirmar que a radical descontinuidade dos paradigmas é, de certa forma, um mito, porque um modelo sempre se estabelece "... a partir de um momento anterior do pensamento...", pela "... ressonância do passado no presente." (GRANGER, G. 1989. p.25) Para o autor,

"Mesmo quando se pretende fazer tábula rasa dos sistemas que o precederam, este não se pode instituir sem levá-los em conta, mais ou menos explicitamente e de modo mais ou menos rigoroso." (GRANGER, G. 1989. p.25)

Granger nos indica, igualmente, a necessidade de delimitar os fatos específicos que se deseja observar:

"Não se poderia, com certeza, exigir das ciências uma definição universal do fato: o conhecimento científico, por sua natureza, repousa numa determinação específica e, por assim dizer, regional do fato e, em constante evolução, porque a cada momento de sua história, cada ramo do pensamento científico delimita a classe de fatos que quer explicar, com os meios materiais e conceituais dos quais pode dispor." (GRANGER, G. 1989. p. 13)

Orientados por esses dois requisitos, passaremos a examinar as referências que possam fornecer "régua e compasso" para traçar o percurso de construção de nosso "modelo abstrato dos fenômenos".

Sendo múltiplos os marcos teóricos e metodológicos a serem examinados, procuraremos evitar duas tentações: a) amalgamar perspectivas entre si irreconciliáveis; b) estabelecer equivalências entre conceitos homólogos ou mesmo homônimos ou assimilá-los de forma precipitada ao nosso campo específico de trabalho.

Nessa perspectiva, procuraremos, primeiramente, fixar alguns conceitos, com o fim de imprimir maior clareza às questões a serem abordadas neste trabalho. O primeiro deles é o conceito de representação documentária.

2.1 A representação documentária

A palavra "representação" evoca inúmeras noções. Para evitar os equívocos que a polissemia do termo possa causar, já que o mesmo é utilizado em contextos variados, procuraremos determinar alguns dos significados que a palavra assume, particularizando, em seguida, seu sentido específico, ao se referir aos processos da Análise Documentária.

O senso comum entende "representação" como reprodução daquilo que se pensa ou como substituição, descrição, e "representar", como o ato de reproduzir, descrever, tornar algo presente, interpretar, etc. (FERREIRA, A. B. H. 1990)

Na tradição filosófica, a palavra "representação" refere-se aos diversos modos de apreensão de um objeto. Assim, no sentido aristotélico, a representação é assimilada à fantasia intelectual ou sensível, enquanto, para os estóicos, ela está relacionada à impressão direta ou indireta. Para Descartes, ela é imaginação, para Spinoza, apreensão sensível, enquanto em Kant ela é apreensão intuitiva ou conceptual. No sentido da percepção, ela está presente em Leibniz, sendo a representação o mesmo que idéia, em Locke e Hume.

Segundo Ferrater Mora,

"A multiplicidade de aplicações do vocábulo representação o faz de uso incômodo, tanto em Filosofia quanto em Psicologia. É, pois, necessário, quando se fala de representação, especificar em que sentido se emprega o conceito." (FERRATERMORA, J. 1971. p. 567)

Desse modo, o mesmo autor sistematiza os diversos sentidos do termo "representação" da seguinte forma:

- representação enquanto apreensão de um objeto efetivamente presente e, deste modo, assimilada à percepção;
- representação enquanto reprodução de percepções passadas na consciência e, portanto, recordação;
- representação como imaginação;

- representação como alucinação.

Na Lingüística, "representação" é um conceito associado ao aparecimento da imagem verbal-mental, no falante. Desse modo, a linguagem está no lugar de outra coisa: "A função do signo é o de representar, de tomar o lugar de outra coisa, evocando-a a título de seu substituto." (BENVENISTE, E. 1974. p.51)

Na Semiótica, representar adquire um sentido bastante próximo ao do conceito lingüístico de representação, definido por Benveniste. Para Peirce, "representar" é

"estar em lugar de, isto é, estar numa tal relação com um outro que, para certos propósitos, é compreendido por alguma mente como se fosse a outra coisa." (PEIRCE, C. S. 1977. p. 61)

Na Ciência da Comunicação, "representação" é um conceito mediador entre o emissor e o receptor. É entendido, também, como o processo e o produto social da construção de sentidos, por meio de qualquer sistema de significação. (O'SULLIVAN et al. 1983)

Na Documentação, o termo "representação" é um conceito pré-teórico, associado, de um lado, à descrição de aspectos que identifiquem materialmente os documentos (catalogação) e, de outro, ao processo e ao produto da condensação de conteúdos de textos, ou seja, à indexação e à elaboração de resumos (processos) e aos próprios índices e resumos (produtos).

Quando da descrição dos aspectos materiais dos documentos, a identificação dos elementos pertinentes dá-se por apreensão instantânea. Já no interior da Análise Documentária, a representação resulta de procedimentos inferenciais mais complexos.

Gardin assimila a operação de construir representações documentárias aos processos realizados no interior das ciências que analisam e interpretam textos:

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

A informação documentária², ao ser abordada como algo resultante de operações de natureza semântica, permite que a observemos como uma forma de representação simbólica. Ela é, contudo, uma representação particular porque, sob as suas diversas formas de manifestação, deve procurar responder às exigências impostas pela recuperação da informação, mantendo, com o texto original, uma relação de similaridade.

Embora a informação documentária seja obtida, de um lado, pela neutralização do poder expressivo do texto e, de outro, seja moldada segundo regras previamente determinadas, prevalece a idéia de algo que, apesar de ser formalmente diferente do original (portanto, "representação"), é equivalente a ele, do ponto de vista do conteúdo informacional. Nesse sentido, a informação documentária é uma representação condensada, construída a partir de um objeto efetivamente presente - o documento.

Segundo Cassirer (CASSIRER, E. 1953-1957) a função de **representação**, é uma das três funções fundamentais da consciência, ao lado da função **expressiva** e da função **significativa**. Essas três funções correspondem, por sua vez, a três esferas: a função expressiva, à esfera do sensível; a função de representação, à esfera da intuição, da percepção das coisas; e a função significativa, à esfera do "sentido" ou "conceito".

Cada uma dessas funções se realiza, por sua vez, de modo mais típico, em uma forma simbólica, respectivamente, no mito, na linguagem e na ciência. Mas, no interior de cada uma delas, encontram-se elementos que são a manifestação das três funções fundamentais. Assim, na linguagem, há elementos "miméticos", característicos da função expressiva, como também dos elementos "puramente simbólicos", característicos da função significativa. (CASSIRER, E. 1953-1957)

Aceita a idéia de que o texto é uma forma de representação simbólica, deve-se aceitar, igualmente, o fato de que a Análise Documentária, ao operar com textos do domínio da ciência e da técnica, opera com objetos que trazem as marcas das várias funções da consciência: representação (mito), função expressiva (linguagem) e função significativa (ciência).

Além disso, se aceitarmos o princípio segundo o qual o homem não entra em contato direto com as

coisas, mas se apropria da realidade através da mediação de estruturas conceituais e lógicas, é lícito afirmar que a geração dos produtos da Análise Documentária - as representações documentárias - não escapa, também, a essa mediação.

2.2 A significação nos processos documentários

A compreensão dos procedimentos específicos da Análise Documentária supõe a discussão dos processos de significação, tais como são propostos pela Semântica. De fato, o reconhecimento da natureza semântica da Documentação, por parte de certos pesquisadores, propiciou o seu diálogo com as Ciências da Linguagem. (GARDIN, J. C. 1969; MONTGOMERY, C. A. 1972; SPARCK JONES, K., KAY, M. 1973)

O diálogo tem sido bastante profícuo, como observa García Gutiérrez (1990); em outras situações, tem sido acompanhado de frustrações. Jean-Claude Gardin, por exemplo, após várias experiências realizadas nas décadas de 60 e 70, chegou à conclusão de que a Lingüística e a Semiologia têm pouco a oferecer à Documentação. (GARDIN, J. C. 1989)

Atualmente, a apropriação conceitual e metodológica que a Análise Documentária vem realizando sobre os avanços alcançados pela Análise do Discurso e pela Lingüística Textual e a forte interseção que começa a se estabelecer entre a Terminologia e a Documentação (LE GUERN, M. 1989; NEDOBITY, W. 1985, 1986; GOMES, H. E. 1990; TÁLAMO, M. F. G. M. et al. 1992, 1992a, 1992b; CINTRA, A. M. M. et al. 1993), indicam que a prática novamente se aproxima.

Neste trabalho, nos limitaremos a abordar alguns aspectos da significação, no âmbito do texto e do código. A semântica do código permitirá discutir as relações de significação entre as unidades constitutivas das linguagens documentárias; a semântica textual apresenta, por sua vez, conceitos fundamentais para se operar com as apropriações parafrásticas que caracterizam o resumo.

As dificuldades da tarefa são previsíveis, pois trata-se de aproximar ciências conceituais, como o são

2. A informação documentária é aqui definida como a representação condensada do conteúdo informacional de documentos, cuja função básica é a de facilitar a circulação da informação e documentos nas várias esferas da atividade humana.

a Lingüística e a Semiologia, de uma disciplina com finalidades pragmáticas, como a Análise Documentária. Nesta última, os modelos teóricos e metodológicos fazem sentido apenas na medida em que são operacionalizáveis.

A primeira dificuldade colocada para os que procuram na Semiologia os referenciais que possam sustentar seus trabalhos é o fato de a mesma desdobrar-se em diferentes "escolas", como já observou Dascal (1978). Embora elas possam, grosso modo, ser agrupadas em duas correntes: a dos seguidores da Semiótica de Peirce e a dos seguidores da Semiologia de Saussure, cada uma delas não é, por sua vez, monolítica. (DASCAL, M. 1978. p. 19)

Semelhante configuração levou Dascal a afirmar que a Semiologia é uma ciência em estágio pré-paradigmático: ela não dispõe de um conjunto de conceitos, métodos e, sobretudo, de problemas e critérios para avaliar soluções que sejam comuns a todos os que se ocupam das questões relativas aos signos:

"A semiologia é sobretudo uma atitude, um método, mais que uma ciência, entendida como conjunto de teses aceitas consensualmente pelos pesquisadores, em um momento dado." (DASCAL, M. 1978. p. 58)

A Semiologia está longe, portanto, de ser uma ciência consolidada, no interior da qual a Análise Documentária possa buscar conceitos ou metodologias de aplicabilidade imediata.

Prosseguindo, examinaremos os princípios fundantes de cada uma das semiologias. O conceito fundamental da Semiótica de Peirce é a **semiose** - processo no qual qualquer coisa age como signo, isto é, significa (PEIRCE, C. S. 1977. p. 46), enquanto, para Saussure, o objeto da Lingüística, enquanto parte da Semiologia, é a língua,

"sistema gramatical que existe virtualmente em cada cérebro, ou mais exatamente nos cérebros de um conjunto de indivíduos, pois a língua não está completa em nenhum, e só na massa ela existe de modo completo." (SAUSSURE, F. 1973. p. 21)

Esses diferentes pontos de vista implicam, naturalmente, a atribuição de distintas tarefas para as semiologias propostas. Assim, para Peirce, a tarefa fundamental da Semiótica é a análise do funcionamento do signo, no ato individual da semiose, enquan-

to, para Saussure, a arbitrariedade do signo é o princípio semiológico fundamental. Esta noção saussureana de que os signos se estabelecem a partir de regras ou convenções, sendo imotivada a relação entre significante e significado, tem um poder explicativo que não deve ser negligenciado para a compreensão das relações de significação instituídas entre as unidades das linguagens documentárias.

A Linguagem Documentária, código comutador - conhecido também por outras denominações: Linguagem de Indexação, Sistema de Classificação, Linguagem de Informação, Listas de Cabeçalhos de Assuntos - normalmente composto por um conjunto limitado de termos, prescreve as formas de entrada e de busca a serem utilizadas pelo indexador ou pelo usuário, num sistema documentário.

Dada a sua função, as unidades constitutivas das Linguagens Documentárias são selecionadas de diversas fontes: Linguagens de Especialidade, Linguagem de Uso Corrente e Terminologias de Área.

Esse fato apresenta questões semânticas complexas para a estruturação das linguagens documentárias, porque elas têm como ponto de partida sistemas de significação em que as relações entre signos e referentes (ou entre significados e significantes) são de natureza bastante diversa. Na linguagem natural, o fenômeno da "plurissignificação" é a regra, enquanto na terminologia e nas linguagens de especialidade, embora em diferentes graus, predomina a univocidade de significação dos termos.

À primeira vista, linguagem documentária e terminologia apresentam características que as tornam intercambiáveis. Contudo, a terminologia tem como objetivo fundamental a seleção e criação de termos para noções de domínios específicos, e sua fixação por meio da definição. (FELBER, H. 1984) As linguagens documentárias operam, por sua vez, em campos do conhecimento especializado, para descrever e tornar recuperáveis os textos aí produzidos. Desse modo, a operação de construção de linguagens documentárias supõe, necessariamente, os instrumentos terminológicos como fontes de referência. (TÁLAMO, M. F. G. M. et al. 1992, 1992a, 1992b; CINTRA, A. M. M. 1993)

As terminologias não se confundem, portanto, com as linguagens documentárias, antes instauram-se como o universo referencial destas últimas. Neste

quadro, as linguagens documentárias operam como classificações intermediárias³ construídas na intersecção de ao menos três paradigmas designacionais: a linguagem de especialidade (núcleo de idéias), as classificações científicas e suas definições (terminologias de áreas) e a linguagem natural. (TÁLAMO, M. F. G. M. et al. 1992b; CINTRA, A. M. M. et al. 1993)

As linguagens documentárias não substituem, portanto, a linguagem especializada, as terminologias e a linguagem natural em sua função comunicativa geral, mas as primeiras - as linguagens documentárias - são mais adequadas que as demais⁴ na tarefa específica de tratar e recuperar informação.

As linguagens documentárias compartilham, com as terminologias, a característica de serem construídas para fins específicos, mas não apresentam a sua precisão; compartilham com a linguagem natural a função comunicativa, mas precisamente, para tornar eficaz essa função⁵, não comportam o poder expressivo desta última.

Até o momento, a análise da significação⁶ foi feita no âmbito do código⁷, procedimento que nos permitiu caracterizar as linguagens documentárias. A tarefa seguinte a ser empreendida, refere-se à significação engendrada ao nível textual. Um primeiro ponto de vista que merece ser discutido é apresentado por Kristeva: a idéia essencial da autora é a de que o "conteúdo informacional" ou "sentido" de um texto não pode ser nunca determinado completamente. Assim, o "resíduo" não explicável no âmbito da função comunicativa é o objeto sobre o qual deve-se elaborar uma estruturação suplementar⁷, por meio das categorias e técnicas da semanálise. O resultado desse trabalho é de fato o significado do texto. (KRISTEVA, J. 1974) A mesma questão é abordada por Granger, que propõe a distinção entre sentido e significação. O sentido é aquilo que é apropriado, estruturado e transmitido pela "grade de codificação" da experiência, que

é a língua. A operação de codificação deixa, por sua vez, certos resíduos que constituem precisamente o que este autor denomina de "significação". (GRANGER, G. 1974)

A Análise Documentária teria dificuldades intransponíveis para operar com semelhantes modelos. Basta lembrar que as operações documentárias com conteúdos de textos partem do pressuposto de que é possível identificar a informação principal dos mesmos, com o objetivo de construir representações que mantenham relação de semelhança com o original. Passaremos, pois, a discutir questões relacionadas aos textos e às várias possibilidades de análise dos mesmos, de modo a estabelecer um paradigma apropriado aos objetivos documentários.

2.3 O texto como objeto: algumas abordagens

Ao se definir a Análise Documentária como uma disciplina de natureza metodológica, que se preocupa com o tratamento do conteúdo de textos, torna-se inevitável sua comparação com outras disciplinas ou ciências que também operam com semelhante objeto. Trataremos, de forma breve, as abordagens que têm subsidiado a Documentação de forma mais direta, a saber: a Análise de Conteúdo, a Análise do Discurso, a Lingüística do Texto, a Análise Logicista e, por fim, a Análise de Textos segundo a Inteligência Artificial.

Embora todas essas áreas observem um mesmo objeto empírico, denominado discurso ou texto, as noções subjacentes aos termos utilizados variam de área para área, sendo igualmente diversos os objetivos, técnicas e instrumentos usados para analisá-lo, como veremos a seguir.

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

4 Um outro ponto de vista, que merece ser citado, é apresentado pelo pesquisador espanhol Antonio Garcia Gutiérrez, o qual propõe a "naturalização" das linguagens documentárias: "El language documental debe ser un language natural sintetizado y una taxionomia dinámica da la realidad." (GARCIA GUTIÉRREZ, A. 1990a. p. 78)

5 Uma interessante proposta de integração das linguagens documentárias e terminologias de área em um único instrumento - o tesouro-terminológico - é apresentada por Le Guern. (LE GUERN, M. 1989)

6 O termo "código" está sendo utilizado aqui no sentido de sistema convencional que se destina a representar e a transmitir informação. (DUBOIS, J. et al. p. 114)

7 A semanálise foi proposta por Kristeva, originalmente em 1969, na obra "Recherches pour une sémanalyse". Ela se coloca, antes de tudo, como uma possível base para a construção de uma semiótica geral, capaz de apropriar-se dos fenômenos da significação engendradas no texto. Partindo do pressuposto de que a significação ao nível textual é distinta daquela regida pelo signo, a semanálise introduz o texto como objeto de conhecimento da semiótica, instaurando-se, desse modo, como uma "ciência que constrói-se como uma crítica do sentido, de seus elementos e suas leis." (KRISTEVA, J. 1974. p. 20-25)

2.3.1 A Análise de Conteúdo

Por Análise de Conteúdo entende-se um conjunto de procedimentos de análise e interpretação de textos de natureza diversa (artigos de jornais, obras literárias, testemunhos, discursos políticos etc.) baseados na descrição e quantificação de palavras, temas ou frases desses textos. (GARDIN, J. C. 1987, p. 54-55; BARDIN, L. 1988, p. 31)

Segundo Henry e Moscovici:

"a Análise de Conteúdo visa não ao estudo da língua e da linguagem, mas à determinação daquilo que denominaremos de **condições de produção de textos**. Na Análise de Conteúdo, o que se procura caracterizar são as condições de produção e não os próprios textos. O conjunto das condições de produção constitui o campo de determinação dos textos." (HENRY, P., MOSCOVICI, S. 1968, p. 37)

Os procedimentos básicos da Análise de Conteúdo são normalmente esquematizados nas seguintes etapas: a) formulação de dados; b) redução dos dados; c) inferência; d) análise. (KRIPPENDORF, K. 1990) Deve-se observar que os dados aqui referidos não são "fatos" absolutos, mas elementos considerados representativos de um dado *corpus* (escritos pessoais, discursos políticos, artigos de imprensa, obras filosóficas, contos populares, questionários etc.), que são, por sua vez, submetidos a um tratamento formal que permita a sua manipulação.

O registro dos dados é um dos problemas metodológicos fundamentais dessa técnica, tendo em vista o princípio de que os fenômenos não são diretamente observáveis. Desse modo, as mensagens presentes nos *corpora* devem ser normalizadas, isto é, transcritas em termos formais, antes de serem processadas. Esse aparato descritivo para os quais se traduz os dados denomina-se linguagem de dados. Segundo Krippendorf, "Toda linguagem de dados é um sistema intermediário entre o mundo dos fenômenos reais e o dos fatos científicos..." (KRIPPENDORF, K. 1990, p. 124)

Para cumprir adequadamente suas funções, os elementos constitutivos dessa linguagem devem exibir duas propriedades fundamentais: serem exaustivos e mutuamente excludentes. A exaustividade refere-se à

capacidade da linguagem de dados para representar todas as unidades de registro consideradas significativas. A propriedade de "exclusão mútua" refere-se, por sua vez, à capacidade da linguagem para estabelecer distinções nítidas entre os fenômenos em análise. (KRIPPENDORF, K. 1990)

A linguagem de dados é, de fato, um código que dá forma ao universo objeto de análise. Enquanto código, é um

"recurso destinado a aumentar o rendimento informativo de uma mensagem" e, como tal, deve apresentar "um conjunto de signos e suas regras de utilização". (COELHO NETTO, J. T. 1980, p. 140)

Os signos de um código devem ser classes de entidades capazes de discriminar e representar, inequivocamente, os dados do objeto em análise. Em outras palavras: uma linguagem de dados não admite categorias ambíguas, porque estas podem comprometer a confiabilidade dos procedimentos descritivos e, por extensão, a interpretação do fenômeno em estudo.

A análise de Conteúdo, embora amplamente utilizada nas chamadas Ciências Sociais e no campo da Comunicação, é vista com restrição por muitos autores. Um de seus problemas cruciais, o calcanhar-de-aquiles, é precisamente a linguagem de dados. Henry e Moscovici (1968) já afirmaram que a Análise de Conteúdo peca por dois motivos: a) por não definir, com o devido rigor, as categorias utilizadas para descrever os dados coletados nos textos; b) pela falta de uniformidade desses dados, na medida em que eles podem ser palavras materialmente presentes nos textos ou unidades maiores de conteúdo (frases ou mesmo páginas), representadas pelas referidas categorias não definidas com rigor, como anteriormente apontado.

A sofisticação dos métodos estatísticos utilizados na Análise de Conteúdo não torna mais confiáveis os resultados obtidos. Segundo Gardin, tal sofisticação não foi acompanhada de mecanismos de representação suficientemente rigorosos, aptos a conferir credibilidade ao produto final da análise, ou seja, às suas interpretações. Nesse tipo de análise, invariavelmente, a constituição dos dados, sua agregação em classes ou categorias e sua correlação com outros dados ou teorias ficam comprometidas, por ser ambígua a linguagem de dados. (GARDIN, J. C. et al. 1987)

8 Paul Henry e Serge Moscovici afirmam: "preferimos a expressão 'condições de produção' àqueles outros comumente utilizados e que fazem referência seja à situação de comunicação, seja ao ambiente sócio-cultural, seja à psicologia individual. Estas últimas expressões têm todas o inconveniente de restringir arbitrariamente o campo das determinações possíveis dos textos, enfatizando este ou aquele aspecto particular." (HENRY, P., MOSCOVICI, S. 1963, p. 37)

Caracterizado, em grandes traços, o que venha a ser a Análise de Conteúdo, e apresentadas as críticas mais comuns a essa técnica, deve-se estabelecer a distinção entre a Análise de Conteúdo e a Análise Documentária. A exposição, acima apresentada, permite concluir que:

- os *corpora* da Análise de Conteúdo podem ser constituídos de textos de natureza diversa, desde os textos científicos até as manifestações estéticas, enquanto a Análise Documentária restringe seu campo de ação à análise de textos de especialidade, de natureza técnico-científica;

- a Análise de Conteúdo extrai dos textos fragmentos (palavras, temas e frases) que possam "revelar" opiniões e valores, ou seja, aquilo que está subentendido no texto; desse modo, a sua preocupação é com o extra texto, isto é, com as condições de produção do texto (HENRY, P., MOSCOVICI, S. 1968);

- a Análise Documentária, por outro lado, preocupa-se com a identificação da estrutura informacional dos textos, ou seja, com o próprio texto, a fim de elaborar representações condensadas que permitam ao leitor identificar seu conteúdo informacional.

Desse modo, embora exista um certo grau de isomorfismo entre as operações da Análise de Conteúdo e da Análise Documentária, os objetivos que cada uma persegue determinam a existência de diferenças fundamentais na natureza dos dados com os quais cada uma delas opera.

2.3.2 A Análise do Discurso e a Lingüística do Texto

Os estudos do discurso e do texto remontam aos anos 50, período no qual Harris realiza os trabalhos pioneiros da área. A partir daí, as pesquisas lingüísticas observam, de forma dominante, a palavra e a frase. O desenvolvimento mais amplo desses estudos, contudo, dá-se a partir dos anos 80. (VAN DIJK, T. 1992)

Uma primeira questão, logo reconhecida pelos que estudam o discurso, é o fato de que sua abordagem não se esgota em uma única disciplina. Desse modo, textos e discursos são objetos pluridisciplinares, cuja apreensão impõe a formulação de um quadro teórico para o qual concorrem o lingüístico, o literário e o

filosófico (BAKHTIN, M. 1984), o sociológico (PÊCHEUX, M. 1969) ou mesmo o psicológico. (KINTSCH, W., VAN DIJK, T. 1978)

O discurso - sendo definido como o lugar da subjetividade, a partir da oposição língua/fala - não se instaura como objeto observável, no interior da teoria saussureana. Dentre as inúmeras experiências de incorporação do discurso ao campo de investigação da Lingüística, destaca-se o modelo de Análise do Discurso, de Pêcheux. Nessa proposta, os discursos, por serem articulados no interior de condições de produção determinadas, são definidos como objetos que exibem traços estáveis de manifestação da ideologia.

Para Pêcheux, a Análise do Discurso é, de fato, a teoria que procura "apreender como no lingüístico se inscrevem as condições sócio-históricas de produção." (BRANDÃO, H. N. N. 1991. p. 83) Dito de outro modo, o modelo proposto por Pêcheux, inscreve-se no conjunto de trabalhos interdisciplinares sobre o discurso e as relações de confronto entre as classes sociais.

Essa perspectiva revela uma característica central da Análise do Discurso: a de operar em dois níveis distintos, de modo a

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

Também Adam, embora sob ótica distinta, afirma não ser possível analisar o discurso do ponto de vista puramente lingüístico:

"A Lingüística e a Pragmática textuais devem, portanto, definir um campo de pesquisa específico, dentro do domínio mais vasto do discurso porque outras disciplinas (como a História, a Sociologia, a Psicossociologia, a Psicanálise, etc.) têm, provavelmente, mais condições para descrevê-lo." (ADAM, J. M. 1989. p. 191)

Como se pode ver, é quase consensual a idéia de que o discurso deva ser abordado em sua dimensão social. Por outro lado, o campo de pesquisa específico, ao qual se refere Adam, é o do texto, e este não é um objeto empírico; ao contrário, é um objeto teórico depurado dos traços que caracterizam o discurso: "O texto é um objeto abstrato, resultante da subtração do

contexto sobre o objeto empírico (discurso)." (ADAM, J. M. 1989. p. 191)

Pode-se distinguir a Teoria do Texto, ou Lingüística do Texto, da Análise do Discurso, a partir da descrição das tarefas que cada uma delas se atribui. No caso da Lingüística do Texto, elas podem ser resumidas em dois pontos:

- determinar o que faz com que um texto seja um texto;
- diferenciar as várias espécies de textos. (FÁVERO, L. L., KOCH, I. G. V. 1988. p. 14)

Já a Análise do Discurso, segundo Diana Barros, deve considerar três questões, que são básicas para a concepção do discurso e sua análise:

- a relação do discurso com sua enunciação, condições de produção e de recepção;
- o discurso como o lugar, ao mesmo tempo, do social e do individual;
- a articulação entre narrativa e discurso, isto é, o discurso constituído sobre estruturas narrativas que o sustentam. (BARROS, D. L. P. 1988. p. 3)

Por maiores que sejam os apelos da Análise do Discurso, deve-se admitir que as diversas tentativas de incorporar seus procedimentos à Análise Documentária não foram bem sucedidas. Duas razões, ao menos, explicam os percalços:

- os dois níveis de análise (externo e interno), propostos pela Análise do Discurso, por serem destinados a desvendar a

"relação do falante com o processo de produção de frases (enunciação) ou a relação do discurso com o grupo social a que ele se destina (sociolingüística)" (DUBOIS, J. et al. 1992. p. 50-51),

não são aplicáveis aos objetivos da Análise Documentária. Esta última não se preocupa com as condições sócio-históricas da produção discursiva; ao contrário, seus objetivos restringem-se à identificação da base temática do texto;

- em segundo lugar, é preciso considerar que a noção de informação relevante de um texto varia de acordo com os objetivos que se persegue. Os procedimentos para selecioná-los são regidos, necessariamente, por essa conceptualização inicial. Desse modo, seria contraditório supor que os mecanismos utilizados no âmbito da Análise do Discurso

possam ser transferidos mecanicamente para a Análise Documentária. Uma experiência prática, realizada por Kobashi (1989), corrobora esse ponto de vista. No citado trabalho, discutem-se os problemas da proposta de Análise Documentária, de inspiração greimasiana, apresentada por Isabel Cunha (1990), cujo modelo não se mostrou capaz de dar conta da análise de textos para fins documentários, por duas razões: ao propor a identificação das informações principais de um dado texto, a partir do mapeamento dos **enunciados de estado** nele presentes, a referida autora baseou-se em uma compreensão equivocada do modelo greimasiano de narratividade. Como se sabe, o enunciado de estado é uma abstração, no sentido de mínimo semântico atribuível a enunciados que se "lexicalizam" de múltiplas formas; desse modo, eles não podem ser identificados apenas através dos verbos "ser" e "estar", materialmente expressos na superfície textual, como propôs Cunha; os fragmentos textuais (enunciados) que se apresentavam materialmente "lexicalizados" pelos verbos "ser" e "estar" não se caracterizaram como informações principais do texto.

A lição mais importante aprendida no percurso da referida experiência foi a de que a apropriação equivocada de conceitos de outras áreas pode produzir modelos inconsistentes. Desse modo, a apresentação de novas propostas deve ser precedida da compreensão rigorosa dos conceitos lógico-lingüístico-discursivos, com os quais se pretende operar.

Levando-se em consideração a possibilidade de isolar do objeto empírico **discurso** o objeto teórico **texto**, tal como é proposto por Adam (1989), procuraremos, a seguir, apresentar os conceitos essenciais relacionados ao objeto texto, discutindo, ao mesmo tempo, sua operacionalidade para fins documentários.

O texto, em uma primeira aproximação, define-se por sua "autonomia e seu fechamento". (DUCROT, O., TODOROV, T. 1972. p. 375) O quadro conceitual no interior do qual se discute o texto pode ser, grosso modo, classificado em duas vertentes: aquelas de natureza estritamente lingüística, centradas em fenômenos "frásticos", e a Pragmática textual, que procura captar, por sua vez, a dimensão global da textualidade, definindo o texto como um todo, e não como a reunião de proposições. É essa segunda vertente que passaremos a discutir.

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

- a condição de progressão, já que a ausência de

informação nova implica paráfrases perpétuas;

- a condição de coesão e de coerência, expressas pela reiteração e repetição de certos conteúdos que reaparecem regularmente ao longo do texto.

Na mesma perspectiva proposta por Ducrot, Adam afirma: "A textualidade pode ser definida como um equilíbrio delicado entre continuidade-repetição, de um lado, e progressão de informação, de outro." (ADAM, J. M. 1989. p. 209)

A coerência e a coesão, ao se referirem à permanência de categorias semânticas no texto, podem ser assimiladas à noção de isotopia proposta por Greimas. É importante observar que os conceitos de coesão, coerência e isotopia articulam-se à idéia de sentido global do texto.

Segundo Adam, "...o tópico ou tema do texto instaura-se como uma relação não-linear de coesão-coerência." (ADAM, J. M. 1989. p. 194-195) Ou ainda, segundo a formulação de Eco,

"...a isotopia pode ser vista como um fenômeno de coesão semântico-referencial, enquanto o tópico do discurso é um fenômeno pragmático, relacionado com a coerência e a pertinência." (ECO, U. 1983. p. 97)

O conjunto de noções, acima exposto, caracteriza o texto como uma unidade de sentido que apresenta redundâncias, reiterações e repetições, estando subjacente, portanto, a idéia de que esse texto poderá ser condensado, suprimindo-se as redundâncias, etc., sem que se descaracterize a sua informação central. Estas idéias serão retomadas mais adiante.

Como vimos anteriormente, a Linguística do Texto tem duas tarefas fundamentais: a primeira refere-se à determinação do que venha a ser um texto; a segunda, à caracterização das várias espécies de textos. Até aqui, discutimos a primeira tarefa; passaremos, agora, à segunda, às tipologias textuais.

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

que é central.

A superestrutura designa uma estrutura convencional. Assim, uma estória apresenta normalmente um esquema narrativo com as categorias convencionais: situação, complicação e resolução, enquanto o discurso científico apresenta, via de regra, um esquema constituído pelas categorias problema, hipótese, metodologia, resultados e conclusão. Enquanto paradigma de organização textual, o esquema ou superestrutura fornece uma base para a interpretação do texto.

No modelo em discussão, a compreensão é descrita, por sua vez, como um processo que supõe a transferência das unidades semânticas identificadas no texto para um esqueleto conceptual, derivado do esquema de conhecimento. O texto-base resultante é, em suma, o produto do casamento entre conhecimento do esquema (superestrutura) e informação textual. (KINTSCH, W., VAN DIJK, T. 1978 p. 48)

Para Kintsch e Van Dijk, a compreensão do discurso é de natureza estratégica. As estratégias, nesse modelo, são parte de nosso conhecimento geral, e como tal, precisam ser aprendidas e reaprendidas, antes de serem automatizadas. Nessa perspectiva, novos tipos de discursos e formas de conhecimento requerem o desenvolvimento de novas estratégias. (KINTSCH, W., VAN DIJK, T. 1978. p. 11)

O que distingue um bom de um mau leitor? O bom leitor, segundo esses autores, é aquele capaz de explorar o texto de forma mais direta, não se preocupando com a decodificação *bottom-up* (letra por letra, ou palavra por palavra). Em suma, bons leitores utilizam melhor os processos *top-down*, nos quais a superestrutura textual assume papel preponderante. (VAN DIJK, T., KINTSCH, W. 1983. p. 23)

O modelo acima apresentado permite elaborar, como hipótese, um esquema de Análise Documentária baseado, em primeira instância, no reconhecimento da superestrutura textual. Este esquema poderá ser o elemento capaz de imprimir rapidez e consistência à coleta e organização de informações para a elaboração de resumos e índices. Esta idéia será, por ora, apenas enunciada. Os conceitos relativos à superestrutura serão retomados, posteriormente, para subsidiar a formulação de procedimentos de elaboração de informações documentárias.

2.3.3 A Análise Logicista

A Análise Logicista comparece como objeto de discussão, ao lado das demais disciplinas que analisam textos, pelo seu grau de influência nas pesquisas realizadas no interior da linha de pesquisa Análise Documentária, da Escola de Comunicações e Artes da USP. Recordemos, apenas, que a produção teórica de Jean-Claude Gardin, tanto as obras referentes, especificamente, à Documentação (GARDIN, J.C. 1964, 1966, 1973, 1979) quanto aquelas mais abrangentes, dedicadas ao par Semiologia/Informática ou mesmo à Epistemologia Prática (GARDIN, J. C. 1974, 1987, 1987a) têm subsidiado, de forma sistemática, os trabalhos da linha acima referida. (SMIT, J. 1973, 1987; CUNHA, I. M. R. F. 1990; KOBASHI, N. Y. 1989; LARA, M. L. G. 1993)

A Análise Logicista, proposta por Gardin, é uma metodologia de análise de textos científicos produzidos pelas ciências ditas humanas. Segundo ele:

"... a expressão concreta de um raciocínio científico é o texto científico, no qual o autor expõe as operações do espírito que o conduziram, da observação de certos fatos empíricos, ao enunciado de proposições denominadas, de forma diversa: teses, hipóteses, interpretações, comentários, conclusões, explicações, etc." (GARDIN, J. C. 1987a. p. 4)

No modelo de Gardin (1987a. p. 25-26) os textos são objetos construídos, delimitados necessariamente pelos pontos de partida e de chegada, sendo os raciocínios (inferenciais e/ou dedutivos) "esquematzáveis" sob a forma de proposições descritivas e interpretativas. A esquematização - processo e produto da Análise Logicista - é, por sua vez, a representação condensada do objeto analisado, do qual foram mantidos tão somente os componentes e as articulações essenciais.

Deve-se observar que, na Análise Logicista, a representação faz-se por meio de um sistema simbólico (linguagem de representação distinta da linguagem natural e da linguagem do texto), cujo produto se expressa sob forma de proposições, que articulam dados e raciocínios do texto.

Certos aspectos desse modelo são análogos aos processos documentários de condensação e representação, mediados por um código comutador. Contudo, a Análise Logicista, ao se propor recuperar a lógica da criação de teorias e interpretações, no âmbito de cada ciência, a

partir da análise de sua produção textual, incorpora-se à Teoria do Conhecimento. O fato de Gardin denominá-la de Epistemologia Prática apenas reforça esse ponto de vista.

Nas palavras do próprio autor, a Análise Logicista ou

"... análise das construções não pretende ser uma ciência nova, mas um exercício de engenharia cognitiva, com um fim próprio: elucidar as organizações que sustentam certos textos ou certas classes de textos produzidos nas Ciências Humanas. À medida que se pode ver, nessas organizações, os 'fundamentos' lógico-semânticos das construções estudadas, a análise delas se aproxima da epistemologia, mas de uma epistemologia prática, aplicada, onde as conclusões não valem senão para os textos ou classes de textos considerados." (GARDIN, J. C. 1987. p.80-81)

Além disso, tendo em vista as exigências de explicitação de cada procedimento utilizado e de validação das inferências derivadas da análise das construções, a Análise Logicista é morosa e aplicável apenas a *corpus* restritos.

São essas duas características que estabelecem a clivagem entre as Análises Logicista e Documentária. Recordemos primeiramente que esta última analisa textos visando à sua condensação, sem se propor a avaliar a consistência ou validade das interpretações apresentadas no texto científico. Em segundo lugar, a Análise Documentária justifica-se, em larga medida, por sua capacidade de acompanhar *pari passu* a vasta escala de produção de textos científicos e técnicos do mundo contemporâneo. Suas técnicas específicas de manipulação de textos não comportam, pois, operações que comprometam a velocidade da análise para a produção de representações.

2.3.4 A Análise de Texto na perspectiva da Inteligência Artificial

A Inteligência Artificial (IA) pode ser definida como a área que se preocupa com o comportamento inteligente e com sua simulação computacional. Do ponto de vista prático, a IA procura descobrir formas de programar computadores para que os mesmos possam executar as funções que caracterizam a inteligência humana. O objetivo final da IA, portanto, é tornar os computadores mais úteis. (SHAPIRO, S. C. 1987)

A IA, enquanto campo de pesquisa autônomo, nasce na década de 50. Contudo, segundo Sabah (1990), as idéias básicas para tornar os computadores "inteligentes" já estavam presentes, em 1936, nos trabalhos de Turing, e em 1943, nos de McCulloch e Pitts. Porém, a consolidação da IA somente se dá a partir da década de 80, quando ela começa a se mostrar aplicável, de fato, aos diversos campos da atividade humana.

A IA é uma área interdisciplinar que se baseia tanto em técnicas computacionais avançadas - apoiadas em linguagens específicas - quanto na apropriação de conhecimentos gerados em outras áreas, em particular na Psicologia Cognitiva, na Lingüística e na Lógica.

A hipótese fundamental da IA, subjacente a qualquer sistema inteligente, é a de que os processos do pensamento são "mecanizáveis", simuláveis em computador. Entretanto, esta hipótese não é unânime. Com base no princípio de que, dada a sua complexidade, certos processos cognitivos (como a semântica) não são formalizáveis, Dreyfus (1972) e Winograd e Flores (1989) consideram que a IA não cumpriu suas promessas iniciais mais otimistas, e nem será capaz de cumpri-las.

Por outro lado, admitindo que o comportamento inteligente, embora seja domínio reservado ao homem, não se esvazia de sua substância ao ser formalizado, Sabah (1990) afirma ser prematura uma resposta definitiva para a hipótese inicial da IA. Considera, no entanto, que a I.A. está em posição extremamente favorável porque os trabalhos atuais procuram combinar as abordagens teóricas com atividades pragmáticas, confrontando os seus modelos com os dados empíricos a serem tratados. Desse modo, o computador é importante para o estudo do próprio cérebro, dada a sua capacidade de simular e manipular símbolos.

Nas pesquisas fundamentais sobre a IA, a linguagem - mais propriamente a linguagem natural - vem sendo importante objeto de experimentação. O tratamento automático da linguagem natural é uma questão complexa. Sua abordagem exige uma postura interdisciplinar, que promova a articulação dos conhecimentos de várias áreas interessadas na linguagem.

Quanto ao tratamento automático das línguas, em seu trabalho sobre a Inteligência Artificial e a

linguagem, Sabah (1990) afirma haver, pelo menos, três questões consideradas fundamentais. São elas:

- o que significa compreender a linguagem? De fato, essa compreensão pode ocorrer em diversos níveis, mobilizando-se, em cada um deles, diferentes tipos de raciocínios. No caso de textos, extrair seu sentido é, em certas circunstâncias, bem diferente de lhe atribuir significação. (SABAH, G. 1990. p. 21) Por outro lado, as pesquisas sobre o tratamento automático da linguagem estão, há longo tempo, limitadas à frase, embora se saiba que o sentido de um texto não advém da simples justaposição do sentido das frases que o compõem. As pesquisas em IA ressentem-se, na realidade, de métodos de representação e de análise globais do texto, que lhe permitam avançar na simulação da compreensão da linguagem.

- a segunda questão é simétrica à anterior: como um programa poderia reconhecer que algo não tem sentido, ou que seu sentido é ambíguo? Além disso, como um programa poderia distinguir entre uma metáfora e expressões metonímicas? Como poderia interpretar elementos novos?

Partindo do princípio de que não se fala sem um objetivo, um programa que realmente compreenda a linguagem deve ser capaz de identificar os objetivos do interlocutor e inferir seu plano, a fim de interpretar corretamente o que ele diz. Esta tarefa é extremamente complexa porque implica considerar os enunciados em situação de comunicação.

Segundo Sabah, a maior limitação dos programas de compreensão das línguas está no fato de os mesmos partirem da suposição de que um enunciado possa ter um sentido literal:

"...supõe-se a existência de um espaço de significações **S** e de uma correspondência em que, a todo elemento **p** de um subconjunto **P** de frases compreensíveis, associa-se um ou mais elementos de **S**: **s(p)** (o sentido ou os sentidos de **p**). Compreender **p** consiste, então, em construir esse elemento **s(p)**, que será denominado **representação interna** da frase. (Sendo a frase concebida como uma seqüência de palavras, a representação é geralmente apresentada sob a forma de um grafo, que liga os sentidos das palavras). Todo formalismo empregado em **S** (quer se trate das estruturas profundas do lingüista,

das fórmulas bem construídas dos lógicos ou das descrições componenciais⁹ do psicólogo) implica então a existência de uma relação de equivalência "ter o mesmo sentido que". Se isso parece ser lingüística e psicologicamente fundamentado (é possível, efetivamente, formular de modos diferentes uma mesma idéia), ele continua sendo discutível porque **o sentido de uma frase parece ser difícil de ser determinado a priori e deve, sobretudo, ser calculado em função do contexto onde essa frase aparece.**" [grifo nosso] (SABAH, G. 1990. p. 21-22)

Apesar dessas limitações, a noção de "representação interna" é hoje amplamente utilizada no tratamento automático das línguas, inclusive por Sabah (1990. p. 22), que procura conferir-lhe maior flexibilidade. Assim, a compreensão "automática" de um texto prevê, inicialmente, a construção de uma representação interna desse texto. Esta representação será, por sua vez, a base de diversas operações, entre outras: ensino assistido por computador, operações inteiramente automáticas de tradução, elaboração de resumos, indexação, recuperação de informação.

Como se pode observar, o tratamento da linguagem, no âmbito da IA, estabelece uma interface nítida com a Lingüística do Texto e a Análise do Discurso.

Essa interface tem despertado a curiosidade da Documentação porque oferece a possibilidade de ampliar as soluções para seus problemas específicos, particularmente aqueles relacionados aos métodos de criação de Bases de Conhecimentos.

O crescimento da oferta de textos integrais legíveis por máquina tende a ampliar tal interesse, já que, em princípio, a IA poderá proporcionar novos meios para explorá-los. Ela fez renascer o interesse da Documentação pela elaboração de sistemas pergunta/resposta (recuperação de dados factuais) gerados pela manipulação automática de textos integrais.

O projeto TOPIC (Text-Oriented Procedures for Information Management and Condensation of Expository Texts), já citado neste trabalho, é um exemplo de Base de Conhecimentos gerada automaticamente, com os recursos da IA. A pesquisadora Linda Smith refere-se, por sua vez, a

outros dois sistemas que, tal como o TOPIC, foram elaborados com base nos conceitos da IA: o RESEARCHER, que analisa e responde a perguntas sobre patentes e o RESEDA, base de biografias, de origem francesa. (SMITH, L. C. 1987)

As hipóteses de simulação do comportamento inteligente relacionadas às estratégias de aprendizagem e modificação do comportamento apresentam um apelo igualmente irresistível para a Documentação. Tais hipóteses permitem pensar em sistemas capazes de resolver, automaticamente, as complexas tarefas de modificar estratégias de busca ou reindexar documentos, em presença de novos dados.

Os melhores resultados de aplicação da IA relacionam-se, incontestavelmente, à automação de tarefas sobre as quais se tem domínio aprofundado.

A Catalogação, por ser uma área baseada em regras explícitas, vem sendo amplamente explorada para a concepção de Sistemas Especialistas capazes de monitorar a busca de formas de entrada de catálogos ou de fornecer pontos de acesso para regras relevantes do AACR2 (Anglo American Cataloging Rules 2).

A IA, contudo, não tem apresentado a mesma *performance* nas áreas relacionadas à análise e indexação de documentos, tendo em vista que esses procedimentos ainda estão pouco formalizados. Acredita-se que é longo o caminho a percorrer, até que as operações da Análise Documentária possam ser transformadas em regras de fato transponíveis para um ambiente de processamento computacional. (KUHLEN, R. 1984)

As experiências recentes da IA indicam que devemos ser cautelosamente otimistas. O otimismo é admissível pelo fato de haver um interesse renovado da IA pelos processos documentários. A cautela é recomendável porque muitos sistemas apresentados até o momento não são senão protótipos laboratoriais, sem possibilidades de aplicação em escala real.

Outro aspecto a ser considerado é o da impossibilidade de se transpor os procedimentos de uma área para outra, de forma mecânica.

A especificidade da Análise Documentária requer a elaboração de modelos próprios, articulados por recortes teóricos, operados em múltiplos domínios. Assim, os módulos de geração de resumos, desenvolvidos pela IA, para suas finalidades

⁹ No original, "descriptions componentielles". Nota do autor.

específicas, como os já conhecidos Cyrus e Frump (SCHANK, R. C., ABELSON, R. T. 1977), embora importantes, são apenas referências teóricas e metodológicas para a elaboração das condensações documentárias.

3. Considerações finais

Procuramos, neste texto, abordar alguns aspectos relacionados à organização e à representação de informações documentárias, na perspectiva da Análise Documentária, campo de conhecimento que opera com textos e, portanto, com a linguagem. Este fato impôs a discussão dos problemas de representação e de significação, desafio que buscamos enfrentar pela apropriação de conceitos construídos pelas ciências da linguagem.

Observamos, também, que a Análise Documentária se aproxima de outros tipos de análise de textos, razão pela qual apresentamos algumas perspectivas de abordagem do objeto-texto, com o fim de determinar a singularidade da Análise Documentária em face das demais disciplinas.

Nessa perspectiva, foi possível observar que, no âmbito da Análise Documentária, os textos são submetidos a um conjunto de operações que visam transformá-los em produtos documentários.

Essas operações não se propõem a gerar interpretações singulares, como a Análise do Discurso, ou a mapear e classificar raciocínios, como a Análise Logicista e sequer a modelar e simular os processos cognitivos, como o faz a Inteligência Artificial. A Análise Documentária não se propõe, também, a

identificar fenômenos lingüísticos, como a Lingüística do Texto, ou mesmo a identificar as condições de produção de textos, objeto da Análise de Conteúdo. A Análise Documentária propõe-se a elaborar diversas modalidades de informações, quer sob a forma de novos textos (resumos) ou de representações padronizadas, mediadas por códigos de conversão, ditos vocabulários controlados (tesauros, sistemas classificatórios). Ao analisar textos, o objetivo específico das operações documentárias é condensá-los, para promover sua circulação nas várias esferas das atividades humanas. Não se trata, portanto, de um mero conjunto de técnicas neutras.

Os problemas relacionados às operações documentárias são complexos, como se pôde observar ao longo deste trabalho. Tendo em vista as limitações das metodologias de Análise Documentária calcadas em modelos meramente estatísticos (que não foram capazes de oferecer soluções adequadas para as questões apresentadas), discutimos a possibilidade de construir novos procedimentos a partir da apropriação de conceitos elaborados pela Lingüística do Texto. No interior desta disciplina, as noções de textualidade e de superestrutura textual mostraram ser, em princípio, parâmetros que poderão ser utilizados para identificar e hierarquizar informações textuais, para os fins específicos da Documentação.

Esperamos, com esta discussão, ter contribuído para sistematizar algumas questões relacionadas à interface Lingüística/Ciência da Informação. É um passo que consideramos necessário para a constituição de um quadro teórico unificado da Ciência da Informação. Como afirmamos anteriormente, não se pode pretender propor algo novo sem a crítica dos conhecimentos construídos até o momento.

Abstract

This paper is a presentation of Documentary Analysis aspects. Documentary Analysis is a subject field related to information organization and retrieval processes. We characterize its most important theoretical and methodological frames of reference and the nature of its relationship with various disciplines that have the text as a research object (Content Analysis, Logicist Analysis, Discourse Analysis, Textual Linguistics, Natural Language Processing). Finally, we discuss the importance of the concepts of representation, of meaning and of text in the context of Documentary Analysis.

Bibliografia

- ADAM, J. M. Pour une pragmatique linguistique et textuelle. In: REICHLER, C. *L'interprétation des textes*. Paris: Minuit, 1989. p. 183-218.
- ARTANDI, S. Document description and representation. *ARIST*, v. 5, p. 144-167, 1970.
- BAKHTIN, M. *Esthétique de la création verbale*. Paris: Gallimard, 1984.
- BARANOW, U. G. Perspectivas na contribuição da Lingüística e de ciências afins à Ciência da Informação. *Ciência da Informação*, Brasília, v. 12, n. 1, p. 23-36, 1983.
- BARCELÓ, J. A. Programming an intelligent database in hypertext. In: COMPUTER APPLICATIONS AND QUANTITATIVE METHODS IN ARCHEOLOGY CONFERENCE, 17. /s. 1./ University of Oxford, 1991. (separata).
- BARDIN, L. *Análise de conteúdo*. Lisboa: Edições 70, 1988.
- BAR-HILLEL, Y. *Language and information*. Reading, Mass.: Addison Wesley, 1964.
- BARROS, D. L. P. *Teoria do discurso: fundamentos semióticos*. São Paulo: Atual, 1988.
- BELLY, N., BORILLO, A., SIOT-DECAUVILLE, N., VIRBEL, J. *Procedures d'analyse semantique appliqué à la documentation scientifique*. Paris: Gauthier-Villars, 1970.
- BENVENISTE, E. *Problèmes de linguistique générale*. Paris: Gallimard, 1974. v. 2.
- BORKO, H. *Subject analysis from a communication point of view*. Atlantic City, N.J.: American Library Association, 1969.
- BORKO, H., BERNIER, C. *Abstracting concepts and methods*. New York: Academic Press, 1975.
- BRANDÃO, H. N. N. *Introdução à análise do discurso*. Campinas: Ed. Unicamp, 1991.
- CASSIRER, E. *The philosophy of symbolic forms*. New Haven: Yale University Press, 1953-1957. 3 v.
- CINTRA, A. M. M. Elementos de Lingüística para estudos de Indexação. *Ciência da Informação*, Brasília, v. 12, n. 1, p. 5-22, 1983.
- _____. Estratégias de leitura em Documentação. In: SMIT, J. (Coord.). *Análise documentária: a análise da síntese*. Brasília: IBICT, 1987.
- CINTRA, A. M. M. et al. Do termo ao descritor: estudo exploratório. *Revista Comunicações e Artes*, São Paulo, 1993. (no prelo)
- COELHO NETTO, J. T. *Semiótica, informação e comunicação*. São Paulo: Perspectiva, 1990.
- CROSS, R. C., GARDIN, J. C., LÉVY, F. *L'automatization des recherches documentaires: um modèle général, le SYNTOL*. Paris: Gauthiers-Villars, 1964.
- CUNHA, I. M. R. F. *Do mito à análise documentária*. São Paulo: EDUSP, 1990.
- CUNHA, I. M. R. F., KOBASHI, N. Y. Análise documentária e Inteligência Artificial. *Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 24, n. 1/4, p. 38-62, 1991.
- DASCAL, M. *La sémiologie de Leibniz*. Paris: Aubier-Montaigne, 1978.
- DREYFUS, H. *What computers can't do: a critique of artificial reason*. New York: Harper & Row, 1972.
- DUBOIS, J. et al. *Dicionário de Lingüística*. São Paulo: Cultrix, 1992.
- DUCLOY, J. et al. *INIST'S: experience in hyper-document building from bibliographic databases*. Nancy: INIST, 1991.
- DUCROT, O. *Dire et ne pas dire*. Paris: Hermann, 1972.
- DUCROT, O., TODOROV, T. *Dictionnaire encyclopédique des Sciences du Langage*. Paris: Seuil, 1972.

- ECO, U. *Leitura do texto literário: lector in fabula*. Lisboa: Presença, 1983.
- EDMUNDSON, H. P. News methods in automatic extracting. *Journal of the ACM*, v. 16, n. 2, p. 264-285, 1969.
- FÁVERO, L. L., KOCH, I. G. V. *Linguística textual: introdução*. São Paulo: Cortez, 1988.
- FELBER, H. *Manuel de terminologie*. Paris: UNESCO/INFOTERM, 1984.
- FERRATER MORA, J. *Diccionario de Filosofia*. Buenos Ayres: Sudameris, 1971.
- FERREIRA, A. B. H. *Novo dicionário da língua portuguesa*. São Paulo: Nova Fronteira, 1990.
- FRANCFORT, H-P., LAGRANGE, M-S., RÉNAUD, M. *Palamede: application des systèmes experts à l'archéologie de civilisations urbaines protohistoriques*. Paris: CNRS: LISH/UPR 315. 1989. 2 v. (Document de travail, 9)
- GARCÍA GUTIERREZ, A. *Estructura lingüística de la documentación: teoría y método*. Murcia: Ed. Un. Murcia, 1990.
- GARDIN, J. C. *On some reciprocal requirements of linguistics and information techniques*. In: CONFERENCE INFORMATION IN THE LANGUAGE SCIENCES. Virginia: Airlie House, 1966. p. 69-93.
- _____. Artificial Intelligence and the future of Semiotics and archeological perspective. *Semiotica*, v. 77, n. 1/3, p. 5-26, 1989.
- _____. De l'analyse logiciste aux systèmes experts. In: GARDIN, J.C. et al. *Systèmes experts et Sciences Humaines: le cas de l'archéologie*. Paris: Eyrolles, 1987. p. 27-55.
- _____. Document analysis and linguistic theory. *Journal of Documentation*, v. 29, n. 2, p. 137-168, 1973.
- _____. L'analyse logiciste. In: GARDIN, J. C. et al. *Systèmes experts et Sciences Humaines*. Paris: Eyrolles, 1987. p. 17-26.
- _____. *La logique du plausible: essais d'épistémologie pratique*. 2. ed. Paris: Maison des Sciences de l'Homme, 1987.
- _____. *Le calcul et la raison*. Paris: EHESS, 1991.
- _____. *Les analyses de discours*. Neuchâtel: Delachaux et Niestlé, 1974.
- _____. On the relation between question-answering systems and various theoretical approaches to the analysis of text. In: CONFERENCE ANALYSIS OF MEANING. *Proceedings*. Oxford: ASLIB Informatics Group, BSC Information Retrieval Specialist Group, 1979. p. 206-220.
- _____. Semantic analysis procedures in the sciences of man. *Social Science Information*, v. 8, n. 1, p. 17-42, 1969.
- _____. *Systèmes experts et Sciences Humaines*. Paris: Eyrolles, 1987a.
- GARDIN, J. C., LAGRANGE, M. S. *Essais d'analyse du discours archéologique*. Paris: CNRS/ Centre de Recherches Archéologiques, 1975. (Notes et Monographies Techniques, 7).
- GARDIN, J. C. et al. *Le SYNTOL: étude d'un système général de documentation automatique*. Bruxelles: Presses Académiques Européennes, 1964.
- GOMES, H. E. (Coord.). *Manual de elaboração de tesouros*. Brasília: PNUB, 1990.
- GRANGER, G. *Filosofia do estilo*. São Paulo: Perspectiva, 1974.
- _____. *Por um conhecimento filosófico*. São Paulo: Contexto, 1989.
- GREIMAS, A. J. *Semiótica do discurso científico*. São Paulo: Difel/SBPL/s.d.l.
- _____. LANDOWSKY, E. *Análise do discurso em Ciências Sociais*. São Paulo: Global, 1976.
- GRICE, H. P. Logic and conversation. In: COLE, P., MORGAN, P. *Syntax and semantics 3: speech acts*. New York: Academic Press, 1975. p. 41-58.
- GRIZE, J. B. (Ed.). *Sémiologie du raisonnement*. Berne: Lang, 1984.
- GUIMARÃES, E. *A articulação do texto*. São Paulo: Ática, 1990.

- HAHN, V., REIMER, V. Heuristic text parsing in TOPIC. Methodological issues in a knowledge-based text condensation system. In: INTERNATIONAL RESEARCH FORUM IN INFORMATION SCIENCE, 5, 1985, Heidelberg. *Proceedings*. Amsterdam: Elsevier, 1984.
- HENRY, P., MOSCOVICI, S. Problèmes de l'analyse de contenu. *Langages*, v. 3, n. 11, p. 36-60, set. 1968.
- HUTCHINS, W. J. *Languages of indexing and classification: a linguistic study of structures and functions*. Herts: Peter Peregrinus, 1975.
- JACOBS, P. S., RAU, L. F. Natural language techniques for intelligent information retrieval. *Journal of ACM*, p. 85-99, 1988.
- KINTSCH, W., VAN DIJK, T. Toward a model of text comprehension and production. *Psychological Review*, v. 85, n. 5, p. 363-394, 1978.
- KOBASHI, N. Y. Análise documentária e tipologias discursivas. In: CUNHA, I. M. F. (Org.). *Análise documentária: considerações teóricas e experimentações*. São Paulo: FEBAB, 1989. p. 31-44.
- _____. Análise documentária: considerações sobre um modelo lógico semântico. In: CUNHA, I. M. F. (Org.). *Análise documentária: considerações teóricas e experimentações*. São Paulo: FEBAB, 1989. p. 45-57.
- KRIPPENDORF, K. *Metodología de análisis de contenido: teoría y práctica*. Barcelona: Paidós, 1990.
- KRISTEVA, J. *Introdução à semântica*. São Paulo: Perspectiva, 1974.
- KUHLEN, R. Some similarities and differences between intellectual and machine understanding for the purpose of abstracting. In: INTERNATIONAL RESEARCH FORUM IN INFORMATION SCIENCE, 5, 1985, Heidelberg. *Proceedings*. Amsterdam: Elsevier, 1984. p. 87-109.
- LARA, M. L. G. *A representação documentária: em jogo a significação*. São Paulo, 1993. Diss. Escola de Comunicação e Artes - USP.
- LE GUERN, M. Sur les relations entre terminologie et lexique. *Meta*, v. 34, n. 3, p. 340-343, 1989.
- LEHNERT, W. G. Narrative complexity on summarization algorithms. In: BARA, G., GUIDA, G. (Ed.). *Computational models for natural language and processing*. Amsterdam: North Holland, 1984. p. 247-260.
- LE ROUX, D. Automatisation de l'activité resumante: essai de typologie. In: COLLOQUE SUR LE RÉSUMÉ DE TEXTE. Pont-à-Mousson. 1990. (separata).
- LESKA, B. M. The connections between systems of concepts and thesauri. In: INTERNATIONAL SYMPOSIUM THEORETICAL AND METODOLOGICAL PROBLEMS OF TERMINOLOGY. *Proceedings*. Moscou, 1979.
- LEVY, P. *La machine universelle: création, cognition et culture informatique*. Paris: Éditions de la Découverte, 1987.
- LIDDY, E. D. The discourse-level structure of empirical abstracts: an exploratory study. *Information Processing and Management*, v. 27, n. 1, p. 55-81, 1991.
- LOSANO, J. et al. *Análisis del discurso: hacia una semiótica de la interacción textual*. Madrid: Ediciones Cátedra, 1989.
- LUHN, H. P. The automatic creation of literature abstracts. *IBM Journal of Research and Development*, v. 2, p. 159-165. 1958.
- _____. *Keyword-in-context for technical literature (KWICK Index)*. New York: ASDD Report IBM Advanced system Development Division, Yorktown Heights, 1959.
- MILSTEAD, J. L. Methodologies for subject analysis in bibliographical databases. *Information Processing and Management*, v. 28, n. 3, p. 407-431, 1992.
- MINSKY, M. *A framework for representing knowledge*. Cambridge, Mass.: MIT, 1974.
- MONTGOMERY, C. A. Linguistics and Information Science. *Journal of the American Society for Information Science*, v. 23, n. 3, p. 195-219, 1972.

- NEDOBITY, W. Terminology and Artificial Intelligence. *International Classification*, v. 12, n. 1, p. 17-19, 1985.
- _____. Classification, Artificial Intelligence, cognitive psychology. In: GAUL, W., SCAHDER, M. (Ed.). *Classification as a tool of research*. Amsterdam: Elsevier, 1986.
- NEWELL, A., SIMON, H. A. *Human problem solving*. Englewood Cliffs: Prentice Hall, 1972.
- O'SULLIVAN et al. *Key concepts in communication*. London: Methuen, 1983.
- PÊCHEUX, M. *Analyse automatique du discours*. Paris: Dunod, 1969.
- PEIRCE, C. S. *Semiótica*. São Paulo: Perspectiva, 1977.
- RONDEAU, G. *Introduction à la terminologie*. Québec: Gaëtan Morin, 1984.
- SABAH, G. *L'Intelligence Artificielle et le langage: representations des connaissances*. Paris: Hermès, 1988.
- _____. *L'Intelligence Artificielle et le langage: processus de compréhension*. Paris: Hermès, 1990.
- SALTON, G. *Automatic text processing: the transformation, analysis and retrieval*. New York: Addison-Wesley, 1989.
- SAUSSURE, F. *Curso de Lingüística geral*. São Paulo: Cultrix, 1973.
- SCHANK, R. C., ABELSON, R. P. *Scripts, plans, goals and understanding*. Hillsdale, N.J.: LEA, 1977.
- SCHANK, R. C., CHILDERS, P. G. *The cognitive computer on language learning and Artificial Intelligence*. New York: Addison-Wesley, 1984.
- SHAPIRO, S. C. (Ed.). *Encyclopedia of Artificial Intelligence*. New York: J. Wiley, 1987.
- SMIT, J. *Les langages documentaires comme metalangages du discours scientifique*. Paris: École Pratique des Hautes Études, 1973. Mémoire.
- _____. *De l'analyse documentaire à l'analyse poétique: a propos des critiques des "Chats" de Baudelaire*. Paris: École Pratique des Hautes Études en Sciences Sociales, 1976. Thèse de 3^e. Cycle.
- _____. (Coord.). *Análise documentária: a análise da síntese*. Brasília: IBICT, 1987.
- _____. Seshat et l'analyse poétique: a propos des "Chats" de Baudelaire. In: GARDIN, J. C. et al. *La logique du plausible: essais d'Épistémologie pratique*. 2. ed. Paris: Maison des Sciences de l'Homme, 1987. p. 103-144.
- SMITH, L. C. Artificial Intelligence and information retrieval. *ARIST*, v. 22, 1987.
- SPARCK-JONES, K., KAY, M. *Linguistics and Information Science*. New York: Academic Press, 1973.
- _____. Linguistics and Information Science: a post script. In: WALKER, D. E., KALGREN, H., KAY, M. (Ed.). *Natural language and Information Science*. Stockholm: Skriptor, 1977. p. 183-192. (FID Publication 551).
- _____. Problems in the representation of meaning in information retrieval. In: CONFERENCE ANALYSIS OF MEANING. *Proceedings*. Oxford: ASLIB Informatics Group, BSC Information Retrieval Specialist Group, 1979. p. 193-201.
- TÁLAMO, M. F. G. M. *Elaboração de resumos*. São Paulo: Escola de Comunicações e Artes, 1987. 14p. [datilografado].
- TÁLAMO, M. F. G. M. et al. A contribuição da terminologia para a elaboração de tesaurus. *Ciência da Informação*, v. 21, n. 3, p. 197-200, 1992.
- _____. Instrumentos de controle terminológico: limites e funções. In: SIMPÓSIO LATINO-AMERICANO DE TERMINOLOGIA, 2. *Anais*. Brasília: IBICT; Paris: União latina. 1992a.
- _____. *A interface análise documentária, lingüística documental e terminologia*. In: SIMPÓSIO IBEROAMERICANO DE TERMINOLOGIA, 3. 1 a 5 de dezembro de 1992b. Espanha.
- ULLMAN, S. *Semântica*. Madrid: Aguillar, 1967.
- VAN DIJK, T. *Cognição, discurso e interação*. São Paulo: Contexto, 1992.
- VAN DIJK, T., KINTSCH, W. *Strategies of discourse comprehension*. Orlando: Academic Press, 1983.

- VITALI, V., LAGRANGE, M. S. An expert system for the provenance determination of archeological ceramics based on INAA. In: CONFERENCE OF THE COMPUTER APPLICATION AND QUANTITATIVE METHODS IN ARCHEOLOGY. *Proceedings*. Birmingham: Birmingham University. Computer Center. 1988.
- WINOGRAD, T. *Understanding natural language*. New York: Academic Press, 1972.
- WINOGRAD, T., FLORES, F. *L'Intelligence Artificielle en question*. Paris: PUF, 1989.
- WITTY, F.J. The beginnings of indexing and abstracting: some notes toward a history of indexing in the antiquity and middle ages. *The Indexer*, v. 8, n. 4, p. 193-198, 1973.