

Aline Santos Jacob

Representação da informação: da linguagem gestual às linguagens documentárias

Monografia apresentada ao Departamento de Ciência da Informação e Documentação (CID) da Universidade de Brasília (UnB) como pré-requisito para a obtenção do título de bacharel em Biblioteconomia.

Orientadora: Professora Doutora Marisa Bräscher Basílio Medeiros

Brasília
2009

Dados Internacionais de Catalogação na Publicação (CIP)

Jacob, Aline Santos, 1982-.

Representação da informação: da linguagem gestual às linguagens documentárias. Brasília, dezembro 2009.

44 p.

Orientadora: Professora Doutora Marisa Bräscher Basílio Medeiros

Monografia em Biblioteconomia e Ciência da Informação - Universidade de Brasília, Departamento de Ciência da Informação e Documentação.

1. Representação da informação. 2. Linguagens documentárias.

Dedico este trabalho aos meus queridos pais que, muitas vezes até com certa dificuldade, sempre se esforçaram para me proporcionar estudo e mostrar o valor deste e, mesmo que por muitas vezes incompreendidos, nunca desistiram do sonho de ver todos os 3 filhos formados no ensino superior. Realizo, por fim, orgulhosa, o sonho de longos anos.

AGRADECIMENTO

Agradeço primeiramente e mais do que a quaisquer outros aos meus amados e incansáveis pai e mãe que acima de tudo me ensinaram a ser “Ser Humano”, e em nenhum momento duvidaram da minha capacidade. A esses seres iluminados devo tudo o que sou hoje.

À minha orientadora Marisa que, sensível e compreensiva, soube respeitar meu ritmo de trabalho.

Ao Fábio, pelas palavras amigas e por sempre se mostrar muito disposto a me ajudar.

Às minhas irmãs e cúmplices Michelle, Milena e Mariana pela companhia, mesmo que à distância, nas madrugadas afora, pelas correções feitas e sugestões dadas, além do apoio incondicional nos momentos difíceis.

Agradeço também aos chefes de estágio e amigos de curso que tiveram toda a paciência e compreensão quando, por vezes, ao longo da redação desta monografia, estive ausente na realização de outros trabalhos.

“Somente o desconhecido deve ter interesse para você, porque isso é o que você ainda não viveu; você ainda não se movimentou neste território. Mova-se! Algo de novo pode acontecer ali. Sempre decida pelo desconhecido, independente do risco, e você crescerá continuamente. Mas permaneça decidindo pelo conhecido e você irá se mover num círculo com o seu passado, de novo e de novo. Você irá continuar a se repetir, você se tornará um disco arranhado.”

Bhagwan Shree Rajneesh (Osho)

RESUMO

O presente estudo aborda algumas questões sobre a linguagem como capacidade humana essencial para a o processo de comunicação que permite a produção do objeto básico da Ciência da Informação. As formas de representação da informação do primitivo, com a linguagem gestual e oral, até os dias atuais com as linguagens documentárias serão brevemente elucidadas. Como a Ciência da Informação é uma área interdisciplinar este trabalho investiga também as interfaces existentes entre esta e uma das ciências às quais ela recorre como forma de trabalhar os problemas de tratamento, disseminação e recuperação da informação provocados pela chamada explosão bibliográfica em meados do século XX, a ciência lingüística.

Palavras-chaves: Representação da Informação; Lingüística Documentária; Análise Documentária; Linguagem Documentária

ABSTRACT

This study will board some questions about language as like as human capacity essential to the process of communication that allows the production of the basic object of Information Science. The forms of information representation since primitive age, with sign language and oral, until present, with the indexing languages will soon be elucidated. As information science is an interdisciplinary area this work will also investigate the interfaces between this and one of sciences which the information science call upon as way to work on problems of processing, dissemination and retrieval of information caused by the explosion called literature in the mid-twentieth century, linguistic science.

Key-words: Representation of Information; Documentary Linguistics; Documentary Analysis;
Documentary Language

LISTA DE SIGLAS

ABNT – Associação Brasileira de Normas e Técnicas

AD – Análise Documentária

CDD – Classificação Decimal de Dewey

CDU – Classificação Decimal Universal

CI – Ciência da Informação

LC – Classificação da Library of Congress

LD – Linguagem Documentária

LN – Linguagem Natural

SRI – Sistema de Recuperação da Informação

SUMÁRIO

1. INTRODUÇÃO	10
2. OBJETIVOS	12
2.1 Objetivo Geral	12
2.2 Objetivos Específicos.....	12
3. CONSIDERAÇÕES SOBRE A LINGUAGEM	13
3.1 Evolução das formas de linguagem.....	15
3.1.1 Linguagem gestual e linguagem oral	15
3.1.2 Linguagem escrita	16
3.2 Língua.....	18
3.3 Signo lingüístico	20
3.4 Relações sintagmáticas e paradigmáticas.....	22
3.5 Categorização.....	23
4. CIÊNCIA DA INFORMAÇÃO.....	25
5. ANÁLISE DOCUMENTÁRIA	28
5.1 Indexação.....	31
6. LINGUAGEM DOCUMENTÁRIA.....	34
7. CONCLUSÃO.....	38
REFERÊNCIAS BIBLIOGRÁFICAS	39

1. INTRODUÇÃO

A linguagem como capacidade humana permite a apreensão dos objetos e das idéias de mundo e a representação deste possibilita que o homem exprima suas idéias de maneira que seja compreendido pelos demais, além de possibilitar a produção de informação, entendida como conhecimento registrado.

Desde que o homem adquiriu a capacidade de se comunicar e de produzir, por meio desse processo, o conhecimento, há uma necessidade natural de criação de métodos capazes de passar esse conhecimento adiante seja oralmente, seja sob a forma escrita. As formas de representação da linguagem se modificam e se ajustam de acordo com o contexto histórico e cultural em que estão inseridas e acompanham a evolução humana. Assim, várias formas de representação da linguagem foram criadas e passaram da arte pictográfica à escrita alfabética como reflexo da evolução por que passou, e ainda passa, a humanidade.

É, contudo, a partir da evolução para um tipo de escrita que pudesse estar acessível todos que o progresso humano foi impulsionado, já que, com o acesso ao que já foi descoberto anteriormente por outros temos a possibilidade de complementar ou até mesmo reformular idéias e ajustá-las ao nosso favor. Portanto, produzimos e buscamos por informação a fim de utilizá-la para conferir nosso bem-estar e preservação de nossa memória.

Desde a invenção da imprensa por Gutenberg, grande massa de informações vem sendo disponibilizada e, durante muito tempo, a preocupação com as formas de armazenamento e recuperação dessas informações foi desconsiderada.

Entretanto, devido à crescente produção e busca por informações, principalmente nas áreas científica e tecnológica, com a chamada explosão bibliográfica ocorrida após a 2ª Guerra Mundial, torna-se mais evidente a necessidade de discutir questões relacionadas à recuperação da informação. Já nesta época, se fazia necessária a criação de métodos eficazes para a organização do conhecimento registrado, sendo esses indispensáveis para tratar dos problemas relacionados à recuperação desse conhecimento. Esses problemas são, em grande parte, amenizados com o uso das linguagens documentárias. Para a formulação e aplicação destas linguagens construídas, a Ciência da Informação toma emprestado, dentre outras ciências, conceitos relacionados à área da Ciência Lingüística.

Pretende-se com este trabalho compreender sinteticamente as questões ligadas à linguagem como forma de representação do pensamento humano e da informação produzida

nesse processo e, ainda, como ela foi evoluindo até chegar às linguagens documentárias atuais. Por fim, apresentam-se as contribuições que a Ciência Lingüística oferece à Ciência da Informação para o tratamento, a representação e a recuperação da informação.

O trabalho está dividido em sete capítulos, dos quais quatro se referem ao desenvolvimento do tema. O capítulo três traz algumas definições de linguagem como o processo cognitivo humano advindo da necessidade de comunicação. Assim, para demonstrar que esta capacidade evolui para suprir essa necessidade de comunicação, que também passa por um processo de evolução devido a um contexto social e cultural, o capítulo aborda também a evolução da linguagem, da forma gestual até a escrita alfabética. A linguagem escrita é determinante para que, bem mais tarde, se institua uma ciência para estudar o fluxo informacional contido em documentos, a Ciência da Informação. Além disso, nesse mesmo capítulo, serão elucidados conceitos lingüísticos aplicáveis à CI.

O capítulo quatro é destinado à contextualização da Ciência da Informação. Para tanto, além de apresentar o contexto histórico de seu surgimento, coloca definições de estudiosos da área sobre o que é, do que trata essa ciência, além de abordar seu caráter interdisciplinar a fim de que se demonstre, por fim, a interface desta ciência com a ciência lingüística.

Como no trabalho pretende-se investigar também os processos envolvidos na representação da informação documentária, o capítulo cinco aborda a análise documentária como processo indispensável para a organização da memória documentária além de discorrer sobre a atividade de indexação considerada também como sendo fundamental para a questão da representação da informação no atual contexto.

Finalmente no capítulo seis são identificados conceitos e aplicações das linguagens documentárias.

2. OBJETIVOS

2.1 Objetivo Geral

Este trabalho visa compreender as formas de representação da informação envolvidas no exercício da linguagem e na Ciência da Informação.

2.2 Objetivos Específicos

- Estudar os fundamentos básicos da linguagem, bem como suas formas de evolução.
- Compreender alguns conceitos lingüísticos aplicáveis à Ciência da Informação.
- Descrever as maneiras de análise e representação da informação no contexto da Ciência da Informação.
- Identificar conceitos e aplicações das linguagens documentárias.

3. CONSIDERAÇÕES SOBRE A LINGUAGEM

A linguagem é uma capacidade que passa por um processo evolutivo e cresce na medida em que evolui também o sujeito que a produz e utiliza: o homem. O desenvolvimento humano e o desenvolvimento da linguagem são complementares. Em Filosofia da comunicação e da linguagem, D'Amaral (1977, p. 33) afirma que

Coexistência é a presença do homem na história. O mundo, como lugar do homem, não é um espaço, mas uma duração. A exploração da essência do homem aponta, portanto, para historicidade e a mundaneidade do existir. A linguagem, lugar em que a coexistência se revela, é, portanto, o fundamento da história.

A capacidade de se comunicar não é única dos seres humanos e a linguagem, enquanto faculdade que permite a interação com o mundo, também não é exclusiva da espécie humana. Porém, é exatamente ela que difere o homem dos outros animais. Diferentemente das outras espécies, a humana é capaz de se comunicar por meio de uma linguagem articulada, ou seja, somos capazes de decompor a linguagem em unidades menores que, combinadas entre si, podem remeter a milhares de objetos. Esta capacidade é o que nos permite um enorme estoque de informações e, por isto também, somos capazes de produzir conhecimento. Mas não é possível que exista produção de conhecimento, uma apreensão real do mundo sem a capacidade de traduzir aquilo que se vê. A linguagem, portanto, é antes o processo cognitivo que permite a interação humana com aquilo que vê e no contexto em que se insere. É por meio dela que o ser humano organiza e dá nome às suas experiências. Por meio dela também somos capazes de nos relacionar com o mundo e fazer inferências e enunciados verdadeiros sobre os objetos que o circundam. (DAHLBERG, 1978)

Mas como surgiu a linguagem? Em que contexto esse ato se torna possível e preponderante para a história da evolução humana?

Conforme Sapir (apud Lyons, 1981, p. 17): “A linguagem é um método puramente humano e não instintivo de se comunicarem idéias, emoções e desejos por meio de símbolos voluntariamente produzidos”. Assim, a linguagem, esse ato humano para captar e transmitir idéias surge com o convívio social e se desenvolve na mesma medida em que se faz necessário interagir com o grupo.

Com a vida em grupo e o desenvolvimento do trabalho multiplicam-se os casos de ajuda mútua e de atividade conjunta – desde os nossos passados simiescos os homens já viviam em congregados. A vida coletiva gerou a necessidade de dizer algo uns aos outros. (KAJI-MARKENFELDT, 2006, p. 20)

A partir desta convivência em conjunto faz-se cada vez mais necessária também a delimitação de regras e a compreensão destas, algo que só é possível por meio da linguagem.

À medida que a humanidade evolui e as relações sociais, seja entre dois indivíduos ou mais, se tornam cada vez mais presentes e necessárias à sobrevivência, revela-se uma necessidade de perpetuação da história como sendo a perpetuação do próprio homem. Mas ultrapassar o tempo não é um processo fácil, pois que tudo se dá devido a um contexto social, cultural e histórico. O que pode ser compreendido por uma sociedade de determinada época, pode não mais fazer sentido para gerações futuras. É por isso que, com o passar dos tempos, o homem naturalmente procurou ser entendido não só pelo seu contexto, mas pelas gerações que o seguiriam.

É devido à existência da linguagem que se torna possível a comunicação, esta entendida como um processo de transferência de informação. Conforme Belkin (1978), esse processo ocorre a partir da decisão de um indivíduo (gerador) de comunicar um aspecto do seu estado de conhecimento por meio de uma estrutura comunicável (texto). Do outro lado, existe um receptor com uma necessidade latente de informação, buscando satisfazê-la por meio do uso de informação registrada numa estrutura comunicável. É a partir dessa relação estabelecida entre o sujeito gerador e o receptor que se identifica um ciclo de comunicação.

A comunicação humana somente será durável se registrada em suportes documentários, caracterizando assim um documento. Especificamente, a comunicação escrita em texto e registrada em suportes documentários é objeto da leitura e das atividades documentárias (FUJITA, 2004). É exatamente nesse ponto que a Ciência da Informação se liga intimamente à questão da linguagem. Tendo como objeto de estudo a informação, seu fluxo e transferência, essa ciência baseia-se na existência de um suporte de comunicação, a própria linguagem e em um suporte de comunicação escrita, o documento.

3.1 Evolução das formas de linguagem

Para garantir que sua história estivesse acessível às gerações futuras, o homem buscou aprimorar os sistemas de comunicação desenvolvendo para isso sistemas universais de significação. Ou seja, cria possibilidades em escala evolutiva para que o conhecimento

produzido por ele possa ser utilizado por seus descendentes. Essa evolução das formas de linguagem será abordada nas seções que se seguem.

3.1.1 Linguagem gestual e linguagem oral

“Afirma-se que no começo era o verbo. Não é esta a cronologia. No começo era o grito, depois o gesto e, em seguida, o canto.” (MELLO, 1972, p. 19)

Antes de desenvolver sistemas complexos de significação foi preciso recorrer a formas de expressão mais simplificadas. A utilização dos gestos, dos sons e até do próprio corpo foi, durante a maior parte da nossa história, a forma utilizada para manifestar uma necessidade latente de comunicação e de transposição de nossa história e pensamentos. Os rituais asseguraram, por longo tempo, a preservação da memória humana e o corpo serviu como suporte de informação quando o homem ainda não detinha os suportes usuais. Afirma Katzenstein (1986, p. 10) que

Muito antes de o homem ser capaz de falar e escrever, ele comunicava suas experiências interiores, pensamentos e sentimentos não-verbais, de um complexo de gestos - movimentos de todo o corpo ou suas partes -, por meio do olhar, do silêncio [...] Cada membro de um grupo entendia o significado dos gestos, palavras, ruídos e odores dos sacrifícios, o perfume do incenso que acompanhavam seus rituais .

Sem a posse de um sistema de signos capaz de expressar suas idéias e registrá-las em um suporte concreto capaz de perpassar a dimensão tempo-espço, durante séculos, a comunicação humana e a conservação da memória foram moderadas pela tradição ritual e oral. Conforme Mc Garry (apud DODEBEI, 2002, p. 36) “para facilitar a memorização e a evolução, a tradição era conservada na poesia ou na prosa rítmica e os mitos e lendas eram usados como uma espécie de enciclopédia tribal.”

Historiadores e cientistas ainda não concordam sobre qual forma de linguagem teria se originado primeiramente, se oral ou gestual, e talvez nunca cheguem a uma resposta, haja vista a impossibilidade de datar o que não pode ser registrado. O que não pode ser datado não pode ser colocado em escala cronológica. Como afirma Bordenave (1982, p. 23) “a comunicação humana tem um começo bastante nebuloso. Realmente não sabemos como foi que os homens primitivos começaram a se comunicar entre si, se por gritos ou grunhidos, como fazem os animais, ou se por gestos, ou ainda por combinações de gritos e gestos.”

Contudo, sabe-se que estas duas formas de comunicação foram os meios utilizados pela humanidade para se comunicar antes que a escrita pudesse tomar o seu lugar na formação da história.

3.1.2 *Linguagem escrita*

Em um passado distante a necessidade primordial do homem era a de comunicar. Com o tempo veio a sofisticação dos métodos e a comunicação, apenas como ato, já não era capaz de suprir o espírito humano enquanto ser que busca aperfeiçoar-se e superar-se a cada dia. Ao vislumbrar sua condição de ser produtor de conhecimento o homem observa que esse conhecimento pode ser utilizado por outros de sua espécie. Surge então a necessidade de registro da comunicação. Imprimir suas capacidades e descobertas em um meio capaz de transpassar o tempo e não perecer com ele substitui aquela necessidade primária de comunicar-se apenas por meio de fala e/ou gestos. Ele começa então a inscrever seu espírito e sua história para que fiquem marcados.

Os registros que materializam os conhecimentos vêm sendo gravados, utilizando-se da enorme gama de suportes físicos, que se tornaram disponíveis ao longo da evolução cultural do homem que se estende dos registros encontrados nas cavernas aos registros digitais de nossos dias (ALVARENGA, 2003, p. 5)

A maior parte da história do homem foi contada na forma oral e, devido a isso, grande parte dela se perdeu dada a dificuldade de transmitir, por longo tempo, o conhecimento armazenado sob esta forma. Esse modo de transmissão e conservação da memória demanda maior vontade e atitude dos indivíduos em passar e manter viva a história de seu grupo, ao passo que, o conhecimento registrado sob a forma escrita, pode tornar-se mais evidente e, portanto, ter sua transmissão facilitada.

Deste modo, o grande marco, a invenção que permitiu que o homem pudesse se tornar efetivamente independente foi a invenção da escrita, e independência exprime, de certo modo, poder. Mais do que isso, a escrita inicia uma nova era na história. Por meio dela o homem achou uma solução para o problema da distância. Por mais impressionante que a comunicação por meio da fala ou dos gestos possa parecer, e ser realmente, a criação de um sistema capaz de perdurar mesmo com o passar dos tempos, a escrita, é realmente o que permitiu e permite ainda hoje que o homem seja manipulador de sua própria história. Como afirma Bordenave

(1982, p. 26): “Apesar da riqueza, a linguagem oral sofre duas sérias limitações: a falta de alcance e a falta de permanência. Daí o fato de que os homens tenham apelado a modos de fixar seus signos e a modos de transmiti-los a distância”. Assim, a linguagem escrita surgiu como resultado da necessidade de contato e de conhecimento entre as comunidades separadas pelo tempo e pelo espaço e, como afirma Saussure (1969), a única razão de ser da escrita é a de representar a linguagem.

O desenvolvimento da capacidade humana de apreensão das idéias de mundo se configura de tal maneira que as formas de colocar à disposição o conhecimento gerado por essa apreensão não acompanha, na mesma escala de tempo e rapidez, a produção de mais informações. A produção do objeto essencial da escrita, a informação, cresce na mesma proporção em que se desenvolve a capacidade humana de utilizar e transformar a natureza a seu favor. Esse crescimento exponencial de informação produz uma necessidade que não existia com a linguagem oral, a de armazenar o conhecimento produzido. Conforme Le Coadic (1996, p. 6):

Com o advento da escrita, a informação passou de oral a escrita. Isto teve por conseqüência, por um baixo custo energético, multiplicar a informação (cópia de manuscritos, imprensa, fotocópia) e memorizá-la, permitindo assim exteriorizar, uma das funções do cérebro humano, que é a memória.

A linguagem escrita tem sua história dividida em 3 fases: pictórica, ideográfica e alfabética. A fase pictórica consiste na representação de objetos por meio de símbolos que correspondam à imagem daquilo que se quer representar. É uma escrita simplificada dos objetos da realidade. Este tipo de escrita, contudo, não oferece grande alcance tendo em vista o grande número de objetos a serem representados. A limitação dessa representação levou o homem a desenvolver outro tipo de linguagem escrita, a ideográfica, conforme podemos constatar em Furtado (1995, p 32): “Em termos de progresso na escrita, procura-se superar as dificuldades da escrita pictográfica, em grande parte devidas à multiplicação dos signos, e caminha-se no sentido do fonetismo”¹.

Os ideogramas, facilmente reconhecíveis ainda nos dias atuais na escrita chinesa e japonesa, são representantes da escrita ideográfica. Nessa fase, percebe-se não mais a representação de objetos apenas, mas também a de idéias e de sons. Este sistema, que aproximou a escrita de uma das suas funções, a de interpretar a língua falada, foi também gradualmente direcionado para o fonetismo. A partir desta escrita é possível também

¹ Sistema onde as palavras passaram a ser decompostas em unidades sonoras

representar por meio de símbolos conceitos mais abstratos, como por exemplo, a idéia de “pressa” representada, para os indígenas da América do Norte, pela figura de um pássaro voando.²

A fase marcada pela escrita alfabética é, talvez, a mais importante, pois foi a partir da criação de um sistema universal de códigos que pudesse estar acessível a todos que o homem pôde, mais tarde, desenvolver a gramática e, mais do que nunca, ampliar seu intelecto e sua rede de comunicação.

Para chegar ao sistema alfabético utilizado atualmente o processo de evolução da escrita foi longo e demorado. Durante esse processo o homem percebeu que a decomposição do som das palavras as reduzia em unidades independentes e diferenciáveis umas das outras. Surge, a partir daí, a escrita alfabética, na qual cada sinal corresponde a uma letra.

Foi a partir da criação do alfabeto fenício, constituído por vinte e dois signos, que a escrita alfabética se expandiu. Inspirados na criação fenícia, outros povos criaram seus próprios alfabetos, como os gregos, que aperfeiçoaram e ampliaram essa escrita, a qual passou a ser composta por vinte e quatro letras, divididas em vogais e consoantes e, a partir de 720 a.C. aparecem os primeiros documentos (em argila) escritos na linguagem alfabética.

Apesar de inicialmente o alfabeto não ter provocado uma maior acessibilidade ao conhecimento registrado, já que durante certo tempo seu uso foi restrito a monges e burguesia, não se pode deixar de conferir a esse sistema a indubitável importância para o progresso e a evolução humana. Conforme Furtado (1995): “O sistema alfabético vai, no entanto, ocasionar uma verdadeira revolução: por um lado alarga o uso da escrita, ultrapassando o círculo restrito dos escribas; por outro, possibilita o aparecimento das religiões do livro.”

3.2 Língua

Para que a linguagem tivesse sua evolução e sua função de comunicar garantidas era necessário um suporte de linguagem. Um código capaz de sustentar o processo de

² BORDENAVE, Juan E. Díaz. O que é comunicação. São Paulo, 1982.

comunicação humana e permitir a exploração do conhecimento que resulta deste processo. Esse sistema, a língua, é o que permite a tradução do pensamento em signos decifráveis tornando possível a troca de mensagens.

Enquanto a fala representa o lado individual da linguagem, a língua, segundo Saussure (1969a), representa o lado social, uma vez que funciona como um conjunto de convenções determinado por uma espécie de contrato social feito por indivíduos de um mesmo grupo. Ou seja, ela só existe em função de uma prática social e, sendo assim, se configura como uma atividade social e histórica extremamente dependente dos indivíduos.

Saussure (1969b) afirma que a língua ao mesmo tempo em que se constitui em um sistema abstrato é também um conjunto de convenções necessárias adotadas pelo corpo social para permitir que essa faculdade se torne possível nos indivíduos. Assim, criá-la ou modificá-la não está ao alcance do indivíduo, que para se comunicar precisa de submeter-se ao convencionalizado. Segundo o autor, “a língua é, a cada momento, tarefa de toda gente; difundida por uma massa e manejada por ela, é algo de que todos os indivíduos se servem o dia inteiro.” (SAUSSURE, 1969c, p. 88).

Ao contrário da linguagem, considerada por muitos como um sistema natural de comunicação, a língua constitui-se em algo que é criado, um sistema para se fazer concreto o exercício da linguagem. Ela se encontra na mente do falante, que associa os conceitos às imagens (signo lingüístico), mas é exteriorizada pela fala a fim de que se possa estabelecer uma comunicação entre sujeitos. Como sugere Kristeva (1969, p. 25) “a existência de um código comum fundamenta a comunicação e torna possível a troca das mensagens”. A língua é, portanto, o suporte que permite o exercício da linguagem, sendo a partir dela que a informação pode ser explorada e disseminada. Sem a criação de um código que possibilitasse o exercício da linguagem não seria possível o registro de informações e, nem tampouco, a própria existência de uma ciência para estudar o fluxo informacional contido em documentos, a Ciência da Informação.

3.3 Signo lingüístico

O signo lingüístico formado na mente do falante quando no exercício da linguagem funciona como o representante de uma idéia expressa no código convencional da língua e caracteriza-se como uma unidade simbólica que une um significante a um significado, isto é, uma espécie de contrato entre a formação mental feita a partir do objeto representado e a forma fônica representante. O signo é composto pela sua forma física e por um conceito mental que está associado a ele, sendo este conceito, por sua vez, uma apreensão da realidade externa. Ele existe em função de um sujeito e apenas se relaciona com a realidade através dos conceitos.

Segundo Saussure (1969), os signos que constituem uma língua não têm nenhum valor absoluto em si, não fazem sentido se considerados isoladamente. Eles só existem em função de alguma coisa e são dependentes de significante e significado, só se fazendo existentes e inteligíveis na presença desses dois elementos. Se apenas o aspecto material ou gráfico de um signo for considerado não é possível encontrar nele nada que remeta ao que reconhecemos como sendo uma palavra. O mesmo ocorre se apenas o aspecto conceitual for considerado, ou seja, tentar reconhecer um signo unicamente por aquilo que ele representa.

A riqueza da linguagem decorre de uma característica fundamental do signo: ele é um símbolo, um traço que está no lugar de outra coisa, a qual pode ser um objeto concreto, ou um conceito abstrato. O signo é uma representação e, como tal, se reduz a um rastro uma vez que não é a presença da coisa ou do conceito. Assim, o sujeito estabelece uma espécie de ilusão ao ver o signo como uma presença, isto é, de ver no signo a presença da coisa ou mesmo do conceito e essa ilusão é necessária para que ele funcione ao mesmo tempo como a presença e a não presença da coisa. A atribuição de significados a determinados signos é precisamente a base da comunicação em geral. Há quem afirme, inclusive, que sem o recurso dos signos, não seríamos capazes de distinguir duas idéias de modo claro e constante. Afirma Condillac (apud FOUCAULT, 1999) que é o signo quem “traça a divisória entre o homem e o animal; ele que transforma em imaginação a memória voluntária, a atenção espontânea em reflexão, o instinto em conhecimento racional”.

São duas as características fundamentais do signo e dizem respeito à sua arbitrariedade e a sua linearidade. Quanto à primeira entende-se que o signo faz parte de um código

convencional (a língua) e se estabelece por acordo entre os elementos de um mesmo grupo lingüístico. Assim, o sujeito não pode, por livre iniciativa, alterar a cadeia de fonemas de um significante, por exemplo, do signo lingüístico “casa”, que representa lingüisticamente a casa que existe na realidade, por que esta representação foi estabelecida por uma comunidade lingüística e está atada à sua língua. A arbitrariedade, portanto, não se refere à idéia de livre escolha, de contrato social à *posteriori*, e sim à *priori* (CINTRA, 1983). É importante ressaltar, contudo, que os signos não são rígidos e estáticos e, como produtos de uma sociedade dinâmica, também são mutáveis. Este é o paradoxo da mutabilidade e imutabilidade do signo, defendido por Saussure. Quanto a seu caráter linear, o signo apresenta-se em uma seqüência onde cada elemento ocupa uma posição determinada. Assim, no signo “casa” cada fonema e cada letra ocupam um espaço na cadeia deste significante e por que se apresentam nessa ordem lógica podem ser compreendidos formando a idéia de casa. Essa característica é o que possibilita o surgimento de um contexto onde as relações de oposição e associação ocorrem, as relações sintagmáticas e paradigmáticas.

No campo da Ciência da Informação, o estudo dos signos lingüísticos leva a compreender melhor o que é uma linguagem documentária, enquanto modo de organização de um conjunto de signos, bem como suas características e forma de funcionamento enquanto sistema semiótico particular. A idéia de signo lingüístico associada à de representação da informação leva também à formulação da idéia de signo documentário, indispensável para as atividades de análise documentária e indexação. Para a AD a característica arbitrária conferida ao signo elucida o fato de que as construções documentárias são essencialmente fundamentadas institucionalmente, isto é, assumem significados dependentes das marcas institucionais a que estão subordinadas devendo ser referidos, portanto, às atividades onde estão contextualizados (LARA, 1993). Além disso, a separação de termos em eixos sintagmáticos e paradigmáticos em uma LD é possível devido a o caráter linear do signo, conforme já exposto neste capítulo.

3.4 Relações sintagmáticas e paradigmáticas

A diferenciação entre as relações sintagmáticas e as relações paradigmáticas existentes entre as palavras que designam simbolicamente os referentes é indispensável para a formulação de campos semânticos que possibilitam a passagem de uma linguagem natural para uma linguagem documentária no campo da Ciência da Informação.

As relações sintagmáticas são compostas sempre de duas ou mais unidades consecutivas e as combinações dessas unidades se dão devido ao caráter linear da língua, que exclui a possibilidade de pronunciar dois elementos ao mesmo tempo. Exigem a presença dos dois elementos relacionados, ou seja, não é possível estabelecer o valor de um desses elementos sem que o outro o preceda ou o suceda (SAUSSURE, 1969). Por relacionar no mínimo dois elementos lingüísticos é um tipo de relação em que os elementos relacionados se encontram em presença um do outro. Por isso, Saussure também define estas relações como relações *in praesentia*. Um termo só passa a ter valor a partir do momento em que ele se contrasta com outro elemento. Essa dependência ou relação que dois elementos consecutivos de uma seqüência mantêm entre si é o que caracteriza uma relação sintagmática.

Ao contrário das relações sintagmáticas, existentes apenas na presença recíproca de dois termos, o paradigma sendo um conjunto de unidades que podem fazer parte de um mesmo contexto, contudo, acontece na seleção de suas unidades, ou seja, se uma está presente, a outra ou as outras deverão estar ausentes. Esse conjunto de elementos se encontra na memória do falante e apenas um fará parte do discurso. As relações associativas ou paradigmáticas, portanto, porque dizem respeito à seleção entre elementos, são um tipo de relação em que o elemento selecionado exclui os demais elementos da relação e, por esta razão, também recebem o nome de relações *in absentia*.

No campo da Ciência da Informação as relações paradigmáticas e sintagmáticas configuram-se como importantes instrumentos para a construção de linguagens documentárias, pois é a partir delas que é possível definir as categorias de termos que irão compor uma LD.

3.5 Categorização

Constantemente estamos identificando novos objetos, fenômenos e fatos através da nossa capacidade de dar sentido às coisas separando delas suas características essenciais e colocando-as em espaços de reconhecimento segundo características comuns. Categorizar é reunir características comuns a vários conceitos de modo que esse conjunto de características resultará em uma característica geral passível de agrupar esses diversos conceitos em uma só classe. Afirma o filósofo Foucault (1999, p. 71) que

O espírito humano é naturalmente levado a supor que existem nas coisas mais ordem e semelhança do que possuem; e enquanto a natureza é plena de exceções e diferenças, por toda a parte o espírito vê harmonia, acordo e similitude [...] não se pode conhecer a ordem das coisas na sua natureza isoladamente, mas, sim, descobrindo aquela que é a mais simples, em seguida aquela que é a mais próxima para que se possa aceder necessariamente, a partir daí, até as coisas mais complexas.

Desse modo, o ato de categorizar é um processo mental do ser humano que, naturalmente, procura organizar as coisas do mundo para melhor compreendê-lo. Mas antes, organizamos coisas a partir de uma ordem já estabelecida, ou seja, ordenamos segundo classes de coisas já convencionadas e, portanto, a categorização parte também de um processo social.

A definição de categorias e sua discussão é base para a organização e representação do conhecimento e, portanto, de fundamental importância para o campo da Ciência da Informação. A tentativa de estruturação da informação com vistas a recuperá-la é precisamente o que caracteriza a categorização no campo da CI, que se propõe a organizar o conhecimento de qualquer área. Para organizar o conhecimento de determinada área, entretanto, estudam-se primeiro os conceitos que compõem esse campo do conhecimento e as relações entre eles. Parte-se do princípio que conceitos e relações são fundamentos do conhecimento e do pensamento, e então necessários para o raciocínio e inferências, base do processo de categorização.

Em Teoria do Conceito, Dahlberg (1978, p. 102) define um conceito como a “compilação de enunciados verdadeiros sobre determinado objeto, fixada por um símbolo lingüístico.” O conceito, portanto, é uma representação mental sem a qual não seria possível organizar informações em classes de coisas, não sendo possível, portanto, o procedimento de categorização. À medida que novos conceitos vão surgindo novas estruturas vão sendo apreendidas e informações vão sendo inseridas nas categorias já existentes ou estas últimas

vão sendo alteradas por meio de um processo de reestruturação. Esta tentativa de inserção de informações em categorias já existentes para abarcar novos conceitos pode, contudo, causar a descaracterização destas categorias. (LARA, 2001)

4. CIÊNCIA DA INFORMAÇÃO

A chamada explosão bibliográfica, com ápice no período que sucede a 2ª Guerra Mundial quando a sobrevivência de uma nação parecia depender essencialmente do conhecimento científico e tecnológico e, conseqüentemente, a informação passou a ser não só um mero produto da inteligência humana, mas um eficaz e almejado instrumento de poder, provocou o interesse dos governos por grandes investimentos em pesquisa e desenvolvimento promovendo, em conseqüência, avanços científicos e tecnológicos nas diversas áreas do conhecimento.

As descobertas e produção científicas, cujas proporções ultrapassam os limites de tempo e espaço, ocasionam, entretanto, um problema relacionado à organização do conhecimento gerado. A pergunta que se faz em tal contexto é: como armazenar tantas informações e organizá-las de modo que estas possam estar acessíveis à comunidade científica geral?

Le Coadic (1996) afirma que toda ciência é uma atividade social determinada por condições históricas e socioeconômicas e que a sociedade da informação necessita de uma ciência que estude as propriedades da informação e os processos de construção, comunicação e uso. Assim, a Ciência da Informação surge da necessidade de se estabelecer uma ordem para a crescente produção de informação, principalmente no pós-guerra da 2ª Guerra Mundial. Conforme Lima (2003, p. 78):

A CI apareceu como uma nova área do conhecimento a partir da revolução técnico-científica posterior à II Guerra Mundial. O grande volume de informações gerado no crescente número de áreas do conhecimento passou a demandar um nível maior de organização informacional

Borko (1968, p. 3) descreve a Ciência da Informação como

uma disciplina que investiga as propriedades e o comportamento da informação, as forças que governam seu fluxo, e os meios de processá-la para otimizar sua acessibilidade e uso. A CI está ligada ao corpo de conhecimentos relativos à origem, coleta, organização, armazenagem, recuperação, interpretação, transmissão, transformação e uso da informação

Para Barreto (2007, p. 1),

A Ciência da Informação se preocupa e se ocupa com os princípios teóricos e as práticas da criação, organização e distribuição da informação [...] mostra a sua essência quando uma linguagem no pensamento de um emissor atinge uma linguagem de inscrição pública colocada em uma estrutura passível de apropriação por receptores e com destino final para gerar conhecimento.

Para Rees e Saracevic (apud COSTA, 1990) é o "estudo dos fenômenos da Comunicação e das propriedades dos sistemas de Comunicação. A Ciência da Informação investiga técnicas e métodos que possibilitam a compreensão melhor das propriedades, do comportamento e fluxo da informação". Ou seja, está diretamente ligada com os processos da comunicação humana detendo-se, contudo, aos processos que governam a comunicação registrada.

Em todo caso, constata-se que o objeto e tema central da Ciência da Informação, como o próprio nome já denota, é a informação, seja ela entendida como conhecimento registrado como propõe Le Coadic (1996) e Farradane³ (1979), seja como dados relevantes para a tomada de decisão conforme Kochen⁴ (1983), ou ainda como "instruções seletivas" redutoras de incerteza conforme Pignatari⁵ (1977).

À Ciência da Informação cabe a tarefa de estudar o fluxo informacional presente nos documentos. Hoje, mais do que em qualquer outra época, este tipo de informação, a documentária, atua como instrumento propulsor do desenvolvimento dos vários ramos do conhecimento humano e, portanto, as formas de representação para posterior recuperação constituem-se em um grande desafio para a Ciência da Informação. Para encará-lo, esta ciência utiliza-se de uma das suas principais características, a de ser uma ciência interdisciplinar que, abrangendo áreas como a lingüística, a tecnologia de computadores, as comunicações, a biblioteconomia, a administração dentre outras, toma emprestado destas diversos conceitos para auxiliá-la a prover meios para garantir, de maneira simples e objetiva, a última etapa do ciclo da informação, a disseminação.

Dentre tais ciências, no atual contexto da sociedade da informação em que os esforços da área de Ciência da Informação se voltam principalmente às questões de organização, representação e recuperação da informação, a Lingüística desempenha um papel fundamental. No que pese as diferenças que permeiam os dois campos há, entretanto, uma proximidade entre as duas ciências, uma vez que a eficiência dos Sistemas de Recuperação da Informação (SRI) é garantida muitas vezes a partir de estudos lingüísticos.

A partir de alguns conceitos lingüísticos, inicialmente citados e comentados neste trabalho, problemas semânticos podem ser amenizados evitando assim a dispersão de

³ FARRADANE, J. The nature of information. *Journal of Information Science*. v.1, n.1, p.13-17, Apr. 1979.

⁴ KOCHEN, M. *Information and Society*. ARIST, v.18, pp. 277-304, 1983.

⁵ PIGNATARI, D. *Informação. Linguagem. Comunicação*. São Paulo: Perspectiva, 1977

vocabulário que pode resultar em uma má recuperação de informações, ou ainda, torna-se possível, por exemplo, a conversão de uma linguagem natural para uma linguagem documentária. Além disso, o conceito de análise documentária em ciência da informação foi, outrora, formulado a partir do conceito das atividades de análise de textos.

5. ANÁLISE DOCUMENTÁRIA

A atribuição de significados e categorização dos objetos e idéias apreendidas pelo cérebro humano é uma prática natural demarcada pela característica intrínseca ao homem de compreender e organizar o conhecimento a fim de que sua história possa ser transmitida no tempo e acessada pelas gerações seguintes. Conforme Cintra *et al.* (1994, p.19)

A prática da linguagem é marcada por uma tendência natural do homem: compreender, governar e modificar o mundo. Com efeito, o homem busca, incansavelmente, encontrar uma ordem para as coisas, já que um mundo caótico seria incompreensível, insuportável; por isso ele busca encontrar em meio à aparência caótica uma ordem mesmo que subjacente, uma estrutura capaz de explicar as coisas.

Demarcar as especificidades de dado objeto em certo contexto é fato que se dá durante o processo de apreensão do real e que exige a assimilação do que já é conhecido para só depois definir as particularidades do objeto. Ou seja, o processo de conhecimento se dá por meio da analogia.

A apreensão dos significados contidos em um documento ou texto é um processo que se dá também por meio de analogias e que requer a definição das particularidades do documento. O leitor, ao principiar a leitura, inicia também, automaticamente, um processo de delimitação de pontos importantes do documento de acordo com o que já conhece e com seus objetivos, procedendo assim a uma análise.

O termo análise, no Novo Dicionário da Língua Portuguesa⁶, dentre outras definições, significa: “exame de cada parte de um todo, tendo em vista conhecer sua natureza, suas proporções, suas funções, suas relações, etc.”. Ou seja, o ato de analisar pressupõe a desconstrução de um todo para submeter as partes a um exame mais detalhado a fim de que se possa ter uma visão global mais clara sobre determinado conjunto.

Do mesmo modo paradoxal, a leitura de um texto para fins de disseminação e recuperação da informação é um processo que requer, inicialmente, a desconstrução do documento para uma extração das idéias principais e uma posterior tradução em uma linguagem capaz de representar essas idéias. É a Análise Documentária (AD). Segundo Coyaud (1967), “a análise documentária é primeiramente uma operação de reconhecimento das unidades lexicais que representam as noções importantes de um documento.”

⁶ BUARQUE DE HOLANDA, A. Novo dicionário da língua portuguesa. 1975. 1499 p.

O termo “Análise Documentária” introduzido na literatura da Documentação por Jean Claude Gardin é utilizado para definir, segundo ele próprio, um processo de extração de significados de um texto, o que não implica na retirada de símbolos encontrados necessariamente nele. Tem como objetivo o isolamento dos sentidos dos textos “com a finalidade de permitir pesquisas retrospectivas de informação realizadas sobre o conteúdo ou a significação de textos.” (GARDIN, 1973, p.67, apud LARA, 1993, p.14).

Este tipo de análise surge com a preocupação com a criação de mecanismos de controle bibliográfico para resolver ou amenizar os problemas de armazenamento e recuperação da informação desencadeados pelo crescimento desordenado do conhecimento científico e tecnológico após a 2ª Guerra Mundial. Conforme Maimone:

As maneiras “empíricas”, ou seja, conhecimento adquirido pela experiência de organizar um acervo, já não eram suficientes para suprir uma quantidade tão ampla e diversificada de materiais. Houve necessidade de desenvolver “sistemas” para melhor organizar e controlar as áreas do conhecimento. Assim surge a Análise Documentária, com fins de indexar e melhor organizar o universo documental que começava a se expandir naquele momento. (MAIMONE, 2007, p. 4)

Pode-se observar daí que a análise documentária é um processo importante para garantir a disseminação e o acesso à informação, objetivos centrais da Ciência da Informação.

Para Cunha (1987, p.68), 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”.

Para Lara (1993, p. 4) a AD se define como “uma atividade específica no interior da Documentação, respondendo pela análise e tratamento da informação com o objetivo de recuperá-la e disseminá-la”. É uma atividade que tem como objetivo genérico a extração da informação documentária para sua representação em códigos próprios e objetivo final a disseminação dessa informação. A autora considera ainda que o processo de AD é constituído por dois tipos distintos de representação da informação, onde o primeiro é construído por meio de um processo de condensação intensiva do texto original gerando um novo texto que reconstrói o sentido do primeiro (os resumos) apresentando assim “uma relação de semelhança e contigüidade” entre ambos; e no segundo tipo, a representação é feita por meio de um sistema que permite a transmissão da mesma mensagem sob uma outra forma lingüística, a de uma Linguagem Documentária (LD), mas que, entretanto, não apresenta necessariamente uma relação de contigüidade com o texto original, como nos resumos. Índices e resumos, entretanto, não são considerados linguagens documentárias e sim produtos da representação documentária.

A mesma idéia sobre a representação produzida pela análise documentária é apresentada por Kobashi (1996). Segundo ela a representação de conteúdos de documentos com vistas à recuperação da informação é o objetivo central da AD. Conforme a autora, o resumo e o índice são os dois tipos básicos de representação da informação dados pela AD. No resumo a semelhança e a contigüidade com o texto original são mantidas, já que se trata de uma reconstrução feita a partir de partes do texto. Já no índice esta relação é quebrada devido ao caráter generalizante assumido pelo próprio processo de indexação, caracterizado pela representação das idéias do texto por meio de palavras-chave ou descritores. Os processos de recuperação da informação representados pela indexação e pelo resumo são complementares. Como as linguagens de indexação designam classes de coisas e não informações particulares, a busca pelo processo de indexação pode oferecer resultados com “certo grau de imprecisão”, sendo necessário recorrer aos resumos para se obter maior grau de relevância nas informações requeridas. (KOBASHI, 1996, p.10)

Para Smit (apud LARA, 1993) o processo de análise documentária compreende duas etapas básicas: uma de descrição e outra de ordenação de dados. A primeira etapa, de descrição, refere-se ao momento em que há a interpretação dos dados e a identificação de Traços Descritivos, unidades portadoras de informação e Argumentos, dados provenientes de fontes externas; enquanto a segunda etapa, de ordenação, funciona como uma validação dessa interpretação por meio da definição de subconjuntos conceituais.

É importante ressaltar que o processo de AD não engloba apenas o reconhecimento de operações semânticas ou sintáticas, mas carrega, como todo processo de análise, um agregado de ideologia do analista do documento, já que a ele cabe a tarefa de decidir quais os pontos mais relevantes do documento examinado, além das condições institucionais a que estão submetidos os documentos. Este é um ponto que gera discussões a respeito do papel do analista documentário enquanto sujeito responsável pela demarcação da validade de um documento e/ou partes dele.

Lucas (1997) afirma que a análise do indexador não deve deter-se no conteúdo informacional propriamente dito. As condições de produção e o interesse do leitor são fatores que influenciam na relevância da informação para um nível institucional e na busca por informações, e a supressão destes durante o processo de análise pode acarretar uma má definição das partes relevantes.

Quando as condições de produção do discurso não são levadas em consideração, a análise não se mostra capaz de explicar o funcionamento discursivo, apontando para

um discurso sem memória em que a contradição entre as várias formações discursivas que ali se entrecruzam fica apagada. (LUCAS, 1997, p.7)

Cintra (1989, p. 31) também desconsidera um posicionamento neutro do indexador ou analista da informação documentária quanto ao contexto da produção da informação: “Sendo a linguagem intrinsecamente comprometida com o cultural e o ideológico, tanto o processo de produção do texto, quanto o de recepção não se isentam destes componentes. E é esse o comprometimento que garante o caráter interativo da leitura”

Ainda em Lucas observa-se que ao bibliotecário ou analista da informação é negada a interpretação até mesmo pelo tempo que este profissional tem para realizar a tradução das idéias centrais dos textos em descritores ao afirmar que

a leitura do bibliotecário simula o modo de produção industrial, buscando produtividade, rapidez, não dando margem à reflexão e ao acúmulo de conhecimento por parte do bibliotecário; esta leitura deseja-se rigorosa, transparente, o sujeito leitor não deve interpretar; a leitura deve ser literal apreendendo o conteúdo do texto e produzindo representações do mesmo, simulacros, dando-lhe unidade por meio palavras-chave. (LUCAS, 1997, p. 8)

5.1 Indexação

A análise documentária descrita acima é um processo indispensável para a atividade de indexação. Pode-se dizer, inclusive, que a AD é a 1º etapa no processo de indexação e, portanto, sem ela não é possível indexar um documento. Para Fujita (2003, p. 62) “a indexação em análise documentária, sob o ponto de vista dos sistemas de informação, também é reconhecida como a parte mais importante porque condiciona os resultados de uma estratégia de busca”, sendo que o bom ou o mau desempenho dessa atividade influencia na recuperação da informação.

De acordo com a NBR 12676 da ABNT, que dispõe sobre os métodos para análise de documentos, indexação é o ato de identificar e descrever o conteúdo de documentos com termos representativos dos seus assuntos, que constituem uma linguagem de indexação, e consiste basicamente em três estágios: o exame do documento e estabelecimento do assunto de seu conteúdo; a identificação dos conceitos presentes no assunto; tradução desses conceitos por meio de uma linguagem de indexação, ou linguagem documentária.

Para Lancaster (2004) indexação é o processo de atribuição de palavras-chave para determinar o conteúdo temático de um documento que facilita na recuperação da informação. Neste processo duas etapas podem ser observadas: a de análise conceitual e a de tradução. A análise conceitual consiste em examinar o documento de forma a decidir seus assuntos e definir seus conceitos. Esta etapa nada mais é do que uma interpretação do conteúdo de um documento, da informação contida nele. Nesta etapa é importante observar que na definição dos assuntos do documento deve-se levar em consideração o objetivo do serviço de informação ao qual está subordinado e o grupo de usuários ao qual ele se destina. Este é um grande desafio para o indexador que precisa antecipar as necessidades do usuário para garantir a eficiência na recuperação da informação.

Quando termina esta etapa de análise documentária com os assuntos já definidos em linguagem natural - expressão que no contexto da recuperação da informação normalmente se refere às palavras que ocorrem em textos impressos e, por isso, considera-se como seu sinônimo a expressão texto livre (LANCASTER, 2004) - chega o momento de inseri-los em um Sistema de Recuperação de Informação (SRI). Começa então a etapa de tradução dos assuntos definidos em termos para uma linguagem documentária. A seleção desses termos pode acontecer por extração, quando são retirados do próprio texto, isto é, termos em língua natural, ou por atribuição, quando os termos são selecionados por uma fonte que não é o documento, mas que possuem coerência com o assunto. Na maioria dos sistemas a indexação por atribuição envolve o uso de vocabulário controlado, uma lista limitada de termos autorizados que podem ser usados pelo indexador para representar os conteúdos principais do documento. Os termos de indexação por atribuição provêm das linguagens documentárias e, portanto, orientados para a univocidade, já que seus significados são prefixados. Por possuírem significados fixos esses termos também podem ser considerados signos documentários.

A partir disso, percebe-se que a indexação é também uma operação de tradução de textos em linguagem natural para uma linguagem documentária (CINTRA *et al.* 1994). O porquê de se fazer uma tradução que mude a linguagem natural do usuário para uma linguagem pré-estabelecida por um sistema é devido às características que tornam essa linguagem natural de difícil utilização para o tratamento da informação. Dentre essas características destacam-se a redundância, a polissemia e a ambigüidade. Essas características conferem aos termos em linguagem natural uma grande possibilidade de significações. Não

que essa determinação seja ruim: a linguagem natural é, invariavelmente, a base para a existência de linguagens construídas, já que estas tomam como modelo as relações paradigmáticas e sintagmáticas estabelecidas pela primeira (DODEBEI, 2002). Contudo, para fins de recuperação de informação documentária, seu caráter polissêmico representa “uma ameaça” para a relevância dos resultados de busca por informações.

6. LINGUAGEM DOCUMENTÁRIA

Principalmente no pós-guerra da 2ª Guerra Mundial a produção científica se deu de tal intensidade que as formas usuais de armazenamento e posterior recuperação da informação já não eram mais capazes de suportar o contingente de produção.

Assim como a escrita pictográfica, que em certa época, já não era capaz de traduzir todos os objetos existentes e, portanto, não satisfazia a necessidade de comunicação por completo, o uso de termos em linguagem natural para a indexação de documentos, à luz da nossa época de incessante produção no campo científico já não é capaz de garantir a eficiência na recuperação de informações pelos usuários. A utilização de termos em linguagem natural para a armazenagem e busca das informações produzidas muitas vezes não é suficiente para que se tenha uma relevância nos resultados de busca. Devido à grande quantidade de informações existentes, a recuperação por meio desses termos também oferece uma quantidade enorme de resultados, além de não estabelecer diferenças conceituais entre palavras, fato que se dá devido o caráter polissêmico desta. Assim, as formas usuais de representação da informação já não se mostravam suficientemente eficientes para garantir uma boa recuperação dessas informações.

A partir desse quadro de constante disponibilização de informações e intensificação de problemas relacionados à sua recuperação o homem sente a necessidade de traçar mais um nível de evolução nas formas de linguagem.

... o fato de a ciência moderna ser caracterizada por um alto grau de especialização, um alto grau de compartilhamento, de fragmentação dos papéis e responsabilidades dos cientistas, engendra, entre outras coisas, uma necessidade vital de estruturas técnicas e de técnicas complexas para cada disciplina. (MARCONDES E SAYÃO, apud. DODEBEI, 2002, p. 22)

Foi então, proposta uma nova forma de linguagem que pudesse atender à necessidade de obtenção de informações relevantes em um curto espaço de tempo e que, ao mesmo tempo, fosse capaz de amenizar ou diminuir os problemas de vocabulário oriundos das linguagens naturais. Essa nova linguagem destinada a traduzir de maneira sintética o conteúdo dos documentos configura-se como um importante instrumento para garantir a eficiência dos sistemas de recuperação da informação tendo como função representar o conhecimento inscrito, é a Linguagem Documentária (LD). Derivadas das operações que tornam documentos úteis a uma pesquisa documentária essas linguagens se constituem em produtos

construídos para proporcionar uma interação mais clara na comunicação entre sistema-usuário, ou na comunicação documentária. Afirma Cintra *et al.* (1994a, p. 24) que

Com efeito, dentro do amplo universo da linguagem, as LDs possuem um status muito particular: através delas pode-se representar, de maneira sintética, as informações materializadas nos textos... são sistemas simbólicos instituídos, que visam facilitar a comunicação. Sua função comunicativa, entretanto, é restrita a contextos documentários, ou seja, as LDs devem tornar possíveis a comunicação usuário-sistema.

Caracterizadas essencialmente por representar o conteúdo de documentos sob a forma de conceitos e termos especializados, linguagens documentárias exigem a existência de regras explícitas para sua utilização, e por este motivo, são consideradas linguagens construídas. Assim, para cada elemento escolhido para compor uma LD existe uma definição única e pré-estabelecida, isto é, a cada significante procura-se fazer corresponder um significado. (CINTRA *et al.* 1994b).

Segundo Gardin (apud CINTRA *et al.* 1994c) uma Linguagem Documentária se compõe de um léxico, de uma rede paradigmática e de uma rede sintagmática. O léxico, ou vocabulário, constitui-se na lista dos elementos descritores que compõem uma determinada LD. A rede paradigmática é dada pelas relações entre as unidades lexicais, determinadas *a priori*, sendo a partir delas que se faz possível a noção de hierarquia e a determinação de elementos na ordem gênero/espécie, já que o significado de cada palavra traz consigo a sua inserção em uma classe de objetos, por exemplo: gás hélio supõe elemento químico e também gás nobre. As inferências nessas relações são finitas, pois estão contidas numa escala entre o maior gênero e a menor espécie. A rede de relações sintagmáticas, em contrário, representa a possibilidade infinita de relações entre palavras em busca de um sentido, ou seja, classes distintas de palavras que intencionalmente reunidas vão compor uma mensagem: /árvore/ e /madeira/ e /casa/ e /incêndio/ (DODEBEI, 2005). Essas relações são estabelecidas, dentro de um sistema de recuperação de informação, pela utilização dos operadores booleanos E, OU, NÃO.

A construção de uma linguagem documentária pressupõe, segundo Dodebei (2002), a identificação e a segmentação do domínio do conhecimento em categorias devendo a identificação destas ser realizada considerando-se a literatura da área. Esse processo permite que, a inclusão de um novo termo em uma categoria adequada.

No campo da Análise Documentária as LDs funcionam como instrumentos intermediários ou comutadores, ou seja, elas são os instrumentos que permitem a tradução da síntese de documentos dada pela AD e das perguntas dos usuários. O usuário faz uma

pergunta em linguagem natural que é formalizada em linguagem documentária e, por isso, estas são consideradas também como instrumentos de comunicação documentária. (CINTRA *et al.* 1994).

Quanto à sua tipologia as linguagens documentárias podem ser divididas em dois grupos: em pré-coordenadas e em pós-coordenadas. As LDs pré-coordenadas são linguagens que combinam ou coordenam termos no momento da indexação. Nelas, os assuntos são tratados como uma unidade, isto é, são representados por um conjunto de termos já combinados para formar assuntos complexos. Podem ser de dois tipos: classificatórias e alfabéticas. As LDs classificatórias possuem subordinação lógica no interior de um sistema de pensamento, indo do genérico para o específico. Incluem-se aqui os sistemas de classificação CDD, CDU, LC e classificação facetada de Ranganathan, que se constituem como as mais antigas linguagens documentárias. Atualmente, têm basicamente a função de ordenar o acervo por assunto. As alfabéticas possuem cabeçalhos de assuntos ordenados alfabeticamente. As listas de cabeçalhos de assunto, que arrolam termos de todas as áreas do conhecimento, são exemplos dessas linguagens.

As linguagens documentárias pós-coordenadas são as que combinam ou coordenam os termos no momento da busca, ou recuperação da informação. Neste tipo de linguagem, o vocabulário torna-se menos extenso uma vez que não é necessário criar um novo termo para cada assunto complexo bastando combinar os que já existem. Uma simples lista de descritores autorizados atualizada pode ajudar na indexação. São exemplos de LDs pós-coordenadas, o sistema Unitermo e os tesouros. Desses, os tesouros se mostram muito eficientes para a precisão do campo de indexação já que incluem diferentes tipos de relacionamento entre os termos e possibilitam diferentes tipos de organização por meio das relações conceituais que permitem a identificação do conceito e não somente a identificação da palavra. Essas estruturas são extremamente dependentes dos conceitos lingüísticos de relações paradigmáticas e sintagmáticas, pois é a partir delas que são estabelecidas as relações de hierarquia e noção de rede de associações, que compõem a estrutura básica de uma LD.

É possível notar a partir das exemplificações que há uma progressão contínua das LDs a caminho da especialização. Se no início a pretensão era cobrir todo o universo do conhecimento, a exemplo dos sistemas tradicionais de classificação, hoje com a ultra-especialização das áreas do saber, para cumprir a função de facilitar a comunicação científica,

as linguagens documentárias precisam voltar-se também a domínios cada vez mais específicos.

7. CONCLUSÃO

Durante muito tempo as discussões que envolviam as questões da linguagem giraram em torno de como e por que esta capacidade humana teria surgido. No contexto da atual sociedade da informação, com o desenvolvimento dos estudos acerca do tema essas discussões passaram a fazer parte não só dos campos das ciências cognitivas e linguística, mas também do corpus discursivo das ciências destinadas à compreensão dos processos de informação. Encontra-se nesta lista a Ciência da Informação que, ademais, não dispensa as formulações já estabelecidas por aquelas ciências, principalmente a ciência linguística, mas as utiliza para complementar e melhorar as atividades de análise, representação e recuperação da informação, conforme exposto ao longo do presente trabalho. A linguagem como forma de representação da informação é capacidade que atua como importante ferramenta na sistematização do conhecimento.

Os avanços científicos e tecnológicos provocam grandes mudanças nos processos de comunicação escrita. Em um passado remoto as formas de representação da informação por meio da linguagem recorriam às faculdades corporais e, em seguida, aos suportes materiais. Hoje com a incontrolável produção de informações e as facilidades que a ciência computacional nos oferece, primamos por desenvolver novos modelos de representação da informação por meio de linguagens capazes de assegurar sua organização de maneira a facilitar a transferência e compreensão de informações relevantes para o processo de produção do conhecimento. A construção e uso de linguagens documentárias, conforme ilustrado neste trabalho, surgem como resposta a esse investimento e permitindo assim garantir a eficácia no processo de comunicação documentária .

REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. **Métodos para análise de documentos - seleção de termos de indexação - NBR 12676**. Rio de Janeiro, 1992.

ALVARENGA, Lidia. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaços digitais. **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 15, 1º sem. 2003. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/viewFile/97/5233>>. Acesso em: 03 nov. 2009.

AMARAL, Márcio Tavares d'; Instituto Nacional do Livro (Brasil). **Filosofia da comunicação e da linguagem**. Rio de Janeiro: Civilização Brasileira, 1977. 136 p. (Coleção Perspectivas do Homem; 108)

AZEVEDO, Willian Azevedo. A construção da ciência da informação na pós-modernidade: dialética histórica. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 6, n. 2, p. 71-82, jan./jun. 2009. Disponível em: <<http://www.periodicos.ufsc.br/index.php/eb/article/view/10359/11023>>. Acesso em: 22 out. 2009.

BARANOW, Ulf Gregor. Perspectivas na contribuição da lingüística e de áreas afins á ciência da informação. **Ciência da Informação**, Brasília, 12(1): 23-25, 1983. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/1527>>. Acesso em: 15 set. 2009.

BARRETO, Aldo de Albuquerque. Glossário sobre a Ciência da Informação. **DataGramaZero: Revista de Ciência da Informação**, v. 8, n. 1, fev. 2007. Disponível em: <http://dgz.org.br/fev07/Ind_com.htm>. Acesso em: 20 out. 2009.

BELKIN, N. J. Information concepts for information science. **Journal of Documentation**, v. 34, n.1, p.55-85, mar.1978.

BORDENAVE, Juan E. Díaz. **O que é comunicação**. São Paulo: Editora Brasiliense, 1982. 105 p. (Coleção Primeiros Passos; 67)

BORKO, H. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5.
CINTRA, Anna Maria Marques *et al.* **Para entender as linguagens documentárias**. São Paulo: Editora Polis/APB, 1994, 72 p.

CINTRA, Anna Maria Marques. Elementos de lingüística para estudos de indexação. **Ciência da Informação**, Brasília, 12 (1): 5-22, 1983. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/1526>>. Acesso em: 30 out. 2009.

COSTA, Antônio Felipe Corrêa da. Ciência da Informação: o passado e a atualidade. **Ciência da Informação**, Brasília, 19(2): 137-43, jul./dez. 1990. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/viewArticle/1388>>. Acesso em: 20 out. 2009.

CUNHA, Isabel Maria Ribeiro Ferin. **Do mito à análise documentária**: a “luso-assimilação”. São Paulo: USP, 1987. 211 p. Tese (Doutorado) - Escola de Artes e Comunicações, Departamento de Biblioteconomia e Documentação, Universidade de São Paulo.

DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro, 7(2) : 101-107, 1978.

DERRIDA, Jacques. **Margens da filosofia**. Campinas: Papirus, 1991. 373 p.

DODEBEI, Vera Lucia Doyle Louzada de Mattos. **Representação, memória e linguagem**. In: Tesouro: linguagem de representação da memória documentária. Niterói: Intertexto, 2002, p. 16-59.

FOUCAULT, Michel. **Palavras e as coisas**: uma arqueologia das ciências humanas. 8 ed. São Paulo: M Fontes, 1981. 407 p.

FUJITA, Mariângela Spotti Lopes. A identificação de conceitos no processo de análise de assuntos para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul./dez. 2003. Disponível em: <<http://www.sbu.unicamp.br/seer/ojs/include/getdoc.php?id=40&article=9&mode=pdf>>. Acesso em: 30 out. 2009.

FURTADO, José Afonso. **O que é o livro**. Lisboa: Difusão Cultural, 1995.

GOMES, Henriette Ferreira. Interdisciplinaridade e Ciência da Informação: de característica a critério delineador de seu núcleo principal. **DataGramZero: Revista de Ciência da Informação**, v. 2, n. 4, ago. 2001. Disponível em: <http://www.dgz.org.br/ago01/Art_04.htm>. Acesso em: 20 set. 2009.

KAJI-MARKENFELDT, Alessandra Ribeiro. **A transformação do macaco em homem: o processo de hominização através da linguagem e do trabalho.** Brasília: UnB, 2006, 196 p. Dissertação (Mestrado) - Instituto de Psicologia, Universidade de Brasília.

KATZENSTEIN, Ursula Ephraim. **A origem do livro: da idade da pedra ao advento da impressão tipográfica no ocidente.** São Paulo: Hucitec, 1986. 455 p.

KOBASHI, N.Y.. Análise Documentária e representação da informação. **Revista Informare**, Rio de Janeiro, v. 2, n. 2, p. 5-27, 1996.

KOBASHI, Nair Yumiko Análise documentária e representação da informação. **INFORMARE – Caderno do Programa de Pós-Graduação em Ciência da Informação**, Rio de Janeiro, v. 2, n. 2, p. 5-27, jul./dez. 1996. Disponível em: <http://www.datagramazero.org.br/dez07/Art_01.htm>. Acesso em: 20 ago. 2009.

KRISTEVA, Julia. **Historia da linguagem.** Lisboa, 1974. 454 p.

LANCASTER, F. W. **Indexação e resumos: teoria e prática.** Brasília: Briquet de Lemos, 2004. 452 p.

LARA, Marilda Lopes Ginez de. **A representação documentária: em jogo a significação.** São Paulo: USP, 1993. Dissertação (Mestrado) - Escola de Artes e Comunicações, Departamento de Biblioteconomia e Documentação, Universidade de São Paulo.

LARA, Marilda Lopes Ginez de. Algumas contribuições da semiologia e da semiótica para a análise das linguagens documentárias. **Ciência da Informação**, Brasília, v. 22, n. 3, p. 223-226, set./dez. 1993. Disponível em: <<http://revista.ibict.br/index.php/ciinf/article/view/1129/778>>. Acesso em: 18 set. 2009.

LARA, Marilda Lopes Ginez de. O processo de construção da informação documentária e o processo de conhecimento. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 7, n. 2, p. 127-139, jul./dez. 2002. Disponível em: <<http://www.eci.ufmg.br/pcionline/index.php/pci/article/viewFile/399/216>>. Acesso em 15 out. 2009.

LARA, Marilda Lopes Ginez de. O Unicórnio (o Rinoceronte, o Ornitorrico...), a análise documentária e a linguagem documentária. **DataGramZero: Revista de Ciência da Informação**, v. 2, n. 6, dez. 2001. Disponível em: <http://www.dgz.org.br/dez01/Art_03.htm>. Acesso em: 10 nov. 2009.

LE COADIC, Yves-francois. **A ciência da informação**. Brasília: Briquet De Lemos, 1996. 119 p.

LIMA, Gercina Ângela Borém. Interfaces entre a ciência da informação e a ciência cognitiva. **Ciência da Informação**, v. 32, n. 1, p. 77-87, jan./abr. 2003. Disponível em: < <http://www.scielo.br/pdf/ci/v32n1/15975.pdf>>. Acesso em 20 ago. 2009.

LUCAS, Clarinda Rodrigues. Biblioteconomia: produção e administração da interpretação. **Ciência da Informação**, Brasília, v. 26, n. 1, p. 46-53, jan./abr. 1997. Disponível em: < http://www.scielo.br/scielo.php?pid=S0100-19651997000100007&script=sci_arttext>. Acesso em: 25 ago. 2009.

LYONS, John. **Linguagem e lingüística**: uma introdução. Rio de Janeiro: Editora Guanabara, 1987. 322 p.

MAIMONE, Giovani Deliberali. Aspectos gerenciais da análise documentária. In: **Proceedings 7º Encontro Nacional de Ensino e Pesquisa da Informação – CIFORM**. Disponível em: < <http://www.cinform.ufba.br/7cinform/soac/papers/91e130babfc0729d29c477e79bcf.pdf>>. Acesso em: 20 set. 2009.

MENDONÇA, Ercília Severina. A lingüística e a ciência da informação: estudos de uma interseção. **Ciência da Informação**, Brasília, v. 29, n. 3, p. 50-70, set./dez. 2000. Disponível em: < <http://revista.ibict.br/index.php/ciinf/article/viewArticle/224>>. Acesso em: 20 ago. 2009.

MOURA, Heronides. Relações paradigmáticas e sintagmáticas na interpretação de metáforas. **Linguagem em (Dis)curso - LemD**, v. 7, n. 3, p. 417-452, set./dez. 2007. Disponível em: < <http://www3.unisul.br/paginas/ensino/pos/linguagem/0703/7%20art%205.pdf>>. Acesso em 28 out. 2009.

NATALI, Johanna W. Documentação e lingüística: inter-relação e campo de pesquisa. **Revista Brasileira de Biblioteconomia e Documentação**, 11(1/2), p. 33-42, jan./jun. 1978.

NOVELLINO, Maria Salet Ferreira. A linguagem como meio de representação ou de comunicação da informação. **Perspectivas em ciência da informação**, Belo Horizonte, v. 3, n. 1, p. 137-146, jul./dez. 1998.

PINHEIRO, Lena Vania Ribeiro. Ciência da informação: desdobramentos disciplinares, interdisciplinaridade e transdisciplinaridade. Rio de Janeiro, IBICT/MCT, 2002. Disponível em: < <http://www.uff.br/ppgci/editais/lenavanialeituras.pdf>>. Acesso em: 16 set. 2009.

RUSSO, M. SANTOS; E. T. G. SANTOS; M. J. V. C. **Produção Científica Brasileira: da comunicação à recuperação via WEB.** Disponível em: <http://www.sibi.ufrj.br/trab_mariza_ago2001.doc>. Acesso em: 26 out. 2009.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996

SAUSSURE, Ferdinand de. **Curso de lingüística geral.** 12. ed. São Paulo: Cultrix 279 p.

SMIT, Johanna W. TÁLAMO, Maria de Fátima G. M. KOBASHI, Nair Y. A determinação do campo científico da Ciência da Informação: uma abordagem terminológica. **DataGramaZero: Revista de Ciência da Informação**, v. 5, n. 1, fev. 2004. Disponível em: < http://www.dgz.org.br/fev04/Art_03.htm>. Acesso em: 20 out. 2009.

SMIT, Johanna W. **Análise documentária:** a análise da síntese. Brasília: IBICT, 1987. 135 p.