

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE ESTUDOS SOCIAIS APLICADOS
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO E DOCUMENTAÇÃO**

**ANÁLISE DO USO DAS LINGUAGENS CONTROLADA E LIVRE
NAS ESTRATÉGIAS DE BUSCA EM BASES DE DADOS**

Ilza Leite de Azevedo Santos Lopes

Dissertação apresentada ao Departamento de Ciência da Informação e Documentação da Universidade de Brasília, como parte dos requisitos para obtenção do título de Mestre em Ciência da Informação.

Professor orientador: Dr. Murilo Bastos da Cunha

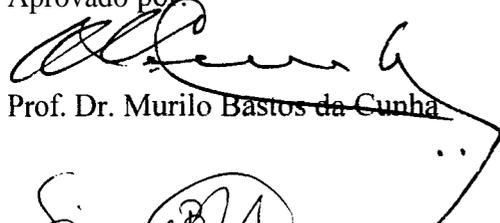
Brasília

2000

Dissertação apresentada ao Departamento de Ciência da Informação e Documentação da Universidade de Brasília como requisito parcial para obtenção do grau de Mestre.

Brasília, 06 de dezembro de 2000.

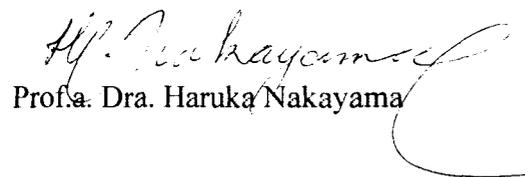
Aprovado por:



Prof. Dr. Murilo Bastos da Cunha



Prof.a. Dra. Simone Bastos Vieira



Prof.a. Dra. Haruka Nakayama

Prof.a. Sueli Angélica do Amaral - Suplente

Ainda que eu fale as línguas dos homens e dos anjos, se não tiver caridade, sou como bronze que ressoa, ou como címbalo que tine.

Ainda que eu tenha o dom da profecia e conheça todos os mistérios e toda a ciência, ainda que possua a fé em plenitude, a ponto de transportar montanhas, se não tiver caridade, nada sou.

Ainda que distribua todos os meus bens em esmolas e entregue o meu corpo a fim de ser queimado, se não tiver caridade, de nada me aproveita.

I Cor 13:1-3

A meus pais
Aristides e Joana
(In memoriam)

AGRADECIMENTOS

Ao Mestre dos Mestres, **Jesus**, a quem confiei as minhas mais profundas dúvidas e incertezas e que não me desamparou ao longo de toda essa caminhada.

À **Mãe Santíssima**, a quem tantas vezes recorri, pedindo forças para continuar a missão que me tinha sido destinada.

Aos amigos da grande jornada da vida que esperaram ansiosamente por este momento, de missão cumprida perante o Criador.

Ao professor Dr. Murilo Bastos da **Cunha**, da Unb, orientador incansável e paciente, pelo incentivo e apoio especial, nossa eterna gratidão.

À minha filha Angélica que acompanhou todo esse percurso de **vida**, acreditando em sua mãe.

Às minhas netas, **Amanda** e **Marina**, e ao meu “**filho**” Eli, a quem dediquei tão pouca atenção nos últimos meses.

Aos colegas das Bibliotecas do **IBICT**, da **UnB** e da **UCB**, pelo apoio e incentivo.

À Gleisse Ribeiro Alves, pelo apoio na obtenção de documentos "quase" impossíveis.

À Gláudia **Canongia**, amiga de todas as horas, principalmente as mais difíceis.

À Cláudia Regina de Araújo **Silva**, amiga inseparável das dificuldades logísticas.

Nossos mais sinceros agradecimentos a todos aqueles que direta e indiretamente colaboraram para a realização deste trabalho.

RESUMO

Esta pesquisa teve como **objetivo** principal o estudo sobre o planejamento da estratégia de busca num ambiente de bases de dados de acesso em CD-ROM, proporcionando uma perspectiva sobre a complexidade para a busca da informação conceitual, incluindo a **seleção** de termos para as estratégias e a função do vocabulário controlado ou da linguagem natural nesse contexto. Estudou, ainda, o processo de entrevista inicial para atendimento à demanda de informação e sua interferência no planejamento da estratégia de busca, além de descrever as etapas do planejamento da mesma e os critérios para seleção de bancos e/ou de bases de dados.

Pretendeu-se estudar os fatores que compõem o planejamento da estratégia de busca, **analisando-se** a influência do uso do vocabulário natural e do controlado na formulação da estratégia, **objetivando** identificar qual das linguagens obteria o melhor desempenho. Para a análise do uso da linguagem controlada na base de dados *Psychological Abstracts Literature (PSYCLIT)*, foram utilizados os conceitos em campos de registros de dados específicos e, comparativamente, para a análise da linguagem natural na estratégia, foram usados conceitos livres nos campos exclusivos onde ocorrem esses termos. Numa avaliação dos resultados os usuários identificaram os melhores itens recuperados em função da precisão obtida, fornecendo portanto, os elementos primordiais para avaliação dos termos que proporcionaram os resultados mais satisfatórios.

ABSTRACT

Analysis on the use of the controlled or natural languages on the search strategies in databases.

The principal purpose of this study is the planning of the search strategy in the CD-ROM database environment providing a perspective on the complexity of searching for conceptual information, including the selection of terms to the search strategies and the role of vocabulary control or the free-text in this context. Describes also, the process of the preliminar interview in order to observe its interference on the planning of the search strategy, describing the steps in the planning of the strategies and the criterias for the selection of the information retrieval systems and the databases.

The goal is intend to study the factors that participate on the composition of the search strategy, analysing the influences on the use of the controlled vocabulary in the planning of the strategy, and trying to identificate which language has the better performance. In order to evaluate the possibility to use controlled language in *the Psychological Abstracts Literature (PSYCLIT)* database, were used the concepts which were registred on specific search fields, and comparatively, for the analysis of the natural language terms that were used in the search strategy, the free concepts in the search fields where these terms occured. The users had identificate on an process of evaluation of the results, which language had offered the better results, giving yours opinions about the process of these evaluation.

SUMÁRIO

Resumo	.v
Abstract	vi
Lista de siglas e/ou abreviaturas	ix
Lista de figuras	xi
Lista de tabelas	xii
1 INTRODUÇÃO	1
1.1 Definição do problema	3
1.2 Objetivos	4
1.2.1 Objetivo geral	4
1.2.2 Objetivos específicos	5
1.2.3 Hipóteses	5
1.2.4 Definições operacionais	6
1.3 Justificativa	7
2 REVISÃO DE LITERATURA	9
2.1 Estudos sobre estratégia de busca	10
2.1.1 Processo de entrevista	18
2.1.2 Técnicas de estratégia de busca	23
2.1.3 Estratégia de busca - etapas	30
2.1.4 Implementação da estratégia de busca	34
2.2 Estudos sobre as linguagens: natural e controlada	37
2.2.1 LC e LN: Vantagens e Desvantagens	48
2.2.2 LC e LN em bases de dados	55
2.2.3 Considerações sobre a Revisão de Literatura	62
3 METODOLOGIA	64
3.1 Introdução	64
3.2 Escolha da Metodologia	64
3.3 Métodos e técnicas	65
3.4 Amostra e população	67
3.5 Coleta de dados	69
3.6 Análise dos dados	70
3.7 Limitações da pesquisa	71
4 ANÁLISE DOS DADOS	72
4.1 Introdução	72
4.2 Planejamento da estratégia de busca	72
4.2.1 Estratégia de busca utilizando linguagem controlada - LC	76
4.2.2 Estratégia de busca utilizando linguagem natural - LN	81
4.2.3 Estratégia de busca utilizando LC e LN simultaneamente	86
4.3 Análise comparativa dos resultados	90
S CONCLUSÕES	94
5.1 Introdução	94
5.2 Testes das hipóteses	95
5.2.1 Hipótese nº 1	95
5.2.2 Hipótese nº 2	96
5.2.3 Hipótese nº 3	96
5.3 Objetivos	97
5.3.1 Objetivo A: Adequação da LC	97
5.3.2 Objetivo B: Adequação da LN	98
5.3.3 Objetivo C: Adequação da LC+LN	99

5.3.4	Objetivo D: Identificação da melhor linguagem	100
5.4	Conclusões finais	100
5.5	Sugestões para estudos futuros	101

	REFERÊNCIAS BIBLIOGRÁFICAS	103
--	-----------------------------------	------------

	ANEXOS	111
--	---------------	------------

Anexo I - Formulário do Serviço de Busca Bibliográfica

Anexo II - Avaliação do Serviço de Busca Bibliográfica - questionário

Anexo III - Formulário de Avaliação das Referências

Anexo IV - Bases de dados: Linguagens controladas

Anexo V - Estratégias de busca: detalhamento

LISTA DE SIGLAS E/OU ABREVIATURAS

ACM	American Computing Machinery
APA	American Psychological Association
APILIT	American Petroleum Institute - Literature
APIPAT	American Petroleum Institute - Patents
ARIST	Annual Review of Information Science and Technology
ASIS	American Society of Information Science
BA	Biological Abstracts
BIOSIS	BioSciences Information Services
BRS/AFTERDARK	Bibliographical Retrieval Services/AfterDark
CA	Chemical Abstracts
CDD	Classificação Decimal de Dewey
CDI	Comprehensive Dissertation Index
CD-ROM	Computer Disk-Read Only Memory
CDU	Classificação Decimal Universal
COMPENDEX	Computerized Engineering Index
EPA	Environmental Protection Agency
ERIC	Educational Resources Information Center
INIS	International Nuclear Information System
INSPEC	Information Services for Physics, Electrotechnology and Control
ISA	Information Science Abstracts
ISO	International Standard Organization
ISSN	International Standard Serial Number
IUPAC	International Union of Pure and Applied Chemistry
LC	Linguagem controlada
LCSH	Library of Congress Subject Headings
LISA	Library and Information Science Abstracts

LN	Linguagem natural
JASIS	J ournal of American Society for Information Science
NASA	National Aeronautical Space Agency
NTIS	National Technical Information Service
OPAC	Online Public Access Catalog
PRECIS	P reserved Context index System
PSYCLIT	P sychological Literature
SCI	Science Citation Index
UCB/SIBI	Universidade Católica de Brasília/Sistema de Bibliotecas
UNISIST	United Nations Information Systems in Science and Technology

LISTA DE FIGURAS

Figura 1	Estrutura da informação para representação de assunto	58
Figura 2	LC e LN: registro na Base PSCYLIT	66
Figura 3	Quadro demonstrativo das referências recuperadas	74
Figura 4	índice comparativo das estratégias	75
Figura 5	LC: referências recuperadas	80
Figura 6	LN: referências recuperadas	85
Figura 7	LC e LN: referências recuperadas	89
Figura 8	Comparação dos resultados das estratégias para o índice Muito Bom	90
Figura 9	Comparação dos resultados das estratégias para o índice Bom	91
Figura 10	Comparação dos resultados das estratégias para o índice Irrelevante	92

LISTA DE TABELAS

Tabela 1	Entrevista <i>versus</i> Aconselhamento	22
Tabela 2	Vocabulário controlado: vantagens e desvantagens	52
Tabela 3	Linguagem natural: vantagens e desvantagens	54
Tabela 4	Base de dados PSYCLIT: campos	65
Tabela 5	Objetivo <i>versus</i> hipóteses	69
Tabela 6	LC: termos de busca <i>versus</i> estratégia	77
Tabela 7	Linguagem controlada - referências recuperadas	78
Tabela 8	LN: termos de busca <i>versus</i> estratégia	82
Tabela 9	Linguagem natural - referências recuperadas	83
Tabela 10	LC e LN: temas de busca <i>versus</i> estratégia	86
Tabela 11	LC e LN - referências recuperadas	87

1. INTRODUÇÃO

O **objetivo** deste estudo é o de examinar o processo de planejamento da estratégia de busca num ambiente de bases de dados de acesso em CD-ROM ou em linha, procurando identificar qual o tipo de vocabulário que deve ser empregado na formulação da mesma. Bases de dados vêm sendo desenvolvidas intensivamente desde a década de 70, cobrindo praticamente todas as áreas do conhecimento humano, incluindo tanto campos multidisciplinares como energia e meio ambiente, quanto disciplinas tradicionais e específicas como Psicologia, objeto do presente estudo de caso.

As bases de dados podem conter informações bibliográficas, numéricas, referenciais como as **comumente** encontradas nos **diretórios** e textuais como os jornais, enciclopédias e alguns periódicos mais representativos. Os produtores dessas bases de dados incluem: associações profissionais, agências governamentais, instituições privadas, organismos internacionais e outros. Cada base de dados pode ser **acessada** em linha em diferentes bancos de dados, ou, ainda, ser comercializada em CD-ROM, para acesso local ou em rede. Por esse motivo, os bancos de dados e os fornecedores de **CD-ROM**, desenvolvem suas próprias linguagens de busca para as bases oferecidas sob suas tutelas, fornecendo também documentação apropriada sobre a estrutura da base e o vocabulário usado na indexação.

Num contexto de recuperação da informação em bases de dados, o uso na estratégia de busca, de conceitos baseados em linguagem natural e/ou linguagem controlada, tem sido objeto de estudos, com o objetivo de aprimorar os sistemas de recuperação da informação. Portanto, a decisão a ser tomada na escolha dos conceitos, para a elaboração das estratégias de busca, vai provavelmente afetar os resultados a serem obtidos numa busca, sendo necessário o conhecimento prévio da estrutura da informação e da política de indexação dos documentos na base de dados a ser consultada.

Num projeto de investigação cujos resultados foram publicados em 1991, **FIDEL** observou que os intermediários que executam as buscas consideram várias alternativas para a **seleção** dos termos da estratégia, baseados em grande parte, pela experiência prévia e pelas **próprias** condições oferecidas no ambiente de trabalho relacionadas com a recuperação de informação. Fundamentada na observação e análise dos protocolos verbais e de busca dos termos para a estratégia, **Fidel** propôs uma série de regras para determinar quando os descritores da linguagem controlada ou os termos da linguagem natural devem ser utilizados. Saliou ainda, que de acordo com o ambiente de trabalho dos intermediários, se prático, teórico ou geral, a escolha dos termos das estratégias de busca eram **selecionadas** de diferentes subconjuntos de regras.

Na estratégia de busca, o uso exclusivo da linguagem controlada, significa que a mesma terminologia utilizada pelo indexador também será utilizada pelo usuário, tanto no estágio de

entrada da pergunta (na estratégia de busca), quanto no **estágio** de saída (na resposta). O uso da linguagem controlada na indexação das bases de dados proporciona certas vantagens, como na resolução dos problemas semânticos da comunicação, uma vez que **permite** a interpretação adequada dos termos e seus contextos ao apresentar as relações hierárquicas existentes entre os termos. Uma outra vantagem é o estabelecimento de vínculos entre assuntos relacionados, reduzindo a ambiguidade entre os termos e orientando o usuário na identificação de todos os termos passíveis de serem utilizados na estratégia de busca.

A linguagem livre ou natural, numa estratégia de busca, tem uma função complementar, que vai ampliar a possibilidade de recuperação de documentos pertinentes, enriquecendo, **dessa** forma, o universo de informações recuperadas. Essa linguagem está, por outro lado, mais próxima do usuário, que desconhece as convenções de uma linguagem controlada.

O uso da linguagem natural no processo de recuperação de informação apresenta, dentre outras vantagens, a rapidez na entrada de dados no sistema, minimizando a influência da **subjetividade** do indexador e ampliando a facilidade de elaboração de estratégias de busca. O uso da linguagem natural, numa estratégia de busca, torna desnecessário o prévio conhecimento da estrutura complexa dos **vocabulários** controlados, para a **interação** com os sistemas de recuperação de informação **e/ou** com as bases de dados.

Na recuperação da informação, utilizando-se bases de dados em **CD-ROM** ou nos sistemas **conversacionais**, a estratégia de busca funciona como elo de ligação entre a informação armazenada e a necessidade de informação do usuário. Portanto, a decisão relacionada com o uso da linguagem controlada ou da linguagem natural pertence ao intermediário e/ou usuário que processa a busca.

Constata-se que os índices das bases de dados são, de certa forma, obstáculos para o usuário final que tem dificuldade em definir sua necessidade de informação na mesma terminologia da base a ser consultada. Além disso, a falta de familiaridade com a linguagem controlada, com o uso da sinonímia e variações de grafia são parte das dificuldades a serem transpostas, tanto pelo usuário, durante a sua solicitação de busca, quanto pelo intermediário, na fase de planejamento da estratégia de busca.

Essas constatações motivaram o desenvolvimento da presente pesquisa, na tentativa de identificar qual a **combinação** ideal de termos e/ou conceitos que deve ser usada no planejamento da estratégia de busca, analisando-se de forma criteriosa a utilização da linguagem natural e da linguagem controlada, além da **combinação** das duas, visando à obtenção de subsídios para o planejamento e organização de programas de treinamento no uso de bases de dados conversacionais e em CD-ROM.

1.1 Definição do problema

No planejamento de estratégia de busca em bases de dados em **CD-ROM** e/ou de acesso em linha, são frequentes algumas questões quanto ao uso de vocabulário controlado ou de linguagem natural na elaboração da estratégia:

- a) quando usar a linguagem controlada na formulação da estratégia de busca?
- b) quando usar a linguagem natural na formulação da estratégia de busca?
- c) quando usar a linguagem natural e a controlada simultaneamente na estratégia de busca?
- d) qual a linguagem mais adequada para a **operacionalização** de estratégias de busca?

A linguagem controlada no contexto da busca em **CD-ROM** e/ou em linha pode abranger, dependendo da base de dados a ser consultada, um conjunto de alternativas, tais como: códigos de classificação (como na base de dados **COMPENDEX** e nas bases de Patentes); códigos de **seção** (como no **CHEMICAL ABSTRACTS**); códigos de produtos industrializados e de eventos (como os das bases do **PREDICASTS**); ou, ainda, listas de cabeçalhos de assunto e os *thesauri* de diversas outras bases de dados.

LANCASTER (1993) define o vocabulário controlado como "um conjunto de **termos** que devem ser usados num sistema de recuperação de informação, tanto no estágio de entrada, quanto no de saída".

Por outro lado, uma lista de cabeçalhos de assunto ou um *thesaurus*, citando apenas dois exemplos de vocabulários controlados, não permanece estática. Ao longo dos anos, são incorporados novos termos e a tendência é aumentar a especificidade, ou, ainda, provocar a substituição desses termos a partir de uma determinada data, como por **exemplo** no *thesaurus* produzido pelo American Petroleum **Institute** para a indexação das bases de dados de petroquímica: **APILIT** e **APIPAT**. A **National Library** of Medicine também atualiza a sua **publicação** *Medical Subject Headings*, incluindo junto a grande parte dos termos, a data em que os mesmos foram **atualizados**. Esses indicadores orientam o intermediário que planeja as estratégias de busca sobre os conceitos que podem ou não ser incluídos nas mesmas a partir de um determinado período, ajustando os termos para atendimento às demandas dos usuários.

Esse dinamismo produz sérias implicações na **formulação** da estratégia de busca, pois as instituições que se propõem a implantar e/ou manter um serviço de consulta a bancos e/ou bases de dados internacionais têm que manter, devidamente atualizadas, todas as linguagens

documentárias e/ou vocabulários controlados utilizadas para a indexação dos documentos nas bases.

O especialista de informação, também chamado intermediário, que vai planejar a estratégia de busca, não pode subestimar esse fato, considerando-se os efeitos negativos que podem repercutir na recuperação da informação.

SIEVERT e **BOYCE** (1983) analisam os vocabulários controlados, **circunscritos**, aos *thesaurie* listas de cabeçalhos de assunto, sob a função tradicional de agrupar os termos variáveis usados na linguagem natural, em termos controlados. Reconhecem, porém, que os recursos da linguagem de busca em linha, modificaram **consideravelmente** a função do vocabulário, por que a busca de um termo **não-qualificado**, isto é, em qualquer campo do registro de informação, pode ocorrer tanto no título e no resumo (campos de linguagem natural) e também no descritor (campos de linguagem controlada).

LANCASTER (1986) aponta os vocabulários controlados das bases como um dos fatores que afetam a recuperação da informação, pois a geração desses vocabulários vai exigir um criterioso planejamento na construção e/ou **atualização** dos mesmos, principalmente no que se refere ao controle de hierarquia e na rede de referências cruzadas. Afirma ainda, que:

o vocabulário controlado deve relacionar sua própria terminologia, com as variações de expressões que ocorrem na literatura, **compatibilizando-as** com as expressões das solicitações de busca feitas ao sistema. (Lancaster, 1986, p. 151).

Finalizando esta análise, convém destacar que o intermediário que operacionaliza a estratégia de busca, conhecedor, portanto, dessas linguagens de comando para a recuperação da informação, muitas vezes necessita refinar o conjunto recuperado, utilizando, **justamente**, o campo de busca que proporciona maior precisão, ou seja, o descritor que foi atribuído da linguagem controlada no processo de entrada de **dados/indexação**.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo geral foi:

Estudar o processo de planejamento de estratégia de busca, num ambiente de bases de dados em CD-ROM e/ou de acesso em linha, verificando qual o tipo de vocabulário que deve ser empregado na mesma.

1.2.2 Objetivos específicos

Os objetivos específicos foram:

- a) identificar a adequação do uso do vocabulário controlado na elaboração da estratégia de busca;
- b) identificar a adequação do uso da linguagem natural na elaboração da estratégia de busca;
- c) identificar a adequação do uso simultâneo do **vocabulário** controlado e do natural na elaboração da estratégia de busca;
- d) identificar qual a linguagem mais adequada para a **operacionalização** da estratégia de busca.

1.2.3 Hipóteses

As hipóteses a serem testadas são:

H1 - Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem controlada (LC) não contribuem para o aumento de citações relevantes recuperadas na base de dados PSYCLIT.

Procura-se verificar por essa hipótese o relacionamento entre as variáveis:

- a) independente: o uso de termos e/ou conceitos em linguagem controlada;
- b) dependente: total de citações relevantes recuperadas na Base PSYCLIT;
- c) intervenientes: lógica e recursos da linguagem de recuperação da base de dados.

H2 - Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem natural (LN) não contribuem para o aumento de citações relevantes recuperadas na base de dados: PSYCLIT (**Psychological Abstracts- Literature**)

Procura-se verificar por essa hipótese o relacionamento entre as variáveis:

- a) independente: o uso de termos e/ou conceitos em linguagem natural;
- b) dependente: total de citações relevantes recuperadas na base PSYCLIT;
- c) intervenientes: lógica e recursos da linguagem de recuperação da base de dados.

H3 - Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza simultaneamente a LC e a LN podem contribuir para o aumento de citações relevantes recuperadas na base de dados PSYCLIT.

Procura-se verificar por essa hipótese o relacionamento entre as variáveis:

- a) independente: o uso simultâneo de **termos/conceitos** de LC e LN;
- b) dependente: total de citações relevantes recuperadas na base PSYCLIT;
- c) interveniente: lógica e recursos da linguagem de recuperação da base de dados.

1.2.4 Definições operacionais

Bases de dados — no contexto dessa pesquisa, é um conjunto de referências bibliográficas armazenados em **forma** legível por computador.

Citações relevantes — citações e/ou referências consideradas exatamente de acordo com o interesse do usuário; categorizadas neste trabalho com o conceito "**Muito Bom**" (MB), no formulário de avaliação (Anexo 3).

Citações periféricas — citações e/ou referências consideradas de âmbito mais genérico do que o interesse **atual** do usuário e que podem vir a ser úteis em futuro próximo, categorizadas neste trabalho com o conceito "Bom" (B), no **formulário** de avaliação (Anexo 3).

Citações irrelevantes — citações e/ou referências consideradas fora do perfil de interesse do usuário, categorizadas nesse trabalho com o conceito "Irrelevante" (I), no formulário de avaliação (Anexo 3).

Linguagem controlada — a linguagem que é utilizada nas bases de dados apenas nos campos: descritor, termos de indexação, identificadores. A linguagem utiliza termos e/ou conceitos representativos do conteúdo do documento a partir de: **thesaurus**, lista de cabeçalhos de assunto, classificações especializadas, relações de códigos e outros.

Linguagem natural — a linguagem do discurso **técnico-científico**; na base de dados, corresponde às palavras contidas no título e no resumo do documento referenciado.

Linguagem de recuperação — a linguagem lógica e/ou de interface, que permite o encontro da informação solicitada com a pergunta formulada na base de dados, mediante o uso de comandos definidos pelo sistema de recuperação.

Estratégia de busca — técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada num sistema de recuperação da informação, num banco de dados e/ou numa base de dados ou, ainda, nos catálogos públicos de acesso em linha.

1.3 Justificativa

O planejamento das estratégias de busca que visam à recuperação de informações sob demanda tem sido uma questão que vem sendo estudada, exaustivamente, há quase três **décadas**, no contexto de bancos e bases de dados de acesso remoto.

Historicamente, uma base de dados é o resultado do avanço tecnológico na área do uso de computadores para registro de grandes volumes de informação. Isso significa que a mesma é um subproduto da geração dos "**indices e abstracts**" e das bibliografias impressas e, portanto,

com algumas **exceções**, os seus índices foram estruturados apenas com a função de organização da informação.

Entretanto a busca em CD-ROM **e/ou** em linha, requer um conjunto de operações que, entre outras incluem: a identificação do tema a ser pesquisado; a **seleção** da base que mais se identifica com o tema; a identificação de termos mais apropriados, os sinônimos e as grafias **alternativas**; o uso de combinações lógicas inclusivas ou exclusivas, limitação de período de tempo; limitação de idiomas não desejáveis e outras.

Num programa de treinamento de usuários para utilização de bases de dados, toma-se imprescindível abordar a questão do uso, de termos extraídos da linguagem natural ou da linguagem controlada e/ou documentária utilizada pelo produtor da fonte de informação. Assim, no planejamento da estratégia de busca, vários fatores devem ser levados em consideração, além dos que estão relacionados simplesmente com a identificação do assunto, **justificando-se** dessa forma o presente estudo.

2. REVISÃO DA LITERATURA

A **revisão** da literatura versou sobre os temas: estratégia de busca, incluindo o processo de entrevista, as etapas de elaboração da estratégia e a implementação da mesma. Incluiu, ainda, estudos sobre as linguagens natural e controlada como instrumentos de recuperação da **informação**.

Como o objetivo da pesquisa é identificar a adequação do uso das linguagens natural e/ou controlada na elaboração da estratégia de busca, a revisão focalizou a literatura que analisa a **operacionalização** da estratégia de busca com ênfase na linguagem escolhida.

A revisão procurou ser exaustiva, baseando-se na busca dos trabalhos mais representativos nas seguintes fontes de informação especializadas em **CD-ROM**, em linha e via Internet:

- **Library and Information Science Abstracts (LISA)**, de 1969-2000;
- **Information Science Abstracts (ISA)**, de 1970-2000;
- **Library Literature**, de 1980-2000;
- **Educational Resources Information Center (ERIC)**, de 1970-2000 ;
- **ACM Electronic Guide to Computing Literature**, de 1980-2000 ;
- **Proquest - General Periodicals On Disc**, de 1990-2000 e,
- **Annual Review of Information Science and Technology (ARIST)**

A revisão de literatura foi dividida em dois tópicos: estudos sobre estratégia de busca e estudos sobre o uso das linguagens natural e controlada na recuperação da **informação**. Em relação aos estudos sobre estratégia de busca, foram mencionados o processo de entrevista, as técnicas de elaboração de estratégia, as etapas de busca e a implementação da estratégia de busca, **objetivando contextualizar** o problema. No tópico relativo aos estudos sobre as linguagens natural e controlada destacam-se, aqueles relacionados com o processo de recuperação de informação em bases de dados.

Na literatura nacional, o tema não tem sido **objeto** de estudo contínuo, motivo pelo qual foi enfatizada, a literatura internacional.

2.1 Estudos sobre estratégia de busca

Objetivando situar a presente pesquisa no contexto da recuperação da informação, alguns conceitos fundamentais são apresentados para o entendimento da questão relacionada com o processo de planejamento das estratégias de busca, tais como:

Estratégia é um termo conhecido por suas aplicações militares e interage com o conceito **tática**, também de aplicação militar, no que se refere à forma de execução da estratégia de busca.

Define-se a seguir os três conceitos, como citados no Dicionário Aurélio.

1) Estratégia:

- Arte militar de planejar e executar movimentos e operações de tropas, navios e/ou aviões, visando a alcançar ou manter posições relativas e potenciais bélicos favoráveis a futuras ações táticas sobre determinados objetivos.
- Arte militar de escolher onde, quando e com que travar um combate ou uma batalha.
- Arte de aplicar os meios disponíveis com vista à consecução de objetivos específicos.

2) Busca:

- Ato ou efeito de buscar.
- Procura com o fim de encontrar alguma coisa;
- Investigação cuidadosa; pesquisa, exame;
- Procura minuciosa; revista, exame;
- Movimento íntimo para alcançar um fim.

3) Tática:

- Parte da arte da guerra que trata da disposição e da manobra das forças durante o combate ou na iminência dele.
- Parte da arte da guerra que trata de como travar um combate ou uma batalha.
- Fig. Processo empregado para sair-se bem num empreendimento.

Para alcançar a resposta pretendida pelo usuário de informação, faz-se necessária a execução de movimentos e operações táticas, ora restringindo os resultados alcançados, ora

ampliando-os para a **obtenção** de informações **mais** relevantes, conforme o pedido de busca demandado.

A estratégia de busca pode ser definida como uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada numa base de dados. Isto significa que, a partir de um **arquivo**, um conjunto de itens que constituem a resposta de uma determinada pergunta será selecionado.

A escolha de qual **banco/base** de dados é o mais adequado, qual a linguagem a ser empregada, se a natural ou se a controlada, bem demonstra que o planejamento e execução de uma estratégia de busca envolve escolhas que irão determinar a obtenção eficaz daquelas informações específicas solicitadas pelo usuário ao serviço de informações.

BOURNE (1977) investigou o impacto dos erros ortográficos nas bases de dados bibliográficas e como eles afetam a produtividade das buscas em linha. Analisou erros ortográficos que ocorrem na linguagem natural, isto é, nos campos de títulos e resumos, assim como os que ocorrem nos campos de descritores das linguagens controladas. Relembra que os grandes bancos de dados oferecem **múltiplas** bases de dados para consulta, algumas com milhões de itens de informação e que esses bancos desenvolveram linguagens de busca que permitem a consulta textual para pesquisa nas palavras contidas nos títulos, resumos e outros campos de dados. Afirma, ainda, que a disponibilidade e uso de todas essas facilidades têm **contribuído** para o **desenvolvimento** de linguagens de indexação associadas com essas bases de dados. Observa, entretanto, que a quantidade de erros de ortografia, formas variantes de palavras e mesmo erros ortográficos dos termos de indexação afetam **consideravelmente** os resultados das buscas. Lembra que, no índice impresso, essa alta incidência de erros ortográficos não é tolerada e, portanto, a mesma atitude deve ser mantida quanto ao índice das bases de dados.

Constatamos, na vida profissional, que o usuário (fina), está pouco habilitado a lidar com a incidência de erros ortográficos em bases de dados. O **intermediário**, que dispõe de maior habilidade para utilizar as linguagens controladas, obtém melhor desempenho em consequência do conhecimento prévio desses e de outros fatores durante o processo de planejamento e **operacionalização** da estratégia de busca.

BATES (1981) afirmou que a estratégia de busca é um plano completo de uma busca de **informação**, porém, alguns autores a confundem com **formulação** de busca. A formulação é apenas um conjunto de frases que expressam a pergunta do solicitante num formato compatível com a **base** e o banco de dados onde a mesma será **realizada**. Entretanto, a estratégia de busca é um plano mais completo que engloba uma ou mais **táticas**, além da própria formulação da busca, que são as frases estruturadas de forma a serem bem sucedidas para a obtenção de resultados **satisfatórios**.

HARTER (1984), num estudo **exploratório**, analisou os diversos estilos de busca em linha, com ênfase no comportamento e atitude dos intermediários. Esse estudo revela que existem enormes diferenças entre os **intermediários** que **operacionalizam** as buscas. Os intermediários de instituições acadêmicas diferem daqueles de outros tipos de biblioteca, principalmente das especializadas, no que se refere ao aprofundamento das estratégias de busca.

Esses resultados apresentados por **Harter** confirmam nossa observação pessoal ao longo de muitos anos atuando em programas de treinamento para o público das universidades brasileiras. **Efetivamente**, a superficialidade na abordagem dos temas solicitados pelos intermediários de bibliotecas **acadêmicas** está diretamente subordinado ao fato de a clientela atendida não exigir um maior aprofundamento nos resultados das **buscas-solicitadas**. Entretanto, tal não ocorre com os intermediários de instituições especializadas onde o usuário, num degrau profissional mais atuante, necessita aprofundar seus conhecimentos, muitas vezes, para solucionar problemas de laboratórios de pesquisa e desenvolvimento. Essa enorme **diferença**, como apontada pelo autor, influencia no planejamento dos programas de treinamento a serem oferecidos.

Noutro estudo de **Harter**, também de 1984, foi apresentado um modelo filosófico e **comportamental** para os especialistas da busca em linha e para o processo de recuperação da informação baseado nos padrões da investigação científica. Descrevendo a natureza da pesquisa científica, traçou uma analogia com os conceitos utilizados na pesquisa: variáveis, formulação de hipóteses e teste, definições **operacionais**, validade, **confiabilidade** e suposições, correlacionando os parâmetros do modelo com a busca em linha.

Num estudo mais recente, em 1990, HARTER propôs uma metodologia para estudar os efeitos de **seleção** de termos e o ajuste dos resultados na recuperação da **informação**. Supondo que o agrupamento dos resultados de várias estratégias de busca formuladas por diferentes intermediários para um dado tema, pode ocultar importantes relações, sugeri então, outras pesquisas para o **aprimoramento** da estratégia de busca.

TENOPIR (1985) analisou o desempenho da LN e da LC na *Harvard Business Review Online* uma base de dados textual, onde foi aplicada uma metodologia de busca, separadamente, por texto completo, por resumo, por título, e nos descritores para 36 questões da amostra da pesquisa. Cada uma das diferentes estratégias de busca produziu documentos únicos e a busca no texto integral produziu maior recuperação e menor precisão do que a busca com LC. Uma outra característica das bases de dados textuais é o número de pontos de acesso. Tipicamente, numa grande base de dados textuais será bem mais difícil encontrar uma recuperação aceitável com uma precisão tolerável. As bases de dados textuais proporcionam uma grande recuperação, porém baixa precisão, enquanto as bases bibliográficas oferecem melhor desempenho.

FIDEL (1986), num estudo sobre sistemas especialistas desenvolvidos para auxiliar os usuários finais na consulta as bases de dados, demonstrou que a maioria desses sistemas são baseados na análise dos textos, e não nos modelos de busca humana. Por esse motivo, tais

sistemas não podem processar os **critérios** relacionados com o pedido de busca e responder às questões relativas aos fatores de **revocação** e precisão. Uma análise do comportamento de busca dos **intermediários** revelou a rotina que utilizam no processo de **seleção** das **chaves-de-busca** quando optam por escolher termos e/ou conceitos da linguagem natural e/ou da linguagem controlada dentro de um conjunto de outras **ações**, que fazem parte do planejamento da estratégia de busca. Sugeriu que certas regras decisórias dos intermediários fossem automatizadas para ampliar a capacidade dos sistemas especialistas, principalmente as que foram detectadas durante a seleção de conceitos para a estratégia. A autora denominou essa etapa de Rotina de Seleção de Termos, afirmando que, com certos ajustes, ela pode ser automatizada, conduzindo o intermediário e/ou usuário para decisões adequadas quanto aos termos a serem incluídos no **planejamento** da estratégia de busca.

BATES (1988) conceituou a estratégia de busca como o "estudo da teoria, princípios e prática de planejar e executar **táticas** e estratégias de **busca**". A autora destaca-se na literatura por ter sido a primeira a definir teoricamente o conceito de estratégia de busca e a **tática** para a sua execução. Adaptando as definições ao ambiente de recuperação da **informação**, deu ênfase ao conceito de comportamento de **busca**, indagando: "o que as pessoas fazem, e como pode ser determinado o que elas pensam, quando estão executando uma busca de informação?" Nesse momento, cabe uma análise semântica dos principais termos que compõem o **objeto** de estudo, lembrando que o planejamento e a execução de uma estratégia de busca é uma "**arte** de escolher onde, quando e com que investigar cuidadosamente" a fonte de informação para alcançar os **objetivos** específicos do solicitante. A escolha de onde buscar, identificando qual a base de dados é a melhor qualificada para atendimento das perguntas específicas do solicitante, depende de uma série de fatores relacionados com a **abrangência** de assunto e **período** da base, dos tipos de documentos indexados, da linguagem de indexação, dos campos de busca **disponibilizados** em determinado banco de dados e outros.

Num outro estudo sobre estratégias de busca, em 1990, BATES, separou o processo de busca em bases de dados em duas partes: uma realizada por indivíduos e a outra pelo sistema de recuperação. Observou que o processo de busca efetuado por pessoas é constituído por diferentes movimentos, táticas, estratégias e estratégias. Definindo o movimento como um pensamento ou **ação** que é parte do processo de busca, demonstrou que a tática é composta por um ou mais movimentos executados no mesmo processo. O estratégia seria constituído pelo uso de múltiplas táticas e/ou movimentos e, a estratégia seria o planejamento que conteria os movimentos, as táticas e os estratégias para a completa busca da informação. O processo de busca realizado pelo sistema de recuperação seria composto dentre outros, pela linguagem de busca do sistema e pela estrutura de informação da base de dados.

WILDEMUTH et al. (1991) analisaram o comportamento do usuário final no que concerne à formulação de suas frases de busca. Um estudo foi realizado para analisar como o usuário final

se comporta quando está executando a sua própria estratégia de busca. Todas as **táticas** de busca foram identificadas visando a obtenção de dados que auxiliassem no desenvolvimento de programas de treinamento para os usuários finais. Observou-se que a preocupação com os problemas que afetam as buscas e com a obtenção de resultados mais satisfatórios vinham de longa data. No entanto, apesar dos estudos realizados para identificação de programas de treinamento que tomassem os resultados de busca dos usuários finais mais **efetivos**, não houve melhora no desempenho dos mesmos.

TENOPIR, NAHL e HOWARD (1991) aplicaram, num experimento com usuários finais, uma medida de avaliação do comportamento dos mesmos durante o processo de planejamento e **execução** de suas próprias estratégias de busca. **Nessa** pesquisa, os usuários identificaram e descreveram como e por que foi escolhida cada estratégia. O comportamento **afetivo** dos usuários finais, exibido durante as buscas, demonstrou que existe um envolvimento mais abrangente do que o cognitivo sobre as **atividades** de busca e destaca o interesse individual dos usuários finais e o cuidadoso planejamento das estratégias de busca. Os autores descreveram, ainda, a avaliação dos próprios usuários sobre a utilidade das referências recuperadas, **comparando-as** com as estratégias formuladas e **operacionalizadas** pelos mesmos.

SPINK e SARACEVIC (1993), numa série de experimentos sobre a **seleção** de termos **e/ou** conceitos para o planejamento da estratégia de busca de 1991 a **1993**, estudaram usuários e intermediários em situações reais de buscas **interativas** em linha. Esses experimentos propunham-se a identificar em qual estágio do processo de busca os termos **e/ou** conceitos eram **selecionados** e de quais vocabulários controlados os mesmos eram obtidos. Analisaram, também, a sequência na qual os termos eram selecionados nas diferentes linguagens controladas, para constatar a frequência das mudanças ocorridas nessas sequências. Afirmaram, baseados na prática profissional e em pesquisas, que a busca em sistemas de recuperação de informação é:

um processo de alta complexidade envolvendo numerosos fatores e variáveis além de decisões e o entrelaçamento dos **sub-processos inter-relacionados** com a busca. Uma das chaves para um dos **sub-processos** é a da seleção de termos para a estratégia de busca que por sua vez é influenciada por outros fatores, particularmente os relacionados com os resultados. Sintetizam essa questão com uma pergunta para as investigações: - Que termos de busca devem ser selecionados para um determinado tema que represente efetivamente o problema de informação do usuário? (Spink e **Saracevic**, 1993, p. 63).

Essas e outras questões foram analisadas para a identificação de: padrões do discurso **interativo** entre os usuários e os intermediários; efeitos de diferentes variáveis no processo de recuperação e de obtenção dos resultados; guias para melhorar o desempenho das **interações** entre os mesmos; padrões de seleção dos termos para a estratégia de busca e da sequência da escolha dos mesmos e das fontes desses termos.

Esses experimentos analisados por **SPINK** e **SARACEVIC** vêm contribuindo para o entendimento da complexidade do processo de busca, porém, diversas outras implicações práticas podem ser mencionadas no contexto de recuperação em bases de dados. As primeiras estão relacionadas com os intermediários que **operacionalizam** as buscas, pois os mesmos devem ter o conhecimento continuamente aprofundado sobre as diferentes fontes de linguagens controladas e suas estruturas hierarquizadas ou não, para que os termos e/ou conceitos da estratégia de busca sejam os termos utilizados pelos indexadores no processo de entrada dos documentos na base. Outra **implicação**, nesse contexto, está diretamente relacionada com os usuários, que devem ser **informados** detalhadamente sobre todo o processo de busca. Dessa forma, eles poderão participar **ativamente** dessas **ações**, compreendendo sua complexidade, as **limitações** das bases e as **interações** que devem ser efetuadas para o alcance dos resultados desejados da busca. Uma terceira **implicação** está relacionada com o processo de entrevista, quando devem ser explicadas ao usuário todas essas negociações com as bases e os sistemas visando assegurar a compreensão do problema e a utilização de todos os termos e/ou conceitos possíveis de serem usados na estratégia de busca para a obtenção de resultados satisfatórios.

A avaliação dos usuários dos serviços de recuperação da informação tem sido mensurada por intermédio das medidas de **revocação** e precisão, como relatadas por SU, num artigo publicado em 1994. A autora conduziu um estudo sobre as medidas tradicionais de avaliação de desempenho na recuperação da informação, dentre outras: relevância, eficiência, **utilidade** e satisfação do usuário. Entre as medidas testadas e que **referem-se** à presente investigação, destacaram-se as de precisão que indicam a satisfação do usuário com os resultados da busca da informação. A preferência dos usuários pela qualidade, mais do que pela quantidade de referências sugere futuros estudos para aperfeiçoamento dos sistemas de recuperação de informação.

OJALA (1995) analisou os diversificados recursos de busca oferecidos pelos sistemas em linha que podem afetar a decisão do intermediário e/ou do usuário final na escolha do melhor sistema para a realização de uma busca de informação numa determinada base de dados. Mencionou **alguns** fatores, como as variações no uso de operadores **booleanos**, o uso de operadores de proximidade, a quantidade de frases da estratégia de busca que pode ser operacionalizada e outras.

WILDEMUTH e **MOORE (1995)** realizaram um estudo sobre as estratégias de busca utilizadas por usuários finais na base de dados **MEDLINE**. Esse estudo tinha o **objetivo** de identificar as técnicas de planejamento da estratégia de busca, **relacionando-a** com a **efetividade** dos resultados obtidos, e revelou, que apesar da satisfação do usuário final com os seus próprios resultados de busca, ocorreram inúmeras falhas no planejamento da estratégia. Os recursos da base de dados foram totalmente subutilizados e os resultados dos intermediários, comparados com o grupo de usuários finais, foram mais **efetivos**.

OJALA (1996) analisou a influência do ritmo da busca dos **intermediários** e as implicações sobre as suas escolhas, em relação a bancos de dados, a bases a serem utilizadas e a própria elaboração da estratégia de busca. Afirmou que o conhecimento da influência do ritmo da busca **permite** aos intermediários serem flexíveis nas suas próprias abordagens sobre as solicitações de busca, **adaptando-as** de acordo com a ocasião a ser executada.

STILL (1996) apresentou um estudo comparativo sobre o planejamento da estratégia de busca em **livros-texto** americanos e ingleses. Alertou para as formas diferenciadas de busca usando **palavras-chave** e/ou vocabulários controlados nos documentos, enfatizando as variações de palavras e as grafias dos termos.

NICHOLAS (1996), **apresentou** uma análise comparativa sobre a estratégia de busca realizada por profissionais de informação e por usuários finais. Comparou os **estilos** de busca de cada grupo e revelou que o usuário final pesquisa de forma muito limitada no uso dos comandos, porém, eles são mais econômicos e rápidos. Revelou, **ainda**, que o usuário final tem um comportamento e estilo de busca bastante **generalista**.

WARD (1996) apresentou um programa de treinamento sobre estratégia de busca a ser utilizado em várias bases de dados referenciais com o objetivo de tornar o usuário final mais independente para a realização de suas próprias buscas.

DIMARTINO e **ZOE** (1996) apresentaram um estudo sobre as técnicas de estratégia de busca utilizadas pelo usuário final treinado, para verificar se eles estão ou não capacitados para executar as buscas **efetivamente**. Concluíram que 55% dos usuários finais ficam insatisfeitos com suas próprias buscas.

NAHL e **TENOPIR** (1996) analisaram o comportamento **afetivo** e cognitivo de usuários finais **inexperientes**, durante o processo de recuperação da informação em bases de dados textuais. A preparação e execução da estratégia de busca de cada usuário final proporcionou uma análise **qualitativa** de cada etapa de busca, **caracterizando** as estratégias de acordo com os atos afetivos, cognitivos e **sensorio-motores** de cada participante dos grupos estudados.

SPINK (1997) investigou uma nova linha de pesquisa sobre a natureza do comportamento progressivo do usuário final durante o processo de recuperação de informação. O estudo caracterizou as mudanças sobre a **interação** com os sistemas de recuperação, incluindo: o contexto **situacional**, o problema de informação do usuário, os critérios de julgamento da relevância o estado afetivo e cognitivo do usuário, as fontes de informação usadas e as estratégias de busca formuladas.

SIMON (1997) investigou um programa **interativo** de estratégia de busca em bases de dados que facilitam a descoberta de **informação** especializada previamente desconhecida, porém, de interesse científico. A partir dos títulos de documentos recuperados, o programa estabelece entradas possíveis de busca na base de dados produzindo, uma série de sugestões heurísticas

que **auxiliam** na **seleção** de um segundo conjunto de artigos complementares ao primeiro resultado.

SCWARZWALDER (1997) propôs aos intermediários uma abordagem que os leva a agregar valores nas buscas realizadas para seus usuários. A partir do processo de entrevista, quando é possível detectar a real necessidade de informação do usuário, esse profissional obtém os resultados da busca e os **analisa**, sintetiza e avalia.

KENNEDY e **COLE** (1997) apresentaram um método para identificar todas as etapas de preparação da estratégia de busca em bases de dados em linha, em CD-ROM e nos OPACs. Esse método é totalmente dependente das necessidades de informação de cada **usuário**, **ensinando-o** a focalizar o seu problema para a elaboração da estratégia de busca.

YUAN (1997) analisou o efeito da experiência de busca do usuário final e o seu comportamento no uso de sistemas de recuperação. A análise de um grupo de usuários novos e de usuários mais experientes não revelou resultados significativos entre os dois grupos, pois, a quantidade de erros na formulação da busca ocorreu igualmente com os novos e com os experientes usuários finais.

PINTO e **LANCASTER** (1999), apresentaram vários critérios para julgamento da qualidade dos resumos nos documentos **técnico-científicos**. Afirmaram que os requisitos essenciais para que os resumos possam ser lidos e considerados úteis pelos pesquisadores, são os mesmos requisitos que se aplicam na busca por computador e citam dentre outros: a qualidade da leitura, a **exaustividade** na abordagem do conteúdo do documento; a coerência com o conteúdo; a brevidade no número de palavras e os custos intelectuais para geração dos mesmos. **Concluíram** que a crescente aplicação dos computadores para processamento de textos não reduz o valor dos resumos para as **atividades** de recuperação da informação. Essas conclusões vêm de encontro a proposta defendida por **FIDEL** (1986), que defendeu o aperfeiçoamento na elaboração dos resumos para ampliar e precisar os resultados de busca em linguagem natural.

Cabe destacar que esses estudos realizados em plena era do acesso à informação por meios **eletrônicos**, em que os usuários navegam na Internet com desenvoltura, ainda revelam as dificuldades inerentes ao processo de recuperação da informação em bases de dados.

Apesar dos intensivos programas de treinamento oferecidos pelos produtores das bases de dados, pelos próprios sistemas de recuperação em linha, de toda a documentação existente sobre as características de cada base de dados e suas respectivas estruturas de informação, dos sistemas amigáveis que oferecem **"menus"** para guiar o usuário em cada etapa do processo de busca, das linguagens de busca com recursos especiais para se aproximarem cada vez mais do usuário inexperiente, o processo de busca continua sendo um fator de dificuldade que ainda não foi minimizado pelas novas tecnologias disponíveis.

Além dos fatores acima mencionados, convém destacar outros que vão afetar a **seleção** do banco e da base de dados mais adequada para a **operacionalização** de uma estratégia de busca, tais como:

- a) Critérios para seleção de bancos de dados;
- b) Linguagem de busca, do banco de dados;
- c) Critérios para seleção das bases de dados.

Esses fatores serão analisados no presente estudo no item sobre implementação da estratégia de busca.

2.1.1 Processo de entrevista

Uma busca de informações depende **intrinsecamente** da **correta** compreensão, por parte do intermediário, das necessidades específicas de informação do solicitante. Assim, uma entrevista realizada antes do planejamento e execução da estratégia de busca é de fundamental importância, principalmente se o solicitante não puder estar presente no momento da operacionalização do seu pedido.

A literatura sobre os serviços de referência registra, frequentemente, as dificuldades do solicitante em explicitar sua necessidade de informação, prejudicando dessa forma a identificação do ponto focal de seu interesse. No processo de entrevista, é imprescindível que sejam negociados com o solicitante, todos os aspectos relacionados com o seu tema. Nesse caso, o intermediário, para ser bem sucedido na entrevista, deve possuir as seguintes características:

- a) habilidade para conduzir o processo de comunicação;
- b) habilidade para compreender os conceitos sobre o tema;
- c) habilidade para analisar o problema em seus aspectos mais abrangentes e/ou específicos;
- d) conhecimento da estrutura da **informação** do banco e/ou base de dados que vai ser utilizado;
- e) compreensão da política de indexação e do uso do **vocabulário** controlado da base a ser consultada;
- f) conhecimento das bases de dados diretamente relacionadas com o tema de busca e, também, das fontes alternativas de informação;
- g) suficiente compreensão do assunto a ser pesquisado, isto é, buscar uma **microcultura** sobre o tema em fontes especializadas.

Assim, a habilidade do **intermediário** em estabelecer empatia com o solicitante e a demonstração de sua sincera vontade de *encontrar a informação* desejada vai proporcionar uma ponte favorável para a obtenção de resultados satisfatórios no momento da execução da estratégia de **busca**.

Durante a entrevista, o **intermediário** deve discutir a pergunta em detalhes com o solicitante, identificando a área geral do assunto a ser pesquisado, tendo em vista a captação dos **sub-tópicos** relacionados ao tema. Essa negociação permitirá um enfoque mais flexível para a utilização dos termos e/ou conceitos na elaboração da estratégia de busca e vai capacitar o intermediário para as prováveis mudanças de **tática** de busca, quando obtiver os primeiros resultados da informação.

HURYCH (1982) enfatizou o papel especial que a entrevista de referência exerce no processo de recuperação de informação em bases de dados. Definiu o usuário e/ou cliente, como a pessoa que recebe serviços de uma agência, nesse caso, a biblioteca. A entrevista de referência tem sido observada em **termos** de apoio ao usuário, com uma atenção **diretamente** voltada para a negociação do tema de busca a ser pesquisado, sendo afetada pelos seguintes fatores, dentre outros: a comunicação **não-verbal**; a **interação** interpessoal e as técnicas de entrevista. Sugere a autora que os profissionais de **informação** devem aplicar métodos científicos para identificação das reais necessidades de informação dos usuários.

AUSTER e **LAWTON** (1984), num trabalho de pesquisa sistemática em recuperação de informação, investigaram as relações existentes entre as técnicas usadas pelos intermediários durante as entrevistas que antecedem o processo de busca propriamente dito e o total de novas informações obtidas pelos usuários como resultado das buscas. Investigaram, ainda, o grau de satisfação do usuário com a qualidade dos itens recuperados. Foram testadas, durante o processo de entrevista, questões abertas e fechadas e a média de satisfação dos usuários foi obtida com as questões abertas. Naturalmente que durante o processo de entrevista as questões abertas proporcionaram um melhor envolvimento do usuário, aproveitando-se essas perguntas para uma melhor captação de sua real necessidade de informação.

ABELS (1996) tratou do processo de entrevista realizado mediante correio **eletrônico** afirmando que, assim como os serviços de referência **eletrônicos** continuam a crescer oferecendo novos tipos de serviço, também, os pedidos de busca complexos podem ser negociados por essa nova via. Esse estudo foi empreendido para discutir as diferenças entre a entrevista realizada pelo correio **eletrônico** e as que utilizam outros "medias". Apresentou uma taxonomia de enfoques para a entrevista por correio eletrônico, propondo o enfoque sistematizado e sugerindo um modelo formal para o desenvolvimento da entrevista. Os resultados de investigação foram satisfatórios, confirmando o uso do modelo para solicitações de busca de **informação** mais complexas.

DERVIN e **DEWDNEY** (1996) descreveram uma nova abordagem para a condução da entrevista de referência preliminar, para a preparação da estratégia de busca, denominada de

"Questionamento Neutro". É um processo de negociação do pedido de busca entre o usuário e o intermediário, com a finalidade de se determinar a real necessidade de informação do solicitante e, nesse processo, o posicionamento neutro do intermediário permite a compreensão do pedido de busca sob o ponto de vista do usuário. O Questionamento Neutro é aberto na sua forma para evitar um diagnóstico prematuro do tema de pesquisa por parte do intermediário de modo a conduzi-lo para mudanças na maneira pela qual ele associa a informação solicitada com as fontes a serem pesquisadas. O uso desse tipo de entrevista preliminar, requer do intermediário um repensar sobre o paradigma da **informação** como "mercadoria", conduzindo-o ao encontro da real necessidade de informação do usuário.

SOMERVILLE (1997) sugeriu que antes do planejamento da estratégia de busca, e logo após a entrevista, o intermediário devia procurar maiores informações sobre o tema solicitado em fontes de informação especializadas. Conveniente e extremamente produtivo é que o intermediário solicite ao cliente, um artigo recente sobre o tema a ser pesquisado. Assim, de posse desse documento, o intermediário deve analisar alguns pontos básicos, tais como: o título e o resumo, procurando identificar termos e/ou conceitos auxiliares para esclarecimento do tema; procurar no corpo do próprio documento, outros indicadores para um entendimento mais completo sobre o tema; as citações desse artigo, com suas datas de publicação, títulos de periódicos, títulos de outros documentos, e autores que já estudaram o mesmo tema. Todas essas indicações favorecem a obtenção de uma microcultura sobre o assunto a ser pesquisado.

O processo de entrevista requer do intermediário uma atitude profissional cuidadosa, sendo desejável o uso de um formulário bem elaborado para registro do pedido de informação onde deverão ser anotados todos os tópicos mencionados pelo solicitante.

Para que o planejamento e execução da estratégia de busca encontrem resultados que satisfaçam o cliente, alguns parâmetros devem ser estabelecidos durante a entrevista, dentre esses:

- a) **Exaustividade** da busca - **deve-se** procurar exaustivamente os resultados em várias bases de dados, ou procurar apenas alguns documentos mais pertinentes ao tema proposto?
- b) Restrições quanto ao idioma - deve-se restringir os resultados da busca para algum idioma **específico**, ou imprimir todas as possíveis respostas em idiomas de difícil acesso para o cliente?
- c) Restrições quanto ao tipo de documento - deve-se restringir os resultados da busca apenas para os documentos de mais fácil acesso em qualquer biblioteca, ou limitar a resposta da busca para um determinado tipo de documento?
- d) Abrangência retrospectiva - deve-se recuperar os documentos no mesmo período de abrangência da base de dados a ser consultada, ou limitar os resultados para os últimos cinco anos, por exemplo?

- e) Restrições quanto a estudos experimentais - **deve-se** limitar os resultados de busca eliminando-se os estudos de laboratório, ou restringir apenas para os estudos referentes a experimentos com seres humanos, homens ou mulheres, idosos ou crianças?
- O Restrições quanto ao custo operacional - deve-se limitar o resultado da busca de acordo com o **provável** custo **final**, ou *restringir apenas a* quantidade de respostas até um determinado número de documentos recuperados?

JANKE (1985) analisou um processo diferente de entrevista, denominado aconselhamento para a busca onde o bibliotecário experiente em buscas em linha orienta o usuário final que será o executor de sua própria estratégia de busca nos bancos e bases de dados. Essa modalidade de utilização dos sistemas em linha pelo usuário final **cresceu em importância** na década de 80, logo após o lançamento de sistema BRS/After Dark e do DIALOG Knowledge Index, na Online National **Conference** em Atlanta, **Georgia**, em 1982.

O desenvolvimento de interfaces amigáveis para a utilização dos grandes bancos de dados trouxe uma nova perspectiva para o uso desses sistemas pelo próprio usuário final. Assim, o intermediário experiente em planejamento e execução de estratégias de busca, pode assumir um novo papel no cenário dos grandes bancos de dados: o de orientador do usuário final. O autor define o aconselhamento preliminar da busca como:

um processo onde o **bibliotecário**, *agindo com a* função de mero conselheiro, assiste o usuário final no desenvolvimento de sua própria estratégia de busca, antes da execução da mesma. (Janke, **1985**, p.15.)

No processo de entrevista tradicional, o bibliotecário e/ou intermediário captam os elementos principais do tema a ser pesquisado nas bases de dados e planejam a estratégia de busca, baseados nesses elementos. Entretanto, no processo de aconselhamento preliminar, a entrevista difere **consideravelmente**, como pode ser observado a seguir, numa adaptação das sugestões obtidas em JANKE (1985).

Apresenta-se na Tabela 1, as principais características da entrevista tradicional e as etapas de aconselhamento realizadas pelo bibliotecário quando o mesmo se coloca no papel de mero orientador de um usuário final que será o responsável pela execução de sua própria estratégia de busca. A experiência **profissional** do autor desta pesquisa, adaptou e acrescentou novos dados nessa fase preliminar do planejamento da estratégia, visando completar o processo de entrevista sugerido por JANKE (**1985**).

Tabela 1 - Entrevista *versus* aconselhamento

Entrevista tradicional	Aconselhamento preliminar
1. O bibliotecário entrevista o usuário que lhe fornece os elementos fundamentais para a preparação da estratégia de busca.	O bibliotecário aconselha o usuário final que será o executor da própria estratégia de busca. Todas as etapas de seleção da base, uso da terminologia adequada, lógica booleana e comandos de impressão são fornecidos e o usuário assume a responsabilidade pela sessão de busca.
2. O bibliotecário utiliza instrumentos de auxílio, como thesaurus , vocabulários controlados e outros, para planejar a estratégia de busca junto com o usuário.	O bibliotecário e o usuário final usam o <i>thesaurus</i> e outros instrumentos de auxílio para elaborar a estratégia de busca. O usuário final é envolvido fortemente em cada etapa do planejamento da estratégia de busca e toma-se responsável tanto pela seleção dos termos e/ou conceitos que vão compor a estratégia como pela sua utilização lógica.
3. O bibliotecário assume o controle total do uso dos recursos específicos de cada base de dados a ser consultada: dos operadores de proximidade, da truncagem de termos, dos descritores, da lógica booleana e outros.	O bibliotecário explica o funcionamento essencial da lógica booleana e, em alguns casos, o uso dos operadores de proximidade e dos campos de busca específicos da base a ser consultada. Quem executa a busca é o usuário.
4. A elaboração da estratégia de busca e a Lógica booleana são formulações complexas com experientes intermediários.	A elaboração da estratégia de busca e a lógica booleana são formulações simples, conforme a percepção do usuário.
5. No processo de execução da busca, a instituição assume todos os problemas que possam ocorrer, incluindo custos imprevistos, pois é o bibliotecário quem está executando a busca.	O usuário final é o responsável pelos custos e pelos resultados obtidos, satisfatórios ou não, pois ele foi o executor da busca.
6. O usuário toma-se um observador passivo do processo de busca, que vai ser executado pelo entrevistador .	A função do usuário é ativa e participativa, pois ele vai executar a busca após o aconselhamento recebido.
7. Ao final da entrevista, o bibliotecário revê todos os dados coletados com o usuário sobre o tema, para verificar se houve uma correta compreensão do mesmo, antes de executar a estratégia de busca. A pressão psicológica fica sobre o bibliotecário.	Ao final da sessão de aconselhamento, o bibliotecário e o usuário final fazem uma revisão de todo o planejamento da busca, antes que a mesma seja executada pelo usuário. A pressão psicológica fica sobre o usuário final.
8. O objetivo da entrevista tradicional é criar um elo de confiança entre o intermediário e o usuário, cabendo ao primeiro a responsabilidade de prover informações sob demanda.	O objetivo do aconselhamento é encorajar os usuários na execução de suas próprias buscas de informação, tornando-os auto-suficientes no uso de bancos e bases de dados.

Fonte: Adaptado de Janke (1985)

A literatura especializada registrou desde os meados da década de 80, um forte movimento no sentido de prover o usuário final com os instrumentos necessários para que o mesmo realizasse suas próprias buscas. Interfaces amigáveis foram desenvolvidas para permitir o uso maciço dos bancos e bases de dados, bem como os programas baseados em "menus" onde o usuário menos experiente era orientado passo a passo. Entretanto, os custos de conexão com as bases e com os bancos de dados e, no caso do Brasil, também os custos por caracteres transmitidos e **recebidos**, inviabilizaram a participação do usuário final no processo de execução da estratégia de busca.

Constantemente estudado e relatado nas revistas especializadas (*Reference Librarian, Reference Quarterly (RQ), Online, Database, JASIS, Library Trends, etc.*), a evolução do processo de entrevista está fortemente entrelaçada com as novas tecnologias de armazenamento e recuperação de informação desenvolvidas ao longo das últimas décadas. Sua influência no processo de recuperação é bastante representativa, necessitando uma especial análise durante a negociação. É por isso que as instituições nacionais utilizadoras de bancos e bases de dados no **exterior** passaram a investir fortemente no treinamento de profissionais de informação com requisitos indispensáveis para busca de informação em linha: conhecimento de idiomas estrangeiros, conhecimento avançado em fontes de informação impressas e/ou **eletrônicas**, conhecimentos específicos na área de **atuação** das empresas e outros.

2.1.2 Técnicas de estratégia de busca

No planejamento da estratégia de busca, algumas **ações** ou **táticas** precisam ser **critériosamente** relacionadas principalmente com a **seleção** de termos e/ou conceitos, com as restrições e/ou limites determinados pelo solicitante e com outras questões.

BATES (1974), conceituou **tática**, como um plano idealizado para a **efetivação** de uma busca, sendo que esse plano pode ser bem sucedido ou não. A tática de busca requer do intermediário um constante julgamento na sua utilização, pois o mesmo tem que tomar a decisão sobre o melhor momento de implementá-la durante o processo de busca. Caso não seja usada no momento **correto**, existe o risco de que a tática de busca se tome **inapropriada**, e isto só pode ser detectado durante a execução do processo de busca.

Nesse artigo, a autora definiu e analisou cerca de treze táticas de busca. Na tática denominada "**registro**", o solicitante deseja uma busca exaustiva sobre um tema, porém ele não deseja encontrar resposta para a mesma, isto é, a resposta desejada é nenhuma. Apesar da utilização de todas as possibilidades de uso de sinonímia, termos **quasi-sinônimos** e outros recursos permitidos pela base de dados, o solicitante não pretende encontrar qualquer informação

sobre o seu **próprio** tema de pesquisa. Afinal, trata-se de uma inovação e a resposta desejada é que realmente não existam outros autores pesquisando o seu tema.

Uma outra **tática** denominada "levantamento" pretende revisar, antes da execução da estratégia de busca propriamente dita, todas as opções disponíveis em outras bases de dados menos usuais para o tema solicitado. Isso porque existe uma tendência natural, em qualquer intermediário, de planejar uma estratégia de busca para a base de dados que melhor venha a responder à pergunta do solicitante. Nesta tática, as perguntas são processadas em bases de dados de menor qualificação para a resposta a ser obtida, porque, nesse caso, estão em jogo outros fatores, tais como o custo final da busca. Portanto, um "**levantamento**" preliminar em outras bases de dados pode quase sempre contribuir para a identificação de itens relevantes **eliminando**-se custos relacionados com o tempo de busca na base mais adequada. A tática de "**levantamento**" é um recurso que deve ser usado para a obtenção de novos **termos** de busca. Assim, os itens recuperados vão apresentando novos termos e/ou conceitos que vão conduzindo a uma nova **seleção** de termos que serão utilizados durante a formulação da estratégia de busca.

A tática denominada "**vizinho**" implica na visualização de termos alfabeticamente próximos nos dicionários em linha oferecidos pelos bancos de dados. Neste caso, singular, plural, **adjetivo**, advérbio e termos compostos, utilizados como descritores e/ou identificadores do conteúdo temático de cada documento registrado na base, podem vir a ser usados de forma complementar durante a execução de estratégia de busca.

Convém destacar que inúmeras bases de dados bibliográficas oferecem para busca temática os campos de título, descritor e resumo. Assim sendo, um **termo** técnico específico sobre, por exemplo, uma determinada espécie de ave, pode não estar representado nesses campos. Isso não quer dizer, entretanto, que a espécie de ave não esteja sendo estudada num determinado documento. Nesse caso, o uso de um **thesaurus** apropriado, isto é, o uso do vocabulário controlado daquela base de dados a ser pesquisada, vai permitir a identificação de termos mais genéricos, os relacionados e/ou específicos que deverão compor a expressão de busca.

A análise de um determinado tema de pesquisa proporciona pelo menos três possibilidades de busca:

- a) o tema pode estar sendo discutido claramente em documentos totalmente devotados ao mesmo;
- b) o tema pode aparecer como **sub-tema** em documentos relacionados com um assunto mais abrangente;
- c) o tema pode ser objeto de estudo em documentos dedicados a assuntos mais específicos.

Essas três possibilidades de registro da informação podem vir a interessar ao solicitante, porém documentos referentes à segunda possibilidade precisam estar devidamente explicitados, caso contrário não serão recuperados.

Num planejamento da estratégia de busca essas possibilidades devem ser exploradas consultando-se preferencialmente o *thesaurus* da própria base de dados a ser utilizada. Convém, ainda, a título de assegurar que a resposta seja a mais completa possível, recorrer a *thesaurus* de outras bases de dados de assuntos **correlatos** para a **identificação** de novos termos e/ou conceitos.

A autora relacionou também as **táticas** denominadas **"ressoletrar"** e **"reespaçar"**, significando que as grafias **alternativas** para cada termo e/ou **conceito** a ser usado na expressão de busca devem ser utilizadas sempre que necessário. Um exemplo de variações de grafia são as palavras que no inglês americano se escrevem com **"s"** e no inglês do Reino Unido com a letra **"z"**, como **polimerização** e outras. O intermediário ao formular a estratégia de busca deve verificar cuidadosamente essas variações, não esquecendo, também, das siglas, abreviações, nomes comerciais de substâncias químicas, ou de drogas, etc. Além disso, as palavras hifenadas **comumente** utilizadas pelos autores em seus trabalhos necessitam também de cuidadosa **verificação** para assegurar uma completa recuperação das informações solicitadas.

MARKEY e **ATHERTON** (1978) apresentaram um manual com o **objetivo** de treinar usuários e intermediários nas estratégias de busca em linha. A base de dados **ERIC** foi utilizada na investigação e o estudo financiado pelo próprio produtor da base. Foi desenvolvido um arquivo especialmente elaborado para permitir que os intermediários da busca pudessem comparar seus resultados com os resultados de uma "busca perfeita". Um grupo desenvolveu **conjuntos-resposta** para mais de 20 pedidos de busca exaustivamente pesquisados no arquivo da base ERIC. Os resultados **conjuntos** de suas estratégias de busca para cada pedido tornaram-se o "perfeito **conjunto-resposta**". Para cada tópico solicitado foram fornecidas várias formulações representativas que respondiam aos diferentes **objetivos** de cada um dos pedidos, comparando os resultados do **intermediário** com um programa automático para cálculo da **revocação** e da precisão e com um conjunto-resposta. Após esses experimentos, o manual de treinamento da base ERIC, via banco de dados **DIALOG**, foi elaborado com o objetivo de fornecer um roteiro seguro para o aperfeiçoamento do experiente intermediário. Esse manual possui várias sessões, sendo constituído por um modelo completo do processo de busca com as técnicas de formulação da estratégia, também citadas em **HAWKINS (1982)** e exercícios variados para cada etapa de busca, conduzindo o intermediário de forma segura no uso e no conhecimento da base.

DOLAN (1980), **SCHIFFMANN** (1985) e **CURNUT** (1991) descreveram uma técnica de busca denominada **"hedges"**, desenvolvida pela National **Library** of Medicine, que consiste no agrupamento de um conjunto de termos de busca, específicos de um assunto para futura utilização e pode ser aplicada em qualquer área de assunto. Esses **"hedges"** são escolhidos para definir

grupos de termos **relacionados** que ocorrem em vários **níveis** de hierarquias das linguagens controladas e, normalmente, não estão reunidos nos diversos **thesauris** sob um mesmo conceito.

DOLAN (1980) exemplificou com o conceito relacionado a "**peçoal**", Ele formou um conjunto de termos de busca contendo mais de 50 (cinquenta) possíveis termos de busca, que são usados por diversas linguagens controladas das bases de dados. O armazenamento desse conjunto de termos, que representam o conceito "peçoal", pode ser feito única vez em uma estratégia de busca, e reutilizado para diferentes temas de busca relacionados com esse conceito.

Uma significativa economia de custos pode ser obtida com essa técnica nas bases de dados em linha, pois esse agrupamento de termos para representar o conceito, fica armazenado indefinidamente sob a senha do usuário. Sua utilidade nas bases em CD-ROM também é visível e **possível**, ampliando **consideravelmente** as possibilidades de recuperação da informação.

Num **artigo**, de revisão no **ARIST**, **BATES** (1981) salientou, a inexistência de estudos sobre técnicas de busca. Definindo os vários aspectos relacionados com o tema, tais como: mecânica da busca, formulação de busca, estratégia de busca e perfil de busca, a autora destacou alguns trabalhos realizados por **FENICHEL** (1981) referentes ao comportamento do intermediário que executa a busca. Por esse motivo, esse artigo foi considerado por **BATES** como uma revisão **sob a ótica psicológica** das **técnicas** de busca, já que sua tentativa de **analisar** o tema, passando por tópicos como: busca física, psicologia de busca, estratégias em linha, uso do vocabulário controlado e termos livres, busca bibliográfica e em índices de citações, heurística de busca e questões de busca e texto, não elucidaram como o intermediário idealiza e executa a técnica de busca.

HAWKINS e **WAGERS** (1982) apresentaram algumas técnicas para formulação da estratégia de busca e foram assim intituladas: construção em blocos, **fracionamento** sucessivo, crescimento da pérola de citação e pesquisa **iterativa**. Essas técnicas representam apenas a parte mecânica da estratégia, e os autores não definem se os termos utilizados podem ser os da linguagem natural ou os da controlada. Apesar disso, são extremamente úteis para ajudar no raciocínio lógico usado com os operadores **booleanos**, mas não substituem o processo associativo feito **anteriormente**, no momento da identificação dos conceitos e suas relações. São técnicas que tentam transpor o raciocínio humano para o raciocínio da máquina, nem sempre com eficiência e eficácia. Os autores utilizam a expressão "crescimento da pérola de **citação**" para descrever uma técnica de busca similar à do "**levantamento**" designada por **BATES** (1974). Em **HAWKINS**, parte-se de um único termo e/ou conceito, identificando-se pelo menos uma citação relacionada com o tema de busca. Nessa citação, outros termos e/ou conceitos são extraídos para ampliar o **espectro** de **informação** sobre o tema solicitado. Dessa forma, a primeira citação é considerada como uma pérola que vai conduzir a outras citações permitindo o crescimento do resultado de busca.

FIDEL (1984), num estudo de caso relacionado com o comportamento do intermediário na **operacionalização** das estratégias de busca em linha, descreveu dois modelos de estilo de busca:

o **operacionalista** e o **conceitualista**. Baseado na sistemática observação de **intermediários** experientes no momento da operação das estratégias de busca, esse estudo pode ser considerado como um ponto de partida para investigar tópicos gerais na recuperação da informação. O modelo de investigação sobre os estilos de busca foram desenvolvidos para um estudo empreendido sobre o conhecimento da interface **homem-máquina**, examinando-se as habilidades empregadas por esses intermediários num processo de **interação** heurística, para melhoramento dos resultados da busca.

FIDEL (1985, p. 62), definiu o modelo operacionalista como "um movimento que usa as características do sistema de recuperação com o objetivo de modificar o conjunto recuperado sem mudar o significado conceitual que ele representa". Ao passo que o modelo conceitualista é definido como aquele que modifica o conjunto recuperado pela mudança de significado do conceito que ele representa. Apresentou, ainda, as diferenças típicas entre os dois estilos de busca, relacionando as limitações e implicações do modelo.

Num estudo sobre a estratégia de busca em bases de dados em linha, **FIDEL** (1985) detectou todas as etapas de **operacionalização** das estratégias, **denominando-a** técnica de **"mover"**. O objetivo desse estudo foi apresentar e descrever várias modificações na formulação original da estratégia de busca, quando o intermediário estivesse executando a mesma. Afirmou que a mais importante vantagem da busca em linha é **"que os intermediários** podem revisar suas estratégias **pré-definidas** anteriormente ao longo do processo de busca em qualquer etapa da mesma". A técnica **"mover"** está **diretamente** relacionada com os dois estilos de busca: operacionalista ou conceitualista, descritos também, num artigo publicado em 1984. São métodos caracterizados pela **forma** usada pelos intermediários para analisar os pedidos de busca, o tipo de termos que eles **selecionam** e suas atitudes em **direção** a novos desenvolvimentos nos sistemas de recuperação utilizados. Caracterizou os intermediários **operacionalistas** como aqueles que habitualmente selecionam movimentos, nos quais confiam, entre os termos de busca livres e os termos da linguagem controlada, com o objetivo de modificar os conjuntos recuperados, sem mudar, **entretanto**, o conceito específico que cada conjunto representa. Já os intermediários **conceitualistas** modificam os conjuntos recuperados modificando o significado dos conceitos representados nos conjuntos e são frequentemente apoiados pelas análises de assunto relacionadas com os pedidos e pela estrutura das linguagens de indexação. A autora observou ainda, que esses "movimentos" são realizados pelos intermediários para melhorar as formulações das estratégias em três situações problemáticas:

- a) quando o conjunto recuperado é muito amplo;
- b) quando o conjunto recuperado é muito pequeno; e,
- c) quando consideram o conjunto recuperado fora do objetivo do usuário.

Todos esses "movimentos" foram examinados e relacionados com as **táticas** de busca definidas por BATES (1974) e com as técnicas de estratégias de busca preconizadas por MARKEY e ATHERTON (1978) e por HAWKINS e WAGERS (1982).

Num trabalho anterior, FIDEL (1984) havia descrito os estilos como **operacionalistas** e **conceitualistas** ressaltando a ideia de que os intermediários empregam estilos pessoais na **operacionalização** das estratégias e, se as características desses estilos podem ser identificadas e descritas, daí surge um guia e/ou modelo para a pesquisa em bases de dados de um modo geral.

KIRKBRIDE (1991) mencionou alguns dos critérios para **seleção** de bancos e bases de dados, alertando para as dificuldades inerentes ao processo de escolha dos termos para a estratégia de busca. Afirmou que a formulação **de** estratégia de busca depende do tema de pesquisa e da linguagem de busca dos diversos bancos de dados em disponibilidade. **Portanto**, é necessário um conhecimento minucioso das técnicas de estratégia de busca, dos **prós** e **contras** da estratégia no texto completo dos documentos, da busca em linguagem controlada, da política de indexação, da cobertura de assunto das bases de dados, e das características das linguagens de recuperação dos diversos sistemas disponíveis. Em algumas ocasiões um tema de pesquisa ou um tipo de formato solicitado exigem uma **pré-seleção** da base e do banco de dados, antes mesmo **de** se planejar qual técnica será usada na formulação da estratégia de busca.

A autora sugeriu determinadas ações para a busca em bases de dados textuais, para as bases que utilizam vocabulários controlados, para a busca em linguagem natural e apresentou como regras de sobrevivência para o intermediário os seguintes requerimentos: flexibilidade, criatividade e habilidade linguística. Por esse motivo, quando o intermediário planeja **taticamente** os passos que serão executados para a operacionalização de uma busca, ele está planejando e executando movimentos e operações para alcançar determinados **objetivos**. Assim, se for necessário produzir um relatório de mercado sobre os competidores de produtos **eletrônicos** para atender a um determinado usuário, o **intermediário** deverá **selecionar** primeiramente um banco e uma base de dados que proporcione a "carga ou **download**" desses dados numa planilha **eletrônica** ou que a própria base, a partir de elementos de dados fornecidos, gere o relatório solicitado.

HARTER e CHENG (1996) apresentaram um novo conceito e técnica para recuperação da informação denominado descritores **"colinked"**. Esses descritores identificam termos de busca que, hipoteticamente podem ser superiores àqueles inicialmente selecionados pelo intermediário. A teoria sobre esses descritores sugere a movimentação automática de um ou mais termos de busca para outros termos que devem ser superiores no desempenho da recuperação, sobre os termos usados **originalmente**. Esse estudo encontra-se ainda, em fase **experimental** e seus **resultados** preliminares têm sido promissores.

Convém destacar que outros procedimentos técnicos permitem refinar e completar a equação da pesquisa estando diretamente relacionados com os recursos da linguagem de busca da base e/ou do banco de dados e ainda com a especificidade do tema da pesquisa a saber:

a) Ponderação ou atribuição de pesos

Consiste em atribuir um coeficiente de ponderação que **distingue** determinado descritor dos demais descritores de um documento. Essa atribuição de peso é **calculada** de acordo com a importância do descritor no documento, sendo efetuada no momento da indexação do mesmo.

b) Vinculação

Consiste em definir a proximidade dos descritores com **sub-descritores** ou **qualificadores**, definidos pelos produtores da base de dados. Neste caso, a especificação desejada pode ser estabelecida na formulação da estratégia de busca assegurando, portanto, a recuperação precisa da informação solicitada.

c) Restrições e/ou Limitações

Consiste em definir, durante o processo de entrevista junto ao usuário, o **período** de tempo desejado na pesquisa, os idiomas de interesse, os experimentos em nível de laboratório, o sexo **e/ou faixa-etária** de interesse do pesquisador e outras restrições que tenham sido oferecidas pelo produtor da base ou pela linguagem de busca do sistema de recuperação da informação.

O tempo de conexão é caro e o tempo do intermediário também custa, então é importante estar conectado pelo menor tempo possível em cada base de dados pois, assim, mais bases de dados podem ser **acessadas** pelo mesmo valor. Existem inúmeras maneiras de economizar tempo de computador. **Assim**, o uso da **truncagem** permite a utilização da raiz da palavra para recuperar todas as possibilidades de expansão do termo e pode ser empregada para reduzir o número absoluto de palavras numa estratégia de busca. A truncagem em estratégias de busca requer menos palavras e, portanto, menos tempo de **input** no terminal.

Pode-se economizar tempo e dinheiro com o uso de uma estratégia universal, quando se pesquisa em várias bases de dados em linha. A estratégia universal, quando empregada com o comando **"save"** e "recair, é muito eficiente. Os comandos específicos para fazer isto dependem do sistema a ser **acessado**, mas esses comandos são fáceis de aprender e podem economizar tempo de conexão pois, as palavras são digitadas apenas uma vez. Nas bases em CD-ROM, do mesmo fornecedor, esses recursos também são oferecidos.

Ao optar pelo uso desta técnica, **deve-se** estar certo de que a estratégia de busca pode ser **transferida** para outras bases de dados em que os campos que identificam cada item de um registro de informação podem ser reutilizados. Essa decisão implica um estudo prévio dos campos de busca de cada base a ser consultada.

A experiência do **intermediário**, do analista de informação **e/ou** do usuário final com as obras básicas de referência combinada com um bom conhecimento das específicas bases de dados e **softwares**, permite ao intermediário a realização de buscas de qualidade nos sistemas em linha. Entretanto, o intermediário deve estar adequadamente preparado para trabalhar em conjunto com o solicitante da busca e, também, dedicar a maior parte do próprio tempo e esforço, na estruturação da estratégia antes da conexão com o banco/base de dados.

A preparação **tática** da estratégia de busca, bem como a **seleção** do **banco/base** de dados a ser consultada para responder a uma pergunta específica, exigem do intermediário toda uma gama de conhecimentos especializados: idiomas, conhecimento da estrutura da informação na base de dados, conhecimento da linguagem de busca do banco de dados, instrumentos de auxílio para **identificação** da terminologia a ser usada, lógica booleana, e demais recursos **disponibilizados** para a obtenção dos resultados esperados pelo solicitante.

A implementação da estratégia de busca, portanto, requer do **intermediário pré-requisitos** intelectuais e pessoais para **operacionalização** da mesma, por exemplo, raciocínio lógico, autoconfiança, serenidade, além dos já citados anteriormente.

2.1.3 Estratégia de busca - etapas

A identificação apropriada dos elementos descritivos de um item e/ou registro de informação contido numa base de dados é fundamental na elaboração da estratégia de busca. Assim, a primeira etapa será verificar a documentação da base a ser consultada, a **fim** de identificar a codificação definida pelo banco de dados para cada campo do item de **informação**.

Uma preparação adequada é indispensável tanto para a busca computadorizada quanto para a busca manual. Tipicamente, o pesquisador não é a pessoa que opera o terminal de computador, por conseguinte, o técnico, intermediário ou dentista da informação deve assegurar uma definição adequada do tópico antes de começar a elaboração de uma revisão especializada da literatura.

Ajudar o pesquisador a definir **clara** e precisamente o problema é muito importante para a formulação de uma **efetiva** estratégia de busca. Apenas em raras ocasiões pode ser desejável conhecer algo acerca de algum problema. A maneira como se assessora o **usuário** na definição ou limitação do problema vai afetar as atividades efetuadas no momento de operacionalização da busca **online**, ou nos índices manuais impressos.

É sabido que a maior parte dos usuários ao dirigir-se a um serviço de informação acredita possuir uma boa compreensão dos próprios problemas; sendo assim, a tarefa imediata do **intermediário** junto aos sistemas é ajudar o usuário a definir e especificar o problema, com termos e conceitos que são apropriados para aquela fonte de informação específica que será utilizada para a busca. Ou seja, os termos usados num campo particular devem ajustar-se àqueles usados nas bases de dados mais relevantes que serão consultadas.

Para a preparação da estratégia de busca, o usuário deve fornecer preferencialmente, os seguintes dados:

- a) escrever um título sucinto;
- b) escrever uma pequena definição do problema;
- c) listar os termos mais apropriados para o problema; e,
- d) listar os termos que não são desejados.

A partir desses dados, o intermediário poderá executar a busca de informação de uma forma relativamente simples. E se vários dos maiores autores que publicam sobre o assunto solicitado puderem ser relacionados, então a busca será muito mais eficiente. Infelizmente, porém, essas condições são raríssimas.

Como definir o problema - etapas

A seguir, são apresentadas várias orientações organizadas em etapas, que podem ajudar o pesquisador a definir o problema, contribuindo, portanto, para acentuar a qualidade da busca.

1ª Etapa: Discussão do tópico geral da pesquisa

Uma entrevista, mesmo uma chamada telefônica, pode eliminar muitos problemas. Se o **tópico** da busca não tiver sido claramente intitulado, definido e descrito pelo usuário, então é tarefa do intermediário ajudar a **fazê-lo**. É útil perguntar como os resultados da busca irão ser aplicados porque a resposta pode mudar a **direção** ou a ênfase da busca.

2ª Etapa: Conhecimentos básicos sobre os instrumentos de busca

Se o pesquisador se familiarizar com alguns dos instrumentos básicos de auxílio para a busca em seu campo, por exemplo, um *thesaurus*, isto pode ajudar a definir o tópico e gerar uma lista das **palavras-chave** a serem usadas na estratégia de busca. O pesquisador deve **selecionar** os termos que especifiquem o problema por causa do seu grande conhecimento do assunto; o intermediário deve ajudar, mas não deve definir o assunto porque, na maioria das vezes, a definição obtida para o tema difere completamente do pesquisador.

3ª Etapa: Formulação "provisória" da estratégia de busca

Após o pesquisador ter fornecido um tópico claro e descritores **válidos**, o intermediário pode formular uma estratégia de busca inicial (tentativa).

A busca está bem definida se o intermediário for capaz de assegurar a recuperação de todas as citações para vários termos. Mas, tipicamente a informação mais complexa, aquela que requer a **interseção** de, pelo menos, dois conjuntos de termos é desejável. Neste caso, o intermediário tem que ajudar o pesquisador a agrupar termos similares num conjunto único e, sucessivamente, pode agrupar outros termos similares em outros conjuntos até alcançar a resposta solicitada. Porém, o que é realmente desejável é que a informação solicitada ocorra em ambos os conjuntos ao mesmo tempo.

4ª Etapa: Compreensão da lógica dos conjuntos de termos

O intermediário deve ajudar o usuário a compreender as propriedades básicas da teoria dos conjuntos tal como é usada nas estratégias de busca por computadores. O uso da interseção de mais de dois conjuntos de termos deve ser evitado porque, embora os resultados possam ser bem precisos, eles são muito limitados e podem provocar uma possível exclusão de informações relevantes.

5ª Etapa: Interdisciplinaridade

Os intermediários que executam a busca nos sistemas precisam conhecer sobre os campos de pesquisa **correlatos** e como usá-los para responder a questões que demandam buscas **interdisciplinares**.

Frequentemente o **pesquisador** não está **familiarizado** com os campos de pesquisa relacionados; cabe, portanto, ao intermediário propor a expansão da busca para estes outros campos aumentando, **conseqüentemente**, as **possibilidades** de documentos de interesse virem a ser recuperados.

6ª Etapa: Eliminação de termos indesejados

Existem várias maneiras para se eliminar termos indesejados de uma busca. Se o pesquisador não está interessado em **citacões** sobre um determinado conceito, o intermediário pode eliminar do resultado final todas as que tiverem o termo citado. Mas existem problemas com este enfoque. Ocasionalmente, ambos os **termos** ocorrerão na mesma citação. Nesse caso, o solicitante deve ser avisado sobre esses possíveis problemas.

Com o solicitante, o intermediário examinará a necessidade de exclusão de termos afins com o tema pesquisado. Os termos indesejados serão excluídos do resultado da busca depois de se ver o impacto dessa exclusão no resultado total da busca. A decisão para excluir termos nem sempre é fácil e, visualmente, depende da especificação do tópico.

7ª Etapa: Especificação dos parâmetros relevantes para a execução da busca

Todos os parâmetros relevantes, devem ser considerados para se determinar os limites da busca. Quanto, em termos de recursos financeiros, pode ser gasto na busca? Deve a busca ser limitada nos anos mais recentes? Quais as bases de dados que provavelmente irão fornecer as mais relevantes **citacões**? O pesquisador quer todas as citações que mencionam uma autoridade particular ou somente as que são autorizadas por uma pessoa particular? O pesquisador deve responder a estas **questões**, devidamente assistido pelo intermediário nessa tomada de decisões, que anotará todos os detalhes para o posterior planejamento da estratégia de busca.

Essas etapas estão baseadas na experiência profissional adquirida e são um roteiro a ser seguido por usuários e/ou intermediários iniciantes nas buscas em bases de dados em CD-ROM ou em linha. Num programa de treinamento no uso de bases de dados torna-se aconselhável seguir uma orientação que inclua todas as fases de negociação com o usuário da informação, a fim de se obter resultados satisfatórios. Essas etapas podem ser categorizadas como **pré-requisitos** indispensáveis para o planejamento das estratégias de busca.

2.1.4 Implementação da estratégia de busca

Uma das primeiras dificuldades com que o intermediário e/ou usuário final de buscas em bancos de dados se depara é a escolha entre vários bancos de dados, onde a base de dados de interesse está hospedada. Assim, os critérios apresentados a **seguir**, que envolvem a linguagem de busca do banco de dados e a **seleção** do **mesmo**, servem como um roteiro seguro para responder à questão: onde buscar a informação?

Critérios para a seleção de bancos de dados:

- 1) tradição na oferta de serviços de informação **eletrônicos**;
- 2) multiplicidade na oferta de bases de dados em todas as áreas do conhecimento;
- 3) linguagem de busca flexível e precisa, permitindo respostas em tempo hábil e pertinentes à pergunta formulada;
- 4) linguagem de busca documentada em manuais para uso do banco e das bases e facilidade de obtenção dos manuais, pelo próprio banco de dados;
- 5) **coleta** e disseminação dos instrumentos de auxílio para o planejamento das estratégias de busca em cada base de dados;
- 6) facilidades para contabilização dos custos de cada busca, permitindo pagamentos correntes e/ou cumulativos;
- 7) recursos próprios e de terceiros para acesso ao documento,
- 8) divulgação do banco e das bases de dados, através de boletins **regulares**, distribuídos gratuitamente aos seus próprios usuários, comunicando as mudanças ocorridas, tais como: aumento de **preço**, novos recursos para estratégia de busca etc;
- 9) **oferta** de cursos de treinamento no banco de dados e em suas principais bases de dados;
- 10) disponibilidade de acesso, por variadas redes de comunicação de dados.

Vários problemas podem **ocorrer** quando da execução da estratégia, caso a linguagem de busca do próprio banco de dados não ofereça, pelo menos, algumas das condições relacionadas com:

aj tempo de resposta razoável, isto é, o processamento da pergunta e a visualização da resposta não deve apresentar um intervalo do tipo **"break coffee"**;

- b) a linguagem de busca deve permitir a interrupção do processamento de uma pergunta mal formulada;
- c) estratégias de busca usando operadores lógicos e de proximidade;
- d) **truncagem** de singular, plural, raiz de **conceitos/palavras** etc;
- e) uso de operadores diferentes na mesma pergunta;
- f) armazenamento de estratégia de busca em uma base de dados, para posterior **aproveitamento**, no mesmo dia ou dias depois;
- g) eliminação de perguntas cujos resultados tenham sido insatisfatórios;
- h) deve **permitir** a busca por diferentes idiomas e períodos de tempo;
- i) planejamento da busca por campos **pré-definidos** ou **pós-definidos**;
- j) visualização parcial e/ou selecionada de itens de interesse, podendo o usuário definir partes específicas para apresentação da resposta;
- k) impressão parcial, selecionada ou total dos resultados obtidos no seu próprio equipamento (no disco rígido, disquete ou impressora), ou em terminal remoto.

Critérios para seleção de bases de dados

São listados a seguir alguns dos critérios para seleção de bases de dados que vão interferir na decisão de onde procurar a informação. Vários tópicos devem ser verificados, a fim de se evitar custos excessivos durante o processo de execução da estratégia de busca.

COBERTURA DE ASSUNTO E PERÍODO

- **Qual a melhor base de dados sobre um determinado tema?**
 - a) Bases de dados **orientadas para** uma disciplina (CA, BA, ERIC...)
 - b) Bases de dados orientadas para uma missão (**NASA...**)
 - c) Bases de dados orientadas para um problema (Energyline, Pollution...)
 - d) Bases de dados **multidisciplinares** (NTIS, SCI, LC, CDI...)?

COBERTURA DE TIPOS DE DOCUMENTOS

- » **Que tipos de documentos são indexados na base de dados?**

Artigos, livros, folhetos, teses, **pré-prints**, relatórios, patentes, normas técnicas e outros.

INDEXAÇÃO E CODIFICAÇÃO

- **Qual a característica da linguagem de indexação?**
 - a) Inclui linguagem livre, ou seja, **palavras-chave** do próprio documento ou **enriquecimento** de termos feitos por especialistas;
 - b) Linguagem controlada, ou seja, uso de **thesaurus**, vocabulário de termos hierárquicos, lista alfabética, cabeçalhos de assunto;
 - c) Codificação, por exemplo: código de classificação privado/público, específico/genérico.

ITENS DE INFORMAÇÃO/CAMPOS DE BUSCA DISPONÍVEIS

- **Que campos de busca estão disponíveis?**
 - a) Inclui todos os dados para identificar o documento original se necessário: AU (autor), TI (título do documento), SO (Fonte onde foi publicado o documento), DE (Descritor), PY (Ano de publicação), AB (Resumo), PA (**Patenteador/Inventor**), PN (Número da Patente), PD (Data de publicação da patente) e outros.
 - b) resumo do autor;
 - c) resumo do produtor da base.

VERSÃO IMPRESSA

- **A base de dados possui correspondente versão impressa**

ACESSO AO DOCUMENTO

- **Qual a facilidade para obtenção de documento citado na base?**

Oferece esse serviço via: produtor da base (?); via terceiros (?).

Essas e outras perguntas devem ser cuidadosamente examinadas pelo intermediário e/ou usuário final para que os resultados a serem alcançados sejam plenamente satisfatórios para o usuário da informação.

2.2 Estudos sobre as linguagens: natural e controlada

Ao longo das três últimas décadas de crescimento, expansão, utilização local e remota da informação armazenada em grandes sistemas de recuperação da informação, inúmeros autores se dedicaram ao estudo das linguagens natural e controlada aplicada à indexação e recuperação de informações. Não pretendemos esgotar o assunto nesse tópico, porém, apresentar apenas os mais significativos estudos levantados nas fontes de informação explicitadas anteriormente.

BHATTACHARYA (1974) analisou os experimentos realizados com a linguagem natural e o seu desempenho para recuperação nas áreas de aerodinâmica, ciência nuclear, física e biologia. Esses testes de avaliação do desempenho da linguagem natural tanto no processo de indexação, quanto no de busca, demonstraram que, o uso da linguagem natural como linguagem de indexação e/ou de recuperação é viável com um controle mínimo de terminologia, ou mesmo com total ausência de controle nessas áreas. A autora demonstrou, ainda, que, nas áreas de química, física, botânica, zoologia e geologia, uma linguagem artificialmente elaborada, com controle de terminologia, torna-se quase impossível de ser **criada** devido ao acelerado desenvolvimento dessas áreas; sugeriu, portanto, o uso da linguagem natural como instrumento de indexação e de recuperação simultaneamente. Citou também, o esforço que, desde 1982, a International Union of Pure & Applied **Chemistry** (IUPAC) vem desenvolvendo na área da química para o controle da nomenclatura dos compostos orgânicos, conforme **extrato** da introdução de publicação editada pela IUPAC:

Essas regras... constituem-se em recomendações para a denominação dos tipos de compostos e de compostos individuais. Elas não são exaustivas, excepto em casos **específicos**...A Comissão deseja que cada nação tente reduzir as variações de **nomenclatura**. (*apud* **Battacharya**, 1974, p.248).

SVENONIOUS (1976) num dos primeiros **questionamentos** sobre o processo de recuperação da **informação** analisou os significados dos vocabulários controlados e da linguagem natural nas bases de dados em linha. Avaliou uma possível utilização dos conceitos que compõem um vocabulário controlado com o uso de medidas nas quais eles possam vir a ser quantificados e tratados como variáveis para serem utilizados em pesquisas teóricas e experimentais. **Afirmou** que o controle do vocabulário implica num processo **classificatório**, com duas etapas distintas. A primeira refere-se à classificação de variantes gramaticais do mesmo termo e/ou conceito, significando singular e plural, variantes gramaticais e diferentes flexões dos tempos verbais. Numa segunda etapa, os termos e/ou conceitos são agrupados por descreverem o mesmo conceito ou um similar, isto é, sinónimos ou palavras que são equivalentes em seus significados.

Existem outras relações que podem ser úteis para o objetivo do controle do vocabulário, especialmente os relacionados com os processos de recuperação da informação. Assim, a

possibilidade de **truncagem** de termos à direita, permite a união automática de termos ortograficamente similares, podendo os mesmos serem ainda explorados em relação aos seus sufixos ou **infixos**. Os **dicionários "online"** oferecidos pelos melhores sistemas de bancos de dados, apresentam em ordem alfabética termos com raízes idênticas e com suas variações gramaticais e respectivos significados diferentes, o que permite a **seleção** dos termos que podem ser **aplicados** na estratégia de busca.

A autora finalizou afirmando que o controle **bibliográfico** não pode existir sem o controle do vocabulário, no que concordamos plenamente, porém, essa questão até o momento, ainda não foi solucionada, mesmo com as sofisticadas tecnologias de computação em vigor. Basta observar o uso dos robôs de busca utilizados atualmente na Internet que bem demonstram o esforço de cada empresa em organizar e classificar o "**caos**" **informacional** circulante nas redes mundiais. Esses robôs criam e agrupam em grandes classes os milhões de sítios oferecidos na **Internet**, numa tentativa de organização desse conhecimento.

CHARTON (1977) investigou a eficiência do *Chemical Abstracts Key Index*, um vocabulário não controlado constituído pelas entradas **selecionadas** dos títulos ou textos dos documentos. **Selecionado** um tema para determinação de entradas no "**Keyword Index**", os resultados foram comparados com a busca em linha e ficou demonstrado que o índice de **palavras-chave** impresso não possibilitou a identificação dos conceitos pesquisados.

Um dos primeiros estudos da década de 80 sobre a utilização da linguagem **natural** e controlada na estratégia de busca, foi o de CALKINS (1980). Comparando as duas linguagens e utilizando as bases COMPENDEX e **ENVIROLINE**, a autora estabeleceu algumas hipóteses para a análise de um tema de interesse da *U. S. Environmental Protection Agency (EPA)*. Na primeira hipótese, a premissa era que a busca em linha, em linguagem natural, recuperaria automaticamente todos os itens indexados (com a linguagem controlada) e, na segunda, que a busca utilizando o vocabulário controlado das bases mencionadas recuperaria virtualmente todos os itens mais pertinentes. Duas estratégias de busca foram elaboradas para a comprovação dessas hipóteses: uma baseada na linguagem natural e outra composta de termos de indexação controlados e códigos. Relatando as dificuldades encontradas para a escolha de termos, tanto para a linguagem natural quanto para a controlada, a autora conclui que não pôde ser comprovada a primeira hipótese. O teste da segunda hipótese comprovou que as melhores citações foram recuperadas, porém, em número bastante limitado. A **conclusão** foi pelo uso das duas linguagens **simultaneamente** na estratégia de busca, **pois**, a **combinação** das mesmas aumentou **consideravelmente** a recuperação, e que o intermediário "**analista da busca**" deve possuir uma grande habilidade em traduzir a necessidade de informação do usuário tanto para a linguagem de busca do sistema quanto para as características de cada base.

HENZLER (1978), comparou amostras do uso de termos de indexação e termos livres na **base** de dados CANCERNET. Numa amostra de 100 títulos, ele encontrou que 35% de todas as

palavras frequentes nos títulos não tinham significados equivalentes, contra os descritores do vocabulário controlado. Confirmando que 50% dos descritores assinalados nos documentos não possuíam representações no texto livre dos documentos, Henzler concluiu que tanto os termos livres, quanto os do vocabulário controlado, deviam estar presentes numa combinação ideal durante a elaboração da estratégia de busca.

RAITT (1980) apresentou a linguagem de indexação como principal componente na estratégia de busca em sistemas "online" de recuperação de **informação**. Abordou os vários artifícios usados pelas linguagens de indexação, portanto com o uso de vocabulários controlados, para o aumento da **revocação** e da precisão. Dentre os que aumentam a **revocação**, nomeou os de controle de sinonímia, os de conexão de termos, controle de formas dos termos e os de agrupamento. Para os que influenciam a precisão, destacou os de coordenação, os de conexão e indicadores de função e os que estabelecem pesos para os descritores. Revisou algumas linguagens de indexação, por exemplo: *NASA Thesaurus*, *Thesaurus of Engineering and Scientific Terms*, *Thesaurus of Metallurgical Terms*, *Subject Headings for Engineering*, *INIS Thesaurus* e *INSPEC Thesaurus* que foram utilizados no sistema QUEST, demonstrando como os artifícios de revocação e precisão foram usados nesse sistema.

MARKEY, ATHERTON e NEWTON (1980), no projeto de estudos sobre a base de dados ERIC (*Educational Resources Information Center*), no final da década de 70, analisaram o uso da linguagem livre e do vocabulário controlado nas estratégias de busca em linha. Nesse projeto foram comparadas as buscas que utilizavam linguagem natural e controlada objetivando **comprovar**, dentre outras, as seguintes possibilidades: que características distinguem o vocabulário de busca? que percentual da estratégia é formulado em linguagem livre? qual a preferência pelo uso do termo livre **quando** existe o **descriptor disponível? neste caso, os conceitos são inexpressíveis** para serem considerados descritores?

As análises relacionadas ao projeto não foram consideradas exaustivas pelas autoras, porém "pelo menos pôde proporcionar algumas indicações referentes a que tipos de tópicos de busca são melhores para uso do termo **livre**", já que um dos objetivos era aprimorar o acesso à base de dados ERIC, criando outros instrumentos de auxílio além do *thesaurus*.

Os resultados das estratégias de busca utilizando termos livres e controlados foram comparados e submetidos a especialistas em Educação. Essa análise revelou que as formulações de busca utilizando a linguagem natural tiveram maior revocação (93%) e menor precisão (71%) do que as estratégias que utilizaram vocabulário controlado, com revocação de 76% e precisão de 95%. Concluíram que a busca com termos livres pode, frequentemente, ser a melhor opção quando se deseja alta revocação, todavia, o uso combinado da linguagem livre e da controlada oferece melhores resultados.

CARROW e NUGENT (1981) apresentaram uma avaliação comparativa dos métodos de busca com termos de indexação *versus* termos livres usando a base *National Criminal Justice*

Reference. O texto pesquisado incluía resumos, títulos dos documentos e anotações. Os resultados mostraram que os dois métodos de busca apresentaram a mesma precisão no desempenho, mas as buscas com termos de indexação produziram uma significativa e melhor recuperação. Os autores propuseram que os dois métodos fossem utilizados como complemento um do outro e afirmaram que o melhor desempenho da estratégia de busca seria aquele que utilizasse os dois métodos **concomitantemente**.

SCHABAS (1982) descreveu os estudos comparativos realizados para **avaliara** eficiência de duas linguagem de indexação: **Library of Congress Subject Headings** (LCSH) e PRECIS, ambas em relação à **revocação** e precisão. O principal **objetivo** desse estudo foi comparar a proporcionalidade das taxas de revocação e precisão das duas linguagens, observando os efeitos de enriquecimento das mesmas com a palavras dos títulos das citações do experimento. Foram identificadas, também, as diferenças de comportamento entre as linguagens de indexação na área de Ciências Sociais e na de Ciências Puras e Aplicadas. Os resultados da análise demonstraram que o acréscimo das palavras em linguagem natural contidas nos títulos de documentos aumentou a **revocação**, com uma perda mínima em **termos** de precisão. Evidenciou-se, ainda, que as linguagens controladas usadas foram eficientes na recuperação de documentos para a área de Ciências Puras e Aplicadas, mas que o PRECIS teve uma clara vantagem sobre a LCSH para a de Ciências Sociais.

Analisando a crescente utilização da editoração **eletrônica** aplicada principalmente aos **jornais** de grande porte, PEREZ (1982) apresentou as vantagens e desvantagens do uso das linguagens natural e controlada no registro de informações em bases de dados textuais. Este autor sugeriu que, através da geração de um pequeno vocabulário controlado, devia ser levado em consideração o enriquecimento das matérias jornalísticas, a fim de aumentar os benefícios da recuperação textual com o uso da indexação, já que a tendência é, cada vez mais, existirem bases de dados com texto completo.

KNAPP (1982) observou que a estratégia de busca em linguagem natural é uma parte essencial no conhecimento dos intermediários que executam as buscas em bases de dados. Analisando as diferentes possibilidades de planejamento de estratégias de busca com a utilização da linguagem natural e com a linguagem controlada, sugeriu algumas técnicas para elaboração das estratégias, focalizando principalmente as que usam a linguagem natural. Relacionou dentre outras: os operadores de proximidade que podem ser usados nas buscas; os casos especiais para uso dos termos livres na estratégia; os problemas que ocorrem com a busca em linguagem natural; as fontes de pesquisa de sinónimos dos termos em LN e as estratégias para encontrar os termos nos bancos e nas bases de dados.

SCHRODER (1983) confirmou, e acrescentou na afirmação de **CALKINS** (1980), que, **além** do vocabulário natural e do controlado, é preciso também usar os acrónimos, quando for necessário. Alertou, porém, para a estrutura de **informação** de cada base. A identificação

apropriada dos elementos descritivos de um item e/ou registro de informação contido numa base de dados é de fundamental importância no planejamento da estratégia de busca. Assim, a primeira etapa a ser verificada será a de identificar na documentação da base a ser consultada, a codificação definida pelo banco de dados para cada campo do item de **informação**.

Vários autores têm se concentrado em estudos de formulação da estratégia de busca, baseados no uso simultâneo das linguagens natural e controlada.

WAGERS (1983) analisou a eficiência da estratégia de busca em linguagem natural através do desempenho unicamente dos termos que compõem o resumo dos documentos. Em experimentos realizados no Sistema **DIALOG** e nas bases de dados ***Energylite, Management e America History and Life*** o autor levantou a hipótese de que os termos livres das frases do resumo podem ser equiparados com os termos de indexação e, nesse caso, podem ser usados eficientemente na estratégia de busca. Sugeriu ainda que, sob certas circunstâncias, o resumo pode concorrer mais eficientemente para busca em linha principalmente quando existe, por parte do produtor da base, interesse em destacar, por exemplo, instituições, dados factuais e outros. Nesse caso específico, baseando-se no uso da linguagem natural enriquecida pelos objetivos básicos do produtor da base.

Esse fator, quase isolado, vem comprovar a necessidade de **aprimoramento** das linguagens de busca, o que já ocorre frequentemente nos bancos de dados, isto é, criação de recursos adicionais para bases de dados especializadas, incluindo busca no texto do resumo, em frases agrupadas por operadores de união ou apenas em campos específicos.

SIEVERT e BOYCE (1983) levantaram um **questionamento** sobre a função **tradicional** do vocabulário controlado como instrumento de recuperação. Analisando a estrutura de informação das bases e do banco de dados **DIALOG**, demonstraram que os mesmos estão sendo **efetivamente** utilizados como um instrumento de precisão, e não de **revocação**.

Nos testes levados a efeito com termos extraídos do ***Thesaurus of ERIC Descriptors***, da base ERIC, já analisada por MARKEY, ATHERTON e **NEWTON** (1980), os autores **concluíram** que o vocabulário controlado pode ser eficientemente utilizado sem a completa entrada dos termos, tendo em vista a possibilidade de combiná-los com os operadores de adjacência do sistema **DIALOG**, e com os recursos do sistema para restrição do uso desse termos, apenas no campo de descritores.

Alertaram, ainda, para a correta utilização do vocabulário controlado e seus limites, pois os mesmos aumentam a revocação com perdas mínimas de precisão dos resultados e, nesse caso, os termos controlados tornam-se claramente um instrumento de precisão.

Dentre as linguagens controladas que podem ser aplicadas na recuperação da informação, **SVENONIOUS (1983)** destacou, como um instrumento de busca, o uso do sistema de classificação definido pelo autor da base. Uma das possíveis formas em que a classificação pode ser usada na

recuperação é a ampliação do número de documentos relevantes recuperados. Essa classificação, segundo a autora, não se restringe às tabelas de classificação tradicionais, antes inclui todos os instrumentos **classificatórios** desenvolvidos e aplicados pelos produtores das bases de dados no processo de indexação. Um bom exemplo desse instrumento é a base correspondente ao *Biological Abstracts* impresso, denominada *Biosis Previews*, que utiliza uma lista de **palavras-chave**, uma de códigos de conceitos e outra de códigos **bio-sistemáticos** para a indexação e/ou recuperação de informação. Por esse motivo, o uso da classificação pode ser aplicado tanto para ampliar a recuperação, quanto para melhorar a precisão dos resultados e, ainda, para minimizar os custos de conexão, refinando os resultados. Finalizando, a autora apontou tendências de pelo menos oito possíveis usos da classificação nos sistemas de recuperação da informação.

PITERNICK (1984) destacou alguns tipos de vocabulários utilizados tradicionalmente como instrumentos de indexação e que, com o advento dos sistemas de recuperação **"online"**, foram transformados em vocabulários de busca. Afirmou que os **"Thesaurie** Listas de Cabeçalhos de Assunto, num passado recente, caracterizavam-se mais como vocabulários de indexação do que como vocabulários de busca", **propondo, então,** a geração de vocabulários voltados para a recuperação da informação e não para a indexação. Alertou, ainda, que embora um vocabulário controlado seja extremamente útil para elaborar estratégias de busca visando especificamente a precisão dos resultados, existem momentos em que se torna necessária e imprescindível a utilização de termos extraídos da linguagem natural. Relembrou que a cada nova **atualização** de uma base de dados, novos termos e/ou conceitos são incorporados e que a busca com termos livres em alguns casos é de extrema importância. Apresentou, também, uma proposta para o enriquecimento dos **thesauri** e listas de cabeçalhos de assunto, denominando-a de Vocabulário **reunido**" ou **"merged vocabulary"** para uso em áreas correlatas.

TENOP1R (1985) analisou o **desempenho** da linguagem natural e da linguagem controlada numa base de dados textual (*Harvard Business Review Online*) onde foi aplicada uma metodologia de busca, separadamente por texto completo, por resumo, por título e nos descritores para 36 amostras da pesquisa para cada uma das linguagens. Cada uma das diferentes estratégias de busca, produziu documentos únicos, e a busca no texto integral com a linguagem natural produziu maior recuperação e menor precisão do que a busca com a linguagem controlada.

Uma outra característica das bases de dados textuais que convém destacar é o número de pontos de acesso ao documento. Tipicamente, numa grande base de dados textual será bem mais difícil encontrar uma recuperação aceitável com uma precisão tolerável. As bases de dados textuais proporcionam uma grande recuperação e mais baixa precisão, enquanto as bases de dados bibliográficas oferecem um melhor desempenho.

AUSTIN (1986) relatou as mudanças ocorridas nas funções específicas de indexação com o uso dos computadores, mostrando que as mudanças ocorreram também com novas **atividades** na geração dos produtos de indexação: os índices e os resumos. Relembrou que o uso de

vocabulário controlado na indexação, e o seu respectivo uso na recuperação da **informação**, vai requerer o estabelecimento de certas regras **terminológicas** recomendadas, tais como:

- a) **conceitos devem ser representados consistentemente para os propósitos de recuperação, por substantivos ou frases substantivadas;**
- b) os indexadores devem trabalhar com um vocabulário de termos preferidos, **designando-se** um dos sinónimos de um determinado conceito, como o termo mais adequado para uso;
- c) a opção pelo singular ou plural dos conceitos e suas **exceções** devem ser registradas claramente nesse vocabulário, visando a consistência da indexação, sua **fidedignidade** e posterior uso na recuperação. (Austin, 1986, p.8).

Essa preocupação de AUSTIN, no auge do crescimento da indústria da informação, com a geração dos **índices** e resumos em forma legível por computador, denominadas bases de dados e sua consequente comercialização com o acesso remoto e, **também**, com o mercado crescente do CD-ROM, denota o necessário desenvolvimento de instrumentos de apoio para a elaboração das estratégias de busca, que são os vocabulários controlados e/ou as linguagens documentárias geradas pelo produtor de cada base de dados. Se o desenvolvimento desses vocabulários controlados para cada base de dados utilizasse alguns dos padrões **mínimos** para sua geração, as chances de melhoramento na recuperação da informação seriam ampliadas sensivelmente. Os usuários não precisariam reescrever suas estratégias de busca, caso necessitassem procurar em duas ou mais bases de dados que cobrissem o mesmo tema.

HARTER (1986) analisou as linguagens para recuperação da informação, abordando suas características e classificando-as como: linguagens de busca, linguagem natural, indexação por assunto, vocabulários controlados e indexação por **citações**. Também apresentou uma breve comparação entre a linguagem natural e os vocabulários controlados, principalmente no processo de recuperação da informação. Segundo o autor, tanto o uso do vocabulário controlado quanto o uso da linguagem natural apresentam vantagens e **desvantagens** na indexação e na recuperação da informação. Enquanto o primeiro é **"rígido, inflexível, mas preciso**, o outro é altamente expressivo, flexível, mas potencialmente ambíguo." Caracterizou como atributos da linguagem natural os seguintes pontos: altamente flexível e expressiva; muito difícil para execução de buscas genéricas; oferece variedade de pontos de acesso; apresenta problemas com sinónimos, homógrafos e falsas coordenações; não requer treinamento para sua utilização; é fácil para a representação de conceitos novos e complexos; não é padronizada; possui liberdade de expressão; não necessita de indexação e oferece um alto grau de exaustividade. Como atributos da linguagem controlada, foram mencionados: não é expressiva; seu uso é relativamente fácil para executar buscas genéricas; permite limitados pontos de acesso; controla os sinónimos e homógrafos; evita falsas coordenações; é altamente inflexível e compactada; requer treinamento para sua utilização; dificulta a representação de conceitos novos e complexos; é precisa e

padronizada, apresenta problemas de **consistência** na indexação e oferece um baixo grau de exaustividade. Finalizando, o autor apresentou algumas sugestões para a tomada de decisão por parte do **intermediário** e/ou usuário final quanto ao uso da linguagem natural ou da controlada.

FIDEL (1986) analisou a importância dos guias para elaboração de **resumos**, desenvolvido pelos produtores de bases de dados, visando ao **aprimoramento** de busca de informações em linguagem natural. Comprovou que o desempenho na recuperação do documento está **diretamente** relacionado com o conteúdo e o estilo da linguagem utilizada no resumo. Num outro estudo, também de 1986, a autora investigou as políticas de resumo usadas pelos produtores das bases de dados com o **objetivo** de identificar a relação entre a ampliação e a recuperação de informação na linguagem natural contida nos resumos dos documentos. Os editores consideram o conteúdo dos resumos e o idioma como um fator preponderante no melhoramento da recuperação. Recomendou que, uma vez decidido pelo **resumista** que conceitos serão utilizados nos resumos e suas grafias, esses termos devem ser coordenados com os termos de indexação registrados nos vocabulários controlados. Finalizou afirmando que um guia para a elaboração de resumos, com seus diversos tipos e determinando ainda, os respectivos tamanhos dos mesmos, pode oferecer um melhor desempenho na recuperação em linguagem livre e/ou natural.

TENOPIR (1987) examinou a postura do intermediário que **operacionaliza** as buscas e reafirmou que, independente da experiência adquirida, o mesmo se posiciona de forma estática no processo de recuperação, **pois**, age como um profissional que não questiona o processo **subjetivo** de indexação, **selecionando** para a estratégia de busca apenas termos do vocabulário controlado. Mencionou que muitas bases de dados não desenvolveram vocabulários controlados para a recuperação, porém as que são subproduto de acervos de grandes **bibliotecas** quase sempre já geraram vocabulários do tipo: cabeçalhos de assunto, **thesaurus** ou, ainda, listas de descritores permitidos para a indexação e/ou recuperação. A maioria dessas bases estabelece um campo de busca para o descritor que inclui termos **selecionados** de alguns desses tipos de vocabulários controlados. Lembrou que, nas áreas de Ciências Exatas, os termos usados nos títulos do documentos apresentam um nível de ambiguidade bem menor do que nas áreas de Ciências Humanas e Sociais, contribuindo, portanto, para que nessas áreas a linguagem natural detenha uma recuperação mais eficaz. A autora apontou várias dificuldades a serem superadas em relação ao uso apenas do vocabulário controlado nas buscas, causadas em grande parte pela **atualização** constante do conhecimento e da morosidade em se manterem atualizados os respectivos vocabulários. Sugeriu que, dependendo do tema de busca solicitado e da base de dados a ser consultada, sejam utilizadas nas mesmas estratégias tanto os termos e/ou conceitos em linguagem natural, quanto os termos da linguagem controlada.

SVENONIOUS (1988), num contexto de recuperação de informação, *analisou* os instrumentos utilizados para indexação, a saber: as classificações, os cabeçalhos de assuntos e os *thesaurus*, apresentando considerações para a construção dos mesmos. Esses vocabulários

controlados e suas respectivas funções nos sistemas de recuperação da informação, contribuem para a ampliação dos resultados relevantes desejados pelos usuários. Sugeriu que um sistema de mapeamento usando **tecnologias** avançadas de computação podem vir a identificar nos termos do usuário em linguagem natural, os termos equivalentes do vocabulário controlado. Recomendou o desenvolvimento de vocabulários controlados, voltados para a recuperação da informação e lembrou que bases de dados podem ser agrupadas em grandes **bases**, porém torna-se necessário **compatibilizar** as regras para a geração de **thesaurus**, visando à normalização do conjunto de vocabulários controlados.

LANCASTER, ELLIKER e **CONNEL** (1989), num artigo de revisão do **ARIST**, levantaram os seguintes subtemas relacionados ao tópico análise de assunto: a presença, identificação e a expressão do assunto nos textos dos documentos, nas bases de dados, nas linguagens natural e controlada, nos pedidos de informação e nas estratégias de busca. A cobertura da revisão no período de **1986-1988** proporcionou o agrupamento dos documentos analisados em seis grandes categorias: teoria e prática da indexação; vocabulários controlados, incluindo classificação e cabeçalhos de assunto; estratégias de busca e métodos de busca; busca em linguagem natural; indexação automática e o uso de citações na recuperação da informação.

BOYCE e **McLAIN** (1989), num estudo *realizado na linguagem de recuperação de bancos de dados*, revelaram que a profundidade da indexação incluindo o total de termos que são assinalados em média por documento e os pontos de acesso disponíveis no banco têm um expressivo efeito nos padrões de desempenho desses sistemas de recuperação da informação. Demonstraram que o uso de termos na estratégia de busca sem a identificação do campo de busca correspondente vai implicar na identificação imediata do mesmo no campo de descritor da LC, por parte da linguagem de recuperação do sistema. Demonstraram ainda que, como os termos da LC aumentam a precisão nos resultados da busca, o uso dos mesmos numa estratégia devem ser criteriosamente examinados, tendo em vista a estrutura de informação oferecida pelo sistema. Analisaram as questões relacionadas com a profundidade de indexação nas bases de dados, concluindo que a função do vocabulário controlado em sistemas de recuperação é, principalmente, ampliar a **precisão** e não, a **revocação**.

NICHOLLS e **HOLTMANN** (1989) relataram a investigação realizada numa base de dados em CD-ROM na qual foram testados o uso da linguagem natural e da controlada na estratégia de busca sobre um tema específico, demonstrando que os dois enfoques ainda são muito polêmicos no que se refere aos fatores de revocação e precisão. Sugeriram, para a estratégia de busca, uma combinação flexível de termos extraídos da linguagem natural e da linguagem controlada da base de dados a ser utilizada para a pesquisa, lembrando que a natureza do pedido de busca, a estrutura de informação de base e a linguagem de busca do sistema são variáveis intervenientes no processo de recuperação.

COMBER e STANFORD (1989) estabeleceram analogias entre o conceito de **texto-livre**, ou linguagem natural, e o conceito de vocabulário controlado, num contexto de busca em linha. Mencionaram as dificuldades inerentes ao processo de produção de bases de dados, no que se refere às **diretrizes** a serem adotadas nas **atividades** de indexação, dentre elas os fatores custo de indexação e de recuperação. Alertaram que raramente os indexadores são especialistas nos assuntos cobertos pelas bases de dados sendo este, um dos principais fatores que afetam a credibilidade do produtor da base e da informação registrada na mesma. Apontaram ainda, as dificuldades encontradas nos métodos tradicionais de indexação onde os assuntos são, por vezes, indexados de uma forma muito geral e, em outras ocasiões, de forma muito específica.

Num **projeto** desenvolvido em 1987, ROWLEY (1990) comparou a indexação e busca em linguagem natural com a busca em linguagem **controlada**. Utilizando uma pequena base de dados contendo títulos e descritores extraídos do ER/C *Thesauruse* os cabeçalhos de assunto da Sears *List of Subject Headings*, a autora realizou uma série de buscas usando termos de linguagem natural e da linguagem controlada. As medidas de **revocação** e precisão foram comparadas no experimento para avaliar a eficácia de cada uma das linguagens durante a busca. Comprovou que o uso de uma linguagem de indexação controlada na busca não ofereceu, necessariamente, o melhor desempenho. Afirmou que a disponibilidade de novas interfaces para o uso de linguagem natural na busca de informações com a introdução dos sistemas **hipermediae hipertexto** poderá provocar um desempenho melhor nas linguagens natural e controlada para a busca da informação.

FIDEL (1991) apresentou um estudo de caso onde relatou o experimento realizado com os intermediários durante o processo de **seleção** da terminologia a ser utilizada na estratégia denominada "**chaves-de-busca**". Analisou o uso de descritores extraídos de vocabulários controlados e também a busca que utiliza texto livre. O método de estudo de caso ofereceu dados acerca da maneira pela qual os dois tipos de busca **interagem** por intermédio da observação de 47 intermediários durante a execução de suas estratégias de busca e pela análise de seus protocolos verbais e de busca no momento em que decidem pela modificação das "**chaves-de-busca**". Esse estudo de comportamento da busca "**online**" realizado por intermediários experientes, **comparando**-os com os inexperientes, é de fundamental importância para o desenvolvimento de sistemas de recuperação orientados para o usuário, além de contribuir para as definições das funções dos termos livres e dos termos controlados na busca em linha.

KABACK (1992) analisou um dos debates mais controversos no contexto de recuperação da informação que tem sido o argumento sobre o uso ou não da linguagem natural ou da controlada nas estratégia de busca. O **questionamento** sobre qual delas tem a melhor performance na recuperação da informação, foi testado num experimento sobre um tema em documentos de patentes. Nessa investigação, foram testadas as hipóteses sobre o uso da linguagem natural e da controlada, e para cada base de dados de patentes consultada sobre o tema foram levantadas as dificuldades relativas aos tipos de vocabulários **possíveis** de serem utilizados na estratégia de

busca. **Sugeri**, como **conclusão**, que para a área de patentes devem ser usadas ambas as linguagens na estratégia de busca, incluindo-se, ainda, a Classificação Internacional de Patentes.

COUSINS (1992) investigou as facilidades da busca por assunto nos catálogos em linha de acesso público. A pesquisa tinha por objetivo identificar se o melhoramento da indexação utilizado nos catálogos em linha ampliaria o desempenho na recuperação da informação. Para um estratégia de busca ser bem sucedida, era necessário que os termos de busca do usuário fossem os mesmos utilizados pelos **indexadores**. Por esse motivo, foram examinadas as perguntas formuladas pelos usuários com o objetivo de traduzir os termos da linguagem natural para um vocabulário controlado. O experimento utilizou os seguintes vocabulários: *PRECIS*, *Library of Congress Subject Headings* e a *Classificação Decimal de Dewey*. Os testes demonstraram algumas perdas de informação como resultado desse processo de indexação, porém o objetivo da representação do conteúdo do documento era permitir que os usuários encontrassem a informação, não o de prover uma representação "**ideal**" do conteúdo dos documentos. Finalizando a investigação, a autora demonstrou que o *PRECIS* obteve melhor desempenho, sendo o mais **efetivo** vocabulário para representar as necessidades de informação dos usuários. A lista de cabeçalhos de assunto da *Library of Congress* obteve um desempenho regular, enquanto que a *Classificação Decimal de Dewey* provou ser inadequada para a recuperação dos catálogos em linha.

ROWLEY (1994) revisou o debate sobre as linguagens natural e controladas utilizadas no contexto de indexação e recuperação das bases de dados em linha dos sistemas de recuperação da informação, dos catálogos em linha de acesso público e das bases de dados em CD-ROM. Definiu a linguagem natural de indexação como a linguagem comum do documento que está sendo indexado, podendo estar contida no título, resumo ou no texto integral do documento, enfocando, principalmente, as linguagens de indexação utilizadas no próprio processo de indexação e no processo de busca por assunto. Definiu ainda, as linguagens controladas de indexação como "linguagens em que os **termos** são utilizados para representar assuntos e os mesmos são atribuídos aos documentos por um **indexador**". Nesse caso, o intermediário que planeja a estratégia de busca deve conhecer essa linguagem e consultá-la a cada momento em que for necessária a formulação da estratégia.

Entretanto, a prática e a experiência profissional têm demonstrado o quão distante os intermediários e/ou usuários finais se encontram desses instrumentos de trabalho, principalmente no que se refere à recuperação de informação. Os próprios bancos de dados e os produtores das bases de dados nem sempre apresentam esses documentos ao público usuários dos seus serviços, o que dificulta sobremaneira o planejamento da estratégia de busca. Existem exceções, como o banco de dados DIALOG e o ORBIT, que incluem no manual do banco de dados um capítulo denominado "**Search Aids**", onde relacionam os diversos vocabulários controlados utilizados na indexação dos documentos. Uma lista desses vocabulários encontra-se no Anexo IV.

Nas bases armazenadas em CD-ROM, frequentemente são apresentados os índices das palavras significativas do documentos, incluindo os termos definidos como descritores pelo produtor da base. A consulta, porém, a esses índices numa negociação de busca em linha ocasiona aumento considerável no custo final da busca, enquanto a consulta aos índices das bases em **CD-ROM** contribui para o **aprimoramento** da **seleção** de termos para as estratégias de busca, implicando, ainda, num tempo maior de execução da estratégia, o que ocasiona custos sobre o salário do intermediário.

Por outro **lado**, a avaliação do uso das linguagens natural e controlada nas estratégias de busca tem sido realizadas, com base nas medidas de **revocação** e precisão inicialmente utilizadas nos experimentos de CRANFIELD.

No projeto CRANFIELD II, CLEVERDON (1977) testou três linguagens de indexação. A primeira, baseada na linguagem natural; a segunda, na linguagem controlada, e uma terceira com conceitos simples extraídos dos títulos dos documentos. Esses testes foram baseados nas medidas de revocação e de precisão e o autor, através dessas experiências, sugeriu uma indexação exaustiva com cerca de 30 descritores para cada documento. Os estudos provenientes desses testes ampliaram as pesquisas relacionadas com a avaliação dos resultados da recuperação em bases de dados para as diversas linguagens documentárias utilizadas no processo de indexação.

Essa alternância no processo de indexação **tradicional**, sem o uso de um vocabulário controlado, especialmente de um **thesaurus**, onde as notas de escopo podem **direcionar** o indexador para o melhor conceito a ser utilizado, provocam falhas tanto na recuperação da informação quanto na precisão dos resultados. Este fato vem sendo demonstrado por diversos outros autores, dentre eles: SVENONIOUS (1988, 2000), LANCASTER (1979,1986,1993), ROWLEY (1990, 1994), FIDEL (1983, 1991) e SARACEVIC (1988,).

2.2.1 LC e LN: Vantagens e Desvantagens

Numa base de dados bibliográfica, frequentemente o índice básico é composto por palavras extraídas do título, do resumo e do campo de descritor. Portanto, se na estratégia de busca não for indicado em qual campo o termo a ser pesquisado será utilizado, a busca será em princípio, nesses três campos nos bancos de dados. Qual a vantagem dessa indexação realizada pelo banco de dados?

Em princípio, a forma exata de uma estratégia de busca é determinada pela natureza da base de dados a ser **acessada** e pela sua respectiva estrutura de informação, isto é, pela formatação de seus campos de identificação do documento e dos campos de identificação do conteúdo temático do mesmo. A identificação dessa estrutura de campos de busca implica no

conhecimento da documentação **básica** fornecida pelos produtores das bases e pelos bancos de dados onde as mesmas estão hospedadas. Nas bases em CD-ROM, a maioria dos produtores armazena nos próprios discos o vocabulário controlado das mesmas, facilitando, dessa forma, o planejamento da estratégia de busca.

LANCASTER (1979) , observou que embora os computadores possibilitem a manipulação de extensas listas de palavras, eles ainda não contribuem para a solução dos problemas intelectuais relacionados com a elaboração das estratégias de busca em linguagem natural, pois:

"O **thesaurus**, ou outro instrumento de controle do vocabulário oferece muita ajuda aos intermediários que executam a busca, incluindo dentre outros: o controle de **sinônimos** e **quase-sinônimos**; a separação dos homógrafos; o uso da **pré-coordenação** para evitar falsas coordenações e relações **incorretas** entre os termos e ainda, a ligação de termos relacionados , todos com suas respectivas **hierarquias**.... O thesaurus pode ser muito específico, mas nunca **tão** específico quanto a linguagem natural, que é a linguagem do discurso dos próprios autores." (Lancaster, 1979, p. 281).

ROTHMAN (1983) levantou um **questionamento** pertinente sobre a busca de informação em linguagem natural, em oposição ao vocabulário controlado. Afirmou que, com a busca em linguagem natural e/ou em **texto-livre**, a base de dados está **efetivamente auto-indexada**, pois, cada palavra no documento indexado é candidata a termo de busca e identifica, ainda, a unidade do texto no qual se encontra. Por esse motivo, a linguagem natural dos documentos se constitui em termos de indexação ou pontos de acesso imediato. Por outro lado, na busca em linguagem natural, os usuários podem interagir diretamente com os itens indexados na base, enquanto que com a indexação utilizando um vocabulário controlado, o indexador é interposto entre os **usuários** e a base de dados, ficando na posição de mediador ou intérprete.

Naturalmente que o uso da linguagem natural tem dificuldades a serem superadas e, da mesma forma o vocabulário controlado apresenta certas desvantagens, pois os termos preferidos pelos indexadores frequentemente não são os termos utilizados pelos usuários em situações específicas de busca. Nesse momento, a figura do intermediário se justifica, sendo o elo de ligação entre a indexação utilizada na base e a necessidade de **informação** do usuário.

Em grandes bases de dados o processo de indexação envolvendo diversos indexadores provoca, com **certeza**, inconsistências na identificação de documentos similares. Essas discrepâncias entre termos assinalados pelos indexadores e os termos utilizados pelos usuários no momento de busca não podem ser considerados genericamente como erros porque, na realidade, o processo de indexação ocorre num outro contexto, o de análise de conteúdo do documento e a tradução desse conteúdo para o vocabulário controlado da base. A busca de informação do usuário, por outro lado, precisa ser traduzida para a linguagem controlada da base, pelo próprio usuário ou por um intermediário. Portanto, as deficiências são inevitáveis, principalmente quando ambos desconhecem as linguagens controladas da(s) base(s) de dados a serem **consultadas**.

As bases de dados, no contexto de recuperação da informação, podem ser **divididas** em duas categorias:

- 1 - as que usam vocabulário controlado;
- 2 - as que usam vocabulário livre ou a linguagem natural.

- **Vocabulário controlado ou linguagem controlada**

Pode ser definido como um conjunto de **termos** organizados de forma hierarquizada **e/ou** alfabética, com o **objetivo** de possibilitar a recuperação de informações temáticas, reduzindo substancialmente a diversidade de terminologia. São também conhecidos como linguagens documentárias ou linguagens controladas.

Uma base de dados, que utilize um vocabulário controlado, possibilita ao **intermediário** no planejamento da estratégia de busca a **recuperação**, no campo específico de descritor, apenas daquelas **palavras-chave** listadas no **thesaurus** e/ou **vocabulário** controlado da base de dados. Nenhum outro termo pode ser usado. Isto significa que o índice será razoavelmente consistente na determinação de itens de informação pertinentes ao tema solicitado. Dificuldades, porém surgem quando se limita a busca apenas ao campo de descritor.

Historicamente, o primeiro sistema a utilizar a linguagem natural com o objetivo de recuperar informações foi o Uniterm, descrito **por Taube** em **1951**, obtendo considerável influência sobre o **desenvolvimento** de outros sistemas de recuperação de informação na década de 50.

Deve-se destacar ainda que a indexação, de acordo com UNISIST (1977), é a operação que descreve e identifica o conteúdo de um documento, mediante o uso de termos. Portanto, o conteúdo temático de um documento pode ser representado por termos selecionados da linguagem natural ou por símbolos.

Na indexação manual, os conceitos são extraídos por um processo de análise intelectual, com base num julgamento **subjetivo**, por parte do **indexador**, dos assuntos tratados no documento. Os termos utilizados na indexação **serão**, com frequência, extraídos de algum tipo de vocabulário controlado, por exemplo, um *thesaurus*, mas **também**, podem ser extraídos do próprio documento e, nesse, caso, são denominados termos **"livres"**.

O processo de indexação manual compreende basicamente três fases:

- 1) compreensão do conteúdo do documento após a leitura completa do título, resumo e de outras partes que compõem o documento;
- 2) identificação dos conceitos, mencionando os processos, as operações, os equipamentos e outros;

3) **seleção** dos conceitos, de acordo com a política de indexação adotada em relação à: exaustividade, especificidade e consistência.

Na elaboração da estratégia de busca, o intermediário, ou mesmo o usuário final, necessita conhecer os critérios de indexação adotados pelo produtor da base de dados, a fim de melhor **identificar** os conceitos a serem empregados durante o processo de busca da informação.

LANCASTER (1986) salientou que a indexação é um processo intelectual **subjetivo**, observando que os indexadores nem sempre **incluem** um mesmo termo/conceito para representar o conteúdo do documento. Esse fator subjetivo do processo de indexação interfere **consideravelmente** na recuperação e, portanto, deve ser levado em conta no momento de se elaborar a estratégia de busca.

Algumas questões relativas aos problemas semânticos da comunicação são solucionados com os vocabulários controlados, porém os resultados obtidos podem trazer uma perda de precisão, caso a linguagem **utilizada** na indexação não seja **selecionada** para a estratégia de busca. A geração de uma linguagem controlada envolve altos investimentos, equipes especializadas e **interdisciplinares**, além de uma consistente aplicação da política definida para a elaboração da mesma e constante **atualização** e revisão.

Reverendo o processo de busca de informação usando o vocabulário controlado, **BRUNDAGE** (1989) traçou um paralelo entre a linguagem científica e a linguagem controlada. Lembrou que os cientistas de uma certa **forma**, estão familiarizados com o vocabulário controlado em suas áreas de especialização e, portanto, os paradigmas dos conhecimentos recebidos nas disciplinas podem ser usados como modelo no ensino do uso, especialmente, dos **thesauri**. Observou que o intermediário que **operacionaliza** as buscas dispõe de conhecimentos prévios sobre o uso das linguagens controladas e da linguagem natural, porém, o usuário final possui uma profunda intuição sobre a terminologia de seu campo de especialização e essas diferenças vão orientar os programas de treinamento para os diversificados grupos de participantes.

Apresentam-se na Tabela 2 as principais vantagens e desvantagens do uso do vocabulário controlado na recuperação da informação.

Tabela 2 - Vocabulário controlado: vantagens e desvantagens

VANTAGENS	DESVANTAGENS
1. Controle total do vocabulário de indexação, minimizando os problemas de comunicação entre indexadores e usuários.	Custos: a produção e manutenção da base de dados terá despesas maiores com a equipe de indexadores. Será necessário ainda, manter pessoal especializado na atualização do <i>thesaurus</i> .
2. Com o uso de um <i>thesaurus</i> e suas respectivas notas de escopo, os indexadores podem assinalar mais corretamente os conceitos dos documentos.	O vocabulário controlado poderá não refletir os objetivos do produtor da base.
3. Se bem constituído, o vocabulário controlado poderá oferecer alta recuperação e relevância e, também, ampliar a confiança do usuário frente a um possível resultado negativo.	Um rígido vocabulário controlado poderá se distanciar das necessidades de informação dos usuários.
4. As relações hierárquicas e as remissivas do vocabulário controlado auxiliam tanto o indexador, quanto o usuário na identificação de conceitos relacionados.	Necessidade de treinamento no uso dos vocabulários controlados tanto para os intermediários, quanto para os usuários finais.
5. Redução no tempo de consulta à base, pois a estratégia de busca será melhor elaborada com o uso do <i>thesaurus</i> .	Desatualização do vocabulário controlado poderá conduzir a falsos resultados .

O uso unicamente da LC, numa estratégia de busca, está **diretamente** relacionado com o tema da pesquisa, sendo objeto, portanto, de decisão operacional do intermediário da busca. Esse intermediário, após o processo de entrevista com o usuário, examinará criteriosamente a documentação da(s) base(s) de dados que serão **utilizadas** para busca da informação. A documentação especializada de cada base, dentre outros, compreende: manual da base de dados, com detalhes específicos sobre a estrutura da informação; **linguagem** controlada utilizada, que pode ser um *thesaurus*, uma lista de descritores, o índice da correspondente versão impressa; o código de classificação utilizado e outros documentos.

- **Linguagem natural ou Vocabulário livre**

Conceitua-se a expressão *linguagem natural* como sinónimo de discurso comum, isto é, a linguagem usada habitualmente na fala e na escrita. No contexto de recuperação de informação, o conceito se refere, normalmente, às palavras que ocorrem em textos impressos, considerando-se como seu sinónimo a expressão **exto livre**".

Se uma base de dados permite o uso de termos não controlados na estratégia de busca, isto significa que qualquer conjunto de caracteres ou termos podem ser pesquisados nessas bases de dados. Muitas bases de dados operarão em ambos os modos, ou seja, desde que o sistema onde estiverem hospedadas **pré-defina** as características de cada campo de busca.

Os "**termos livres**", que podem identificar o conteúdo temático de um item de informação, ocorrem naturalmente nos seguintes campos de busca: título, **resumo**, ou no texto integral do próprio item, quando a base for caracterizada como textual.

KNAPP (1982) resumiu brevemente os casos em que a busca em LN pode obter um melhor desempenho: para tópicos específicos; para temas atuais; para novas terminologias ainda não incluídas nas LC; para uma busca retrospectiva onde o conceito da LC é muito recente e não cobre os anos anteriores; quando o termo da LC é muito abrangente ou muito específico; para pesquisa em várias bases de dados; para identificação imediata de palavras de títulos dos documentos e para **complementação** de citação bibliográfica incompleta.

LANCASTER (1986) relatou que os princípios do sistema Uniterm provocaram uma imediata atenção, pois o conteúdo temático dos documentos podia ser representado mediante o uso de palavras simples (**unitermos**), extraídas do próprio texto a ser indexado. Nesse caso, o indexador extraía do texto os termos simples que iriam identificar o conteúdo do documento, lançando cada termo em fichas especialmente desenhadas. Os documentos eram numerados e as buscas eram **realizadas** mediante a identificação dos números constantes em duas ou mais fichas. Esse processo de indexação utilizando termos livres foi utilizado inicialmente para a busca manual.

Dentre os vários problemas relacionados com o Uniterm, pode ser citado o da dispersão de conteúdos temáticos similares, ocasionando uma demorada e cansativa manipulação de fichas com termos alfabeticamente distantes. Assim, documentos com assuntos relacionados, apareciam sob diferentes unitermos, e uma busca completa sobre um determinado tema exigia a identificação de todas as formas possíveis como esse tema estaria representado. Esse e outros problemas levaram à busca de soluções para **minimização** dessas dificuldades, e por consequência, aos vocabulários controlados e ao desenvolvimento de **thesaurus** para a recuperação de informações.

Nas bases de dados bibliográficas, os campos de busca em que se pode pesquisar usando apenas termos e/ou conceitos da linguagem natural **normalmente** são os do título e resumo dos documentos. Nesses campos, cada palavra é automaticamente candidata a ser pesquisada, **excetuando-se** aquelas designadas pelos sistemas como **não-significativas**, as quais vão compor as listas de palavras proibidas. Assim, as palavras remanescentes são usadas para criar índices que podem ser pesquisados em linguagem natural. Esse recurso pode ser utilizado para **rastrear** temas e seus respectivos conceitos **terminológicos** que ainda não foram incluídos nas **linguaagens**

controladas, ou seja, nos *thesauri*, nas listas de cabeçalhos de assunto, nas tabelas de classificações especializadas, listas de descritores, códigos **taxonômicos**, nomenclaturas e outros.

Cabe ao intermediário que operacionaliza as estratégias de busca, para atendimento às demandas de informação, a decisão **tática** de utilizar esses recursos no planejamento da estratégia de busca, para alternativamente ampliar ou restringir os resultados que se pretendem, de acordo com o definido pelo usuário da informação.

Apresenta-se na Tabela 3 as principais vantagens e desvantagens da linguagem natural no processo de recuperação da **informação**.

Tabela 3 - Linguagem natural: vantagens e desvantagens

VANTAGENS	DESVANTAGENS
1. Permite o imediato registro da informação numa base de dados, sem necessidade de consulta a uma linguagem de controle.	Os usuários da informação, no processo de busca, precisam fazer um esforço intelectual maior para identificar os sinônimos , as grafias alternativas, os homônimos etc.
2. Processo de busca é facilitado com a ausência de treinamentos específicos no uso de uma linguagem de controle.	Haverá uma alta incidência de respostas negativas ou de relações incorretas entre os termos usados na busca (por ausência de padronização).
3. Termos de entrada de dados são extraídos diretamente dos documentos que vão constituir a base de dados.	Custos de acesso tendem a aumentar com a entrada de termos de busca aleatórios.
4. Temas específicos citados nos documentos podem ser encontrados.	Uma estratégia de busca que arrole todos os principais conceitos e seus sinônimos deve ser elaborada para cada base de dados (ex: nomes comerciais de substâncias químicas não ocorrem no Chemical Abstracts).
5. Elimina os conflitos de comunicação entre os indexadores e os usuários, pois, ambos terão acesso aos mesmos termos.	Perda de confiança do usuário numa possível resposta negativa.

A decisão quanto ao uso da linguagem natural na estratégia de busca depende frequentemente dos recursos oferecidos pelo sistema e das restrições impostas pelo tema da pesquisa.

Uma considerável literatura devotada a recuperação da informação, em bases de dados em linha e/ou em **CD-ROM** tem sido encontrada em revistas especializadas como, por exemplo: *Online Review & CD-ROM, Databases, Journal of the American Society for Information Science, Journal of Documentation, Special Libraries, Aslib Proceedings, Information Processing & Management*, além de revisões de literatura publicadas no *Annual Review of Information Science (ARIST)*, e em trabalhos de Conferências como as da *American Society for Information Science (ASIS)*.

Essa revisão sugere fortemente que o debate sobre o desempenho da linguagem natural e da linguagem controlada nos sistemas de recuperação da informação continua sendo uma área de grande interesse e **atividade**. O impacto crescente das tecnologias de armazenamento, acesso e recuperação da informação afetou profundamente os estudos sobre os aspectos complexos da busca tanto no que se refere ao processo de entrevista (a primeira etapa), quanto à **seleção** de termos para busca e, também, quanto à execução da respectiva estratégia. Nas últimas três décadas, os estudos e análises dessas questões o avanço em certas áreas, por exemplo: a indexação automática; os sistemas especialistas para auxílio na negociação da busca; os estudos sobre **interação homem-máquina**, tomando as linguagens de recuperação mais amigáveis para os usuários inexperientes; a pesquisa no texto completo; o acesso público aos catálogos em linha; o **hipertexto** e o acesso aos sistemas pelo usuário final.

Como sugeriu **ROWLEY** (1994), o debate pode ser visto sob dois ângulos: o da pesquisa e o da prática. Assim, avanços da pesquisa podem conduzir à mudanças significativas nos sistemas de recuperação da informação; e avanços relacionados com a prática sugerem que os usuários desses sistemas devem continuar a exercitar suas habilidades de elaboração de estratégias de busca, utilizando os termos e/ou concertos da linguagem natural e da controlada.

A solução para essas questões ainda não foi definida apesar dos inúmeros estudos e **experimentos** efetuados nas duas últimas décadas.

2.2.2 LC e LN em bases de dados

Os diversos sistemas de recuperação da informação denominados bancos de dados, também participam do processo de indexação das bases de dados hospedadas nos mesmos. Assim, os bancos como **DIALOG, ORBIT, BRS** e outros, preparam índices em linha para cada base de dados disponibilizada para consulta. Os produtores de bases de dados em **CD-ROM** também participam do processo de indexação das bases fornecidas com suas interfaces de busca, oferecendo índices constituídos pelos próprios registros das bases de dados.

ANDERSON e WILSON (1983) **confirmaram** que os índices assim constituídos frequentemente oferecem as seguintes possibilidades:

- a) habilidade para busca de informação num campo específico;
- b) habilidade para busca em todos os campos de assuntos simultaneamente;
- c) habilidade para evitar referências de campos não desejados;
- d) políticas de análise gramatical para cada campo visando a maximização da capacidade de busca;
- e) consistência no tratamento de cada campo, incluindo apresentação dos registros em formatos **pré-definidos**;
- f) facilidades para busca precisa de **subcampos**;
- g) consistência no tratamento de campos similares;
- h) habilidade para pesquisar simultaneamente o equivalente a múltiplos volumes impressos de diversos anos anteriores.

Por outro lado, a maneira pela qual um banco de dados e/ou um produtor de bases em CD-ROM formata os vários elementos de dados para a realização da busca pode, frequentemente, oferecer diferentes resultados. Assim, um **determinado** tema executado na mesma base, em CD-ROM ou em linha, ou, ainda, em bancos de dados distintos vai produzir resultados finais diferenciados.

KNAPP, COHEN e JUEDES (1998) salientaram que a pesquisa em bases de dados apresenta dificuldades especiais para a área de humanidades, porque os assuntos podem ser abordados de inúmeras maneiras sendo que vários sinónimos podem ser usados para descrever um só conceito e os termos podem apresentar variações relativas à precisão. Apontaram como causa significativa na baixa recuperação de **informação**, a inabilidade dos intermediários e/ou dos usuários finais em selecionar todos os possíveis termos que os autores utilizaram nos seus **trabalhos**, o que já tinha sido observado por LANCASTER e FAYEN (1973). Apresentaram como proposta, os resultados de um estudo utilizando estratégias de busca em LC e LN, no qual ficou demonstrado que a combinação dos dois modelos de estratégia usadas em conjunto, ofereceu maior recuperação. Sugeriram que instrumentos de auxílio para o planejamento de estratégias devem ser desenvolvidos para apoio tanto aos **intermediários**, quanto ao usuário final.

Num estudo realizado na área de mecânica dos solos, MUDDAMALLE (1998) demonstrou que o uso concomitante da LC e da LN nas estratégias de busca apresentou melhor desempenho na recuperação do que as estratégias que foram realizadas com cada tipo de linguagem individualmente. Recomendou que a fim de serem obtidos resultados **satisfatórios** na

recuperação, uma **combinação** na estratégia, de termos da LC e da LN deve ser adotada. Salientou que:

"a LC e a LN não podem mais ser tratadas como técnicas de busca **separadas**, mas devem sempre ser tratadas em conjunto, como uma combinação ideal para ampliar os resultados das buscas de informação." (Muddamalle, 1998, p. 887)

Com o **objetivo** de identificar as similaridades semânticas no processo de recuperação da informação, QIN (2000) escolheu um assunto específico na área de saúde: resistência à antibióticos. Para essa investigação foram definidas duas bases de dados: uma em linguagem controlada e outra em linguagem natural. A que utiliza apenas a LN, enriquecida com palavras dos títulos das citações referenciadas nos artigos, foi a *Science Citation Index*. A base selecionada com a LC, foi a MEDLINE que indexa seus documentos utilizando o *Medical Subject Headings (MESH)*. Os primeiros resultados do teste da estratégia com a LC e a LN sobre o tema escolhido, demonstraram as limitações da indexação da base em LC, além de comprovar quão ampliada pode ser a **dissimilaridade** entre os diferentes métodos de indexação do mesmo documento, fator este que com certeza vai afetar os resultados finais da estratégia de busca.

Os termos que constituem o **vocabulário** de uma linguagem controlada são de dois tipos: termos derivados e termos assinalados. Os primeiros são extraídos dos próprios documentos e os últimos são os termos utilizados para normalização bibliográfica, também denominados, metadados de organização da informação. Os termos derivados descrevem a própria linguagem do autor, isto é, termos da LN e, os assinalados são os termos que constituem os vocabulários controlados estabelecendo as relações hierárquicas entre os componentes da LC. Nas linguagens controladas das bases de dados eles são denominados por: **palavras-chave**, descritores, termos de indexação, códigos de classificação e outros.

SVENONIOUS (2000), afirmou que a LC é derivada da literatura que pretende descrever e portanto, toma-se necessário a delimitação de suas fronteiras. Definiu o vocabulário controlado como:

"uma linguagem normalizada que **seleciona** da linguagem natural um conjunto restrito de palavras e frases, tratando esse conjunto semanticamente para o estabelecimento das relações com outros termos." (Svenonious, 2000, p. 134).

Salientou ainda, que existem dois tipos principais de linguagens controladas: a linguagem de assunto alfabética e a linguagem de assunto **classificatória**, apontando três diferenças entre elas. A primeira utiliza: expressões verbais, **thesaurie** listas autorizadas de termos e ordena os

assuntos alfabeticamente. A segunda, usa: expressões verbais e notações; esquemas de classificação e ordena assuntos sistematicamente, primeiro por disciplina e dentro desta, hierarquicamente por tópicos.

A estrutura da informação para a representação do conteúdo temático dos documentos nas bases de dados em linha, nos respectivos bancos e nos produtos em CD-ROM permite a consulta tanto de **termos** da linguagem natural quanto de termos da linguagem controlada, conforme apresentado na Figura 1.

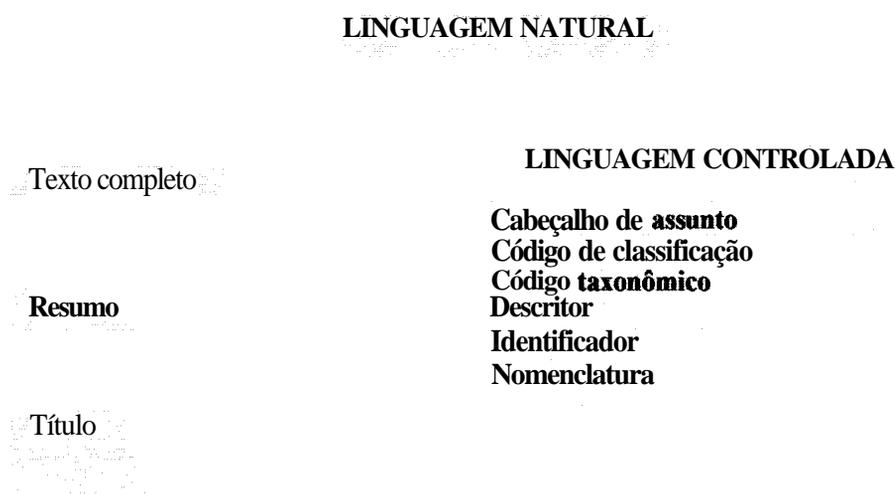


Figura 1 - Estrutura da informação para representação do assunto

Em bases de dados, a representação do conteúdo dos documentos pode estar contida nos campos de busca da LN e da LC; portanto, considerando que esta é um subconjunto daquela, conforme sintetizado na Figura 1, a visualização desses campos permite **inferir** que, em determinados contextos, os termos da LN e da LC podem ser superpostos.

A utilização de termos e/ou conceitos da LN na recuperação da informação está diretamente relacionada com os **termos** contidos no texto completo dos documentos, ou com um conjunto de termos escolhidos livremente das palavras dos títulos ou dos resumos. O uso do vocabulário controlado na busca da informação implica a **seleção** de termos de uma lista autorizada, por exemplo, um *thesaurus*, que fornece os **termos** e/ou conceitos com suas inter-

relações hierárquicas, referências cruzadas, notas de escopo, **qualificadores** de termos e outros recursos. Esses vocabulários frequentemente são disponibilizados em linha ou em CD-ROM, e a sua estrutura hierarquizada pode ser **consultada** com a finalidade de se fazer **seleção** dos termos e/ou conceitos da estratégia de busca. Uma lista de bases de dados e **suas** respectivas linguagens controladas é apresentada no Anexo 4. Nesse rol são mencionadas as bases de dados do DIALOG, cabendo ressaltar que os tipos mais frequentes de **vocabulários** controlados são os **esauri**, seguidos de listas de cabeçalhos de assunto e códigos específicos de **classificação** desenvolvidos pelos próprios produtores das bases de dados.

Vários **sistemas** de recuperação de informação e produtores de bases em **CD-ROM** oferecem o acesso a múltiplas bases de dados e para cada área do conhecimento humano, várias bases estão disponibilizadas no contexto da indústria da informação. Entretanto, cada base utiliza o seu próprio vocabulário controlado; uma vez que não existe padronização ou compatibilidade entre esses vocabulários, mesmo naqueles que são específicos de um **determinado** assunto, o mesmo conceito pode ser designado por diferentes termos nos distintos vocabulários. **SVENONIOUS** (1986) confirmou essa falta de padronização e alertou para a função do **thesaurus** como instrumento imprescindível na recuperação da informação.

No planejamento de uma estratégia de busca, esses fatores devem ser levados em consideração no momento da seleção dos termos para a estratégia e, também, na escolha das bases de dados que serão pesquisadas.

Thesaurus da Base PSYCLIT - Histórico

No contexto desta pesquisa foi utilizado o *Thesaurus of Psychological Index Terms*, quinta edição, produzido pela *American Psychological Association* em 1988, para levantamento dos termos da linguagem controlada. Numa revisão histórica sobre esse *thesaurus*, WALKER e **MULHOLLAND** (1992), relataram o início do registro da literatura em Psicologia, com o lançamento da publicação: *Psychological Abstracts*, em 1927, na qual o primeiro instrumento de auxílio para a recuperação foi o arranjo dos resumos por classificação de grandes áreas de Psicologia. Adicionalmente, termos de indexação e breves frases descritivas foram sendo acrescentadas, porém nenhum controle dessa terminologia foi estabelecido durante quase 40 anos.

Com o advento do uso dos computadores para o registro de grandes volumes de informação, o cenário dos sistemas de indexação e de recuperação da informação foi revisto. Assim, em 1966, o primeiro volume do *Psychological Abstracts* foi produzido num formato legível por máquina e a base de dados **tomou-se** disponível em linha, a partir de 1967, em grandes bancos de dados. Nesse novo panorama, tornou-se evidente que era necessário o controle do vocabulário utilizado na indexação para **facilitar** a recuperação dos documentos citados na base.

Palavras-chave foram extraídas dos documentos a serem indexados, gerando uma lista de aproximadamente 800 termos. Essa lista, revisada anualmente, foi o início do *Thesaurus of Psychological Index Terms*. Muitos dos termos dessa lista estão registrados em edições atualizadas do Thesaurus com a data **inicial** de **1967**. A comprovação de que era necessário uma **atualização** constante e regular do **thesaurus**, mobilizou a equipe de especialistas no sentido de dar continuidade ao desenvolvimento de um vocabulário controlado para indexação e posterior recuperação dos documentos incluídos na base. Com recursos assegurados por financiamento da *National Science Foundation*, a *American Psychological Association* iniciou o processo de desenvolvimento do *thesaurus* envolvendo, numa determinada fase, mais de mil especialistas.

A primeira etapa foi a **seleção** de termos candidatos para o *thesaurus*, a partir da lista dos 800 termos usados **internamente** para a indexação dos documentos. Em seguida, foram **selecionados** dados da frequência de ocorrência de termos dos títulos e dos resumos, durante um período de 5 anos, fornecendo novos termos extraídos da linguagem natural. **Posteriormente**, foram extraídas frases e termos **pré-coordenados** de uma lista do índice **KWIC (Key-Word-in-Context)**, produzida a partir de 10.000 títulos de **artigos**, livros, dissertações e separatas. Essas três fontes de **extração** de termos proporcionaram cerca de 13.000 termos candidatos, que foram re-examinados por especialistas e padronizados de acordo com as **normas** vigentes para controle **terminológico** e de construção de *thesaurus*.

A primeira edição, em 1974, testada durante dois anos, passou pelo crivo de uma equipe de especialistas que estabeleceu regras de **inter-relações** entre os termos, com definição de critérios de avaliação e adequação dos termos, além de **diretrizes** para inclusão e exclusão dos mesmos. A sexta edição foi publicada em 1992 e os responsáveis pela atualização concordam que, embora os **thesauri** sejam um dos mais úteis instrumentos de indexação e de busca, não se pode esperar que eles sejam um produto completo e acabado. Toma-se imprescindível, portanto, uma análise e revisão contínua do vocabulário, das notas de escopo e da hierarquia estabelecidos em edições anteriores, tendo em vista a necessidade de uma atualização que **reflita** o desenvolvimento da área que o mesmo representa.

A maioria das bases de dados são pesquisadas utilizando-se os termos em **texto-livre** e/ou linguagem natural, a partir dos campos de título, do resumo e do texto completo dos documentos registrados na base. Usando a seleção de termos para a estratégia na LN, o intermediário e/ou usuário **final** pode especificar quais desses termos **devem aparecer** na mesma frase, **no parágrafo**, no campo, no título ou no resumo. Duas técnicas especialmente úteis para a busca em LN são a de **truncagem** de termos e a busca com operadores de proximidade.

A truncagem de termos permite ao intermediário que **operacionaliza** a busca, usar a raiz do termo sem especificar todas as possíveis variações desse termo (prefixos e/ou sufixos). Já a técnica de busca com operadores de proximidade ou de adjacência, permite especificar na estratégia, a posição relativa de dois ou mais termos entre eles próprios. Para as buscas relativas

ao levantamento das **últimas** tecnologias numa **determinada** área, ou a um novo assunto, ou para um novo produto, e ainda, a busca em documentos de patentes, o uso da estratégia em linguagem natural pode ser o melhor caminho para o encontro da informação desejada. A terminologia pode ser muito **atual**, as aplicações ainda não são tão significativas para serem indexadas e, portanto, os novos termos não foram incorporados a nenhuma lista autorizada de LC.

Segundo **LANCASTER** (1986), a LN "não tem atraso, nem vocabulário específico, mas tem a garantia bibliográfica total", pois, os termos são derivados e/ou extraídos **diretamente** dos documentos.

O exame do documento "**Search Aids for use with DIALOG Databases - 1987**", revelou que menos de 30% das bases de dados do DIALOG, **disponibilizam** linguagens **controladas** (ver Anexo 4). Além disso, nem todos os **thesauri** estão disponíveis para consulta em linha, e nem todas as bases em **CD-ROM** possuem **thesauri** para **seleção** dos termos para a estratégia de busca.

HARTER (1986), comparando os problemas de ambiguidade da linguagem natural em áreas especializadas, como Física e Química, e em áreas de Ciências Sociais, como Educação, ou em bases de dados **multidisciplinares**, confirmou o valor da linguagem natural para a pesquisa nas Ciências Exatas e alertou para as dificuldades a serem superadas nas áreas Sociais e Humanas. Complementando essas análises, o autor ressaltou que, na indexação com LC o número de conceitos indexados é relativamente menor quando se comparam com os pontos de acesso permitidos pela LN e que:

"tipicamente, apenas alguns poucos termos são assinalados para representar o conteúdo dos documentos. Entretanto, mesmo a indexação exaustiva como a usada pela *National Library of Medicine* não supera a **exaustividade** proporcionada pela linguagem natural, especialmente aquelas oferecidas pelo texto completo dos documentos." (Harter, 1986, p.54).

As discussões e estudos sobre o uso da LN e da LC na recuperação da informação vêm se estendendo há bastante tempo e a solução ainda não foi encontrada. O desenvolvimento de linguagens controladas para indexação e recuperação da informação requer trabalho em equipe, altos investimentos e tempo para gerar, testar e sedimentar um instrumento de trabalho que deve ser continuamente atualizado. Com o foco da busca centrado no usuário final das bases em CD-ROM e dos OPAC's, **reconhece-se** que ele tem excessiva dificuldade em formular estratégias de busca apropriadas e, portanto, não se beneficia do uso das linguagens controladas, ainda que as mesmas estejam documentadas e associadas com as bases de dados.

Por outro lado, a recuperação de informação **envolve**, na prática, uma mistura de LN e LC para a busca em grande variedade de diferentes tipos de bases de dados, incluindo as de texto integral e as bibliográficas, além de outros sistemas baseados em CD-ROM e os OPAC's.

HARTER (1986), ROWLEY (1994), LANCASTER (1993) e outros estudiosos do assunto sugerem que a prática deve nortear a decisão da escolha dos **termos** da LN e da LC. Muitos intermediários e/ou usuários finais usam, ambas as linguagens no momento de formulação da estratégia de busca, principalmente porque na maioria das bases de dados é possível a busca simultânea por campos de busca que podem ser combinados entre si. Assim, os campos de resumo, de títulos, de identificadores, de descritores ou cabeçalhos de assunto e de códigos de **classificação** podem ser amplamente utilizados visando à obtenção de um resultado mais satisfatório, independentemente da verificação, no momento de operação da busca, de qual dessas linguagens terá um melhor desempenho. O foco, portanto, está na obtenção de resultados satisfatórios e não, no instrumento utilizado para alcançar esses resultados.

2.3 Considerações sobre a revisão de literatura

Inúmeros estudos sobre o uso da LN e da LC na recuperação da informação têm se concentrado na possível utilização das duas linguagens na estratégia de busca simultaneamente. Tal fato demonstra que o uso de uma ou outra ainda não foi definido, nem por experimentos, nem por consenso.

Observou-se, também, a predominância de investigações que descrevem as vantagens e desvantagens do uso da LN e da LC nas estratégias de busca em bases de dados. Essa observação **reflete** a ampla difusão do tema, sem contudo apontar especificamente a melhor **linguagem**, pois vários outros fatores intervenientes afetam os resultados finais das buscas em bases de dados.

Por outro lado, muitos autores desenvolveram estudos abordando as técnicas de planejamento das estratégias de busca, suas etapas e **operacionalização** sem focar, entretanto, as implicações relacionadas com as próprias **linguagens** de recuperação do banco de dados onde as bases estão hospedadas, deixando de mencionar a estrutura de informação de cada base num determinado sistema de recuperação de informação.

Destacam-se, ainda, os estudos comparativos sobre o desempenho dos **intermediários** e dos usuários finais em situações reais de buscas interativas em linha, visando à identificação do processo de **seleção** dos termos para a implementação das estratégias de busca, além de identificar em que momento essa seleção se efetua e se os resultados finais satisfazem os requisitos solicitados.

Nos estudos sobre uma das fases que antecedem a elaboração das estratégias de buscas, que é o processo de entrevista preliminar com o usuário, observou-se uma forte tendência para a utilização das novas tecnologias de informação acopladas a essa fase de negociação do pedido de busca.

As dificuldades comuns, identificadas por **ROWLEY (1990, 1994)**, **SVENONIOUS (1976, 1986, 2000)**, **HARTER (1986)**, **LANCASTER (1979, 1986, 1993)** e **FIDEL (1987, 1991)**, dentre outros, relacionadas com as regras de seleção dos termos para a estratégia de busca, indicam que a prática deve nortear a decisão final; porém, o debate sobre o uso da LC ou da LN permanece sendo questionado até os dias atuais.

Considera-se que a complexidade do problema deve proporcionar novos estudos que venham a **contribuir** para o **aprimoramento** do processo de decisão relativo à seleção de termos que vão compor as estratégias de busca em bases de dados, seja em **CD-ROM** ou em linha, seja nos catálogos públicos de acesso em linha, que é o objetivo dessa investigação.

3. METODOLOGIA

3.1 Introdução

Os estudos **tradicionais** sobre o uso dos **vocabulários** controlados e dos **termos** livres na elaboração de estratégias de busca para acesso às bases de dados em linha ou em CD-ROM, aos catálogos em linha de acesso público e em sistemas de recuperação da informação revelam a inexistência de soluções que possam ser adotadas de imediato. Apesar do significativo avanço nas tecnologias de armazenamento e recuperação de informações, ainda é um tema controverso. Autores como **ROWLEY** (1993) e **SVENONIOUS** (1986), dividem esses estudos em "**fases**" ou "eras" de debates sobre o uso de linguagens de indexação/recuperação natural **versus** linguagens de indexação/recuperação controlada. Assim, a razão para esta pesquisa é proceder ao levantamento do uso das linguagens natural e controlada na formulação das estratégias de busca, visando identificar em qual delas **pode-se** obter um melhor desempenho na recuperação da **informação**.

3.2 Escolha da Metodologia

A natureza desta pesquisa é o estudo de caso, e foi usada a abordagem qualitativa para analisar o contexto no qual os usuários e/ou intermediários tomam decisões sobre qual o vocabulário será utilizado durante a formulação da estratégia de busca. A natureza do problema e os **objetivos** da pesquisa apontaram naturalmente o uso de métodos de pesquisa qualitativa como os mais adequados para prover meios sobre a compreensão das práticas utilizadas no planejamento das estratégias de busca.

Os **intermediários** da busca, normalmente, estabelecem processos associativos para identificar quais fontes de informação vão atender a determinada necessidade de informação de um **usuário**. Portanto, essas avaliações pessoais sobre a **seleção da(s)** base (s) de dados, em qual banco de dados, **atualização** dos itens de **informação**, custos de conexão e de impressão dos resultados, cobertura de assunto e período de abrangência, flexibilidade da linguagem de recuperação e outras questões, contribuem para a tomada de decisão final sobre o planejamento e a **operacionalização** da estratégia de busca.

Cabe destacar que os resultados da busca de informações estão **diretamente** relacionados com a linguagem escolhida pelo intermediário para a formulação da estratégia de busca. O ponto de vista do usuário sobre a relevância dos itens recuperados foi avaliado mediante o uso de um questionário e de um formulário de avaliação.

3.3. Métodos e técnicas

Os procedimentos usados na elaboração da estratégia de busca **utilizando-se** a linguagem controlada, a linguagem natural ou o uso simultâneo das duas, foram pesquisados na base de dados PSYCLIT, produzida pela American **Psychological Association** (APA), em CD-ROM, **levando-se** em conta os seguintes elementos:

1. Linguagem controlada (LC)

Campos: descritor (DE): mais relevante (**MAJ**) e menos relevante (MIN); e códigos de classificação de assunto (CC), todos atribuídos pela APA.

2. Linguagem natural (LN)

Campos: título (TI), **frase-de-busca** (KP) e resumo (AB)

3. Uso simultâneo da LC e LN

Campos: conjunto dos campos da LC e LN

Na base de dados PSYCLIT, cada registro é constituído em princípio, pelos campos citados na Tabela 4.

Tabela 4 - Base de dados PSYCLIT: campos

CAMPOS	DESCRIÇÃO
NA	Número de acesso da APA, correspondendo ao ano e ao número do resumo na versão impressa do <i>Psychological Abstracts</i>
TI	Título
AU	Autor
AF	Afiliação do primeiro autor referenciado
JC	Código do periódico
PY	Ano de publicação do documento
IB	Código para identificação de livros
IS	Código do ISSN
JN	Indica o volume e fascículo no <i>Psychological Abstracts</i>
LA	Indica o idioma de origem do documento
DT	Tipo de documento indexado na base
AG	Indica agrupamentos de pessoas por idade
AB	Resumo do artigo
DE	Descritor, subdividido em MAJ e MIN
KP	Frase-de-busca (termos extraídos da linguagem natural)
TC	Sumário (dos livros)
CC	Códigos de classificação de assunto, atribuído pela APA

A base, cuja interface é produzida pela **SilverPlatter**, dispõe de um **software** de recuperação de informação extremamente elaborado, oferecendo as seguintes possibilidades:

1. recuperação de termos simples ou compostos de todos os campos de busca de um registro,
2. recuperação dos radicais etimológicos usando-se o recurso de **truncagem** à direita;
3. recuperação de termos compostos com o uso dos operadores de proximidade e adjacência;
4. recuperação por descritores, utilizando-se a linguagem controlada da base, inclusa no próprio CD-ROM ;
5. uso dos operadores **booleanos** AND, OR, NOT, para combinação entre os conjuntos formulados na estratégia de busca;
6. combinação de termos contidos no mesmo campo de busca, com o **operador WITH**;
7. refinamento por campos específicos, com o operador IN;
8. limitação dos resultados da busca por ano de **publicação**, faixa etária, idioma e outros segmentos;
9. **seleção** de termos para a formulação da estratégia de busca, a partir do próprio INDEX ou THESAURUS contido na própria base;
10. refinamento dos resultados, utilizando-se códigos de classificação (CC), definidos pelo produtor da base.

A oferta de todas as possibilidades citadas anteriormente, permitiu a elaboração da estratégia de busca por diversas rotas: uso da **linguagem** controlada, da linguagem natural, uso simultâneo das duas, ampliação ou redução do resultado da pesquisa, uso de vários processos para seleção de termos e ainda, possível utilização de várias técnicas de **formulação** de estratégias de busca.

Na figura 2 é apresentado um exemplo de registro da base PSYCLIT com a contribuição dos vocabulários controlados e da linguagem natural.

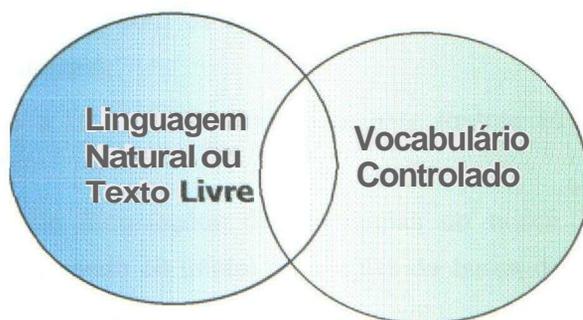


Figura 2 - LC e LN: registro na Base PSYCLIT

O exemplo de um registro da base PSYCLIT é apresentado com o número médio de termos de indexação para demonstrar a contribuição do vocabulário controlado e da linguagem natural num item de informação. Os termos do vocabulário controlado foram obtidos dos seguintes

campos: descritor (21), **cabeçalhos de seção** de assuntos (3) e códigos de classificação de assuntos (4). Os termos da linguagem natural foram obtidos dos seguintes campos: título (19), identificadores (1) e resumo (**115**). A combinação desses dois elementos forneceu uma **interseção** de 5 termos obtidos da LN e 7 termos da LC.

Para os propósitos deste trabalho de pesquisa, foram considerados na estratégia de busca os campos que permitem identificar o uso da LN e da LC separadamente.

3.4 Amostra e população

O tamanho **correto** de uma amostra segundo **afirmativa** de BAILEY (1978) "depende da natureza do universo a ser investigado e do propósito da **pesquisa**" ..."**e** que... cerca de 30 casos pode ser considerado como um número mínimo indispensável em pesquisas, nas quais a análise de dados estatísticos deva ser efetuada, embora algumas técnicas possam ser utilizadas com menos de 30 **casos**."

Com base nessa orientação, a presente pesquisa foi constituída por três conjuntos de 10 (dez) buscas solicitadas por usuários da **UCB/SIBI**, na área de Psicologia, registradas em formulário próprio do Serviço de Busca Bibliográfica da UCB/SIBI (Anexo 1), as quais foram testadas de forma que, para cada um dos conjuntos de 10 buscas, a estratégia fosse efetuada em três etapas, a saber

- a) estratégia **utilizando** a LC da base de dados, com os **termos/conceitos selecionados** pelo intermediário, com a **participação** do usuário (**10** formulações de estratégias);
- b) estratégia utilizando a **LN**, com os **termos/conceitos** determinados pelo usuário (outras 10 formulações de estratégias);
- c) estratégia utilizando a LC e a LN simultaneamente (outras 10 formulações de estratégias distintas).

Essas etapas de formulações de estratégias de busca resultaram na composição definitiva da amostra, totalizando 30 (trinta) estratégias de busca diferenciadas umas das outras, pelo tipo de vocabulário que foi empregado.

É uma amostra **não-probabilística**, do tipo acidental, conforme caracterização de RUDIO (1990) e contemplou a demanda de buscas na área de Psicologia da UCB. A população total dessa amostra, foi constituída por professores e alunos dos cursos de Graduação e Pós-Graduação em Psicologia sendo composta por cerca de 400 indivíduos.

Por motivos de economia, foram considerados os resultados de busca limitados pelos últimos dois **anos**, de janeiro de 1998 até abril de 2000, e com a impressão de, pelo menos, 30 (trinta) referências por busca, totalizando em princípio, 900 referências que foram avaliadas e medidas em relação à **revocação** e à precisão.

Os resultados de cada tipo de estratégia foram então remetidos aos solicitantes acompanhados num primeiro momento, do **Formulário** de Avaliação de Referências (Anexo 3) e, num segundo momento, do Formulário de Avaliação do Serviço de Busca Bibliográfica - Questionário (Anexo 2). O solicitante da busca analisou os resultados, **registrando-os** nos respectivos Formulários de Avaliação, **devolvendo-os** para a continuidade desta pesquisa. Cada item foi identificado pelo solicitante como MB (Muito bom), B (Bom) e I (**Irrelevante**). O Formulário de Avaliação - Questionário foi aplicado pelo investigador em outro momento posterior à pesquisa, para validação dos resultados aferidos no Formulário de Avaliação de Referências (Anexo 3).

Após a avaliação realizada pelos usuários, foram tabulados os dados para cada uma das 3 (três) etapas da estratégia de busca, **objetivando** a identificação do tipo da estratégia que obteve o melhor desempenho e para a **mensuração** dos **fatores** de revocação e precisão dos **resultados**. Assim, as medidas de revocação e precisão que foram obtidas no processo de avaliação dos resultados pelos usuários foram utilizadas para demonstrar se a melhor formulação de estratégia de busca é a que utiliza:

- apenas a linguagem controlada?
- apenas a linguagem natural?
- simultaneamente as duas linguagens?

Os resultados foram devidamente aferidos em relação às hipóteses da presente pesquisa objetivando sua comprovação.

Apresenta-se na Tabela 5, um quadro contendo os objetivos dessa investigação **correlacionando-os** com as respectivas hipóteses de trabalho.

Tabela 5 - Objetivos *versus* hipóteses

OBJETIVOS ESPECÍFICOS	HIPÓTESES		
	H1	H2	H3
1. Identificar a adequação de uso do vocabulário controlado na elaboração da estratégia de busca.	X		
2. Identificar a adequação de uso da linguagem natural na elaboração da estratégia de busca.		X	
3. Identificar a adequação de uso simultâneo do vocabulário controlado e da linguagem natural na elaboração da estratégia de busca.			X
4. Identificar qual a linguagem mais adequada para a operacionalização da estratégia de busca.	X	X	X

3.5 Coleta de dados

A **coleta** de dados foi efetuada junto aos professores e alunos da Universidade Católica de Brasília (UCB), que oferece cursos de Psicologia nos níveis de: graduação, com cerca de 300 alunos, **pós-graduação lato sensu**, com cerca de 55 alunos e os de **stricto sensu**, com aproximadamente 25 alunos. São 15 professores com titulação de doutores e mestres, e 380 alunos, totalizando 400 indivíduos.

A demanda por informações em Psicologia para os projetos de pesquisa desenvolvidos na universidade é atendida pelo Sistema de Bibliotecas, cujo Serviço de Pesquisa Bibliográfica utiliza as mais modernas tecnologias para o atendimento aos seus **usuários**, incluindo a base de dados PSYCLIT em **CD-ROM**; o **site** da *American Psychological Association*; o sistema PROQUEST com mais de 2000 títulos de periódicos indexados, sendo 40% dos títulos com texto integral, produzido pela *University of Microfilms*, contendo uma **coleção** de periódicos virtuais retrospectivos com uma média de cobertura de cerca de cinco anos; a base de dados **ERIC**, que indexa documentos na área de Educação e áreas correlatas, incluindo Psicologia, e outras bases que cobrem as áreas de Ciências Humanas, Ciências Sociais e Ciências da Saúde.

Cada **solicitação** de busca bibliográfica **dirigida** ao Sistema de Bibliotecas é feita em um formulário próprio (Anexo 1), **objetivando** a obtenção de todos os detalhes que **caracterizam** o tema a ser pesquisado: um título geral, descrição detalhada, **fornecimento** de **palavras-chave/termos** técnicos em inglês, limitações por tipo de idioma, por anos de pesquisa e outras. Cada busca registrada no serviço foi processada após a **identificação** do conjunto de termos e/ou conceitos que foram utilizados na formulação das estratégias em: LC, em LN e **utilizando-se** simultaneamente a LC e a LN.

Cada solicitação de um usuário gerou três estratégias de busca diferentes, as quais foram utilizadas para a comprovação das hipóteses de trabalho.

3.6 Análise dos dados

Por motivos de economia, pretendeu-se executar as buscas na base de dados PSYCLIT, em CD-ROM, limitando-se os resultados aos últimos dois anos, de 1998 à 2000, e com a impressão de até 30 referências por cada formulação de estratégia de busca, sendo que as mesmas foram impressas conforme definido pela linguagem de recuperação, isto é, a apresentação das mesmas obedeceu a uma ordem cronológica decrescente por ano de publicação. Foi assegurado, portanto, ao usuário, um resultado recente e **atualizado** sobre o tema da pesquisa.

Os resultados obtidos nas buscas bibliográficas foram submetidos aos solicitantes acompanhados de um Formulário de Avaliação das Referências, onde os usuários registraram o grau de satisfação encontrado. Após essa etapa, um Formulário de Avaliação do Serviço de Busca - Questionário (Anexo 2) foi aplicado pelo investigador para validação dos dados anteriores.

A análise dos dados foi efetuada **tomando-se** como base a classificação determinada pelo usuário de cada busca, e separadamente, por cada tipo de estratégia executada. Assim, os resultados foram avaliados pelos usuários em três momentos distintos:

- 1) primeiro momento: a resposta com o uso de termos e/ou conceitos extraídos da LC (***Thesaurus of Psychological Index Terms***) e determinado pelo **intermediário** da busca, com a **participação** do usuário;
- 2) segundo momento: a resposta quando da utilização de termos e/ou conceitos em LN fornecidos pelos usuários;
- 3) **terceiro** momento: a resposta com o uso simultâneo da LC e LN, também determinado pelo intermediário da busca.

Essas **avaliações** segundo o tipo de linguagem **utilizada** na estratégia de busca, geraram, via Excell, arquivos de dados para a criação de tabelas, que permitiram a análise sobre o vocabulário que melhor desempenho obteve na formulação da estratégia de busca.

Os resultados de cada estratégia de busca foram codificados e tabelados pelos temas de busca, e pelo tipo de linguagem utilizada na estratégia: LC, LN e LC+LN. As tabelas foram ordenadas por número de referências recuperadas, número de referências relevantes (*MB*), número de referências periféricas (B) e pelo número de referências irrelevantes (I). As tabelas apresentaram, também, o somatório do resultado de todas as estratégias realizadas para cada busca bibliográfica, além dos dados referentes ao grau de satisfação dos **usuários**. Assim, com os dados devidamente computados, foi possível uma análise de desempenho de cada uma das estratégias de busca utilizadas, para a comprovação ou não, das hipóteses formuladas na presente pesquisa.

3.7 Limitações da pesquisa

Essa pesquisa foi realizada especificamente na área de Psicologia, utilizando a base de dados em CD-ROM: *Psychological Abstracts Literature (PSYCLIT)*, produzida pela SilverPlatter, no período de janeiro de 1998 até março de 2000.

Buscando identificar qual a linguagem mais adequada para o planejamento da estratégia de busca foram realizadas buscas nas seguintes modalidades: com a LC, com a LN e com a LC+LN, utilizando-se os recursos de **software** do produtor, apenas no que se refere aos campos de busca que correspondessem as especificações das linguagens utilizadas nas estratégias.

Foi dada ênfase na aferição dos resultados realizados pelos usuários, por intermédio de formulários apropriados para avaliação do processo de planejamento das estratégias de busca, visando a identificação da linguagem de melhor desempenho na recuperação da informação.

4. ANÁLISE DOS DADOS

4.1 Introdução

Os dados desse estudo foram reunidos a partir das estratégias de busca **operacionalizadas** na base de dados Psyclit em CD-ROM e, dos formulários de **avaliação** das referências (Anexo III) respondidos pelos usuários do Serviço de Busca Bibliográfica da **UCB/SIBI**, na área de Psicologia. Os dados relativos ao grau de satisfação dos usuários foram computados separadamente por tipo de estratégia utilizada com a linguagem controlada, com a linguagem natural e com a aplicação das duas simultaneamente e os respectivos resultados obtidos. Um formulário de avaliação de pesquisa **bibliográfica-questionário** (anexo II) foi aplicado aos usuários após a primeira fase de avaliação individual das referências, objetivando a validação das respostas fornecidas pelos mesmos.

O presente trabalho procurou examinar o efeito do uso da linguagem controlada, da linguagem natural e a combinação de ambas simultaneamente, no planejamento das estratégias de busca na base de dados em CD-ROM de Psicologia, produzida pela American Psychological Association - APA, no período de janeiro de 1998 até março de 2000.

4.2 Planejamento da estratégia de busca.

O **objetivo** geral desse estudo foi analisar o processo de planejamento da estratégia de busca num ambiente de bases de dados em CD-ROM, verificando qual o tipo de vocabulário deve ser empregado no planejamento da mesma.

Em artigos publicados em **1981**, 1988 e 1990, BATES afirmou que a estratégia de busca envolve um planejamento completo de uma busca de informação, englobando uma ou mais **táticas**, além da própria formulação da busca, que são as frases estruturadas que expressam a pergunta dos solicitantes. Essas frases compostas por termos **e/ou** conceitos da LC e/ou da LN, devem levar em consideração, a natureza do pedido de busca, a estrutura da informação da base e a linguagem de busca, ambas variáveis intervenientes no processo de recuperação da informação.

LANCASTER (1979, 1986) relacionou três critérios de avaliação que identificam como um sistema de **informação** ou uma base de dados, deve responder às necessidades de seus usuários.

Os critérios eram: qualidade dos resultados, tempo de resposta e custo. Nesse estudo foi avaliada a eficácia da **estratégia de busca pela qualidade dos resultados da busca obtido e posteriormente avaliados** pelos usuários.

Para avaliar as referências bibliográficas obtidas em cada modalidade de estratégia de busca, a saber: LC, LN e LC+LN, os usuários foram orientados a aplicar três classificações codificadas como segue:

- Muito bom (MB) - citações altamente relevantes
- Bom (B) - citações periféricas
- Irrelevante - citações irrelevantes

Essas categorias foram **deliberadamente** escolhidas por sua facilidade relativa de uso por parte dos usuários. Assim, os documentos altamente relevantes e os irrelevantes são rapidamente assinalados. A categoria de citações periféricas necessita uma análise mais minuciosa por parte do usuário porque o assunto de interesse é abordado, porém, de forma mais genérica, podendo ser J numa etapa posterior da pesquisa do solicitante.

Em um artigo publicado em 1991, **KIRKBRIDE** alertou para as dificuldades inerentes ao processo de escolha dos termos para a estratégia de busca, salientando que o planejamento da estratégia depende do tema a ser pesquisado e da linguagem de busca dos diversos sistemas de recuperação de informação onde a base de dados estiver hospedada. Essa observação, confirma o presente estudo **devendo-se** acrescentar ainda, dentre outros, a estrutura da informação da base de dados e a sua respectiva linguagem de indexação.

TENOPIR (1987), NICHOLLS e HOLTSMANN (1989) concordam que a discussão sobre o uso da LC, e da LN na estratégia de busca depende de pelo menos quatro pontos, a saber: a) da base de dados; b) do vocabulário usado para indexação e da política de atualização desse vocabulário; c) do tema de busca a ser pesquisado e d) do nível de demanda do solicitante, complementando dessa forma as afirmações de **Kirkbride**.

Apresenta-se na Figura 3, uma comparação do total de referências recuperadas em cada índice: Muito bom, Bom e Irrelevante.

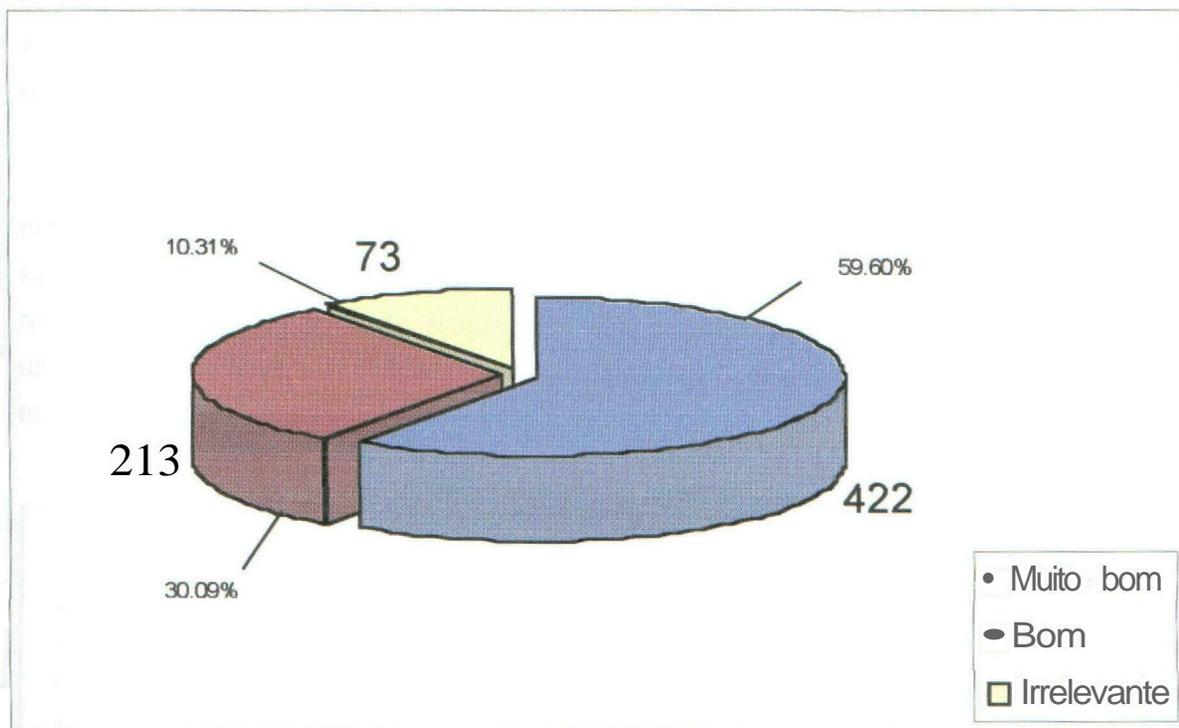


Figura 3 - Quadro demonstrativo das referências recuperadas

As três modalidades de estratégia de busca utilizadas ofereceram num cômputo geral, 708 referências bibliográficas sobre todos os temas solicitados, sendo que dessas, 422 foram classificadas como **Muito bom**, correspondendo a 59,60% do total. Com o critério definido como Bom foram recuperadas 213 referências, correspondendo a 30,09% do total alcançado. O conjunto de citações classificadas como **Irrelevantes** atingiu um total de 73 referências, correspondendo a **10,31%** do total geral recuperado.

Os dados consolidados pela avaliação dos usuários e obtidos no formulário de avaliação de pesquisa *bibliográfica-questionário* (anexo II) revelaram um índice de satisfação por parte dos mesmos, entre **81%** e 100% sobre os resultados das buscas, validando as respostas **obtidas** com a avaliação individual das referências. O **formulário** oferecia uma pergunta com cinco variações de percentuais para a avaliação consolidada das **referências** recuperadas com as três modalidades de estratégias de busca. Todos os usuários participantes do presente estudo, preencheram os

formulários demonstrando uma **aceitação** geral para com os **resultados** apresentados pelo Serviço de Busca Bibliográfica da UCB/SIBI.

SU (1996) num estudo sobre as medidas de avaliação de desempenho de bases de dados sugeriu que o valor como um todo, dos resultados encontrados nas buscas é um indicador excelente de avaliação. Afirmou que esse valor é uma medida fundamentada na resposta dos usuários sobre a utilidade dos resultados das buscas em sua totalidade.

NICHOLLS e HOLTMANN (1989) investigando o uso da LC e da LN nas estratégias de busca numa base em CD-ROWI, demonstraram que os dois enfoques ainda são muito polêmicos no que se refere aos fatores relacionados com a **revocação** e a precisão, o que foi observado com os resultados obtidos no presente estudo, que não proporcionaram na média, uma significativa diferença no quantitativo de citações relevantes tanto da LC, quanto da LN, conforme demonstrado na Figura 4.

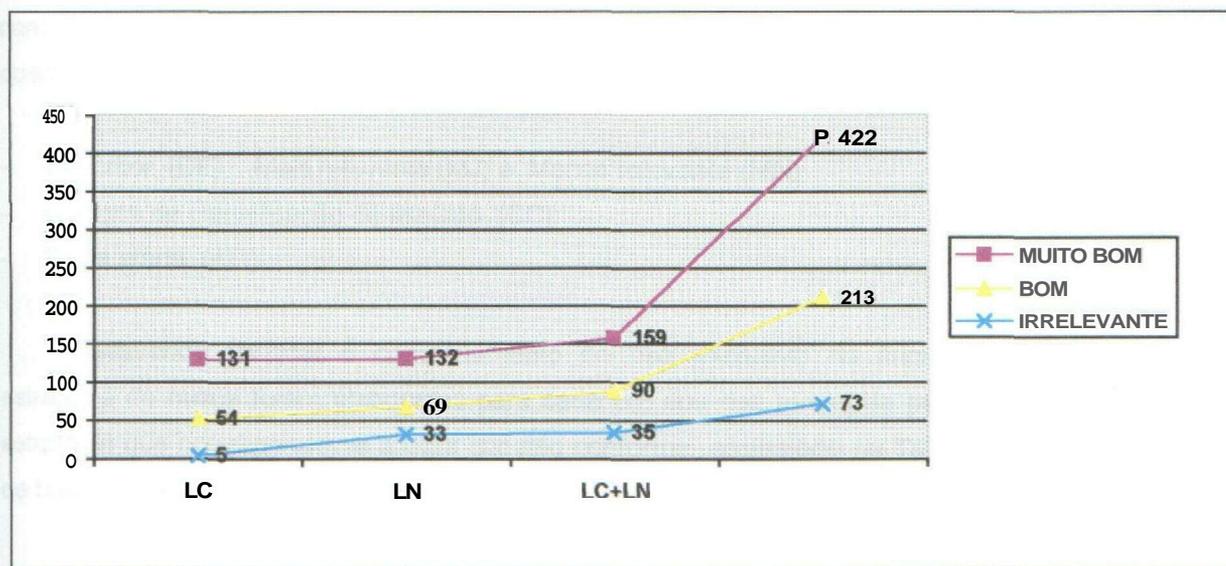


Figura 4 - Índice comparativo das estratégias

O índice comparativo das três modalidades de estratégias de busca utilizadas no presente estudo, revelou que não houve significativa diferença na recuperação da informação para o índice Muito bom, tanto da LC quanto da LN como demonstrado na Figura 4, quando apenas 131 referências para a LC e 132 para a LN, foram classificadas pelos usuários como citações relevantes.

Analisando os instrumentos utilizados para a indexação, num contexto de bases de dados, SVENONIOUS (1988), recomendou o desenvolvimento de vocabulários controlados voltados, porém, para a recuperação da informação, salientando as funções de controle **terminológico** dos mesmos, no que concordamos, tendo em vista especialmente o uso das bases de dados com seu pleno potencial de recuperação de informação.

4.2.1 Estratégia de busca utilizando linguagem controlada - LC.

As linguagens controladas, também chamadas de linguagens documentárias apresentam certos pontos fortes como por exemplo: o controle de sinónimos e **sinónimos** próximos levando a termos preferenciais; proporciona notas de escopo; qualifica os homógrafos; mostra termos mais abrangentes, mais específicos e relacionados; supera problemas de sintaxe com termos compostos e outros dispositivos. No Anexo IV é apresentada uma relação de linguagens controladas com suas respectivas bases de dados.

A estrutura da informação para busca na LC da base de dados PSYCLIT e a **formata dos** campos de identificação do conteúdo temático dos documentos, apresenta a seguinte **configuração** que foi utilizada no planejamento de cada estratégia de busca:

- Descritor (**DE**) : Mais relevante (MJ) e Menos relevante (MN)
- Códigos de classificação de assunto (CC)
- Faixa etária (AG)

Objetivando verificar o comportamento do uso exclusivo da linguagem controlada na estratégia de busca foram elaboradas para cada um dos dez temas da pesquisa da amostra, a estratégia que correspondesse a essa questão conforme assinalado na Tabela 6. As estratégias de busca detalhadas estão no anexo V.

Tabela 6 - LC: temas de busca *versus* estratégia

TEMAS	ESTRATÉGIA DE BUSCA: TERMOS USADOS
1. Fenomenologia de distúrbios afetivos	Phenomenology ; affective disturbances; obsessive-compulsive-neurosis
2. Doenças renais: aspectos psicológicos	Kidney-diseases ; dialysis; anxiety; major-depression
3. Denegação e religiosidade	Denial; Defense-mechanisms; rational* ; intellectual* ; sublimat* ; relig*
4. Agressividade na infância e adolescência	Aggressive-behavior; AG=childhood; AG=adolescence
5. Divórcio na meia idade: crises e conflito	Divorce; divorced-persons; AG=middle-age
6. Casamento e satisfação conjugal	Marriage; marital satisfaction
7. Dependência química e função familiar	Drug addiction; parental; family; attitude* ; role* ; function* relation* ; permissiveness* ; structur*
8. Adoção de crianças: problemas de comportamento	Adopt* child* ; behav* ; problem* ; disorder* ; change*
9. Obesidade: aspectos psicológicos	Obesity; self-esteem; self-concept ; self-confidence; emotional adjustment; emotional-control; body-image
10. Depressão e stress na meia idade	Depression,-emotion; stress; AG=middle-age

Os termos **e/ou** conceitos utilizados nas estratégias de busca da LC foram identificados no *Thesaurus of Psychological Index Terms*, 5.ed. produzido pela APA, e contido na versão comercializada pela **SilverPlatter**. Para cada tema de pesquisa, os termos foram selecionados do próprio *thesaurus* da base, que apresenta duas **seções** principais: a lista de termos permutados e a lista de termos detalhados.

A primeira **seção** apresenta uma lista alfabética de todos os termos do vocabulário controlado, denominados descritores e que são usados para a indexação dos documentos na base. A lista de termos detalhados contem: termos específicos, termos relacionados, termos gerais, notas de escopo e remissivas. Nessa lista pode ser usada a função "explode" que busca automaticamente o descritor solicitado para a estratégia de busca e todos os seus termos específicos. Essa função foi utilizada em algumas das estratégias de busca, minimizando a entrada de dados na **operacionalização** da estratégia.

Segundo **LANCASTER** (1993), o vocabulário controlado é usado para solucionar especialmente as duas questões a seguir: a) para assegurar, tanto quanto possível, uma representação consistente do conteúdo de assunto dos documentos, tanto no que se refere à indexação, quanto no que se refere à busca; b) para facilitar a execução das buscas nas bases de dados, organizando os termos e/ou conceitos hierarquicamente.

O vocabulário controlado pode ser considerado como um instrumento de auxílio quase imprescindível em alguns casos especiais de busca, sugerindo e orientando o intermediário **e/ou** usuário final, na **seleção** dos termos para a formulação da estratégia de busca. Ele oferece

padronização para: a seleção dos termos de indexação e/ou busca; para uso dos termos; para a interpretação da abrangência dos mesmos, com referência ao seu escopo, além da data em que teve início na base, o respectivo uso desses termos, na própria linguagem controlada.

Os resultados alcançados com as estratégias de busca utilizando termos apenas da linguagem controlada são apresentados na Tabela 7.

Tabela 7 - Linguagem controlada - referências recuperadas

LINGUAGEM CONTROLADA - REFERÊNCIAS RECUPERADAS				
PESQUISAS	MUITO BOM	BOM	IRRELEVANTE	TOTAL
1 Fenomenologia	06	07	-	13
2 Doenças renais	11	05	-	16
3 Denegação	08	-	-	08
4 Agressividade	20	09	01	30
5 Divórcio	21	08	01	30
6 Casamento	15	08	01	24
7 Dependência química	10	01	-	11
8 Adoção de crianças	11	03	01	15
9 Obesidade	16	05	-	21
10 Depressão e stress	13	08	01	22
TOTAL	131	54	05	190

O coeficiente de precisão, que é a relação entre o total de itens recuperados e o total dos itens considerados úteis e classificados pelos usuários como Muito bom foi de 68.95%. Das 190 referências recuperadas com o uso exclusivo da LC, 131 foram consideradas citações relevantes, isto é, **exatamente** de acordo com o interesse dos usuários. Desse grupo de referências obtidas com o uso restrito da LO, 54 citações, ou seja, 28.42% do total, foram classificadas como Bom pelos usuários, que as consideraram citações periféricas, isto é, citações mais genéricas do que o interesse **atual** dos mesmos podendo vir a ser úteis em futuro próximo. Apenas 5 referências, ou seja, 2.63% do total, foram classificadas como irrelevantes. Esses resultados denotam um alto índice de citações relevantes com o uso exclusivo da LC, na estratégia de busca, em relação às citações classificadas como boas e mesmo as obtidas como irrelevantes.

Um estudo de BOYCE e McLAIN (1989), salientou que o uso exclusivo dos termos da linguagem controlada numa estratégia de busca aumenta a precisão nos resultados da busca confirmando o resultado da presente investigação. Os autores confirmaram ainda, que a função **principal** da LC em bases de dados é a de ampliar a precisão e não a **revocação**, o que vem de encontro aos resultados obtidos nesse estudo, pois, de um total de 422 referências classificadas

pêlos usuários como Muito bom, 131 foram obtidas com o uso exclusivo da LC, na estratégia de busca.

Dentre as dez estratégias da LC, um dos temas, o terceiro: denegação e religiosidade, apresentou um resultado de 100% de citações relevantes com o uso apenas da linguagem controlada, porém, a revocação foi baixa, tendo em vista o patamar estabelecido de 30 citações para cada tema de busca e sua respectiva estratégia diferenciada. Os temas de números 2, 7, 8 e 9 apresentaram uma resposta acima de 70% em termos de precisão, mantendo a revocação abaixo do esperado.

A primeira hipótese, H1, tinha como premissa que:

"os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem controlada não contribuem para o aumento de citações relevantes recuperadas na base de dados PSYCLIT". A variável dependente correspondia ao total de citações relevantes recuperadas na base e a **variável** independente ao uso de termos e/ou conceitos em linguagem controlada. A variável interveniente correspondia à lógica e os recursos da linguagem de recuperação da base de dados. Ambas as variáveis foram testadas conforme o previsto e os resultados obtidos, de certa forma, não contribuíram para o aumento de citações relevantes de maneira significativa, pois o desempenho da LC, comparado com os resultados totais ofereceram um índice de aceitação de **31.04%**.

Em artigo publicado em 1990, ROWLEY comparou a busca da LC com a da LN, e analisou as medidas de revocação e de precisão. Comprovou que o uso da LC na estratégia de busca, não ofereceu necessariamente o melhor desempenho, o que foi observado também no presente estudo. A linguagem controlada contribuiu sem dúvida, para a precisão dos resultados e para ampliar a confiança dos usuários nos serviços oferecidos, tendo em vista a participação **ativos** mesmos na escolha dos termos do *thesaurus* da própria base PSYCLIT e, na avaliação final citada anteriormente.

SIEVERT e BOYCE em 1983, e PITTERNICK em 1984, alertam para o fato de que os vocabulários controlados foram desenvolvidos tradicionalmente como instrumento de indexação, **mas** que vêm sendo utilizados como um instrumento de precisão na maioria das bases de dados. Entretanto, as diversas linguagens documentárias utilizadas para indexação, podem ser consideradas como preciosos instrumentos de auxílio na formulação das estratégias de busca, devido as peculiaridades de seus controles **terminológicos** que possibilitam uma efetiva identificação dos termos adequados e permitidos para uso em cada base de dados.

Apresenta-se na Figura 5, dados que demonstram o desempenho dos temas de busca com o uso exclusivo da linguagem controlada.

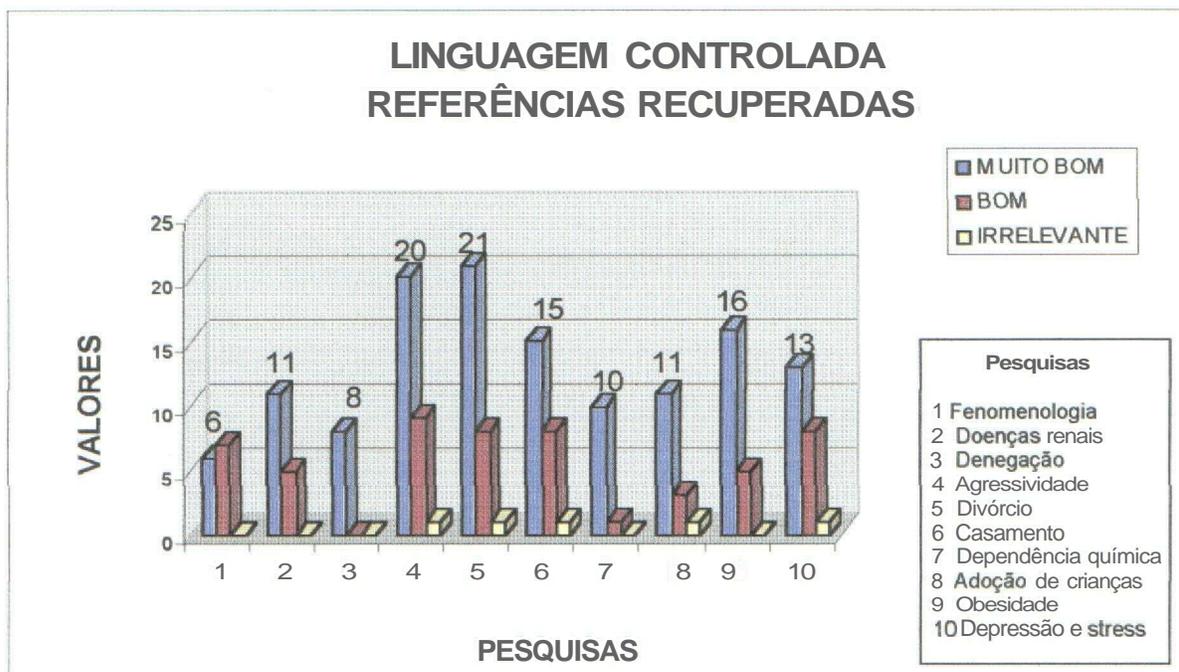


Figura 5 - LC: Referências Recuperadas

Pode-se inferir a partir da Figura 5, que alguns pontos ficaram evidentes em relação ao desempenho das estratégias da LC com os temas de busca, revelando que os temas 5 e 4 foram os melhores, ocupando a primeira e a segunda posição na avaliação de itens **relevantes**. O terceiro tema de busca, apesar de ter obtido apenas oito referências, foi avaliado pelo usuário, em 100% como muito bom, o que de certa forma demonstra que a LC atingiu plenamente a expectativa desse **usuário**, traduzindo suas necessidades de informação **para** a linguagem de indexação da base de dados PSYCLIT, de forma ideal. Entretanto, a **estratégia** para o tema 1, teve o pior desempenho com a LC, pois os itens considerados como de interesse periférico pelo usuário, ultrapassaram os itens considerados relevantes. Como tinha sido **previsto** um total máximo de 30 referências para cada modalidade de estratégia, pode-se observar que apenas os temas 4 e 5 atingiram esse patamar **estabelecido**. Todavia, com exceção do tema 1, todos os outros temas de busca obtiveram pela avaliação dos **usuários**, um resultado acima de suas respectivas médias, com o uso restrito na estratégia de busca, da linguagem controlada.

4.2.2 Estratégia de busca utilizando linguagem natural - LN

As estratégias de busca formuladas em linguagem **natural** contam com alguns pontos fortes, oferecendo dentre outros: termos extraídos diretamente dos documentos; alto grau de **exaustividade**; uso de novas terminologias ainda não incluídas nas linguagens controladas; alta expressividade e flexibilidade, porém com ambiguidades, e alta variedade de pontos de acesso. Na LN cada palavra tem um valor de recuperação igual e é ponto de acesso em potencial. Não existe demora na incorporação de novos termos em qualquer vocabulário, porque um novo termo já é parte integrante da LN tão logo ele seja registrado num documento.

A estrutura da informação na base de dados PSYCLIT em CD-ROM e a formatação dos campos para identificação dos termos que **reflitam** o interesse dos usuários na formulação de cada estratégias de busca apresenta a seguinte configuração:

- título (TI)
- resumo (AB)
- frase-chave (KP)

O campo de título apresenta o título do documento original. Títulos em idiomas diferentes do inglês são seguidos por barra (*/*) e a sua respectiva tradução. O campo de resumo inclui cerca de 75 à 175 palavras que **sumarizam** os tópicos principais. Os registros anteriores ao ano de **1967**, não possuem resumo, apesar da **retrospectividade** da base alcançar o ano de 1887. O campo de frase-chave contem termos livres oriundos da terminologia do autor e que descrevem o conteúdo dos documentos. Esse campo não foi utilizado no presente estudo, por sua excessiva duplicidade com as palavras contidas nos próprios títulos e resumos.

Com o **objetivo** de verificar o desempenho do uso exclusivo da linguagem **natural**- LN na estratégia de busca, foram elaboradas para cada um dos dez temas de pesquisa da amostra, a estratégia que correspondesse a essa questão, conforme assinalado na Tabela 8.

Tabela 8 - LN: temas de busca versus estratégia

TEMAS	ESTRATÉGIA DE BUSCA: TERMOS USADOS
1. Fenomenologia de distúrbios afetivos	Phenomenology; affectiv* ; disturbance* ; disorder* ; neurosis ; obsessive* ; compulsive*
2. Doenças renais: aspectos psicológicos	Kidney; renal; dialysis; anxiet* ; depress*
3. Denegação e religiosidade	Denial; defense near mechanism*; rational*; intellectual* ; sublimat* ; relig* ; christian* ; spiritual*
4. Agressividade na infância e adolescência	Aggressive; behav* ; AG=childhood; AG=adolescence
5. Divórcio na meia idade: crises e conflitos	Divorce* ; AG=middle age
6. Casamento e satisfação conjugal	Marriage ; marital; satisfaction
7. Dependência química e função familiar	Drug ; addiction* ; dependency; parental; famil* ; attitude* ; role* ; function* ; relation* ; structur* permissiviness*
8. Adoção de crianças: problemas de comportamento	Adopt*; child* ; behav* ; problem* ; disorder* ; change*
9. Obesidade: aspectos psicológicos	Obesity; self ; steem ; concept* ; confidence ; emotional* ; adjustment* ; control* ; body image
10. Depressão e stress na meia idade	Depressi* ; emotion* ; stress ; AG=middle-age

Os termos e/ou conceitos utilizados na formulação da estratégia de busca da LN foram identificados a partir da própria terminologia empregada pelos usuários quando da solicitação da pesquisa à **UCB/SIBI** - Serviço de Busca bibliográfica. Alguns dos recursos do **software** de recuperação de informação da base em CD-ROM foram utilizados, tais como: recuperação de termos simples e/ou compostos, nos campos de título (TI) e resumo (AB); uso de **truncagem** na raiz do termo e recuperação dos termos compostos com o uso de operadores oferecidos pela linguagem de busca da própria base.

Um recurso auxiliar da linguagem de busca na base de dados, denominado pesquisa com fragmentos de palavras, foi utilizado durante a formulação da estratégia de busca na LN. A utilidade das buscas, com fragmentos de palavras foi descrita por vários autores, entre eles, **LANCASTER** (1993, p. 204). Esse recurso de truncamento à esquerda das palavras, permitiu que os termos da estratégia da LN, fossem pesquisados na base de dados, com todas as possíveis variações de singular e/ou plural, substantivo e/ou **adjetivo** e outros casos, ampliando a possibilidade de recuperação dos termos, conforme pode ser observado sinteticamente na Tabela 8 e, de maneira mais detalhada no Anexo V.

Os resultados alcançados com as estratégias de busca utilizando termos apenas da linguagem natural são apresentados na Tabela 9.

Tabela 9 - Linguagem natural - referências recuperadas

LINGUAGEM NATURAL - REFERÊNCIAS RECUPERADAS				
PESQUISAS	MUITO BOM	BOM	IRRELEVANTE	TOTAL
1 Fenomenologia	03	04	-	07
2 Doenças renais	13	13	04	30
3 Denegação	17	07	06	30
4 Agressividade	20	08	02	30
5 Divórcio	16	07	07	30
6 Casamento	18	10	02	30
7 Dependência química	06	07	01	14
8 Adoção de crianças	17	04	09	30
9 Obesidade	10	05	02	17
10 Depressão e stress	12	04	-	16
TOTAL	132	69	33	234

O coeficiente de precisão, que é a relação entre o total de itens recuperados e o total dos itens considerados úteis e classificados pelos usuários como Muito bom foi de 56.41%. Do total de 234 referências recuperadas com o uso exclusivo da LN, **132** correspondem às citações **relevantes**, isto é, as que alcançaram exatamente o interesse expresso pelos usuários. As estratégias de busca operacionalizadas com o uso restrito da LN, proporcionaram um resultado de 69 referências, ou seja, 29.49% do total de citações periféricas, classificadas de Bom, e que podem vir a ser úteis no futuro. Do total geral da LN de 234 referências recuperadas, 33 citações, ou seja, 14.10% desse total, foram consideradas **irrelevantes** pelos usuários.

Os estudos que sugerem o uso da linguagem natural como instrumento eficiente para a recuperação da informação como WAGERS (1983) e FIDEL (1986), mostram, entretanto, a necessidade de um aperfeiçoamento por parte do produtor da base, no controle dos textos dos resumos, com o **objetivo** de ampliar o conhecimento sobre o conteúdo dos documentos. Vale destacar, que os resultados com a LN não foram considerados significativos, tendo em vista o alto percentual de citações não desejadas pelos usuários de 43,59%. Além disso, o resultado de Muito bom com a LC, foi de 131 referências contra 132 com a LN, sendo que com a LN, o índice de não satisfação dos usuários foi bastante ampliado e o da LC, ficou em **31,05%**, abaixo, portanto, dos resultados obtidos com a linguagem natural. Esses resultados sugerem que a ambiguidade da linguagem natural pode ter contribuído para a baixa precisão das citações dos seguintes temas: **fenomenologia**, **doenças renais**, **denegação**, **divórcio**, **casamento**, **dependência química**, **adoção** de crianças e **obesidade**.

Os estudos de KNAPP (1982) sobre o uso na estratégia, com melhor desempenho para a LN, mostram casos específicos de estratégias de busca, incluindo, dentre outros: o uso de novas **terminologias** para uma busca retrospectiva quando o termo da LC é muito recente e não cobre os anos **anteriores**; para a **complementação** de uma citação bibliográfica incompleta e outros, não incluindo, porém, a solução da ambiguidade da LN que provocou um índice de insatisfação bastante elevado.

A segunda hipótese, H2, tinha como premissa que:

- "os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem **natural**, não contribuem para o aumento de citações relevantes recuperadas na base de dados **PSYCLIT**" pôde ser confirmada, devido ao elevado índice de referências consideradas não satisfatórias pela avaliação dos usuários. A variável dependente correspondia ao total de citações relevantes recuperadas na base e a variável independente ao uso de termos e/ou concertos em linguagem natural. A variável interveniente correspondia à **lógica e os recursos** da linguagem de recuperação da base de dados. Ambas as variáveis foram testadas conforme o previsto sendo que os resultados obtidos, de certa forma, não contribuíram para o aumento de citações de maneira significativa.

WAGERS (1983) em estudo sobre a eficiência da estratégia de busca em LN, por intermédio unicamente dos temas que compõem o resumo dos documentos, sugeriu como FIDEL (1986), que o produtor da base enriqueça os resumos, dentre outros com: dados factuais, nomes de instituições e localidades geográficas.

Em estudo publicado em 1986, FIDEL analisou a importância dos guias para a elaboração de resumos visando o **aprimoramento** da busca em LN, comprovando que o desempenho da recuperação está diretamente relacionado com o conteúdo e o estilo da linguagem usada no resumo dos documentos. Tendo em vista que a estratégia de busca utilizando a LN, somente pode ser formulada, nas bases de dados bibliográficas, com os campos: de título, constituído por cerca de 30 palavras, e com o campo de resumo, constituído por aproximadamente 250 palavras, **seria** de extrema validade, o investimento por parte dos produtores das bases de uma melhoria no conteúdo dos mesmos.

TENOPIR (1987) questiona o planejamento da estratégia de busca em linguagem natural ou controlada, alertando para as ocorrências de casos específicos de uso da LN e/ou da LC, para uma elaboração de estratégia, que alcance resultados satisfatórios do ponto de vista do **usuário**.

Apresenta-se na Figura 6, dados que demonstram o desempenho das estratégias de busca nos respectivos temas, com o uso exclusivo da LN.

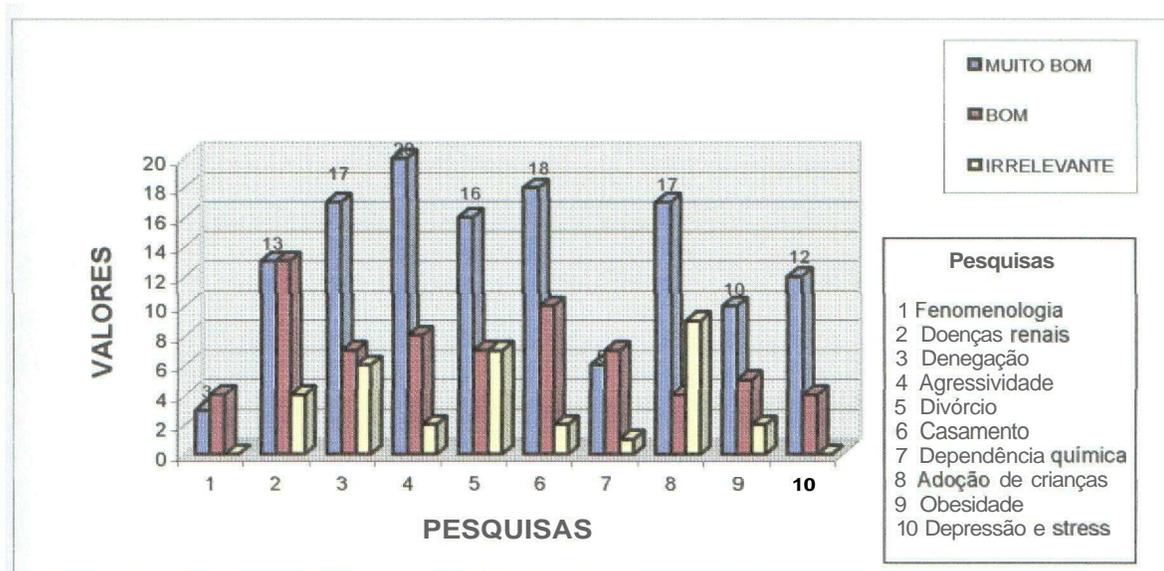


Figura 6 : LN - Referências recuperadas

Observa-se nos dados da Figura 6, que os temas de busca: 4, 6, 3 e 8, foram os melhores na avaliação dos itens relevantes. A estratégia planejada para o tema 5, obteve um resultado satisfatório, porém a soma dos itens considerados periféricos e irrelevantes quase alcançou o total de itens recuperados e considerados como muito bons pelo usuário. Os temas: 1, 7, 9 e 10 apresentaram resultados recuperados, abaixo do patamar previsto de 30 referências para cada tema de busca, embora, os itens 9 e 10 tenham obtido resultados categorizados como muito bom, acima da média, portanto, dos itens 1 e 7. Os temas de busca: 5 e 8, obtiveram um elevado resultado de itens irrelevantes com a estratégia de busca, ocasionados, possivelmente, pelas falsas coordenações com os termos da LN. O tema de pesquisa com o pior desempenho foi o tema 1, sobre fenomenologia, com apenas 7 referências recuperadas e apenas 3 classificadas como muito bom pelo usuário.

Os estudos de HARTER (1986, p. 54) sobre as características da LC e da LN, salientaram que as aparentes vantagens da LN, também podem ser vistas como desvantagens, devido a ambiguidade inerente à própria linguagem natural que apesar de permitir uma variedade de pontos de acesso, oferece problemas relacionados dentre outros com, os sinónimos e homógrafos, além das falsas coordenações.

4.2.3. Estratégia de busca utilizando LC e LN simultaneamente

As estratégias de busca formuladas com a soma das frases lógicas da LC e da LN pretendiam verificar se o uso **simultâneo** das mesmas poderia contribuir para o aumento de citações relevantes recuperadas na base de dados, conforme apresentado na hipótese H3. O detalhamento dessas estratégias encontra-se no Anexo V.

A estrutura da informação para busca na LC e LN da base de dados PSYCLIT implicava no uso dos mesmos campos já definidos para a estratégia da LC e da LN e, consistiu na soma dos resultados das frases lógicas utilizadas para as estratégias com a LC e com a LN separadamente. Esta soma lógica elimina os itens duplicados oferecendo portanto, uma resposta única com o uso simultâneo das duas linguagens: controlada e natural.

Objetivando a análise de desempenho do uso simultâneo da linguagem controlada e da linguagem natural na estratégia de busca, foram elaboradas para cada um dos dez temas de pesquisa da amostra, a estratégia que correspondesse a essa questão conforme assinalado na Tabela 10.

Tabela 10 - LC e LN: temas de busca versus estratégia

TEMAS	ESTRATÉGIA DE BUSCA: TERMOS USADOS
1. Fenomenologia de distúrbios afetivos	Soma lógica dos resultados da LC+LN com os termos: phenomenology; Affectiv*; disturbance*; disorder*; neurosis; obsessive*; compulsive*
2. Doenças renais: aspectos psicológicos	Soma lógica dos resultados da LC+LN com os termos: kidney; renal*; dialysis; anxiet*; depressi*
3. Denegação e religiosidade	Soma lógica dos resultados da LC+LN com os termos: denial; defense near mechanism*; rational*; intellectual*; sublimat*; christian*; spiritual*
4. Agressividade na infância e adolescência	Soma lógica dos resultados da LC+LN com os termos: aggressiv*; behav*; AG=childhood; AG=adolescence
5. Divórcio na meia idade: crises e conflitos	Soma lógica dos resultados da LC+LN com os termos: divorce*; AG=middle age
6. Casamento e satisfação conjugal	Soma lógica dos resultados da LC+LN com os termos: marriage; marital satisfaction*
7. Dependência química e função familiar	Soma lógica dos resultados da LC+LN com os termos: drug; addiction*; dependenc*; parental*; famil*; attitude*; role*; function*; relation*; permissiveness*; structur*
8. Adoção de crianças: problemas de comportamento	Soma lógica dos resultados da LC+LN com os termos: adopt*; child*; behav*; problem*; disorder*; change*
9. Obesidade: aspectos psicológicos	Soma lógica dos resultados da LC+LN com os termos: obesit*; self; steern; concept*; confidenc*; emotional*; adjustment*; control*; body image
10. Depressão e stress na meia idade	Soma lógica dos resultados da LC+LN com os termos: depress*; emotion*; stress; AG=middle-age

Os termos e/ou conceitos utilizados nas estratégias de busca com a LC e a LN foram extraídos do *thesaurus* contido na própria base em CD-ROM, e também, os que foram extraídos dos campos de título e resumo. A soma lógica dos resultados obtidos nas estratégias para a LC e nas estratégias para a LN, constituíram os resultados da LC+LN.

Os resultados alcançados com as estratégias de busca utilizando termos tanto da linguagem controlada, quanto os termos da linguagem natural são apresentados na Tabela 11.

Tabela 11 - LC e LN - referências recuperadas

LCeLN - REFERÊNCIAS RECUPERADAS				
PESQUISAS	MUITO BOM	BOM	IRRELEVANTE	TOTAL
1 Fenomenologia	09	10	-	19
2 Doenças renais	14	11	05	30
3 Denegação	14	07	09	30
4 Agressividade	18	11	01	30
5 Divórcio	12	10	08	30
6 Casamento	22	07	01	30
7 Dependência química	17	07	01	25
8 Adoção de crianças	17	05	08	30
9 Obesidade	19	10	01	30
10 Depressão e stress	17	12	01	30
TOTAL	159	90	35	284

O coeficiente de precisão, que é a relação entre o total de itens recuperados e o total dos itens considerados úteis e classificados pelos usuários como Muito bom foi de 55.98%, com 159 referências recuperadas de um total de 284 citações.. As estratégias de busca **operacionalizadas** com o uso simultâneo da LC e LN, proporcionaram um resultado de 31.70% de citações periféricas, com 90 referências, consideradas como de interesse futuro pelos usuários e avaliadas pelos mesmos como **Bom**. Do total de 284 citações recuperadas com a LC e a LN, 35 citações, ou seja, apenas 12,32% foi categorizada como **irrelevante** pelos usuários.

HENZLER, num estudo sobre o uso da LC e da LN, em 1978, concluiu que tanto os termos da linguagem natural, quanto os da linguagem controlada deviam estar presentes numa combinação ideal durante a elaboração das estratégias de busca. Os resultados alcançados no presente estudo confirmam essa assertiva concordando também, com CALKINS (1980) que, num estudo comparativo do uso da LC e da LN na estratégia de busca concluiu que a combinação das mesmas aumentou consideravelmente a recuperação de citações de interesse dos usuários.

Numa avaliação comparativa com os métodos de elaboração de estratégias de busca usando a LC e a LN, CARROW e NUGENT (1981), demonstraram que os dois métodos deveriam ser utilizados como complemento um do outro, salientando que o melhor desempenho da estratégia de busca seria aquele que utilizasse os dois métodos **concomitantemente**, confirmando dessa forma os resultados presentemente alcançados.

A terceira hipótese, H3, tinha como premissa que:

*os resultados de uma estratégia de busca quando se utiliza simultaneamente a LC e a LN podem contribuir para o aumento de citações relevantes recuperadas na base de dados PSYCLIT pôde ser confirmada devido ao elevado número de referências **classificadas** como muito bom pelos usuários. A variável dependente correspondia ao total de citações relevantes recuperadas e a **variável** independente ao uso simultâneo de termos e/ou conceitos da LC e da LN. A variável interveniente correspondia à lógica e aos recursos da linguagem de recuperação da base de dados. Ambas as variáveis foram testadas conforme o previsto, sendo que os resultados obtidos demonstraram que o uso combinado da LC e da LN **contribuíram** para o aumento de citações relevantes na base de dados PSYCLIT.

Em recente investigação, **MUDDAMALLE** (1998) **afirmou** que o uso combinado da LC e da LN na estratégia de busca ofereceu melhores resultados, sugerindo que ambas fossem usadas em conjunto, numa combinação ideal para a recuperação da informação, confirmando os resultados do presente estudo.

ROWLEY (1994, p. 117) salientou que existe um grande número de **fatores** que podem contribuir para a determinação de uma combinação ideal do uso de **termos** da LC e da LN numa estratégia de busca, incluindo:

- » o ambiente de busca, constituído por: área de assunto, o escopo e a cobertura da base de dados; busca em CD-ROM; busca em linha; e a linguagem de busca do sistema;
- o intermediário, incluindo: área do conhecimento na sua formação e a **"expertise"** em busca ;
- facilidades de recuperação e estratégias formuladas;
- natureza da busca, incluindo: necessidade de precisão; necessidade de **revocação**; natureza da estratégia e tipo de busca.

Acrescentem-se à essas observações, as de TENOPIR (1987), NICHOLLS e HOLTSMANN (1989) , KIRKBRIDE (1991) e outros autores, que tendo realizado estudos comparativos com o uso da LC e da LN em diferentes bases de dados, recomendaram o uso de todos os recursos para que os resultados finais fossem de encontro as reais necessidades de **informação** dos usuários.

Uma combinação de termos de busca baseados na LC e na LN é frequentemente considerado o melhor enfoque para a recuperação da informação em bases de dados, pela ampliação do número de citações **relevantes**, **evitando-se** as desvantagens inerentes a cada uma dessas linguagens. A facilidade de acesso à terminologia oficial da base de dados, contida **regularmente** no próprio CD-ROM, ou mesmo disponível em linha, oferece um sólido respaldo para a seleção dos termos e/ou conceitos que vão compor a estratégia de busca. É bem verdade, que o **desenvolvimento** de novas **tecnologias** de registro e acesso à informação, tem contribuído de forma inquestionável no **aprimoramento** do processo de recuperação da informação, porém, o polêmico debate sobre o uso da linguagem controlada ou da natural nas estratégias de busca em bases de dados permanece em questão até o presente momento, sugerindo desta forma, novos estudos que auxiliem na solução desses problemas.

Com o propósito de analisar o comportamento da LC e da LN nos temas de busca, apresenta-se na Figura 7, as referências recuperadas com o uso conjunto das duas linguagens.

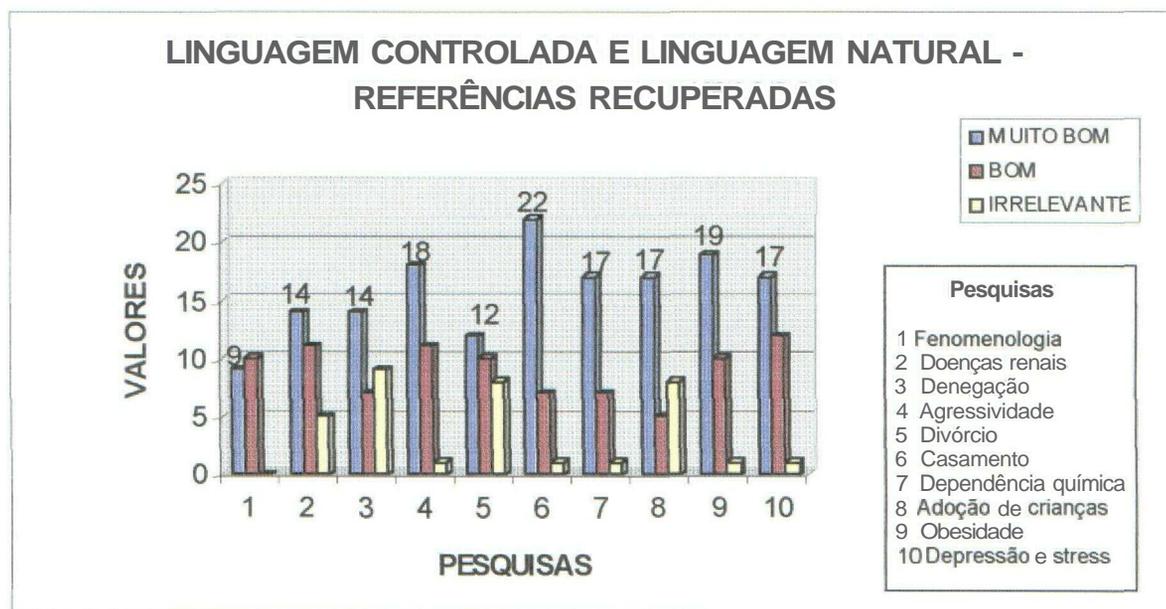


Figura 7 - LC e LN: referências recuperadas

Pode-se inferir pelos dados da Figura 7, que o desempenho da LC e da LN no que se refere ao índice muito bom foi significativo. Os resultados categorizados como relevantes para os temas de busca: 6, 9, 4, 7, 8 e 10 **destacaram-se** com o uso simultâneo na estratégia da LC e da LN, obtendo-se um desempenho acima da média. As pesquisas: 1 e 7, não atingiram a totalidade das 30 referências **pré-estabelecidas**, recuperando respectivamente 19 e 25 citações, **entretanto** o tema 7, obteve uma classificação de relevância de 72% do total de referências recuperadas.

4.3 Análise comparativa dos resultados

Os dados obtidos com a **aplicação** das três modalidades de formulação de estratégias de busca, a saber: com a linguagem **controlada**, com a linguagem **natural** e com o uso simultâneo da linguagem controlada e da linguagem natural, apresentaram resultados que confirmaram as hipóteses formuladas no presente estudo.

Os dados foram analisados separadamente para cada um dos **critérios** de avaliação **obtidos**, a saber: **Muito bom**, para as citações altamente relevantes; **Bom**, para as citações periféricas e **Irrelevantes** para as citações julgadas insatisfatórias. Os mesmos foram coletados nos formulários de avaliação das referências (Anexo III), realizadas pelos usuários solicitantes das buscas. Na Figura 8, apresenta-se uma comparação dos resultados das estratégias de busca considerando-se apenas os dados do índice **Muito bom**, com as três modalidades de estratégias: LC, LN e LC+LN.

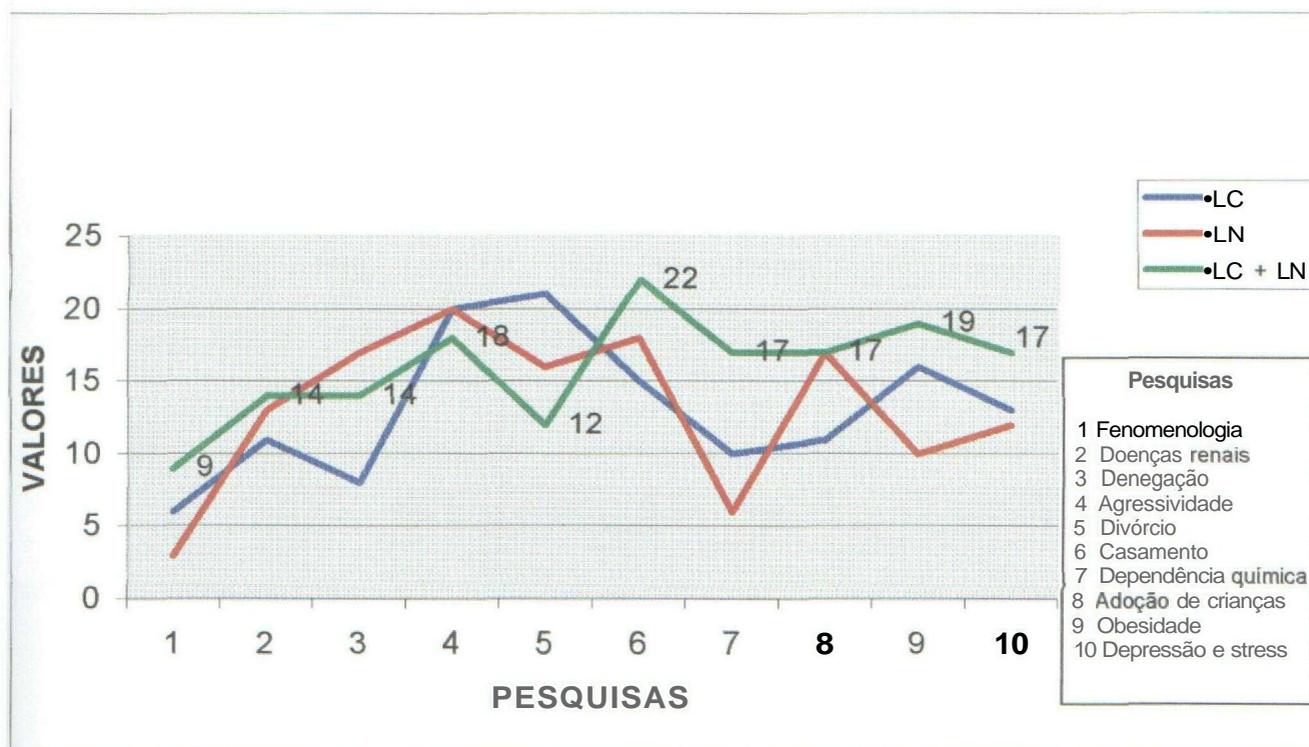


Figura 8: Comparação dos resultados das estratégias para o índice Muito bom.

O coeficiente de precisão entre o total de itens recuperados e o total dos itens considerados relevantes e classificados pelos usuários como **Muito bom**, foi de:

- Linguagem controlada 31.04%
- Linguagem natural 31.27%
- Linguagem controlada + Linguagem natural 37.67%

A estratégia de busca formulada com a LC e a LN apresentou um desempenho melhor de recuperação e satisfação para os usuários com 159 citações recuperadas de um total de 422 referências classificadas como relevantes pelos usuários. A estratégia com a LC e com a LN obtiveram um resultado semelhante, com o índice de 31.27% para as 132 citações classificadas como MB para a LN, e o índice de 31.04% para as 131 citações da LC.

A Figura 3 demonstra que de um total de 708 referências recuperadas, 422 ou 59.60% foram classificadas como relevantes ou Muito bom pelos usuários. Desse resultado a LC+LN teve um desempenho maior do que a LC e a LN, comprovando a hipótese 3:

"H3 - Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza simultaneamente a LC e a LN podem contribuir para o aumento de citações relevantes recuperadas na base de dados PSYCLIT".

Na Figura 9, são apresentados os dados relativos ao índice Bom.

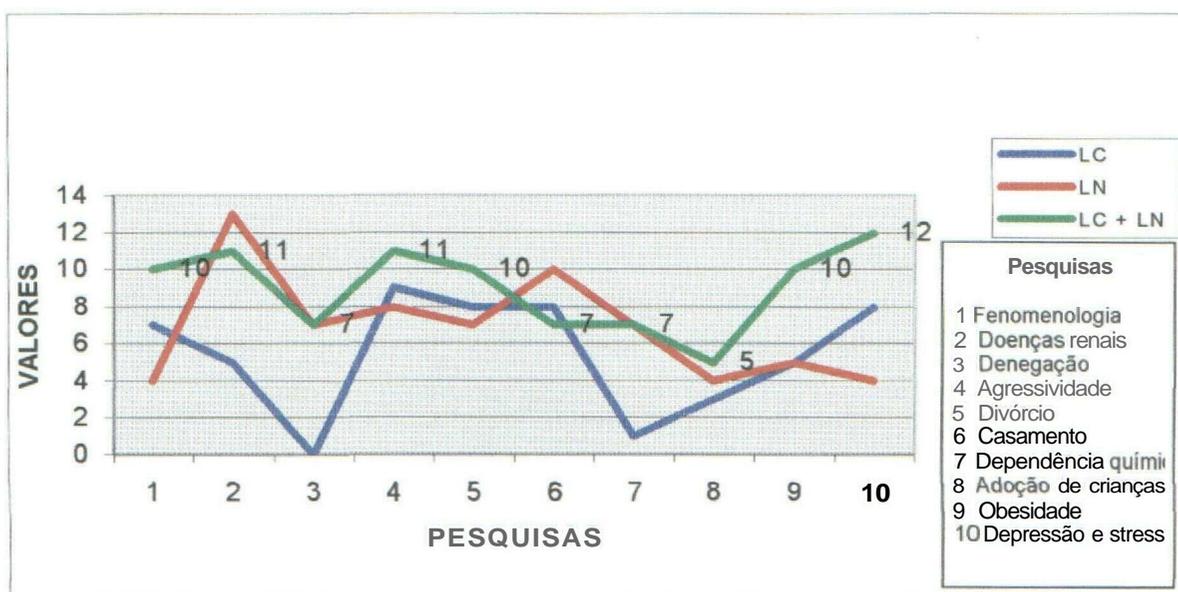


Figura 9 - Comparação dos resultados das estratégias para o índice Bom.

O coeficiente de precisão entre o total de itens recuperados e o total dos itens considerados periféricos e classificados pelos usuários como Bom, foi de:

- Linguagem controlada 25.35%
- Linguagem natural 32.40%
- Linguagem controlada + linguagem natural 42.25%

A estratégia de busca formulada com a LC+LN apresentou 90 itens recuperados, de um total de 213 referências classificadas de periféricas pelos usuários, confirmando a observação de LANCASTER (1993)

"Uma característica tamentável inerente à recuperação de informação é uma melhoria de revocação em geral, implicar uma perda de precisão e *vice-versa*." (Lancaster 1983, p. 4).

Os dados classificados de Bom pelos usuários comprovam que ao aumento de revocação correspondeu uma perda na precisão.

Apresenta-se na Figura 10, os dados comparativos dos resultados das estratégias para o índice irrelevante.

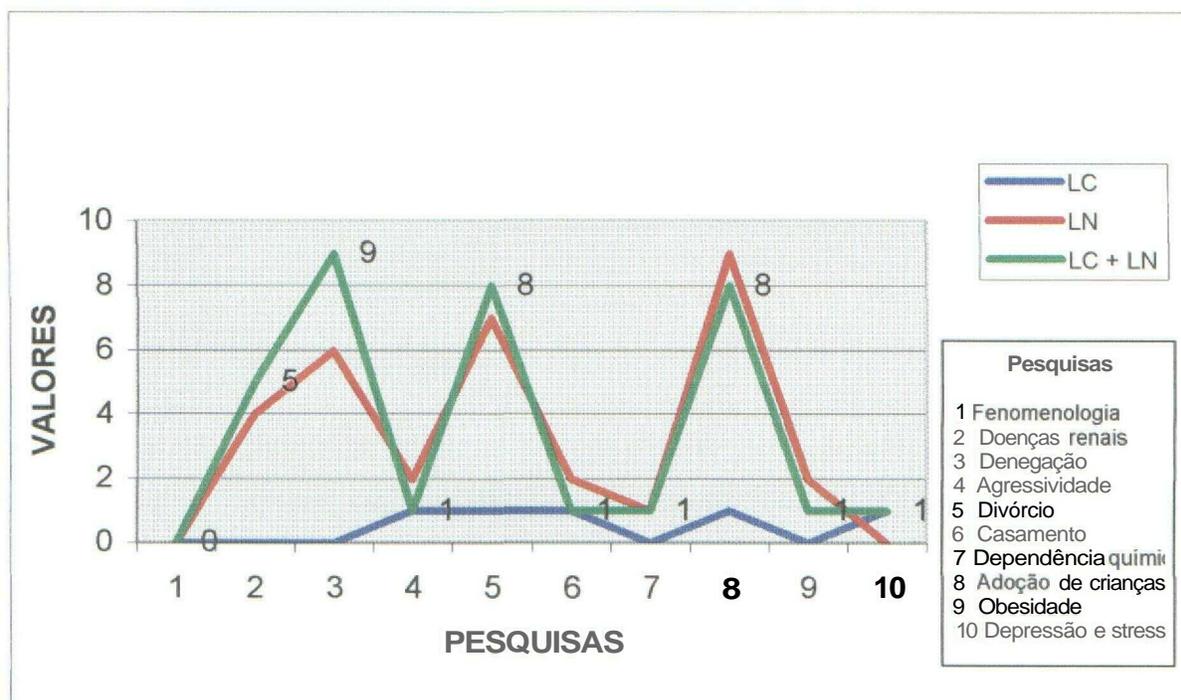


Figura 10 - Comparação dos resultados das estratégias para o índice Irrelevante.

O coeficiente de precisão entre o total de itens recuperados e o total dos itens considerados irrelevantes pêlos usuários foi de:

- Linguagem controlada 6.85%
- Linguagem natural 45.20%
- Linguagem controlada + linguagem natural 47.95%

A estratégia de busca formulada com a LC + LN apresentou um resultado de não relevância maior do que a LC e a LN, com um índice de 47.95% correspondendo à 35 das 73 referências recuperadas, classificadas como irrelevantes pêlos usuários.

Num cômputo geral, das 708 referências recuperadas, os itens classificados como irrelevantes pêlos usuários, nas três modalidades de estratégias de busca executadas, corresponderam a 73 referências, sendo: 5 para a LC; 33 para a LN e 35 para o uso concomitante da LC+LN. A precisão da LC na estratégia de busca foi confirmada, com os índices maiores de irrelevância ocasionados pelo uso da LN, ocorrendo também, na LC+LN, quando do resultado da soma lógica dos resultados da LC com a LN. A desvantagem de uso da LN, em relação a LC, fica evidente pois, ocorreu uma significativa incidência de respostas negativas, ou de relações incorretas entre os termos usados na estratégia, provavelmente, por falta de normalização dos termos da linguagem natural. Assim, nos temas de busca: doenças renais, denegação, divórcio e adoção de crianças, o índice de irrelevância foi de quase 50% em relação ao total de citações recuperadas, o que pode ser considerado altamente insatisfatório para os usuários.

Pode-se inferir pêlos dados apresentados nas Figuras 8, 9 e 10, que o melhor desempenho de formulação de estratégia de busca ocorreu com o uso simultâneo das linguagens controlada e natural, embora tenha ocorrido um elevado índice de citações consideradas periféricas pêlos usuários com o uso dessas linguagens em conjunto.

Como bem ressalta HARTER (1986), FIDEL (1991), LANCASTER (1986, 1993), ROWLEY (1994), SVENONIOUS (1986, 2000) , e outros autores, a prática e a experiência dos intermediários ainda comanda a seleção dos termos para as estratégias de busca. Embora a experimentação persistente possa resolver em parte, a questão de qual tipo de linguagem deva ser utilizada nas estratégias de busca, existe uma crescente evidência de que buscas com termos na LC e na LN, na verdade se complementam mutuamente, o que veio de encontro aos resultados alcançados no presente estudo.

5. CONCLUSÕES

5.1 Introdução

A análise efetuada neste estudo possibilitou a consecução dos objetivos pretendidos, fornecendo subsídios para um melhor entendimento do complexo processo de recuperação de informação em bases de dados em CD-ROM ou em linha. O propósito do estudo era o de examinar o processo de planejamento de estratégia de busca, num ambiente de bases de dados em CD-ROM, verificando qual o tipo de vocabulário deveria ser empregado na mesma. Como objetivos específicos foram estudados: identificar a adequação do uso da linguagem controlada na elaboração da estratégia de busca; identificar a adequação do uso da linguagem natural e identificar o uso simultâneo da linguagem controlada e da natural na elaboração da estratégia de busca.

Três hipóteses foram testadas. A primeira (H1) foi a de que os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem controlada não contribuem para o aumento de citações relevantes. A segunda (H2) foi a de que os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem natural não contribuem para o aumento de citações relevantes. E finalmente a terceira (H3) foi a de que os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza simultaneamente a LC e a LN podem contribuir para o aumento de citações relevantes recuperadas na base de dados PSYCLIT. A metodologia empregada neste trabalho pode ser aplicada à bases de dados de outras áreas do conhecimento, tendo em vista as peculiaridades de terminologia de outras áreas do campo científico e tecnológico.

Os dados para esse estudo foram obtidos a partir do atendimento às solicitações de busca dirigidas à UCB/ SIBI - Serviço de Busca Bibliográfica por usuários da área de Psicologia. Foram elaborados três conjuntos de dez buscas com os seus respectivos planejamentos das estratégias, que foram operacionalizadas especificamente com termos extraídos da LC, da LN e com a soma lógica proveniente dos resultados da LC+LN. Essas formulações de estratégias resultaram na composição definitiva da amostra, totalizando trinta estratégias de busca diferenciadas umas das outras pelo tipo de vocabulário que foi empregado. Os resultados das buscas foram limitados aos últimos dois anos, de janeiro de 1998 até março de 2000, e com o máximo de trinta referências para cada estratégia.

Foram obtidas com as três modalidades de estratégias de busca um total de 708 referências, sendo que 422 referências, ou seja, 59.60% desse total, foram classificadas pelos usuários como citações relevantes. Desses resultados 213 citações foram classificadas como periféricas,

correspondendo a 30.09% do total de citações, com âmbito mais genérico do que o interesse atual do usuário podendo vir a ser úteis em futuro próximo. As referências classificadas como irrelevantes por todos os usuários participantes desse estudo foram na ordem de 73, para as três modalidades de estratégias de busca, correspondendo à 10.31% de participação no resultado final.

5.2 Testes das hipóteses

Com base na operacionalização das estratégias de busca e nas avaliações efetuadas pelos usuários, para cada modalidade de estratégia, foram geradas avaliações das referências individualmente proporcionando primeiramente, um conjunto de dados com os resultados das estratégias na linguagem controlada; a seguir, um segundo conjunto de dados com os resultados das estratégias apenas com a terminologia da linguagem natural e um terceiro conjunto de dados com os resultados de busca efetuados utilizando-se a soma lógica dos resultados da LC e da LN simultaneamente. Além disso, foi efetuada uma avaliação posterior por cada usuário, para validade dos dados. Diversas conclusões são colocadas, como a seguir.

Os dados foram tabulados para proporcionar uma análise mais acurada que permitisse determinar a aceitação ou rejeição das seguintes hipóteses:

5.2.1 Hipótese nº 1

A hipótese H1 tinha a seguinte premissa:

"Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem controlada (LC) não contribuem para o aumento de citações relevantes recuperadas na base de dados PSYCLIT".

Conclusões:

Utilizando-se na formulação da estratégia de busca os termos que ocorrem apenas no campo de descritor, códigos de classificação de assunto e faixa etária, foram encontrados um total de 190 citações, com um coeficiente de precisão de 68.95% para 131 referências consideradas relevantes para os usuários. Desse total, 54 referências foram classificadas como bom, isto é, citações periféricas que não interessavam aos usuários no momento da pesquisa, totalizando 28.42% de citações de futuro interesse. Esses dados estão apresentados na Tabela 7. Os resultados das estratégias com a LC produziram apenas 2.63% de citações irrelevantes, que correspondem a cinco referências. O baixo índice de citações irrelevantes comprovou que houve

uma perda mínima de itens não relevantes, porém a **revocação** não proporcionou uma diferença considerável perante a LN, confirmando a hipótese.

5.2.2 Hipótese n° 2

A hipótese H2 tinha a seguinte premissa:

"Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza a linguagem natural (LN) não **contribuem** para o aumento de citações relevantes recuperadas na base de dados PSYCLIT."

Conclusões:

Usando-se no planejamento da estratégia de busca os termos que ocorrem apenas nos campos de título e resumo dos documentos, foram encontrados um total de 234 citações, com um coeficiente de precisão de **56.41%** para 132 referências consideradas relevantes pelos usuários. Desse **total**, 69 referências foram consideradas periféricas, isto é, de âmbito mais genérico do que o **atual** interesse do usuário perfazendo 29.49% de citações classificadas como bom e apresentadas na Tabela 9. Apenas 14.10% foram consideradas irrelevantes pelos usuários. A diferença de itens relevantes em relação aos resultados obtidos com as estratégias para a linguagem controlada foi de apenas uma citação. Houve um aumento não significativo de revocação em relação à LC, porém com perda de precisão, confirmando a hipótese.

5.3.3 Hipótese n° 3

A hipótese H3 tinha a seguinte premissa:

"Os resultados de uma estratégia de busca de informação em bases de dados quando se utiliza simultaneamente a LC e a LN podem contribuir para o aumento de citações relevantes recuperadas na base de dados PSYCLIT."

Conclusões:

Usando-se na formulação da estratégia de busca os termos que ocorrem simultaneamente, na LC, nos campos de descritor, códigos de classificação de assunto e faixa etária e na LN, nos campos de título e resumo, foram encontrados um total de 284 citações. O coeficiente de precisão

que é a **relação** entre o total de itens recuperados e o total dos itens considerados relevantes pelos usuários, foi de 55.98%, com 159 referências classificadas como muito boas pelos mesmos. As citações periféricas, classificadas como de interesse futuro pelos usuários foram da ordem de 90 referências, com um percentual de **31.70%**. Esses dados estão apresentados na Tabela 11. Do total de 284 referências recuperadas com a LC+LN usadas simultaneamente, foram encontradas 35 citações categorizadas como irrelevantes pelos usuários, com um percentual de 12.32%. Os resultados obtidos com o uso **simultâneo** de termos da LC e da LN, nas estratégias de busca, comprovaram a **hipótese**, pois o coeficiente de precisão entre as três modalidades de estratégia de busca foi de 37.67% para a LC+LN, enquanto que a LC apresentou um coeficiente de 31.04% e a LN um coeficiente de **31.27%**.

5.3 Objetivos

O **objetivo** geral proposto para a pesquisa, qual seja, o de estudar o processo de planejamento de estratégia de busca, num ambiente de bases de dados em CD-ROM ou de acesso em linha, verificando qual o tipo de vocabulário que deve ser empregado na mesma foi alcançado, tendo em vista a consecução dos **objetivos** específicos relacionados nos itens analisados a seguir.

Foram examinados quatro objetivos no presente estudo que proporcionaram as seguintes conclusões:

5.3.1 Objetivo A : adequação da LC

O objetivo A tinha o seguinte propósito:

- identificar a adequação do uso do vocabulário controlado na elaboração da estratégia de busca.

Conclusões:

A busca utilizando **termos** da linguagem controlada é um processo que requer do intermediário e/ou do usuário final, um conhecimento antecipado da terminologia utilizada pelo produtor da base para a indexação dos documentos. Atualmente os sistemas de recuperação da informação oferecem inúmeras bases de dados que apresentam os respectivos vocabulários

controlados em diferentes níveis de tratamento da terminologia. Não existe padronização ou compatibilização entre os vocabulários usados como bem alertou **Svenonious** (2000).

O processo de busca de informação envolve em princípio duas etapas bem definidas dentre outras, a saber: a etapa de análise de conteúdo ou análise conceitual da solicitação da busca, e a etapa de tradução desses conteúdos para a linguagem da base de dados a ser pesquisada. Os usuários desse estudo foram envolvidos na fase de **análise** de conteúdo contribuindo para a **seleção** dos termos que deveriam ser utilizados nas estratégias de busca, utilizando o **thesaurus** da base PSYCLIT. Os resultados em termos quantitativos de referências recuperadas não diferiram significativamente dos resultados encontrados com as estratégias executadas com a linguagem natural, não oferecendo portanto, o melhor desempenho em termos de citações consideradas pelos usuários como relevantes.

Pelo exame dos resultados foi identificado que não existem evidências neste estudo para supor que a terminologia a ser usada na estratégia de busca deva ser apenas a registrada nos campos específicos de controle de assunto definido pelo produtor da base de dados. Sugere-se que a mesma terminologia da LC possa também, ser utilizada em todos os outros campos, ampliando-se as possibilidades de recuperação da informação e, possivelmente, a satisfação do usuário.

5.3.2 Objetivo B: adequação da LN

O **objetivo B** tinha o seguinte propósito:

- identificar a adequação do uso da linguagem natural na elaboração da estratégia de busca.

Conclusões:

A busca utilizando termos da linguagem natural é um processo idêntico, pelo qual a base de dados é auto-indexada. Cada palavra contida nos documentos da base de dados é um ponto de acesso em potencial e identifica a unidade do texto na qual está inserida. Nesse caso, nenhum termo é assinalado como termo de indexação e nenhuma referência cruzada ou qualquer outro tipo de controle de vocabulário é inserido. Na busca em linguagem natural o intermediário e/ou usuário final, interage com os documentos da base diretamente, porém, a dificuldade maior está em questionar a base de dados com todas as possíveis terminologias que se aplicam a um determinado conceito. A busca na linguagem natural tem suas limitações já relacionadas no

capítulo 2. Cabe destacar, porém, a decisão que deve ser tomada em relação aos problemas contextuais, no momento de formulação da estratégia de busca. Palavras existem, em todas as línguas, cujos significados variam conforme as circunstâncias na qual são usadas. Exemplificando com a palavra "**Stress**", numa base de dados **multidisciplinar**, o contexto no documento pode estar relacionado com fadiga de material, de interesse da área de Engenharia, mas, também podem existir documentos na mesma base de dados, cujo contexto interesse ao profissional de saúde. O refinamento a ser feito para evitar possíveis "ruídos" no resultado final da busca, deve prever todas as possibilidades de **contextualização** para essa palavra, o que de certa forma aumenta o custo final da busca e o tempo de resposta. Por outro lado, os sistemas de recuperação da informação e/ou as bases de dados que permitem a varredura dos registros em linguagem natural apresentam certas vantagens, pois, não impõem restrições no vocabulário utilizado pelo solicitante da busca. O uso da LN tem suas vantagens imediatas e dependendo do pedido de busca, podem realmente vir a ser, a única solução possível para um determinado pedido de informação.

A análise dos resultados identificou que não existem evidências neste estudo para supor que a terminologia a ser usada no planejamento da estratégia de busca deva ser apenas pesquisada nos campos **específicos** da LN definido pelo produtor da base de dados. **Sugere-se** que a mesma terminologia da LN possa ser usada em todos os campos, ampliando-se as possibilidades de recuperação de informação e, possivelmente, a satisfação do usuário.

5.3.3 Objetivo C - Adequação da LC+LN

O **objetivo C** tinha o seguinte propósito:

- identificar a adequação do uso simultâneo do vocabulário controlado e da linguagem natural na elaboração da estratégia de busca.

Conclusões:

A busca utilizando **termos**, tanto da linguagem natural quanto da linguagem controlada, é um recurso passível de execução devido aos dispositivos desenvolvidos pelos sistemas de recuperação e pelas bases de dados. Cada palavra no item de informação é candidata a ponto de acesso, independente de se **pré-definir** ou não, o campo em que a mesma estiver registrada. Nesse estudo foram definidos previamente os campos de pesquisa, tanto para a linguagem natural, quanto para a linguagem controlada, **executando-se** cada estratégia de busca em separado, para posteriormente efetuar-se a soma lógica dos respectivos resultados. Evidenciou-se que o uso conjunto dessas duas linguagens proporcionou um resultado superior ao obtido com as

estratégias executadas, tanto com a LC quanto com a LN separadamente. Por esse motivo, sugere-se que tanto a terminologia da LC, quanto a da LN, possam ser utilizadas em todos os campos de um registro de informação, **ampliando-se** as **possibilidades** de recuperação da informação e, possivelmente, a satisfação do usuário.

5.3.4 Objetivo D - Identificação da melhor linguagem

O objetivo D tinha o seguinte propósito:

- identificar qual a linguagem mais adequada para a **operacionalização** da estratégia de busca.

Conclusões:

O alcance deste objetivo foi atingido, uma vez que: o melhor desempenho com as três modalidades de estratégia de busca, foi obtido com o uso simultâneo da linguagem controlada e da linguagem natural o que veio de encontro não somente as hipóteses do presente estudo, como também aos autores que têm se destacado na literatura especializada sobre o tema.

5.4 Conclusões finais:

A análise do uso da LC e da LN para recuperação da informação, numa área específica do conhecimento como é o caso de Psicologia, soma-se a outras investigações relatadas na literatura para identificar o melhor desempenho dessas linguagens no que se refere ao grau de satisfação dos usuários.

A experiência profissional e os estudos relatados na literatura demonstram que a abordagem de uma busca para atendimento a demanda de informações na área acadêmica é menos profunda do que a necessidade de informação expressa por usuários da área industrial. Tendo em vista o constante intercâmbio e **interação** formal e informal entre os membros da Academia, observa-se que a busca de informação obedece mais aos interesses científicos do que aos tecnológicos por parte desses usuários.

Por outro lado, a demanda de informação na área industrial obedece aos ditames mais fortes dos segredos industriais, aliados ao clima sempre crescente de competição e globalização

da economia. Por esse motivo, a procura incessante por novos processos e produtos por parte do segmento **industrial**, tem requerido dos produtores das bases de dados um constante controle de qualidade nos seus produtos de **informação**, gerando uma **recuperação** que atenda às expectativas desse segmento economicamente mais exigente.

HARTER . em estudo publicado em 1986, mensurou os pedidos de informação dos usuários acadêmicos **comparando-os** com a demanda de informações dos solicitantes da área tecnológica, concluindo que estes são mais profundos em suas abordagens de informação, exigindo qualidade dos dados obtidos, **enquanto** que aqueles aceitam a quantidade como **fator** de satisfação de suas necessidades de **informação**.

KNAPP, COHEN e JUEDES (1998), afirmaram que a pesquisa em bases de dados oferece dificuldades especiais para a área de **humanidades**, dado que os assuntos podem ser abordados com diferentes variações **terminológicas** afetando, portando, os resultados finais, como asseverou **também, LANCASTER (1993)** no que concordamos plenamente.

Ao longo dos últimos anos, os estudos e as investigações realizadas sobre o uso controverso na estratégia de busca, da linguagem controlada e da **linguagem** natural, têm sugerido a utilização conjunta dessas linguagens como **forma** ideal para a recuperação da informação. Entretanto, a escolha dos termos para a formulação de uma estratégia de busca não é uma decisão trivial, exigindo frequentemente, conhecimentos mais profundos sobre a **seleção** de terminologia que represente os conceitos do tema de busca solicitado. Por esse motivo, sugere-se a inclusão nos programas de treinamento para uso de bases de dados, dos conceitos da linguagem controlada e da natural e seus respectivos **fatores** de impacto nos resultados finais das estratégias de busca.

5.5 Sugestões para estudos futuros

O presente estudo sugeriu os seguintes tópicos adicionais para pesquisa:

1. estudos comparativos de utilização do vocabulário controlado e da linguagem natural em outras bases de dados, utilizando as diferentes linguagens documentárias na área de Ciências Humanas e Sociais;
2. estudos comparativos de utilização da linguagem controlada e da linguagem natural em outras bases de dados da área tecnológica.

3. estudos comparativos entre as linguagens controlada e natural no que diz respeito a flexibilidade de planejamento das estratégias de busca em diferentes bases de dados;
4. estudos comparativos de utilização das linguagens controlada e natural, no que diz respeito as medidas de precisão e **revocação**, relacionadas com a demanda e satisfação do usuário.

Futuros estudos são sugeridos devido a **discussão** intensa que tem sido realizada nas últimas décadas sobre o uso da LC e da LN na recuperação da informação. Apesar do contínuo avanço das novas tecnologias aplicadas ao registro e controle da informação, ainda são necessários desenvolvimentos de bases de dados e programas de recuperação de informação que viabilizem a utilização dos recursos **informacionais** disponíveis de forma mais **satisfatória**.

REFERÊNCIAS BIBLIOGRÁFICAS:

ABELS, E.G. The e-mail reference interview. *RQ*, v. 35, n.3, p. 345-357, Spring, 1996

ANDER-EGG, E. *Introducción a las técnicas de investigación social, para trabajadores sociales*. 7.ed. Buenos Aires, **Humanitas**: 1978.

ANDERSON, J.D., WILSON, C.B. Essential decisions in indexing systems design. IN: FEINBERG, H. *Indexing specialized formats and subjects*. London: **Scarecrow Press**, 1983, **cap. 1.**

ARMSTRONG, C. J. , LARGE, J. A, (ed). *Manual of online search strategies*. Boston : G.K. **Hall & Co.**, 1988.

AUSTER, E., LAWTON, S. B. Search interview techniques and information gain as antecedents of user satisfaction with online bibliographic retrieval. *JASIS*, v. 35, n. 2, p. **90-103**, 1984.

AUSTIN, D. Vocabulary control and information technology. *Aslib Proceedings*, v. 38, n. 1, p. **1-15**, Jan. 1986.

BARDIN, Laurence. *Análise de conteúdo*. Lisboa : Edições 70, 1995.

BATES, Maria J. How to use controlled vocabulaires more effectively in online searching. *Online*, v.12, n. 6, p. **45-56**, Nov. 1988.

------. How to use information search tactics online. *Online*, v.11, n. 5, p. 47-54, May 1987.

----- Information search tactics. *JASIS*, v.30, n. 4, p. 205-14, 1974.

----- Search techniques. *ARIST*, v.16, p. 139-169, 1981.

------. Where should the person stop and the information search interface start? *Information Processing and Management*. v. 26, n. 5, p **575-591**, 1990.

BATTACHARYA, K. The effectiveness of natural language in Science indexing and retrieval. *Journal of Documentation*, v. 30, n. 3, p. 235-293, Sept. 1974.

BAYLEY, K.D. *Methods of Social Research*. New York: Free Press, 1978. P. 84

- BOURNE, C. Frequency and **impact** of spelling errors in **bibliographic** databases. *Information Processing & Management*, v. 13, p. 1-12, 1977.
- BOYCE, B. R. , **McLAIN**, J. P. **Entry** point depth and online search using a controlled vocabulary. *JASIS*, v. 40, n. 4, p. 273-276, 1989.
- BRUNDAGE, **Christina** A. Teaching controlled vocabulary and natural language to **end-users** of scientific online and **CD-ROM** databases. *Science & Technology Libraries*, v. 10, n. 1, p. 3-13, Fall 1989.
- BUSHA, C. H., **HARTER**, S. P. *Research methods in librarianship: techniques and interpretation*. New York : **Academic Pres**, 1980.
- CALKINS, Mary L. Free text or controlled vocabulary? A case history **step-by- step-analysis...plus** other **aspects** of search **strategy**. *Online*, v.3, n. 2, p. 53-65, June, 1980.
- CARROW, D., **NUGENT**, J. **Comparison** of free-text and index search abilities in an operating information **system**. IN: INFORMATION MANAGEMENT IN THE **1980's**, 40th, 1977. New York. Proceedings... New York: **ASIS**, 1977. v. 14, p. 232-238.
- CHAMIS, A. Y. *Vocabulary control and search strategies in online searching*. New York : Greenwood Press. 1991.
- CHARTON, B. Searching the literature for **concepts**. *Journal of Chemical Information and Computer Sciences*, v. 17, p. 45-46, 1977.
- CLEVERDON, C.W. The **Cranfield tests** on index language **devices**. *AslibProceedings*, v. 19, p. 173-194, 1967.
- . *A comparative evaluation of searching by controlled language and natural language in the NASA database*. **Bedford** : Cranfield Institute of Technology, 1977. ESA **Contract** Number 1/432. (cópia de microfilme). 1 v.
- COMBER, D., **STANFORD**, J. A **comparison** between free text and a **thesaurus** controlled vocabulary in searching an online **records** management database. *Records Management Journal*, v. 1, n. 3, p. 113-120, **Aut**. 1989.
- COUSINS, S. A. **Enhancing** subject access to OPACs: controlled vocabulary vs natural language. *Journal of Documentation*, v. 48, n. 3, p. 291-309, Sept. 1992.
- CURNUTT, G. L. Reactive **chemical** hedges: a search **tool** for **comprehensive** retrieval of Chemical safety data. *Journal or Chemical Information and Computer Sciences*, v. 31, n. 1, p. 116-119, 1991.

DERVIN, B., DEWDNEY, P. **Neutral questioning: a new approach to the reference interview.** *RQ*, v. 25, n. 4, p. 506-513, Summer 1996.

DICIONÁRIO Aurélio Eletrônico. Rio de Janeiro, Nova Fronteira, 1996. Versão 2.0

DIMARTINO, D., ZOE, L. R. **End-user full-text searching: access or excess?** *Library and Information Science Research*, v.18, n. 2, p.133-49, 1996.

DOLAN, D. R. **Hedges for online searching.** *Database*, v. 3, n. 1, p. 79-82, Mar. 1980.

DRABENSTOTT, K. M. , VIZINE-GOETZ, D. **Usina *subject headings* for online *retrieval: theory, practice and potential.*** New York : Academic Press, 1994.

DUBOIS, C. P. R. **Free text vs. controlled vocabulary, a reassessment.** *Online Review*, v. 11, n. 4, p. 243-251, Aug, 1987.

D'UNRUG, M.C. **Analyse de *contenu* et *acte de parole.*** Paris : Editions Universitaires, 1974.

FENICHEL, C.H. **The process of searching online bibliographic databases, a review of research.** *Library Research*, v. 2, p. 107-127, 1980.

FIDEL, Raya. **Controlled vocabulary and free-text searching: searches selection of search keys.** IN: ASIS ANNUAL MEETING, 50th, 1987, New York. **Proceedings...**New York: ASIS, 1987. p. 71-73.

----- Moves in online searching. *Online Review*, v. 9, n. 1, p. 61-74, 1985.

----- Online searching styles: a case-study-based model of searching behavior. *JASIS*, v. 35, n. 4, p. 211-221, 1984.

----- The possible effect of abstracting guidelines on retrieval performance of tree-text searching. *Information Processing and Management*, v. 22, n. 4, p. 309-316, 1986.

----- Qualitative methods in information retrieval... research. *Library and Information Science Research*, v. 15, n. 3, p. 219-247, 1993.

----- Searchers selection of search keys: I. The selection routine. *JASIS*, v. 42, n. 7, p. 490-500, 1991.

----- Searchers selection of search keys. II. Controlled vocabulary of free-text searching. *JASIS*, v. 42, n. 7, p. 501-514, 1991.

----- Searchers selection of search keys. III. Searching styles. *JASIS* v. 42, n. 7 p. 515-527, 1991.

- Towards expert systems for the selection of search keys. *JASIS*, v. 37, n. 1, p. 37-44, 1986.
- Writing abstracts for free-text searching. *Journal of Documentation*, v. 42, n. 1, p. 11-21, March 1986.
- FIDEL, R., SOERTEL, D. Factors affecting online bibliographic retrieval: a conceptual framework for research. *JASIS*, v. 34, n. 3, p. 163-180, 1983.
- HARTER, S.P. Online searching styles: an exploratory study. *College & Research Libraries*, v. 45, n. 4, p. 249-258, July, 1984.
- Online Information retrieval: concepts, principles and techniques. London : Academic Press, 1986. p. 22-63.
- Scientific inquiry: a model for online searchig. *JASIS*, v. 35, n. 2, p. 110-117, Mar. 1984.
- Search term combinations and retrieval overlap: a proposed methodology and case study. *JASIS*, v. 41, n. 2, p. 132-146, 1990.
- HARTER, S. P., CHENG, Y. R. Colinked descriptors: improving vocabulary selection for end-user searching. *JASIS*, v. 47, n. 4, p. 311-25, Apr. 1996.
- HAWKINS, D. T., WAGERS, R. Online bibliographic search strategy development. *Online*, v. 6, n. 3, p.10-15, May 1982.
- HENZLER, R. G. Free or controlled vocabularies: some statistical user-oriented evaluations of biomedical information systems. *International Classification*, v. 5, n. 1, p. 21-26, 32, Mar. 1978.
- HOLLEY, R. et al. (Ed.). *Subject indexing: principles and practices in the 90's.* Munchen : K.G. Sauer, 1995.
- HURYCH, Jitka. The professional and the client: the reference interview revisited. *Reference Librarian*, n. 5/6, p. 199-205, Fall/Winter 1982.
- JAHODA, G., BRAUNAGEL, J. S. The librarian and reference queries: a systematic approach. New York : Academic Press, 1991.
- JANKE, R. V. Presearch counseling for client searchers (end-users). *Online*, v.8, n. 9, p. 13-26, Sept. 1985.

- KABACK, S.M. Online Patent Information: **who** needs indexing? **We** do, **naturally**. *World Patent Information*, v. 14, n. 3, p. 198-199, **Aug.** 1992.
- KENNEDY, L, COLE, C. **Connecting** online search **strategies** and **information** needs: a user-centered, **focus-labeling** approach. *RQ*. v. 36, n. 4, p. 562-568, Summer 1997.
- KIRKBRIDE, P. **Full text**, free **text** and controlled **vocabulary**: **strategic** search planning. IN: **ONLINE/CD-ROM 91**. Chicago : Online Inc. 1992. p. 73-78.
- KNAPP, Sara D. Free-text searching of online databases. *Reference Librarian*,n. 5/6, p. **143-153**, **Fall/Winter 1982**.
- KNAPP, S. D., COHEN, L.B., JUEDES, D.R. A natural language **thesaurus** forthe **humanities**:the need for a database search aid. *LibraryQuarterly*,v. 68, n. 4, p. **406-30**, **Oct. 1998**.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brasília, **Briquet de Lemos/Livros**, 1993.
- . *Information retneval systems: characteristics, testing and evaluation*. 2nd.ed., New York **Wiley**, 1979,
- . *Vocabularycontrol for information retrieval*. 2nded. 1986. Arlington : **IRP**, 1986.
- LANCASTER, F.W., FAYEN, E.G. *Information retrieval on-line*. Los Angeles: **Melville Publishing**, 1973.
- LANCASTER, F. W., ELLIKER, C., CONNEL, T. H. *Subject analysis*. IN: **ARIST**, v. 24, p. 35-93, 1989.
- MARCONI, M. A., LAKATOS, E. M. *Técnicas de pesquisa*. 2ed. rev. e ampl. São Paulo : **Atlas**, 1990. Cap. 3.
- MARKEY, K., ATHERTON, P. *Online training and practice manual for ERIC data base searchers*. Syracuse: **Syracuse University**, 1978.
- MARKEY, K., ATHERTON, P., **NEWTON**, C. An analysis of controlled vocabulary and free text search **statements** in online searches. *Online Review*, v. 4, n.,**3**,p. 225-323, Sept. 1980.
- MANSUR, O. Free text and controlled vocabulary, *Information Processing & Management*, v. **16**,n. 4, p. 129-37, 1980.

- MUDDAMALLE, M. R.. Natural language **versus controlled vocabulary** in information **retrieval**: a case study in soil mechanics. **JASIS**, v. 49, n. 10, p. 881-887, Oct. 1998.
- NAHL, D., TENOPIR, C. **Affective and cognitive searching behavior** of novice end-users of a **full-text database**. **JASIS**, v. 47, n. 4, p. 276-86, Apr. 1996.
- NICHOLLS, P., HOLTSMANN, S. **Women's issues searching with DIALOG ERIC**: natural language and controlled vocabulary **strategies**. **Laserdisk Professional**, v. 2, n. 3, p. 97-103, 1989.
- NICHOLAS, D. *Are information professionals really better online searchers than end-users?* Oxford : Leamed information, 1996. p. 383-387.
- OJALA, M. The **rhythm of searching**. *Online*, v. 20, n. 4, p. 52-54, July/Aug. 1996.
- . **Who's hosting this search?** *Online*, v. 19, n. 4, p. 88-92, 1995.
- PEREZ, Ernest. **Text enhancement**: controlled **vocabularies** vs free text. **Special Libraries**, v. 72, n. 3/4, p. 183-192, 1982.
- PINTO, M., LANCASTER, F. W. **Abstracts and abstracting in knowledge discovery**. **Library Trends**, v. 48, n. 1, p. 234-248, Summer 1999.
- PITERNICK, Anne B. **Searching vocabulaires**: a **developing category** of online search **tools**. *Online Review*, v.8, n. 5, p. 441-453, Oct. 1984.
- QIN, Jian **Semantic similarities between a keyword database and a controlled vocabulary database**: na investigation in the antibiotic **resistance literature**. **JASIS**, v. 51, n. 3, p. 166-180, Mar. 2000.
- RAITT, D. J. **Recall and precision devices in interactive bibliographic search and retrieval systems**. **Aslib Proceedings**, v. 32, n. 7/8, p. 281-301, July/Aug. 1980.
- RICHARDSON, R. J. et al. *Pesquisa social: métodos e técnicas*. 2.ed. São Paulo : Atlas, 1989. Cap.11.
- ROTHMAN, J. **Is indexing obsolete? Keyword indexing and free-text searching**. IN: FEINBERG, H. *Indexing specialized formats and subjects*. London : Scarecrow Press, 1983. Cap. 2, p. 22-34.
- ROWLEY, J. E. **A comparison between free language and controlled language indexing and searching**, *Information Services & Use*, v. 10, n. 6, p. 147-155. 1990.

- . The **controlled versus natural indexing** languages debate revisited: a perspective on information **retrieval practice** and research. *Journal of Information Science*, v. 20, n. 2, p. 108-119, 1994.
- SCHABAS, A. H. **Postcoordinate** retrieval: a comparison of **two indexing languages**. *JASIS* v. 33, n.1, p. 32-37, Jan. 1982.
- SCHRODER, J. J. Study of strategies used in online searching: 4 Acronyms - the missing element in your searching? *Online Review*, v. 7, n. 6, p. 475-83, Dec. 1983.
- SCHIFFMANN, G.N. "Hedges"... the **sometimes ignored search** technique that can save a lot of time. *Online*, v. 9, n. 6, p. 40-41, Nov. 1985.
- SCWARZWALDER, R. **Adding value** to your online results. *Database*, v. 20, n. 1, p. 47-49, Feb/March, 1997.
- SIEVERT, M. E., BOYCE, B. R. Hedge **trimming** and the **resurrection** of the controlled **vocabulary** in online searching. *Online Review*, v. 7, n. 6, p. 489-495, Dec. 1983.
- SIMON, H. A., VALDES-PEREZ, R. E. **Scientific discovery** and **simplicity** of method. *Artificial Intelligence*, v. 91, n. 2, p. 183-203, Apr. 1997.
- SMITH, L. C. Artificial **intelligence** in information retrieval. *Information Processing & Management*, v. 12, p. 189-222, 1976.
- SOMERVILLE, A. The place of the **reference** interview in **computersearching**: the **academic setting**. *Online*, v.1, n. 4, p. 14-27, 1977.
- SPINK, A e SARACEVIC, T. Dynamics of search **term selection** during mediated online searching. IN: ASIS ANNUAL MEETING, 56th, 1993. Columbus, OH. *Proceedings...New York: ASIS* , 1993. v. 30, p. 63-72.
- . Sources and use of search **terms** in online searching. IN: ASIS ANNUAL MEETING, 55TH, 1992. Pittsburg, Proceedings,,, **Pittsburg: ASIS**, 1992. P.249-255.
- **Interaction** information retrieval: selection and **effectiveness** of search terms. *JASIS*, v. 48, n. 8, p. 741-761, Aug. 1997.
- STILL, J.A A comparison of online search **strategy formation** as presented in **British and American** tecbooks. *Microcomputers for Information Management*, v. 13, n. 3-4, p. 301-10, 1996.

SU, Louise T. The **relevance** of **recall** and precision in user **evaluation**. *JASIS*, v. 45, n. 3, p. 207-217, Mar. 1994.

----- **Value** of search results as a **whole** as a measure of information **retrieval** performance. IN: **ASIS ANNUAL MEETING**, 59th, 1996, Baltimore, **Maryland**. **Proceedings...** Baltimore: **ASIS**, 1996. V. 33, p. 226-237.

SVENONIOUS, E. Design of controlled **vocabularies**. IN: *ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE*. New York : Mareei Dekker, 1988. v. 45, **supp.10**, p.82-108.

----- Design of controlled vocabularies in the context of **emerging** technologies. *Library Science*, v. 25, n. 4, p. **215-227**, Dec. 1988.

----- **The intellectual foundation of information organization**. Cambridge: MIT Press, 2000.

----- Natural language vs controlled vocabulary. IN: CANADIAN CONFERENCE ON INFORMATION SCIENCE, 4th, **Ontario**, 1976. **Proceedings...** p. 141-150.

----- Unanswered **questions** in the design of controlled vocabularies. *JASIS* v. 37, n. 5, p. 331-340, 1986.

----- Use of **classification** in online retrieval. *Library Resources & Technical Services*, v. 27, n. 1, p. **76-80**, Jan./Mar. 1983.

----- Vocabularies for online **subject** searching. IN: *ENCYCLOPEDIA OF LIBRARY AND INFORMATION SCIENCE*. New York : Mareei Dekker, 1988. v. 45, **supp. 10**, p. **399-420**.

TENOPIR, C. **Full text** database retrieval performance. *Online Review*, v. 9, n. 2, p. 149-164, Apr. 1985.

----- Searching by controlled vocabulary or free text? *Library Journal* v. **119**, n. **15**, p. 58-59, 1987.

----- Searching **Harvard Business Review**. Online... lessons in searching a **full text** database. *Online*. v. 9, n. 2, p. 71-78, Mar, 1985.

TENOPIR, C., NAHL, J. D., HOWARD, D. L. Strategies and assessments online: **novices's experience**. *Library and Information Science Research*, v. **13**, n. 3, p. **237-66**, Jul./Sep. 1991.

TRIVIÑOS, A. N. S. *Introdução à pesquisa em Ciências Sociais: a pesquisa qualitativa em Educação*. São Paulo : Atlas, 1992. Cap. 4

THE UNISIST draft of indexing principles: text and comments. *International Classification*, v. 4, n. 1, p. 29-34, May, 1977.

WALKER, A. , MULHOLLAND, S. N. The Thesaurus of Psychological Index Terms: a historical review. *Behavioral and Social Sciences Librarian*v. 11, n. 2, p. 39-57, 1992.

WARD, D. M. More to research than ERIC: a quick attack on database dependency. *Research Strategies*, v. 14, n. 2, p. 108-11, Spring 1996.

WAGERS, R. Effective searching in database abstracts. *Online*, v. 7, n. 5. p. 60-77, Sept. 1983.

WILDEMUTH, B. M., et al. A detailed analysis of end-user search behaviours. IN: ASIS ANNUAL MEETING, 54th, 1991, New York. Proceedings... New York, ASIS, 1991, v. 28, p. 302-31.

WILDEMUTH, B. M., MOORE, M. E. End-user search behaviors and their relationship to search effectiveness. *Bulletin of the Medical Library Association*, v.. 83, n. 3, p. 294-304, Jul. 1995.

YUAN, W. End-user searching behavior in information retrieval: a longitudinal study. *JASIS*.v. 48, n. 3, p. 218-234, March, 1997.

ANEXOS

ANEXO I

Formulário do Serviço de Busca Bibliográfica



UNIVERSIDADE CATÓLICA DE BRASÍLIA - UCB

SISTEMA DE BIBLIOTECAS - SIBI

PESQUISA BIBLIOGRÁFICA

IDENTIFICAÇÃO DO REQUISITANTE

NOME	MATRICULA	TELEFONE	E-MAIL
CATEGORIA			
<input type="checkbox"/> PROFESSOR UCB	<input type="checkbox"/> PÓS-GRADUAÇÃO UCB	<input type="checkbox"/> DGRADUAÇÃO UCB	<input type="checkbox"/> FUNCIONÁRIO UCB D OUTROS (ESPECIFICAR)
CURSO/LOCAL DE TRABALHO			

DADOS RELATIVOS A PESQUISA

ASSUNTO E RESUMO (informe de maneira sucinta o seu pedido)

Area for writing the subject and summary of the research request.

PALAVRAS-CHAVE

PORTUGUÊS

INGLÊS

IDIOMAS QUE LÊ

INGLÊS FRANCÊS ESPANHOL ALEMÃO OUTROS _____

FINALIDADE DA PESQUISA

DISSERTAÇÃO/TESE MONOGRAFIA PESQUISA OUTROS _____

DATA/ASSINATURA

____/____/____
DATA

FUNCIONÁRIO

BASES CONSULTADAS/NUMERO DE REFERÊNCIAS RECUPERADAS

--

OBSERVAÇÕES/POSIÇÃO DE PENDÊNCIA

DATA ____/____/____	_____
	FUNCIONÁRIO
DATA ____/____/____	_____
	FUNCIONÁRIO
DATA ____/____/____	_____
	FUNCIONÁRIO
DATA ____/____/____	_____
	FUNCIONÁRIO
DATA ____/____/____	_____
	FUNCIONÁRIO
RESPOSTA FINAL ____/____/____	_____
DATA	FUNCIONÁRIO

ANEXO II

Avaliação do Serviço de Busca Bibliográfica - Questionário



UNIVERSIDADE CATÓLICA DE BRASÍLIA - UCB
SISTEMA DE BIBLIOTECAS - SIBI
BIBLIOTECA CENTRAL – BC

Serviço de Pesquisa Bibliográfica

Nº. Pedido: _____

Recebido: _____

Respondido: _____

Atendido por: _____

Usuário: interno

externo

PESQUISA BIBLIOGRÁFICA - AVALIAÇÃO

Nome: _____
Matricula: _____ Curso: _____
Título/Tema: _____

Nº. de Referência(s): _____

Com relação ao seu pedido de busca bibliográfica, o resultado atendeu:

- 100-81% 80-61% 60-41% 40-21% abaixo de 20%

Relacione abaixo suas criticas e Sugestões: _____

Responda, o sucesso do serviço depende da sua contribuição!
Favor devolver este formulário a Biblioteca Central/SCRM.

Assinatura do Usuário: _____ Data: ____/____/____

ANEXO III

Formulário de Avaliação das Referências



UNIVERSIDADE CATÓLICA DE BRASÍLIA – UCB
SISTEMA DE BIBLIOTECAS - SIBI
BIBLIOTECA CENTRAL - BC

Serviço de Pesquisa Bibliográfica

Nº do Pedido: _____

AVALIAÇÃO DAS REFERÊNCIAS

TÍTULO: _____

BASE: _____ Nº DE REF.: _____

ASSINALE PARACADA ITEM:

MB – MUITOBOM

B – BOM

I - IRRELEVANTE

Nº REF.	CONCEITO			Nº REF.	CONCEITO		
1.	MB	B	I	31.	MB	B	I
2.	MB	B	I	32.	MB	B	I
3.	MB	B	I	33.	MB	B	I
4.	MB	B	I	34.	MB	B	I
5.	MB	B	I	35.	MB	B	I
6.	MB	B	I	36.	MB	B	I
7.	MB	B	I	37.	MB	B	I
8.	MB	B	I	38.	MB	B	I
9.	MB	B	I	39.	MB	B	I
10.	MB	B	I	40.	MB	B	I
11.	MB	B	I	41.	MB	B	I
12.	MB	B	I	42.	MB	B	I
13.	MB	B	I	43.	MB	B	I
14.	MB	B	I	44.	MB	B	I
15.	MB	B	I	45.	MB	B	I
16.	MB	B	I	46.	MB	B	I
17.	MB	B	I	47.	MB	B	I
18.	MB	B	I	48.	MB	B	I
19.	MB	B	I	49.	MB	B	I
20.	MB	B	I	50.	MB	B	I
21.	MB	B	I	51.	MB	B	I
22.	MB	B	I	52.	MB	B	I
23.	MB	B	I	53.	MB	B	I
24.	MB	B	I	54.	MB	B	I
25.	MB	B	I	55.	MB	B	I
26.	MB	B	I	56.	MB	B	I
27.	MB	B	I	57.	MB	B	I
28.	MB	B	I	58.	MB	B	I
29.	MB	B	I	59.	MB	B	I
30.	MB	B	I	60.	MB	B	I

ANEXO IV

Bases de dados: Linguagens controladas

ANEXO IV
RELAÇÃO DE BASES DE DADOS E RESPECTIVAS LINGUAGENS CONTROLADAS

BASE	LINGUAGEM CONTROLADA
ABI/INFORM	Classification Code Chart Search Inform
AGRÍCOLA	Agricultural Terms. Extensive revision of the Thesaurus Thesaurus of Agricultural Terms FNIC Controlled Vocabulary
AGRIS	Agrovoc Descriptor Codes
AQUACULTURE	Aquaculture Thesaurus
AQUATIC SCIENCES AND FISHERIES ABS.	ASFA Thesaurus ASFIS Geographic Authority List ASFIS Guide to Names and Acronyms of Organizations ASFIS Subject Categories and Scope Descriptions World List of Aquatic Sciences and Fisheries Serials Titles
ART LITERATURE INT.	RILA Subject Headings
BIOSIS PREVIEWS	Biosis Previews: keywords, concept codes and byosystematic codes
BOOKS IN PRINT	Subject Guide to Books in Print Subject Guide to Childre Books in Print Subject Guide to Forthcoming Books Subject Thesaurus for Bowker Online Satabases
CA SEARCH	CA Headings List CAS Chemical Substance Name Selection Manual CAS Standard Abbreviations CASSI KWOC Index Naming and Indexing of Chemical Substances for Chemical Abs. Natural Language Term List Registry Handbook-Common Names Rotated Title Phrase List Subject Coverage and Arrangemens of Abstracts by Sections
CAB ABSTRACTS	CAB Thesaurus
CANCERLIT	Medicai Subject Headings-Annotated Alphabetical List. Medicai Subject Headings-Tree Structures
CHEMICAL REGULAT.	CRGS Thesaurus.
CHILD ABUSE & NEGL.	Child Abuse and Neglected Thesaurus
CIS INDEX	CIS Online User Guide and Thesaurus of Index Terms

BASE	LINGUAGEM CONTROLADA
COFFEELINE	Coffeeline Search Guide and Thesaurus
CLAIMS PATENTS	Classification Definitions Compound Term List Fragment Term List International Patent Classification Thesaurus of General Terms
COMPENDEX	CAL Code Brochure SHE: Subject Headings for Engineering Subject Heading Guide to Engineering Categories
COMPUTER DATABASE	Online User's Guide to IAC Databases : Thesaurus
CRIS/USDA	Keyword Bank in CRIS Manual of Classification of Agricultural and Forestry Research
DE HAEN DRUG DATA	Abbreviations used in De Haen Databases De Haen Therapeutic Classifications National Adverse Drug Reaction Directory
DOE ENERGY	Guide to Abstracting and Indexing Subject Categories and Scope Subject Thesaurus
ECONOMIC LIT. IND.	index of Economic Articles with Classification System and Schedule
ELECTRIC POWER	EPRI Thesaurus
EMBASE	Guide to Excerpta Medica Classification and Indexing System
ENVIROMENTAL BIB.	EPB Online Vocabulary Aid
ERIC	ERIC Identifier Authority List Thesaurus of ERIC Descriptors
ECER	Thesaurus of ERIC Descriptors for Exceptional Child Educ. Res.
FINIS	Thesaurus of Financial Services Marketing Terms
FLUIDEX	Thesaurus for Fluid Engineering
FSTA	Food-Multilingual Thesaurus FSTA Thesaurus Thesaurus for VITIS - Viticulture and Enology Abstracts
FOREIGN TRADE	Online User Manual with Thesaurus based onn UDC
GEOARCHIVE	Geosaurus: Geosystems Thesaurus of Geoscience
GEOREF	GeoRef Thesaurus and Guide to Indexing GeoRef Serials List and Kwoc Index

BASE	LINGUAGEM CONTROLADA
GPO MONTHLY CAT.	Library of Congress Subject Headings
HARVARD BUSINESS	HBR Online Thesaurus
ISA	Information Science Abstracts User Guide with Controlled vocabulary
INSPEC	INSPEC Classification INSPEC Thesaurus Concordance to the INSPEC Classification
INSURANCE ABS.	Thesaurus of Insurance Abstracts
IPA	International Pharmaceutical Abstracts Search Aid Kit IPA Thesaurus of Subject Terms Trade Name Cross Reference List
LC MARC	Library of Congress Subject Headings
LIFE SCIENCES	Thesaurus of Life Sciences Collection
LINGUISTICS	Linguistics and Language Behavior Abstracts User's Manual LLBA controlled vocabulary terms
MAGAZINE INDEX	Geographic Codes for IAC Databases Subject Guide to IAC Databases
MANAGEMENT CONT.	Subject Guide to IAC Databases
MATERIALS BUS.FILE	Thesaurus of Metallurgical Terms
MATHSCI	Mathematical File: Index of Subject Classification Words Subject Classification Schemes
MEDLINE	Medicai Subject Headings--Annotated Alphabetical List Medicai Subject Headings--Supplementary Chemical Records Medical Subject Headings--Tree Structures Permuted Medicai Subject Headings
METADEX	Metadex Terms Frequency Records Thesaurus of Metallurgical Terms
MIDDLE EAST	Thesaurus of Indexing Terms for the Middle East Abstracts & Index.
NEWSPAPER INDEX	Geographic Codes for IAC Dataabases Subject Guide to IAC Databases
NCJRS	National Criminal Justice Thesaurus
NEWSEARCH	Geographic Codes for IAC Databases Subject Guide to IAC Databases

BASE	LINGUAGEM CONTROLADA
GPO MONTHLY CAT.	Library of Congress Subject Headings
HARVARD BUSINESS	HBR Online Thesaums
ISA	Information Science Abstracts User Guide with Controlled vocabulary
INSPEC	INSPEC Classification INSPEC Thesaurus Concordance to the INSPEC Classification
INSURANCE ABS.	Thesaurus of Insurance Abstracts
IPA	International Pharmaceutical Abstracts Search Aid Kit IPA Thesaurus of Subject Terms Trade Name Cross Reference List
LC MARC	Library of Congress Subject Headings
LIFE SCIENCES	Thesaurus of Life Sciences Collection
LINGUISTICS	Linguistics and Language Behavior Abstracts User's Manual LLBA controlled vocabulary terms
MAGAZINE INDEX	Geographic Codes for IAC Databases Subject Guide to IAC Databases
MANAGEMENT CONT.	Subject Guide to IAC Databases
MATERIALS BUS.FILE	Thesaurus of Metallurgical Terms
MATHSCI	Mathematical File: Index of Subject Classification Words Subject Classification Schemes
MEDLINE	Medicai Subject Headings--Annotated Alphabetical List Medicai Subject Headings--Supplementary Chemical Records Medical Subject Headings--Tree Structures Permuted Medicai Subject Headings
METADEX	Metadex Terms Frequency Records Thesaurus of Metallurgical Terms
MIDDLE EAST	Thesaurus of Indexing Terms for the Middle East Abstracts & Index.
NEWSPAPER INDEX	Geographic Codes for IAC Dataabases Subject Guide to IAC Databases
NCJRS	National Criminal Justice Thesaurus
NEWSEARCH	Geographic Codes for IAC Databases Subject Guide to IAC Databases

BASE	LINGUAGEM CONTROLADA
NTIS	Computer Sciences Microthesaurus COSATI Subject Category List DDC Retrieval and Indexing Terminology Posting Terms with KWOC Energy Information Database: Subject Thesaurus Environmental Microthesaurus Health Care Microthesaurus Medical Subject Headings NASA Thesaurus NTIS Subject Classification Social Science and Business Microthesaurus TEST : Thesaurus of Engineering and Scientific Terms
NURSING&ALLIED HEALTH	CINAHL Subject Heading List
PACKAGING S&T ABS	Thesaurus for Packaging Science and Technology Abstracts
PAIS INTERNATIONAL	PAIS Subject Headings List
PAPERCHEM	Keyword Frequency List Thesaurus of Pulp and Paper Terms
P/E NEWS	Petroleum and Energy News Index Guide/Keyword List
PNI	Pharmaceutical News Index User's Guide with controlled vocabulary
PHILOSOPHER'S	Philosopher's Index Thesaurus
POLLUTION ABS.	Nanogen Index: A Dictionary of Pesticides and Chemical Pollutants
POPULATION BIB.	Population/Family Planning Thesaurus
PREDICASTS	PTS Company Directory PTS MARS Users Guide PTS PROMT Users Guide PTS Users Manual
PSYCINFO	Thesaurus of Psychological Index Terms
RELIGION INDEX	Religion Index Thesaurus
RILM ABSTRACTS	RILM International Thesaurus
SPIN	Energy Database: Subject Thesaurus Physics and Astronomy Classification Scheme
SPORT	The Sport Thesaurus
SUPERTECH	Supertech User's Manual with EIC Keyterm List Telegenline User's Manual with EIC Keyterm List

BASE	LINGUAGEM CONTROLADA
TEXTILE	TEXTILE Technology Digest--Keyterms
THOMAS DATABASES	Thomas Online Thesaurus and Search Guide
TRADE&INDUSTRY	Geographic Codes for IAC Data bases Subject Guide to IAC Databases
TRADE OPPORTUNITIES	Standard Industrial Classification Manual
TRIS	Air Transportation Research Information Service Subject Terms Highway Research Information Service Thesaurus Term List Maritime Research Information Service Subject Term List Railroad Research Information Service Subject Term List
TSCA INITIAL INVENT.	Naming and Indexing of Chemical Substances for Chemical Abstracts Naming Organic Compounds
ULRICH'S	Subject Thesaurus for Bowker Online Databases
US POLITICAL SCIENCE	Political Science Thesaurus
WATER RESOURCES	Abstractings and Indexing Guide The Water Resources Thesaurus
WATERNET	Glossary: Water and Wastewater Control Engineering WATERNET Thesaurus
WELDASEARCH	International Welding Thesaurus
WORLD ALUMINUM	Thesaurus of Aluminum Technology
WORLD PATENTS	International Patent Classification & List of Standard Thesaurus Terms Company Code Manual Derwent CPI Registry Compounds Chemical Code Dictionary Plasdoc Code Dictionary
WORLD TEXTILES	Register of Keyterms
ZOOLOGICAL RECORD	Zoological Record Search Guide

ANEXO V

ESTRATÉGIAS DE BUSCA: DETALHAMENTO

PESQUISA Nº 1 - BASE PSYCLIT

TEMA:

**FENOMENOLOGIA DOS DISTÚRBIOS AFETIVOS DO TIPO:
NEUROSES OBSESSIVA OU COMPULSIVA.**

Nº	Records	Request
1	234	explode "Phenomenology"
2	5197	explode "Affective-Disturbances"
3	526	explode "Obsessive-Compulsive-Neurosis"
4	20	#1 and (#2 or #3)
5	87625	DT="JOURNAL-ARTICLE"
6	13	#4 and (DT = "JOURNAL-ARTICLE")
7	472	phenomenology
8	314	phenomenology in ti, ab
9	2781	affective*
10	2278	disturbance*
11	18953	disorder*
12	896	(affective* and (disturbance* or disorder*)) in ti, ab
13	746	neurosis
14	890	obsessive*
15	1034	compulsiv*
16	7	(neurosis and (obsessive* or compulsiv*)) in ti, ab
17	9	#8 and (#12 or #16)
18	87625	DT = "JOURNAL-ARTICLE"
19	7	#17 and (DT = "JOURNAL-ARTICLE")
20	19	#6 or #19

PESQUISA Nº 2 - BASE PSYCLIT

TEMA:

DOENÇAS RENAIIS: ASPECTOS PSICOLÓGICOS DOS PACIENTES RENAIIS.

Nº	Records	Request
1	44	explode "Kidney-Diseases"
2	31	explode "Dialysis"
3	15	#2 and #1
4	1523	explode "Anxiety"
5	2	#4 and #1
6	3907	"Major-Depression" in DE
7	4	#6 and #1
8	18	#3 or #5 or #7
9	87625	DT = "JOURNAL-ARTICLE"
10	16	#8 and (DT = "JOURNAL-ARTICLE")
11	80	kidney
12	103	renal
13	119	(kidney or renal) in ti, ab
14	52	dialysis
15	4855	anxiet*
16	8311	depressi*
17	10096	(dialysis or anxiet* or depressi*) in ti, ab
18	40	#13 and #17
19	87625	DT = "JOURNAL-ARTICLE"
20	36	#18 and (DT = "JOURNAL-ARTICLE")
21	41	#10 or #20

PESQUISA Nº 3 - BASE PSYCLIT

TEMA:

DENEGAÇÃO E RELIGIOSIDADE

(ESPÉCIE DE MECANISMO DE DEFESA QUE INCLUI COMO TIPOS ESPECÍFICOS A INTELLECTUALIZAÇÃO, A RACIONALIZAÇÃO E A SUBLIMAÇÃO)

Nº	Records	Request
1	50	explode "Denial"
2	513	explode "Defense-Mechanisms"
3	1445	rational*
4	1255	intellectual*
5	46	sublimat*
6	154	(rational* or intellectual* or sublimat*) in de
7	631	#6 or #1 or #2
8	1533	relig*
9	6	#7 and (relig* in de)
10	136	explode "Christianity"
11	322	explode "Spirituality"
12	3	#7 and (#10 or #11)
13	8	#12 or #9
14	317	denial
15	296	denial in ti, ab
16	890	defense
17	1833	mechanism*
18	66	(defense near mechanism*) in ti, ab
19	1445	rational*
20	1255	intellectual*
21	46	sublimat*
22	2502	(rational* or intellectual* or sublimat*) in ti, ab
23	2836	#15 or #18 or #22
24	1533	relig*
25	1283	relig* in ti, ab
26	701	christian*
27	708	spiritual*
28	817	(christian* or spiritual*) in ti, ab
29	1802	#25 or #28
30	96	#23 and #29
31	87625	DT = "JOURNAL-ARTICLE"
32	70	#30 and (DT = "JOURNAL-ARTICLE")
33	74	#13 or #32
34	72	#33 and #31

PESQUISA Nº 4 - BASE PSYCLIT

TEMA:

AGRESSIVIDADE NA INFÂNCIA E NA ADOLESCÊNCIA

Nº	Records	Request
1	791	aggressive-behavior
2	757	aggressive-behavior in de
3	14865	AG = "CHILDHOOD"
4	250	#2 and (AG = "CHILDHOOD")
5	12828	AG = "ADOLESCENCE"
6	114	#4 and (AG = "ADOLESCENCE")
7	87625	DT = "JOURNAL-ARTICLE"
8	95	#6 and (DT = "JOURNAL-ARTICLE")
9	1677	aggressive*
10	23716	behav*
11	586	(aggressiv* and behav*) in ti, ab
12	179	#11 and #3
13	74	#12 and #5
14	87625	DT = "JOURNAL-ARTICLE"
15	64	#13 and (DT = "JOURNAL-ARTICLE")
16	127	#8 or #15

PESQUISA Nº 5 - BASE PSYCLIT

TEMA:

DIVÓRCIO NA MEIA IDADE: CRISES E CONFLITOS

Nº	Records	Request
1	294	explode "divorce"
2	23	explode "Divorced-Persons"
3	309	#2 or #1
4	16457	AG = "MIDDLE-AGE"
5	39	#3 and (AG = "MIDDLE-AGE")
6	609	divorce*
7	525	divorce* in ti, ab
8	85	#4 and #7
9	88	#5 or #8

PESQUISA Nº 6 - BASE PSYCLIT

TEMA:

CASAMENTO E SATISFAÇÃO CONJUGAL

Nº	Records	Request
1	1011	marriage
2	375	marriage in de
3	1591	marital
4	2466	satisfaction
5	134	marital satisfaction in de
6	24	#5 and #2
7	663	#1 in ti, ab
8	1591	marital
9	2466	satisfaction
10	199	(marital and satisfaction) in ti, ab
11	61	#10 and #7
12	87625	DT = "JOURNAL-ARTICLE"
13	49	#11 and (DT = "JOURNAL-ARTICLE")
14	69	#13 or #6

PESQUISA Nº 7 - BASE PSYCLIT

TEMA:

**DEPENDÊNCIA QUÍMICA: ATITUDES E FUNÇÕES FAMILIARES EM
RELAÇÃO À DROGA.**

Nº	Records	Request
1	426	explode "Drug-Addiction"
2	2849	parental
3	9293	family
4	10451	attitude*
5	14197	role*
6	13953	function*
7	24147	relation*
8	39	permissiveness*
9	9542	structur*
10	2878	((parental or family) and (attitude* or role* or function* or relation* or permissiveness* or structur*)) in de
11	9	#1 and #10
12	867	explode "Drug-Dependency"
13	11	#10 and #12
14	12	#13 or #11
15	87625	DT = "JOURNAL-ARTICLE"
16	11	#14 and (DT = "JOURNAL-ARTICLE")
17	10866	drug
18	1333	addiction
19	955	dependency
20	354	(drug and (addicion or dependency)) in ti, ab
21	2849	parental
22	9293	family
23	10451	attitude*
24	13953	function*
25	24147	relation*
26	39	permissiveness*
27	9542	structur*
28	4181	((parental or family) and (attitude* or function* or relation* or permissiveness* or structur*))in ti, ab
29	14	#20 and #28
30	25	#16 or #29

PESQUISA N° 8 - BASE PSYCLIT

TEMA:

ADOÇÃO DE CRIANÇAS: PROBLEMAS DE COMPORTAMENTO

N°	Records	Request
1	1428	adopt*
2	18273	child*
3	123	adopt* child* in de
4	23716	behav*
5	13054	problem*
6	18953	disorder*
7	13312	change*
8	2689	(behav* and (problem* or disorder* or change*)) in de
9	15	#3 and #8
10	1428	adopt*
11	18273	child*
12	310	(adopt* and child*) in ti, ab
13	4744	(#4 and (#5 or #6 or #7)) in ti, ab
14	39	#12 and #13
15	87625	DT = "JOURNAL-ARTICLE"
16	33	#14 and (DT = "JOURNAL-ARTICLE")
17	12	#9 and #15
18	37	#17 or #16

PESQUISA N° 9 - BASE PSYCLIT

TEMA:

ASPECTOS PSICOLÓGICOS DA OBESIDADE

N°	Records	Request
1	204	explode "Obesity"
2	627	explode "Self-esteem"
3	1950	explode "Self-concept"
4	99	explode "Self-confidence"
5	762	explode "Emotional-adjustment"
6	161	explode "Emotional-control"
7	8	#1 and (#2 or #3 or #4 or #5 or #6)
8	309	explode "Body-image"
9	20	#1 and #8
10	25	#9 or #7
11	87625	DT = "JOURNAL-ARTICLE"
12	21	#10 and DT = "JOURNAL-ARTICLE")
13	319	obesity
14	206	obesity in ti, ab
15	14299	self
16	1383	esteem
17	4513	concept
18	834	confidence
19	2238	(self and (esteem or concept or confidence))in ti,ab
20	5	#19 and #14
21	7058	emotional
22	2592	adjustment
23	10168	control
24	650	(emotional and (adjustment or control)) in ti, ab
25	2	#24 and #14
26	7	#25 or #20
27	3652	body
28	1459	image
29	262	body image in ti, ab
30	11	\$29 and #14
31	17	#30 or #26
32	31	#31 or #12

PESQUISA N° 10 - BASE PSYCLIT

TEMA:

DEPRESSÃO EMOCIONAL E STRESS NA MEIA IDADE

N°	Records	Request
1	381	explode "Depression, - Emotion"
2	2655	explode "Stress"
3	59	#2 and #1
4	16457	AG = "MIDDLE-AGE"
5	22	#3 and (AG = "MIDDLE-AGE")
6	87625	DT = "JOURNAL-ARTICLE"
7	22	#5 and (DT = "JOURNAL-ARTICLE")
8	7772	depression
9	8907	emotion*
10	357	(depression and emotion*) in ti, ab
11	5580	stress
12	4646	stress in ti, ab
13	41	#12 and #10
14	16	#13 and #4
15	34	#14 or #7