

Cabeçalho de assunto como linguagem de indexação

Subject Headings as Indexing Language

MARIA AUGUSTA DA NÓBREGA CESARINO *

MARIA CRISTINA MELLO FERREIRA PINTO *

O papel das linguagens de indexação como instrumento controlador de vocabulário nos sistemas de recuperação de informação. Características dos cabeçalhos de assunto como linguagem de indexação: pré-coordenação, vocabulário controlado, não hierarquia, linguagem enumerativa, função prescritiva, aplicação a SRI, unidimensionais, agrupamento alfabético dos termos, comparação entre o cabeçalho de assunto e outras linguagens de indexação quanto à expressividade, ambigüidade, compacidade e ao custo. Comentários sobre as listas de cabeçalhos de assunto mais conhecidas.

1. INTRODUÇÃO

Ao darem entrada em qualquer sistema de recuperação de informação os documentos serão analisados de duas maneiras:

* Professoras da Escola de Biblioteconomia da UFMG

a) bibliograficamente ou objetivamente — este tipo de análise pretende a descrição do documento através de suas características físicas, com o objetivo de dar resposta à questão: «Qual a aparência física deste documento?»

b) intelectualmente ou subjetivamente — este tipo de análise pretende a descrição do documento em termos de suas características de conteúdo, com o objetivo de dar resposta à questão: «Sobre o que é este documento?»

Para produzirmos uma descrição única de um documento é necessária a descrição de suas características físicas baseando-nos num conjunto ordenado de dados bibliográficos. A descrição bibliográfica ou objetiva deve resultar idêntica, apesar do documento poder ser processado em diferentes lugares. Isto por que a descrição é baseada em fatores facilmente identificáveis e que por si representam a obra a que se referem. Por exemplo:

- o número de especificação de uma patente
- o autor ou o título de uma obra
- o local, editor e data de publicação

são fatores que realmente identificam uma obra. É óbvio que existem problemas relativos ao processamento de descrição física de obras:

- vários autores
- obras anônimas
- títulos alternativos
- etc.

No entanto, poderemos conseguir um controle desses problemas através da adoção de regras formais que nos indicarão soluções adequadas a cada tipo de problema. A formulação de regras para descrição bibliográfica foi iniciada por Cutter, em 1876, com sua obra «Rules for a Dictionary Catalogue». Esta obra serviu de base para os modernos códigos de catalogação que hoje são usados.

Mas o problema maior permanece na descrição do conteúdo dos documentos. Talvez por ser menos reduzível a regras e por trabalharmos com fatores que não identificam realmente uma determinada obra. Um usuário que peça a um bibliotecário um livro sobre «plantações de café» não está realmente identificando uma obra «X», mas dando características que podem ser comuns a várias delas. Este processamento de análise e determinação do conteúdo informativo de obras é comumente chamado **indexação**. Para Hilda Feinberg (2) «indexação consiste na indicação do assunto contido em um documento e a determinação de um ou mais **termos** que representarão o assunto desse documento».

A palavra **termo** é usada de forma generalizada para representar qualquer forma de classe, subclasse, cabeçalhos de assunto, unitermos, palavras ou frases compostas.

2. A INDEXAÇÃO E O CONTROLE DE VOCABULÁRIO

As palavras não têm significado preciso e apesar disto, na recuperação de informações temos de agir como se tivessem. Por isso, se impõe a necessidade de reduzirmos ao máximo o elemento «interpretação pessoal». A determinação do significado das palavras deve passar da fase da indexação para a da construção do vocabulário. Considerando a operação efetiva de um sistema de informações, um aspecto-chave é que, se uma informação não foi adequadamente identificada ao entrar no sistema, ela não será nunca recuperada.

Para Lancaster (6) o termo «Recuperação de Informações» é mal empregado, por que os sistemas de recuperação de informações são procedimentos para recuperar documentos, já que, para ele as informações são intangíveis. O termo Recuperação de Informações foi primeiramente empregado por Moors em 1950.

Todo procedimento de Recuperação de Informações é ligado à manipulação de «classes». Quando indexamos um documento, estamos colocando-o em uma classe determinada. Para facilitar o processo, cada classe recebe um «nome», que é chamado «termo indexador». Ao conjunto de termos indexadores chamamos Linguagens de Recuperação de Informações ou Linguagens de Indexação.

Para Vickery (12) podemos identificar três estágios básicos no processo de indexação:

1º) estágio analítico ou interpretativo, onde o indexador identifica tópicos na obra;

2º) um estágio de sumarização, onde estabelecemos a importância dos tópicos tendo em vista:

- o objetivo do serviço de informação e necessidade dos usuários;
- objetivos para os quais o texto foi escrito.

3º) um estágio de tradução, onde os tópicos selecionados são transformados nos termos da linguagem de indexação usada.

Da mesma forma, quando vamos efetuar uma busca num sistema, Lancaster (7) visualiza três estágios:

1º) decidimos quais as classes que, provavelmente, nos servirão, ou seja, selecionamos termos indexadores prováveis de nos serem úteis;

2º) examinamos essas classes;

3º) recuperamos alguns ou todos os membros.

O objetivo da indexação é, pois, em essência, a transformação de uma expressão da **Linguagem Natural** em uma **Linguagem de Indexação**. Parte de um texto compreendido por palavras, frases, parágrafos em seqüência linear e termina numa condensação do todo em linguagem documental. A eficiência dos sistemas de recuperação de infor-

mações depende muito da estrutura e composição dessa linguagem. Se ela for composta de classes muito abrangentes, gerais, o sistema terá um grau muito superficial de especificidade, levando conseqüentemente a uma alta revocação e baixa precisão. Já se a linguagem se compuser de um número muito grande de classes, estas serão muito específicas e o sistema terá um maior grau de precisão e conseqüentemente baixará o grau de revocação.

Já se tentou o uso de sistemas sem controle do vocabulário. A idéia foi de Mortimer Taube, quando propôs seu sistema unitermo, ou sistema de indexação coordenada. No entanto, vários foram os fatores que levaram a um consenso geral da necessidade de controle de vocabulário nos sistemas de recuperação de informações:

a) fatores humanos — diferenças de cultura, de experiência dos autores e domínio de terminologia entre indexadores, autores e usuários;

b) fatores referentes à própria linguagem natural onde:

- vários termos podem designar o mesmo conceito (sinônimos);
- um mesmo termo pode designar conceitos diferentes (homógrafos);
- o sentido dos termos se alteram pela posição que ocupam na frase (sintaxe);

c) fatores hierárquicos — um conceito implica em outros mais amplos e/ou mais restritos.

Normalmente esse controle tem sido exercido através de:

- seleção de termos preferidos;
- distinção dos homógrafos;
- estabelecimento formal das relações entre os termos.

3. O CABEÇALHO DE ASSUNTO COMO LINGUAGEM DE INDEXAÇÃO

Todas as linguagens de indexação exercem a mesma função nos sistemas de recuperação da informação:

- a) representam o assunto de uma maneira consistente;
- b) permitem a coincidência entre linguagem do indexador e do pesquisador;
- c) possibilitam ao indexador modular o nível de sua pesquisa, do mais geral para o mais específico, e vice-versa, segundo a necessidade do usuário.

Embora o objetivo de qualquer linguagem de indexação seja o mesmo, a maior ou menor eficiência em atingir plenamente tal objetivo está sujeita a variações, segundo as características próprias de cada uma.

Na literatura sobre o assunto, encontramos duas divisões para as linguagens de indexação, partindo de critérios diferentes. A divisão mais comum separa as linguagens em:

- sistemas alfabéticos
- sistemas classificados.

Os sistemas alfabéticos usam termos da própria linguagem natural e por isso determinam uma ordenação alfabética para os arquivos. Os sistemas classificados se baseiam em classificações arbitrárias do conhecimento humano, dando normalmente uma notação simbólica para as classes, e determinam uma ordenação com base lógica, de acordo com os símbolos usados.

Talvez a forma mais antiga de sistema alfabético seja a de cabeçalhos de assunto. O conceito mais difundido de cabeçalhos de assunto o define como «palavra ou grupo de palavras que expressam o conteúdo de um documento».

A primeira forma de organização de bibliografias listava as obras arranjadas de acordo com seus autores. Logo se sentiu a necessidade da listagem de assunto. As primeiras listas de assunto vinham geralmente no final das listagens de autores. Com a sistematização dessas apresentações de assunto, as listas passaram a ser organizadas alfabética ou classificadamente.

Alguns dos fatores que determinaram o aparecimento e formalização dos cabeçalhos de assunto foram:

- os títulos das obras não representavam adequadamente o assunto das mesmas;
- problemas relativos às subdivisões de assunto;
- as obras de mais de um assunto;
- livro de assuntos relacionados;
- obras que relacionavam os assuntos a lugares e épocas diferentes.

As primeiras regras para construção de catálogos alfabéticos de assunto foram elaborados por Cutter, em 1876. Anteriormente eram dados os cabeçalhos de acordo com o julgamento do próprio catalogador.

Para Cutter, se a indexação não fosse feita de acordo com regras formais, de uma maneira consistente não haveria razão pela qual o público devesse encontrar a entrada que necessitasse sob determinado cabeçalho e não sob outro qualquer. Cutter elaborou então princípios fundamentais a serem seguidos na elaboração de um catálogo alfabético de assunto. Podemos resumir esses princípios em três:

1º) **princípio específico** — Os assuntos devem dar entrada pelo termo mais específico e não pela classe a que está subordinado. Apesar de parecer óbvio, este princípio causou impacto porque na época era comum a adoção de cabeçalhos bastante gerais, que pouco representavam o assunto da obra em questão;

2) **princípio de uso** — Para ele «os cabeçalhos serão aqueles sob os quais é provável que a maioria dos americanos educados irão procurar, com referências cruzadas para outras formas de cabeçalhos relacionados». É o princípio da conveniência, de acordo com a necessidade dos usuários;

3º) **princípio sindético** — Por se basearem no alfabeto dos cabeçalhos de assunto fazem aproximações absurdas de assuntos e ao mesmo tempo, separam assuntos relacionados. Assim Cutter propôs o desenvolvimento nas listas de cabeçalhos de assunto de estruturas sindéticas que, através de uma rede bem construída de referências cruzadas, poderiam ajudar ou mesmo superar este problema.

Outra divisão para as linguagens de indexação, classifica-as segundo a sua evolução histórica em:

- linguagens tradicionais
- linguagens modernas.

A característica que realmente separa as primeiras das segundas é o conceito de coordenação que, segundo Lancaster, representa um marco revolucionário nos sistemas de recuperação da informação.

Entre as linguagens conhecidas como tradicionais, podemos citar:

- os cabeçalhos de assunto
- os sistemas de classificação.

Ambos ligados, no seu início, aos instrumentos tradicionais de recuperação da informação dentro da biblioteca: catálogos alfabéticos de assunto e catálogos classificados, bem como a idéia de arranjo dos documentos nas estantes.

As linguagens modernas apareceram a partir de 1950, embora suas primeiras manifestações tenham surgido

antes. São as linguagens pós-coordenadas: coincidência ótica, cartões perfurados nas margens, unitermos, com o seu mais conhecido instrumento de controle: o thesaurus.

3.1. Características dos Cabeçalhos de Assunto como Linguagem

Segundo as abordagens de Toman (10) e Lancaster (7), podemos dizer que, assim como todas as linguagens buscam atingir o mesmo objetivo, apresentam também, em maior ou menor grau, certas características que vão moldar o seu desempenho e permitir que sejam avaliadas e comparadas umas às outras. Essas características são as seguintes:

pré coordenadas	←————→	pós-coordenadas
vocabulário controlado	←————→	vocabulário não controlado
não hierarquia	←————→	hierarquia
enumeração	←————→	síntese
sistema fechado	←————→	sistema aberto
sistema prescritivo	←————→	sistema sugestivo
sistema unidimensional	←————→	sistema multidimensional
agrupamento alfabético dos termos	←————→	agrupamentos sistemático dos termos

Façamos uma descrição dos cabeçalhos de assunto, de acordo com essas características:

1. Os cabeçalhos de assunto representam uma linguagem muito estruturada e pré-coordenada, oferecendo ao pesquisador possibilidades limitadas de modular a sua pesquisa.

Entretanto, existem hoje vocabulários que se intitulam listas de cabeçalhos de assunto e são utilizadas em sistemas pós-coordenados, como é o caso de Mesh: Medical Subject Headings, usado com o sistema MEDLARS. Harris (4) nega ao MeSH a condição de lista de cabeçalhos de assunto, tratando-o como thesaurus.

Lancaster (7) apresenta uma posição mais aberta, acatando a idéia de que pode haver uma evolução teórica necessária sem a indispensável alteração de terminologia.

2. os cabeçalhos de assunto, obedecendo ao princípio de pré-coordenação, representam um modelo de **vocabulário controlado**, onde os termos devem ser escolhidos em um dicionário já existente. São portanto **sistemas fechados** de recuperação da informação;

3. estas duas características conduzem a uma terceira idéia: de uma linguagem onde o instrumento de controle de vocabulário, no caso as listas de cabeçalhos de assunto, exerce uma **função prescritiva**, ao contrário, por exemplo, de um thesaurus que teria apenas uma **função sugestiva**;

4. os cabeçalhos de assunto são sistemas onde o princípio de hierarquia está presente em pouquíssimas ocasiões. São linguagens **não hierárquicas**;

5. os cabeçalhos de assunto são, basicamente, **enumerativos**, oferecendo poucas possibilidades de síntese, já que esta deve estar prevista no instrumento de controle da linguagem. A síntese, nas listas mais tradicionais, é representada pelos sub-cabeçalhos;

6. o arranjo do cabeçalho de assunto é, via de regra, **alfabético**. Esta característica não é inerente à condição de pré e pós-coordenação, embora os thesauri apresentem, quase sempre, paralelamente ao arranjo alfabético, um arranjo sistemático (ou facetado).

Na verdade, o arranjo de uma linguagem é uma característica que exerce pouca influência no desempenho de um sistema de recuperação de informação, sobretudo quando este sistema é automático;

7. como característica final dos cabeçalhos de assunto destaca-se o seu aspecto de **linearidade**, o que o torna aplicável apenas a pesquisas **unidimensionais**.

Isto representa a maior limitação dos cabeçalhos de assunto, face a nova tecnologia aplicada aos sistemas de recuperação da informação.

A partir destas características, podemos comparar algumas linguagens:

	Cabeçalhos de Assunto	Sistema Unitermo (Thesaurus)	Classificação Facetada	Classificação Hierárquica
Pré-coordenação	X	—	X	X
Pós-coordenação	—	X	—	—
Hierarquia	—	—	—	X
Não-hierarquia	X	X	X	—
Sistema fechado	X	—	X	X
Sistema aberto	—	X	—	—
Sistema prescritivo	X	—	X	X
Sistema sugestivo	—	X	—	—
Sistema unidimensional ..	X	—	X	X
Sistema multidimensional ..	—	X	—	—
Arranjo alfabético	X	X	—	—
Arranjo sistemático	—	—	X	X

3.2. Componentes Básicos da Linguagem de Indexação

Segundo Lancaster (7), qualquer linguagem de indexação apresenta os seguintes componentes básicos:

a) vocabulário

b) gramática

c) regras para uso da linguagem e para controle das mudanças a serem efetuadas.

Meadow (8) propõe quatro medidas para estudar as linguagens de indexação: expressividade, ambigüidade, compacidade e custo, e faz uma comparação interessante entre várias linguagens, inclusive os cabeçalhos de assunto. Salienta, entretanto, que as linguagens podem ser muito complexas, difíceis de medir ou comparar, sendo perigosa qualquer generalização.

As figuras 1, 2, 3, 4 e 5 (8) mostram que os cabeçalhos de assunto não permitem muita variabilidade, quer no seu vocabulário ou na sua sintaxe.

Também não apresentam grande riqueza ou expressividade de vocabulário, perdendo para a linguagem natural, para os sistemas facetados e os termos livres de documentos. Em contra-partida são menos ambíguos que estas três linguagens. A constituição do vocabulário é de maior custo do que a linguagem facetada ou de termos livres, sendo, entretanto uma linguagem mais barata do que os sistemas hierárquicos e a linguagem natural.

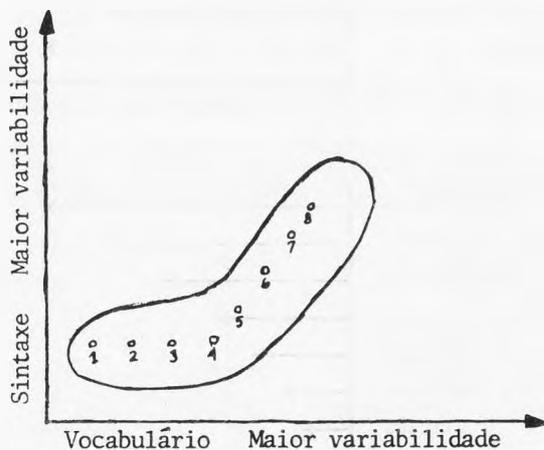


FIGURA 1 — Alcance útil das linguagens: (1) Classificação hierárquica, (2) Cabeçalhos de assunto, (3) Palavras-chave controladas, (4) Palavras-chave livres, (5) Descritores, (6) Termos facetados, (7) Frases, (8) Linguagem Natural. (8:47)



FIGURA 2 — Expressividade Relativa das Linguagens (8:48)



FIGURA 3 — Ambigüidade Potencial Relativa das Linguagens (8:49)

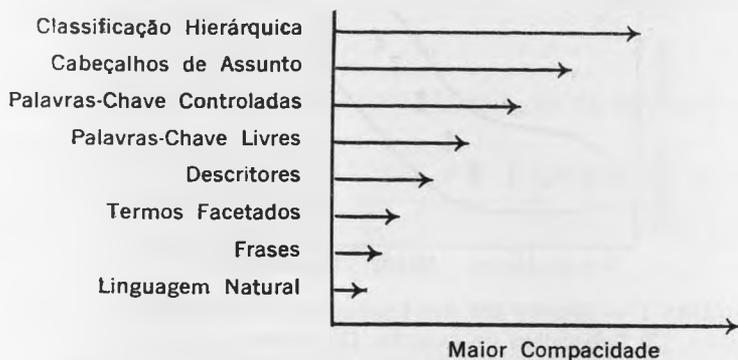


FIGURA 4 — Compacidade Relativa das Linguagens (8:50)

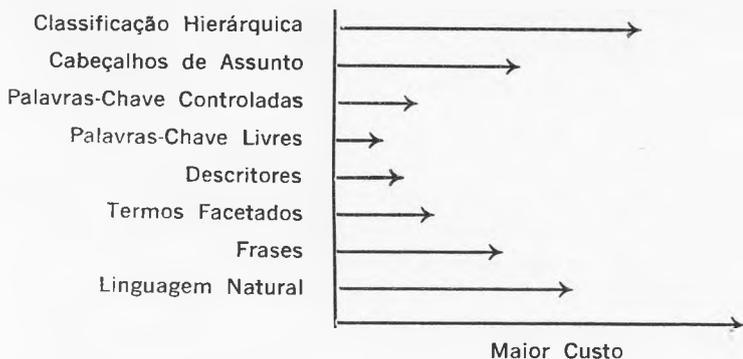


FIGURA 5 — Custo Relativo de Cada Item na Seleção de Termos (8:51)

Os cabeçalhos de assunto apresentam uma gramática bem estruturada, com rígido controle de sinônimos, quase-sinônimos, homógrafos.

A rede de relações que inclui o tipo: **ver também** é mais pobre do que a apresentada em um thesaurus. Uma grande limitação dos cabeçalhos de assunto é que estes fazem referência apenas do geral para o específico e de alguns assuntos correlatos, o que diminui a sua capacidade de revocação.

Por sua vez, o fato de ser pré-coordenado permite atingir maior precisão, dispensando artifícios complexos, onerosos e nem sempre eficientes, necessários em linguagens pós-coordenadas. tais como elos de coordenação e indicadores de função.

Além do vocabulário e da gramática, o cabeçalho de assunto, sendo uma linguagem controlada, prevê regras fixas apresentadas nas listas de controle que dizem respeito, sobretudo a:

- forma de entrada direta ou invertida
- formação de síntese através de subcabeçalho
- uso de singular e plural

- uso de palavras abreviadas ou termos por extenso
- uso de termos traduzidos e transliterados
- uso de sinais de pontuação.

Embora estas regras sejam fixas apenas dentro de um mesmo sistema, podendo variar de um sistema para outro, seguem quase todas o mesmo consenso geral, uma vez que as listas de controle foram feitas baseadas umas nas outras. Este fato é de se lamentar pois indica uma grande passividade por parte daqueles que usam cabeçalhos de assunto.

3.4. As Listas de Cabeçalhos de Assunto

Qualquer linguagem de indexação pode ser elaborada a partir de duas abordagens distintas:

- abordagem analítica, baseada sobretudo na garantia literária e participação do usuário;
- abordagem formal, baseada sobretudo no trabalho e opinião de especialistas.

O que se vê na prática é uma mistura das duas abordagens, com predominância da segunda e, infelizmente, pouquíssima participação dos usuários. Com exceção dos cabeçalhos de assunto elaborados para a Biblioteca do Congresso que apresentam grande garantia literária, os outros sistemas que adotam cabeçalhos de assunto fizeram adaptações de listas já existentes. O sistema de recuperação da informação que escolhe como linguagem o cabeçalho de assunto deve adotar um instrumento de controle. As listas de cabeçalho de assunto são estes instrumentos. Algumas das mais conhecidas e usadas são:

1. **Library of Congress Subject Headings** (Início dos estudos: 1897 — 1ª edição: 1911 — 8ª edição: 1975 —

2 vol.). Reflete o acervo da Biblioteca do Congresso dos EUA, sendo a mais conhecida e difundida naquele país. Embora seja útil para o objetivo específico para o qual foi criada, e o fato de seus cabeçalhos aparecerem nas fichas impressas da LC, contribui para a grande inércia dos indexadores em relação ao uso de cabeçalhos. Prevê-se modificações profundas nesta lista face aos sistemas automatizados;

2. **Sears List of Subject Headings.** Está na 9ª edição (1965) sob a responsabilidade de Westby. Podemos considerar esta lista como uma condensação da lista da Library of Congress, inclusive porque seu objetivo foi o de adaptação da mesma para bibliotecas pequenas e médias. Embora americana, teve mais aceitação nas bibliotecas britânicas.

No Brasil, temos muito difundidos os trabalhos de:

3. C. Rovira e J. Aguayo. **Lista de encabezamientos de materiales para bibliotecas** (1967). É uma adaptação das listas anteriormente citadas para bibliotecas latino-americanas;

4. W. Ferraz. **Relação de assuntos para cabeçalhos de fichas.** (2 ed. 1972. Reeditada em 1977). É uma lista muito limitada, com cabeçalhos gerais, com poucas possibilidades de se atingir um grau razoável de especificidade. Aplicável a pequena coleções;

5. IBBD. **Lista geral de cabeçalhos de assunto.** A idéia inicial deste trabalho era a publicação de uma lista nacional de cabeçalhos, a fim de facilitar a tarefa de uniformização dos cabeçalhos usados pelo SIC (Serviço de Intercâmbio de Catalogação) nas suas fichas bibliográficas. O IBICT retomou o desenvolvimento deste trabalho, visando ao lançamento da edição completa da obra.

Segundo Horner (5), são geralmente quatro os pontos críticos das listas de cabeçalhos de assunto convencionais:

1. não são bastante específicas para assuntos de muitos documentos, principalmente para assuntos que cada dia se tornam mais complexos;

2. os cabeçalhos de assunto não são construídos sistematicamente;

3. as referências também não são construídas sistematicamente. As referências do específico para o geral são usadas ocasionalmente;

4. os arranjos dos subcabeçalhos não são homogêneos.

O que se pode dizer a respeito das listas de cabeçalhos de assunto é que, ao serem adotadas por uma determinada biblioteca, devem ser adaptadas a ela. Não se deve exigir nenhuma rigidez quanto ao emprego de cabeçalhos, e a adaptação desses e a adoção de novos deve ser prática normal num serviço de biblioteca.

Como toda linguagem de indexação, uma lista de cabeçalhos de assunto é composta de vocabulário e sintaxe. E portanto deve apresentar um vocabulário, ou seja, uma lista de termos aprovados para representar assuntos e regras formais para uso destes termos. O vocabulário e a sintaxe irão variar de lista para lista, de acordo com os objetivos do próprio sistema a que irá servir, como também de acordo com as necessidades dos usuários, e usuários em potencial.

Um ponto importante a ser observado nas listas de cabeçalhos de assunto é o grau de especificidade que ela permite.

Por serem linguagens baseadas em pré-coordenação, quanto mais específica for uma lista, maior grau de pré-coordenação irá adotar. No entanto, a previsão de todas as possíveis coordenações de assunto é inviável. Como nos catálogos classificados, para minimizar o problema do acesso unidimensional, poderemos adotar a solução de entradas múltiplas, arcando, no entanto, com os proble-

mas advindos disso: aumento do volume do arquivo e a viabilidade econômica desta solução. Só poderemos chegar a um grau de especificidade grande quando abandonarmos a noção de enumeração.

3.4.1. Listas de Cabeçalhos de Assunto e Thesauri — O aparecimento e a adoção de sistemas pós-coordenados de recuperação de informações influenciaram muito as linguagens de indexação. Esta influência se fez sentir também nas listas de cabeçalhos de assunto. Houve a necessidade de adaptação dos sistemas que usavam linguagem natural para indexação.

Surgiram os thesauri que, para muitos, nada mais são do que a adaptação das listas de cabeçalhos de assunto para uso em sistemas pós-coordenados. Lancaster (7) falando do aparecimento dos thesauri cita organizações que são suficientemente honestas para chamar suas listas autorizadas de listas de cabeçalhos de assunto. É o caso do sistema Medlars, cuja lista se chama Mesh «Medical Subject Headings». Realmente a diferença entre lista de cabeçalhos de assunto e thesauri está mais no modo de se usar as linguagens do que na linguagem propriamente dita. Os thesauris, desenvolvidos para uso em sistemas pós-coordenados, desvincularam seus cabeçalhos de pré-coordenação. Deveríamos alertar os bibliotecários, sedentos de novidades, contra o uso de thesauri em instrumentos tradicionais, como os catálogos de assunto em fichas. Os cabeçalhos dos thesauri, chamados descritores, são muito menos informativos do que os cabeçalhos de uma tradicional lista de cabeçalhos de assunto. E tendo em vista o modo de usar os catálogos em bibliotecas, muito mais eficiente é a adoção dos tradicionais cabeçalhos de assunto, por serem muito mais informativos, usando, como usam, pré-coordenação. Darão ao usuário uma visão muito mais clara do assunto das obras indexadas.

O que se pode pensar em fazer é a adoção da estrutura de um thesaurus na indicação do relacionamento entre termos. Realmente, o relacionamento de termos mostrados nas listas de cabeçalhos de assunto, é muito confuso através de «see also», «ver também». A estrutura adotada pelos thesauri de BT (Broader Term), NT (Narrower Term) e RT (Related Term) é muito mais clara da visualização do tipo de relacionamento. Assim também como o uso de UF (Used For) e US (Use) para termos equivalentes.

4. CONCLUSÃO

Se acompanharmos a evolução histórica dos estudos feitos sobre cabeçalhos de assunto ou aplicáveis a esta linguagem, perceberemos uma tendência muito pragmática dos bibliotecários e lamentamos, como Harris (4) a existência de poucos estudos teóricos sobre o assunto e o grande número de trabalhos que são meros códigos práticos.

Entretanto, de **Cutter** (1876), o pioneiro na tentativa de sistematizar a indexação alfabética de assunto, apresentando seus dois princípios básicos: especificidade e garantia de uso; passando por **Kaiser** (1911) que, em seu «Systematic Indexing» já discute os problemas de ordem de citação nos cabeçalhos compostos e arranjo sistemático nos subcabeçalhos; por **Ranganathan** (1945) que, em seu «Dictionary Catalogue Code» e em trabalhos posteriores, reelaborou a idéia inicial de Kaiser, criando suas famosas categorias fundamentais (Personalidade, Matéria, Energia, Espaço e Tempo) para estabelecer uma ordem de citação nos cabeçalhos compostos (Ranganathan desenvolveu também a idéia do índice em cadeia que minorou a perda de revocação em linguagens formadas de cabeçalhos de assunto); analisando o trabalho de **Coates** (1960) que também apresentou valiosa contribuição teórica (aplicada ao British Technology Index) desenvolvendo a idéia de ca-

tegorias fundamentais e ordem de citação; observando as propostas do **Precis** (Preserved Context Indexing System) desenvolvido pelo Classification Research Group, sob a direção de Derek Austin, para aplicação ao índice de assunto da British National Bibliography, usando cabeçalhos de assunto e operadores lógicos, notamos que novas abordagens têm sido propostas à indexação alfabética de assunto: rotação, permutação, índice em cadeia, operadores relacionais, índices articulados que, à primeira vista, parecem se relacionar apenas a linguagens sofisticadas quando, em essência, são evoluções dos tradicionais cabeçalhos de assunto.

As pesquisas que tentaram avaliar o papel das linguagens de indexação nos SRI (Sistemas de Recuperação da Informação) mostram que, com a tendência de proliferação dos sistemas automatizados, as linguagens estão passando de uma grande sofisticação para linguagens mais simples. Com isto, se considerarmos cabeçalhos de assunto como linguagem estritamente pré-coordenada, não vemos muito futuro para eles.

Entretanto, se adotarmos uma posição mais elástica (aceitando, por exemplo, o Medical Subject Headings como cabeçalhos de assunto), veremos que a tendência futura é a de se fundir o que há de melhor nos sistemas pré e pós-coordenados. O Mesh representa uma linguagem tradicional que evoluiu, no sentido de incorporar a flexibilidade da pós-coordenação. Isso representa um vocabulário mais elástico embora bem menor.

As linguagens construídas automaticamente, embora ainda apresentem muitas falhas, também caminham no sentido de maior flexibilidade, com termos livres e com pós-coordenação.

Indexing languages as a device for the control of vocabulary in the information retrieval systems. The characteristics of subject headings as

an indexing language: pre-coordination, controlled vocabulary, no hierarchy, enumerative language, prescriptive function, application to onedimensional systems, and alphabetic grouping of terms, comparison between subject heading and others indexing languages in relation to expressiveness, ambiguity, compactiveness and cost. Comments on the main subject heading lists.

5. BIBLIOGRAFIA

1. CONRAD, C. C. Status of indexing and Classification systems and potential future trends. I. *Chem. Inf. Comput. Sci.*, 16(4):197-201, Nov. 1976
2. FEINBERG, Hilda *The derivative indexing techniques* Metuchen, Scarecrow, 1973.
3. FOSKETT, A. C. *A abordagem temática da informação* São Paulo, Polígono, 1973.
4. HARRIS, J. D. *Subject analysis; computer implications of rigorous definition*. Metuchen, Scarecrow, 1970
5. HORNER, John. Criticism and evaluation of «conventional» subject headings listas. In: — *Cataloguing*. London, Association of Assistant Librarians, 1970, p. 136-42.
6. LANCASTER, F. W. & OWEN, J. M. *Information retrieval by computers*. In: HAMMER, D. F. *The information age* Metuchen, Scarecrow, 1976
7. LANCASTER, F. W. *Vocabulary control for information retrieval*. Washington, Information Resource Press, 1972.
8. MEADOW, C. T. *The analysis of information systems* 2ed Los Angeles, Melville, 1973.
9. NEEDHAM, C. D. *Organizing knowledge in libraries* 2ed London, Andre Deutsch, 1973
10. TOMAN, J. The influence of information retrieval on the structure of indexing and classification systems. In: FOSKETT, D. J., ed *Library systems and information services*. London, C. Lockwood, 1970. p. 57-66.
11. WELLISCH, H. & WILSON, T. *Subject retrieval in the seventies*. Westport, Greenwood, 1972.
12. VICKERY, B. C. *Techniques of information retrieval*. London, Butterworths, 1970.