

# Construção de linguagens de indexação: aspectos teóricos

ALEXIS RIVIER \*

PALAVRAS-CHAVE: Linguagens de Indexação

Esta contribuição é resultado de uma pesquisa sobre fundamentos teóricos das linguagens de indexação em estudos existentes e na literatura especializada. Foram reunidas mais de cem referências, transcritas no fim do artigo. Depois de apresentar seu método de investigação, o autor examina, sintetizando, os trabalhos escolhidos, a evolução, em quarenta anos, dos princípios teóricos que regem a concepção das classificações, das linguagens de indexação em cadeia e dos tesouros, bem como as implicações das teorias lingüísticas e matemáticas nas linguagens de indexação.

De acordo com alguns autores, como por exemplo J. Maniez, os estudos teóricos sobre os fundamentos das linguagens de indexação (LI) não mereceram a atenção devida dos pesquisadores. Considerando-as simplesmente como um instrumento, julga-se que elas se justificam pela sua eficiência. A literatura apresenta grande quan-

---

\* Alexis Rivier é especialista em Ciência da Informação em Lausanne, Suíça. Colaborou na informatização da Bibliothéque Cantonale du Valais, à Sion.

Este artigo é o resumo de um DESS em informação documentária, preparado e defendido na Université de Lyon 1, sob a orientação de Richard Bouché.

Tradução de Marysia Malheiros Fiuza e Maria Cristina Mello Ferreira Pinto.

tidade de manuais práticos para construção de LI, bem como sobre avaliação de seu desempenho, enquanto que os textos sobre princípios teóricos são menos numerosos e menos conhecidos. Este fato já é uma razão suficiente para que se procure reuni-los. Além disto, percebe-se claramente, pela leitura destes textos, que um melhor conhecimento interno das LI leva muitas vezes a aperfeiçoar sua concepção.

## MÉTODO DE PESQUISA

### Terminologia

É necessário definir os termos de nosso estudo para se justificar e reiterar a pesquisa bibliográfica executada, concernente aos aspectos teóricos das LI. Aqui temos duas palavras que pertencem a realidades diferentes. O termo **Linguagem de Indexação** tem um sentido bastante preciso no domínio da ciência da informação.<sup>1</sup> **Teoria**, ao contrário, pertence ao vocabulário corrente e é comum a todas as disciplinas científicas, no que concerne a sua plurivocação.

**As linguagens de indexação**<sup>2</sup> são linguagens artificiais — isto é, construídas a partir de um conjunto de regras — que servem para representar abreviadamente o conteúdo de um documento. Mas nem todas as linguagens de indexação são rigorosamente padronizadas: há níveis de controle. Inspirando-nos em J. Maniez (17), podemos representá-las em dois eixos: o da coordenação e o do controle. (ver QUADRO 1).

1. "Ciência que estuda a comunicação da informação", de acordo com Yves Le Coadic (Bulletin des bibliothèques de France, 1984, v. 29, n. 2, p. 168-171). Este artigo situa a ciência da informação como uma "interdisciplina" entre as ciências humanas e as ciências puras.
2. Escolhemos este termo entre os vários sinônimos (linguagens documentárias, etc.). Ele corresponde aos equivalentes em língua inglesa: indexing languages, documentary languages, information retrieval languages, etc.

As LI pouco controladas (como os descritores escolhidos livremente para representar o conteúdo de um documento) se aproximam das linguagens naturais (LN). O interesse deste artigo é especialmente dirigido às LI mais controladas, uma vez que o controle é o fator que as diferencia claramente da LN. Trataremos das linguagens pós-coordenadas, nas quais a combinação dos descritores se faz por ocasião da pesquisa documentária (como os tesouros), assim como das LI pré-coordenadas, nas quais a combinação dos termos é fixada na hora da indexação (como as classificações, e sobretudo, as linguagens em cadeia PRECIS, SYNTOL, POPSI, dotadas de uma sintaxe controladora).

O termo **Teoria** tem o sentido de construção intelectual ou especulativa, opondo-se seguramente à **prática**. Um tipo importante de abordagem teórica que guiou nossa pesquisa foi "o uso de modelos". O **modelo** é um instrumento intelectual que serve de princípio explicativo

**QUADRO 1: Representação das LI em função de seus graus de coordenação e de controle**

	Classificações hierárquicas	Classificações facetadas	Tesouros
+			
Controle	linguagens em cadeia (PRECIS, POPSI...)		Terminologias
-			
	indexação de assunto livre		descritores livres
			palavras-chave (unitermos)
	Pré-coordenação	Coordenação	Pós-coordenação

de um domínio do conhecimento. Desta forma, vários modelos lingüísticos ou matemáticos foram empregados para representar a estrutura das LI.

### **Delimitações**

Nosso trabalho não pretende ser exaustivo, mas sim fazer uma síntese da literatura que, às vezes, é redundante. Os textos foram selecionados por seu valor inovador, sua importância para a pesquisa teórica ou, precisamente, por sua qualidade de síntese. Considerando-se que a nossa pesquisa objetiva aspectos teóricos, uma seleção de documentos segundo sua data de publicação torna-se menos importante. Ao contrário das práticas e das técnicas (embora cientes das mudanças produzidas pela informática na documentação) o interesse das teorias das LI não é limitada pelo tempo. A bibliografia aqui apresentada terá portanto uma visão retrospectiva.

Nossa pesquisa se limitou às línguas alemã, francesa e inglesa. Elas representam o essencial de nossa literatura ocidental. A importante contribuição dos países do Leste só nos é conhecida através de traduções.<sup>3</sup>

### **Instrumentos**

Esta pesquisa bibliográfica utilizou a consulta simultânea a repertórios impressos e a bancos de dados. Os instrumentos que utilizamos para a pesquisa manual estão apresentados no QUADRO 2, de acordo com a ordem de elaboração intelectual dos tipos de busca: bibliografias correntes e retrospectivas, bibliografias de assunto específicos, sínteses bibliográficas.

---

3. Alguns títulos anglo-saxões são traduções de revistas soviéticas, tais como *Automatic Documentation and Mathematical Linguistics* ou *International Forum on Information and Documentation*.

No que concerne a pesquisa automatizada, consultamos quatro bancos de dados: PASCAL, INSPEC, LISA e ISA.<sup>4</sup>

A grande dificuldade de consulta a bancos de dados reside na maneira de se estabelecer as estratégias de busca. Desta forma, a representação dos descritores do "aspecto teórico" é difícil, uma vez que o termo "teoria" é utilizado de maneira muito abrangente pelos indexadores. Os bancos de dados consultados apresentaram limitações quanto à pertinência terminológica. Indicamos, abaixo, as estratégias de busca empregadas em cada banco de dados.

PASCAL (Questel +) 1. LIM SCIENCE  
INFORMATION/FG.

2. (LANGUAGE DOCUMENTAIRE ou  
INDEXATION ou THESAUR +/T ou  
CLASSIFICATION ET THEORI +/DE/T.

3. 2 ET (FRE/LA ou GER/LA ou ENG/LA)

INSPEC (Questel +) 1. LIM C + 72 + /CC/T.  
2. (INFORMATION ANALYSIS ou THESAUR /+/ DE/  
T) ET (THEORY ou THÉORIES ou MODEL?/T)

LISA e ISA (Dialog) S1. INDEXING (W) LANGUAGE  
OR THESAU?

S2. S1 AND (THEORY OR THEORIES OR MODEL?)

S3. S2/ENG OR (S2 AND LA=FRENCH)  
OR (S2 AND LA = GERMAN.

---

4. Para indicações práticas consultar o *Repertoire des banques de données professionnelles*. iied. Paris: ADBS-ANRT, 1989.

## QUADRO 2: Instrumentos utilizados na pesquisa manual

### BIBLIOGRAFIAS CORRENTES

Bulletin signalétique du C. N. R. S. 101 Science de l'information documentation. - 1970-1983. - Mensuel.

Devient:

PASCAL Thema T205. Sciences de l'information, documentation. - 1983- .- Mensuel

LISA: Library information science abstracts.- 1969- Mensuel.

### BIBLIOGRAFIAS RETROSPECTIVAS

Library literature.- 1981- .-Vol. cumulatifs pluriennaux, puis annuels.

### BIBLIOGRAFIAS ESPECIALIZADAS

Dahlberg, Ingetraut (ed.). - Classification and indexing e systems: theory, structure, methodology: 1950-1982. - Frankfurt Indeks Verlag, 1985. - (International classification and indexing bibliography, ICIB 3)

Wellisch, Hans H. (ed.) - Indexing and abstracting: an international bibliography. - Santa Barbara (Cal.); Oxford: ABC-Clío 1980.

Maccafferty, Mazime. - Thesauri and thesauri construction.- Dondon: Aslib, 1977. - (Aslib bibliography, 7)

### REVISÕES BIBLIOGRÁFICAS

Annual review of information science and technology. - Washington: American society for information science, 1966.- . - Annuel.

Des surveys font le point sur certains thèmes précis (comme PRECIS). Ils sont souvent accompagnés d'une liste très complète de références. On trouvera les plus recents dans notre bibliographie.

## A RENOVAÇÃO DAS CLASSIFICAÇÕES

### O Classification Research Group (CRG)

Sobretudo após 1945, as falhas das classificações tradicionais se tornaram cada vez mais evidentes. Mesmo sem o advento dos computadores, necessitou-se de uma gestão mais racional dos fundos documentários que proliferaram nos países industrializados. Na América do Norte, as pesquisas se voltaram para outros métodos, como a indexação coordenada (23) que conduziria aos tesouros. Na Europa, tentou-se elaborar classificações com bases novas.

Em 1948, foi fundado, em Londres, o CRG. Um de seus membros, D. J. Foskett (5) descreve o panorama dos temas abordados pelo grupo até 1970. A classificação não seria mais considerada apenas como um instrumento de arranjo físico dos livros nas estantes, mas como uma linguagem autêntica, cuja finalidade seria a pesquisa documentária.

As teorias de S. R. Ranganathan, na Índia, tiveram uma influência considerável no CRG e, de uma maneira geral, continuam a influenciar a maioria dos interessados nas ciências da informação. Seu método de **classificação analítico-sintética** (18:109) consiste em analisar o assunto por diferentes pontos de vista ou **facetar** e depois sintetizá-lo, organizando as facetar de acordo com seqüência canônica que ele denominou PMEST, representando as iniciais inglesas das facetar **personalidade, matéria, energia, espaço e tempo**. Comparada com as classificações enumerativas e rígidas, a possibilidade de se coordenar as facetar apresenta uma flexibilidade maior na indexação e as relações polierárquicas geradas entre as facetar permitem representar melhor a multiplicidade de elementos que compõem os assuntos. O CRG criou diversas

classificações especializadas em facetas, dentre as quais a **Classification of library and information science** de 1971, utilizada pelo serviço bibliográfico LISA.

O CRG pretende, também, esboçar uma nova classificação universal numa base naturalista, que reflita a ordem do mundo real. Para tal, o grupo adotou a famosa tese biológica dos **níveis integrativos** de J. K. Feiblemann: o universo se desenvolveu em níveis sucessivos de complexidade crescente, cada nível integrando o precedente (partículas fundamentais, átomos, moléculas, etc.) Todas as áreas do conhecimento podem ser ordenadas de acordo com este princípio.

Esta pesquisa, de um fundamento quase ontológico da classificação, se opõe frontalmente à tendência pragmática americana que se apóia no princípio de "garantia literária" (literary warrant) segundo o qual um termo só tem razão de ser quando potencialmente útil na classificação de, pelo menos, um documento real. (F.W. Lancaster, 14:86)

A teoria dos níveis integrativos inspirou a noção de contextualização aplicada por D. Austin ao PRECIS (89,90). D. J. Foskett (6) generaliza esta tese na **teoria geral dos sistemas**: um sistema é um todo que não é senão a soma das partes que contém, mas ele próprio é parte de um sistema maior e assim por diante.

#### **Classificação a priori cu a posteriori**

O problema pode ser visto sob outro ponto de vista. A construção das classificações deve ser feita **a priori**, impondo-se uma ordem teórica, ou **a posteriori**, partindo-se das relações concretas entre as coisas"?

Desde 1950, J. E. L. Farradane (100,101) defendeu que uma classificação deve ser elaborada com o apoio de considerações epistemológicas sobre o próprio processo de aquisição do conhecimento. O desenvolvimento científico é uma alternância de **fases dedutivas e indutivas**.

Ora, as classificações tradicionais não são científicas mas puramente dedutivas. Farradane descreveu, então, uma base indutiva: os **isolados** (conceitos) seriam ligados entre si por meio de termos relacionais (operadores) para formar um "analetos", representando um assunto. Combinando-se os analetos através de seus isolados comuns, obtém-se uma estrutura ordenada, coextensiva ao conjunto do saber. Entretanto, Farradane se interessou também pelas propriedades dos analetos isolados. Através de regras de arranjo e permutação, eles podem servir para a criação de índices alfabéticos. Seu método, denominado **indexação relacional** (102) é muito próximo às linguagens de indexação em cadeia.

Na década de sessenta, quando apareceram as máquinas calculadoras sofisticadas, R. M. Needham e K. Sparck Jones (66) tentaram construir automaticamente uma classificação, tendo por base os próprios documentos. Eles ampliavam a noção de classes para a de aglomerados (**Clumps**): se duas palavras-chave figuram na indexação de um documento, elas são equivalentes do ponto de vista de busca desse documento em particular. Medindo-se o **grau de co-ocorrência** de todos os pares possíveis, poder-se-á determinar conjuntos (**clumps**) onde cada palavra-chave tem um grau de co-ocorrência mais elevado com os membros de um conjunto do que com os não-membros.

A partir da década de 80, um exame crítico dos trabalhos do CRG fez com que alguns autores assumissem uma posição apriorista. D. Austin (60) postula a existência de uma parte de classes de **conceitos de base** (coisas, ações, propriedades) e, por outro lado, de classes de **relações de base** (lógicas, gramaticais).

B. P. Frohmann (70) critica a escolha do CRG de estruturar a **posteriori** as relações semânticas e de rejeitar a fórmula rangianathiana predefinida PEST. Cada área

gera suas próprias relações e, então, torna-se impossível criar um sistema geral de categorias. Uma **base semântica** deve ser resgatada **a priori** uma vez por todas, segundo as indicações de D. Austin.

### **A revista International Classification**

Na Alemanha, I Dahlberg foi responsável por um novo impulso no estudo das classificações, ao fundar, em 1974, a revista "International Classification" e se tornar editora da coleção "Studien zur Klassifikation" publicada pelo "Gesellschaft für Dokumentation". A referida autora (4) fez um estudo muito vasto e muito rico para se resumir aqui (ele contém mais ou menos 600 referências bibliográficas!) sobre a pesquisa dos **fundamentos universais da ordenação do conhecimento** (Grundlagen universaler Wissensordnung).

A ordenação do conhecimento é vital em uma época na qual as disciplinas fragmentam-se cada vez mais. A noção de classificação ultrapassa o domínio da biblioteconomia. Conduziu-se, então, uma pesquisa interdisciplinar onde apareciam a filosofia, a epistemologia, a lingüística, as teorias científicas, etc. O resultado se apresenta como um esboço formal de uma nova classificação universal.

T. T. Ballmer (26,27) procurou atingir o mesmo objetivo, restringindo sua atenção às expressões lingüísticas. Desta forma, o verbo tem um importante papel na classificação do conhecimento. Por exemplo, "comer" determina toda uma classe de conhecimentos "comestíveis": frutas, legumes, carnes, etc.

### **Do Leste Europeu**

V. A. Schreider (53,54) lança uma pergunta epistemológica fundamental: a ordem é uma forma determinada, pelo homem ao caos ou é uma propriedade das

próprias coisas? A lógica e a classificação aristotélica oferecem uma base para reflexão. Uma idéia pode ser descrito em **extensão** ou em **intensão**. A extensão é a classe de objetos determinados pela idéia (chamada **taxon** pelo autor). A intensão (ou abrangência) é o conjunto de características da idéia (chamadas **merons**). A classificação apóia-se em duas abordagens: a taxonomia, onde os objetos são apresentados com base em suas similaridades, e a "méronomie", que é a identificação das características essenciais comuns a um **taxon**.

## EVOLUÇÃO DAS LINGUAGENS DE INDEXAÇÃO EM CADEIA

A indexação em cadeia consiste em analisar um assunto através de uma seqüência linear e rígida de termos, com a finalidade de construir índices alfabéticos manuais.

J. O. Kaiser, S. R. Ranganathan e E. J. Coates foram, por diferentes caminhos, os precursores da técnica que se desenvolveu, sobretudo, no fim da década de 60. Em 1960, o sistema KWIC de H. P. Luhn foi o primeiro índice deste tipo inteiramente mecanizado.

Os anglo-saxões (B. W. Mineur, 88) consideram em lados opostos a **indexação em cadeia** baseada no **procedimento em cadeia** de Ranganathan e a indexação em fileira (String) que designa os sistemas modernos.

A obra de T.C. Craven (84) é uma síntese atualizada que descreve um grande número de sistemas.

Em 1968, a British National Library, que produz a British National Bibliography (BNB), decidiu substituir a produção do índice (realizada pelo procedimento em cadeia) por uma técnica moderna em colaboração com o projeto UK-MARC. A partir de 1971, foi testado um protótipo do PRECIS.

Na mesma época, a "escola indiana" com G. Bhat-tacharyya, lançou POPSI. Nos anos 70, apareceram outras linguagens: NEPHIS e LIPHIS de T. C. Craven.

Recentemente foram feitas comparações destas linguagens (A. Cheti, 94). Farradane (85) comparou PRECIS, POPSI e NEPHIS com sua indexação relacional. Pesquisas tipo **survey** estudaram PRECIS (M. Mahapatra, S. C. Biswas, 96) e POPSI (S. C. Biswas, 93).

Estas linguagens são pré-coordenadas. Como serviam para editar índices impressos tradicionais, cada termo deveria aparecer no início da cadeia para poder ser recuperado. Desta forma, era necessário resolver dois problemas:

- 1) ordem de sucessão dos termos (uma "gramática"); e
- 2) as regras para permutação dos termos.

### **Ordem dos termos**

A **indexação sistemática** de J. O. Kaiser (104, E. Svenonius, 106) consiste na análise de todos os assuntos de acordo com duas grandes categorias: **os concretos** (ou objetos) e os **processos** (ou ações). Sintaticamente, os concretos precedem sempre os processos.

S. R. Ranganathan formula regras sintáticas muito precisas para dispor as facetas de um assunto em uma seqüência fixa, como, por exemplo, o famoso **wall-picture Principle** (18:425).

➔ Mais tarde, E. J. Coates (61), retomando as idéias de Kaiser, busca fundamentos psicológicos para definir a precedência da idéia mais "clara", para melhor atender às necessidades do usuário: desta forma, as **coisas** (estáveis e determinadas) precederiam os **materiais** (estáveis mas não determinados) e as **ações** (instáveis e dinâmicas).

G. Bhattacharyya retoma os princípios de Ranganathan (11,92) que ele chama de **postulados** (dando origem ao nome de POPSI). A seqüência fundamental é a seguinte: **a base** (categoria elementar ou disciplina) e o **núcleo (core)** dos conceitos que se relacionam com a base.

No PRECIS, a ordem é determinada pelo princípio de dependência contextual (**context dependency**) (90). Cada termo segue e especifica o termo que representa seu contexto ("Ásia-Índia", por exemplo). A cadeia sintagmática é reforçada pela atribuição de operadores de função (**role**) a cada termo. O sistema foi concebido astuciosamente de maneira que a ordem de citação dos operadores associados aos termos conduz automaticamente a uma cadeia de dependências contextuais.

B. W. Mineur (86), entretanto, critica o PRECIS porque ele privilegiou uma ordem de base única ocultando todas as relações secundárias entre os termos que não ocupavam posições contíguas na cadeia. A indexação relacional de Farradane, se bem que semelhante, não é uma linguagem em cadeia pois os "analetos" podem ter uma estrutura bidimensional.

Em comparação aos sistemas formais como POPSI e PRECIS, o sistema NEPHIS (98) se apresenta muito empírico. O assunto é analisado e representado em termos simples e/ou complexos aninhados (**nested**) em diversos níveis uns nos outros. O sistema, portanto, organiza esses elementos em estrutura de árvore. Para alguns assuntos, esta estrutura simplificada não é conveniente. O sistema LIPHIS, uma versão melhorada do NEPHIS, dá a possibilidade de **ligar (link)** lateralmente os elementos de um assunto cuja estrutura não se reduz a uma hierarquia pura (99).

## Permutação de termos

A partir de uma cadeia principal, é necessário determinar as cadeias derivadas que colocarão em posição de acesso cada um dos termos que a compõem. Podem ser adotadas várias soluções formais (W. J. Hutchins, 34: cap. 6).

O processo em cadeia, de Ranganathan, consiste em suprimir sucessivamente os termos à esquerda. A partir da cadeia ABCD, pode-se derivar BCD, CD e D. Cada termo aparece em posição de acesso, mas ele é seguido somente por uma parte dos outros termos que representam o assunto. O defeito do sistema é que um assunto específico (ABCD) aparecerá sob entradas muito gerais (CD, D). Este problema é denominado cadeia em desaparecimento (**disappearing chain**) (S. C. Biswas, 93) e foi uma das razões do abandono desta técnica utilizada pela BNB de 1950 até 1970.

J. E. Armitage e M.F. Lynch (81) formalizaram a produção de um **índice de assuntos articulados** por permutação de elementos sintaticamente ligados na linguagem natural por preposições, sem introduzir ambigüidades.

A permutação dos termos permite que o princípio de especificidade seja respeitado. Mas ela multiplica as entradas e dispersa a ordem e as relações da cadeia principal.

Os sistemas mecanizados do tipo KWIC trabalham com **rotações** das palavras-chave existentes em uma frase. O PRECIS utiliza a rotação e coloca a cadeia derivada em duas linhas de maneira que a cadeia principal e suas relações possam ser sempre reconstruídas.

## DESENVOLVIMENTO DOS TESAUROS

O termo **tesauro**, escolhido para designar um tipo de LI controlada, se derivou do "Thesaurus of English words and phrases" (1852) de autoria de R. M. Roget (J. Maniez, 65). Ao contrário dos tradicionais dicionários de língua onde se parte do **termo** para descobrir sua definição (portanto seu **significado**), Roget parte do significado (os conceitos) para chegar a todos os termos que o representam, sem definições. O objetivo dos tesauros modernos é, igualmente, representar primeiramente os conceitos, e ainda especificar suas relações.

### Os tesauros na América do Norte

Nos Estados Unidos, os pesquisadores desenvolveram, rapidamente, novas técnicas documentárias, visando substituir os métodos classificatórios tradicionais.

Nos anos 50, M. Taube (23,58) elaborou sua **indexação coordenada** (atualmente diríamos "indexação pós-coordenada"). Cada palavra-chave ou **unitermo** de um assunto era o cabeçalho de uma ficha, sob o qual são registrados os números de referência de todos os documentos pertinentes. No processo de busca, o usuário consulta as fichas que correspondem à sua demanda e detecta os números iguais — representados nas fichas. Por este método, o número de combinações possíveis é muito grande, os usuários não necessitam se preocupar com a ordem dos elementos em um assunto complexo e as referências seriam supérfluas.

A chegada dos computadores facilitou a manipulação da indexação pós-coordenada, com o sucesso que se conhece. Como os unitermos de M. Taube eram esco-

lhidos de uma maneira ainda empírica, logo se sentiu a necessidade do controle do vocabulário e do tratamento das relações paradigmáticas (sobretudo a sinonímia).

Segundo F. W. Lancaster (14), autor de um estudo geral sobre as LI controladas, o primeiro tesouro moderno foi publicado em 1959 pela Sociedade Dupont de Nemours. Apareceram, na mesma época, as noções de **precisão** e de **revocação**, relacionadas com o desempenho da busca. Durante a década de 60, os tesouros foram sistematicamente aperfeiçoados. O fato de as **relações sintáticas** não serem apresentadas nos tesouros provocou coordenações imperfeitas, prejudicando o resultado das buscas. Em 1964, o Engineers Joint Council apresentou o primeiro tesouro com **indicadores de função (role indicators)** que permitiam identificar uma função para cada descritor. J. C. Costello (78) comparou este procedimento com as linguagens naturais que usavam declinações (latim, russo, alemão) como meio de ligação sintática. Mesmo em outras LN que usam meios diferentes para representar a sintaxe (ordem dos termos, preposições, etc), os indicadores de função são certamente os únicos controles sintáticos possíveis nas LI.

J. Aitchinson apresentou, em 1969, um tesouro de gênero novo: o **thesaurofacet** (77). Este novo conceito integra, em um só sistema, as vantagens da apresentação sistemática dos termos, para análise de uma área em facetas e as dos tesouros alfabéticos. O **thesaurofacet** é um instrumento universal que pode ser usado tanto como LI pré-coordenada na produção de cadeias para um índice, como para um sistema pós-coordenado de busca. As duas tradições de indexação muito diferentes (americana e indo-européia) assim se encontram.

## Estudo crítico dos tesouros na Europa

Depois de 1970, os especialistas tomaram consciência de inconvenientes fundamentais. O aumento dos acervos documentários, e a crescente especialização do conhecimento incharam o léxico, criando problemas relativos à sua atualização. O arranjo alfabético é incapaz de representar bem as relações úteis entre os termos. Foram repensados o tesouro e a natureza do processo de indexação.

Na perspectiva de uma pesquisa fundamental sobre as linguagens de indexação, R. Fungmann (7) apresentou novas idéias com relação à tarefa do próprio indexador. A precisão na escolha do termo mais apropriado dependerá da ordem na qual os termos estão dispostos. A ordem alfabética dispersa os termos de significado semelhantes, que só serão ligados através de remissivas. Quanto mais difícil de se encontrar o termo apropriado (principalmente quando a listagem se alonga) mais remota a sua possibilidade de acesso (**predictability**). Inspirando-se no método de Ranganathan, uma LI deve analisar os termos (a fim de limitar o léxico) mas também deve resintetizá-los a fim de preservar as ligações entre eles.

Sob estas bases teóricas, R. Fugmann desenvolveu um sistema de indexação bidimensional chamado TOSAR, utilizado para a documentação automática de química (103). Os tesouros tradicionais propunham, como ajuda à indexação, "esquemas flechados" (**arrowgraphs**) representando seus descritores e suas relações semânticas. O TOSAR vai bem mais longe. O sistema explora a estrutura e as propriedades da teoria de gráficos: um **gráfico** é um conjunto de pontos, chamados **vértices**, ligados entre si por **arcos**. Pode-se então definir as noções de orientação, de distância entre os vértices, etc.

Nas aplicações documentárias, um assunto é representado por meio dos gráficos. Os termos são os vértices e as relações (analíticas ou sintéticas) são os arcos. Os parâmetros da demanda são também formulados por um gráfico cujos vértices e arcos são mais ou menos determinados. Por ocasião da busca, um programa compara o gráfico da demanda com os gráficos dos documentos registrados na memória e reterá os que têm a mesma estrutura.

O método TOSAR foi descrito formalmente, por meio da lógica dos predicados, por H. e I. Nickelsen (105).

Recentemente, R. Fugmann (10) sintetizou sua teoria sob a forma de cinco axiomas ou postulados de indexação. Introduziu uma dimensão ética através do conceito de **indexação imperativa (mandatory indexing)**: o indexador é obrigado a escolher "o" termo mais apropriado, ao contrário da indexação simplesmente controlada, que dá sempre a possibilidade de se escolher vários termos mais ou menos adequados.

D. Soergel (74) definiu, de maneira geral, uma LI como um conjunto de descritores, de relações e de regras para a formação de expressões. O tesauro contém, ainda, um vocabulário de não-descritores que remetem (**lead-in-vocabulary**) aos termos reconhecidos. Sendo os usuários diferentes, a indexação deve ser de modelo múltiplo: os contextos, as abordagens diferentes devem levar ao mesmo descritor. A análise de assuntos por facetas — ou precisamente **pontos de vista** — é um exemplo de indexação de modelo múltiplo.

Numerosos estudos tratam de aspectos específicos dos tesouros. M. Willets (67) compara o uso de diferentes tipos de relações em vários tesouros. Tomando como modelo as categorias gerais do CRG e os nove operadores de Farradane, estuda, em particular, as relações cha-

madas **afinitivas**. Este tipo de relação reagrupa habitualmente todas as espécies de relações que os pesquisadores desistem de distinguir, por insuficiência de análise.

R. Jansen (72) mostra a utilidade de um novo tipo de relação chamada **relação de dependência (Zugehörigkeitsrelation)** que não deve ser confundida com as **relações hierárquicas**. Ela faz a ligação de um termo com todos os outros descritores pré-coordenados em cuja composição esse termo entra.

A Ghose e A. J. Dawle (71), na Índia, propõem uma técnica pragmática (derivada do método dos aglomerados) de estruturação de um léxico para ciências humanas, por meio de questões propostas a especialistas nas disciplinas.

Para C. Kim (36), o processo do dicionário (palavras ligadas a conceitos expressos) e o do **Thesaurus** de P. M. Roget (conceitos ligados a palavras) são úteis na elaboração e atualização de tesouros. De fato, o significado de uma palavra não depende somente da idéia ou do objeto que ela representa (teoria referencial) mas também das relações que ela estabelece com as outras palavras (teoria operacional ou contextual).

## LINGUAGENS DE INDEXAÇÃO E TEORIAS LINGÜISTICAS

Numerosas são as razões que levaram os documentalistas a se interessar pelas teorias lingüísticas. As propriedades das LI se assemelham bastante às linguagens naturais e algumas se derivam profundamente delas. É, portanto, natural que se pergunte se a lingüística não seria adequada para descrevê-las ou mesmo formalizá-las. Como as LI aplicam-se principalmente — mas não exclusivamente — a textos expressos em LN, deve-se considerar o problema da passagem de uma linguagem para outra.

Nos anos 60, alguns pesquisadores depositavam grande esperança na automatização dos tratamentos documentários. Por isto, se dedicaram seriamente a analisar as entidades que eles manipulam: palavras, frases, resumos, descritores, etc. Esse período é marcado pela multiplicação de estudos linguísticos das LI.

K. Spark Jones e M. Kay (41) analisam os trabalhos produzidos entre 1965 e 1973, orientados particularmente para os procedimentos de automatização.

Na França, a tese de M. Coyaud (30) é uma descrição e comparação de várias LI em particular. O esforço teórico se situa na elaboração de uma grade de análise uniforme, cujos constituintes se originam da terminologia lingüística: os monemas (as menores unidades lingüísticas possíveis de serem decompostas) são os **lexemas** (termos) e **morfemas** (relações). A **sintaxe** compreende a **palavra** (um lexema e um morfema) o **sintagma** (diversas palavras relacionadas) etc. O interessante deste método é de permitir a aproximação estrutural de LI aparentemente diferentes.

O trabalho de W. J. Hutchins (34) é uma boa introdução às estruturas lingüísticas gerais das linguagens de indexação: ele as compara com as linguagens naturais sob todos os aspectos (formais, semânticos, pragmáticos...).

### **Linguagens de indexação e linguagens naturais**

Na primeira parte de sua tese, J. Maniez (80) levanta uma questão geral: quais são as diferenças e as semelhanças fundamentais entre LN e LI? Para ele, o elemento comum é a universalidade da função referencial, isto é, a natureza simbólica de signo. O usuário não procura termos de indexação por eles mesmos, mas pelos

documentos onde eles representam o assunto. A estrutura das LI é, aliás, calcada (de forma "empobrecida") na das LN.

J. Cl. Gardins (32) denomina **analyse documentaire** a extração do significado dos documentos. Ele reagrupa vários tratamentos possíveis (resumo, indexação, etc.) mais ou menos elaborados. Em relação às LN, as LI são qualificadas como **metalinguagem**, ou seja uma linguagem ou sistema de símbolos adequados, utilizados para exprimir o conteúdo do documento redigido em LN.

R. Fugmann (31) considera as LI e as LN como complementares, que podem ser usadas tanto uma como a outra na indexação de acordo com a espécie de que se trata. Os **conceitos** individuais, ligados a um único objeto, são expressos na LN através de uma única expressão léxica. Os **conceitos gerais**, ao contrário, ligados a diversos objetos, se representam muitas vezes na LN através de várias expressões léxicas (sinônimos) ou mesmo por expressões não léxicas (perífrases) sobretudo no domínio científico onde a terminologia não acompanha as idéias com a mesma rapidez. Neste caso, uma LI controlada é mais vantajosa que as LN, porque ela tem, realmente, a tarefa de fazer corresponder uma única expressão a um dado conceito.

Farradane (102), por sua vez, considera que a abordagem lingüística não oferece interesse para as teorias de indexação. A lingüística é somente um instrumento para exprimir idéias e que não pode, ao contrário da psicologia, nos esclarecer sobre as relações em jogo, efetivamente, no processo do conhecimento.

### **Os dois eixos de referências**

No início deste século, o lingüista F. de Saussure fez a distinção entre o **eixo sintagmático** que relaciona **in praesentia** dois termos do enunciado enquanto que o

**eixo associativo** (diríamos hoje paradigmático) liga **in absentia** um termo do enunciado, com uma série virtual de sinônimos, termos genéricos, etc.

J. Maniez (80) considera que as relações sintagmáticas pertencem **ao discurso** e participam no processo da criação de uma frase, enquanto que as relações paradigmáticas pertencem à **língua** e existem fora de qualquer contexto.

No início da década de 60, Gardin integrou esta oposição essencial na construção de uma nova LI. O SYNTOL é um "modelo geral" porque pode ser aplicado em qualquer área científica. O termo **linguagem** se justifica pela importância que dá à sintaxe, cujo elemento de base é o sintagma: par de palavras-chave ligadas por uma relação e representando por  $R_i(x,y)$ . As relações se limitam numericamente em três. Regras formais permitem manipulá-las e compô-las.

Estes dois eixos permitem ainda a Gardin (32) classificar as LI (ou metalinguagem) de acordo com a estrutura que apresentam: uma estrutura sintagmática (as relações de Farradane), uma estrutura paradigmática (os tesouros) ou nenhum tipo de estrutura, quando seus termos são totalmente desorganizados (como os uni-termos). Ele mostra o caráter fundamental do sintagma  $R_i(x,y)$  que revela a identidade estrutural dos procedimentos sintáticos tais como os operadores de função ou os relatores de Farradane. Por combinação de sintagmas pode-se representar as relações referentes a 3,4... n posições.

No caso do tesouro, K. T. Bivins (29) insiste na influência do eixo paradigmático materializado pela **estrutura sindética** (o jogo das referências) no resultado da busca.

O tesouro contém, entretanto, relações que não são lingüísticas, nem paradigmáticas, nem sintagmáticas, denominadas por J. Maniez (65) relações **extra-semânticas**: elas ligam termos e fatos reais e correspondem a certas relações denominadas habitualmente "associativas".

### O triângulo semântico

O triângulo semântico dos lingüistas C. K. Ogden e I. A. Richards constitui-se das **expressões** (ou significantes) dos **conceitos** (ou significados) e dos **objetos** (ou referentes) e, também, de suas relações.

Este modelo é muito importante em documentação. Para R. Fugmann (10), o pólo central da busca é o conceito, não o objeto nem a expressão. Dois documentos podem ser pertinentes a uma demanda mesmo que seus títulos sejam distintos. Ele lembra que Ranganathan já distinguia o nível das idéias (**idea plane**) do nível verbal (**verbal plane**).

B. Long (37) modifica um pouco o modelo, distinguindo ainda o significante gráfico e o significante fonológico. Seu estudo consiste, então, em avaliar sistematicamente todas as relações a partir desses quatro pólos. A face semântica é predominante na documentação.

### A universalidade da linguagem

Na década de 60, a linguagem estruturalista (N. Chomsky, C. J. Fillmore, B. Pottins, J. Lyons, etc.) buscou modelar uma **estrutura profunda** das linguagens que permitisse reconhecer **estruturas de superfície** tão diversas como as das línguas naturais que conhecemos. Esses modelos interessariam às ciências da informação se fossem suficientemente fundamentais a ponto de explicar, também, a estrutura das linguagens artificiais, como as LI.

Mais especialmente, os documentalistas se interessavam por esses tratados — como, por exemplo, a gramática de casos de C. J. Fillmore — para conseguir uma lista de categorias fundamentais utilizáveis para a sintaxe das LI (J. Maniez (80); W.J. Hutchins (97). J. Sorensen e D. Austin (97) interessaram-se pelas estruturas profundas dos lingüistas e as compararam com a indexação no PRECIS. Demonstrando que o PRECIS já se situa em um alto nível de generalidade, torna-se logicamente possível empregá-lo num “contexto multilingue” para indexação em cadeia através de outras línguas além do inglês.

Os documentalistas, por seu lado, realizaram pesquisas sobre a universalidade das LI. As facetas de Ranganathan são um exemplo disto. G. Bhattacharyya (91,92) descreve uma estrutura de base da qual pode-se derivar diversas SIL (**subject indexing language**) reais: indexação sistemática, C. C., D. D. C., etc. A partir deste modelo, foram tiradas as estruturas profundas do POPSI.

F. J. Devadason (95) visualiza a utilização do POPSI como uma metalinguagem que permite gerar tesouros, automaticamente. Seriam usadas cadeias formadas em POPSI e enriquecidas com códigos especiais e o computador produziria o léxico do tesouro com sua estrutura sindética.

### **A terminologia**

W. Nedobity (39) considera opostos os conceitos de **lingüística** e **terminologia**. A lingüística ficaria reduzida ao estudo das LN, que são línguas “leves”, onde o significado dos termos depende muito do contexto. A terminologia, ao contrário, atribui ao termo um conceito distinto e definido. É assim mais útil para as ciências da informação que se ocupam das linguagens especializadas.

O objetivo da terminologia é racionalizar o processo da comunicação (G. Wersig (25)). Seu trabalho assemelha-se ao de elaboração de um tesauro: deve-se identificar em uma área do saber, primeiramente os conceitos, depois lhes atribuir um termo (em uma ou em diversas línguas) controlando, ao mesmo tempo, suas relações de sinonímia, homonímia, etc. Além disso, segundo um método dedutivo, dar uma definição rigorosa de cada conceito, como os **tesauro** que Wersig tomou como exemplo.

### O problema do assunto

As tendências mais recentes parecem girar em torno da idéia que os anglo-saxões denominam **aboutness**, que, traduzido, literalmente, é "do que" trata um texto, isto é, seu **assunto**.

J. Maniez (17), em recente obra de síntese, parte da oposição que os lingüistas estruturalistas fazem entre **tema** (thème) e **comentário** (rhème) tanto quando falamos, como quando ouvimos. A **tematização** por meio das LI é uma atividade essencial em documentação. Em oposição ao enunciado, ela não tem valor informativo, o que se constata pela ausência de verbo.

Para K. P. Jones (13), a indexação consiste em identificar o **aboutness** do documento em questão. Este assunto pode ser apreendido (de acordo com as necessidades) através de diversos níveis de análise, um dos quais é a estrutura textual.

C. Beghtol (28) trata detalhadamente as implicações lingüísticas de **aboutness**. Tradicionalmente, a lingüística considera a frase como unidade de base. Mas, para elucidar os mecanismos de compreensão do assunto, é

necessária uma **lingüística textual** levando em consideração a totalidade das frases de um documento. No caso de uma indexação ideal, um texto, seu resumo e sua indexação deveriam representar exatamente o mesmo **aboutness**.

## LINGUAGENS DE INDEXAÇÃO E TEORIAS MATEMATICAS

O interesse da modelização matemática é fornecer bases sólidas às LI, ao definir rigorosamente, passo a passo, os termos empregados: descritores, classes, relações.

A parte da matemática empregada é quase sempre a teoria de conjuntos, cuja terminologia se assemelha, às vezes, à das LI. O nível de conhecimento necessário para acompanhar estes estudos é variável. Os de G. Beling (45), de V. Diodato (46) e de C. N. Mooers (48) são muito acessíveis. Ao mesmo tempo, o nível de formalização é mais ou menos arrojado.

A área trabalhada pode ser muito geral e cobrir todas as LI, como um sistema geral de informação; ou se limitar a um aspecto específico desse mesmo sistema. B. C. Vickery (24) examina alguns destes modelos.

### Modelos gerais

Um dos primeiros modelos, o de J. W. Perry e A. Kent (49), é a parte teórica de um estudo consagrado à realização prática de um sistema de documentação. Trata-se realmente de um "modelo reduzido". Inicialmente, tem-se um léxico limitado de termos separados em categorias ou facetas. Por meio de **simulação** observam-se as propriedades deste sistema fechado e simples de ser manipulado. Por exemplo, é fácil de se mostrar

que "n" facetas permitem definir "n!".<sup>5</sup> Classificações hierárquicas diferentes. Novas noções são definidas: o grau de caráter geral de uma classe (relação do número de descritores subordinados ao número total de descritores), o grau de sinonímia, etc.

Na mesma época, C. N. Mooers (48) definiu um modelo que poderia representar um grande número de sistemas de indexação existentes e permitia até mesmo deduzir, logicamente, sistemas futuros mais aperfeiçoados. O problema da indexação passou a ser visto do ponto de vista do usuário: como fornecer um sub-conjunto "bom" de documentos, entre os existentes no sistema, em resposta a uma demanda? Matematicamente, o sistema de indexação se exprime por uma **transformação** entre o conjunto de termos da demanda e o conjunto dos documentos disponíveis. O modelo distingue, então, várias famílias diferentes.

D. Soergel (55) criou um modelo para todas as LI. Paralelamente à busca da universalidade lingüística, ele propõe uma base lógico-matemática geral. O estudo consiste na construção de um **sistema formal**. A metalinguagem matemática e a linguagem da área da documentação que ela representa são bastante distintas. A ligação entre uma e outra é a **interpretação** ou a **significação** do modelo. Ao nível de metalinguagem, criam-se um léxico e regras para formação das expressões e se deduzem teoremas. Restringindo-se o modelo a casos particulares, pode-se criar LI reais.

S. Reball (50) constata que as LN são expressivas por sua riqueza semântica, mas pouco formalizadas; ao passo que as LI são bastante formalizadas, mas pouco

---

5.  $n!$  símbolo matemático para "fatorial", definido da seguinte forma:  
 $n! = 1 \times 2 \times 3 \dots \times (n-2) \times (n-1) \times n$ .

expressivas. Usando a **lógica dos predicados**, ele constrói um modelo de LI reunindo a expressividade e a formalização.

O estudo de V. Diodato (46) apresenta, de maneira didática, os principais conceitos geralmente usados para representar a estrutura sindética das LI. A relação de sinonímia entre dois descritos é, matematicamente, uma **relação de equivalência**. Todos os sinônimos de um conceito formam, então, uma **classe de equivalência**. As remissivas de um termo a outro são representadas por aplicações e as relações hierárquicas por **relações de ordem parcial**. Da mesma maneira, G. Boeing (45) estuda a estrutura das classificações e M. S. Sridhar (56) detalha as características matemáticas destas relações nos tesouros.

### Modelos específicos

No Leste europeu, estudos mais detalhados foram feitos particularmente sobre tesouros.

V. D. Sydorchenko (57.75) definiu o descritor de um tesouro como a classe de equivalência de termos sinônimos em LN. A eficácia de uma LI depende de seu **poder semântico (semantic power)**, que é o número de expressões não sinônimas bem formadas que ela pode produzir. Pela flexibilidade de sua capacidade combinatória, as LI pós-coordenadas têm um poder semântico superior ao das LI pré-coordenadas.

W. M. Turski (59) dá um modelo de tesouro que compreende um conjunto de **descritores** e de **ascritores** ligados por relações de sinonímia.

N. A. Stokolova (42) aprofunda os aspectos semânticos. A pertinência de um sistema de informação se mede comparando a significação de dois textos: o da

demanda e o do documento. O objeto das LI é representar esta significação. A **pertinência estrita** é uma relação de equivalência entre os dois textos. Ele define também **graus de pertinência**. Em outro trabalho (43) o mesmo autor considera os diversos instrumentos sintáticos (ligações, funções) e mede matematicamente sua influência sobre o poder semântico da LI.

Na França A. Dewèze (69) formaliza a representação das **relações semânticas**. Na perspectiva de construir tesouros multilíngües, ele adota uma **teoria semântica extra-léxica** que se situa em um nível superior ao das LN. Um significado é definido como um conjunto de **semes** (obtidos através de análise proposta por Ranganathan) e aos quais se pode, ulteriormente, atribuir **etiquetas léxicas** em diversas línguas. As relações de uma LI são descritas ao se estudar, por meio de gráficos, as **relações**, entre estas **configurações semânticas**. Por exemplo, um nível em uma monoierarquia terá, em sua configuração sêmica, um seme a mais que o nível lhe é imediatamente superior.

Recentemente R. Schauble (52) propôs uma nova estrutura da informação: o **espaço conceitual (concept space)**, que permite construir uma teoria de tesouros, exposta como um sistema formal.

Este estudo mostra a vitalidade das pesquisas teóricas sobre as LI, levadas a efeito por mais de quarenta anos nos países anglo-saxões, no Leste europeu e na Índia, em número bem mais expressivos do que nos países de língua francesa onde se nota modesta contribuição do ponto de vista quantitativo. O aumento progressivo de crescimento das publicações estimulou, certamente, as maneiras de considerar a indexação, chave de acesso à documentação.

É significativo que, durante o período analisado, todas as linguagens de indexação tradicionais assim como as mais modernas (tesauros) não escaparam às críticas (classificações, cabeçalhos de assunto...) foram sistematicamente examinados.

#### VISÃO CRONOLÓGICA

- 1852 **Grã-Bretanha.** P. M. ROGET publica o *Thesaurus of English words and phrases*.
- ± 1900 **Estados Unidos.** J. O. KAISER introduz o conceito de indexação sistemática.
- 1911 **Estados Unidos.** E. W. HULME formula a noção de garantia literária.
- 1915 **Suiça.** F. de SAUSSURE concebe a teoria de dois eixos de referência: relações sintagmáticas e relações associativas (paradgmáticas).
- 1923 **Grã-Bretanha.** C. K. OGDEN e I.A. RICHARDS; triângulo semântico (expressões — conceitos — objetos).
- Década de 30 — **Índia.** S. R. RANGANATHAN expõe o processo em cadeia e o método analítico-sintético de classificação.
- 1948 **Grã-Bretanha.** Fundação do **Classification Research Group** (CRG) em Londres.
- 1950 **Grã-Bretanha.** J.E. L. FARRADANE propõe uma teoria científica para as classificações, ponto de partida para a indexação racional.
- ± 1951 **Estados Unidos.** M. TAUBE concebe um sistema de indexação coordenada, usando unitermos.
- 1954 **Grã-Bretanha.** O biólogo J. K. FEIBLEMAN descreve a teoria dos níveis de integração.
- 1955 **Estados Unidos.** Fundação do Centro de Pesquisa em Documentação e Comunicação da **Western Reserve University** (WRU) em Cleveland.

- 1957-1958 **Estados Unidos.** Primeiros modelos matemáticos das linguagens de indexação de J. W. PERRY, A. KENT. e C. N. MODERS.
- 1959 **Estados Unidos.** Primeiro tesouro, na acepção moderna do termo, da Sociedade Dupont de Nemours.
- Fim da década de 50 — Aparecimento das noções de precisão e revocação.
- Década de 60 — Trabalhos dos linguistas estruturalistas: N. CHOMSKY, C. J. FILLMORE, J. LYONS.
- 1960 **Estados Unidos.** H. P. LUHN desenvolve o sistema KWIC inteiramente mecanizado, precursor das linguagens de indexação em cadeia.
- 1962 **Índia.** Fundação do Centro de Pesquisa e Treinamento em Documentação, em Cangalore, núcleo da escola indiana, iniciada por S. R. RANGANATHAN.
- 1963 **Estados Unidos.** Tesouro do **Engineers Joint Council** (TEJC) com indicadores de funções (role indicators)
- 1964 **França.** J. C. GARDIN elabora o modelo geral SYNTOL.
- 1964 **Grã-Bretanha.** R. N. NEEDHAN e K. SPARCK JONES: abordagem automática das classificações e noções de **Clumps.**
- 1967 **Alemanha (RFA).** D. SOERGEL: formalização matemática das linguagens de indexação.
- 1969 **Grã-Bretanha.** J. AITCHISON apresenta o Thesourofacet, tesouro que apresenta os conceitos analisados em facetas.
- ± 1969 **Índia.** G. BHATTACHARYYA desenvolve a linguagem de indexação POPSI.
- 1971 **Grã-Bretanha.** D. AUSTIN testa um protótipo do PRECIS, na BNB.

- 1974 **Alemanha (RFA)**. I. DAHLBERG: pesquisa interdisciplinária dos fundamentos universais de ordenação do conhecimento e criação da revista *Internacional Classification*.
- 1974 **Alemanha (RFA)**. R. FUGMANN analisa os princípios dos tesouros e desenvolve um sistema de indexação por meio de gráficos, TOSAR.
- 1977-78 **Canadá**. T. C. CRAVEN: linguagens de indexação em cadeia NEPHIS e LIPHIS.
- 1979 **Índia**. G. BHATTACHARYYA pesquisa os fundamentos comuns das linguagens de indexação com o **Subject Indexing Language (SIL)**.
- 1981 **França**. A. DEWEZE: redes semânticas extra-léxicas aplicadas à constituição de tesouros.
- 1986 **Grã-Bretanha**. C. BEGTOL analisa a noção de assunto (**aboutness**) por meio da linguística textual.

## BIBLIOGRAFIA

As referências citadas, estão classificadas metodicamente e sub-classificadas alfabeticamente por nome dos autores. Guardamos três grandes divisões:

- os principais modelos e teorias cujos termos definidos no texto;
- a expressão das relações entre os tesouros de indexação. Como toda linguagem, as LI possuem ligações semânticas e sintáticas;
- alguns sistemas de indexação originais considerados a partir de reflexões teóricas que lhes serviram de base.

## MODELOS TEÓRICOS

### Generalidades

1. BHATTACHARYYA, G. Fundamentals of subject indexing languages. In: *Ordering systems for global information networks: proceedings / of the third international study conference on classification research, 1975, Jan. 6-11, Bombay, India*; A. Neelameghan ed. Bangalore: Interna-

tional federation for documentation, Committee on classification research (FID/CR): Sadara Ranganathan endowment for library science, 1979. p. 83-89. (Sadara Ranganathan endowment for library sciences series, 13; FID publication, 533).

2. COOPER, Williams S. Indexing documents by Gedanken Experimentation. **Journal of the American society for information science**, 1978, vol. 29, n. 3, p. 107-119.
3. CROS, René-Charles; GARDIN, Jean-Claude; LÉVY, Francis. L'automatisation des recherches documentaires: un modèle général "le SYNTOL". Paris: Gauthier-Villars, 1964. 260 p.
4. DAHLBERG, Ingetraut. Grundlagen universaler Wissensordnung. München: Verlag Dokumentation, 1974. 366 p. (DGD-Schriftenreihe, 3).
5. FOSKETT, Douglas J. Classification for a general index language. London: Library association, 1970. 48 p. (Library association research publication, 2).
6. FOSKETT, Douglas J. Systems theory and its relevance to documentary classification. **International classification**, 1980, vol. 7, n. 1, p. 2-5.
7. FUGMANN, Robert. The glamour and the misery of the thesaurus approach: treatise VI in information retrieval theory. **International classification**, 1974, vol. 1, n. 2, p. 76-86.
8. FUGMANN, Robert. Toward a theory of information supply and indexing: treatise VI on retrieval system theory. **International classification**, 1979, vol. 6, n. 1, p. 3-15.
9. FUGMANN, Robert. On the practical of indexing and its theoretical foundations. **International classification**, 1980, vol. 7, n. 1, p. 13-20.
10. FUGMANN, Robert. The fiveaxiom theory of indexing and information supply. **Journal of the American society for information science**, 1985, vol. 36, n. 2, p. 116-129.
11. GARDIN, Jean-Claude. Eléments d'un modèle pour la description des lexiques documentaires. **Bulletin des bibliothèques de France**, 1966, tome 11, n. 5, p. 171-182.

12. GROLIER, Eric de. Etude sur les catégories générales applicables aux classifications et codifications documentaires. Paris: UNESCO, 1962. 262 p. (Documentation et terminologie scientifiques).
13. JONES, Kevin, P. Towards a theory of indexing. **Jornal of documentation**, 1976, vol. 32, n. 2, p. 118-123.
14. LANCASTER, F. W. Vocabulary control for information retrieval. 2nd ed. Arlington (Virg.): Information resources press, 1986. XVII, 270 p.
15. LANDRY, Bertrand C; RUSH, James E. Toward a theory of indexing. In: Proceedings, vol. 5 / American society for information science, 31st annual meeting, Oct. 20-24, 1967, Columbus, Ohio. New York: Greenwood press, 1968. p. 59-64.
16. LANDRY, Bertrand C.; RUSH, James E. Toward a theory of indexing. II. **Journal of the American society for information science**, 1970, vol. 21, n. 5, p. 358-367.
17. MANIEZ, Jacques. Etude générale des langages documentaires. In: Les langages documentaires et classificatoires: conception, construction et utilisation dans les systèmes documentaires. Paris: Ed. d'organisation, 1987. p. 235-286.
18. RANGANATHAN, Shiyali Ramamrita. Prolegomena to library classification. 3rd. ed. Bombay; Calcutta: Asia publishing house, 1967. 640 p. (Ranganathan series in library science, 20).
19. ROTHMANN, John. Index, indexer, indexing. In: Encyclopedia of library and information science, vol. 11. New York: Dekker, 1974. p. 286-299.
20. SOERGEL, Dagobert. Some remarks on information languages, their analysis and comparison. **Information storage and retrieval**, 1967, vol. 3, n. 4, p. 219-291.
21. SOERGEL, Dagobert. Theoretical problems of thesaurus building with particular reference to concept formation. In: Fachsprache, Umgangssprache / J. S. Petöfi, A. Podlech, E. V. Savigny eds. Kronberg/Taunus: Scriptor-Verlag, 1975. p. 355-381.

22. SVENONIUS, Elaine F. Directions for research in indexing, classification and cataloging. **Library resources and technical services**, 1981, vol. 25, n. 1, p. 88-103.
23. TAUBE, Mortimer; GULL, C. D.; WACHTEL, Irma S. Unit terms in coordinate indexing. **American documentation**, 1952, vol. 3, n. 4, p. 213-218.
24. VICKERY, B. C. Retrieval language models. In: **Information systems**. London: Butterworth, 1973. p. 203-222.
25. WERSIG, Gernot. Eine neue Definition von "Thesaurus". **Nachrichten für Dokumentation**, 1969, vol. 20. n. 2, p. 53-61.

### MODELOS DAS TEORIAS LINGÜÍSTICAS

26. BALLMER, Thomas T. Zur Gewinnung einer Fundamental-klassifikation des menschlichen Wissens. II. **International classification**, 1983, vol. 10, n. 3, p. 127-134.
27. BALLMER, Thomas T. Zur Gewinnung einer Fundamental-klassifikation des menschlichen Wissens. IV. **International classification**, 1984, vol. 11, n. 1, p. 21-27.
28. BEGHTOL, Clare. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. **Journal of documentation**, 1986, vol. 42, n. 2, p. 84-113.
29. BIVINS, Kathleen T. An approach to an atomic theory of indexing. In: **New trends in documentation and information: proceedings / of the 39th FID congress**, University of Edingurgh, 25-28 Sept. 1978; Peter J. Taylor ed. London: ASLIB, 1980. p. 50-55. (FID publication, 566).
30. COYAUD, Maurice. Introduction à l'étude des langages documentaires. Paris: Klincksieck, 1966. 148 p. (T. A. Documents, 1).
31. FUGMANN, Robert. The complementarity of natural and indexing languages. **International classification**, 1982, vol. 9, n. 3, p. 140-144.
- 32a. GARDIN, Jean-Claude. Document analysis and linguistic theory. **Journal of documentation**, 1973, vol. 29, n. 2, p. 137-168.

Traduit dans:

- 32b. GARDIN, Jean-Claude. Analyse documentaire et théorie linguistique. In: Les analyses de discours. Neuchâtel: Delachaux et Niestlé, 1974. p. 120-168.
33. HENRIKSEN, Tor. A linguistic description model for indexing languages. In: Ordering systems for global information networks: proceedings / of the third international study conference on classification research, 1975, Jan. 6-11, Bombay, India; A. Neelameghan ed. Bangalore: International federation for documentation, Committee on classification research (FID/CR): Sadara Ranganathan endowment for library science, 1979. p. 3-8. (Sadara Ranganathan endowment for library science series, 13; FID publication, 533).
34. HUTCHINS, W. J. Languages of indexing and classification: a linguistic study of structures and functions. Stevenage (Herts.): Peregrinus, 1975. 148 p. (Librarianship and information studies, 3).
35. JONES, Kevin P. The environment of classification. Part II. How we classify. *Journal of the American society for information science*, 1974, vol. 25, n. 1, p. 44-51.
36. KIM, Chai. Theoretical foundations of thesaurus-updating. *Journal of the American society for information science*, 1973, vol. 24, n. 2, p. 148-156.
37. LONG, B. Linguistique et indexation. *Documentaliste — Sciences de l'information*, 1980, vol. 17, n. 3, p. 99-106.
- 38a. MELTON, Jessica S. A use for the techniques of structural linguistics in documentation research. Cleveland (Ohio): Western reserve university, Center for documentation and communication research, 1964, 20 p.  
Egalement dans:
- 38b. Classification research: Proceedings / of the second international study conference, Elsinore, Denmark, Sept. 14-18, 1964; Pauline Atherton ed. Copenhagen: Munksgaard, 1965. p. 466-480.
39. NEDOBITY, Wolfgang. Terminology and its application to classification, indexing and abstracting. *UNESCO Journal of information sciences, librarianship and archives administration*, 1983, vol. 5, n. 4, p. 227-234.

40. SCHREIDER, Yu A. The thesaurus in information science and theoretical semantics. **Automatic documentation and mathematical linguistics**, 1971, vol. 5, n. 1, p. 52-57.
41. SPARCK JONES, Karen; KAY, Martin. Linguistics and information science. New York; London Academic press, 1973. 244 p. (FID publication, 492).
42. STOKOLOVA, N. A. Elements of a semantic theory of information retrieval. I. The concepts of relevance and information language. **Information processing and management**, 1977, vol. 13, n. 4, p. 227-234.
43. STOKOLOVA, N. A. Syntactic tools and semantic power of information languages. (PL II of: Elements of a semantic theory of information retrieval). **International classification**, 1976, vol. 3, n. 2, p. 75-81.

#### MODELOS E TEORIAS MATEMATICAS

44. BARNES, R. F. Mathematicological foundations of retrieval theory: general concepts and methods. Bethlehem (Pa): Lehigh university center for the information science, 1965. 29 p.
45. BELING, Gerd. Towards a set theoretical foundation of classification. In: Ordering systems for global information networks: proceedings / of the third international study conference on classification research, 1975, Jan. 6-11, Bombay, India; A. Neelameghan ed. Bangalore: International federation for documentation, Committee on classification research (FID/CR): Sadara Ranganathan endowment for library science, 1979. p. 67-82. (Sadara Ranganathan endowment for library science series, 13; FID publication, 533).
46. DIODATO, Virgil. Modern algebra and information science. **Journal for the American society for information science**, 1983, vol. 34, n. 4, p. 257-261.
47. HELLPRIN, Laurence B. Mathematical model of indexing. Washington: Documentation incorporation, 1957. 26 p.
48. MOOERS, Calvin N. A mathematical theory of language symbols in retrieval. In: Proceedings of the international conference on scientific information. Washington, 1958. p. 57-94.

49. PERRY, James W.; KENT, Allen. A mathematical model system. In: Documentation and information retrieval: an introduction to basic principles and cost analysis. Cleveland (Ohio): The Press of Western reserve university and interscience publishers, 1957. p. 4-25.
50. REBALL, Siegfried. On the application of predicate calculus in information indexing and the degree of indexing languages formalisation. **International forum on information and documentation**, 1978, vol. 3, n. 3, p. 14-17.
51. RICHMOND, Phyllis A. Classification in syndetic structures. In: Universal classification I: Subject analysis and ordering systems: proceedings / 4th International study conference on classification research; 6th Annual conference on Gesellschaft für Klassifikation e. V., Augsburg, 28 June - 2 July 1982; Ingetraut Dahlberg ed. Frankfurt: Indeks Verlag, 1982. p. 149-156.
52. SCHAUBLE, P. Thesaurus based concept spaces. In: Proceedings / of the tenth Annual international ACM. SIGIR conference on research and development in information retrieval. New York: ACM press, 1987. p. 254-262.
53. SCHREIDER, Yu A. The logic of classification. **Automatic documentation and mathematical linguistics**, 1973, vol. 7, n. 2, p. 30-36.
54. SCHREIDER, Yu A. The duality of classification: taxonomy and meronomy. **International forum on information and documentation**, 1981, vol. 6, n. 1, p. 3-10.
55. SOERGEL, Dagobert. Mathematical analysis of documentation systems: an attempt to a theory of classification and search request formulation. **Information storage and retrieval**, 1967, vol. 3, p. 129-173.
56. SRIDHAR, M. S. A mathematical approach to relations in thesauri. **Journal of library and information science (India)**, 1980, vol. 5, n. 1, p. 76-87.
57. SYDORCHENKO, V. D. On the construction of a descriptor model. **Annals of library science and documentation**, 1974, vol. 21, n. 4, p. 142-145.
58. TAUBE, Mortimer; WACHTEL, Irma S. The logical structure of coordinate indexing. **American documentation**, 1953, vol. 4, p. 67-68.

59. TURSKI, W. M. On a model of information retrieval based on thesaurus. **Information storage and retrieval**, 1971, vol. 7, n. 2, p. 89-94.

#### RELAÇÃO ENTRE TERMOS DE INDEXAÇÃO

60. AUSTIN, Derek. Basic concept classes and primitive relations. In: **Universal classification 1: Subject analysis and ordering systems: proceedings / 4th International study conference on classification research; 6th Annual conference of Gesellschaft für Klassifikation e. V., Augsburg, 28 June - 2 July 1982; Ingetraut Dahlberg ed. Frankfurt: Indeks Verlag, 1982. p. 86-94.**
- 61a. COATES, E. J. Significance and term relationship in compound headings. In: **Subject catalogues. London: Library association, 1960. p. 50-64.**  
Repris dans:
- 61b. Theory of subject analysis: a source book / Lois Mai Chan, Phyllis A. Richmond, Elaine Svenonius eds. Littleton (Col.): Libraries unlimited, 1985. p. 183-195.
62. COATES, E. J. Some properties of relationships in the structure of indexing languages. **Journal of documentation**, 1973, vol. 29, n. 4, p. 390-404.
63. GULL, C. D. Structure of indexing authority lists. **Library resources and technical services**, 1966, vol. 10, n. 4, p. 507-511.
64. KERVÉGANT, Désiré. Classification et analyse des relations. **Bulletin des bibliothèques de France**, 1959, vol. 4, n. 11, p. 495-511.
65. MANIEZ, Jacques. Relationships in thesauri: some critical remarks. **International classification**, 1988, vol. 15, n. 3, p. 133-138.
66. NEEDHAM, R. M.; SPARCK JONES, Karen. Keywords and clumps. **Journal of documentation**, 1964, vol. 20, n. 1, p. 5-15.
67. WILLETS, M. An investigation of the nature of the relation between terms in thesauri. **Journal of documentation**, 1975, vol. 31, n. 3, p. 158-184.

## RELAÇÕES SEMANTICAS

68. BEGHTOL, Clare. Semantic validity: concepts of warrant in bibliographic classification systems. **Library resources and technical services**, 1986, vol. 30, n. 2, p. 109-125.
69. DEWÈZE, André. Réseaux sémantiques: essai de modélisation — Application à l'indexation et à la recherche de l'information documentaire. Lyon, 1981.  
Th. univ.: Lyon 1: 1981.
70. FROHMANN, Bernard P. An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG. **Cataloging and classification quarterly**, 1983, vol. 4, n. 1, p. 11-27.
71. GHOSE, A.; DHAWLE, A. S. Problems of thesaurus construction. **Journal of the American society for information science**, 1977, vol. 28, n. 4, p. 211-217.
72. JANSEN, R. Die Bedeutung der Zugehörigkeitsrelation für Thesaurusstruktur, Indexierungsmethode und Recherche. In: Deutscher Dokumentartag, 1974. Band 1, Bonn, 7-11 Okt 1974 / Deutsche Gesellschaft für Dokumentation. München: Verlag Dokumentation, 1975. p. 353-374.
73. POPOVSKA, H. Synonymic and hierarchical relations in descriptor languages. **Automatic documentation and mathematical linguistics**, 1985, vol. 19, n. 1, p. 74-87.
74. SOERGEL, Dagobert. Index language structure. I: conceptual. In: Organizing information: principles of data base and retrieval systems. Orlando (Fl.): Academic press, 1985. p. 251-287.
75. SYDORCHENKO, V. D. Typological classification of thesaurus semantic structures. **Automatic documentation and mathematical linguistics**, 1975, vol.9, n. 2, p. 16-23.
76. YU, Clement T.; RAGHAVAN, Vijay V. Single-pass method for determining the semantic relationships between terms. **Journal of the American society for information science**, 1977, vol. 28, n. 6, p. 345-354.

## RELAÇÕES SINTÁTICAS

77. AITCHINSON, J. Thesaurfacet: a new concept in subject retrieval schemes. In: Subjects retrieval in the seventies: new directions: proceedings / of an international symposium, University of Maryland, 1971; H. Wellisch, T. D. Wilson eds. Westport (Ct.): Greenwood press, 1972. p. 72-98.
78. COSTELLO, J. C. A basic theory of roles as syntactical control devices in coordinate indexes. **Journal of chemical documentation**, 1964, vol. 4, p. 116-124.
79. HUTCHINS, W. J. Facets, roles and cases. In: Informatics 1. Proceedings / of a conference held by the ASLIB co-ordinate indexing group on 11-13 April 1973 at Durham University. London: ASLIB, 1974. p. 89-97.
80. MANIEZ, Jacques. Le rôle de la syntaxe dans les systèmes de recherche documentaire. Tome I, Aspects linguistiques. Tome II, Etude critique de quelques SRD. Dijon: IUT de Dijon, Département carrières de l'information, 1976, VI-184 + 182-XXX p.

## LINGUAGENS E SISTEMAS DE INDEXAÇÃO

### Indexação em cadeia

81. ARMITAGE, Janet E.; LYNCH, Michael F. Some structural characteristics of articulated subject indexes. **Information storage and retrieval**, 1968, vol. 4, n. 2, p. 101-111.
82. BATTY, C. D. Chain indexing. In: Encyclopedia of library and information science, vol. 4. New York: Dekker, 1970. p. 423-434.
83. BHATTACHARYYA, G. Chain procedure and structuring of a Subject. **Library science with a slant to documentation**, 1972, vol. 9, n. 4, p. 585-635.
84. CRAVEN, Timothy C. String indexing. ORLLANDO, Fl. Academic Press, 1986, 246 p.
85. FARRADANE, Jason E. L. Acomporison of some (computer-produced), permuted alphabetical subjectindexes. **International classification**, 1977, vol. 4, n. 2, p. 94-101.

86. WINEUR, B. W. Relations in trains. **Journal of librarianship**, 1973, vol. 5, n. 3, p. 175-202.
87. RAJAN, T. N. From chain to string indexing. **Library herald**,, 1981-1982, vol. 20, n. 2-4, p. 105-110.
88. RAJAN, T. N., ed. Indexing sugstems: concepts, models and techmiques. Calcuta: Indian association ob special libraries and information centres, IASLIC, 1981, 270 p.

#### PRECIS E POPSI

89. AUSTIN, Derek. An information retrieval language for MARC. **Aslib proceedings**, 1970, vol. 22, n. 10, p. 481-491.
- 90a. AUSTIN, Derek; DIGGER, Jeremy A. **PRECIS: the preserved context index system. Library resources and technical services**, 1977, vol. 21, n. 1, p. 13-30.
- Repris dans:
- 90b. **Theory of subject analysis: a source book / Lois Mai Chan, Pyllis A. Richmond, Elaine Svenonius eds. Littleton (Col.): Libraries unlimited, 1985. p. 372-389.**
91. BHATTACHARYYA, G. POPSI: its fundamentals and procedure based on a general theory of subject indexing languages. **Library science with a slant to documentation**, 1979, vol. 16, n. 1, p. 1-34.
92. BHATTACHARYYA, G. Some significant results of current classification research in India. **International forum on information and documentation**, 1981, vol. 6, n. 1, p. 11-18.
93. BISWAS, Subal C. Two decades of POPSI, 1969-1988: a literature review. **Libri**, 1988, vol. 38, n. 4, p. 279-296.
94. CHETI, A. Sistemi d'indicizzazione preordinati: indicizzazione a catena, PRECIS, POPSI, NEPHIS. **Bolletino d'informazioni / Associazione italiana biblioteche**, 1984, vol. 24, n. 1, p. 19-44.
95. DEVADASON, F. J. Postulate-based permuted subject indexing language as a metalanguage for computer-aided generation of information retrieval thesaurus. **International forum on information and documentation**, 1983, vol. 8, n. 1, p. 22-29.

A. RIVIER

96. MAHAPATRA, M.; BISWAS, S. C. **PRECIS; its theory and application.** An extented state-of-the-art review from the beginning up to 1982. **Libro**, 1983, vol. 33, n. 4, p. 316-330.
97. SORENSEN, Jutta; AUSTIN, Derek. **PRECIS in a multilingual context. Part 2. A linguistic and logical explanation of the syntax.** **Libri**, 1976, vol. 26, n. 2, p. 108-139.

#### OUTROS SISTEMAS

98. CRAVEN, Timothy C. **NEPHIS: a nested phrase indexing system.** **Journal of the American society for information science**, 1977, vol. 28, n. 2, p. 107-114.
99. CRAVEN, Timothy C. **Linked phrase indexing.** **Information processing and management**, 1978, vol. 14, n. 6, p. 469-476.
100. FARRADANE, Jason E. L. **A scientific theory of classification and indexing and its practical applications.** **Journal of documentation**, 1950, vol. 6, n. 2, p. 83-99.
101. FARRADANE, Jason E. L. **A scientific theory of classification and indexing: further considerations.** **Journal of documentation**, 1952, vol. 8, n. 2, p. 73-92.
102. FARRADANE, Jason E. L. **Relational indexing and classification in the light of recent experimental work in psychology.** **Information storage and retrieval**, 1963, vol. 1, n. 1, p. 3-11.
103. FUGMAN, Robert; NICKELSEN, Herbert; NICKELSEN, Ingeborg et al. **Representation of concept relations using the TOSAR system of the IDC. Treatise III on information retrieval theory.** **Journal of the American society for information science**, 1974, vol. 25, n. 5, p. 287-307.
- 104a. KAISER, Julius Otto. **Systematic indexing.** In: **Report of proceedings / ASLIB, 3rd conference, Oxford, Sept 24-27, 1926.** London: ASLIB, 1927. p. 20-33.  
Repris dans:
- 104b. **Readings in library cataloging / R. K. Olding ed.** London: Crosby Lockwood, 1966. p. 145-162.

105. NICKELSEN, Herbert; NICKELSEN, Ingeborg. Mathematische Analyse des TOSAR-Verfahrens. *Information storage and retrieval*, 1973, vol. 9, n. 2, p. 95-119.
106. SVENONIUS, Elaine. Facet definition: a case study. *International classification*, 1978, vol. 5, n. 3, p. 134-141.

**OBSERVAÇÃO:** Pode-se notar que mais da metade dos artigos referenciados está concentrado em quatro periódicos:

- *Journal of the American Society for Information Science*
- *International Classification*
- *Information Processing and Mangement (continuação de Information Storage and Retrieval)*
- *Journal of Documentation*

Recebido para publicação em 28/05/92.

### ABREVIÇÕES UTILIZADAS NO TEXTO

- BNB — British national bibliography
- CC — Colon classification
- CRG — Classification Research Group
- DDC — Dewey Decimal Classification
- KWIC — Keyword in context
- LCC — Library of Congress Classification
- LI — Linguagem de indexação
- LIPHIS — Linked phrase indexing system
- LN — Linguagem natural
- NEPHIS — Nesthed phrase indexing system
- POPSI — Postulate-based permuted subject indexing
- PRECIS — Preserved contex indexing system
- SIL — Subject indexing language
- SYNTOL — Syntagmatic organization language
- TOSAR — Topological presentation of synthetic and analytical