

# arquivo & administração

ÓRGÃO OFICIAL DA ASSOCIAÇÃO  
DOS ARQUIVISTAS BRASILEIROS

V. 2 — N.º 1 — ABRIL 1974



55 Clas. PER  
& Administração

ex. 2

PER. Currículo Superior de Arquivo

359

# ASPECTOS SEMÂNTICOS PESQUISA DOCUMENTAL

Prof. Antonio Garcia de Miranda Netto  
(da Fundação Getúlio Vargas)

Para análise dos documentos e sua colocação em memória, precisamos de um sistema capaz de permitir, com mínimo de erro e máximo de economia, "classificar" a informação dentro de um modelo que possa ser posto em correspondência com um modelo de "busca" ou "recuperação". Esses "termos", isolados ou correlacionados, podem ser chamados, de um modo geral, como o faz Costello (1) do Battelle Memorial Institute, de "localizadores" (*locators*). Eles indicam, sejam os descritores (*Moers*), as palavras-chave (*Luhn*), os unitermos (*Taube*) ou os termos de estrutura (*Whaley*) a "posição" de conceitos unitários utilizados na recuperação. Já os *Thesauri*, como veremos, constituem um grau a mais de complexidade. Vamos apresentar três tipos de ordenação, de flexibilidade e complexidade crescente:

a) Ordem **linear** em que os elementos da documentação obedecem a uma seqüência, mediante um critério exterior à sua situação semântica: ordem alfabética, número de sílabas ou qualquer outro critério semelhante.

b) Ordem em **ramificação** — bidimensional, com a presença de categorias e subcategorias, de que são exemplo clássico as classificações de Dewey e a de Olet e Lafontaine. Também poderemos chamá-la de *árvore*.

c) Ordem em **reticulado** — pluridimensional, que leva em conta não só a hierarquia dos elementos em um repertório (ordem em ramificação) mas as associações entre esses mesmos elementos. É, por excelência, a ordem dos *Thesauri*.

## UM POUCO DE HISTÓRIA

Peter Mark Roget (1779-1869) ao publicar o seu *Thesaurus of English Words and Phrases*, em 1852, torna-se, historicamente, o pioneiro deste tipo de ordenação, embora não o primeiro a usar a palavra para indicar repertório ordenado de idéias ou conhecimentos. Brunetto Latini (1220? - 1249?) também chamado Brunettus Florentinus, mestre de Dante, usara a palavra como título de um poema didático alegórico, composto em "settenari accoppiati"

Il *Tesoretto* — que narra uma viagem pelos três reinos da natureza. Este poema é modelo evidente da obra futura do florentino, *La Commedia* mais tarde chamado *A Divina Comédia*. exilado em França, depois da derrota dos guelfos, Brunetto lá escreve, além de Il *Tesoretto*, um tratado, *Les Livres dou Tresor*, primeira

enciclopédia em vulgar, que era ordenada por assuntos correlacionados, e não, alfabeticamente.

Entre *Les Livres dou Tresor* de Brunetto Latini e o *Thesaurus* de Roget, houve em toda a Europa uma verdadeira inundação de *Thesauri*, entre eles os: *Thesaurus Musicus*, editado em 1546 em Nuremberg; o *Thesaurus Sacrorum Rituum* de Bartolomeo Gavanti (1622); o *Thesaurus Novus Anedotorum* de Martène e Durand (1717); e ainda o *Thesaurus Commentationum* de Voldebing (1846).

No *Thesaurus*, Roget, ao contrário de seus ilustres predecessores, teve o cuidado de estabelecer uma relação de correspondência semântica entre vocábulos. É um "dicionário de termos e idéias", que nos parece a característica essencial de um *Thesaurus*. Depois dos trabalhos de Noam Chomsky no campo da lingüística gerativa, ou como ele a chama — lingüística cartesiana (2) — nota-se uma tendência a eliminar a semântica, para insistir na formalização completa da linguagem natural, com a esperança de dar-lhe a univocidade das linguagens de informação. Mas aqui deveremos ter em mente a advertência de René Thom: a formalização, em si mesma, separada de um conteúdo inteligível, não pode ser fonte de conhecimento.

UM NOVO CONCEITO

Q **Ausdehnungslehre** (Tratado da Extensão) de Hermann Guenther Grassmann (1844 e o **Thesaurus** de Roget (1852) foram o ponto de partida para uma nova concepção de ordem, baseada na correlação semântica.

Grassmann traz à matemática (e à lógica, poderíamos acrescentar), um conceito novo: cada construto (construção ideal do pensamento) será ou um ato simples de produção (**Erzeugen**), ou um ato complexo de encadeamento (**Verknüpfung**).

O **Erzeugen** é a **poiesis** que situa o objeto em um ponto dos parâmetros de tempo e espaço.

A **Verknüpfung** é o estabelecimento de uma rede ligando os diferentes **Erzeugnisse**, em um espaço que tem um grau a mais de complexidade. Com isso se passa da noção intuitiva de **posição** e de **métrica** a uma noção mais abstrata de **estrutura**.

Dando um grau maior de refinamento à essa definição de estrutura, poderemos dizer com Warusfel (o **Dictionnaire** — 1966) que, “a estrutura de um conjunto é definida pelas operações e relações nele utilizadas” (3).

Surge assim a estrutura como uma espécie de **propriedade** de um conjunto, o que é extraordinariamente importante para fixar o sentido semântico dos **Thesauri**. O que disse Grassmann, em 1844, sem o rigor de definição de Warusfel, mais de um século depois, nada mais é do que a formalização da idéia de Roget.

No **Thesaurus** introduz-se um novo conceito, que equivale a uma nova dimensão no espaço de um dicionário comum e que poderíamos chamar de semântica cruzada. Em Roget, as idéias são classificadas em uma ordem em reticulado e a cada uma das idéias faz corresponder não só os vários termos que a ela se correlacionam, como ainda a idéia oposta.

O projeto de Roget era construir colunas tríplexes, com a idéia positiva, a idéia negativa e a idéia neutra, mas não pôde transformar o pensamento em realidade pelas dificuldades que explica no prefácio da obra (4). Roget nos dá um exemplo do que desejaria fazer, alinhando em tres colunas

Positivo	Neutro	Negativo
Identity	Difference	Contrariety
Beginning	Middle	End
Past	Present	Future

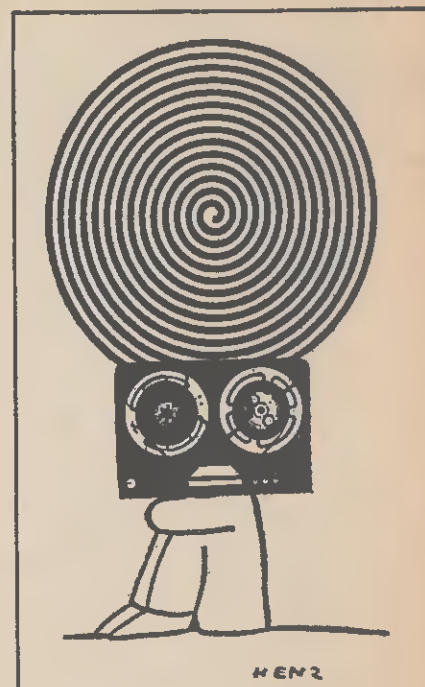
Em outros casos, diz Roget, a palavra intermediária é simplesmente a negação de cada um dos opostos:

Convexity	Flatness	Concavity
Desire	Indifference	Aversion

Outras vezes, a palavra intermediária representa o padrão para comparar os extremos, como no caso de:

Insufficiency	Sufficiency	Redundance
---------------	-------------	------------

onde o termo médio é igualmente oposto, de um lado a insuficiência, de outro a redundância.



Embora o extraordinário desenvolvimento dos estudos de signo, sinal e significado, que marca o pensamento contemporâneo, nos possa fazer crer que os conceitos de Roget são insuficientes para traduzir a complexidade do real, não se pode negar que representaram, em seu tempo, uma extraordinária visão do que se pode fazer no campo da classificação semântica dos elementos da documentação, que são as palavras.

A CHAVE PARA UM TESOURO

Um dos **Thesauri** mais simples e mais gerais, que apareceram nos últimos tempos, é o que foi organizado pelo New York Times, como tentativa de criar um repertório de assuntos gerais. No prefácio podemos ler: “... a espécie de **Thesaurus** que se desenvolveu na última década não é tanto um tesouro como a chave para um tesouro.



A verdadeira riqueza está nos arquivos de informação — livros, panfletos, relatórios, fotografias ou jornais — e o **Thesaurus** representa um instrumento para sua plena exploração.

Nada mais é do que mecanismo para ordenar e controlar o arquivo, de modo a permitir que novos itens possam ser adicionados, logicamente, à relação dos itens já existentes e, mais ainda, que todos os itens importantes se tornem acessíveis, rápido e adequadamente." (5).

Os sistemas de recuperação da informação tem como finalidade a transformação de um repertório finito de conhecimentos, criado no passado, em tempo finito, sob a forma de registro inscrito (memória) em um outro repertório específico e finito, produzido em tempo futuro e indeterminado, para satisfazer os requisitos formulados pelo usuário da informação (6).

Ora, para tal processo é necessária a coordenação dos conceitos contidos no **index** do documentalista.

Os **Thesauri** visam pois, fundamentalmente, a diminuir o que Maloney chama de **perda semântica** no processo (7).

Concluimos, apresentando a definição de **Thesaurus** dada pelos autores de uma das linguagens de

informação mais refinadas, o SYNTOL (Syntagmatic Organization Language): "**Thesaurus** (8) sentido geral: dicionário das palavras de uma língua natural, agrupada por afinidades **semânticas**, segundo critérios variáveis. Sinônimo: dicionário conceitual. Sentido particular (automação documental): dicionário de equivalências absolutas ou condicionais, entre as palavras ou expressões de uma língua natural e os **descritores** de uma linguagem documental" (Grifo nosso)

Poderemos pois dizer que o aspecto semântico é fundamental em um **Thesaurus**, tanto no sentido acima apontado, que é geral, como no sentido mais restrito. Convém, entretanto, deixar bem claro que, diante da importância e do progresso cada vez maior do processamento eletrônico, a construção dos **Thesauri** de transformação (língua natural — linguagem de informação) é um dos capítulos fundamentais da documentalística •

#### BIBLIOGRAFIA

- (1) Costello, Jr., John C. **Indexing in Depth: Practical Parameters in Information Handling, First Principles**. Washington, Spartan Books, 1963.
- (2) Chomsky, Noam. **Cartesian Linguistics: A Chapter in the History of Rationalistic Thought**. New York, Harper & Row, 1966.
- (3) Para a definição da palavra **Structure**, ver: Warusfel, André. **Dictionnaire Raisoné de Mathématiques**. Paris, Seuil, 1966.
- (4) Roget, Peter. **Thesaurus of English Words and Phrases**. Introduction, London, J. M. Dent & Sons, 1946. vol. I.
- (5) The New York Times **Tresaurus of Descriptors**. New York, New York Times Co., 1968.
- (6) Ver referência 8
- (7) Maloney, Clifford J. **Semantic Information**. **American Documentation**, Jul. 1962, p. 276-287.
- (8) Para um desenvolvimento maior de **Linguagens de Informação**, ver: Miranda Netto, A. de. **Terminologia Científica e Técnica, barreira à Comunicação e à Automação**. Rio de Janeiro, 2.º Congresso Regional sobre Documentação, 9.ª Reunião da FID/CLA, 1969.