Contribuição da terminologia para a elaboração de tesauros

Maria de Fátima G. Moreira Tálamo
Marilda Lopes Ginez de Lara
Nair Yumiko Kobashi

Resumo

O tesauro é visto tradicionalmente como instrumento de representação e de controle terminológico, no qual, as unidades de entrada são organizadas segundo princípios lógico-semânticos. Contudo, a ausência de uma definição precisa das propriedades dessas unidades compromete o tesauro na sua função de representação e de controle. Sugere-se estabelecera interface entre terminologia e análise documentária para subsidiar a formulação de metodologias de elaboração de linguagens documentárias mais eficazes.

Palavras-chave

Interface terminologia/linguagens documentárias; Terminologia/tesauros; Análise documentária/terminologia.

O TESAURO: USOS, FUNÇÕES E PRINCÍPIOS DE ORGANIZAÇÃO

As linguagens documentárias são tradicionalmente consideradas instrumentos de controle terminológico que atuam em dois níveis: a) na representação da informação obtida pela análise e síntese de textos; b) na formulação de equações de busca de informação.

Sob esse enfoque, pode-se considerar que a questão crucial – e também a mais controvertida – entre documentos e linguagens documentárias expressa-se nos princípios que regulam as operações de representação. Tal relação será, neste trabalho, observada e discutida a partir de uma linguagem específica: o tesauro.

A construção dos tesauros apóia-se, basicamente, em dois conjuntos referenciais: de um lado, no conhecimento categorizado em assuntos e, de outro, em um *corpus* discursivo do qual são retirados os termos considerados significativos. De um modo geral, supõe-se que a adoção desses procedimentos possa assegurar a eficácia dos tesauros enquanto instrumentos de controle terminológico para representar e recuperar informação.

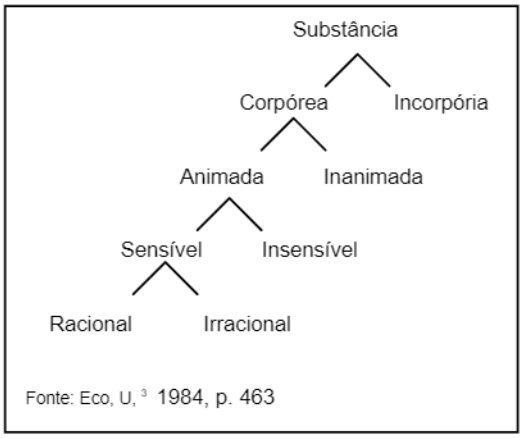
Especificamente, o primeiro conjunto anteriormente referido pretende garantir, de forma classificatória e preditiva, a organização do universo objeto de representação; no segundo, a "garantia literária"[1] procede à seleção dos conceitos mais freqüentes para a representação da informação. Garante-se, desse modo, a contigüidade e semelhança entre o vocabulário a ser utilizado para a representação e o universo do conhecimento expresso nos textos.

Pode-se afirmar que ambos os procedimentos pressupõem, de algum modo, operações de denominação, já que não se resumem no mero reconhecimento de palavras, mas visam, antes de tudo, a determinar o arranjo das significações dos termos selecionados. Nesse sentido, o controle terminológico se faz através do reconhecimento das propriedades dos termos, mesmo que tal reconhecimento se faça, na maioria das vezes, de forma intuitiva.

O processo de denominação em análise documentária envolve, por sua vez, operações de equivalência lexical que visa a assegurar a invariante da informação. Jean-Claude Gardin chama a atenção sobre esse fato, ao afirmar que "a irregularidade entre significante e significados dá lugar a uma normalização dos primeiros fundada sobre a invariância relativa que se empresta aos segundos"[2].

Uma vez selecionados, *os* termos que irão compor o tesauro submetem-se a uma organização determinada. Em larga medida, a macroorganização dos termos pressupõe uma divisão hierárquica do conhecimento, tal como se verifica nos sistemas tradicionais de classificação bibliográfica.

Costuma-se dizer que essa organização inspira-se na lógica aristotélica, na qual as subdivisões das classes obedecem a uma diferenciação que postula a inclusão ou não-inclusão de uma dada qualidade. O princípio de divisão dicotômica utilizado faz derivar apenas duas classes: A e não-A, tal como ocorre na árvore de Porfírio. Essa organização pressupõe, basicamente, as relações de gênero/espécie e todo/parte e, conseqüentemente, relações de implicitação, em que "os nós baixos da classificação implicitam necessariamente os nós altos"[3]. Exemplificando: "... se uma substância for racional, será necessariamente animada; se for animada, será necessariamente corpórea e, se corpórea, será substância"[3]. Em forma de árvore, veríamos as seguintes articulações:



Fonte: Eco, U, [3] 1984, p. 463

A árvore de Porfírio, entretanto, é uma construção que depende da escolha da qualidade que subsumará as subdivisões sucessivas: "... posso pôr o homem na árvore de Porfírio na qual se considere sua natureza orgânica ou inorgânica, e não na natureza moral. O homem é necessariamente mortal somente numa árvore de Porfírio particularmente focalizada no problema da duração da vida"[3].

Portanto, a árvore de Porfírio é selecionada "sob pressão" contextual, podendo ser considerada verdadeira apenas em relação a um determinado código, e não em relação às propriedades dos objetos em si mesmos. Ela é uma árvore que estrutura a substância semântica, representando-a como traços hierarquizados. Tal hierarquia supõe apenas um contexto, considerado o suporte prévio para a interpretação discursiva, qualquer que seja o seu objeto textual.

Embora o tesauro possa se prestar a uma representação gráfica semelhante à da árvore de Porfírio, nada nos autoriza a proceder a uma associação direta entre ambos. Isso, porque, enquanto a árvore de Porfírio supõe uma definição clara das qualidades envolvidas em cada nó hierárquico (materialidade, ou sensibilidade, ou acionalidade etc.), o tesauro, muitas vezes, relega a classificação proposta de início – preditiva, portanto – a um segundo plano.

Ao privilegiar o enfoque centrado na própria relação existente entre os descritores, o tesauro aproxima-se dos procedimentos utilizados nas teorias de classificação mais modernas, ditas facetas (Ranganathan, E. De Grolier, J. Kaiser, B.C. Vickery). Tais procedimentos, por permitir associações de descritores de diferentes naturezas (já que prevê a assimilação dos princípios da análise facetada, com a inclusão dos predicativos aristotélicos), ampliam as possibilidades de análise dos conteúdos. Por outro lado, a ausência de uma definição clara de seus fundamentos, decorrente da coexistência desses dois princípios de organização, explica a dificuldade do tesauro de proceder a uma efetivo controle terminológico.

Desse modo, a organização hierárquica fundamental do tesauro funciona de forma dicionarial. Ela pressupõe uma definição do tipo "se > então", calcada num código de natureza institucional: seus limites dependem dos contratos de natureza social, supondo modelos socioculturais e ideológicos determinados. No extrato a seguir, World Bank é necessariamente um Banco de Desenvolvimento, ou seja, se World BanK, então Banco de Desenvolvimento.

DEVELOPMENT BANKS
BANQUES DE DEVELOPMENT/BANCOS DE DESARROLLO — 11.02.02
TT: AID INSTITUTIONS
FINANCIAL INSTITUTIONS
BT: AID INSTITUTIONS
BANKS NT: AFRICAN DEVELOPMENT BANK
ASIAN DEVELOPMENT BANK
CABEI
CARIBANK
IDB
WORLD BANK
RT: AGRICULTURAL BANKS
AID FINANCING
INDUSTRIAL BANKS

Fonte: Macrothesaurus,[4],1985, p. 444

Na verdade, a inclusão do Banco Mundial entre os bancos de desenvolvimento pressupõe a aceitação da definição de "desenvolvimento" proposta pela instituição responsável pela elaboração do Macrotesauro (OCDE e as Nações Unidas). Outra conceituação de "desenvolvimento" poderia implicar a exclusão do referido banco do conjunto das instituições voltadas para o desenvolvimento.

As questões até aqui apresentadas permitem-nos afirmar que o tesauro é um objeto cultural que registra e representa o conhecimento segundo parâmetros estáveis e previamente determinados. Tais parâmetros, que se materializam sob a forma de redes de relações entre descritores, determinam não só o modo de organização e disseminação da informação, mas também a análise dos textos. Neste último caso, dado o volume de informação a ser indexado, tradicionalmente se utiliza o tesauro como grade interpretativa dos textos. Descritores únicos passam, então, a representar informações de natureza diferente. Como resultado, cada descritor acaba por recobrir o espectro bastante amplo de significados, no qual cabem informações contraditórias e contrárias. Compromete-se, assim, não só a representação e a recuperação da informação específica de cada texto, mas também a eficácia do próprio instrumento na sua função representativa.

Na prática, o uso do tesauro fica comprometido pelo aparecimento de qualidades do texto individual que não são passíveis de serem enquadrados em parâmetros prévios e preditivos. Assim, no lugar de uma análise da significação discursiva com referência às circunstâncias de emissão, supõe-se uma interpretação amarrada em definições conceituais (das propriedades da palavra) quase sempre obscuras ou intuídas, já que, muitas vezes, as relações semânticas entre os descritores não são suficientemente claras e rigorosas. Perde-se, desse modo, a informação específica e individual do texto, em prol de uma atribuição de sentido prevista e sedimentada fora das circunstâncias de enunciação.

Além disso, e por essas razões, a representação em forma de dicionário implica, também, a discussão sobre o verdadeiro eu falso. A veracidade de uma asserção depende de um código que pressuponha sua aceitação. Ela reflete os recortes privilegiados no campo conceitual considerado. No caso dos tesauros, esse recorte é delimitado, inicialmente, pela escolha de um princípio de organização do conhecimento. À organização derivada dessa escolha, são acoplados os termos e conceitos resultantes de um novo recorte feito na literatura da área – que funciona como fonte para a seleção dos termos e conceitos mais recorrentes – a já citada "garantia literária".

Ao se realizarem esses sucessivos recortes para a construção dos tesauros, opera-se simultaneamente uma escolha subjacente de natureza institucional. Essa seleção fica evidente, por exemplo, no caso do descritor "Bancos de Desenvolvimento", mencionado anteriormente, em que uma determinada concepção define e garante a inclusão dos tipos de instituições que dele farão parte.

AS LIMITAÇÕES DO TESAURO PARA CONTROLAR TERMINOLOGIA E REPRESENTAR INFORMAÇÃO

Os recortes sucessivos, já referidos, são também consubstanciados no momento da construção dos tesauros, através de signos lingüísticos (que passam a constituir os signos documentários), resultantes de um processo onomasiológico. Esse procedimento, segundo Hjelmslev, "parte da consideração da substância do conteúdo (conceito) para chegar à forma de conteúdo (signos lingüísticos que correspondem ao recorte do campo conceitual[5]).

A identificação da palavra com o conceito, entretanto, tem seu valor teórico fundado "na ilusão ingênua de uma equivalência absoluta – para não dizer identidade – entre signo lexical e conceito"[6], pressupondo também uma estabilidade e coerência do conceito que só se verifica ao nível terminológico.

Desse modo, a rede de relações prescritas pelos tesauros supõem a sua manifestação nos textos que estão por aparecer. Como, via de regra, tal materialização só ocorrerá nos discursos assentados em recortes semelhantes, a análise e representação dos textos que participam de outros paradigmas (por exemplo, contraditórios) introduzem um processo de semiose, cujos interpretantes não previstos de início

passam a ser articulados quase que livremente (sem controle), convivendo com hierarquias preestabelecidas (supostamente controladas). Confundem-se, desse modo, "predicados livres" com "predicados relacionados", conceitos propostos por Le Guern[7], ao observar a relação entre terminologia e léxico.

Como instrumento de controle terminológico de natureza dicionarial, espera-se do tesauro uma definição precisa das propriedades dos desertores. No entanto, não é isso o que ocorre na prática. Os tesauros estruturam seus desertores (termos preferenciais) de forma relacional, compreendendo três espécies principais de relações semânticas: hierarquia, equivalência e correlação. A partir de uma classificação geral – identificada na estrutura hierárquica e nas listas de grupos de descritores, que variam segundo a especificidade da área objeto e de um tesauro para outro –, os termos são dispostos segundo a extensão relativa dos conceitos correspondentes aos desertores: TT (Top Term), TG (Termo Genérico) e TE (Termo Específico), correspondendo à aplicação das divisões gênero/espécie e todo/parte, principalmente. As relações de equivalência operam ao nível da sinonímia lingüística e documentária, representadas pelos símbolos Use e UP (Usado para); as correlações agrupam, sem explicitar o tipo de relação em jogo, as associações entre os termos, sendo representado pelos símbolos TR (Termo Relacionado) ou TA (Termo Associado).

A organização dos termos escolhidos para *a* representação dos conceitos de área é feita de forma a permitir a inferência de seu significado (sua extensão e intensão), através da posição do termo/conceito na estrutura do tesauro. Assim, o significado de cada unidade será expressa pela sua relação com as demais unidades participantes de um mesmo domínio. Portanto, as unidades de entrada dos tesauros têm seus significados atrelados à posição relacional que ocupam no interior do instrumento. Supõe-se, nesse caso, que uma definição prévia já tenha sido atribuída a essas unidades. No entanto, raramente os tesauros são acompanhados por glossários. Na prática, apenas algunas unidades de entrada têm notas explicativas (notas de escopo) relativas ao seu âmbito de aplicação.

A partir do que foi exposto, é lícito questionar as funções de representação é de controle terminológico conferidas ao tesauro. Apenas ficando em relações lógico-semânticas, é ele capaz de normalizar efetivamente a representação dos conceitos/informação de uma área do conhecimento? A nosso ver, a resposta a essa questão é negativa, pois o tesauro somente exercerá as funções anteriormente citadas, caso se faça acompanhar de um conjunto de definições, pelo menos dos termos específicos da área considerada.

A INTERFACE ENTRE TERMINOLOGIA E ANÁLISE DOCUMENTÁRIA

O tesauro, para se constituir como instrumento efetivo de controle terminológico e de representação de informação, deve sustentar-se em um dicionário terminológico, a partir do qual a rede de relações lógico-semânticas entre os desertores definir-se-ia com maior clareza, beneficiando tanto o trabalho de indexação, quanto a elaboração de equações de busca. Para isso, é preciso estabelecer a interface entre terminologia (estudo dos léxicos especializados) e análise documentária, cujo estudo deve subsidiar de maneira crescente a formulação de metodologias para o uso e a elaboração das linguagens documentárias.

A primeira contribuição desse campo interdisciplinar diz respeito à natureza específica da unidade significante documentária e sua relação com a unidade significante terminológica. Ao contrário do que ocorre com as unidades da linguagem documentária, a definição das unidades terminológicas – denominadas termos – é feita a partir do discurso, e não da língua. Isto quer dizer que o termo tem um conceito específico e relacional, impondo-se como predicado relacionai. Já as unidades das linguagens documentárias definem-se como predicados livres. Cabe à terminologia, desse modo, operar ao nível sintático-semântico, produzindo terminologias específicas de acordo com o estado-da-arte de cada campo considerado. Tais repertórios ou listas de termos especializados de um domínio particular são acompanhados de definições que remetem o termo ao seu referente. Podem ser, por sua vez, reagrupados segundo uma classificação alfabética ou temática. Sob esse ponto de vista, as terminologias exercem a função comunicativa. Destinam-se à consulta de especialistas, servindo de intermediárias entre os referentes ou a realidade dos objetos e os utilizadores[8].

Já os tesauros são repertórios ou listas de termos autorizados, constituídos por unidades – desertores e não desertores – pertencentes a um domínio particular do conhecimento, relacionadas semântica e logicamente. São utilizados para caracterizar tanto o conteúdo de um documento, quanto o conteúdo das questões propostas pelos usuários. Sua função é, por isso, a de servir como intermediária entre os documentos e os usuários.

Além disso, e talvez seja esse o fato mais importante, os termos pertencem à língua especializada, enquanto os desertores não o são necessariamente. Na verdade, enquanto os termos são extraídos obrigatoriamente do discurso, os desertores são unidades preferencialmente conhecidas dos usuários. De uma maneira mais precisa, tem-se que o desertor não representa de fato o conhecimento, ou não o faz integralmente, já que não é uma unidade monoreferencial – como é o termo –, mas sim, preferencial. Decorre daí que o desertor não tem a precisão do termo. E, se concordamos que a elaboração e a operação com linguagens documentárias não prescindem de um conhecimento da área em que se aplicam, devemos supor sempre a terminologia como subsídio para a atividade documentária.

Se aceita a substituição do tesauro pelo tesauro-terminológico, podem-se elaborar procedimentos metodológicos mais rigorosos para a indexação e a recuperação da informação, atendendo-se, assim, com maior eficiência às demandas dos usuários.

AGRADECIMENTOS

Este artigo é um dos produtos gerados no projeto *A especialidade da análise documentária:* parâmetros lógico, lingüístico e documentária para a elaboração de representações documentárias. Agradecemos a colaboração de Regina Keiko O. F. Amaro na redação final deste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

1. FOSKETT, A. C. *A abordagem temática da informação.* São Paulo: Polígono, Brasília: UnB. 1973.

2. GARDIN, J-C. Procédures d'analyse semantique dans les Sciences Humaines. In: POUILOUX, J., MARANDA, P. (ed.) *Échanges et communications:* mélanges offerts à Claude Levy-Strauss. La Haye, 1970, p. 628-657,1970.

3. ECO, U. *Conceito de texto.* São Paulo: T. A, Queiroz/EDUSP. 1984.

4. MACROTHESAURUS for information processing in the field of economic and social development. New York: United Nations, 1985.

5. DUBOIS, J. et alii. *Dicionário de lingüística.* São Paulo: Cultrix, 1978.

6. REY, A. *Le lexique, images et modèles:* du dictionnaire à la lexicologie. Paris: Armand Colin, 1977.

7. LEGUERN, M. Sur les rélations entre terminologie et lexique. *Meta, v.* 34, n. 3, p. 340-343, 1989.

8. BAGGE. C. Équivalence lexicale et traduction. *Meta,* Montreal, v. 35, n. 1, p. 61 -73,1990.

Artigo aceito para *publicação em 25 de novembro de 1992*

Maria de Fátima Gonçalves Moreira Tálamo

Doutora em Ciências da Comunicação pela Universidade de São Paulo (USP), Professora do Departamento de Biblioteconomia e Documentação da Escola de Comunicações e Artes da USP.

Nair Yumiko Kobashi

Mestre em Ciências da Comunicação pela Universidade de São Paulo. Professora do Departamento de Biblioteconomia da Escola de Comunicações e Artes da USP.

Marilda Lopez Ginez de Lara

Aluna do Curso de Pós-graduação em Ciências da Comunicação da ECA — USP. Professora do Departamento de Biblioteconomia da Escola de Comunicações e Artes da USP.

Contribution of the terminology for the development of thesaurus

Abstract

Thesaurus is traditionally considered as a tool for information description and terminology control. Prescriptions to link descriptors in the thesaurus are laid in logical and semantic principles, but are not suppported by precise definition of descriptor's properties. Such problems allow us to discuss thesaurus functions above refered to. Interface between terminology and content analysis is proposed as a profitable way to develop methods to perform efficient indexing languages.

Key words

Interface terminology and content analysis; Terminology and thesaurus; Content analysis and terminology.