

O *Thesaurus* Brasileiro da Educação*

Gaetano Lo Monaco

Palavras-chave: *Thesaurus* Brasileiro da Educação (Brased); linguagem documentária; terminologia.

Resumo

Apresenta a história da construção do *Thesaurus* Brasileiro da Educação e como esse instrumento de indexação e recuperação da informação pode ajudar na padronização da linguagem documentária dessa área do saber.

Introdução

As dificuldades geradas pela ausência de um *thesaurus* brasileiro de educação foram mapeadas pela pesquisa "O perfil dos sistemas de indexação dos documentos utilizados nas bibliotecas e centros de documentação voltados à educação na América Latina e países de língua portuguesa", coordenada por Maria da Graça Camargo Vieira (1999) e executada por um grupo de trabalho organizado pelo Comitê dos Produtores da Informação Educacional (Comped). A partir dos resultados dessa pesquisa constatou-se que: o capital de informações educacionais produzido na América Latina é grande e bastante significativo; falta comunicação, intercâmbio e interação entre os vários centros de produção e disseminação dessas informações; necessita-se, para tanto, de um sistema unificador que, usando a mesma linguagem e os mesmos canais de comunicação, coloque os dados educacionais à disposição de todos.

Também foi identificada a necessidade premente de um instrumento comum que permita o diálogo entre as diversas bases de dados e os centros de informação atuantes na área da educação no Brasil.

Assim, o *Thesaurus* Brasileiro da Educação (Brased) pretende responder, hoje, a uma demanda concreta por "um instrumento consistente e atualizado para o trabalho de indexação, capaz de evitar problemas para o usuário chegar ao documento específico" (Em busca..., 1999, p. 354). Destina-se, portanto, a ser utilizado principalmente pelos serviços de documentação institucionais, pelos centros de análise e de indexação e pelos produtores de bancos de dados documentais na área de educação, além de pesquisadores, professores e demais estudiosos da área.

A sociedade da informação exige um sistema dinâmico no processamento e na disseminação das informações educacionais e, nesse sentido, nos oferece as condições favoráveis, ou seja, o apoio tecnológico da informática e o estímulo do momento histórico em que estamos engajados.

A linguagem documentária *thesaurus*

Um sistema de informação e de documentação especializado supõe, como instrumento básico, um esquema temático que

* O A. agradece a colaboração de Doracy Rodrigues Farias e Maria Angela Torres Costa e Silva.

defina a sua área de especialização. A partir desse esquema surge a linguagem documentária que, estruturada, expressa o sistema conceitual dessa área. A linguagem documentária vem a ser o quadro referencial e o critério para se efetuar a análise da informação em suas várias fases.

O *thesaurus*, como qualquer linguagem documentária, tem sua origem na análise das informações de uma determinada área do conhecimento. O processo fundamental para aproveitar todas as informações significativas da massa documental de uma área e construir um *thesaurus* denomina-se "análise da informação", que é a análise dos assuntos contidos em um documento, com o objetivo de selecionar suas características.

A finalidade da linguagem documentária é propiciar rapidez no processamento das informações, precisão na indexação, especificidade de acordo com a política de indexação do sistema, confiabilidade na recuperação e agilidade na disseminação.

A experiência e a literatura parecem comprovar que a forma mais lógica e objetiva para se elaborar um *thesaurus* é a partir da definição que o termo assume dentro do sistema conceitual da área. "A definição é a chave para todo trabalho científico" (Felber, 1984, p.160).

De acordo com as diretrizes da Unesco/ISO para a elaboração e desenvolvimento de *thesauri* monolíngües, um *thesaurus* é definido,

[...] segundo sua função, como um instrumento de controle terminológico utilizado para traduzir a linguagem natural dos documentos, dos indexadores e dos usuários para uma linguagem sistêmica mais contida (linguagem documentária, linguagem de informação) [...] do ponto de vista de sua estrutura, o *thesaurus* é um vocabulário controlado e dinâmico de termos que possuem entre si relações semânticas e genéricas, baseadas num sistema de conceitos, e que abrange de maneira exaustiva um campo específico de conhecimento.

Também pode ser definido como uma linguagem documentária especializada e

[...] deve refletir com precisão as informações contidas no conjunto dos documentos de uma coleção à qual se aplica o *Thesaurus*. Deve conter os termos e as remissivas que sejam apropriados ao conteúdo temático, levando em consideração tanto a linguagem da coleção de documentos quanto a linguagem e as necessidades de informação dos usuários.

Ao contrário de um dicionário, que oferece definições de palavras ou termos de acordo com o seu significado, um *thesaurus* fornece palavras ou termos capazes de expressar significados mediante relações entre os termos e o conjunto das relações conceituais entre todos os termos define o sistema conceitual da área-objeto.

O elemento principal do *thesaurus* é o descritor, designação dada ao termo ou símbolo autorizado e formalizado que servirá para representar, sem ambigüidade, as informações contidas no documento. Cada descritor deverá representar apenas uma unidade conceitual do sistema a que pertence.

Para a elaboração de um *thesaurus* (especialmente nas áreas das ciências humanas e sociais), é necessário basear-se num estudo crítico do sistema conceitual da área e, a partir daí, analisar a informação em busca de conceitos e termos. Concretamente, o primeiro passo será o de elaborar uma matriz conceitual da área.

Função do *thesaurus* num sistema de informações

Falar em ontologia, características e estrutura do objeto (*óntos*) é falar da essência do *thesaurus* e de sua função na análise e no processamento das informações.

O *thesaurus* de uma área, bem definido em suas bases ontológicas e lógicas, norteia todo o processo de entrada e de saída das informações de um banco ou rede de bancos de dados.

Na entrada, norteia a análise, definindo a seleção e a síntese (indexação e resumo) das informações a serem armazenadas. Na saída, norteia a análise, definindo as necessidades informacionais dos usuários e as chaves (descritores) de busca nos bancos de dados, além de orientar a criação de produtos e serviços de acordo com os objetivos do centro de informação e com as expectativas dos seus usuários.

Thesaurus Brasileiro da Educação

Desde 1974, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) vem tentando entrar na era

do *thesaurus*. Naquele ano, foi traduzida para o português a primeira edição do *Thesaurus Multilíngüe para o Processamento da Informação em Educação* (Eudised). Essa versão brasileira serviu até a metade da década de 1980 como instrumento básico para o processamento das informações educacionais no Inep e, em geral, no Brasil.

Porém, tanto o Eudised como o *Thesaurus de la Educación* da Unesco, utilizados por várias instituições brasileiras, não atenderam às exigências de caráter local e nacional, dados os objetivos para os quais foram elaborados. Perante tais exigências, o próprio Inep, após a mencionada versão do Eudised, elaborou a idéia de um *thesaurus* brasileiro da educação. Vale lembrar que, no começo da década de 70, tinha sido iniciada a elaboração de um fichário conceitual que chegou a cerca de 600 termos educacionais, os quais foram reduzidos a 116 descritores essencialmente brasileiros. Esse trabalho foi realizado sob a coordenação da servidora Regina Helena Tavares e do consultor Durmeval Bartolomeu Trigueiro Mendes. Por vários motivos de caráter institucional, o trabalho referente à linguagem foi interrompido, mas a idéia continuou viva, à espera de realização.

Com a instalação e a ampliação da biblioteca do Inep na sede do Ministério da Educação em Brasília, em 1980, e com a criação do Sistema Nacional de Informações Bibliográficas em Educação (Sibe) por meio da Portaria MEC nº 612, de 11 de novembro de 1981, a idéia da construção do *thesaurus* foi reativada. De fato, uma vez criado o Sibe, era urgente definir seu âmbito temático e estabelecer uma linguagem documentária para o processamento das informações a serem administradas pelo sistema, que era formado por cinco bibliotecas especializadas, como unidades regionais, sendo três de Faculdades de Educação (UFMG, UFRJ, UFRGS) e duas de instituições de pesquisa (Fundação Carlos Chagas e Fundação Joaquim Nabuco), tendo o Cibec como unidade central.

Tentou-se, inicialmente, continuar a elaboração do *thesaurus* "em serviço", isto é, durante o processamento das informações pelos próprios analistas-indexadores. Mas, devido à urgência dessa linguagem para a organização e a ativação do Sibe, em 1983, por proposta do Inep, foi criada uma comissão interinstitucional, por meio da Portaria MEC nº 431/83, para que, no prazo de um ano, fosse elaborado o *Thesaurus* Brasileiro

da Educação (Brased), e estava formada por: Gaetano Lo Monaco (coordenador) e Lidia Alvarenga Nery, do Inep; Diogo José Ayrimoraes, do Conselho Federal de Educação (CFE); Lauro de Barros Silva Filho, da Secretaria de Ensino de Primeiro e Segundo Graus (SEPS); e Elza de Oliveira, da Funarte. A comissão elaborou um esquema conceitual do âmbito educacional, a partir do qual seria criada uma metodologia para a construção do *Thesaurus*.

A complexidade da tarefa, as inúmeras discussões e debates e a pouca disponibilidade de tempo por parte dos membros da comissão atrasaram os trabalhos. Cumprido o período estipulado para o funcionamento da Comissão, um primeiro *corpus* documental para a coleta dos termos já estava definido e uma primeira matriz conceitual, composta por termos coletados pelos analistas, começou a ser testada. Esses termos, organizados dentro da estrutura proposta, constituíram a primeira listagem do *Thesaurus* Brased.

Os trabalhos continuaram no âmbito do Inep, sob a coordenação deste autor, à espera de avaliação por consultores especialistas.

A primeira fase da avaliação, de caráter semântico, feita pelo lingüista Sérgio Ricardo Pereira dos Santos, constatou que, na redação apresentada, o aspecto epistemológico predominava sobre o semântico, fugindo dos padrões de uma metalinguagem que deveria estar baseada nos princípios lingüísticos.

A segunda fase da avaliação, de caráter técnico, feita por Hagar Espanha Gomes e Marisa Bräscher Basílio Medeiros, bibliotecárias e mestres em Ciência da Informação, constatou a necessidade de, primeiramente, conceituar cada termo para melhor estabelecer as relações entre eles.

A terceira fase da avaliação, de caráter epistemológico, feita pelo sociólogo Francisco Salatiel de Alencar Barbosa, convalidou, de modo geral, a matriz conceitual do *Thesaurus*, mas sugeriu algumas modificações na estrutura temática e a introdução de novos termos para completar algumas cadeias conceituais.

Acatadas as sugestões dos avaliadores, fez-se necessária uma profunda revisão da matriz conceitual e, conseqüentemente, uma reestruturação das relações entre os termos. Não foi possível, porém, eliminar as deficiências originadas pela falta de definição conceitual de cada termo, ponto fundamental para a elaboração de um *thesaurus*.

Em 1989, para atender às necessidades do Sibe, foi apresentada a primeira versão do Brased.

Em 1997, inicia-se a elaboração da segunda versão do Brased, buscando o aperfeiçoamento da linguagem documentária com dois objetivos principais: sua informatização, para facilitar a sistematização e a recuperação das informações, e sua disponibilização para o público, que obedeceu às seguintes fases:

- análise de um corpo de documentos primários e secundários, para seleção e coleta de novos termos;
- introdução de novos termos e preparação da proposta da nova versão;
- análise dessa proposta pela equipe de analistas, que teve como resultado uma nova estrutura e novos descritores relacionados;
- informatização da nova versão.

Nessa nova versão, para facilitar a compreensão do *Thesaurus*, buscou-se maior clareza e transparência na lógica da estrutura conceitual.

Versão atual

O *Thesaurus* Brased, como qualquer outra linguagem natural ou documentária, para atender a todas as expectativas de seus usuários, considera a educação não somente na sua natureza e na sua praxe, mas também no seu contexto, sem o qual não é possível entendê-la.

Por essa razão, ao se definir o âmbito temático ou o domínio da área, foi preciso ter em consideração também todas as outras áreas que diretamente são relacionadas com a Educação e que formam o seu contexto, constituindo um conjunto único, interdisciplinar.

Poderá parecer que o Brased é pluridisciplinar, mas, na realidade, os vários aspectos que ele aborda são essenciais para uma compreensão total da educação, que é um fenômeno humano pluri e interdisciplinar. Um *thesaurus* assim concebido poderá atender a todas as exigências teóricas e concretas do pensar e do fazer educação dentro de uma sociedade em desenvolvimento.

O Brased é um *thesaurus* nacional, porque é uma linguagem documentária elaborada especificamente para o processamento das

informações educacionais brasileiras, isto é, para indicar de forma específica a realidade nacional, sem desconhecer a realidade internacional.

E por que é necessário um *thesaurus* nacional de educação? Em primeiro lugar, porque na área de ciências sociais ou humanas – e a educação faz parte dessa área –, as peculiaridades locais ou nacionais são muito características. Em segundo lugar, porque os *thesauri* internacionais não descem ao nível de especificidade necessário para o estudo da realidade educacional do Brasil.

O próprio *Thesaurus da Educação da Unesco* recomenda a sua adaptação às exigências nacionais, ou a elaboração de um *thesaurus* nacional. Neste caso, sugere que os responsáveis pela elaboração se coloquem em intercâmbio com a Unesco.

Estrutura conceitual do Brased

Embora o *thesaurus* seja uma metalinguagem, é inevitável o recurso aos princípios da epistemologia para definir o sistema conceitual de sua área. Na área das ciências exatas, a coincidência do epistemológico com o semântico é usual, porque nelas prevalece a objetividade concreta dos termos. Por isso é simples a sua sistematização.

Já na área das ciências humanas e sociais, predominam o sentido analógico, os modismos, as influências das ideologias e das conjunturas culturais e históricas, que chegam até a alterar o sentido dos termos. Isso faz com que a definição do sistema conceitual da área e dos termos-chave que o representam seja infinitamente mais difícil, se não houver uma combinação harmônica entre os aspectos semânticos e epistemológicos. Uma tentativa para superar essa dicotomia é feita pelos *thesauri* que partem da definição do sistema conceitual da área para identificar, dentro dele, as cadeias conceituais, os conceitos específicos e os respectivos termos.

Com base nessa lógica, a matriz conceitual que fundamenta a estrutura e a elaboração do *Thesaurus* Brased tem no centro o homem como indivíduo e como ser social, que evolui no tempo e no espaço, interagindo constantemente com o seu meio: é a Educação propriamente dita com seus princípios, conteúdo e processo (campo 400).

Partiu-se, portanto, de uma cosmovisão dialética que considera a educação como um processo pelo qual o ser humano (indivíduo e coletividade), interagindo com a realidade total, aprende a desenvolver suas potencialidades, cria cultura, atende às suas necessidades, torna-se agente da sua história. Dessa visão surgiu uma matriz conceitual que coloca o homem (ser pessoa, histórico e social) no centro do sistema Educação, que é pluri e multidisciplinar, isto é, muitas ciências fundamentam o processo e a ação educativa (campo 300 – Fundamentos da educação).

A escola é a educação institucionalizada e, na sociedade politicamente organizada, de fato, é onde se encontram todas as condições para que a educação do homem aconteça (campo 200 – Escola como instituição social).

A educação do homem se realiza dentro da realidade global e em interação com esta; fora desta não há educação (campo 100 – Contexto da educação).

Necessariamente, essa visão leva a uma abordagem mais epistemológica do que semântica da área da educação, e essa abordagem norteia a análise da informação, a elaboração da terminologia e a própria estrutura do *Thesaurus*.

Portanto, o *Thesaurus* Brased divide-se em quatro grandes campos substantivos e um campo complementar, que são subdivididos em grupos, facetas e subfacetas. Os campos substantivos são: Contexto da Educação; Escola como Instituição Social; Fundamentos da Educação; e Educação: princípios, conteúdo e processo.

No campo complementar – Identificadores e Especificadores de Informação –, estão organizados todos os termos que não fazem parte do conteúdo da educação, mas que são necessários para especificá-lo, complementá-lo ou identificá-lo.

A conjugação dos campos substantivos com o campo complementar permitirá a indexação e a recuperação mais precisas e específicas, objetivo fundamental de qualquer *thesaurus*.

Considerações finais

A sociedade atual tem buscado desenvolver instrumentos para otimizar o uso estratégico das informações em seu crescimento socioeconômico e político. Para tanto, é fundamental que tais instrumentos permitam a organização e a recuperação das informações de forma ágil e confiável.

O *Thesaurus* Brased foi criado com o objetivo de permitir a indexação e a recuperação de informações de modo sistematizado, possibilitando a otimização de seu uso por pesquisadores, gestores e estudiosos. Permite também desenvolver as relações conceituais entre os termos utilizados, além de possibilitar a busca por novas formas de análise do contexto educacional. Por ser um *thesaurus* nacional, o Brased está voltado para as peculiaridades da educação brasileira. Entretanto, não abandona as relações entre a realidade nacional e o cenário internacional, uma vez que tem em sua lógica estrutural a idéia de que a educação é algo que tem de ser permanentemente contextualizado.

O mais importante, entretanto, é que apesar de atender a uma demanda atual, o *Thesaurus* Brased não é um instrumento desenvolvido às pressas ou no sentido de oferecer uma resposta institucional a essa demanda específica. Ele é o resultado de um longo trabalho de estudo e pesquisa, embasado em anos de análise documental, discussão e avaliação de especialistas. Trata-se de uma ferramenta consistente, uma referência conceitual na área educacional que abre caminho para uma nova forma de trabalhar a gestão da informação educacional no Brasil.

O *Thesaurus*, assim como a educação, está em constante evolução e se modifica de acordo com o trabalho, a pesquisa e a dedicação de todos aqueles que estão envolvidos com a educação.

O Brased está disponível no endereço eletrônico: www.inep.gov.br/pesquisa/thesaurus

Referências bibliográficas

ADAMS, Katherine. The semantic web: differentiating between taxonomies and ontologies. *Online*, v. 26, n. 4, p. 20-24, July/Aug. 2002.

AUSTIN, D. *Diretrizes para o estabelecimento de tesouros monolíngües*. Brasília: IBICT, Senai, 1993.

EM BUSCA do Tesouro Brasileiro de Educação. *Revista Brasileira de Estudos Pedagógicos*, Brasília, v. 80, n. 195, p. 352-354, maio/ago. 1999.

FELBER, H. *Terminology manual*. Paris: Unesco/Infoterm, 1984.

GOMES, Hagar Espanha (Coord.). *Manual de elaboração de tesouros monolíngües*. Brasília: Programa Nacional de Bibliotecas das Instituições de Ensino Superior, 1990.

LARA, Marilda Lopes Ginez de. *Representação e linguagens documentárias: bases teórico-metodológicas*. São Paulo, 1999. Tese (Doutorado em Ciências da Comunicação) – Universidade de São Paulo (USP).

LIMA, Vânia Mara Alves. *Terminologia, comunicação e representação documentária*. São Paulo, 1998. Dissertação (Mestrado em Comunicação) – Universidade de São Paulo (USP).

LO MONACO, Gaetano. *Consolidação das propostas: "Brasil: sugestões para a execução" e "Argentina: cronograma tentativo"; ficha terminológica bilíngüe: referência de estudo*. Relatório elaborado no âmbito do SIC Mercosul Educacional, GGP Terminologia, em 11 de novembro de 2004. Não publicado.

UNISIST. Princípios de indexação. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 10, n. 1, p. 83-94, mar. 1981.

VIEIRA, Maria da Graça Camargo (Coord.). *Perfil dos sistemas de indexação e documentos utilizados nas bibliotecas e centros de documentação voltados à educação na América Latina e países de língua portuguesa: relatório final Projeto BRA/97/019*. São Paulo: FCC, 1999. Disponível em: http://www.inep.gov.br/download/comped/mapeamento/Relat%C3%B3rio_Final_linguagem_documental.doc.

ZAVITOSKI, Maria Teresa. *Exploração do uso do tesouro como instrumento de recuperação da informação*. São Paulo, 2001. Dissertação (Mestrado em Ciências da Comunicação) – Universidade de São Paulo (USP).

Gaetano Lo Monaco, licenciado em Filosofia pela Faculdade Dom Bosco de Filosofia Ciências e Letras de São João Del-Rei (MG), bacharel em Pedagogia pelo Centro de Ensino Unificado de Brasília (Ceub), Curso de Terminologia no Departamento de Linguística da Universidade de Brasília (UnB), técnico em Assuntos Educacionais, ex-diretor do Centro de Informação e Biblioteca em Educação (Cibec).

Abstract

The Brazilian Thesaurus of Education

This paper presents the compilation of the Brazilian Thesaurus of Education and how this indexation and retrieval information instrument can help the standardization of the documentary language in this specific knowledge area.

Keywords: Brazilian Thesaurus of Education; documentary language; terminology.
