

Tesouro e normalização terminológica: o termo como base para intercâmbio de informações

Thesaurus and terminological standardization: term as basis for information exchange
por [Hagar Espanha Gomes](#) e [Maria Luiza de Almeida Campos](#)

Resumo: Uma análise da Norma Internacional de elaboração de Tesouros Monolíngues (ISO 2788) é aqui apresentada, focalizando a investigação nos aspectos relativos ao termo de indexação como base para intercâmbio de informações. A partir de princípios da Teoria da Terminologia de Wüster e da Teoria do Conceito de Dahlberg, discutem-se as soluções apresentadas na ISO 2788 para a forma do termo. Devido às facilidades de integração de informações e graças às tecnologias de informação vigentes é fundamental que o profissional de informação se pautem em princípios teóricos consistentes para o tratamento terminológico, visando garantir qualidade e consistência nos processos de intercâmbio de informação. Atualmente, o crescimento da utilização de ambientes distribuídas, especialmente na Web, de interfaces abertas de acesso a bancos de dados, de tecnologias de mediadores e de padrões de formato para troca de dados contribuiu para que informações originalmente isoladas sejam disponibilizadas para serem utilizadas de maneira integrada. Neste sentido, o Tesouro assume papel fundamental nesta integração, viabilizando a interoperabilidade semântica de sistemas distribuídos heterogêneos.

Palavras-chave: Tesouro; Normalização Terminológica; Norma Internacional de Elaboração de Tesouros (ISO 2788); Termo de indexação.

Abstract: An analysis of the International Standard for thesaurus construction (ISO 2788) is presented focusing on aspects relating to indexing term as basis for information exchange. Wüster's Theory of Terminology and Dahlberg's Theory of Concept provide the foundations to discuss solutions present in ISO Standard about the form of the term. Due to easiness of information integration and to current information technology, the information professional must work based on sound theoretical principles for the terminological treatment of the term, aiming at quality and consistency in the process of information exchange. Nowadays, the growing activity in distributive environment, specially in the Web, open interfaces of access to data banks, of intermediate technologies and of format standards for data exchange has contributed to originally isolated information be available for use in an integrated way. In this aspect, the thesaurus assumes fundamental role in this integration, providing semantic feasibility among heterogeneous distributed systems.

Keywords: Thesaurus; Terminological standardization; International Standard for Thesaurus Construction (ISO 2788); Indexing terms.

1. Introdução

O uso consistente de princípios na indexação é um fator de qualidade e pode contribuir para a eficácia dos serviços, desde que a análise do assunto seja adequada. Neste contexto, o vocabulário utilizado é um requisito essencial e deve ser devidamente controlado e normalizado. Esta normalização se efetiva não apenas no interior de um serviço de recuperação de informação, mas entre diferentes serviços.

As facilidades de intercâmbio, graças às tecnologias de informação vigentes, passam a requerer maior atenção dos profissionais de informação ao planejarem seus serviços, levando em conta os padrões ou em desenvolvimento ou já instituídos para as diferentes atividades que ocorrem no interior de um serviço de recuperação de informação.

Atualmente, bases de dados de todos os tipos têm proliferado com a disponibilização de informações em rede e principalmente na Web. A recuperação dos conteúdos informativos ainda não é realizada de forma satisfatória, devido à falta de ferramentas de acesso adequadas, que viabilizem, por exemplo, o controle terminológico. O que vem ocorrendo de modo mais freqüente é a utilização de métodos automáticos de indexação e recuperação de informações, que, por si só, estão muito distantes de serem uma solução apropriada, pois se baseiam na freqüência de ocorrência dos termos, não atentando para questões ligadas à sua natureza semântica.

Desta forma, a questão que se coloca relativa ao tratamento e à recuperação de informações diz respeito à qualidade e à precisão no tratamento das informações e à adequação a uma dada solicitação de busca através de um controle terminológico que venha garantir precisão nas informações recuperadas em meio eletrônico. É sempre bom lembrar que, empregando terminologia normalizada em seu discurso, o autor está, também ele, contribuindo para a eficácia dos sistemas de recuperação, além de melhorar a precisão de sua comunicação.

Na área de Terminologia, a ISO desenvolveu uma Norma de Harmonização de conceitos e termos (ISO-860), que pode ser útil em serviços de uma mesma língua ou entre serviços de línguas diferentes. As soluções de problemas terminológicos oferecidas por esta norma e o padrão de metadados compõem um quadro de tecnologia essencial, mas não completo, para a interoperabilidade, seja em termos de software seja em termos de significados. Estas questões são percebidas mais fortemente quanto se trata de organização de informações em rede, o que requer mecanismos de intercâmbio automático de dados.

No âmbito das atividades intelectuais de tratamento da informação, o quadro de tecnologia se completa com a elaboração de tesouros, aproveitando-se de princípios para a produção de terminologias, sobre os quais podem ser aplicados os princípios de harmonização de conceitos e termos entre tesouros de diferentes serviços de recuperação de informação inter- e intra-línguas, e esta produção é campo de atuação dos bibliotecários e documentalistas.

A normalização de vocabulários, nomenclaturas, terminologias ou assemelhados tem sua relevância não apenas em serviços de informação; é de grande valor para as trocas internacionais, nas quais as atividades de exportação/importação focalizam produtos ou tecnologias, nem sempre designados pelos mesmos termos numa mesma língua falada por diferentes países ou por termos cognatos.

Se os princípios de construção de terminologias e tesouros forem incorporados a normas - internacionais e/ou nacionais - outro ganho é a possibilidade de melhores resultados nos serviços que utilizam meta-buscadores, pois diferentes

tesauros de áreas afins, baseados nos mesmos princípios configurar-se-ão como instrumentos compatíveis e harmônicos.

Além disso, sabe-se que os buscadores atuam nos textos, ou em pedaços de textos, utilizando a linguagem natural como base para a recuperação e aqui, também, sobreleva o uso, pelos autores, de terminologia padrão, essencial para uma comunicação sem ambigüidades e, como consequência, facilitadora de melhor recuperação pelos buscadores, como ocorre nas ciências básicas.

O tesouro tem sido considerado como uma linguagem de indexação. Isto é verdade, quando ele está em uso num serviço de recuperação de informação, e destinado ao uso para o qual é desenvolvido, com regras para sua utilização. Mas, se pensarmos no tesouro como uma tecnologia independente, ainda não integrada a um serviço - por exemplo, quando alguém deseja avaliar um tesouro desenvolvido por outros ou quando ele ainda vai ser desenvolvido -, um planejamento, uma fundamentação teórica, que não aceite exceções, como toda linguagem artificial, tornam-se imprescindíveis. Estes aspectos propiciam condições seguras na etapa de manutenção, quando da inclusão de novos termos na estrutura vigente, evitando criações anômalas, como é o caso de decisões ad-hoc, ou seja, decisões voltadas para a solução de um caso específico, não generalizável. O que se deve almejar são regras genéricas que contemplem os casos particulares. Isto é: regras prescritivas.

Estudos voltados para questões relativas as bases teóricas de linguagem documentária ([CAMPOS 1986](#); [CAMPOS, 1995, 2001, 2004](#); [CAMPOS & GOMES, 2003](#); [GOMES, 2003](#); [GOMES&CAMPOS&MOTTA, 2004](#); [MOTTA, 2000](#)), desenvolvidos no Brasil, tem sido para um grupo de pesquisadores uma das condições para a realização de inúmeras atividades relacionadas ao controle terminológico como base para a integração de informações.

Recentemente, o tesouro tem sido objeto de interesse da Informática, nomeado como uma ontologia, sendo considerado como um mecanismo para se alcançar a interoperabilidade. Neste caso, cada metadado, então, é visto como um termo, um conceito, e não um grupo de palavras que podem ser 'fatoradas'. Neste ambiente, os princípios para organização de tesouros com base em conceitos se tornam de fundamental importância, ainda mais que a estruturação dos termos possibilita a interoperabilidade semântica. ([Hunter](#))

Apesar da importância do tesouro, o Brasil não dispõe de norma brasileira própria. Em nível internacional existe a "Norma Internacional de Elaboração de Tesouros Monolíngues (ISO 2788)", que apesar dos esforços empreendidos no sentido de apresentar princípios para uma normalização terminológica consistente, verifica-se, ainda, muitos problemas relativos à unidade de representação, ou seja, ao termo de indexação.

O objetivo deste estudo é apresentar uma análise à luz de teorias como a Teoria do Conceito (Dahlberg) e a Teoria da Terminologia ([WUESTER, 1981](#) e [FELBER, 1984](#)) dos princípios relativos ao termo da Norma ISO 2788, visando municiar o profissional de informação de uma visão crítica para lidar com

questões relativas a intercâmbio de informações no momento atual.

2. Princípios Fundamentais

A análise da norma ISO 2788 que aqui se faz, toma por base alguns princípios que a seguir são apresentados de forma bastante resumida, relacionados à definição de 'termo'.

O Tesouro, como instrumento de controle da indexação, surgiu em serviços de recuperação em ambiente computacional. Naquele momento - no final dos anos 50 e início dos 60 - ainda não haviam sido desenvolvidos procedimentos computacionais para reconhecer termos constituídos por grupos de palavras. Então, a ênfase recaiu em unidades verbais e princípios foram desenvolvidos para ajustar-se a aquela limitação tecnológica. É fácil compreender a preocupação com conceitos como 'fatoração sintática' e 'fatoração morfológica' presentes na literatura da época e abordadas por alguns autores da década de 60. No entanto, se esta é a motivação para o uso de tais técnicas, elas hoje não se justificam. Se pensarmos apenas na recuperação de palavras tais como se encontram nos textos, tais soluções podem ser satisfatórias. Mas quando tentamos aplicar estes princípios à construção de tesouro, ou seja, ao instrumento de controle do vocabulário de indexação, vemos que eles não se ajustam. Por exemplo, um usuário pode preparar uma busca com as palavras 'poluição' e 'atmosférica'. E certamente receberá algo relativo a 'poluição atmosférica'. Mas, e no tesouro, como fica estruturado o assunto? Um tesouro não comporta adjetivos; por outro lado, qual a estrutura da expressão 'poluição atmosférica'? É um termo específico de 'poluição'? O ar é um elemento poluidor? Ou o ar 'carrega' elementos poluidores? Ou a poluição (química, física ou biológica) ocorre no meio 'ar'?

A linguagem natural é defendida e aceita por muitos, mas ela tem armadilhas como a do exemplo acima. Trabalhar com a palavra (extraída da linguagem natural, tout court) parece não ser a solução. É aqui que surgem as contribuições da Terminologia ([WUESTER, 1981](#); [FELBER, 1984](#)) e da Teoria do Conceito ([DAHLBERG, 1978](#)): a unidade do tesouro não deve ser a palavra, mas o conceito, materializado no termo. O conceito é visto, dessa forma, como uma unidade indivisível formada pelo referente, característica e forma verbal.

Se a palavra tem seu significado evidenciado no contexto frasal, o termo, por seu turno, carrega consigo o significado/conteúdo. O termo assim é constituído por uma forma verbal e um conteúdo conceitual. A homonímia fica, em geral, eliminada, nos tesouros, porque estes cobrem domínios específicos do conhecimento e é no contexto do domínio que o termo evoca, para a comunidade do domínio, o conteúdo que tem que evocar. Um postulado da Teoria da Terminologia é que os termos guardam uma monorreferencialidade relativa.

Qual a contribuição da Terminologia para o Tesouro documentário? Ela se ocupa do conceito e das relações entre eles, exatamente como deve ocorrer num tesouro documentário. Outro postulado da Terminologia é que os conceitos se relacionam como um sistema, ou seja, eles se definem uns em relação aos

outros. Fica difícil apreender um conceito isoladamente, pois ele precisa de outros conceitos para poder ser apreendido. Por exemplo, como dominar o conceito de 'verniz', sem ter o conceito de 'resina'? Para dominar ambos os conceitos precisamos dominar as características de cada um dos objetos a que eles se referem, as quais são, por sua vez, também conceitos.

Ou, então, como estruturar adequadamente no Tesouro os termos 'conhaque' e 'conhaque de gengibre'? Se seguirmos a forma verbal, seremos tentados a afirmar que 'conhaque' é o termo geral e 'conhaque de gengibre' é um termo específico. Como nos apropriarmos corretamente do conteúdo de tais termos? Se analisarmos as características dos objetos a que os termos se referem (análise do objeto), vamos saber que o objeto 'conhaque' é um 'tipo de bebida alcoólica fermentada-destilada a partir do vinho', e que o objeto 'conhaque de gengibre' é um 'destilado de cana-de-açúcar adicionado de gengibre, como substância aromática'. A partir daí podemos estruturar os termos do tesouro adequadamente, ou seja, não cairemos na armadilha de considerar o termo 'conhaque' como o termo da classe geral, da qual 'conhaque de gengibre' seria o termo específico, mas estabeleceremos entre eles uma relação associativa. O que fizemos foi uma análise terminológica dos termos, a qual permite o estabelecimento de relações mais adequadas entre eles.

A análise do referente permite estabelecer relações lógicas (de superordenação/subordinação) e relações ontológicas (partitivas e associativas).

As primeiras ocorrem quando da análise das propriedades do objeto/referente e as segundas ocorrem quando observamos sua relação com outros, como, por exemplo, sua função na realidade empírica ou sua contigüidade espacial ou temporal.

Aqui incorporamos, como apresentado anteriormente, elementos da Teoria do Conceito de Dahlberg, que considera o conceito como constituído de três elementos: o objeto/o referente, as características desse objeto/referente e a forma verbal que designa esse objeto/referente. O conceito não prescinde destes três elementos: faltando um deles não se tem o conceito, tem-se uma 'idéia' mais ou menos vaga.

A versão anterior da Norma ISO 704 (Princípios de Terminologia), definia 'conceito' como 'unidade de pensamento'. Concordamos com a Dra. Dahlberg que rejeita esta definição porque o pensamento sem o uso da linguagem não pode ser comunicado, está no quadro cognitivo de um indivíduo; então, não se pode garantir consistência na comunicação. Durante a reunião da ISO-TC 37, em que se aprovaria a nova versão da Norma, foi possível aprovar a definição de conceito como 'unidade de conhecimento'. O termo é a síntese, pois 'carrega' em si as características do objeto/referente. De acordo com a norma ISO 1087 - Vocabulário da Terminologia, 'termo' é definido como 'designação de um conceito definido numa língua especializada, por uma expressão lingüística.'

o termo é essa materialização. Esse par Conceito-Denominação constitui o Termo. Se adotarmos a palavra denominação, em vez de termo, teremos mais facilidade em identificar a expressão lingüística que designa um conceito, ou

seja, para nomearmos um objeto utilizaremos tantas palavras quantas forem necessárias para tal.

Ora, se o 'conceito' é uma unidade, então é indivisível. Se o termo denota o conceito, então não existe termo composto, pois isso equivaleria dizer que um termo composto tem um conceito composto, o que não faz sentido. Então, afirmamos, no que respeita à forma, que o termo é constituído por uma palavra ou por um grupo de palavras. E não se pode, portanto, considerar o conceito de 'termo composto'.

Ainda que muito resumidamente, temos aqui os principais fundamentos que utilizamos na determinação em um tesauro documentário do que consideramos unidade de representação - o termo de indexação. E é a partir destes fundamentos são analisados alguns aspectos da norma ISO 2788.

Entretanto é necessário evidenciar o caminho que adotamos para esta análise. A análise da Norma ISSO 2788 se apóia na investigação do que considera termo de indexação, através da interpretação das definições apresentadas na Norma.

A forma como abordamos cada definição/explicação apresentada, se consolida, também, por uma postura teórica e metodológica. Desta forma, cabe ressaltar uma das propostas da Teoria Geral da Terminologia, no que se refere a elaboração de definições:

"A característica que constitui um conceito é também um conceito. Através dela pode-se comparar conceitos, classificá-los em um sistema de conceitos, sintetizá-los através da definição, e denominá-los através dos termos. O agregado das características que constituem o conceito determina sua intensão. A partir da determinação da intensão do conceito, ou seja da delimitação de suas características, é possível determinar também conceitos que possam ser relacionados ao conceito em análise por possuírem características semelhantes; em outras palavras, é possível determinar a totalidade ou o número de conceitos que este conceito abarca, o que constitui sua extensão". (FELBER, 1984, p. 58)

Assim, nossa análise trata de recuperar em cada definição/explicação apresentada os elementos que compõe os conceitos analisados, que são considerados também conceitos materializados através de termos.

3. A Norma Internacional de Elaboração de Tesouros Monolíngues (ISO 2788)

O objetivo do tesauro, segundo a Norma, é 'assegurar uma prática consistente num serviço de indexação ou entre diferentes serviços, como, por exemplo, numa rede'.

Antes de entrarmos especificamente na questão relacionada ao termo de indexação, já detectamos na definição de "linguagem de indexação" e "tesauro" problemas que implicam em inconsistências.

Vejam os cada definição:

linguagem de indexação

conjunto controlado de termos selecionados da linguagem natural usados para representar, de modo resumido, o assunto dos documentos.

Comentário:

A definição contém o que é e a função. O que se entende por 'controlado', em que consiste o controle? Controle de quê?

Na língua comum 'termo' é o mesmo que 'palavra'. E palavra precisa de um contexto para se depreender seu sentido. Ora, num tesauro este contexto não existe e o significado pode se tornar ambíguo. Por exemplo, seja a palavra 'vinagre': qual o seu significado? Se, num documento ele é usado referindo-se a 'vinagre de arroz' e noutra, a 'vinagre de vinho' (aparentemente um pleonasma), o significado de 'vinagre' é o mesmo em ambos os casos? A palavra 'vinagre' está presente nos dois termos, mas, no primeiro caso, 'arroz' não é determinante de 'vinagre', o que produziria um termo específico de 'vinagre'; em suma, uma análise lingüística do termo não resolve a questão. É preciso uma análise terminológica

Então, o que é o 'termo'? A Norma não define.

Quanto à função, a linguagem documentária serve para 'representar, de modo resumido, o assunto de um documento'. 'Assunto' é também um conceito que não está definido, embora isso não tenha maiores conseqüências nesta análise.

tesauro

vocabulário de uma linguagem de indexação controlada organizada formalmente explicitando 'a priori' relacionamentos entre os conceitos (por exemplo, 'mais geral' e 'mais específico').

Comentário:

Esta definição é circular na primeira parte, ou seja, naquela parte que declara o que o objeto é. Em 'linguagem de indexação' já está explícita a característica 'conjunto de termos' ou seja, um 'vocabulário' (conjunto de termos ou de vocábulos); e também a característica 'controlado'. Isto fica mais claro se aplicarmos o princípio da substituição. Usando este princípio, teríamos como definição de tesauro: vocabulário de um 'conjunto controlado de termos...'

A segunda parte da definição diz o que o tesauro possui: relações explícitas entre os conceitos. Mas não define o que seja 'conceito'.

Como podemos observar estas definições incluem expressões que não estão

definidas, como 'termo' e 'conceito'. A falta de uma metodologia consistente para a construção de definições vai implicar em inconsistências no texto.

3. 1 Termo de indexação

A Norma define 'termo de Indexação' como: *representação de um conceito preferencialmente sob a forma de uma palavra ou expressão*

Como observamos anteriormente, em nenhum momento ela define o que é conceito. Entretanto, admite que ele (o termo) pode ser representado sob forma de uma palavra ou expressão, acrescentando uma nota que diz: "um termo de indexação pode consistir em mais de uma palavra e então é conhecido como termo composto"

Ao introduzir o conceito de 'termo composto' ela contradiz diversos avanços já apresentados pela Teoria do Conceito e a Teoria da Terminologia. Desta forma, a análise que apresentaremos a seguir tem por objetivo evidenciar as contradições apresentadas na Norma, quanto à determinação da unidade de representação dos tesouros documentários, analisadas à luz das Teorias citadas. Especificamente, vamos tomar por base a expressão 'termo composto' e as soluções apresentadas para sua fatoração.

termo composto

termo de indexação que pode ser fatorado morfológicamente em componentes separados, cada um dos quais pode ser expresso ou re-expresso como um substantivo capaz de servir independentemente como termo de indexação.

Comentário:

Utilizando o princípio da substituição, substituiremos termo de indexação por sua definição e teremos, então, que 'termo composto' é a *'representação de um conceito preferencialmente sob a forma de uma palavra ou expressão', que pode ser fatorado morfológicamente, etc. etc.*

Mais uma vez encontramos o termo 'conceito', não definido. Ao propor fatoração morfológica, a Norma está claramente privilegiando o aspecto verbal do termo de indexação, pois se o termo representa o conceito, quando fatoramos o termo isolamos as palavras que o constituem. E isto pode significar a destruição do conceito, ou seja, o conceito representado por aquele grupo de palavras, deixaria de estar representado num tesouro, em nome da fatoração morfológica.

Há uma nota extensa:

"As partes da grande maioria de termos compostos podem ser distinguidas como segue:

a) O foco, isto é, o substantivo componente que identifica a classe geral de conceito à qual o termo como um todo se refere.

Exemplo 1: o substantivo componente 'indexes' no termo composto 'printed indexes'.

Exemplo 2: o substantivo 'hospitals' na expressão 'hospitals for children'.

b) A diferença ou o modificador, isto é, um ou mais componentes que servem para diminuir a extensão do foco e assim especificar uma de suas subclasses

Exemplo 1: o adjetivo 'printed' no termo composto 'printed indexes'.

Exemplo 2: a preposição + substantivo 'for children' no termo composto 'hospitals for children'."

Diz, ainda, a Norma:

"O foco e suas diferenças podem se apresentar com palavras separadas, como em "dining rooms" ou "soup spoons" ou podem vir concatenadas numa palavra, como em "bedrooms" ou "teaspoons..."

Comentário:

A nota *a)* assume que não se fatora uma expressão quando ela designa o termo específico de uma '*classe geral de conceito à qual o termo como um todo se refere*'. Neste caso, privilegia-se o sentido, ou o conceito [como preferimos].

A nota *b)* focaliza a diferença ou o modificador, mas, quando apresenta o mesmo exemplo, fica numa explicação circular: em *a)* o foco caracteriza o todo, em *b)* a diferença específica.

A análise 'morfológica' continua no último parágrafo.

No capítulo 6 da Norma, os termos de indexação são classificados do ponto de vista conceitual, ou seja, os termos designam:

1) Entidades concretas

* coisas e suas partes físicas

* materiais

2) Entidades abstratas

* ações ou eventos

* entidades abstratas e propriedades de coisas, materiais ou ações

* disciplinas ou ciências

* unidades de medida

3) Entidades individuais, ou 'classes-de-um', expressas por nomes próprios

No entanto, a falta de uma definição ou explicação do que seja 'conceito' leva a Norma a desconsiderar tal classificação e a fazer uma abordagem lingüística do termo, ao incluir um capítulo sobre a forma verbal, fornecendo orientação sobre o uso de substantivos, adjetivos, advérbios e verbos e, ainda, sobre o uso de singular e plural.

Todo o capítulo 7 trata de fatoração.

Em 7.1 encontramos:

"A fatoração semântica aplica-se a termos simples e compostos. Um termo que representa uma noção complexa é re-expressa na forma de um termo mais simples ou da definição, cada um dos quais pode também ocorrer em outras combinações para representar uma série de diferentes conceitos. Esta técnica não é recomendada."

Comentário:

De fato, a fatoração semântica é uma técnica que só tem sentido quando se considera o conceito. Isto, porque a fatoração resulta em uma seleção de características do conceito, as quais são, também, conceitos, por isso que podem 'também, ocorrer em outras combinações'. Contrariamente ao que recomenda a Norma, esta técnica é a única que convém, justamente porque as características são também conceitos e, como tal, podem ser estruturados no sistema de conceitos. Por exemplo, num instituto de tecnologia alimentar o termo de uso comum 'leite de soja' não é usado, mas apenas duas de suas características, a saber, os conceitos 'soja' e 'extrato protéico'.

A respeito da fatoração sintática, a Norma diz "*ela pode ocorrer na análise morfológica em componentes separados, cada um dos quais pode ser aceito como um termo de indexação por si só*". Não recomenda esta técnica para termos sincategoremáticos.

Comentário:

A linguagem utilizada está coerente. Observamos que a redação se refere a 'termo de indexação' e não a 'conceito'. Como afirmamos no início do trabalho, se pensarmos apenas em um vocabulário controlado, não há maiores considerações a fazer. Mas se pensarmos num tesouro, que exige relação entre os termos, então o termo 'composto', quando fatorado, deixa de existir e, portanto, não é incluído na estrutura. Dito de outro modo, se o termo 'composto' designa um conceito, este deixará de estar

representado no tesauro (deixa de existir) pois apenas seus componentes estarão incluídos.

Quanto aos termos sincategoremáticos, não há um argumento para não se recomendar esta técnica. Quando a análise é terminológica, e não lingüística, como a proposta neste item da Norma, o termo sincategoremático denota um conceito e, portanto, é indivisível.

Em 7.2, a orientação é voltada para a não-fatoração.

Devem ser mantidos como compostos:

"a) termos muito familiares no uso comum". Exemplo: 'data processing'.

Comentário:

O conhecimento é dinâmico e, como um corolário, também a língua. Termo 'familiar' para quem? Quando? Que parâmetros adotar para aceitar a 'familiaridade'?

"b) perda de sentido' quando se combinam os termos. Exemplo "food + plants" pode significar tanto "food for plants", como "plants as food".

Comentário:

Mais uma vez, a decisão gira em torno das palavras. Ora, "food for plants" denota uma classe de substâncias para nutrir as plantas. É um conceito. Quanto a "plants as food" tem-se igualmente um termo que denota uma classe específica de plantas, com certas características que a tornam comestíveis. Os elementos do termo são indivisíveis porque denotam um conceito e cada um de seus elementos denota outros conceitos.

Na Subseção 7.2.1 há orientação para os casos em que os 'termos compostos' não devem ser separados, a saber :

a) nomes próprios

b) termos em que a diferença perdeu seu significado original. Exemplo: 'lawn tennis' e 'trade winds'.

Comentário:

Ora, se os termos denotam conceitos, por que se preocupar com o significado original de um dos elementos do termo? O importante é saber a qual objeto o termo se refere.

c) Termos que contêm uma diferença que sugere uma semelhança, como um *símile*, com uma coisa ou evento não relacionado. Exemplo: 'tree structures', 'wing nuts'

d) Termos que não podem ser re-expressos (por exemplo, definidos), sem o uso de um substantivo extra que esteja presente no termo composto apenas por implicação. Exemplo: 'free escapes' (escapes for people from fire)

e) Termos que contêm nomes sincategoremáticos. Exemplo: 'artificial limbs'; 'fossil fishes'

Comentário:

Embora não explicitamente, o que está sendo considerado aqui são designações de conceitos, embora, claramente o discurso se refira à questão de isolar ou não as palavras constitutivas do termo.

A Seção 7.3.1 trata de termos que devem ser fatorados sintaticamente:

a) um termo composto deve ser fatorado se o foco se refere a uma propriedade ou parte (inclusive materiais) e a diferença representa o todo ou o possuidor daquela propriedade ou parte. Exemplo: 'aircraft engines'; 'hospital personnel'; 'soil acidity'

Comentário:

Os exemplos acima deixam claro que a norma atua no plano da língua; a estrutura das expressões acima não existe em língua portuguesa, a saber substantivo seguido de substantivo. Se traduzirmos as expressões, verificaremos que o resultado se enquadra numa situação em que a Norma não recomenda fatorar.. 'Motor de avião', em que pese a presença da preposição 'de' significa 'motor para avião' (exatamente como em 'hospital for children'). Em português, a preposição 'de' é uma verdadeira armadilha para este tipo de análise.

b) ao contrário, o nome do todo pode ser modificado pelo nome de sua parte ou propriedade. Exemplo: 'acid soils'. 'concrete bridges'; 'skilled personnel'

Comentário:

Na verdade, o adjetivo não pode ser separado do substantivo porque tesouros só incluem substantivos. Isolar o adjetivo equivaleria a eliminar o termo/conceito.

A subseção 7.3.2 trata de expressões relativas a ações ou eventos:

a) o nome de uma ação transitiva não deve ser modificado pelo nome do objeto sobre o qual a ação se efetua. Exemplos: 'office management'; 'steel casting'; textile printing

Comentário:

Mais uma vez estamos diante de uma estrutura inexistente na língua portuguesa, a saber substantivo seguido de substantivo. Mas noutra parte da norma há uma referência a uma expressão em francês "pollution de l'air", que corresponde a 'air pollution'. Ali não se propõe fatorar. Então, temos uma inconsistência. Este fatoramento é possível do ponto de vista lingüístico, pois se trata de dois substantivos. Do ponto de vista conceitual, porém, deixa a desejar.

b) ao contrário, o nome de uma coisa ou material pode ser modificado pelo nome de uma ação efetuada sobre ela/ele. Exemplo: 'cast steels'; 'extruded plants'; 'printed textiles'

Comentário:

O adjetivo não pode ser considerado 'nome de uma ação'. Ações são designadas por substantivos ou verbos.

Ainda com relação a ações ou eventos, segue a Norma:

a) o nome de uma ação intransitiva não deve ser modificado pelo nome do realizador da ação. Exemplos: 'bird migration'; 'plant growth'; 'metal deterioration'

Comentário: mais uma vez temos o modelo substantivo seguido de substantivo, inexistente na língua portuguesa. Isto invalida o caráter internacional da Norma

b) ao contrário, o nome de uma coisa pode ser modificado pelo nome da ação intransitiva na qual ela está ou foi envolvida. Exemplos: 'degraded plants'; 'migrating birds'; 'moving pavements'

Comentário: Fica claro que o modelo da análise é lingüística: o adjetivo não pode ser considerado um termo.

3.1.1 Comentário geral sobre a questão do fatoramento de termos compostos

O primeiro tipo de fatoração discutido é o da fatoração semântica, que não é

recomendado. No entanto, o exemplo relativo a 'leite de soja' mostra que, do ponto de vista conceitual, este fatoramento faz sentido.

Quanto aos demais casos, é provável que as justificativas talvez não cubram todos os casos que venham a surgir. Uma coisa, no entanto, é certa, se cada uma das expressões incluídas nos exemplos denota um conceito, se se referem, cada uma, a um objeto abstrato ou concreto, então são termos, e nada mais. Portanto, não cabe qualquer discussão sobre fatoramento sintático ou morfológico.

Outro aspecto a considerar: se adotarmos a técnica de fatoramento sintático ou morfológico, o termo que denota o conceito fica sem estrutura no tesauro, ou seja, aquele conceito expresso por um grupo de palavras, deixa de figurar no tesauro.

4. Considerações finais

A adoção da Teoria do Conceito e da Teoria da Terminologia, como fundamentos para a elaboração de tesouros documentários fornecem as condições para se construir instrumentos de trabalho consistentes, porque a análise tem início no conceito. Como se pôde perceber, na Norma 2877 a análise tem início na expressão verbal. Daí decorrem todas as inconsistências apontadas.

De fato, não se está negando a possibilidade de recuperação de informação através de indexação automática, vendo-a, no entanto, como ação complementar, pois um usuário busca por palavras ou expressões com as quais está familiarizado e o sistema deve responder a isto. No entanto, o tesauro é mais do que simplesmente disciplinar o uso de termos: ele é uma estrutura de conceitos e, no momento em que tanto se fala em organização de domínios de conhecimento e sua representação, a estrutura conceitual não pode ser desconsiderada.

Estas questões são pouco discutidas seja em aula seja nos encontros profissionais ou mesmo em listas de discussão. Para os informáticos, no entanto, os princípios teóricos do tesauro começam a ganhar importância nas discussões sobre metadados, interoperabilidade e outras questões assemelhadas, pois o que se pretende estruturar, naquelas áreas, não são palavras, mas objetos, cuja manipulação se torna possível através de suas denominações (ou termos).

Ferramentas semânticas como tesouros precisam ser construídas respeitando princípios teóricos consistentes, abarcando domínios de estudos e pesquisa em língua portuguesa, para serem utilizadas em bibliotecas digitais e virtuais, em sistemas para gestão de conhecimento, para viabilizarem processos de integração de informações entre pesquisadores, como auxílio para as ferramentas de busca de um modo geral, e principalmente como um instrumento para a melhoria do tratamento e da recuperação de informação na rede. Além disso, o uso de conceitos como base para o estabelecimento dos termos em um tesauro permite maior facilidade para o estabelecimento de equivalentes em outras línguas.

Atualmente, o crescimento da utilização de arquiteturas distribuídas, especialmente no ambiente da Web, de interfaces abertas de acesso a bancos de dados, de tecnologias de mediadores e de padrões de formato para troca de dados contribuiu para que informações originalmente isoladas sejam disponibilizadas para serem utilizadas de maneira integrada. Tesauro assume papel fundamental nesta integração, viabilizando a interoperabilidade semântica de sistemas distribuídos heterogêneos.

Se o profissional de informação quiser atuar de forma competente nesta área, há que dominar amplamente as bases teóricas aqui defendidas, para não ficar refém de uma dada aceitação de padrões internacionais, que em muitas vezes, não possuem um princípio teórico consistente.

Notas e Referências Bibliográficas:

CAMPOS, Astério T. (1986) Linguagens documentárias. *Revista de Biblioteconomia de Brasília*, v. 14, n. 1, p. 85-88, jan./jun.1986.

CAMPOS, Maria Luiza de Almeida. Linguagens documentárias: núcleo básico de conhecimento para seu estudo *R Esc. Biblioteconomia UFMG*, Belo Horizonte, v.24, n.1, jan./jun. 1995. p.52-62.

CAMPOS, Maria Luiza de Almeida. *Linguagem documentária: teorias que fundamentam sua elaboração* Niterói, RJ: Eduff, 2001.

CAMPOS, Maria Luiza de Almeida. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. *Ciência da Informação*, Brasília, v.33, n.1, jan./abril 2004.

CAMPOS, Maria Luiza de Almeida & GOMES, Hagar Espanha. Organização de domínios de conhecimento e os princípios ranganathianos. *Perspectiva em Ciência da Informação UFMG*, Belo Horizonte, v.8, n.2, jul./dez. 2003.

DAHLBERG, Ingetraut (1978). A referent-oriented analytical concept theory of INTERCONCEPT. *International Classification* v. 5, n. 3.p. 142-150.

FELBER, H. (1984) *Terminology manual* Paris: UNESCO.

GOMES, Hagar Espanha. Classificação, tesauro e terminologia: fundamentos comuns. (online) Capturado em 4 ago. 2004. Disponível na Internet <http://www.conexaorio.com/bititertulia>

GOMES, Hagar Espanha; CAMPOS, Maria Luiza de Almeida; MOTTA, Dilza Fonseca da. Tutorial para elaboração de Tesouros. (online) Capturado em 4 ago. 2004. Disponível na Internet <http://www.conexaorio.com/bititesauro>

HUNTER, Jane. MetaNet- A Metadata term thesaurus to enable semantic interoperability between metadata domains. *Journal of Digital Information*, Special Issue on Networked Knowledge Organization Systems, Volume 1, issue

8, April 2001. Disponível, também em
http://archive.dstc.edu.au/RDU/staff/jane-hunter/harmony/jodi_article.html

MOTTA, Dilza Fonseca da. Método relacional como nova abordagem para a construção de tesouros. Capturado em ago. 2004. Disponível na Internet
<http://www.conexaorio.com/bit/dilza/index.htm>

WÜSTER, E.(1981) L'Etude scientifique générale de la Terminologie, zone frontalière entre la Linguistique, la Logique, l'Ontologie, l'Informatique et les Sciences des Choses. In: RONDEAU, G. & FELBER, H. (org.) *Textes choisis de Terminologie. I. Fondements théoriques de la terminologie*. Québec, GIRSTERM, p. 57-114.

Sobre as autoras / About the Authors:

Hagar Espanha Gomes
Livre-docente pela Universidade Federal Fluminense; Consultora independente.

Maria Luiza de Almeida Campos
Doutora em Ciência da Informação pelo IBICT/UFRJ; Professora Adjunta do Departamento de Ciência da Informação da Universidade Federal Fluminense/Professora do Programa de Pós-Graduação UFF/IBICT.