

Fernanda Ribeiro e Maria Elisa Cerveira, org.

I Congresso ISKO Espanha e Portugal

XI Congreso ISKO España

7 a 9 de novembro de 2013

Informação e/ou Conhecimento:
as duas faces de Jano

Atas



Faculdade de Letras da Universidade do Porto
CETAC.MEDIA
ISKO



A TEORIA DO CAMPO SEMÂNTICO NO DESENVOLVIMENTO DE VOCABULÁRIOS ESTRUTURADOS
PARA A WEB SEMÂNTICA

VÂNIA MARA ALVES LIMA
Universidade de São Paulo
vamal@ups.br

CIBELE ARAÚJO CAMARGO MARQUES DOS SANTOS
Universidade de São Paulo
cibeleac@ups.br

MICHELY JABALA MAMEDE VOGEL
Universidade de São Paulo
michelyvogel@ups.br

Resumo As normas de tesouros da década de 1970 não atendiam a organização e recuperação da informação de recursos interconectados, e não deixavam claras as diferenças entre conceitos e termos. Na escolha do termo preferido através de critérios pré-estabelecidos, nem sempre o termo escolhido é o termo utilizado na busca. Os tesouros e vocabulários controlados resolveram as questões linguísticas que envolvem a sinonímia com a relação de equivalência, mas a polissemia precisou do artifício do uso de qualificadores dificultando ao usuário a identificação do termo correto. O processamento por máquina nesta solução também é insatisfatório. Os vocabulários estruturados para a Web Semântica precisam de modelos de dados com a distinção entre termo/conceito pela explicitação do campo semântico. Este trabalho pretende relacionar a teoria referente ao campo semântico apoiado nas definições dos conceitos e de suas denominações, os termos, como fundamento para o desenvolvimento de vocabulários estruturados para a Web semântica. Pesquisa teórica desenvolvida através de levantamento bibliográfico sobre vocabulários estruturados e semântica estrutural. Entre 2005 e 2012, as normas foram revisadas, porém as estruturas lidam com as mesmas dificuldades semânticas. As normas de vocabulários estruturados permitem mapeamento entre vocabulários, interoperabilidade e modelo de dados em XML que separa as classes de conceito e de termos de forma mais rigorosa. O W3C (World Wide Web Consortium) desenvolveu a SKOS (Simple Knowledge Organization Systems) para vocabulários estruturados com modelo de dados a distinção entre conceito e as etiquetas usadas para representá-los. Na relação entre conceitos e suas etiquetas são explicitados os campos semânticos, a área coberta ou seu domínio de significação com indicando a carga semântica que o delimita e seus contextos. As relações semânticas dos vocabulários estruturados devem ser desenvolvidas por especialistas com aportes da linguística, da lógica e da terminologia e estudos nessas áreas continuam sendo fundamentais para estes formatos, mesmo quando interoperáveis.

Palavras-chave Linguagens documentárias. Web semântica. Terminologia.

Abstract The theory of semantic field in the development of structured vocabularies for the Semantic Web. The standards for thesaurus development of the 1970s did not propose effective solutions for organization and retrieval of information from interconnected resources, and they did not make clear the differences between concepts and terms. In choosing the preferred term by using pre-established criteria, lots of times the chosen term was not used in the search. The thesauri and controlled vocabularies solved the linguistics questions on synonymy using equivalence relation, but the polysemy required the trick of using qualifiers, making more difficult for the users the identification of the correct term. Similarly, the machine processing in this kind of question was unsatisfying. In the quest for interoperability between those tools within the Semantic Web, arises the need for data models that operate with the distinction between term/concept by showing up the semantic field. This work related the theory referred to semantic field as the foundation for the development of structured vocabularies to the Semantic Web. Theoretical research done by bibliographical survey on

structured vocabularies and structural semantics. In 2005 to 2012, the standards were reviewed, but their frameworks deal with the same semantic difficulties. The structured vocabularies standards allow mapping between vocabularies, interoperability, and data model in XML, which separate the concept and terms groups in a more strict way. The W3C (World Wide Web Consortium) developed the SKOS format (Simple Knowledge Organization Systems) as a data model that consider the distinction between concept the labels used for its representation. In this relationship, the semantic fields, the covered area or signification dominion are shown up to indicate the semantic load that limits it and in which contexts it occurs. The semantic relations of structured vocabularies must be developed by specialists with contributions from Linguistics, Logic, and Terminology and studies of these areas are still fundamental to those formats, even when they are interoperable.

Keywords Indexing languages. Semantic Web. Terminology.

Introdução

Nos últimos dois séculos os bibliotecários tem desenvolvido linguagens documentárias, como vocabulários controlados e tesouros, para a busca e a indexação nos mais diversos sistemas de informação. Ocorre que ainda que desenvolvidos de forma estruturada, utilizando-se dos conhecimentos da linguística e da lógica e procurando estabelecer relações semânticas entre os termos que as compõem, estas linguagens documentárias não tem sido capazes de operar no nível conceitual quando se pensa na interoperabilidade entre elas no âmbito da web.

A busca na Web caracteriza-se pelo apoio de sistemas automáticos, os quais utilizam algoritmos estatísticos baseados em busca inteligente em texto completo (Vatant, 2010). Por este motivo, para os especialistas na área de informática, os motores de busca não precisariam de vocabulários controlados, pois utilizariam tecnologias de busca e indexação automática baseados em algoritmos que podem alcançar um alto nível de rendimento.

Por outro lado, os bibliotecários e especialistas em organização do conhecimento defendem que os motores de busca não tratam de forma adequada a ambiguidade dos termos e que a busca inteligente deveria aproveitar o conhecimento estruturado e o valor agregado destes vocabulários (Vatant, 2010).

Ainda que nem sempre o descritor escolhido pelo bibliotecário na indexação seja o mesmo utilizado pelo pesquisador na busca da informação, o tesouro, usado em contextos controlados por humanos ou com mediação de processo de pesquisa tem boa resposta, pois o usuário consegue interpretar os resultados realizando a inferência entre o termo e conceito, ainda que as diferenças entre os conceitos que envolvem uma informação e os termos expressos na busca por esta informação não estejam claras (Clarke e Zeng, 2012).

As normas de tesouros como a ANSI/NISO Z39.19 e a ISO 2788, desenvolvidas na década de 1970, não atendem mais as necessidades da organização e recuperação da informação em um mundo de recursos interconectados. Verifica-se isso pela maneira como buscam resolver as questões linguísticas.

No caso da sinonímia, por exemplo, onde vários termos remetem a um mesmo conceito, os vocabulários controlados e os tesouros utilizam-se da relação de equivalência. Já no caso da polissemia onde um mesmo termo pode remeter a mais de um conceito, as normas recorrem ao uso de qualificadores para indicar o campo semântico a que o termo pertence. Este recurso, em alguns casos, não só dificulta a identificação do termo correto para a busca, por parte do usuário, como também se mostra insatisfatório no processo de recuperação automática. Cabe, portanto, verificar se as atualizações que tem sido propostas para estas normas conseguem estabelecer as relações entre conceitos, entre conceitos e termos e entre termos, adequadamente no âmbito da Web semântica.

Na Web semântica, o software que a operacionaliza necessita de modelos de dados explicitados que indiquem a distinção entre termo/conceito e as relações que podem ser estabelecidas entre eles.

O vocabulário controlado e o tesauro, por serem linguagens documentárias estruturadas a partir do relacionamento entre seus descritores, ainda constituem-se nas ferramentas mais adequadas para representação e acesso à informação, quando seus descritores têm como referência os termos, isto é, as denominações dos conceitos do domínio que pretendem representar. Em contraposição, a abordagem da pesquisa por texto completo, apesar dos avanços, ainda não consegue operar de maneira efetiva quando o conceito, expresso por diferentes termos ou combinações de termos, ocorre em diferentes contextos.

De qualquer maneira, apesar dos esforços tanto na normalização, para a elaboração dos vocabulários controlados, quanto no desenvolvimento de ferramentas, como as ontologias, que buscam especificar e explicitar formalmente os conceitos compartilhados com o objetivo de garantir uma comunicação livre de ambiguidades, ainda é necessário chegar a um consenso com relação ao significado de cada elemento que operacionaliza a representação do conhecimento na web semântica.

A partir de aportes teóricos e metodológicos das áreas da Linguística, da Terminologia e da Organização do conhecimento, este trabalho apresenta a teoria referente ao campo semântico, apoiada nas definições dos conceitos e suas denominações, os termos, como um dos fundamentos que devem ser levados em consideração quando do desenvolvimento de vocabulários estruturados para a Web semântica.

1 Vocabulários controlados e normalização

O acesso ao conteúdo informacional de documentos individuais, tradicionalmente tem sido feito tanto pela superfície expressiva léxica e pelo contexto de referência, utilizando-se a terminologia do domínio, quanto pelo uso de vocabulários controlados formando redes semânticas. Essas redes semânticas incluem os conceitos, cuja designação pode ser feita por mais de uma palavra ou termo e devem indicar sinônimos e variações gramaticais com remissivas, conter subdivisões normalizadas e os rótulos explicativos para evitar ambiguidade (Moreiro González, 2011, p. 29).

As linguagens documentárias, ou seja, os sistemas de classificação e cabeçalhos de assunto, desenvolvidos nos Estados Unidos a partir do final do século XIX tem como fruto o tesauro através do uso do Unitermo em 1951, como uma ruptura ou um novo modelo do sistema pré-coordenado para o sistema pós-coordenado. Na década de 60 surge o tesauro do Centro de Informação do Ministério da Defesa dos Estados Unidos, o *Chemical Engineering Thesaurus* e baseado neste em 1964 é publicado o *Thesaurus of Engineering Terms*. No ano seguinte, estes dois tesauros são reunidos pelo projeto Lex (1965-1967) que tem por meta estabelecer princípios comuns para construção e uso, preparando um manual de construção de tesauros que foi recomendado como fonte e servindo de base para as diretrizes e normas produzidas pela ANSI (*American National Standardization Institute*) em 1981 e pela Unesco em 1973, que também deu origem a normas nacionais e internacionais na Europa (Campos & Gomes, 2006, p. 351-353).

A criação do Classification Research Group – CRG em Londres representa vertente diferenciada da vertente norte americana, segundo Campos & Gomes (2006, p. 353). O Grupo desenvolveu estudos e pesquisas voltadas para a classificação facetada de Ranganathan e influenciou uma geração inteira de profissionais. As bases classificatórias do tesauro foram desenvolvidas no Thesaurofacet por Aitchison que constrói o esquema de classificação com orientação para uma sintaxe tendo um índice alfabético decorrente do próprio sistema de classificação. No entanto, embora esta vertente tenha avançado quanto à apresentação sistemática, as autoras acima lembram que os tesauros adotam a base linguística na construção do vocabulário, o que pode ser comprovado nas normas nacionais e internacionais.

Na organização e representação da informação e do conhecimento, retomando Gardin (1966), Moreiro González (2011, p. 16) transpõe para a Web a ideia de que as linguagens documentárias consideradas para indexação podem ser apresentadas de acordo com a perspectiva do léxico

utilizado e das relações estabelecidas entre as palavras que as compõem. As linguagens podem ser listas de palavras que utilizam inclusive a folksonomia, listas de termos como glossários, listas de nomes e dicionários entre estes os anéis semânticos, classificações e taxonomias, bem como grupos de relações baseados em associações entre termos e conceitos como tesouros automáticos, tesouros conceituais, *topic maps* e ontologias.

A evolução das normas de tesouro levou, a partir de 2005, à quarta revisão da norma ANSI/NISO Z39.19 para vocabulários controlados pois com a era digital surgiram diferentes tipos de vocabulários, às vezes trabalhando de forma paralela aos tesouros. Esta revisão da norma da ANSI cobre vários tipos de vocabulários cujas abordagens e estruturas lidam com as mesmas dificuldades que os tesouros como as listas de termos controlados, os anéis de sinônimos e as taxonomias, mas esta norma ainda não traz o modelo de dados que torne clara e específica a relação termo/conceito (Clarke e Zeng, 2012).

Mudanças importantes trazidas por esta norma compreendem a referência a *objetos de conteúdo*, entidades que contêm informação no lugar de documentos, o uso de *termo* uma palavra que representa um conceito para referir-se aos descritores e termos preferenciais (Moreiro González, 2011, p. 71). As finalidades dos vocabulários controlados apresentadas na norma são a tradução do vocabulário natural aos termos, a consistência através da uniformidade na atribuição dos termos, indicação de relações semânticas entre os termos, etiquetagem, visualização e recuperação. Também a complexidade dos vocabulários controlados aumenta em relação a controle de ambiguidade, controle de sinônimos, relações hierárquicas e relações associativas (Moreiro González, 2011, p. 72; Santos e Mazini, 2011).

A norma britânica BS8723, que equivale a norma BS5723, voltou-se para vocabulários estruturados e foi publicada em 5 partes entre 2005 e 2008. Trata com profundidade do tesouro com menor ênfase em outros vocabulários, na quarta parte apresenta um guia para mapeamento entre vocabulários e na quinta parte traz um modelo de dados explícito com esquema em XML para intercâmbio de um tesouro inteiro ou suas partes. Este modelo de dados separa as classes de conceito e de termos (Clarke e Zeng, 2012).

Os tesouros que, segundo Moreiro González (2011, p. 69) apresentam vantagens como a existência em muitas áreas do conhecimento, sendo alguns muito testados e com eficácia comprovada, passam a ter mecanismos de compartilhamento na Web após a criação de SKOS (*Simple Knowledge Organization System*). Como desvantagens das versões tradicionais o autor lembra que os tesouros pertencem a domínios restritos, existe dificuldade de definir relações, é um vocabulário sem relação direta com os objetos, sua criação atende consenso prévio de criadores e não de usuários e possui custo elevado na criação, manutenção e funcionamento evitáveis somente com automação ou semiautomação de processos. No entanto, existem dificuldades na automação, devido às estruturas semânticas dos textos.

Em 2005, também o W3C (World Wide Web Consortium) apresentou sua Recommendation SKOS (Simple Knowledge Organization Systems) desenvolvida para publicação de vocabulários estruturados como os tesouros na WEB e traz no seu modelo de dados explicitamente a distinção entre conceito e as etiquetas usadas para representar conceitos. Segundo Clarke e Zeng (2012) no SKOS os relacionamentos semânticos entre conceitos correspondem de forma bem próxima aos relacionamentos hierárquicos e associativos recomendados nas normas de tesouro. O SKOS Core Specification foi revisado em 2009 e nesta versão o SKOS Core Data Model básico foi complementado com um apêndice contendo várias etiquetas que estendem as etiquetas para relações entre termos, denominado SKOS-XL.

Nesse movimento internacional, as normas ISO (2788/1986 – Tesouros Monolíngues e 5964/1985 – Tesouros Multolíngues) também foram revisadas por um grupo de trabalho com representantes de 15 países, as normas antigas foram revogadas e criou-se a Norma ISO 25964 *Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval* (publicada em 2011) e *Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies* (publicada em 2013). Esta norma, segundo Clarke e Zeng (2012) é mais rigorosa na

distinção de termos e conceitos. Embora a norma não esteja disponibilizada de forma gratuita na Internet o modelo de dados e o esquema XML estão disponibilizados gratuitamente no site da NISO.

Embora atualmente tenhamos modelos de metadados que permitem interoperabilidade entre vocabulários, as relações semânticas continuam sendo desenvolvidas através da análise cuidadosa de profissionais da informação quanto aos campos semânticos a que os termos e conceitos pertencem, o nível de relação existente entre eles, preferências de termos, identificação de sinônimos e resolução de termos ambíguos.

As linguagens que fazem parte dos sistemas de organização e representação do conhecimento KOS (*Knowledge Organization Systems*) para a Web, que podem ser vocabulários controlados ou não e pertencer a um mesmo domínio do conhecimento ou a domínios diferentes, apresentam grande complexidade disciplinar e de uso, com fundamentos conceituais de diversos campos como a Lógica Formal e a Estatística, a Computação, a Retórica, a Linguística do texto, a Semiótica e a Lexicografia (Moreiro González, 2011, p. 16-17).

Portanto, torna-se necessário discutir as questões linguísticas que envolvem a elaboração e o uso de linguagens documentárias e vocabulários controlados na web semântica.

2 Questões linguísticas e terminológicas

No contexto da Web Semântica, vocabulários estruturados são considerados Linguagens Documentárias, ou LDs (Caro-Castro, 2011; Campos, 2001, BS 8723 part 1, 2005, Clarke e Zeng, 2011). Essa nova web, de caracterização semântica, “requer linguagens de representação do conhecimento baseadas em ontologias para localizar e acessar os recursos” (Moreiro Gonzalez, 2011, p. 34). A ideia, para o autor, é controlar novamente a polissemia, a ambiguidade, muito presentes na web 2.0 com as folksonomias e tags.

Entre suas funções, as LDs são usadas para normalizar as descrições realizadas com metadados, e para recuperar informações (Caro-Castro, 2011, p. 9). Apresentam uma tríplice função, a saber:

Conseguir que cada descritor represente um só conceito (evitando a polissemia, homonímia e homografia) e que cada conceito esteja representado por um único descritor (controlando a sinonímia); Organizar e relacionar os conceitos de uma área de conhecimento geral ou especializada; Assegurar a consistência do sistema, de maneira que diferentes indexadores utilizem sempre o mesmo termo para expressar o mesmo conceito (Caro-Castro, 2006, p. 3).

Além disso,

a comunicação da informação só é possível através de uma linguagem em que haja significação para os envolvidos. Dar acesso à informação e promover sua circulação são ações que fazem necessário o uso de uma linguagem que a represente baseada em uma hipótese de organização que permita a construção de significação – a Linguagem Documentária. Dessa forma, a elaboração dessa linguagem requer metodologias que tenham como referência conceitos linguísticos e também terminológicos (Vogel, 2007, p. 56).

Assim, procuraremos relembrar tais parâmetros linguísticos e terminológicos que amparam as teorias das LDs, brevemente.

Em ensaio de 1997, para introduzir o conceito de Linguagens Documentárias, Tálamo aponta como base as propostas da Linguística Estrutural para o desenvolvimento daquelas linguagens e da própria área de Análise Documentária. De acordo com a autora,

problemas relativos à comunicação passam a integrar o quadro de discussões da área no momento em que se torna clara a importância da noção de linguagem. Evidencia-se que operar com o universo da linguagem não se confunde com a verbalização de denominações de classes pré-determinadas (Tálamo, 1997, p. 3).

É preciso entender, portanto, que é intrínseco à palavra significar de maneira própria a cada ocorrência (...). As LDs, ao contrário, porque são elaboradas para o exercício estrito da função informativa, compreendem unidades capazes de representar informação. Não é suficiente que tais unidades signifiquem. É necessário que elas signifiquem de maneira determinada (Cintra et al, 2002, p. 69).

Tálamo alerta para o fato de que

imagina-se que ao agrupar uma série de palavras típicas de uma área de conhecimento garante-se a forte credibilidade desse conjunto de palavras, tornando representativo. Ora, é somente a rede de relações das unidades de uma linguagem que pode contribuir para o exercício de qualquer representação (Tálamo, 1997, p. 4).

Quando optamos pela utilização de uma Linguagem Documentária para representação das informações de um sistema, desejamos, entre outras coisas, reduzir a redundância, as ambiguidades, polissemias, variações de seus elementos, ou seja, buscamos univocidade interpretativa via controle do vocabulário: escolhemos quais termos serão as unidades preferenciais, explicitamos ou indicamos a correspondência unidades/conceitos, e procuramos controlar a sinonímia. Dessa forma, também estabelecemos um controle sobre a significação, como um modo de orientar a busca de informações (Vogel, 2007, p. 62-63).

Por ter como objetivo a comunicação, a Linguagem Documentária é construída sob forma de linguagem. Mais que isso, a Linguística Documentária reconhece a natureza da comunicação documentária e afirma a necessidade de composição de quadros de referência para a análise, avaliação e construção dessa modalidade de linguagem construída (Tálamo, 2001, p. 142).

No entanto, não basta uma estrutura e relações para que a Linguagem Documentária seja bem sucedida; é preciso que, durante sua elaboração, seja estabelecido um princípio de classificação que se desenha de acordo com os usuários e a área de conhecimento ou atividade selecionada. Do mesmo modo, é necessário que se considere os objetivos da instituição ou comunidade que de fato fará uso dessa Linguagem Documentária. A elaboração deve levar em conta os conceitos utilizados nesse âmbito (Vogel, 2007, p. 63).

No ambiente digital e na web, um obstáculo comum é o fato de que os sistemas baseiam suas investigações em correntes de signos e não em conceitos. “Assim a palavra “cão” é tratada como a sucessão de signos “c, ã, o” e não como um conceito traduzível (dog, chien, perro, ...), que pertence à subclasses dos mamíferos, animais domésticos etc” (Moreiro Gonzalez, 2011, p. 34).

Em relação aos signos nas Linguagens Documentárias, Lara afirma que os “descritores não são etiquetas, mas o resultado de uma construção julgada pertinente para a significação” (Lara, 1999, p. 66). Para Le Guern (1989), palavras são independentes das coisas e os termos se referem às coisas, localizando-as em áreas do conhecimento. Para o autor, as palavras têm significado, mas não têm referência e representam propriedades e qualidades de um objeto e não o objeto. Uma palavra passa a ser um termo quando remete a uma classe que foi atribuída a um objeto dentro de um universo determinado. Este processo é denominado por Cabré (1993, p. 119) de terminologização, ou seja, as palavras da língua comum adquirem significados específicos, pertinentes a determinado campo do saber científico ou técnico.

Uma forma de compreender o que foi dito acima é comparar um dicionário de língua e um dicionário especializado. No primeiro, as palavras têm inúmeros significados; no segundo, têm o significado que se localiza numa área de especialidade, correspondendo, portanto, ao conceito. Assim, quando se pensa na organização de conteúdos informacionais por meio de descritores, são os conceitos sua referência, não as palavras (Vogel, 2007, p. 78). E aqui, apoiamos-nos na Terminologia, pois a conversão do conceito em termo de uma área específica vai garantir a função comunicativa de uma linguagem documentária, como é o caso dos vocabulários estruturados, isto é, vai garantir a representação da informação de maneira efetiva e adequada.

A relação entre Terminologia e Documentação, ou mais especificamente entre Terminologia e Linguagens Documentárias existe e é positivamente reconhecida por diversos autores, tanto da Terminologia como da Ciência da Informação (Le Guern, 1989; Lara 1993, 1999; Sager, 1990, 1993; Cabré, 1999; Tálamo et al, 1992; Tálamo 2001; entre outros).

A relação entre termo e conceito pode ser determinada segundo o conteúdo semântico de cada conceito, por meio de relações lógicas de gênero/espécie, causa/efeito, ou pelas relações associativas abstrato/concreto, parte/todo, objetivo/meio e relações de oposição. As relações do segundo tipo servem mais para explicar os conceitos do que para classificá-los. Aqui fica evidente, também, a presença da noção de estrutura, como sistema de relações mútuas entre os conceitos, garantia de sua significação (Vogel, 2007, p. 80).

Recorremos às normas terminológicas para entender Conceito, Características, Definição, e Designação:

- *Conceito*: unidade de conhecimento criada pela combinação única de características (ISO 1087-1, 2000, p.2) ou “representações mentais de objetos de dentro de um campo ou contexto especializado” (ISO 704, 2000, p.2);
- *Característica*: abstração de uma propriedade de um objeto ou de um conjunto de objetos. Existem categorias essenciais (indispensáveis ao objeto), e delimitadoras (usadas para distinguir objetos) (ISO 704, 2000, p. 3);
- *Definição*: representação de um conceito por meio de um enunciado descritivo que serve para diferenciá-lo de conceitos relacionados (ISO 704, 2000, p. 6). “A combinação única de características criando a intensão deverá identificar o conceito e diferenciá-lo de outros conceitos” (ISO 704, 2000, p. 15);
- *Designação*: representação de um conceito por um signo que o denote (ISO 704, 2000, p. 6). Pode ser símbolo, nome (designação verbal), e termo (designação verbal num campo específico).

Os conceitos contextualizados na linguagem especial do campo de conhecimento pode ser expresso nas várias formas de comunicação humana de acordo com o sistema utilizado. Na linguagem natural, conceitos podem assumir a forma de termos, nomes, definições ou outras formas linguísticas; na linguagem artificial, pode assumir a forma de códigos ou fórmulas, enquanto em gráficos, podem ter a forma de ícones, figuras, diagramas ou outras representações gráficas. Conceitos também podem ser expressos com o corpo humano como é feito na linguagem de sinais, expressões faciais ou movimentos do corpo (ISO 704, 2000, p. 2).

A formação do conceito acontece pelo reconhecimento de objetos e seu agrupamento em unidades significativas em um dado campo. Objetos que partilham as mesmas propriedades são agrupados em unidades. E essas propriedades são as características abstraídas desses objetos, e combinadas com um conjunto na formação do conceito. Características estão constantemente sendo combinadas a fim de criar conceitos, embora diferentemente em diferentes culturas, campos ou escolas de pensamento (ISO 704, 2000, p. 3).

Características são consideradas essenciais se elas forem indispensáveis para o entendimento do conceito num campo particular de conhecimento; a ausência de uma característica essencial fundamentalmente muda o conceito. A ausência de uma característica fundamental no curso de uma análise levará ao entendimento pobre ou mesmo errôneo do conceito (ISO 704, 2000, p. 4).

“Deve-se notar que a mesma propriedade de um dado objeto pode ser abstraída como uma característica essencial de um conceito em um campo de conhecimento mas não essencial em outro” (ISO, 2000, p. 5).

Essa discussão busca trabalhar a questão da ambiguidade. Para reduzi-la, uma das possibilidades seria colocar o termo, como, por exemplo, o descritor em inglês “Table” em categorias ou classes que o definam – “moveis para o lar” ou “bases de dados”. Como lembra Vatant:

Estes contextos têm de ser formalmente identificados e declarados já como sendo um conceito genérico ou relacionado, ou como um esquema conceitual (em função do nível de granularidade). Não é suficiente o simples indício do contexto por um qualificador como na prática dos tesouros (Vatant, 2010).

Sobre o uso de qualificadores, Vatant comenta que em casos assim, como por exemplo, “Atores – Psicologia”, “ainda que a solução pareça lógica aos usuários humanos, a descrição formal deste conceito não proporciona enlaces nem a “Atores” e nem a “Psicologia”. Para aplicações da Web semântica, esta relação deve ser explícita para ser utilizável” (Vatant, 2012). Ele prossegue:

Uma consequência indireta da elaboração destes exemplos para o gestor de vocabulário é obrigar a uma reflexão sobre a semântica subjacente da sintaxe, que deveria ter sido levada em conta há anos. Para a migração automática ao formato da Web Semântica, as construções sintáticas idênticas devem ser interpretadas e processadas de mesma

maneira. Mas parece que averiguar uma semântica geral para tais “—“ separadores, vai mais além de um geral e vago “SKOS: related”, e que tal semântica seja válida totalmente em um vocabulário grande, como RAMEAU ou LCSH, não é uma tarefa óbvia. Mesmo havendo pautas para os administradores de vocabulários explicando em linguagem natural o que dever ser a utilização correta de cada construção sintática, em geral não há maneira de fazer cumprir essas diretrizes no sistema de gestão de vocabulários, por não falar da avaliação das consequências de seu uso incorreto (Vatant, 2010).

Completando essa noção, Cabré e Codina (2001, citados por Caro-Castro, 2006, p. 1) afirmam que:

Tanto o conhecimento armazenado nos documentos como a necessidade de informação dos usuários se representam mediante conceitos, seriam estes a via ideal para estabelecer relações entre ambos os extremos. Porém, como não podemos manejar diretamente os conceitos, temos de fazê-lo por meio dos signos que os representam.

Exemplo interessante é dado por ECO no texto Marco Polo e o Unicórnio (Eco, 1997, p. 55-56). Em suas viagens, Marco Polo depara-se com um rinoceronte, animal que não conhecia de nenhuma forma. Cruza algumas características que vê com algumas que ouviu falar: possui quatro patas, e tem um único chifre. Chega à noção de unicórnio. No entanto, esta noção possui outros traços além desses: unicórnios são brancos, ágeis, belos e dóceis, ao passo que o animal que vê diante de seus olhos possui pelos de búfalo, pés de elefante, chifre negro e desgracioso, e é feio e arisco. Assim, faltam a Marco Polo características para constituir o conceito.

Voltando à Web Semântica, e sua orientação ao conceito e não a descritores, entendemos que a análise linguística e terminológica dos traços desses conceitos se faz necessária para podermos criar ambientes de interoperabilidade realmente funcionais. Ou seja, é possível afirmar que as implicações linguísticas e terminológicas continuam as mesmas já observadas nas LDs antes da Web Semântica. O controle do vocabulário, ponto de vista da significação não se resolve pela simples interoperabilidade. Novamente, nesse processo, apenas cadeias de significantes ou de letras estão sendo lidos, não conceitos.

É nesse contexto que inserimos a discussão sobre a teoria do campo semântico apoiada na definição dos conceitos e suas denominações, os termos do domínio, como um dos fundamentos a serem considerados quando do desenvolvimento de vocabulários estruturados para a web semântica e a efetiva interoperabilidade lexical, semântica e terminológica desses instrumentos.

3 O Campo semântico, a definição e as ontologias

Determinar o campo em linguística é, segundo os pressupostos epistemológicos, procurar descobrir a estrutura de um domínio dado, ou propor a ele uma estruturação (Dubois, et al, 2011 p. 95).

A noção de campo semântico aparece sempre relacionada ao trabalho de Jost Trier na década de 30, sobre o vocabulário alemão do conhecimento do século XIII e XIV, ainda que o autor preferisse utilizar os termos campo lexical, campo conceitual, campo linguístico de signos, etc. (Germain, 1981, p. 39). O mérito de Trier, para Germain (1981, p.42) seria o estudo sistemático do

campo da inteligência, reunindo sob um todo coerente, as ideias de outros autores, o que permite identificá-lo como o criador da noção de campo semântico.

Como bem coloca Lopes (1987, p. 242) Trier, observou que as unidades léxicas de uma língua encontram-se organizadas em grupos estruturados de tal modo que cada unidade fica ali definida pelo lugar que ocupa respectivamente à posição das demais, isto é, o sentido de elementos lexicais é especificado somente por sua semelhança e sua diferença em relação aos outros elementos relevantes do campo.

A teoria de Trier pode ser resumida em: o léxico de uma língua aparece constituído por um conjunto hierarquizado de palavras (campos léxicos) de modo que cada grupo de palavras recobre exatamente um âmbito bem delimitado ao nível das noções (campos conceituais) e cada um destes campos, tanto léxico quanto conceitual, está formado de unidades justapostas como as pedras irregulares de um mosaico (Germain, 1981, p. 42).

Esta hipótese parte das seguintes premissas, 1) existe anteriormente à estruturação de uma linguística formal, uma certa organização não linguística da experiência humana. 2) a totalidade do léxico de uma língua se apresenta com uma hierarquização de campos léxicos justapostos, sem lacunas e nem sobreposições; 3) uma palavra só adquire seu significado por oposição as outras unidades do mesmo campo (Germain, 1981, p. 44).

Na medida em que a estruturação constitui a essência fundamental da linguagem, todos os elementos linguísticos são resultado dessa estruturação. Assim, Trier denomina campo conceitual as organizações ao nível do pensamento e campo lexical as organizações ao nível da língua (Germain, 1981, p. 46).

Já Matoso Câmara Jr. distingue a organização dos elementos do léxico no nível lexical e no nível semântico, onde o campo lexical é formado por famílias etimológicas, isto é, constituído por palavras cognatas, já o campo semântico é formado por famílias ideológicas, isto é, constituído por palavras do mesmo universo de significação (Bastos, 2002, p. 327).

Genouvrier e Peytard (1974, p. 318) definem campo lexical como o conjunto de palavras que a língua agrupa ou inventa para designar os diferentes aspectos, ou os diferentes traços semânticos, de uma técnica, de um objeto, de uma noção: o campo lexical do automóvel, da álgebra, etc. No que se refere ao campo semântico, o definem como o conjunto de empregos de uma palavra (ou sintagma, ou lexia) onde e pelos quais a palavra adquire uma carga semântica específica e a delimitação destes empregos se daria pelo levantamento de todos os contextos imediatos que a palavra recebe num texto dado.

De acordo com o Dicionário de Linguística (Dubois, et al., 2011, p. 532) denomina-se campo semântico a área coberta, no domínio da significação, por uma palavra, ou por um grupo de palavras da língua. Assim, podemos identificar com pertencentes ao campo semântico da palavra *mesa* as seguintes palavras: *mesa de refeições*, *mesa redonda*, *mesa da Assembleia*, *mesa de operação*, *mesa de som*, etc.

Neste trabalho utilizaremos a denominação campo semântico para indicar os limites, no plano da significação, do emprego dos descritores de um vocabulário estruturado no processo de representação e de recuperação de uma informação no âmbito da web semântica.

A estruturação de um vocabulário para fins de representação e recuperação da informação inicia-se com a identificação das categorias de um dado domínio, como por exemplo, no caso da *Medicina*, no Vocabulário Controlado da USP, onde constam como grandes categorias: *Diagnóstico*, *Doenças*, *Equipamentos e Provisões Hospitalares*, etc..., sob as quais estão agrupados todos os demais termos do domínio.

Da mesma maneira, segundo Hernando Cuadrado (1995, p. 98-99) as palavras que integram um campo semântico devem pertencer à mesma categoria, pois o campo semântico de uma palavra está constituído por todas as que se relacionam com ela no plano do significado, correspondendo a grupos de associações como:

- significante e significado (calor, caloria, calorífico)
- significado (calor, ardor, energia, fervor)
- significado e referente (fresco, frio, quente)

O autor indica que esta caracterização deve ser restrita por algumas limitações, do tipo: a condição mínima para que as palavras pertençam a um mesmo campo é que tenham um traço significativo (sema) comum (quanto maior o número de semas tanto mais coerente será o campo semântico e no geral um menor número de palavras o integra); uma palavra pode tomar parte de todos os campos semânticos que se construam sobre qualquer dos traços significativos que se descubram nela; quando uma palavra possui vários significados, cada um deles pertence a um campo semântico distinto.

Umberto Eco em seu *Tratado de Semiótica Geral* (1981, p. 141-156) coloca que os campos semânticos são segmentários e mostram a visão de uma cultura e que nessa cultura podem existir campos semânticos complementares ou contraditórios. Da mesma maneira, uma mesma unidade cultural pode, dentro de uma mesma cultura, formar parte de campos semânticos diferentes. No âmbito de uma cultura, um campo semântico pode desfazer-se e voltar a reestruturar-se em um novo campo.

Da mesma maneira, a hierarquia interna dos termos de um campo semântico pode ser alterada a partir de mudanças externas que afetam seu contexto. Por exemplo, com relação ao termo **calças** que antes tinha o traço **só para homens**, o qual foi perdido, já que a peça passou a integrar o guarda roupa feminino (Hernando Cuadrado, 1995, p.99).

No caso da linguagem natural, a dificuldade em operacionalizar a constituição de um campo semântico está em que cada língua expressa uma cultura própria, uma visão de mundo, que se manifesta através do léxico. Um grupo pode ter interesses particulares que o levam a distinguir alguns aspectos significativos que não tem importância para os demais. Por esse motivo, no caso da linguagem documentária, como os vocabulários estruturados para a web semântica, estes devem ter como referente à terminologia do domínio, pois esta não só é reconhecida pela comunidade no processo de busca pela informação, mas também obedece a critérios precisos de denominação e definição, apoiados nas normas internacionais.

Segundo Aguado de Cea et al (2009, p. 231) as categorias utilizadas para etiquetar o campo semântico estão estritamente relacionadas com outros estudos em semântica que recorrem ao léxico e categorizam o mundo real como os dicionários, os tesouros e as redes semânticas. Os tesouros agrupam as palavras de acordo com um conjunto padronizado de relações como as hierárquicas, as associativas e as de equivalência. As redes semânticas agrupam os termos em um domínio de conhecimento de acordo com as relações reais que ocorrem nesse domínio. Em qualquer um dos casos, se coloca a necessidade de desambiguação do sentido, tema de estudos na linguística computacional que tem como objetivo a elaboração de etiquetagem automática do significado, seja no caso de uma palavra com vários significados, seja na coincidência formal de duas palavras diferentes.

Com o desenvolvimento da web semântica o autor coloca a necessidade de novos tipos de anotações capazes de indicar o significado ou conteúdo semântico de um documento para que seja processado por computadores. Essa marca semântica que inclui o conteúdo informacional da página a ser publicada na web semântica passa pela escolha da linguagem (XML, RDF, OWL), pela adição de comentários ou anotações semânticas (metadados) e pela racionalização com estes metadados. Para chegar ao consenso do significado desenvolvem-se as ontologias (Aguado de Cea et al, 2009, p. 245).

No sentido filosófico nós podemos nos referir à Ontologia como um sistema particular de categorias a partir de uma determinada visão de mundo (Guarino, 1998). De acordo com Breitman (2005, p. 7) ontologias são especificações formais e explícitas de conceptualizações compartilhadas.

Modelos conceituais que capturam e explicitam o vocabulário utilizado nas aplicações semânticas, servem como base para garantir uma comunicação livre de ambiguidades na chamada web semântica.

Para Moura (2011, p. 167) as ferramentas ontológicas são dispositivos informacionais contextualizados e desenvolvidos sob um ponto de vista específico com o objetivo de orientar a organização da informação temática em um dado contexto informacional. Ao mapear uma dada área do conhecimento, indicando suas categorias gerais e específicas esta ferramentas também operam no âmbito do campo semântico.

Segundo Degoulet et al citado por Moura (2011, p. 168) a interoperabilidade semântica requer a articulação de três elementos: a terminologia, a explicitação das relações semânticas dos elementos da terminologia e a representação dos conceitos nas mensagens. A terminologia seria o vocabulário controlado do domínio. As relações semânticas articulam a rede de conceitos deste domínio e expressam a natureza de cada conceito que a integra. A representação dos conceitos nas mensagens visa compatibilizar as diferentes formas através das quais as organizações se referem aos dados semelhantes. Essa representação só é possível com a inclusão das definições de cada conceito na ontologia, definição essa que deve compilar os enunciados verdadeiros, isto é, as características essenciais e acidentais a respeito de determinado objeto ao qual o conceito se refere (Dahlberg, 1978, p. 106).

Como a rede de relações entre os conceitos de um domínio é estruturada a partir destas características é possível mapear o campo semântico deste domínio. Da mesma maneira, estas características elencadas em uma definição expressam a natureza de cada conceito, indicando a sua posição no interior do campo semântico e possibilitando inúmeras articulações semânticas.

Parece claro que para a interoperabilidade entre instrumentos que pretendam representar o conhecimento na web semântica deve ser considerado o mapeamento dos vários campos semânticos que constituem um domínio, ou mais especificamente as características elencadas nas definições dos conceitos que constituem esse campo. Dessa maneira é possível estabelecer um vínculo entre termo, conceito e significado.

A definição resulta de uma análise e descrição de grandezas signicas, situando-se, pois, no nível semiótico pelo menos no que diz respeito ao seu ponto de partida. Se o conceito é resultado de uma interpretação de fatos naturais e/ou culturais, a definição é o resultado de uma interpretação de unidades lexicais. Conceituar é o processo de um modelo mental que corresponde a um recorte cultural e, em seguida, de escolha da estrutura léxica que pode manifestá-la de maneira eficaz. Este processo parte do mundo natural. Definir é o processo de analisar e descrever o semema linguístico para reconstruir o modelo mental: o seu ponto de partida é a estrutura linguística manifestada (Barbosa 2001, p. 153).

Ao tratar da problemática da definição a autora também distingue campo conceptual, campo lexical e campo semântico e suas respectivas unidades: conceito, vocábulo/termo e semema. O campo conceptual, conjunto de conceitos, é o resultado do processo de conceptualização do “saber sobre o mundo” é pré-linguístico. O campo lexical, conjunto de vocábulos/termos que tem um núcleo sêmico comum, resulta da conversão da informação conceptualizada em significação linguística. O campo semântico constitui um conjunto de sememas e resulta da intersecção do significado das unidades lexicais (vocábulos/termos) de um campo lexical. As relações entre os três não são simétricas, visto que um campo lexical pressupõe e contém necessariamente seus correspondentes no campo conceptual e no campo semântico, entretanto um campo conceptual pode não ter ainda os campos lexicais e semânticos que lhes correspondem. Assim, a definição resulta do fazer interpretativo do sujeito que articula os semas que identificam o conceito, o que em um discurso terminológico é eminentemente paradigmático, como processo e como produto (Barbosa, 2001, p. 154).

A definição permite ainda, situar o conceito dentro do sistema conceitual a que pertence relacioná-lo e delimitá-lo a respeito de outros conceitos graças a suas características (Sager , 1993 apud Azarian, 2011, p. 3).

Azarian e Tebe (2011, p. 13) ao discutirem a metodologia para a elaboração da definição terminológica citam Cabré (1993) e De Besse (1997) que consideram a definição terminológica como um ponto intermediário entre a definição linguística e a definição ontológica: a definição linguística inclui as características que permitem diferenciar uma noção das demais no sistema linguístico; a definição ontológica consiste na enumeração exaustiva das características de um conceito e tem como finalidade incluir toda a informação possível relativa a este conceito, ainda a definição terminológica pretende identificar um conceito no âmbito temático dado mediante a enumeração das características essenciais que devem ser somente as características necessárias e suficientes para esta identificação.

Desmet (2002, p.177) ao analisar a problemática da definição terminológica indica que elaborar uma definição é o primeiro nível de análise do sentido dos termos. Segundo a autora a definição do dicionário segue o modelo aristotélico de inclusão lógica ou hiperonímica em que uma classe de objetos é incluída em outra classe de objetos. A especificidade é marcada pela inclusão de traços distintivos do referente, que marcam a diferença entre duas unidades lexicais classificadas sob o mesmo hiperônimo.

Esta definição, segundo Desmet (2002, p. 181) é a base tanto dos dicionários quanto das terminologias. As principais diferenças ocorrem na escolha do includente e das diferenças específicas. Por exemplo, “pinheiro” pode ter como hiperônimo tanto “árvore resinosa” (gênero comum a todos os falantes para os dicionários) quanto “conífera” (gênero próximo que é um termo técnico da Botânica para a terminologia).

O termo, segundo Cabré (2000) é uma unidade singular e às vezes similar a outras unidades de comunicação, dentro de um esquema global de representação da realidade, admitindo variação conceptual e denominativa, e levando em conta a dimensão textual e discursiva dos termos.

A norma ISO 1087 (1990) só reconhece a definição por compreensão, também denominada definição intensional e a definição extensional. Temos aqui dois aspectos da definição, sua **extensão** e sua **intensão** ou compreensão. A **extensão** da palavra consiste na classe de todas as coisas a que o objeto se aplica. Já a **intensão** ou **compreensão** de uma palavra é a classe das propriedades que determinam os objetos a que a palavra se aplica. Consiste nas propriedades que um objeto precisa possuir a fim de estar na extensão do vocábulo.

A questão da extensão e da intensão também está presente nas ontologias, ao afirmar que uma ontologia é uma especificação formal e explícita de uma conceptualização compartilhada, Gruber (1993) considera que a conceptualização reside no nível das relações extensionais. Já Guarino e Giarretta (1995) que afirmam que o mais adequado seria investir na noção de que a conceptualização reside no nível das relações intensionais.

Isto é corroborado pelo fato de que a definição por extensão é imprecisa, pois supõe que o sujeito pode reconhecer outros elementos da extensão por similaridade em relação ao exemplo selecionado. No caso, quando se exhibe objetos reais da extensão da palavra, por exemplo, para fixar o significado da palavra *cão* se exhibe uma grande variedade de cães. O mesmo não seria possível se tivéssemos que utilizá-la para definir paixão, dor, justiça, azul, amarelo, etc. Já a definição por intensão vai indicar o gênero próximo e a diferença específica relacionando os critérios, traços, características que permitem dizer se um determinado objeto cabe ou não dentro dessa classe. Assim, para definirmos essencialmente algo precisamos: enunciar o gênero próximo ao qual pertence o conceito e enunciar a diferença específica. Ao enunciarmos o gênero próximo indicamos o campo semântico, ao enunciarmos a diferença específica estabelecemos o seu limite.

Em outras palavras, delimitamos o campo semântico em que o conceito/termo/descriptor se inscreve ao instaurarmos o traço comum a toda uma classe de termos, que estarão, por essa razão, associados, ou seja, é a classe em que o sentido específico de um conjunto de termos se define. Ao fornecermos os traços específicos do conceito/termo/descriptor de modo a individualizá-lo no interior de uma classe estabelecemos a disjunção entre os conceitos/termos/descriptores dentro do mesmo campo.

A partir disso, coloca-se a necessidade de buscar uma metodologia que permita operacionalizar a demarcação do campo semântico, apoiado nas definições intensionais dos conceitos/termos/descriptores de maneira que estas definições sejam incluídas como metadados na modelagem de ontologias interoperáveis que permitam não só a organização do conhecimento do domínio, mas também a recuperação da informação.

Considerações finais

Entre 2005 e 2012, as normas foram revisadas, porém as estruturas lidam com as mesmas dificuldades semânticas. As normas de vocabulários estruturados permitem mapeamento entre vocabulários, interoperabilidade e modelo de dados em XML que separa as classes de conceito e de termos de forma mais rigorosa. O W3C (World Wide Web Consortium) desenvolveu a SKOS (Simple Knowledge Organization Systems) para vocabulários estruturados com modelo de dados a distinção entre conceito e as etiquetas usadas para representá-los. Na relação entre conceitos e suas etiquetas, ou seja, entre conceitos, termos e descritores, os quais acabam transformados em metadados, é necessário explicitar os campos semânticos, a área coberta ou seu domínio de significação, indicando a carga semântica que o delimita e seus contextos. Essa delimitação só é possível através das características presentes nas definições de cada elemento do campo.

As relações semânticas dos vocabulários estruturados portanto, devem ser desenvolvidas por especialistas com aportes da linguística, da lógica e da terminologia e estudos nessas áreas continuam sendo fundamentais para estes formatos, mesmo quando interoperáveis.

Sem esse aporte, continuaremos trabalhando no nível sintático.

Referências

- AGUADO DE CEA, et al. (2009). Una vision interdisciplinar de la anotación semântica. In A. Alcina, E. Valero, E. Rambla (Eds.), *Terminologia y sociedad del conocimiento* (pp. 219-254). Bern: Peter Lang.
- AZARIAN, J., Tebe, C. (2011). *La metodologia de elaboración de definiciones terminológicas en vocabulários normalizados: analisis em normas UNE*. Debate Terminologico(7), 2-23.
- BARBOSA, M. A. (2001). A construção do conceito nos discursos técnicos-científicos, nos discursos literários e nos discursos sociais não-literários. *Revista Brasileira de Linguística*, 11(1), 31-60.
- BARBOSA, M. A. (2001b). Da formação do conceptus à estruturação semântica lexical. *Abralin: Boletim da Associação Brasileira de Linguística*, 1, 153-155.
- BREITMAN, K. (2005). *Web semântica: a internet do futuro*. Rio de Janeiro: LTC.
- CABRÉ, M. T. (1993). *La terminologia: teoria, metodologia y aplicaciones*. Barcelona: Antartida/Empuries.

- CABRÉ, M. T. (1999). Situación y aspectos generales de la terminología. *La terminología: representación y comunicación: elementos para una teoría e base comunicativa y otros artículos* (pp. 21-68). Barcelona: IULA/Universitat Pompeu Fabra.
- CAMPOS, M. L. A. (2001). *Linguagem documentária: teorias que fundamentam sua elaboração*. Niterói: Ed. UFF.
- CAMPOS, M. L. A.; GOMES, H. E. (2006). Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. *Perspectivas em Ciência da Informação*, 11(3), 348-359.
- CARO-CASTRO, C. (2006). Léxico y documentación: del lenguaje natural al lenguaje documental. *Nuestras palabras: entre el léxico y la traducción* (pp. 127-145). Frankfurt: Veuver.
- CARO-CASTRO, C. (2011). Vocabulários estruturados, Web Semântica y Linked Data: oportunidades y retos para los profesionales de la documentación. Paper presented at the *II Seminário de Estudos da Informação : Arquivologia, Biblioteconomia e Ciência de Informação : Identidades, Contrastes e Perspectivas de Interlocução*, Niterói, Brasil.
- CINTRA, A. M. M, et al. (2002). *Para entender as linguagens documentárias* (2ª ed. rev. e ampl ed.). São Paulo: Polis.
- CLARKE, S. G., Zeng, M. L. (2012). From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. *Information Standards Quarterly*, 24, 20-26.
- DAHLBERG, I. (1978). Teoria do conceito. *Ciência da informação*, 7(2), 101-107.
- DESMET, I. (2002). A análise do sentido em terminologia: teoria e prática da definição terminológica. *Trad/Term*(8), 169-188.
- ECO, U. (1997). *Kant e o ornitorrinco*. Rio de Janeiro: Record.
- GARDIN, J. C. (1966). Éléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*(5), 171-182.
- GENOUVRIER, E., PEYTARD, J. (1974). *Linguística e ensino do português*. Coimbra: Almedina.
- GERMAIN, C. (1981). *La sémantique fonctionnelle*. Paris: PUF.
- GRUBER, T. (1993). Toward principles for the design of ontologies used for knowledge sharing. In N. Guarino & R. Poli (Eds.), *Formal ontology and conceptual analysis and knowledge organization* (pp. 1-22). Padova: Kluwer Academic Publishers.
- GUARINO, N. (1998). *Formal ontology and information systems*. Paper presented at the FOIS'98, Trento, Italy. <http://www.loa.istc.cnr.it/Papers/FOIS98.pdf>

- GUARINO, N., GIARETTA, P. (1995). Ontologies and knowledge bases: towards a terminological clarification *Towards very large knowledge bases* (pp. 25-32). Amsterdam: IOS Press.
- HERNANDO Cuadrado, L. A. (1995). *Introducción a la teoría y estructura del lenguaje*. Madrid : Editorial Verbum.
- INTERNATIONAL STANDARD ORGANIZATION. (2000). *ISO 1087-1: Terminology work - vocabulary theory and application = travaux terminologiques – théorie et application* . Geneve: ISO.
- INTERNATIONAL STANDARD ORGANIZATION. (2000). *ISO 704: Terminology work – principle and methods = Travail terminologique – principes et méthodes* . Geneve: ISO.
- LARA, M. L. G. (1993). Linguagens documentárias, instrumentos de mediação e comunicação. *Revista Brasileira de Biblioteconomia e Documentação*, 26(1/2), 72-80.
- LARA, M. L. G. (1999). *Representação e linguagens documentárias: bases teórico-metodológicas*. (doutorado), Universidade de São Paulo, São Paulo.
- LE GUERN, M. (1989). Sur les relations entre terminologie et lexique. *Meta*, 34(3), 340-343.
- LOPES, E. (1987). *Fundamentos da linguística contemporânea*. São Paulo: Cultrix.
- MOREIRO GONZÁLEZ, J. A. (2011). *Linguagens documentárias e vocabulários semânticos para a web: elementos conceituais*. Salvador: EDUFBA.
- MOURA, M. A. (2009). Informação, ferramentas ontológicas e redes sociais *ad hoc*: a interoperabilidade na construção de tesouros e ontologias. *Informação & Sociedade. Estudos*, 19, 59-74.
- PAIS, C. T. (1994). Percurso gerativo da enunciação: produtividade léxica e discursiva. *Confluência. Boletim do Departamento de Linguística*, 3(n. esp), 162-181.
- PASTOR-SANCHEZ, J. A., MARTINEZ-MENDEZ, F. J., RODRIGUÉZ-MUNÓZ, J. V. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information research*, 14.
- SAGER, J. C. (1990). *A practical course in terminology processing*. Amsterdam: John Benjamins.
- SAGER, J. C. (1993). Prólogo: la terminología, puente entre varios mundos. In M. T. Cabré (Ed.), *La terminología: teoría, metodología, aplicaciones* (pp. 11-17). Barcelona: AntárticaEmpúries.

- SANTOS, C. A. C. M. dos, MAZINI, E. S. (2011). Organização do conhecimento: das classificações e vocabulários controlados às taxonomias e ontologias na web *Tendências contemporâneas na gestão da informação* (p. 125-142): Editora Sociologia e Política.
- TÁLAMO, M. F. G. M. (1987). A definição para a elaboração de glossários. In J. W. Smit (Ed.), *Análise documentária: a análise da síntese* (pp. 87-98). Brasília: CNPq/IBICT.
- TÁLAMO, M. F. G. M. (1997). *Linguagens Documentárias*. São Paulo: APB.
- TÁLAMO, M. F. G. M. (2001). Terminologia e documentação. *TradTerm*(7), 141-151.
- TÁLAMO, M. F. G. M. et al (1992). Contribuição da terminologia para a elaboração de tesouros. *Ciência da Informação*, 21(3), 197-200.
- VATANT, B. (2010). Conversión de los vocabulários bibliotecarios a la Web Semántica, y vice-versa: un viaje en el que todos ganan. Paper presented at the Reunión: 149. Tecnologías de la Información, Catalogación, Clasificación e Indización con la Gestión del Conocimiento. WORLD LIBRARY AND INFORMATION CONGRESS: 76^a Conferencia General y Asamblea de IFLA, Gotemburgo, Suecia. <http://conference.ifla.org/past/ifla76/149-vatant-es.pdf>.
- VOGELL, M. J. M. (2007). *A noção de estrutura linguística e de processo de estruturação e sua influência no conceito e na elaboração de linguagens documentárias*. (mestrado), Universidade de São Paulo, São Paulo.