



XVI Encontro Nacional de Pesquisa em Ciência da Informação (XVI ENANCIB)
ISSN 2177-3688

GT 2 - Organização e Representação do Conhecimento
Comunicação Oral

REFINAMENTO DE RELAÇÕES EM TESAuros: REENGENHARIA DO THESAGRO¹

REFINING THESAURUS RELATIONSHIPS: REENGINEERING OF THESAGRO

Benildes C. M. S. Maculan, UFMG
benildes@gmail.com

Gercina A. B. O. Lima, UFMG
limagercina@gmail.com

Ivo Pierozzi JR., EMBRAPA
ivo.pierozzi@embrapa.br

Leandro H. M. Oliveira, EMBRAPA
leandro.oliveira@embrapa.br

RESUMO: Apresenta resultados de pesquisa que teve como objetivo estudar um modelo de reengenharia de tesauros tradicionais para transformar a sua estrutura em um sistema de conceitos formado por relações semanticamente mais refinadas. Como fundamento teórico foram utilizados conteúdos sobre sistemas de organização do conhecimento (SOCs) e seus relacionamentos semânticos em tesauros, o conceito e a definição que o delimita, a teoria da Valência Verbal e dos papéis *Qualia*. A modelagem foi aplicada no tesouro brasileiro THESAGRO, do domínio da Agricultura, tendo como recorte temático a subárea da Intensificação Agropecuária. A metodologia foi aplicada a partir do modelo proposto em dois artigos: Soergel *et al.* (2004) e Lauser *et al.* (2006). O modelo é composto por três etapas e sua aplicação consistiu na explicitação formal das relações de equivalência e das relações conceituais na estrutura do THESAGRO. Na implantação do modelo foram utilizados dois *softwares*: (1) o E-Termos, que foi o sistema colaborativo utilizado na gestão terminológica, e (2) o Extrator de Termos, que permite a comparação automática da terminologia de diferentes tesauros. Os resultados demonstraram que o modelo aplicado permitiu o refinamento semântico da estrutura conceitual do THESAGRO, demonstrando a viabilidade de seu uso na conversão de um tesouro tradicional em uma estrutura mais formalizada. Concluiu-se que esse refinamento tem papel basilar na organização do conhecimento, podendo facilitar a sua utilização pelo utilizador final, assim como é um elemento

¹ O conteúdo textual deste artigo, os nomes e e-mails foram extraídos dos metadados informados e são de total responsabilidade dos autores do trabalho.

importante para obter a interoperabilidade entre os distintos tesouros e entre diferentes sistemas de recuperação de informação (SRI).

Palavras-chave: Reengenharia de tesouro. Refinamento de relacionamentos semânticos. Sistema de organização do conhecimento. Sistemas de recuperação de informação.

Abstract: Presents researches results that aimed to study a model of reengineering thesauri making it a conceptual structure semantically enriched. We used theoretical bases about: Knowledge Organization Systems (KOSs), the semantic relationships in thesaurus, concept, definitions, the Valency Theory and the Qualia role. The modeling was applied to the Brazilian thesaurus Thesagro from the Agriculture domain more specifically it was applied to the thematic frame of the agriculture intensification. The methodology was applied with the model developed by Soergel et al. (2004) and Lauser et al. (2006). This model consists of three stages and consists of the formal explanation of the equivalence and conceptual relationships in the structure of the THESAGRO thesaurus. In methodological course some tools were used: (a) e-Terms system: for the terminology management; and (b) Terms Extractor, software that allows automatic comparison of the terminology of different thesauri. The results showed the model permits the semantic refining of the THESAGRO structure, and demonstrated the viability of the use of the analyzed model in the reengineering of tradition thesauri in a structure more formal. It was concluded that this refinement has fundamental role in the organization of knowledge and can facilitate their use by the end user, and is an important element for interoperability between different thesauri and between different information retrieval systems (IRS).

Keywords: Reengineering thesauri. Refining semantic relationships. Knowledge Organization Systems. Information Retrieval Systems.

1 INTRODUÇÃO

No âmbito da CI, a atividade de representação da informação conta com instrumentos (sistemas de classificação bibliográfica, taxonomias, tesouros, redes semânticas, ontologias) que, em geral, são utilizados para a padronização dessa representação. Eles têm em comum a tarefa de exercer o controle terminológico de um determinado domínio e inter-relacionar os diferentes conceitos, minimizando os problemas de polissemia (pluralidade de significados) e ambiguidade (imprecisão do significado) encontradas na linguagem natural (LN).

Os diferentes tipos de instrumentos de representação do conhecimento vêm sendo agrupados e designados pelo termo “Sistema de Organização do Conhecimento” (SOC). Carlan (2010) considera que os SOCs são uma “denominação nova para as linguagens documentárias que agregam elementos incorporados nas inovações tecnológicas da era digital” (CARLAN, 2010, p. 29-30). Hodge (2000, p. 9) acrescenta que os SOCs permitem “organizar conteúdos para apoiar a recuperação de itens relevantes, disponibilizados na base de dados de uma biblioteca digital”. Dessa forma, eles têm um importante papel nos processos de representação, classificação, ordenação e armazenamento de recursos informacionais, e visam auxiliar a recuperação e comunicação de informações em ambiente digital.

Soergel (1999) declara que a função de um SOC é ser um dicionário mono, bi ou multilíngue, podendo ser utilizado por um humano ou como base de conhecimento em

aplicações computacionais. Segundo o autor, os principais objetivos dos SOCs são: (1) mapear domínios individuais: servir de instrumento de referência e de mapa semântico capaz de indicar os relacionamentos entre conceitos; (2) dar apoio a professores e alunos: criar materiais didáticos que possam melhorar a comunicação e o aprendizado; (3) apoiar a implantação de projetos de pesquisa e de atividades profissionais; (4) oferecer classificações para diferentes finalidades; (5) fornecer uma base de conhecimento: para apoiar a construção de mecanismos de buscas para a recuperação de informação; (6) auxiliar o desenvolvimento de *software*: permitir a definição de elementos de dados e de hierarquias de objetos a partir de uma base conceitual.

Soergel et al. (2004) estabelecem que, para cumprir de forma eficiente essas funções, é necessário que o SOC estabeleça relações semanticamente mais expressivas, diversificadas e explicitadas. Os autores afirmam que é preciso contextualizar os conceitos e termos, de forma que o sistema possa solicitar a desambiguação, antes de oferecer respostas às consultas dos usuários. Em adição, o sistema poderá indicar também os conceitos relacionados com a consulta realizada, expandindo a busca. Com isso, o sistema pode ficar mais eficaz e o usuário mais satisfeito, uma vez que irá recuperar itens de informação mais relevantes à sua consulta.

Também Svenonius (2000) afirma que a estrutura semântica usada na normalização de SOCs é obtida através do estabelecimento de relações entre os termos e seus referentes. Concordando com Soergel et al. (2004), a autora acredita que em ambiente *web* há maior necessidade de especificar os significados das palavras, assim como suas variações e granularidade. Deste modo, Svenonius (2000) recomenda que a desambiguação não deva ser realizada apenas no nível da determinação do domínio, pois isso pode limitar a compatibilidade entre distintos SOCs. Para tanto, é preciso buscar respaldo do campo da Linguística, Terminologia e Semântica, visando a identificação de relações (sinonímia, antonímia, hiperonímia, hiponímia, ambiguidade lexical, polissemia, homonímia, metáfora e metonímia) que podem auxiliar na formação do conceito e do sistema de conceitos para a construção de SOCs, assim como amparar a elaboração das definições dos conceitos.

Dentre os diferentes tipos de SOCs, o tesouro tradicional é um instrumento de controle terminológico de uma área específica, composto por conceitos relacionados semântica e genericamente entre si. As bases do tesouro tradicional advêm de duas vertentes básicas: (1) da América do Norte, com arranjo alfabético, remotamente proveniente dos estudos de Charles Ammi Cutter, e também influenciado pela indexação coordenada, conhecida como Sistema Unitermo; e (2) da Europa, sobretudo do Reino Unido, que fundamenta-se na categorização e organização de um domínio em classes básicas, tendo sido influenciada pelos princípios da Teoria da Classificação Facetada, que foi aplicada pelo *Classification Research Group* na

elaboração de tesouros facetados (LANCASTER, 1986). Ressaltam-se, também, as contribuições de Jean-Claude Gardin, que apontou os fundamentos da Terminologia como suporte na construção de linguagens documentárias e, entre elas, os tesouros. Na década de 1970, encontram-se os subsídios da Teoria do Conceito, desenvolvida por Dahlberg (1978a), que influencia a criação de tesouros tradicionais com uma base conceitual.

Estruturalmente, a composição de um tesouro é a seguinte: (1) uma terminologia (descritores preferidos e não-preferidos); (2) uma estrutura gramatical (forma de apresentação e de composição dos descritores); (3) uma rede paradigmática (estabelecida *a priori*); (4) uma rede sintagmática (estabelecida *a posteriori*). Juntos, esses elementos determinam a não existência de qualquer descritor que não esteja relacionado à significação de outro descritor na estrutura de um tesouro (SVENONIUS, 2000). Dessa forma, a estrutura do tesouro tradicional permite reunir assuntos correlatos e evitar a dispersão de temas associados. O tesouro tradicional possui três relações básicas: de equivalência (controle de termos em sinonímia e controle de variações linguísticas), hierárquicas (agrupamentos constituídos por conceitos ordenados em níveis diferentes de generalidade e de especificidade) e associativas (atribuição de ligações não-hierárquicas entre conceitos). Contudo, essas três relações básicas, expressas nos tesouros tradicionais, ainda podem ser consideradas insuficientes para representar o conhecimento de um determinado domínio, uma vez que não anunciam, explicitamente, o significado dos relacionamentos, requisito que consideramos importante para a recuperação de informações em ambiente *web*.

As metodologias de construção da estrutura de tesouros vêm evoluindo (MOTTA, 1987; CAMPOS, 1995; CAMPOS; GOMES, 2003; CAMPOS, 2004, entre outros), já com apontamentos para o desdobramento dos três tipos básicos de relações existentes em tesouros tradicionais. Dentro desta perspectiva, existe a proposta de reengenharia de tesouros tradicionais, desenvolvido por Soergel et al. (2004) e Lauser et al. (2006), que proporciona um modelo para o refinamento das suas três relações semânticas básicas, que permite tornar a estrutura do tesouro tradicional semanticamente mais enriquecida. Com base nessa proposta, este artigo apresenta os resultados de uma pesquisa que aplicou esse modelo de reengenharia em um tesouro tradicional para torná-lo um instrumento com uma estrutura semanticamente mais enriquecida, explicitando o significado das relações em sua estrutura conceitual para o usuário, tornando-o um Sistema de Organização do Conhecimento (SOC), mais adequado para ser utilizado na recuperação de conteúdos em ambiente *web*. Para tanto, houve o refinamento das relações que formam o sistema de conceitos do tesouro tradicional selecionado para a remodelagem. No percurso metodológico

foram utilizados fundamentos teóricos que deram respaldo ao desenvolvimento da pesquisa, sobretudo sobre o conceito e sua definição, e sobre os relacionamentos entre conceitos.

2 FUNDAMENTOS TEÓRICOS

A linguagem natural (LN) utilizada em diferentes domínios decorre de fatores histórico-culturais, o que pode trazer problemas de polissemia e sinonímia, por exemplo. Se a LN for utilizada para representar domínios de especialidade, implicará em “ruído na comunicação” e no compartilhamento de conhecimento. Dessa forma, ao modelar um domínio de especialidade é importante a identificação dos conceitos que o representam.

De modo geral, o “conceito” pode ser definido como um conjunto de características, cujos elementos determinam os atributos dos objetos (DIAS; NAVES, 2007). A questão do conceito pode ser observada a partir do pensamento e estudos de autores tais como Ranganathan (1967), Dahlberg (1978a), Bakhtin (1995) e Vygotsky (1998). Em sua Teoria da Classificação Facetada, Ranganathan (1967) apresenta o conceito como uma unidade do pensamento, que é

[...] um corpo de ideias organizado ou sistematizado, cujas extensão e intensão devem ser coerentes com o domínio de interesse e ajustadas à competência intelectual e ao campo especializado de qualquer indivíduo (RANGANATHAN, 1967, p. 82).

Para o autor, o conceito é formado no plano das ideias, quando é aplicada uma análise conceitual, durante a qual acontece a representação mental do objeto, tendo no termo sua expressão verbal (no plano verbal). Nesse momento, é preciso apreender o objeto, sob o ponto de vista lógico, concebendo uma ideia mental, visando a compreensão e a extensão de seu conteúdo. Nesse caso, a compreensão (mais geral) é o conjunto de elementos que constitui uma ideia e a extensão (mais restrita) é o conjunto de elementos à qual a ideia convém.

Por outro lado, Dahlberg (1978a) indica o conceito como unidade do conhecimento, sendo esse conhecimento constituído pelo conjunto de enunciados verdadeiros sobre um objeto e representado por uma forma verbal, que é o termo. Dahlberg (1978a) desenvolveu a Teoria do Conceito tendo por base os estudos analíticos de Aristóteles e os princípios analítico-sintéticos da teoria de Ranganathan. A fase analítica consiste em fatorar o objeto em suas partes, identificando seus elementos individuais (hierarquia das proposições verdadeiras sobre o conceito), que culminam em uma categoria genérica. Já a fase sintética consiste na integração dos elementos sistêmicos, dando origem ao termo que representa o conceito. Nesse sentido, o

[...] conceito é uma unidade do conhecimento, compreendendo afirmações verdadeiras sobre um dado item de referência, representado numa forma verbal [sendo que:] afirmação verdadeira é a componente de um conceito que expressa um atributo do seu item de referência; item de referência é o componente de um conceito para o qual sua afirmação verdadeira e sua forma verbal estão diretamente relacionadas, sendo assim seu ‘referente’; forma verbal

(termo/nome) de um conceito é o componente que resume convenientemente ou sintetiza e representa um conceito com o propósito de designar um conceito em comunicação (DAHLBERG, 1978b, p. 147, grifos da autora citada).

Assim, a autora apresenta o conceito representado pela tríade: (1) referente (objeto a ser conceitualizado); (2) características (todos os enunciados verdadeiros a respeito do referente); (3) forma verbal (termo).

Por seu turno, Bakhtin (1995) aponta que não pode haver comunicação verbal desligada de uma situação concreta, acrescentando que os enunciados ou conceitos produzidos estão sempre em um contexto ou domínio social, no qual são organizados. Já Vygotsky (1998) pondera que quando se trata do pensamento verbal, os conceitos não são inatos, pois são construções culturais, criados como consequência da socialização (contextualizados). Assim, Vygotsky (1998) sugere que as palavras têm uma função mediadora entre os homens e as coisas do mundo e, assim, cada palavra se refere a uma classe de objetos, porque é uma forma de representação do conceito dessa classe. O autor aponta que os elementos necessários para definir um conceito é dependente daqueles encontrados no mundo real e também dos considerados relevantes para esse dado domínio.

Na área da Linguística, Dubois et al. (1997, p. 135) indicam que “dá-se o nome de conceito a toda representação simbólica, de natureza verbal, que tem uma significação geral conveniente a toda uma série de objetos concretos que possuem propriedades comuns”. Para Saussure, fundador da Linguística moderna, o conceito é sinônimo de significado, estando no plano das ideias como uma representação de um objeto (ou realidade social) (NASCIMENTO, 2008). Saussure também considera os elementos sincrônico (descrição sistemática) e diacrônico (descrição histórica) na formação do conceito. Segundo Maimone e Tálamo (2011), esses elementos são importantes na elaboração de tesouros, pois os elementos sincrônicos auxiliam na estruturação das relações semânticas, na determinação dos agrupamentos (conjunção) e das divisões (disjunção) dos signos contidos em dado domínio. Quanto aos elementos diacrônicos (históricos), consideramos que dão apoio à atualização da terminologia e à indicação histórica (por meio de notas de escopo) dos termos que ficaram obsoletos na estrutura dos tesouros.

Já no campo da Terminologia, Barros (2004) afirma que o conceito (ou noção) é essencial na elaboração de produtos terminológicos, uma vez que o conteúdo semântico de um termo é obtido a partir de uma análise nocional ou análise conceitual. Nesse contexto,

Ao proceder com o funcionamento social de uma área especializada do conhecimento, a Terminologia trabalha com o estudo científico dos conceitos e respectivos termos, que constituem um conjunto expressivo e comunicativo, possibilitando a transferência do conhecimento especializado. Neste sentido, os

conceitos não existem isoladamente, mas sempre uns em relação aos outros (MAIMONE; TÁLAMO, 2011, *on-line*).

Dessa forma, percebemos que a Terminologia é importante no processo de comunicação especializada, pois normalizam e padronizam a linguagem utilizada em um domínio de especialidade.

A partir dos estudos de Wüster, com o desenvolvimento da Teoria Geral da Terminologia (TGT), o conceito é entendido como um elemento do pensamento, sendo formado pelas características comuns de um objeto (ou referente). Barros (2006) aponta que

Para a TGT, o termo, enquanto signo linguístico, não se compõe de conteúdo e de expressão indissociavelmente ligados. Para essa corrente teórica, o conceito precede a expressão e essas duas facetas são independentes uma da outra. Devido à tônica normalizadora dessa linha teórica, o conceito deve ser estabelecido de antemão e, só então, deve-se procurar a expressão linguística que mais adequadamente designe o conteúdo terminológico em questão (BARROS, 2006, p. 22).

Consideramos que essa é uma abordagem positivista do conceito, entendido como “algo universal e imutável” (BARROS, 2006, p. 22). Nesse sentido, o conceito é monorreferenciado e visa a obter uma univocidade comunicacional que tem como objetivo a extinção da ambiguidade e polissemia. Essa natureza analítica e lógico-positivista também está presente na Teoria do Conceito de Dahlberg, razão pela qual vem sendo criticada por alguns autores (HJORLAND, 2007; 2009; KOBASHI; FRANCELIN, 2011; MACULAN, 2015).

Cabré (1999) assinala que as críticas à univocidade comunicacional criada com a teoria de Wüster são pertinentes. Porém, ela afirma que, em algumas áreas, principalmente nas quais envolvem eventos fatais, tais como a Medicina e a Aviação, essa característica normalizadora ainda pode ser considerada necessária. A partir de 1990, essas críticas levaram Cabré a desenvolver a Teoria Comunicativa da Terminologia (TCT), na qual o conceito é uma unidade de conhecimento e cujas bases estão fundamentadas em três teorias: (1) do conhecimento (natureza cognitiva): determinação dos conceitos, suas designações e as suas inter-relações; (2) da comunicação (situações comunicativas): circunstâncias de uso; (3) da linguagem (comum e de especialidade): entendimento das unidades léxicas em diferentes contextos de uso (Cabré, 1999). Na TCT, o conceito se materializa em uma situação comunicativa intencional, em um dado contexto de uso, na qual

[...] o termo é uma entidade poliédrica, ou seja, uma unidade linguística (no sentido que lhe deu Ferdinand de Saussure), é uma unidade de comunicação e uma unidade cognitiva. Cabré formula a *Teoria das portas*, segundo a qual os termos podem ser vistos de diferentes ângulos e analisados segundo os modelos das ciências linguísticas, cognitivas e da comunicação, que se considerarem adequados à pesquisa em pauta (BARROS, 2006, p. 23).

Essa abordagem foi adotada na pesquisa que apresentamos neste artigo, pois apresenta uma visão multidimensional e poliédrica do conhecimento de especialidade. Assim, esse conhecimento é entendido como multifacetado, convivendo com distintas correntes e teorias, simultaneamente. Nesse sentido, a relação do termo com o conceito não é de “um para um”, pois também os domínios de especialidade estão sujeitos à variação terminológica e à polissemia da LN. Ademais, Cabré afirma que os conceitos e os termos que os representam “não pertencem a um domínio, mas são usados em um domínio com um valor singularmente específico” (CABRÉ, 1999, p. 124).

Outro ponto que deve ser ressaltado, também abordado por Dahlberg (1978b), é sobre a importância da definição do conceito, que tem o propósito de delimitar o uso do conceito em um dado domínio. A autora garante que as definições são indispensáveis na construção de um tesouro, pois facilitam a estruturação do conhecimento do domínio modelado. Para Campos (2001), a definição auxilia na determinação do termo e estabelece os limites da intensão, a fixação do significado e a posição do conceito no sistema de conceitos. Por sua vez, Lara (2004) aponta que um “dos instrumentos fundamentais para a estruturação dos campos lógico-semânticos dos tesouros é a definição, que permite [...] a determinação do campo de interpretação do termo e sua inserção em um campo temático” (LARA, 2004, p. 91). Portanto, consideramos que é nos fundamentos da Terminologia que a CI encontra respaldo para a elaboração das definições dos conceitos.

Segundo Desmet (2002, p. 171), em geral, “em todos os trabalhos terminológicos e terminográficos, a análise do sentido em Terminologia começa pela definição dos termos”, percurso de análise que pode ser realizado a partir de diferentes orientações. Assim, ressaltamos que as teorias que respaldam a elaboração de definições estão sempre atreladas às distintas concepções de conceito. Dessa forma, o que determina a opção por um tipo de definição depende da vertente teórica adotada, uma vez que, de certa maneira, a definição será sempre arbitrária. Apesar da aceção adotada, Finatto (2002, p. 74) afirma que é o “enunciado que define uma noção, processo ou objeto [pois] é um elemento-chave na constituição e na veiculação do conhecimento [, que expressa] um segmento de relações de significação de uma determinada área do saber”. Portanto, é possível perceber o importante papel da definição no processo comunicacional do conhecimento de um domínio, pois indica a sua terminologia e as inter-relações dos conceitos de especialidade.

2.1 RELACIONAMENTOS SEMÂNTICOS

Em geral, todo tesouro apresenta três tipos básicos de relações semânticas: de equivalência, hierárquica e associativa. Contudo, essas relações abarcam uma pluralidade de conexões entre objetos e, nesse sentido, Green (2001) afirma que qualquer

[...] relacionamento é uma associação entre duas ou mais entidades ou entre duas ou mais classes de entidades. Para especificar um relacionamento, temos de ser capazes, em primeiro lugar, de designar todas as partes vinculadas pelo relacionamento e, em segundo lugar, de especificar a natureza dessa relação (GREEN, 2001, p. 3).

Assim, a autora sugere tanto a determinação das propriedades (reflexivo, transitivo, simétrico) quanto da natureza da relação para se chegar ao seu significado. Para a pesquisadora, as ligações entre conceitos e termos dá origem a redes paradigmáticas (relações de gênero/espécie, todo/parte e de oposição, que são atribuídas *a priori*) e a redes sintagmáticas (relações validadas no contexto, que são atribuídas *a posteriori*). Ela acrescenta que as relações paradigmáticas são um conjunto finito de classes (hierárquico e de equivalência) e que as relações sintagmáticas são uma classe aberta à grande variedade de tipos possíveis de relacionamentos associativos.

As bases para a determinação de relações semânticas advêm de contribuições desde Aristóteles, que apresentou o raciocínio lógico a partir das premissas relacionais: se todo B é A e se todo C é B, todo C é A. Nesse sentido, a lógica não indica a validade do conteúdo, mas a validade da conclusão advinda dos argumentos propostos. Para essa análise, Aristóteles propôs a utilização dos métodos dedutivo (do geral para o particular) e do indutivo (do particular ao geral).

Por sua vez, Farradane (1980) estudou a vinculação de conceitos aos pares sob o aspecto sintático de organização e com base na psicologia do pensamento. Para o autor, as relações podem gerar qualquer combinação de ligações que são capazes de expressar todas as possíveis vinculações existentes entre dois conceitos, em um dado domínio. O pesquisador aponta dois eixos de análise: (1) horizontal: com três estágios de vinculação (consciência, associação temporária e associação fixada); e (2) vertical: com três estágios de vinculação (conceitualização concomitante, conceitualização não-distinta e conceitualização distinta). Dessa forma, o conteúdo semântico de um conceito pode ser expresso pelas relações estabelecidas entre pares de conceitos.

No âmbito dos fundamentos para a construção de tesouros, Dahlberg (1978b) afirma que haverá relação entre dois conceitos toda vez que, na definição de um conceito, houver referência a outro conceito. A autora também recomenda, de forma semelhante a Aristóteles, que seja realizada uma análise híbrida (usando os métodos dedutivo e indutivo) para a identificação dos

conceitos e suas relações. Ademais, ela sugere a aplicação de uma abordagem relacional para a modelagem das relações existentes em um domínio, pois a autora considera ser “fácil de ser construído, reconhecido e utilizado [...] [garantindo] uma maior objetividade” (DAHLBERG, 1979, *on-line*).

Nessa mesma abordagem, Motta (1987) desenvolveu uma proposta de método relacional para ser aplicado na identificação de relações entre conceitos, que “consiste na análise das definições dos conceitos que integrarão o sistema, de forma a identificar suas características e a estruturação desses conceitos, tendo em vista as relações existentes entre eles” (MOTTA, 1987, p. 39). Assim, percebemos que nesse método (analítico-relacional) a identificação das relações é realizada a partir das definições dos conceitos, que também podem ser utilizadas na estruturação do sistema de conceitos.

Toda relação semântica possui um significado que é representado por expressões verbais. Dessa maneira, para a apreensão desse significado é preciso entender o valor semântico que os verbos carregam, conforme salientado por Soergel et al. (2004). Sendo assim, buscamos respaldo em duas teorias: na Teoria da Valência (TV), criada por Francisco S. Borba (1996), e em parte da Teoria do Léxico Gerativo, desenvolvida por Pustejovsky (1995), especificamente da definição dos papéis *Qualia*. Borba (1996, p. xxi) define a TV como um “conjunto de relações estabelecidas entre o verbo e seus argumentos ou constituintes indispensáveis”. Essa teoria foi desenvolvida a partir da combinação de duas outras teorias: a Gramática de Valências (TESNIÈRE, 1966; CHAFE, 1970; VILELA, 1992) e a Gramática de Casos (FILLMORE, 1968, 1969, 1977; ANDERSON, 1971; JACKENDOFF, 1972; COOK, 1979, 1989).

Para Borba, a valência de um verbo representa o número de argumentos que ele necessita preencher para completar seu sentido e pode ser classificada em três níveis: (1) Valência quantitativa (lógica ou lógico-semântica): número de argumentos que um predicado pode ter: avalente (zero), monovalente (um), divalente (dois), trivalente (três) e tetravalente (quatro); (2) Valência qualitativa (sintática ou morfossintática): características dos actantes (relações gramaticais e/ou funcionais), das propriedades morfológicas, das funções sintáticas, das propriedades sintáticas e das classes que preenchem os argumentos; (3) Valência semântica: traços semânticos das categorias (+humano, +animado, +contável), das funções ou dos papéis temáticos (agente, causativo, beneficiário) e das restrições relacionais de coocorrência ou exclusão.

Borba (1996) também determinou uma classificação sintático-semântica para os verbos: (a) verbos de ação: expressam atividade realizada por um sujeito Agente, indicando um FAZER por parte do objeto, que tem controle sobre o ato, podendo ser humano ou não, afetando física

ou psicologicamente um Paciente ou Experimentador; possuindo, no mínimo, um argumento. Exemplo: O pássaro voa hoje; (b) verbos de processos: expressam um evento ou sucessão de eventos, indicando um ACONTECER ou um EXPERIMENTAR, que afeta física ou psicologicamente um Paciente ou Experimentador ou um Beneficiário; têm, pelo menos, um argumento. Exemplos: “A chuva parou” (Paciente); “Ana sente frio” (Experimentador); “Rosa ganhou uma flor” (Beneficiário); (c) verbos de atividade: ou ação-processo ou mudança de estado ou causativos; expressam uma ação ou um processo realizado por um Agente ou uma causação levada a efeito por um Causativo, sempre apontando uma mudança de estado (afetado), condição (afetado), posição (afetado) ou algo que passa a existir (efetuado); têm, no mínimo, dois argumentos: um agente/causativo e outro afetado/efetuado. Exemplos: “José quebrou o vaso” (afetado), “José escreveu um romance” (efetuado); (d) verbos de estado: expressam uma propriedade (estado, condição ou situação), localizada em um sujeito (ou objeto) que é apenas seu suporte, com o verbo indicando um SER/ESTAR/EXISTIR em relação ao sujeito que é o seu Experimentador ou Beneficiário; têm, obrigatoriamente, um argumento que é inativo (não é Agentivo, nem Causativo e nem Paciente). Exemplo: “Fernando tem três filhos”.

Para cada determinada situação é possível determinar uma classe de verbos (ação, processo, atividade e estado). Assim, a ausência de um agente permite que se distinga um verbo de processo de um verbo de ação, ou um verbo de atividade de um verbo de estado (SCHWARZE, 2001, p. 97).

Os verbos podem exprimir diversas operações, que se referem às funções ou propósitos de objetos reais, imaginários ou abstratos. Dada essa multiplicidade de dimensões de significados dos verbos, a estrutura dos papéis *Qualia*, que é parte da Teoria do Léxico Gerativo, pode ser utilizada como passo inicial na identificação do significado semântico de um verbo. A determinação dos papéis *Qualia* para as expressões verbais, que representam as relações semânticas na estrutura de um tesouro, permite identificar os fenômenos de hiperonímia/hiponímia (relação de gênero-espécie) e de holonímia/meronímia (relação todo-parte).

A estrutura *Qualia* é composta pelos seguintes papéis: (a) Formal: generalização de uma operação descrita através de outra operação representada; distingue um objeto em um domínio mais amplo ou geral; (b) Constitutivo: constituição de uma operação descrita, expondo-a por meio de outras operações que são necessárias para efetivá-la; indica uma relação entre um objeto e suas partes constituintes; (c) Agentivo: especificação da entrada de uma operação, na forma de argumentos, representados por objetos reais, imaginários ou abstratos; indica elementos ou fatores que estão envolvidos na origem do objeto ou as causas para o objeto

acontecer, existir ou ocorrer; (d) Téliico: especificação da saída de uma operação, na forma de objetos reais, imaginários ou abstratos; expressa o propósito e a função do objeto.

Tanto o sentido verbal (entre o verbo e o sujeito) quanto o sentido nominal (entre o substantivo e o adjetivo) dos papéis *Qualia* estabelecem as relações existentes entre os termos em um dado contexto comunicativo. Assim, os papéis *Qualia* que interessam ser representados no refinamento das relações semânticas em tesouros são os sentidos verbais das relações que expressam a ligação existente entre dois conceitos, conforme a síntese no Quadro 1.

QUADRO 1 – O sentido verbal dos papéis *Qualia*

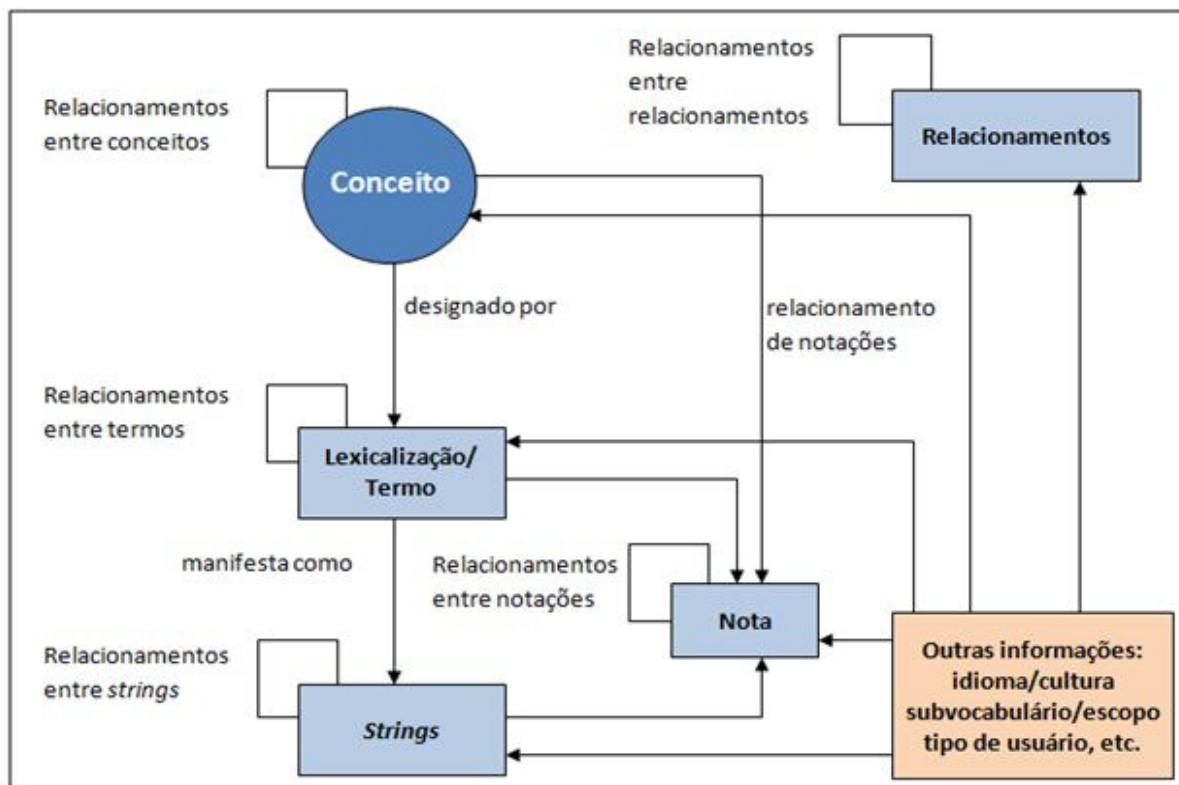
| RELAÇÃO |
|--|
| substância = causa formal (<i>Qualia</i> FORMAL) |
| ter = causa material (<i>Qualia</i> CONSTITUTIVO) qualidade quantidade lugar posição tempo |
| sofrer = causa eficiente (<i>Qualia</i> AGENTIVO) |
| agir = causa final (<i>Qualia</i> TÉLICO) |

Fonte: Com base em PUSTEJOVSKY (1995).

As relações incluídas no *Qualia* Formal são, sobretudo, as relações lexicais de equivalência (sinonímia, antonímia). As relações denominadas *semântico-conceituais* são aquelas classificadas com o *Qualia* Constitutivo, representadas, principalmente, pelas hiperonímias, seguidas pelas meronímias. Já as relações classificadas com o *Qualia* Agente e com o *Qualia* Téliico se referem, respectivamente, a uma ação responsável por uma mudança de estado em um objeto e por uma ação sofrida pelo objeto.

3 METODOLOGIA

O modelo de reengenharia de tesouros tradicionais aplicado foi inicialmente desenvolvido por Dagobert Soergel, e está apresentado em dois artigos: (1) artigo de Soergel et al. (2004); e (2) artigo de Lauser et al. (2006), cuja estrutura conceitual pode ser observada na Figura 1.



Fonte: Traduzido de SORGEL et al. (2004, p. 11).

Analisando a Figura 1 é possível entender que o modelo de reengenharia de tesauro proposto pelos autores tem como principal característica a separação da modelagem em cinco distintos níveis de entidades: conceito, termo (ou lexicalização), *string* (ou variações linguísticas), notas de escopo e relacionamentos. Nessa proposta de modelagem, o nível dos conceitos é considerado o ponto central do modelo, com informações associadas à semântica (significado) conceitual do domínio modelado. No nível dos termos, as informações estão associadas sob o ponto de vista da Linguística ou da etimologia dos termos. Quanto ao nível dos *strings*, as informações englobam as variantes linguísticas ou o nome de uma marca (em um idioma específico, por exemplo). No nível das notas de escopo estão incluídas informações que facilitam o entendimento do usuário sobre a estrutura do tesauro e sobre a terminologia utilizada, sendo que elas podem estar associadas a qualquer um dos níveis de entidade que necessitar de esclarecimento. Finalmente, o nível dos relacionamentos é considerado a espinha dorsal do tesauro, já que permite a construção de uma rede de relações interligando os distintos níveis de entidades.

Ressaltamos outra importante peculiaridade do modelo é a formal distinção entre as entidades termo e *string*: o primeiro designa o conceito e, o segundo, as manifestações desse termo (singular/plural; variação regional). Esse é um elemento novo no modelo, pois essa distinção não é usualmente encontrada na elaboração de tesauros tradicionais.

Na aplicação desse modelo, as relações semânticas podem ocorrer entre entidades de mesmo tipo e entre entidades de tipos diferentes:

FIGURA 2 – Estrutura relacional do modelo de reengenharia

| | |
|--------------------------|--|
| conceito para conceito | é_uma (hierarquia); praga_de |
| termo para termo | é_sinônimo_de; é_tradução_de |
| conceitos para termos | tem_lexicalização (liga os conceitos a seu representante lexical) |
| termo para <i>string</i> | tem_acrônimo; tem_variação_ortográfica; tem_abreviatura (liga os termos com suas formas variantes) |

Fonte: Com base em LAUSER et al. (2006).

Dessa maneira, o estabelecimento de relacionamentos entre os diferentes níveis de entidades torna possível a formação de um sistema semântico-conceitual, importante para a representação e para o entendimento do domínio.

A aplicação dos procedimentos do modelo de reengenharia de tesouro exigiu uma etapa de Planejamento Inicial, no qual houve a descrição do domínio modelado – Agropecuária –, assim como a formação da equipe de trabalho. Essa equipe multidisciplinar foi formada por profissionais de distintas especialidades, a saber: (a) modelador: a proponente da pesquisa, como profissional da informação; (b) terminólogos: as terminólogas Gladis Maria de Barcellos Almeida e Dayse Simon Landim de Souza Kamikawachi, ambas da UFSCar; (c) especialista do domínio: o biólogo Ivo Pierozzi Júnior, da EMBRAPA; (d) tecnólogo: o cientista da computação Leandro Henrique Mendonça de Oliveira, da EMBRAPA. O planejamento também incluiu a seleção do tesouro a ser reformulado pelo modelo de reengenharia: o *Thesaurus* Agrícola Nacional (THESAGRO), único vocabulário controlado do domínio da Agropecuária, em português brasileiro, para a aplicação do de tesouro.

Foram utilizadas duas ferramentas no desenvolvimento da aplicação: o sistema Termos Eletrônicos (e-Termos) e o Extrator de Termos e Estruturas Conceituais Agrícolas Multilíngue (ETECAM). O sistema e-Termos é um ambiente computacional colaborativo *web*, empregado para apoio à criação e ao desenvolvimento de produtos terminológicos (ensino, glossários, vocabulários controlados). O caráter colaborativo do sistema e-Termos é seu maior diferencial, uma vez que torna possível a interação de diferentes profissionais, especialistas e instituições, sem limitação geográfica. Ele é composto por procedimentos automatizados e semiautomatizados, sendo constituído por seis etapas que simplificam e harmonizam o trabalho terminológico. As bases teóricas do sistema e-Termos possuem orientação descritiva, em conformidade com a Teoria Comunicativa da Terminologia (TCT), desenvolvida por Cabré (1999). Por sua vez, o ETECAM foi desenvolvido pela Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) para seu uso exclusivo. Esse *software* foi utilizado para realizar

uma comparação automática entre as terminologias de distintos tesouros, visando a descobrir a existência de termos coincidentes. O ETECAM executa as seguintes tarefas: (a) entrada de uma lista de termos, (b) verificação da existência de termos coincidentes à lista, (c) extração de termos e de seus *clusters* semânticos e (d) extração de termos equivalentes em outros idiomas.

3.1 APLICAÇÃO DAS ETAPAS DO MÉTODO DE REENGENHARIA E RESULTADOS

O modelo de reengenharia de tesouro aplicado é composto por três etapas básicas: (1) definição da estrutura do tesouro, (2) coleta de terminologia e (3) edição do tesouro.

Primeira Etapa – definição da estrutura do tesouro

Essa etapa incluiu a determinação da subárea da Intensificação Agropecuária como recorte temático para a remodelagem e o mapeamento das características do tesouro THESAGRO. O conceito de Intensificação Agropecuária (IA) empregado para a modelagem seguiu a teoria da economista dinamarquesa Ester Boserup (1910-1999), que desenvolveu uma proposta que opõe-se à Teoria Malthusiana, formulada no final do século XVIII, que se baseia no princípio da escassez. De acordo com Boserup (1965), a IA é a relação entre a população (que engloba a densidade populacional: tamanho absoluto e crescimento), o meio ambiente (recursos da terra e outros fatores tais como os climáticos e a qualidade do solo) e a tecnologia (estreita relação com ferramentas, técnicas e insumos utilizados na agricultura). Para Lambin (1994), a teoria de Boserup estabelece que existirá um sistema mais extensivo de plantio sempre que a densidade populacional exigir um aumento da produção de alimentos.

A análise e o mapeamento do THESAGRO permitiu identificar que ele possui cerca de 9.400 descritores, todos identificados por um ID (identificador) numérico. Ademais, a sua estrutura conceitual é representada pelas três relações semânticas básicas de um tesouro tradicional: de equivalência, hierárquicas e associativas. Apesar de ser um tesouro em português, os símbolos utilizados para representar os relacionamentos estão expressos em inglês:

FIGURA 3 – Estrutura relacional do THESAGRO

| | |
|---------------------------|--|
| Relações de equivalência: | USE e USED FOR (UF) |
| Relações hierárquicas: | BROADER TERM (BT) e NARROWER TERM (NT) |
| Relações associativas: | RELATED TERM (RT) |

Fonte: Elaborado pelos autores.

Segundo os resultados apresentados em Maculan (2015), na estrutura conceitual do THESAGRO foram identificadas: (1) cerca de 5.000 relações genéricas (BT) e cerca de 9.000 relações específicas (NT), o que permite considerar o THESAGRO um tesouro com alta especificidade; (2) cerca de 12.000 representações de relações associativas (RT), demonstrando

a intencionalidade dos desenvolvedores de manter o maior número possível de associações entre conceitos que não somente as hierárquicas; (3) cerca de 2.000 descritores ligados apenas por relações associativas (TR), sendo considerados descritores órfãos; (4) 41 notas de escopo (SN, *Scope Note*, em inglês), sempre ligadas a um descritor preferido, utilizadas para explicar o significado desse descritor e/ou para determinar o seu uso; (5) não possui um glossário com definições dos conceitos.

A reengenharia do THESAGRO teve como ponto de partida uma taxonomia já existente, elaborada pelos especialistas do domínio, cuja criação teve por base os termos extraídos de um *corpus* formado por 393 textos completos e 283 resumos de artigos científicos recuperados da base de dados *Web of Science*, utilizando como estratégia de busca a expressão “agricultural intensification”. A taxonomia é composta por 639 conceitos da subárea temática da Intensificação Agropecuária e está estruturada em nove classes básicas: (1) agricultura extensiva; (2) agricultura intensiva; (3) material e métodos; (4) ambiente; (5) agronomia; (6) território e paisagem; (7) socioeconomia; (8) espaço e tempo; (9) instituições. O conjunto dessas classes acolheu e estruturou todos os conceitos que compõem a terminologia da subárea modelada.

O estabelecimento das classes básicas e o agrupamento dos conceitos da área temática da Intensificação Agropecuária foram executados tendo por base uma modelagem dedutiva, segundo os princípios da Teoria da Classificação Facetada, juntamente com os métodos indutivo, sugeridos por Dahlberg (1978a), e o e relacional, recomendado por Dahlberg (1979) e Motta (1987): (a) método dedutivo: como mecanismo de abstração para a análise do campo temático da Intensificação Agropecuária como um todo, a partir dos princípios da teoria de Boserup (1965), que permitiu subdividir o domínio por suas partes componentes, identificando suas classes básicas; (b) método indutivo: representação de cada um dos elementos particulares do contexto modelado, separadamente, a partir da análise dos resultados obtidos no projeto INTAGRO² para a soja e cana de açúcar; (c) método relacional: entendimento da área temática de uma maneira orgânica, envolvendo três elementos: (1) o conceito; (2) a análise conceitual; e (3) a definição do conceito, que permitiu a identificação das relações entre os conceitos e a criação e a estruturação do sistema de conceitos: existência de uma inter-relação entre a dinâmica da população, seu crescimento ou não, o meio ambiente e o uso da tecnologia, com o aumento ou manutenção da produção agrícola em um dado tempo e lugar.

² O projeto INTAGRO foi desenvolvido a participação da EMBRAPA (Campinas/SP) com o principal objetivo de mapear a realidade das atividades de Intensificação Agropecuária em regiões brasileiras, buscando identificar se o crescimento da produção de grãos está associado à expansão da área agricultável ou à adoção de processos produtivos mais intensivos.

Por ser uma terminologia formada por muitos termos, foi necessário estabelecer uma amostra intencional, composta por 30 termos representativos da subárea da Intensificação Agropecuária. Para a seleção da amostra foi aplicada a técnica de amostragem não-aleatória (ou não-probabilística), cujo resultado do processo de seleção permitiu o controle da variabilidade amostral, conforme Figura 4.

| | |
|--|---|
| 1. INTENSIFICAÇÃO AGROPECUÁRIA 2. AGRICULTURA EXTENSIVA 3. AGRICULTURA INTENSIVA | MATERIAL E MÉTODOS 4. cultura 5. sensoriamento remoto 6. sistema de informação geográfica |
| AMBIENTE 7. meio ambiente 8. ambiente físico 9. solo | AGRONOMIA 10. manejo da cultura 11. manejo do solo 12. pousio 13. período de pousio 14. pesticida 15. ciclo da cultura 16. cultura anual 17. pecuária 18. adubo verde 19. produto agropecuário 20. biomassa |
| TERRITÓRIO E PAISAGEM 21. posse da terra 22. escassez de terra 23. cobertura da terra 24. mudança de cobertura da terra | |
| SOCIOECONOMIA 25. densidade demográfica 26. crescimento populacional 27. pressão populacional | ESPAÇO E TEMPO 28. mudança agrícola 29. análise de séries temporais |
| INSTITUIÇÕES 30. Embrapa | |

Fonte: Elaborado pelos autores.

A equipe de trabalho considerou que esse conjunto de 30 termos possui uma representatividade satisfatória e suficiente do recorte temático da Intensificação Agropecuária, uma vez que ele adere aos princípios da teoria de Boserup.

Segunda Etapa – coleta de terminologia

Para a coleta de terminologia foram selecionados a taxonomia da Intensificação Agropecuária (usada como ponto de partida e já citada anteriormente) e outros três tesouros: o Thesagro, o Agrovoc e o *National Agricultural Library* (NAL). Para apoiar a estruturação do novo sistema de conceitos e da elaboração das definições dos conceitos, também foram selecionadas as seguintes fontes de informação: AGROBASE (base referencial gerenciada pela Biblioteca Nacional de Agricultura (BINAGRI); Base de Dados da Pesquisa Agropecuária – BDPA; Dicionário Agrícola CERES – Agronomia Século XXI; Glossário de Termos usados na Agricultura (tradução de termos da área da Agropecuária, em português/inglês e inglês/português); e Glossário de Termos Usados em Atividades Agropecuárias, Florestais e Ciências Ambientais.

A coleta de terminologia foi realizada utilizando a ferramenta ETECAM, com a qual houve a comparação automática entre os 30 termos da amostra e a terminologia existente em cada um dos três tesouros selecionados, a partir de duas listas básicas: (1) contendo os 30 termos da amostra, em português brasileiro, acrescentando-se as expressões desses termos no singular e plural; (2) contendo os 30 termos da amostra, traduzidos para o inglês, acrescentando-se as expressões desses termos no singular e plural, e a sua forma inversa (adjetivo + substantivo). Essa atividade permitiu recuperar os termos coincidentes com as duas listas básicas, existentes nos três tesouros, separadamente, assim como os *clusters* semânticos de cada termo. Depois da comparação automática, foi executada também uma comparação intelectual, com a qual foram adicionados os termos coincidentes não resgatados de forma automática. Os resultados da comparação, automática e intelectual, resultaram em uma estrutura conceitual com cerca de 600 conceitos.

Terceira Etapa – edição do tesouro

A edição, gestão e desenvolvimento da terminologia do THESAGRO foi realizada utilizando o sistema e-Termos. Esta etapa foi composta pelos seguintes procedimentos: compilação da Base Definicional, elaboração de glossário, confecção das Fichas Terminológicas, determinação de Notas de Escopo e construção do sistema de conceitos.

A Base Definicional foi criada a partir da compilação de contextos explicativos e contextos definitórios sobre cada um dos 30 termos da amostra. As definições dos conceitos deram origem ao glossário, que foi elaborado a partir das informações da Base Definicional e com a validação da equipe de trabalho. O glossário foi composto pelos seguintes tipos de definições: (i) definição inicial: com maior exaustividade; (ii) definição do especialista; (iii) definição final: elaborada em conjunto pela equipe de trabalho; (iv) definição do modelador.

O enunciado definitório elaborado pelo modelador, sempre que possível, foram a partir dos seguintes elementos: <termo + conceito genérico mais próximo (conceito superordenado) + diferença específica + natureza do conceito (processo, produto, propriedade, medida, equipamento, método, matéria-prima) + características diversas que identificam o conceito definido (função, origem, uso, composição, forma, cor) + conceitos específicos (conceitos subordinados) + remissivas de equivalência + sigla, acrônimo ou variante>.

Foi elaborada uma Ficha Terminológica para cada termo da amostra. Cada ficha foi composta por 38 campos semânticos, alguns de preenchimento obrigatório e outros opcionais, entre os quais destacamos: definições (do especialista, modelador e final), informações enciclopédicas, Notas de Escopo, termos e conceitos relacionados (de equivalência, de variação linguística, hierárquica e associativa).

As Notas de Escopo (NE) foram criadas sempre que a avaliação da equipe de trabalho julgou necessário inserir informações sobre o significado ou sobre a utilização do descritor. Seguindo essa orientação, foram criadas NE para oito descritores: cultura, ambiente físico, solo, manejo da cultura, pousio, adubo verde, produto agropecuário e biomassa.

O sistema de conceitos foi construído a partir dos conteúdos das definições e dos registros das fichas terminológicas. Para representar as relações semânticas entre termos e conceitos foram utilizados a compilação disponível na documentação do tesauro AGROVOC (*Agrontology*), composto por 170 distintos tipos de relações para a área da Agropecuária (de equivalência, hierárquicas e associativas), incluindo os relacionamentos inversos. Desse total, foram aplicados, na estrutura conceitual da amostra, 44 diferentes tipos de relações, incluídos alguns relacionamentos inversos. Como resultado, foram representadas 286 relações hierárquicas, sendo 225 relações de gênero-espécie (com 52 termos gerais e 173 termos específicos) e 61 relações todo-partes (com 22 termos gerais e 39 termos específicos), o que garantiu a natureza específica da estrutura original do THESAGRO. Ademais, foram representadas 232 relações associativas, demonstrando a complexidade da subárea da Intensificação Agropecuária.

4 CONSIDERAÇÕES FINAIS

Os resultados apresentados neste artigo fazem parte de um convênio firmado entre a Embrapa Informática Agropecuária (Embrapa), Unidade Campinas/SP, a Universidade Federal de Minas Gerais (UFMG) e o Grupo de Pesquisa Protótipo Mapa Hipertextual (MHTX). O desenvolvimento da pesquisa ora apresentada partiu do pressuposto de que as relações semânticas nos tesouros tradicionais ainda são bastante genéricas, indicando a necessidade de estudos sobre o refinamento das mesmas. Dessa forma, a pesquisa priorizou a explicitação dos relacionamentos na estrutura conceitual do tesauro THESAGRO.

Com o refinamento das relações estabelecidas entre os pares de entidades (conceitos, termos, *strings* e notas de escopo), que explicitou o tipo de ligação atribuído entre elas, percebemos que facilitou entendimento do domínio modelado. A abordagem subjacente ao método de reengenharia aplicado permitiu combinar os princípios de construção de tesouros com as novas e emergentes tecnologias da *web* semântica, oferecendo a possibilidade de tornar a estrutura dos tesouros mais formalizada e passível de ser representada em formatos tais como *Simple Knowledge Organization System* (SKOS). Consideramos, também, que a aplicação do método de reengenharia de tesauro trará subsídios para facilitar a interoperabilidade entre diferentes tesouros ou sistemas.

AGRADECIMENTOS

Agradecimento ao apoio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Empresa Brasileira de Pesquisa Agropecuária (Embrapa).

REFERÊNCIAS

- ANDERSON, J. M. **The grammar of case: towards a localist theory**. London: CUP, 1971.
- BAKHTIN, M. **Marxismo e filosofia da linguagem**. 12. ed. São Paulo: Hucitec, 2006.
- BARROS, L. A. **Curso básico de Terminologia**. São. Paulo: EdUSP, 2004, 296p.
- BARROS, L. A. Aspectos epistemológicos e perspectivas científicas da terminologia. **Ciência e Cultura**, São Paulo, v. 58, n. 2, abr./jun. 2006.
- BORBA, F. S. **Uma gramática de valências para o português**. São Paulo: Ática, 1996.
- BOSERUP, E. **The conditions of agricultural growth: the economics of agrarian change under population pressure**. Chicago: Aldine, 1965. 128 p.
- CABRÉ, M. T. **La terminología: representación y comunicación**. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 1999. 369p.
- CAMPOS, M. L. A. Linguagens documentárias: núcleo básico de conhecimento para seu estudo. **R. Esc. Biblioteconomia UFMG**, Belo Horizonte, v. 24, n. 1, p. 52-62, jan./jun. 1995.
- _____. **Linguagem documentária: teorias que fundamentam sua elaboração**. Rio de Janeiro: EUFF, 2001.
- _____. Modelização de domínios de conhecimento: uma investigação de princípios fundamentais. **Ci. Inf.**, Brasília, v. 33, n. 1, p. 22-32, jan./abr. 2004.
- _____ ; GOMES, H. E. Organização de domínios de conhecimento e os princípios rangathianos. **Perspect. Ci. Inf.**, Belo Horizonte, v. 8, n. 2, jul./dez. 2003.
- CARLAN, E. **Sistemas de organização do conhecimento: uma reflexão no contexto da Ciência da Informação**. 2010. 195 f. Dissertação (Mestrado em Ciência da Informação)–Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Universidade de Brasília, Brasília, 2010.
- CHAFE, W. L. **Meaning and the structure of language**. Chicago: University of Chicago, 1970.
- COOK, W. A. S. J. **Case grammar: development of the matrix model (1970-1978)**. Washington, D.C.: Georgetown University, 1979.
- _____. **Case grammar theory model**. Washington, D.C.: Georgetown University, 1989.
- DAHLBERG, I. Teoria do conceito. Tradução Astério Tavares Campos. **Ci. Inf.**, Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978a.
- _____. A referent-oriented, analytical concept theory of Interconcept. **International Classification**, v. 5, n. 3, p. 122-151, 1978b.

_____. Teoria da classificação, ontem e hoje. Tradução de Henry B. Cox. In: CONFERÊNCIA BRASILEIRA DE CLASSIFICAÇÃO BIBLIOGRÁFICA, 1972, Rio de Janeiro. **Anais...** Brasília, IBICT/ABDF, v. 1, p. 352-370, 1979.

DESMET, I. A análise do sentido em terminologia: teoria e prática da definição terminológica. **TradTerm**, v. 8, p. 169-188, 2002.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

DUBOIS, J. et al. **Dicionário de linguística**. São Paulo, Cultrix, 1997.

FARRADANE, J. Relational indexing: part I e part II. **Journal of Information Science**, n.1, p. 267-276; 313-324, 1980.

FILLMORE, C. J. The case for case. In: BACH, E.; HARMS, R.T. (Org.) **Universals in linguistic theory**. New York: Rinehard and Winston, 1968. p. 1-88.

FILMORE, C. J. Types of lexical information. In: KIEFER, F. (Ed.). **Studies in syntax and semantics**. Dordrecht-Holland: D. Reidel, 1969.

FILMORE, C. J. **A semântica na linguística moderna**: o léxico. Tradução de Lúcia M. Lobato. Rio de Janeiro: Francisco Alves, 1977.

FINATTO, M. J. B. O papel da definição de termos técnico-científicos. **Revista da ABRALIN**, v. 1, n. 1, p. 73-97, jul. 2002.

GREEN, R. Overview of relationship in knowledge organization. In: BEAN, C. A.; GREEN, R. (Ed.). **Relationship in knowledge organization**. Dordrecht: Kluwer, 2001. Chapter 1, p. 3-18.

HJORLAND, B. Concept in knowledge organization (KO). **The epistemological lifeboat**, 2008. Disponível em:

<http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/concept_in_knowledge_organizatio.htm>. Acesso em: 12 dez. 2014.

HJORLAND, B. Concept theory. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1519-1536, 2009.

HODGE, G. **Systems of knowledge organization for digital libraries**: beyond traditional authorities files. Washington, DC: Council on Library and Information Resources, 2000.

JACKENDOFF, R. **Semantic interpretation in generative grammar**. Massachusetts: MIT Press, Cambridge, 1972.

KOBASHI, N. Y.; FRANCELIN, M. M. Conceitos, categorias e organização do conhecimento. **Informação e Informação**, Londrina, v. 16, n. 3, p. 1-24, jan./jun. 2011.

LAMBIN, E. F. **Modeling deforestation processes a review**. Luxemburgo: B. Research Report, 1994. 113p. (Trees series)

LANCASTER, F. W. **Vocabulary control for information retrieval**. 2nd ed. Arlington: Information Resources, 1986.

LARA, M. L. G. Diferenças conceituais sobre termos e definições e implicações na organização da linguagem documentária. **Ci. Inf.**, Brasília, v. 33, n. 2, p. 91-96, ago. 2004.

LAUSER, B. et al. From AGROVOC to the Agricultural Ontology Service: Concept Server an OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. **Proceedings...** México: DCMI, 2006.

MACULAN, B. C. M. S. **Estudo e aplicação de metodologia para reengenharia de tesouro: remodelagem do THESAGRO**. 2015. 350 f. Tese (Doutorado em Ciência da Informação)-Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, 2015.

MAIMONE, G. D.; TÁLAMO, M. F. G. M. Linguística e terminologia: contribuições para a elaboração de tesouros em ciência da informação. **DataGramZero - Revista de Ciência da Informação**, v. 12, n. 2, abr. 2011.

MOTTA, D. F. **Método relacional como nova abordagem para a construção de tesouros**. 1987. 89 f. Dissertação (Mestrado em Ciência da Informação)-Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1987.

NASCIMENTO, E. M. F. S. Saussure: o estruturalista antes do termo. **Diálogos Pertinentes** – Revista Científica de Letras, Franca/SP, v. 4, n. 4, p. 259-276, jan./dez. 2008.

PEDROSA, C. E. F. Recursos para uma desambiguação das “frases” veiculadas pelas revistas *Veja* e *Isto É*. In: CONGRESSO NACIONAL DE LINGUÍSTICA E FILOLOGIA, 5., 2001, Rio de Janeiro. **Anais...** Rio de Janeiro: UERJ, Cadernos do CNLF, v. 5, n. 11, 2001.

PUSTEJOVSKY, J. **The generative lexicon**. Cambridge, MA: MIT Press, 1995.

RANGANATHAN, S. R. **Prolegomena to library classification**. Bombay: Asia Publishing House, 1967.

SCHWARZE, C. La sémantique do verbe. In: __. **Introduction à la sémantique lexicale**. Tübingen: Narr, 2001. p. 89-113.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society of Information Science**, v. 50, n. 12, p. 1119-1120, 1999.

_____ et al. Reengineering thesauri for new applications: the AGROVOC example. **Journal of Digital Information**, v. 4, n. 4, 2004.

SVENONIUS, E. **The intellectual foundations of information organization**. Cambridge: The MIT Press, 2000.

TESNIÈRE, L. **Éléments de syntaxe structurale**. 2. ed. Paris: Klincksieck, 1966.

VILELA, M. **Gramática de valências: teoria e aplicação**. Coimbra: Almedina, 1992.

VYGOTSKY, L. S. **Pensamento e linguagem**. 2. ed. São Paulo: Martins Fontes, 1998.

WÜSTER, E. **Introducción a la teoría general de la terminología y a la lexicografía terminológica**. Barcelona: IULA, 1998.