

---

# Relações conceituais como ponto de inflexão entre linguagens documentais, terminologia e ontologias

*Las relaciones conceptuales como punto de inflexión entre los lenguajes documentales, la terminología y las ontologías*

*Conceptual relations as an inflection point among indexing languages, terminology and ontologies*

---

**Walter MOREIRA**

Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília (Brasil)  
walter.moreira@marilia.unesp.br

## Resumen

Se reflexiona sobre la relación de complementariedad entre los lenguajes documentales, la terminología y la ontología en lo que se refiere al estudio de las relaciones conceptuales empleadas con el fin de organizar el conocimiento. A partir de textos seleccionados de la literatura internacional, se analiza la relación entre los lenguajes documentales y diversas soluciones tecnológicas, tomando el SKOS como modelo, así como las relaciones entre los tesauros, las ontologías y la terminología, con el apoyo del concepto teórico de la representación documental. Se aboga por la adopción de los principios de la terminología, como los ha recibido la ciencia de la información, en la construcción de ontologías; y se defiende el aspecto innovador que las ontologías pueden ofrecer al tratamiento de las relaciones conceptuales en los lenguajes documentales.

**Palabras clave:** Relaciones conceptuales. Lenguajes documentales. Terminología. Ontologías.

## Abstract

A reflection is offered on the complementary relations among documentary languages, terminology, and ontologies, in regard to the study of conceptual relations used to organize knowledge. Building on selected texts from the international literature, the relations between documentary languages and different emerging technology solutions are analyzed, taking SKOS as a model; and also the relations among thesauri, ontologies and terminology, with the support of the theoretical concept of documentary representation. The adoption of the principles of terminology, as understood in Information Science, in building ontologies is supported; and the innovative opportunities offered by ontologies to the improvement and processing of conceptual relations in indexing languages is defended.

**Keywords:** Conceptual relations. Documentary languages. Terminology. Ontologies.

## 1. Introdução

O debate recente sobre ontologias reacende as discussões sobre os modelos de organização do conhecimento, notadamente no que se refere à identificação das relações conceituais. Considerados os aspectos de formalização destas relações, as ontologias, para sua realização eficaz, dependem também da produção teórica e metodológica gerada no interior da Ciência da Informação, especificamente no campo que envolve as Linguagens Documentais.

Comparado aos esquemas de classificação bibliográfica, principalmente os anteriores à abordagem facetada, o tesauro representou um avanço considerável na direção de linguagens documentais mais flexíveis, capazes de dar conta da enorme massa documental que começa a se acumular, principalmente, no pós-guerra.

A busca por flexibilidade e os imperativos da telemática, aliás, também têm norteado a evolução dos instrumentos de representação documental e, neste sentido, o campo da Ciência da Informação tem olhado atentamente para as ontologias.

Pretende-se, com o presente trabalho, refletir sobre as relações de complementaridade existentes entre as linguagens documentais, a terminologia e as ontologias no que se refere ao estudo das relações conceituais para fins de organização do conhecimento.

## 2. Linguagens documentais e soluções tecnológicas

Ainda são necessários cuidados para que não se tomem as tecnologias como panaceia. A advertência em si pode soar anacrônica, ou mesmo ressentida, pois este risco sempre exis-

tiu. Ganha, contudo, novo fôlego quando as tecnologias da informação tornam-se sinônimas, por excelência, de tecnologias. Há quatorze anos, García Gutiérrez, em resposta aos que acreditavam que a solução para os problemas decorrentes da organização da informação envolvia apenas esta questão, já advertia:

[...] Muitos atribuem os problemas a pouca idade da teleinformática [...], mas os erros que se referem à recuperação da informação e satisfação do usuário na era do computador multimídia são idênticos aos já conhecidos há trinta anos. Estamos novamente diante da falta de aproveitamento dos recursos da linguística aplicados à documentação (García Gutiérrez, 1998, tradução livre).

A insatisfação com o triunfo do suporte sobre o conteúdo pronuncia-se em vários países como desdobramentos das propostas de Gardin (1966; 1973) na França, e, posteriormente, de García Gutiérrez (1990; 1998), na Espanha e de Hutchins (1975), na Inglaterra. No Brasil destaca-se, nesta linha, a produção do *Grupo Temma* (Cintra; Tálamo; Lara; Kobashi, 1994; Fujita, 1986; 1992; 2003; Guimarães, 1989; 1994; 2000; Kobashi, 1988; 1994; Lara, 1993; 1999; 2009; Tálamo, 1997). Estes pesquisadores, dentre outros, rompem com a abordagem tecnicista e agrupam-se em torno dos estudos sobre as aplicações da linguística e da terminologia na construção e no uso de representações documentais.

As recentes tecnologias da informação oferecem um novo contexto para o conjunto de técnicas desenvolvidas pela Documentação para a organização e recuperação da informação. Destaca-se o *Simple Knowledge Organization System* (SKOS) como exemplo deste tipo de solução.

Desenvolvido pelo consórcio W3C, o SKOS é uma linguagem formal, baseada em *Resource Description Framework* (RDF), e é utilizada para a representação de linguagens documentais, incluindo-se tesouros, sistemas de classificação e listas de cabeçalhos de assuntos. O princípio orientador do SKOS é o fato de que as linguagens documentais compartilham estruturas similares e são também utilizadas em aplicações similares. O propósito, então, do projeto é capturar esta similaridade e explicitá-la, em linguagem de máquina, com o fim de facilitar os processos de interoperabilidade na produção de metadados (Greenberg; Losee; Pérez Agüera; Scherle, White, Willis, 2011).

A julgar pela quantidade de trabalhos que o discutem, o SKOS parece ainda não ter se solidificado como proposta, pelo menos ainda não

teoricamente. Em pesquisa de levantamento, com cobertura cronológica total, realizada nas bases *Information Science & Technology Abstracts* (ISTA), *Library, Information Science & Technology Abstracts* e *Library and Information Science Abstracts* (LISA), foram encontrados apenas 16 trabalhos compreendidos entre 2006 e 2011. Considerando-se os mesmos critérios de busca, pode-se atestar que o tema é praticamente desprezado na literatura científica brasileira da Ciência da Informação, a julgar pela nulidade de retorno da pesquisa realizada na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI).

Os relatos sobre as aplicações do SKOS (Miles; Pérez-Agüera, 2007; Eito-Brun, 2011) reafirmam a necessidade de discutir a eficácia das linguagens documentais, *stricto sensu*, no processo de representação e localização dos recursos de informação disponíveis, principalmente, na *web*.

Da mesma forma, questiona-se a eficácia das ontologias como instrumento no processo de representação do conhecimento sem o conluio das teorias e metodologias criadas para o desenvolvimento de tesouros, voltadas, prioritariamente, para o universo da comunicação científica e tecnológica.

Por sua característica de linguagem dinâmica, o tesouro admite atualizações constantes e incorporação de novos conceitos e relações, mas é, por força de definição, circunscrito a domínios específicos do conhecimento. Não sendo assim, arrisca-se a comprometer a qualidade da rede conceitual que lhe sustenta.

Pode-se, contudo, e este é um ponto de complementaridade, construir ontologias de fundamentação, que descrevem conceitos mais gerais, cujos sentidos sejam equivalentes em domínios mais amplos do conhecimento. Dentre as funções mais importantes de uma ontologia de alto-nível está a de fornecer bases conceituais que suportem a interoperabilidade semântica entre ontologias de aplicação, o que pode beneficiar diretamente na construção de tesouros.

### 3. Tesouros e ontologias

Um documento que define formalmente as relações entre os termos. Essa foi a definição que Berners-Lee, Hendler e Lassila (2001), ofereceram para o termo ontologia em seu artigo basal para quem quer compreender a *web* semântica. A definição, por sua vagueza, ajusta-se também tanto à taxonomia quanto ao tesouro e mesmo a outros dispositivos. Nesta definição, o termo ontologia foi simplificado a ponto do comprometimento terminológico e, se não dá origem, cola-

bora, pela autoridade de quem a pronuncia, para as inúmeras confusões conceituais que geraram uma infinidade de textos cujo fim principal é dirimi-las.

Pela proximidade de estrutura e de finalidade, é natural que surjam comparações entre tesouros e ontologias, busca-se, portanto, discuti-las com o fim de compreender melhor suas especificidades. É comum adotar a via fácil das comparações e julgar um tesouro melhor do que uma ontologia ou vice-versa, conforme o ponto de vista do observador e o lugar de onde fala, mas só se pode considerar um instrumento melhor do que o outro se eles são designados para resolver efetivamente o mesmo problema, o que não é o caso.

Enquanto a produção de inferências lógicas é função explícita das ontologias, por meio de uma estrutura de relações, nos tesouros, estas inferências são menos explícitas ou menos formalizadas; sendo mormente sugeridas.

O tesouro, como atesta o contexto histórico em que surge, ainda que se preste a inferências ou outras instâncias epistemológicas, possui vertente mais prática: auxiliar na representação e na recuperação de documentos (notadamente os científicos). As ontologias surgem em outro contexto, quando se dispõe de maiores recursos de tecnologias da informação e também possuem como foco as relações entre os conceitos, não apenas na instância da comunicação científica e tecnológica, mas também naquela que envolve os recursos digitais de modo mais amplo, como se pode depreender das taxonomias de ontologias apresentadas por Maedche (2002) e Almeida e Bax (2003). Considerando-se o seu alto grau de formalização, é mais viável promover a interoperabilidade em diferentes níveis de tratamento das relações conceituais nas ontologias do que nos tesouros.

Esta mesma amplitude semântica tem provocado debates no interior da Ciência da Informação e da Ciência da Computação em torno da precisão do termo “ontologias”. A popularização deste termo nos meios acadêmicos e nos veículos de comunicação científica é mais recente, comparando-se com o termo “tesouro”. Sendo o conceito de tesouro um conceito mais tradicional na Ciência da Informação (utilizado, pelo menos, desde a década de 1950), é natural que apresente no seu interior maior clareza terminológica do que na Ciência da Computação. Isso é menos pacífico no se refere às ontologias, tomadas algumas vezes, em ambos os campos, como tudo o que modela algum aspecto da realidade.

Estas questões de terminologia são destacadas por conta da necessidade cada vez mais premente de precisão terminológica, de modo especial na comunicação científica.

A tarefa de identificar o núcleo das especializações é cada vez mais inglória e as diversas especializações do conhecimento obrigam a recortes cada vez mais delimitados e concepções variantes de objetos que guardam entre si mais relações de proximidade do que de diferenciação, interferindo nas ações de categorização.

#### 4. Terminologia

Não existe conhecimento sem terminologia. A proposição é de Benveniste (1989, p. 252), para quem “uma ciência só começa a existir ou consegue se impor na medida em que faz existir e em que impõe seus conceitos”. Na especificação mesma de seu objeto, sustenta o autor, é preciso denominá-lo e, desse modo, a denominação é, ao mesmo tempo, “a primeira e última operação de uma ciência”.

Ao afirmar que os problemas relacionados à obtenção de informação nos “novos” sistemas de informação são problemas de linguagem, García Gutiérrez (1998) aproxima a questão terminológica da documentação, cujos processos e produtos serão estudados pela linguística documental.

Considerando-se que as delimitações entre os domínios do conhecimento são cada vez mais tênues, suas terminologias também não são mais tão facilmente demarcadas (como poderia fazer supor a terminologia clássica), mas também por esta mesma razão sua identificação torna-se condição para os diálogos que a ciência e a tecnologia estabelecem entre si e com seus públicos.

Mesmo a distinção entre termo e vocábulo (uma das dicotomias que sustentam a proposta wüsteriana e que já foi considerada mais nítida), também é afetada pela intensidade da comunicação científica. A circulação mais franca do conhecimento científico propiciada pelo avanço das tecnologias de informação tem favorecido cada vez mais a vocabularização de termos antes considerados especializados, os quais deixam o domínio da ciência e da tecnologia e alcançam outras esferas da sociedade. O movimento contrário, da terminologização, é igualmente influenciado.

No cenário dos sistemas de recuperação da informação em que é preciso garantir não só acesso e apropriação, mas igualmente regularidade e precisão nos processos que envolvem a

produção do conhecimento, a precisão terminológica torna-se ela mesma motor do desenvolvimento da ciência normal, conforme a compreensão paradigmática de Kuhn (1978, p. 40):

Quando um cientista pode considerar um paradigma como certo, não tem mais necessidade, nos seus trabalhos mais importantes, de tentar construir seu campo de estudos começando pelos primeiros princípios e justificando o uso de cada conceito introduzido.

Seria ingênuo, se não fosse simplesmente impraticável, tentar engessar o conceito num princípio de univocidade de referência. Esta, aliás, tem sido a maior crítica à terminologia wüsteriana, que expressa conceitos (estáveis, paradigmáticos, universais) e não significados (linguísticos, variáveis).

Krieger (2001, p. 69) critica esta concepção, pois, segundo a autora:

está fundamentada num modelo positivista de ciência que ainda compreende a linguagem das ciências como lugar homogêneo e transparente, dada sua função única de expressar as verdades científicas.

A univocidade comunicacional desconsidera os diferentes sentidos que o discurso pode emprestar aos termos. O problema, contudo, não é anacrônico e nem é verificado exclusivamente em Wüster, pois está materializado também nas normas de elaboração de tesouros que, como observou Lara (2009), sequer mencionam a polissemia.

Ainda que desenhadas prioritariamente para ambientes computacionais, as ontologias não podem prescindir da validação humana, por este motivo estão indelevelmente relacionadas às metodologias propostas pela Terminologia para identificar e expressar as relações entre termos e conceitos.

## 5. Ontologias, informação e representação documental

Os tesouros, na condição de linguagem documental, são instrumentos que visam a orientar as relações entre os conceitos, a atribuição de descritores pelos agentes nos sistemas de informação documental e seu posterior uso para fins de recuperação. Na condição de código comutador, o tesouro lida, necessariamente, com a redução semântica. Entender isto como algo solucionável apenas pela tecnologia pode conduzir (como ocorreu na década de 1970) à falsa concepção de que a linguagem natural, dispondo-se de tecnologias adequadas ao seu tratamento, seria mais eficaz no trato destas questões.

As experiências com o uso da linguagem natural podem até mesmo melhorar as condições de recuperação da informação (embora ainda não se conheçam resultados satisfatórios sobre isso), mas escapam das questões da linguística documental, porque não envolvem a construção de informações documentais.

A produção de informações documentais envolve, entre outras instâncias, a da representação do conhecimento. Subjacente a esta proposição encontra-se uma das questões que fundamentam os estudos sobre organização do conhecimento: pode-se efetivamente representar o conhecimento? Este questionamento tem unido os esforços de alguns dos atuais pesquisadores da Ciência da Computação e da Ciência da Informação, entre outros campos, mas tem origem mais remota, nasce na filosofia, mais precisamente estudada pela Teoria do Conhecimento. Do ponto de vista da Ciência da Informação, entretanto, é preciso completá-la com objetivos pragmáticos, pois é preciso perguntar também pela qualidade e pela viabilidade desta possível representação.

A questão sobre o que integra o universo da representação nas ontologias pode ser resolvida com um axioma: em inteligência artificial o que existe é o que pode ser representado (Gruber, 1995). Nesta perspectiva, parte-se de um modelo para construir uma representação, de modo semelhante ao que ocorre nas linguagens documentais, invariavelmente desenhadas para atender objetivos institucionais. Veja-se, por exemplo, as histórias da Classificação Decimal de Dewey (CDD), da Classificação Decimal Universal (CDU) ou, caso mais notável, da *Library of Congress Classification* (LCC), que não previu em sua gênese a expansão do seu uso.

É sempre importante esclarecer que o conceito de representação do conhecimento não se atrela à ideia de conhecimento como representação do real. A representação no caso dos tesouros e também das ontologias é operacional e escapa do que compreende o conceito de mimese. Defender a relação de fidedignidade entre conteúdo do documento e a expressão utilizada para representá-lo transformaria a linguagem documental em nomenclatura (Lara, 2001).

## 6. Considerações finais

A história recente registra diversos momentos em que certos dispositivos tecnológicos sequer têm tempo para amadurecer e são substituídos. Considera-se que debate sobre as relações de complementaridade entre linguagens documentais e ontologias é útil para que se evitem ideias apressadas de sobreposição ou mesmo que se

despendam energias em retrabalhos. As relações conceituais formam o estofado das linguagens documentais e têm sido estudadas há pelo menos dois séculos.

Acredita-se que este conhecimento acumulado em teorias e metodologias pode contribuir para que se evitem erros primários no desenvolvimento de ontologias, principalmente no que se refere ao mapeamento de campos conceituais e a sua expressão terminológica. Por outro lado, a proposta mais adequada à interoperabilidade das ontologias, que precisam compreender os sistemas de informação de modo mais abrangente do que aquele que o universo mais tradicional das linguagens documentais realizava, oferece mecanismos para que se possam compreender de modo mais formalizado as relações conceituais, o que pode resultar na melhoria da qualidade das relações semânticas eventualmente estabelecidas.

## 7. Referências

- Almeida, M. B.; Bax, M. P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. // *Ciência da Informação*. 32:3 (Set./Dez.) 7-20.
- Benveniste, E. (1989). *Problemas de linguística geral II*. Campinas: Pontes, 1989.
- Berners-Lee, T., Hendler, J.; Lassila, O (2001). The semantic web. // *Scientific American*. (May). <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21> (2007-10-15).
- Cintra, A. M. M.; Tálamo, M. F. G. M.; Lara, M. L. G.; Kobashi, N. Y. (1994). *Para entender as linguagens documentárias*. São Paulo: Polis/APB, 1994.
- Eito-Brun, R. (2011). Integration of SKOS and SRU in a distributed collaboration environment for archival material description. // Gilchrist, A.; Vernau, J. (eds.). *Facets of knowledge organization: proceedings of the ISKO UK Second Biennial Conference, 4th-5th July 2011, London*. Bingley: Emerald, 2011.
- Fujita, M. S. L. (1986). *PRECIS na língua portuguesa: uma experiência com a indexação de artigos de periódicos brasileiros*. Dissertação (Mestrado em Ciências da Comunicação). São Paulo: Universidade de São Paulo, 1986.
- Fujita, M. S. L. (1992). *Linguagem documentária em odontologia: uma aplicação do sistema de indexação PRECIS*. Tese (Doutorado em Ciências da Comunicação). Universidade de São Paulo, 1992.
- Fujita, M. S. L. (2003). *A leitura documentária do indexador: aspectos cognitivos e linguísticos influentes na formação do leitor profissional*. Tese (Livre-docência). Universidade Estadual Paulista Júlio de Mesquita Filho, 2003.
- García Gutiérrez, A. L. (1990). *Estructura lingüística de la documentación: teoría y método*. Murcia: Universidad, Secretariado de Publicaciones, 1990.
- García Gutiérrez, A. L. (1998). *Elementos de lingüística en sistemas de información y documentación*. // *Revista Latina de Comunicación Social*. 7. <http://www.ull.es/publicaciones/latina/a/66ant.htm> (2012-01-10).
- Gardin, J. C. (1966). *Éléments d'un modèle pour la description des lexiques documentaires*. // *Bulletin des Bibliothèques de France*. 5 171-182.
- Gardin, J. C. (1973). *Document analysis and linguistic theory*. // *Journal of Documentation*. 29:2 (Jun.) 137-168.
- Greenberg, J.; Losee, R.; Pérez Agüera, R.; Scherle, R.; White, H.; Willis, C. (2011). *HIVE: Helping interdisciplinary vocabulary engineering*. // *Bulletin of the American Society for Information Science & Technology*. 37:4 (Apr./May) 23-26.
- Gruber, T. R. (1995). *Toward principles for the design of ontologies used for knowledge sharing*. // *International Journal Human-Computer Studies*. 43:5/6. 907-928. [http://www.itee.uq.edu.au/~infs3101/\\_Readings/OntoEng.pdf](http://www.itee.uq.edu.au/~infs3101/_Readings/OntoEng.pdf) (2010-02-28).
- Guimarães, J. A. C. (1989). *Recuperação temática da informação em Direito do Trabalho no Brasil: propostas para uma linguagem de indexação na área*. Dissertação (Mestrado). Universidade de São Paulo, 1989.
- Guimarães, J. A. C. (1994). *Análise documentária em jurisprudência: elementos para uma metodologia de indexação de acórdãos trabalhistas brasileiros*. Tese (Doutorado em Ciências da Comunicação) - Universidade de São Paulo, 1994.
- Guimarães, J. A. C. (2000). *Condensação documentária em legislação e jurisprudência: elementos instrumentais para elaboração de ementas*. Tese (Livre-docência). Universidade Estadual Paulista Júlio de Mesquita Filho, 2000.
- Hutchins, W. J. (1975). *Languages of indexing and classification: a linguistic study of structures and functions*. Herts: Peter Peregrinus, 1975.
- Kobashi, N. Y. (1988). *Política científica e tecnológica: da análise do discurso à análise documentária*. Dissertação (Mestrado em Ciências da Comunicação) - Universidade de São Paulo, 1988.
- Kobashi, N. Y. (1994). *A elaboração de informações documentárias: em busca de uma metodologia*. Tese (Doutorado em Ciências da Comunicação) – Departamento de Biblioteconomia e Documentação, Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.
- Krieger, M. G. (2001). *O termo: questionamentos e configurações*. // Krieger, M. G.; Maciel, A. M. B. (orgs.). *Temas de Terminologia*. Porto Alegre: Editora da UFRGS, 2001. 62-81.
- Kuhn, T. S (1978). *A estrutura das revoluções científicas*. São Paulo: Perspectiva, 1978.
- Lara, M. L. G. (1993). *Representação documentária: em jogo a significação*. Dissertação (Mestrado em Ciências da Comunicação) – Universidade de São Paulo, 1993.
- Lara, M. L. G. (2001). *O unicórnio (o rinoceronte, o ornitorrinco...), a análise documentária e a linguagem documentária*. // *DataGramaZero*. 2:6 (Dez). [http://www.datagramazero.org.br/dez01/Art\\_03.htm](http://www.datagramazero.org.br/dez01/Art_03.htm) (2005-02-22).
- Lara, M. L. G. (2009). *Linguística Documentária: seleção de conceitos*. Tese (Livre-docência) – Universidade de São Paulo, 2009.
- Lara, M. L. G. *Representação e linguagens documentárias: bases teórico-metodológicas*. Tese (Doutorado em Ciências da Comunicação). Universidade de São Paulo, 1999.
- Maedche, A. (2002). *Ontology learning for the semantic web*. Boston: Kluwer Academic Publisher, 2002.
- Miles, A.; Pérez-Agüera, J. R. (2007). *SKOS: Simple Knowledge Organisation for the Web*. // *Cataloging & Classification Quarterly*. 43:3/4. 69-83.
- Tálamo, M. F. G. M. (1997). *Linguagem Documentária*. São Paulo: Associação Paulista de Bibliotecários, 1997.

Enviado: 2012-04-14. Aceptado: 2012-06-23.

