

AVALIAÇÃO DO ESPECTRO SEMÂNTICO DE INSTRUMENTOS PARA ORGANIZAÇÃO DA INFORMAÇÃO

EVALUATION OF THE SEMANTIC SPECTRUM FOR INFORMATION ORGANIZATION INSTRUMENTS

Maurício Barcellos Almeida

Professor adjunto da UFMG, Depto de Teoria e Gestão da Informação
mba@eci.ufmg.br

Renato Rocha Souza

Professor adjunto da UFMG - Depto de Organização e Uso da Informação
rsouza@eci.ufmg.br

Resumo

O termo "semântica" tem sido amplamente utilizado nos últimos anos em diversas áreas de pesquisa e, em particular, em áreas relacionadas à tecnologia da informação. Um dos motivadores de tal apropriação é a visão da *Web Semântica*, originada na expansão da *web* e nas limitações dos instrumentos de busca baseados em sintaxe. Cabe questionar, entretanto, qual é esta "semântica" da qual trata a Web Semântica, uma vez que o estudo do assunto é complexo e controverso. O presente artigo se insere nesse contexto, analisando o que é semântica, discutindo-a no âmbito da tecnologia da informação e propondo alternativas interpretativas. Avalia-se criticamente um *espectro*, o qual propõe a ordenação de instrumentos (modelos, linguagens, estruturas taxonômicas, etc.) de acordo com uma escala semântica. Espera-se tornar claro o uso do termo em contextos diversos, bem como propor um novo espectro a partir das considerações do presente artigo.

Palavras-chave: Web Semântica. Ontologias. Semântica. Representação do conhecimento.

1 INTRODUÇÃO

Segundo as definições encontradas na literatura, o termo "semântica", em geral, trata-se do estudo do significado ou de ciência que estuda a significação. Segundo o Dicionário Aurélio, semântica vem "do grego. *semantiké* i.e., *téchne semantiké*, a arte da significação" (AURÉLIO, 1999, p. 1832). Nos últimos anos, o termo tem sido utilizado em campos de pesquisa diversos, e suas acepções variam de acordo com concepções e os construtos variados. Tradicionalmente estudado na Linguística e na Filosofia, ganhou popularidade em áreas relacionadas à tecnologia da informação, principalmente em função da visão conhecida como *Web Semântica* (BERNERS-LEE, 1998).



Esta obra está licenciada sob uma [Licença Creative Commons](https://creativecommons.org/licenses/by/4.0/).
DOI 10.5007/1518-2924.2011v16n31p25

O principal objetivo da Web Semântica (WS) é promover melhorias nos processos de representação e recuperação da informação na *web*, a qual, desde 1990, é caracterizada pelo uso de linguagens de marcação que objetivam a apresentação e a leitura por pessoas e por mecanismos de busca baseados em algoritmos com orientação à sintaxe. A motivação a melhorar os processos de recuperação na *web* é resultado da expansão de suas funcionalidades e do aumento significativo no volume de dados disponíveis.

Segundo Souza e Alvarenga (2004), a WS pretende propor melhorias nas linguagens de marcação utilizadas para confecção de páginas *web*, de modo a beneficiar a interação entre tais páginas e os sistemas. Advogam que, com o uso da semântica, se pode aumentar a possibilidade de associações dos documentos a seus significados por meio dos metadados descritivos. A questão do significado de semântica é assim fundamental na proposta da WS. O conjunto de tecnologias inserido nessa proposta busca resolver problemas relacionados ao fato de que “computadores não tem uma forma confiável de processar a semântica”¹ (BERNERS-LEE, 2001, p.1).

Dada a complexidade inerente ao estudo da Semântica e à diversidade de abordagens ao assunto apresentadas ao longo dos anos, nem sempre excludentes e sim complementares, cabe avaliar que tipo de semântica é considerado na visão da WS. O presente artigo se insere nesse contexto. Após um breve levantamento das abordagens existentes no estudo da Semântica em sua área de origem, a Linguística, apresentam-se conotações para o termo usadas na WS e em áreas correlatas. Destaca-se uma proposta de espectro (OBRST, 2004; DACONTA, 2005), a qual faz uso de uma escala semântica para ordenar instrumentos diversos, como taxonomias, tesouros, bancos de dados, dentre outros; além daqueles utilizados na WS, como as ontologias, linguagens de representação, etc. Pretende-se avaliar criticamente essa proposta, esclarecer a acepção de semântica em cada caso e propor um novo espectro.

O restante do artigo está organizado como segue. A seção dois discorre sobre o estudo da Semântica em sua acepção mais comum, ou seja, a semântica da linguagem natural, desde os primeiros estudos até as abordagens ditas contemporâneas, descrevendo as principais linhas de pesquisa na Linguística. A seção três apresenta a semântica proposta para uso através de computadores e sistemas na *web*, em particular, a visão da WS. A seção quatro discute a escala semântica do espectro avaliado e propõe um novo espectro, consolidado através das

¹ [...] computers have no reliable way to process the semantics [...]

considerações do presente artigo. Finalmente, a seção cinco apresenta considerações finais e indica possibilidades para trabalhos futuros.

2 SEMÂNTICA NA LINGUÍSTICA

Parece consenso que semântica diz respeito ao estudo da significação. A dificuldade reside em definir o que é significação. O conceito de significado não é claro e nem consensual, aparecendo em variações que vão desde o *realismo*², professado pelos que acreditam que a língua se superpõe como uma nomenclatura a um mundo no qual as coisas existem objetivamente, até formas de *relativismo*³, segundo as quais é a estrutura da língua que determina a capacidade humana de perceber o mundo.

Ilari e Geraldi (1985, p.6) advertem que “[...] a semântica é um domínio de investigação de limites movediços; semanticistas de diferentes escolas utilizam conceitos e jargões sem medida comum, explorando em suas análises fenômenos cujas relações não são sempre claras [...]”. Apesar da dificuldade em se definir exatamente o que é semântica, Cançado (2005) delimita os principais aspectos que a caracterizam: i) a *composicionalidade*; ii) as *propriedades semânticas*; iii) a *referência e a representação*.

A *composicionalidade* diz respeito à capacidade da teoria semântica de atribuir significado às palavras e às sentenças, de acordo a linguagem. As *propriedades semânticas* dizem respeito à caracterização das relações sistemáticas entre palavras e sentenças de uma língua. As noções de *referência e representação* estão relacionadas à natureza do significado: a noção de referência é explicada como a conexão entre as expressões linguísticas e o mundo; a noção de representação diz respeito à associação do significado a uma representação mental.

Esses aspectos vão nortear as principais vertentes de estudo da Semântica. A seção 2.1 apresenta um breve histórico das abordagens semânticas, e a seção 2.2 enfatiza a *Semântica Formal*, em função de sua importância para os objetivos do presente artigo.

² O realismo filosófico corresponde a presunção de que a realidade é independente de esquemas conceituais, práticas linguísticas e crenças humanas.

³ No relativismo filosófico, acredita-se que a verdade de uma proposição depende de quem a interpreta, em função da variedade de possibilidades morais e culturais.

2.1 Breve histórico e principais abordagens

A presente seção descreve abordagens inerentes à Semântica, desde as primeiras iniciativas até as contemporâneas. Cabe destacar que um levantamento exaustivo não é o nosso objetivo, e este nem seria possível, dada a complexidade e o volume de material sobre o assunto. Dessa forma, muitos autores e abordagens não são citados. Espera-se, porém, apresentar alguns dos principais aspectos e possibilidades do estudo da Semântica, de forma a contextualizar o assunto de acordo com os objetivos do artigo.

Uma das primeiras referências ao estudo do significado é o diálogo escrito por *Platão*, do qual participam *Sócrates*, *Crátilo* e *Hermónegos*⁴, cujo assunto é a origem dos nomes. Nesse contexto, o termo *nomes* diz respeito a: i) um termo geral para palavras; ii) substantivos ou adjetivos; iii) nomes próprios (SEDLEY, 2006). Os filósofos gregos discutiam se os nomes são convencionais ou naturais, ou seja, se a linguagem é um sistema arbitrário de signos ou se as palavras têm relações intrínsecas com os objetos que representam.

Segundo Sowa (2000), o estudo do significado, já estabelecido por *Aristóteles* em sua obra *Categorias*, tem sido estudado ao longo dos séculos sobre o rótulo de lógica ou de teoria dos signos. Apenas no século XIX, *Bréal*⁵ sugere o estudo das leis que governam o significado, atribuindo-lhe o nome *semântica*, termo derivado do verbo grego. Nesse contexto, a Semântica era uma disciplina diacrônica, ou seja, de caráter histórico. Segundo Noth (1995), *Saussure*⁶ fez a distinção entre a linguística *diacrônica*, estudada através das mudanças do significado em uma língua ao longo do tempo, e a *sincrônica*, a qual diz respeito aos fenômenos em um ponto específico do tempo. A Semântica moderna tem orientação *sincrônica*.

Ainda no século XIX, temos como um importante marco no estudo do significado a distinção concebida por *Frege*⁷ sobre dois elementos presentes no significado de uma sentença: o *sentido* e a *referência*. Segundo Dummet (1981), os termos originais do alemão utilizados por *Frege* eram *sinn*, para *sentido*; e *bedeutung*, um substantivo para o verbo *bedeuten*, traduzido como *significação* ou *significado*. O *sentido* diz respeito apenas ao que é importante para determinar a verdade ou falsidade de uma sentença. Qualquer outra característica que não afete essa determinação pertence à sua *referência*. Segundo Morris (2007), a *Teoria da*

⁴ *Crátilo*. Disponível em: <<http://www.gutenberg.org/etext/1616>>. Acesso em: 20 ago. 2008.

⁵ *Michel J. A. Bréal*, filologista francês (1832-1915), em sua obra *Essai de Sémantique* (1897).

⁶ *Ferdinand de Saussure*, linguísta suíço (1857 - 1913).

⁷ *Friedrich L. G. Frege*, matemático e filósofo (1848 - 1925).

Referência de Frege explica as operações básicas sobre categorias de expressões linguísticas ao atribuir-lhes coisas as quais se referem, ou seja, seus *referentes*.

Nirenberg e Raskin (2004) observam a existência de diferentes soluções ao longo dos anos para a tarefa de representar o significado e destacam: i) a *análise componencial*; ii) os *campos semânticos*; iii) o *uso de metalinguagem*. Tais abordagens surgem no período entre o estabelecimento da Teoria da Referência e as iniciativas ditas contemporâneas.

A abordagem conhecida como *análise componencial* é baseada em conjuntos de características semânticas, denominadas *componentes*, as quais são usadas para descrever termos afins em diferentes sociedades. A combinação dos componentes permite obter o significado de termos comuns à maioria das culturas. Exemplos de autores representativos dessa abordagem são *Kroeber*⁸ e *Goodenough*⁹.

A abordagem conhecida como *campos semânticos* consiste na criação de grupos de palavras com significados relacionados. As palavras são localizadas próximas a outras em função de uma combinação de fatores intuitivos, incluindo, dentre outros, relações paradigmáticas e relações sintagmáticas. Exemplos de autores representativos dessa abordagem são *Trier*¹⁰ e *Weisgerber*¹¹.

A abordagem dos *campos semânticos* explora a semântica sem a utilização de uma metalinguagem. A importância do uso de uma metalinguagem é enfatizada por Cançado (2005, p.19): “a adoção de uma metalinguagem diferente da própria língua elimina prováveis distúrbios na análise linguística”. Uma extensão da *análise componencial* para cobrir todo um léxico pode gerar uma metalinguagem para descrever o significado de palavras. O *uso de metalinguagem* para a análise componencial, baseada em Lógica de 1ª Ordem (LPO), resultou na atribuição de entidades lógicas (predicados, argumentos, funções) aos componentes. Exemplos de autores pioneiros nessa abordagem são *Greenberg*¹² e *Bar-Hillel*¹³.

Dentre as abordagens contemporâneas, identificam-se três vertentes principais (CANÇADO, 2005): i) a abordagem *pragmática*; ii) a abordagem *mentalista*; iii) a abordagem *referencial*.

⁸ *Alfred L. Kroeber*, antropólogo norte-americano (1876-1960).

⁹ *Ward H. Goodenough*, antropólogo norte-americano (1919-?).

¹⁰ *Jost Trier*, linguísta alemão (1894 - 1970).

¹¹ *Leo Weisgerber*, linguísta alemão (1899 - 1985).

¹² *Joseph H. Greenberg*, linguísta e antropólogo norte-americano (1915 - 2001).

¹³ *Yehoshua Bar-Hillel*, matemático e linguísta (1915 - 1975).

A *abordagem pragmática* é representada pela *semântica argumentativa* de Ducrot ¹⁴ e pela *Teoria dos Atos da Fala* de Austin ¹⁵. A *semântica argumentativa* é uma abordagem originada na França, a qual advoga que as sentenças são pronunciadas como parte de um discurso em que um orador tenta convencer seu interlocutor a respeito de uma hipótese. A linguagem não é usada para falar algo sobre o mundo, mas para convencer o ouvinte a entrar em um jogo argumentativo. A *Teoria dos Atos da Fala* enfatiza a ideia de que parte do sentido de uma sentença tem função social.

A *abordagem mentalista* afirma que o sentido acontece em um nível intermediário entre o mundo e as palavras, que corresponde ao nível de representação mental. As principais iniciativas dentro dessa abordagem são: i) a *semântica cognitiva*, representada, por exemplo, por Lakoff ¹⁶; ii) a *semântica representacional*, representada por Jackendoff ¹⁷; e iii) a *semântica lexical*, representada, por exemplo, por Levin ¹⁸.

A *semântica cognitiva* pressupõe a relação da linguagem com uma representação através de esquemas e imagens, mapeando domínios conceituais distintos, onde o uso da metáfora corresponde a um processo cognitivo fundamental. A *semântica representacional* tem compromisso com a forma das representações mentais internas que constituem a estrutura conceitual e com as relações formais entre esse nível e outros níveis de representação (sintático, fonológico, visual, etc.). A *semântica lexical* explora a noção de *papéis temáticos*, os quais são funções semânticas que os argumentos de um verbo (sujeito e complementos) desempenham em uma sentença. Outras iniciativas da semântica lexical propõem estudar o léxico para viabilizar a aplicação da linguística em práticas computacionais.

A *abordagem referencial* tem origens no estudo da Lógica e da Filosofia da Linguagem, em particular, na distinção proposta por Frege. Segundo Oliveira (2001), Frege utilizava como exemplo as expressões *Estrela da Manhã* e *Estrela da Tarde*, as quais tem significado diferente, mas se referem à mesma entidade, o planeta Vênus. Sowa (2000) explica que Frege atribuiu as entidades de sua teoria aos três vértices de um triângulo, denominando-os *símbolo*, *sentido* e *referência*. Essa mesma distinção foi introduzida na Semântica linguística por Ogden et al (1972, p. 32) através do *triângulo do significado*: “[...] um diagrama em que os

¹⁴ Oswald Ducrot, linguista e filósofo francês (1930 - ?)

¹⁵ John L. Austin, filósofo inglês (1911- 1960).

¹⁶ George P. Lakoff, pesquisador em linguística cognitiva (1941 - ?).

¹⁷ Ray Jackendoff, linguista norte-americano (1945 - ?).

¹⁸ Beth Levin, professora do Depto de Linguística da *Stanford University*.

três fatores envolvidos, sempre que se declara que se entende algo, são colocados no vértice do triângulo, sendo as relações existentes entre eles representadas pelos lados [...]"

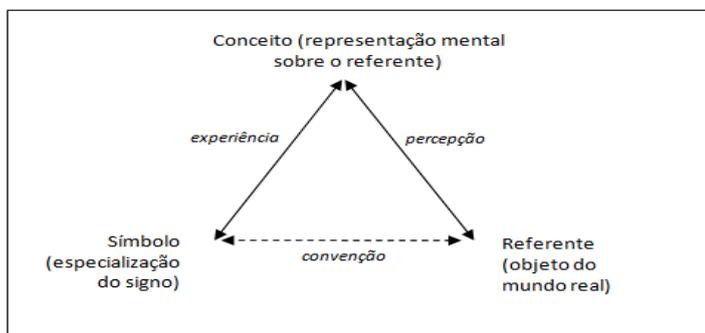


Figura 1: triângulo do significado
Fonte: adaptado de Ogden et al (1972)

No Triângulo do Significado (FIG. 1), as arestas relacionam as entidades objeto e signo, signo e conceito, objeto e conceito. Convencionou-se, culturalmente, identificar um objeto do mundo real por um signo (aresta convenção); tal signo, quando percebido por uma pessoa, gera um conceito em sua mente (aresta percepção); o conceito é formado na mente da pessoa a partir de sua vivência no mundo, por ela ter reconhecido no objeto outro objeto que já conhecia (aresta experiência). A aresta convenção é apresentada como uma linha pontilhada para explicar que um símbolo não se conecta diretamente como o seu referente (objeto do mundo), mas através de uma representação mental desse elemento no mundo.

Dahlberg (1978) apresenta o *Triângulo do Conceito*, diferente do Triângulo do Significado de Ogden et al (1972), no qual o conceito é um dos vértices do triângulo e o seu significado pode assumir várias possibilidades. Em tríade, Dahlberg (1978) apresenta o conceito como a soma dos enunciados verdadeiros e essenciais sobre um referente, e o termo / léxico como a forma comunicável e representável do conceito. Nesta representação, percebe-se que o significado de um termo, ou seja, de uma forma verbal é decorrente da interpretação de um conjunto de enunciados verdadeiros (características) que se podem atribuir a um referente, ou seja, a um objeto, fenômeno, processo ou entidade.

Uma importante iniciativa, no âmbito da abordagem referencial, é aquela conhecida como *Semântica Formal*. Exemplos de autores importantes que atuaram nessa área são Carnap¹⁹, Tarski²⁰, Kripke²¹ e Montague²². Em função de sua importância para os objetivos do presente trabalho, a seção 2.2 é dedicada a descrever suas principais características.

¹⁹ Rudolf Carnap, filósofo alemão (1891 - 1970).

²⁰ Alfred Tarski, lógico e matemático polonês (1901-1983).

²¹ Saul Aaron Kripke, lógico e filósofo norte-americano (1940 - ?).

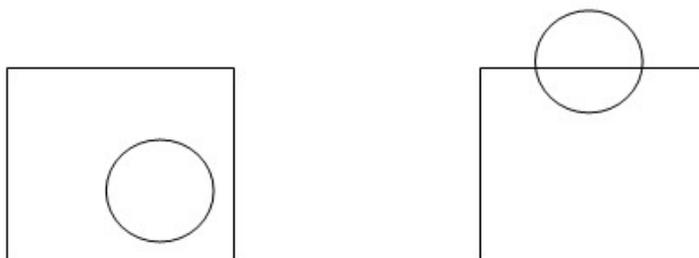
2.2 A Semântica Formal

Os principais aspectos que permeiam o estudo da Semântica Formal ao longo de sua evolução são: i) ênfase no *princípio da composicionalidade*; ii) uso da *condição de verdade* para explicar o significado; iii) concepção de *modelos em semântica* (OLIVEIRA, 2001). Essas três linhas de estudo são descritas no restante da presente seção.

O *princípio da composicionalidade* estabelece que o significado das sentenças depende do significado das palavras que as compõem. Assim, o significado do todo é função do significado das partes e da combinação sintática entre elas. Para deduzir o significado de uma sentença, é preciso conhecer o significado de suas partes, bem como as regras que definem como combinar tais partes.

O uso da *condição de verdade* para explicar o significado diz respeito a determinar em que condições tal sentença é verdadeira. Segundo Morris (2007), essa visão tem raízes na *Teoria da Verdade Semântica de Tarski*²³: um conjunto de regras para aplicação do conceito de verdade a sentenças de sistemas formais baseados em lógica, sem o risco da ambiguidade inerente à linguagem natural. Uma vez que a noção de verdade é central na lógica, a preocupação reside em explicar como as sentenças dos sistemas formais são consideradas verdadeiras ou falsas. Nesse contexto, saber o significado de uma sentença equivale a conhecer suas *condições-verdade*, o que não é o mesmo que saber o seu *valor-verdade*, ou seja, se o fato é verdadeiro ou falso.

Para esclarecer o que envolve conhecer as condições para que uma sentença seja verdadeira ou falsa, Portner (2005) apresenta como exemplo a sentença: *o círculo está dentro do quadrado*. Segundo o autor, conhecendo a sentença, uma simples avaliação da FIG.2 permite dizer em quais condições a sentença é verdadeira. Em Semântica Formal, tais situações são denominadas *situações possíveis* (*possible situations*) ou *estados das coisas* (*state of affairs*).



²² Richard Merett Montague, matemático e filósofo norte-americano (1930 - 1971).

²³ A *Semântica da Verdade Condicional (Truth-Conditional Theory)* usa o mesmo princípio lógico de Tarski, porém aplicado a linguagem natural. Trata-se de teoria desenvolvida por *Donald Herbert Davidson*, filósofo norte-americano (1917 - 2003).

(a) a sentença é verdadeira

(b) a sentença é falsa

Figura 2: condições verdade para a sentença *o círculo está dentro do quadrado*.

Fonte: adaptado de Portner (2005)

A existência de um conjunto de cenários, como os mostrados na FIG. 2, permite denominar um *conjunto verdadeiro* e um *conjunto falso*. Tais conjuntos são denominados *mundos possíveis* (*possible worlds*). Nesse contexto, o termo *situação* diz respeito a um cenário incompleto, uma parte do universo limitado no espaço e no tempo. Já o termo *mundo* é usado quando as pessoas têm uma ideia completa do que o mundo deve ser, com todos os seus componentes e habitantes, os quais não se alteram ao longo do tempo.

Segundo Nirenberg e Raskin (2004), o uso da lógica no âmbito da Semântica Formal permitiu a aplicação da *noção de proposição* ao estudo da sentença. O significado de uma sentença passa assim a corresponder a uma proposição, o que corresponde a um deslocamento no triângulo do significado, do nível da palavra para o nível da sentença (FIG. 3a). Os lógicos renomearam os vértices do triângulo de acordo com termos dos seus sistemas (FIG. 3b). Assim, nenhum dos elementos do triângulo (b) se relaciona diretamente com a linguagem natural, uma vez que uma proposição é resultado da tradução de uma sentença declarativa em uma metalinguagem baseada em lógica.

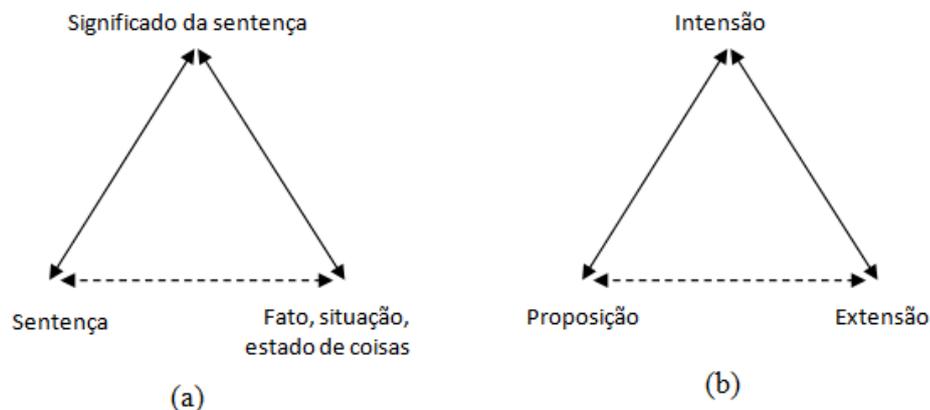


Figura 3: triângulo do significado (a) para sentenças e (b) usando termos lógicos

Fonte: Adaptado de Nirenberg e Raskin (2004)

Na abordagem de *modelos em semântica*, constrói-se um sistema mais simples em relação a um sistema complexo que se deseja estudar, o qual serve de modelo para o estudo. Em seguida, constrói-se uma teoria para o modelo, em geral uma *teoria lógica*. Se os resultados são razoavelmente positivos em relação ao sistema complexo, diz-se que o sistema simples é um bom modelo, caso contrário, ele é abandonado. Estudar a semântica das linguagens formais auxilia o estudo das linguagens naturais, pois ambas tem características comuns e a linguagem formal é mais simples do que a linguagem natural. Portanto, um procedimento

usual dos semanticistas formais tem sido propor modelos simples em linguagens formais e em seguida interpretar através deles o maior número possível de sentenças das linguagens naturais.

Segundo Neto (2003), um modelo simples para interpretar uma linguagem é a *Teoria dos Conjuntos*, esse procedimento consiste em associar expressões a objetos do mundo. Assim, nenhum modelo para interpretar a realidade pode prescindir de objetos. Nesse caso, os objetos que compõem os modelos de interpretação da linguagem natural são denominados *objetos teóricos do modelo*.

Portner (2005) explica que a *Semântica Teórica dos Modelos (Model-Theoretic Semantics)* – uma instância da *Teoria da Verdade Semântica* de Tarski – considera elementos que fornecem modelos da realidade úteis para propósitos semânticos. Tais modelos são compostos por: i) um conjunto de mundos possíveis; ii) um conjunto de indivíduos; iii) um conjunto de períodos de tempo; iv) uma descrição de quais indivíduos habitam quais mundos; v) uma descrição de quais períodos são anteriores a quais outros períodos.

Além dos elementos citados, existe uma *interpretação* que descreve o significado de cada palavra, frase ou sentença em termos de um modelo específico. Dado um modelo, o significado de um nome deve pertencer a um indivíduo daquele modelo e o significado de uma sentença deve ser um conjunto de mundos possíveis naquele modelo. Uma interpretação semântica pode ou não dizer algo sobre como é a realidade. Na verdade, declarações são feitas sobre o modelo e então se avalia como tal modelo se conecta à realidade. Modelos que representam de forma precisa a realidade são denominados *modelos pretendidos (intended models)*. A interpretação que conecta partes de linguagem aos modelos pretendidos, da mesma forma que a linguagem natural se liga a realidade, é denominada *interpretação pretendida (intended interpretation)*.

Finalmente, cabe citar que importância da perspectiva formal reside no fato de que outras teorias semânticas estabeleceram-se a partir da semântica formal, ao tentar solucionar problemas não resolvidos por esse modelo ou propondo modelos alternativos. A FIG. 4 apresenta um quadro sinótico das abordagens semânticas, conforme apresentadas ao longo da seção dois.

Evolução	Abordagem	Breve descrição
Antiguidade	Platônica	Origem dos nomes
	Aristotélica	Categorias
Século XIX	Diacrônica	Semântica como disciplina
	Sincrônica	Origem da orientação linguística moderna
	Referencial	Sentido, significado, referência
Século XIX e XX	Campos semânticos	Palavras com significados relacionados
	Análise componencial	Conjuntos de características semânticas
	Metalinguagem	Componentes como entidades lógicas
Século XX	Pragmática	Uso de jogo argumentativo
		Sentença tem função social
	Mentalista	Linguagem e esquemas mentais
		Outros níveis de representação mental
		Papéis temáticos dos argumentos de um verbo
	Referencial	Triângulo do significado
Semântica formal	Teorias de origem Lógica Filosófica	

Figura 4: Quadro sinótico das abordagens semânticas em Linguística

3 A SEMÂNTICA NA WS

No âmbito da WS, considera-se que algo tem semântica quando pode ser processado e compreendido por um computador. Essa ideia parece não considerar as origens do termo para explicá-lo e, apesar de seu viés claramente tecnológico, não é consensual, mesmo se considerado apenas o escopo da Ciência da Computação (SHETH, RAMAKRISHNAN E THOMAS, 2005). Conforme mencionado na seção um, pretende-se verificar qual semântica é considerada na WS. Uma vez descritas, na seção dois, as principais abordagens da Semântica do ponto de vista de autores e de disciplinas diversas, como Linguística, Lógica e Filosofia, cabe agora descrever os diferentes usos para o termo na área de tecnologia. O restante da presente seção se ocupa dessa tarefa, destacando a variedade de significados, interpretações dúbias e interseções com teorias apresentadas na seção dois.

Sowa (2004) lista os modelos propostos na segunda metade do século XX para compreensão da linguagem e para sua implementação em computadores: o modelo *estatístico*, o modelo *sintático*, o modelo *lógico*, o modelo *léxico* e o modelo *neural*. Segundo o autor, cada uma dessas abordagens se fundamenta em uma teoria específica, como estatística, matemática, regras gramaticais, dentre outras, mas ignora aspectos da linguagem, aos quais a tecnologia não consegue se adaptar. Os modelos lógicos, baseados na lógica filosófica, produziram teorias semânticas formais de qualidade superior em relação às abordagens concorrentes. Tais teorias têm sido amplamente utilizadas nas pesquisas sobre ontologias, estruturas que tem papel relevante no contexto da WS. Entretanto, como na teoria semântica formal, esses modelos padecem da impossibilidade de lidar como um texto normal, escrito por pessoas para

Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 16, n. 31, p.25-50, 2011. 35

fins de comunicação. A linguagem fica assim restrita a sentenças deliberadamente escritas em uma notação que é apenas similar à linguagem natural.

A *Teoria Semiótica de Peirce* ²⁴ define os três componentes principais de uma linguagem: *sintaxe*, *semântica* e *pragmática* (SOWA, 2000). Essa noção de linguagem é utilizada em Representação do Conhecimento ²⁵, onde a linguagem corresponde, em geral, a um tipo de lógica. Branchman e Levesque (2004) definem semântica da linguagem nesse contexto como a especificação do significado que se presume para expressões sintaticamente bem formadas. A especificação semântica, assim, não corresponde precisamente ao significado dos termos, mas apenas ao significado de sentenças de acordo com uma *função interpretação*, a qual remete à noção de interpretação mencionada na seção 2.1. Ainda assim, para obter especificações sobre o significado das sentenças é preciso adotar uma visão simplista de mundo, a qual considera apenas que: i) existem objetos no mundo; ii) para cada predicado P, de aridade um ²⁶, alguns objetos satisfazem P e outros não, sendo que a decisão é obtida através da função interpretação; iii) outros aspectos do mundo não interessam.

Uschold (2001) distingue e cita os tipos de semânticas presentes no mundo dos sistemas computacionais, classificando-as em *semântica do mundo real*, *semântica axiomática* e *semântica Teórica dos Modelos*. Segundo o autor, a semântica na WS se identifica principalmente com a *semântica do mundo real*, uma expressão utilizada para indicar o mapeamento de objetos do mundo para um modelo computacional.

A *semântica axiomática* é uma abordagem bastante específica e definida no âmbito das linguagens e padrões da WS, em particular o *Resource Description Framework Schema* (RDFS). Segundo Fikes e McGuinness (2001), o objetivo da semântica axiomática é proporcionar a tradução de descrições RDFS para lógica, ou seja, estabelecer regras para mapear RDFS em LPO. Esse mapeamento é realizado especificamente com vistas a proporcionar capacidade de inferências automáticas à linguagem de representação considerada.

A *Semântica Teórica dos Modelos* está relacionada à Teoria dos Modelos, e (W3C, 2004. p.2)

[...] assume que a linguagem se refere ao ‘mundo’ e descreve condições mínimas que o mundo deve satisfazer, de forma a atribuir significado apropriado às

²⁴ Charles Sanders Peirce, lógico e matemático norte-americano (1839 - 1914).

²⁵ Ramo da Inteligência Artificial, associado ao desenvolvimento de sistemas especialistas.

²⁶ Em lógica, um *predicado* descreve propriedades de objetos ou relacionamentos representados por variáveis; *aridade* é o número de argumentos de uma função (uma função binária tem dois argumentos).

expressões da linguagem. Um mundo particular é chamado interpretação, e assim a teoria dos modelos é também conhecida por teoria da interpretação.²⁷

Trata-se de uma definição similar àquela apresentada na seção dois referente à Semântica Formal. Entretanto, a continuação dessa mesma definição do W3C esclarece os objetivos da semântica naquele contexto:

A principal utilidade da teoria da semântica formal não é fornecer uma análise profunda da natureza das coisas descritas pela linguagem [...], mas sim fornecer uma forma técnica de determinar quando processos de inferência são válidos, isto é, preservar a verdade.²⁸

É importante observar que a definição, mesmo que adequada às propostas da W3C (sistemas), tem raízes na teoria da Semântica Formal, como admite a própria instituição: “a Teoria dos Modelos é uma teoria da Semântica Formal, a qual relaciona expressões a interpretações”²⁹ (W3C, 2004, p.44). Isso também fica claro na proposição de “preservar a verdade”, que remete à *Teoria da Verdade Semântica* de *Tarski*, também já mencionada na seção dois. Hodges (2005) corrobora essa visão e explica a noção de verdade, considerada a Teoria de Modelos: quando se diz que a declaração D é verdadeira de acordo com a interpretação I, está se afirmando que D, *quando interpretada a partir da I*, é verdadeira. Segundo o autor, em seu sentido amplo, a Teoria dos Modelos tem pontos de contato com a Filosofia e com os estudos da Semântica linguística. Em função disso, existe a possibilidade de interpretação dúbia entre a semântica, como definida pelo W3C (linguagem formal), e a semântica da linguagem natural.

Uschold (2001) apresenta outra classificação para a Semântica, a qual lança alguma luz sobre a confusão no uso do termo, pois está relacionada a forma *como a semântica é expressa e para quem se destina*. Nesse sentido, o autor distingue quatro tipos de semânticas alinhadas sequencialmente em um *continuum* semântico, enfatizando que, dentre esse tipos, os três primeiros não são adequados ao processamento por máquinas.

- *Semântica implícita*: o significado é comunicado a partir do entendimento comum obtido por consenso entre pessoas, por exemplo, definições de marcações *Extended Markup Language* (XML) obtidas por consenso em comunidade que usa um aplicativo;

²⁷ Model theory assumes that the language refers to a 'world', and describes the minimal conditions that a world must satisfy in order to assign an appropriate meaning for every expression in the language. A particular world is called an interpretation, so that model theory might be better called 'interpretation theory'.

²⁸ The chief utility of a formal semantic theory is not to provide any deep analysis of the nature of the things being described by the language or to suggest any particular processing model, but rather to provide a technical way to determine when inference processes are valid, i.e. when they preserve truth.

²⁹ Model Theory (n.): A formal semantic theory which relates expressions to interpretations.

- *Semântica informal*: o significado é explícito e expresso informalmente, por exemplo, como em glossários, em especificações de requisitos para sistemas, etc. (o autor não define claramente o que chama de “significado explícito” ou mesmo “semântica explícita”, a não ser por oposição a “semântica implícita”);
- *Semântica formal para processamento humano*: trata-se de uma semântica explícita e expressa em linguagem formal, mas que serve à comunicação humana e não a sistemas, por exemplo: pesquisa de semântica em categorias ontológicas;
- *Semântica formal para processamento por máquinas*: trata-se de semântica explícita, especificada formalmente, a qual pode ser utilizada por computadores para processamento direto via mecanismos de inferências para derivar novos dados a partir de existentes.

Mesmo no caso de termos idênticos, não existe consenso entre os autores em algumas definições. A definição de Uschold (2001) para *semântica implícita* considera a semântica em uma linguagem e difere da definição de Sheth, Ramakrishnan e Thomas (2005), a qual considera a semântica implícita em todos os tipos de conjuntos de dados, não apenas na linguagem. Os autores distinguem outras duas categorias para a semântica, as quais denominam *semântica formal* e *semântica nebulosa*³⁰:

- *Semântica implícita* : se refere ao significado inserido em padrões de dados não explicitados em sintaxe legível por computadores; está presente em textos não estruturados e em repositórios de documentos com estrutura pouco formal; para analisar essas fontes são utilizadas técnicas da Recuperação da Informação e da Linguística computacional;
- *Semântica formal*: dados são representados por linguagem formal baseada em estruturas sintáticas estabelecidas e por regras que definem as combinações possíveis, associadas às interpretações semânticas; declarações expressas para sistemas, utilizadas em Representação do Conhecimento, Inteligência Artificial e Bancos de Dados;
- *Semântica nebulosa*: corresponde às implementações estatísticas que permitem explorar relações não estabelecidas explicitamente; a possibilidade de composição hierárquica associada à análise estatística traz vantagens, pois permite a formalização da combinação em linguagens passíveis de uso em mecanismos de inferência; aplicada à extensão de modelos de Bancos de Dados e de Representação do Conhecimento, em casos em que a informação é imprecisa ou incompleta.

³⁰ Os autores usam dois termos originalmente: *Soft Semantics* e *Powerful Semantic*.
Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 16, n. 31, p.25-50, 2011.

A FIG. 5 resume as abordagens da semântica para sistemas computacionais, em particular para a WS, descritas ao longo da presente seção.

Abordagem	Breve descrição	Semântica linguística
Repr. do conhecimento	A semântica é formal e baseada em teorias lógico-filosóficas	formal
Repr. do conhecimento	A semântica é o significado de sentenças através de interpretação	formal
Semântica da <i>web</i>	A semântica possibilita interpretação por um computador	formal
Semântica mundo real	A semântica mapeia objetos do mundo para o sistema	indeterminada
Semântica axiomática	A semântica mapeia linguagens da WS para a Lógica	formal
Teoria dos Modelos	A semântica valida processos de inferência automáticos	formal
Semântica implícita	A semântica transmite o consenso obtido entre as pessoas	indeterminada
	Semântica inserida em padrões de dados não legível para máquinas	indeterminada
Semântica informal	A semântica é explícita e informal	indeterminada
Formal para humanos	Semântica explícita e expressa em linguagem formal, para pessoas	formal
Formal para máquinas	Semântica explícita e expressa em linguagem formal, para máquinas	formal
	Semântica definida por regras sintáticas mais interpretações	formal
Semântica nebulosa	Semântica baseada em estatística	formal

Figura 5: Quadro sinótico das abordagens à semântica na WS

A última coluna, denominada *semântica linguística*, classifica as abordagens da WS em relação ao tipo de semântica em seu campo de origem (FIG. 4, Seção Dois). A maioria das interpretações para a semântica, descritas no âmbito da WS, são nada mais do que tipos de *Semântica Formal*, existindo algumas exceções. Tais exceções são indeterminadas, pois suas descrições não possibilitam verificar sua origem e classificá-las com o mesmo critérios. Entretanto, algumas considerações podem ser feitas:

- semântica do mundo real: a prática de mapear objetos do mundo real para um sistema teve origem na pesquisa em Banco de Dados, quando os modelos de dados passaram a ser conhecidos como *modelos semânticos* (SAYÃO, 2001); semântica diz respeito a melhoria de entendimento dos modelos (os anteriores eram orientados a implementação);
- semântica implícita: trata-se de uma semântica não necessariamente formal, e o que parece caracterizá-la é a obtenção de consenso entre um grupo de pessoas ou instituições, de forma a estabelecer um padrão; em Bibliotecas Digitais, por exemplo; tal abordagem é denominada *federação* (FOX, 2002);

Além das exceções, cabe ainda destacar a FIG. 5, a *semântica formal para humanos*, de forma a esclarecer que a Semântica Formal não é utilizada apenas no contexto dos sistemas informatizados, mas também por pessoas.

4 AVALIAÇÃO DO ESPECTRO

Uma vez descritas, na seção dois, as principais abordagens sobre a semântica da linguagem natural e, na seção três, as formas como o termo aparece em pesquisas da WS, cabe agora analisar criticamente a proposta de espectro. Apresenta-se ainda um novo espectro baseado nas considerações do presente artigo.

Diversas versões de espectros e *continuums* semânticos são encontrados através de uma simples busca pela *web*. Entretanto, a maioria são variações sobre o mesmo tema. Considera-se a versão (FIG. 6) apresentada em Daconta (2005) e em Obrst (2004).

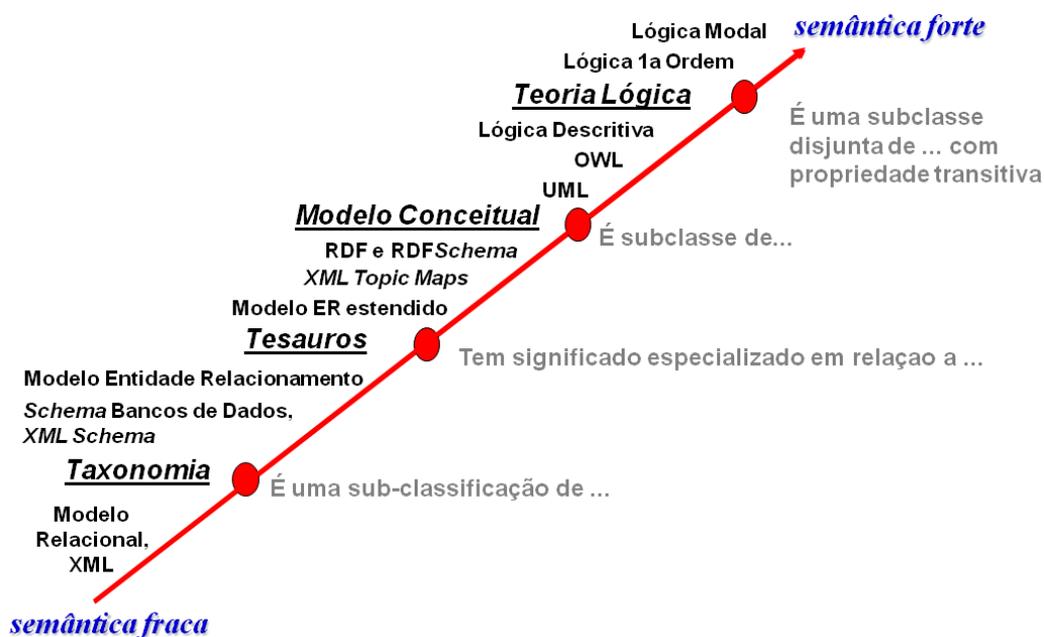


Figura 6: espectro analisado

Observa-se que o espectro é um tipo de escala que começa em *semântica fraca* e termina em *semântica forte*. Entre a semântica fraca e a forte estão dispostos *instrumentos*, sendo que um detém semântica mais forte em relação a outro, de acordo com sua posição na escala. A semântica forte e semântica fraca são caracterizações da maior ou menor expressividade do instrumento, a qual torna possível representar melhor a realidade, com vistas ao funcionamento de um sistema automatizado.

Entretanto, os instrumentos apresentados no espectro são distintos e variados: tesauros, esquemas de bancos de dados, linguagens de modelagem, linguagens declarativas orientadas a *web* e outros. A consideração conjunta desses instrumentos compromete sua avaliação e localização no novo espectro. Propõem-se separá-los em grupos, de acordo com seu uso. As FIG. 7, 8, 9 e 10 apresentam, respectivamente, os instrumentos organizados em grupos, conforme a seguinte orientação: i) relacionados à WS (e a Representação do Conhecimento);

Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf., ISSN 1518-2924, Florianópolis, v. 16, n. 31, p.25-50, 2011.

ii) relacionados a sistemas baseados na *web*; iii) relacionados a sistemas de informação automatizados; iv) relacionados a organização da informação em documentos e em registros bibliográficos.

Esses grupos também refletem as possibilidades de representação, desde aquelas voltadas diretamente ao uso humano (FIG. 10), àquelas voltadas para a representação em sistemas automatizados. No caso dos sistemas, os grupos refletem três fases evolutivas: i) sistemas de informação procedurais, baseados em bancos de dados (FIG. 9); ii) sistemas orientados a *web* (FIG. 8); iii) sistemas orientados à WS (FIG. 7). Cabe ainda citar que a lista de instrumentos não é exaustiva e que muitos dos instrumentos presentes no espectro não pertencem ao conjunto da WS. No restante da presente seção são descritos os instrumentos do espectro, destacando sua relação com a semântica e finalmente é proposto um novo espectro.

A expressão *Lógica Modal* é usada para designar uma família de lógicas (lógica modal, deôntica, lógica temporal, dentre outras), as quais tem regras similares, mas uma variedade de símbolos distintos. Segundo Garson (2008), a Lógica Modal estuda deduções que envolvem as expressões *necessariamente* e *possivelmente* e os seus operadores são: *probabilidade*, *possibilidade* e *necessidade*. A semântica na Lógica Modal é definida através da noção de *mundos possíveis*. A LPO é uma linguagem para expressar conhecimento, em que a sintaxe estabelece as regras de formação das declarações, e a semântica é capturada por interpretações. Tanto a Lógica Modal quanto a LPO não apresentam eficiência do ponto de vista das inferências e por isso são chamadas de *indecidíveis* (BRANCHMAN e LEVESQUE, 2004).

As origens da *Lógica Descritiva* remontam ao final dos anos 80, quando surgiram as *linguagens conceituais* ou *terminológicas* (BAADER et al, 1992) para representação de relações entre conceitos e relações entre conceitos e indivíduos. Segundo Baader et al (2003), a Lógica Descritiva é uma família de lógicas baseadas em estruturas hierárquicas, que apresentam um bom equilíbrio entre expressividade e tratabilidade. Tem sido utilizada na área de Banco de Dados e WS. A OWL é, na verdade, um tipo de Lógica Descritiva.

De forma a melhorar as possibilidades de uso da *web*, a WS usa recursos, como metadados, ontologias, lógica, protocolos, dentre outros. Nesse sentido, a OWL é uma tentativa de padronizar as linguagens para construção de ontologias (ANTONIU e VAN HARLEMEN, 2004). A divisão da linguagem em três dialetos busca atender a necessidades de

expressividade (no caso da OWL *Full*) e de inferências (no caso da OWL *Lite*)³¹. A OWL foi construída sobre o padrão de metadados RDFS, de forma que não é fácil separar o início das linguagens da WS. O critério utilizado aqui é a possibilidade de inferência, de forma que a RDFS foi classificada como linguagem orientada à *web*.

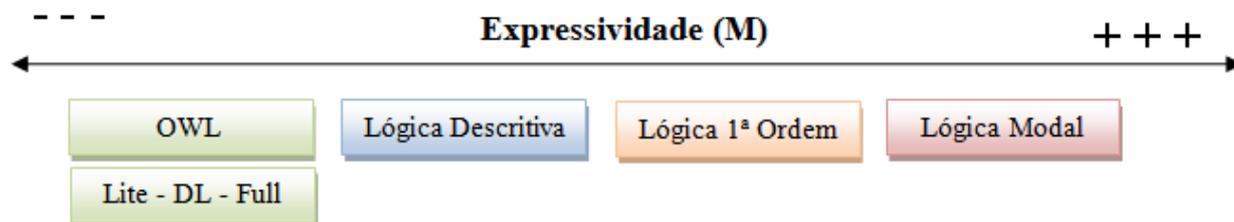


Figura 7: WS e Representação do Conhecimento, (M) = máquinas

O RDFS é um padrão de metadados que utiliza a sintaxe XML, desenvolvido para solucionar limitações do RDF (AHMED et al, 2001). O RDFS propõe como primitivas a definição de classes e subclasses, além de conter a noção de tipos de dados. Possui ainda recursos para determinar as propriedades adequadas a uma classe (marcação *domain*) e valores adequados a uma propriedade (marcação *range*). O RDF é um padrão de metadados que evoluiu a partir do XML, definindo um modelo de dados baseado em *recursos*, *propriedades* e *valores*. Esse modelo reduz a ambiguidade das declarações ao especificar um local para definição de cada elemento considerado, através do mecanismo XML *namespaces*. O W3C deixa claro o tipo de semântica envolvida na introdução da especificação RDF: “esse documento usa um técnica denominada teoria dos modelos para especificar a semântica de uma linguagem formal”³² (W3C, 2004, p.1).

O XML corresponde a uma primeira tentativa de reduzir os problemas verificados na década de 90 com o uso do HTML baseado em marcas fixas para apresentação de dados. No XML, a estrutura do documento é expressa por marcas flexíveis, tornando o conteúdo acessível a sistemas e computadores. A origem do XML é o *Standard Generalized Markup Language* (SGML), um padrão internacional para marcação de textos eletrônicos. A SGML define um *Document Type Definition* (DTD), cujo objetivo é estruturar documentos através das marcações, ou seja, definir como as marcas são distribuídas pelo texto. Cabe destacar que o

³¹ Para maiores detalhes sobre as diferenças entre os três tipos de linguagens OWL ver Horridge et al (2004).

³² This document uses a basic technique called model theory for specifying the semantics of a formal language.

W3C apresenta especificações para a semântica do RDF e do RDFS, mas não apresenta um documento similar para o XML.

As linguagens de marcação e os padrões de metadados orientados à *web* corresponderam a uma mudança de paradigma na área de sistemas, no qual a ênfase passou de um modelo baseado em dados estruturados (Bancos de Dados) para um baseado em dados semiestruturados da *web* (ABITEBOUL, BUNEMAN e SUCIU, 2000).

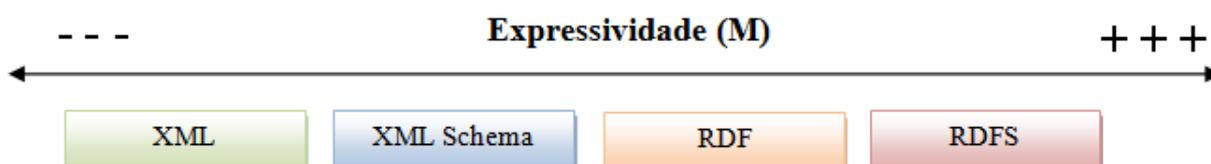


Figura 8: orientados a *web*, (M) = máquinas

O *Modelo Relacional* (Codd, 1970) foi um dos primeiros instrumentos para modelagem conceitual. Entretanto, nesse modelo o conceito *relação* é usado para representar tanto uma *entidade* quanto *relacionamentos entre entidades* (PECKHAM e MARYANSKI, 1988). Esse fato gera dificuldades de entendimento, torna difícil o mapeamento de conceitos do mundo e induz a erros de modelagem. No Modelo Relacional, uma relação corresponde a uma *tabela*, com linhas denominadas *tuplas* e colunas denominadas *atributos*. Segundo Kroenke (1998), esses termos têm origem da *Álgebra Relacional*, um ramo da LPO. O conceito de tabela é similar aos conceitos de relação na matemática. A semântica do Modelo Relacional diz respeito à Semântica Formal.

Silberchatz, Korth e Sudarshan (1999) explicam que um *schema* de banco de dados corresponde ao seu esquema lógico, ou seja, ao modelo conceitual após a aplicação das formas normais (regras para normalização do banco de dados). Como exemplo, explicam que o conceito correspondente seria o de uma *variável*, a qual se pode atribuir *instâncias*. A semântica relacionada ao *schema* também é a Semântica Formal.

Em função dos problemas do modelo relacional, desenvolveu-se o *Modelo Entidade-Relacionamento* (E-R) (CHEN, 1976), o qual fornece termos extras para modelagem usados como primitivos. Além disso, o E-R elimina a sobrecarga do constructo *relação*. Segundo Silberchatz, Korth e Sudarshan (1999), a maior expressividade do E-R diz respeito à tentativa de representar o significado dos dados através de um mapeamento da realidade em um modelo conceitual. O E-R é um *modelo semântico*, no qual, entretanto, a semântica

referenciada é a Semântica Formal, pois um relacionamento “[...] é a relação matemática com n^2 conjuntos de entidades” (SILBERCHATZ, KORTH e SUDARSHAN, 1999, p. 25)

A *Orientação ao Objeto*, uma proposta para modelagem conceitual de sistemas, propõe novas formas de lidar com eventos, os quais definem o estado de um modelo (OLIVÉ, 2007). Na verdade, as propostas de Orientação ao Objeto apresentam a mesma teoria e diferentes notações. Uma tentativa de uniformizar essas notações é a *Unified Modeling Language* (UML). A definição de semântica na UML é um tanto vaga, uma vez que o termo é citado em vários contextos e as suas especificações informam que “existem, é claro, outras semânticas relevantes para a UML [...]”³³ (OMG, 2007, p.10). Apesar da variedade, observam-se duas situações específicas a se considerar.

A UML utiliza uma linguagem formal, a *Object Constraint Language* (OCL), para descrever expressões em modelos UML em que a semântica é definida como “[...] o mapeamento das expressões OCL para valores do domínio semântico”³⁴ (OMG, 2006, p. 95). Nesse contexto, um *domínio semântico* é o conjunto de valores que as expressões podem produzir. A especificação possui ainda um anexo sobre semântica, no qual se define formalmente o que chama de *modelo de objeto* usados na OCL, a partir de definições da *Teoria dos Conjuntos* e sua *interpretação*. Na seção dois, esse elemento são descritos como parte da Semântica Formal.

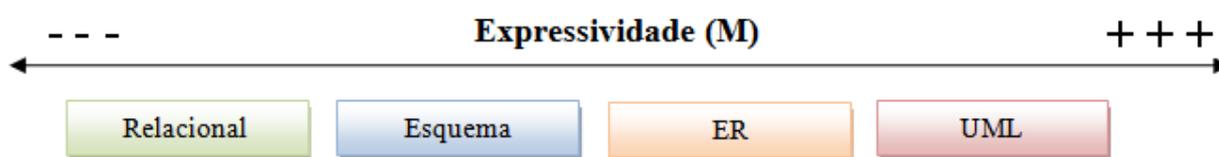


Figura 9: orientados a sistemas de informação, (M) = máquinas

A FIG.10 apresenta um conjunto de instrumentos diferentes dos anteriores. A expressividade desses instrumentos é voltada para pessoas e não para sistemas. Nesse contexto, a semântica é a semântica da linguagem natural. Um *tesauro* é um instrumento de controle de vocabulário que permite relacionar termos representativos do conteúdo de documentos de acordo com três tipos de relações: termo genérico, termo específico e termo relacionado. Uma *taxonomia* corresponde a uma estrutura básica de organização da informação baseada em relações de classe/subclasse. Um *dicionário* é uma lista de termos alfabeticamente ordenados em um

³³ There are, of course, other semantics relevant to UML specifications.

³⁴ [...] represent the mapping from OCL expressions to values from the semantic domain.

idioma específico, que apresenta, além de definições, informações, como figuras, pronúncia, etimologia dos termos, dentre outros. Um *glossário* é uma lista de termos relacionados a um domínio específico e de suas definições.

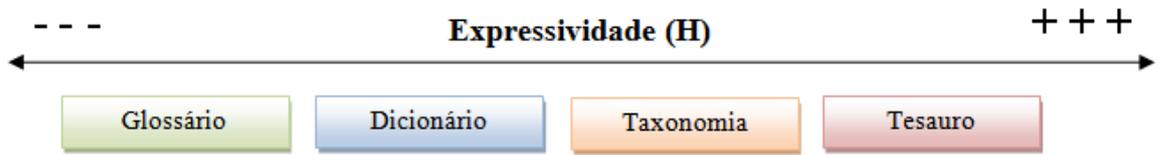


Figura 10: orientado a organização da informação, (H) = humanos

Após as considerações sobre os instrumentos, apresenta-se a proposta para o *novo espectro*, o qual permite visualizar todos os instrumentos conjuntamente (FIG. 11). Um eixo orientado por uso divide o espectro em dois quadrantes: uso da semântica por humanos e uso por máquinas; em cada quadrante a expressividade semântica do instrumento é considerada, conforme o caso, em semântica da linguagem natural ou em semântica Formal.

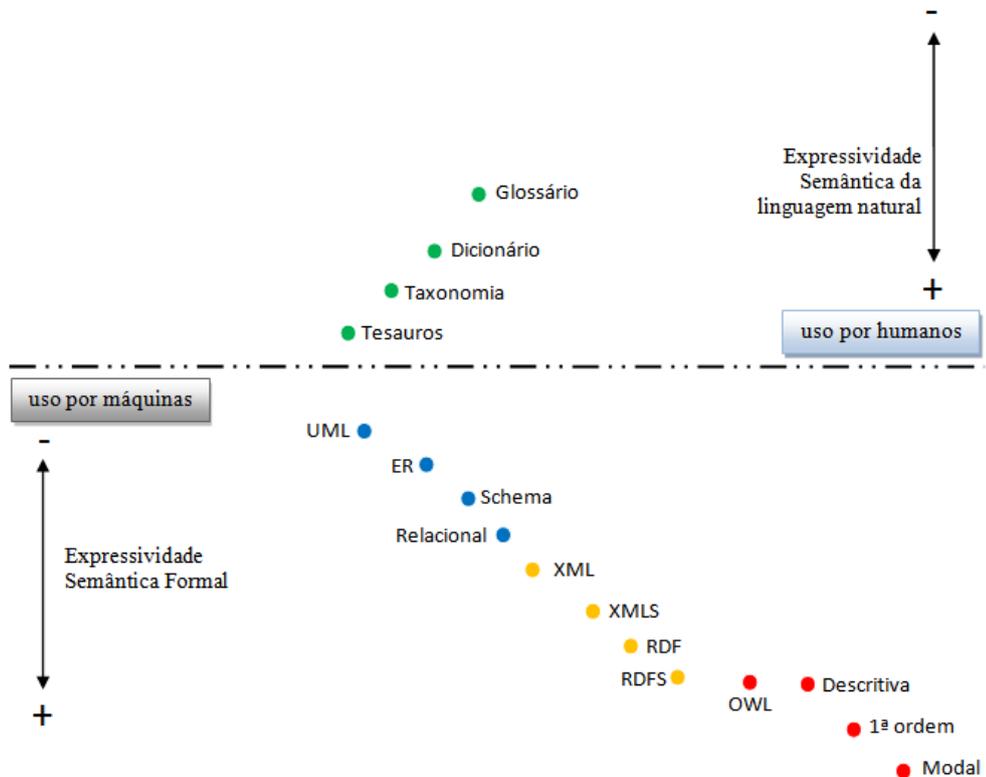


Figura 11: Proposta de espectro modificado

5 CONSIDERAÇÕES FINAIS

O presente artigo apresentou uma breve revisão de literatura sobre semântica em Linguística e buscou entendimento sobre o uso do termo semântica na área de tecnologia da informação, em particular, na visão da WS. Discutiu um espectro semântico e apresentou nova proposta, a qual leva em conta o uso do instrumento (por computadores ou por pessoas). Enfatiza-se novamente que a revisão de literatura não é exaustiva e, com certeza, pesquisadores importantes da Filosofia e Linguística não foram citados. Salienta-se ainda que a breve explicação sobre os instrumentos para uso humano não indica que estes são menos importantes. Apenas buscou-se manter os objetivos do artigo, ou seja, apresentar a Semântica em seus vários contextos de uso.

No final da seção 2.1, mencionou-se a importância da abordagem linguística da Semântica Formal para os objetivos do artigo. Tal importância é comprovada pela verificação de que todas as abordagens relacionadas à tecnologia da informação são, na verdade, tipos de Semântica Formal. Com isso em mente, torna-se mais simples entender o uso do termo. Esse entendimento é importante no âmbito da Ciência da Informação para que não haja, por exemplo, algum tipo de confusão entre a semântica usada em tesouros e a semântica usada em ontologias.

Dentre as possibilidades de interpretação errôneas, as quais o presente artigo pretende reduzir, existe uma que merece destaque. Trata-se da confusão entre a semântica, ou a definição do significado, no *planejamento de um sistema* e a semântica *que é usada pelo sistema*. Em muitos casos, como nos modelos semânticos e na UML, as especificações citam como definir o significado dos objetos de forma que sejam representados. Um caso típico é o *dicionário de dados*, em Bancos de Dados e em atributos XML, onde se definem entidades, como, por exemplo: *autor é o responsável intelectual por uma obra; aluno é a pessoa que está matriculada em uma instituição educacional*; e assim por diante. De fato, nesse caso o que se define é o que realmente se entende por *autor* e por *aluno* em certo contexto, por pessoas, algo próximo do significado na linguagem natural.

Entretanto, a questão relevante aqui não é exatamente o significado de coisas no mundo. Essa definição é importante, sejam os sistemas procedurais, orientados a *web*, a WS, etc. Sempre existe uma fase em que as pessoas definem o que o sistema vai abranger, fase essa conhecida por modelagem de negócios, modelagem conceitual, análise de sistemas, aquisição de conhecimento, dentre outras denominações. A ênfase do artigo não está nessa fase, e sim na

semântica do ponto de vista do uso pelo sistema no computador. A partir dessa consideração, organizou-se o espectro, ou seja, o instrumento que apresenta capacidade de dedução, de inferência, é o mais expressivo (quadrante inferior, FIG. 11). Essa característica leva o sistema a ser considerado mais evoluído em relação a outros, o que se chama de *sistema inteligente*, o tipo de sistema que a WS busca.

A WS parece uma iniciativa promissora e as possibilidades que traz em suas propostas são inovadoras. Entretanto, a validade dessas propostas não é consenso. A WS prioriza a pesquisa para melhorias na capacidade de expressão e na computabilidade da lógica, em particular da Lógica descritiva. Tal compromisso fica claro na ênfase da pesquisa em linguagens de representação, como, por exemplo, a OWL. Na verdade, semântica e significado são assuntos bem mais complexos e abrangentes do que as implementações em LPO e Teoria dos Modelos que a WS promove. Essa constatação leva a um descrédito em relação à proposta. Gärdenfors (2004, p.2), por exemplo, afirma que “[...] a WS não é semântica”³⁵.

É preciso admitir que pesquisadores da área de WS deixam claro o uso da Semântica Formal, mas apresentam explicações demasiadamente pobres para defini-la. Antoniu e Van Harmelen (2004, p.110) definem Semântica Formal como uma semântica que “descreve precisamente o significado do conhecimento. Por ‘precisamente’, entende-se aqui que a semântica não se refere a intuições subjetivas, nem é aberta a interpretações diferentes por pessoas (ou máquinas)”³⁶. As definições do W3C citadas na seção três seguem a mesma linha.

Em relação ao novo espectro, observa-se que sua organização não é tão diferente em relação ao avaliado. Conclui-se que o espectro avaliado traz uma visão razoável sobre a semântica dos instrumentos, menos do detalhe essencial relativo ao uso da semântica. Observa-se ainda que o novo espectro precisa de ajustes, por exemplo: i) o espaço entre os instrumentos não segue uma escala que expressa as diferenças semânticas de forma quantitativa; ii) a Lógica Modal está localizada no extremo mais expressivo, no quadrante das máquinas, sendo que é não decidível (não gera inferências eficientes), além de ser utilizada em pesquisas sobre linguagens formais para uso por pessoas (GUARINO, 2004). Espera-se assim, em trabalhos futuros, buscar os ajustes necessários para que o espectro seja o mais representativo possível da semântica dos instrumentos.

³⁵ [...] the Semantic Web is not very semantic.

³⁶ A formal semantics describes the meaning of knowledge precisely. Precisely here means that the semantic does not refer to subjective intuitions, nor is open to different interpretations by people (or machines).

REFERÊNCIAS

- ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. **Data On The Web: from relations to semistructured data and XML**. San Francisco: Morgan Kaufmann, 2000. 258p.
- AHMED, K. et al. **Professional XML Meta Data**. Birmingham: Wrox, 2001. 567p.
- ANTONIU, G. e VAN HARLEMEN, F. **A semantic Web Primer**. Cambridge: MIT, 2004, 238p.
- AURÉLIO, B. H. F. **Novo Aurélio Século XXI: o dicionário da língua portuguesa**. 3ed. Rio de Janeiro: Nova Fronteira, 1999. 2128p.
- BAADER, F. et al. **Terminological Knowledge representation: a proposal for a terminological logic**. Disponível em: <<https://eprints.kfupm.edu.sa/68607/1/68607.pdf>>. Acesso em: 11 jan. 2003.
- _____. **The Description Logic Handbook: Theory, Implementation and Applications**. Cambridge: Cambridge Press, 2003. 568p.
- BRANCHMAN, R.J.; LEVESQUE, H. J. **Knowledge Representation and Reasoning**. San Francisco: Morgan Kaufmann, 2004. 381p.
- BERNERS-LEE, T. **Semantic Web Road map**. (1998). Disponível em: <<http://www.w3.org/DesignIssues/Semantic.html>>. Acesso em: 20 Jan. 2003.
- _____. **The Semantic Web; A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities**. (2001). Disponível em: <<http://www.sciam.com/article.cfm?id=the-semantic-web>>. Acesso em: 16 jan. 2005.
- CHEN, P. The Entity-Relationship Model; Toward a Unified View of Data. *ACM Transactions on Database Systems*, v.1, n.1, p. 9–36, 1976.
- CODD, E. F. A Relational Model for Large Shared Data Banks. *Communications of the ACM*, v.13, n.6. p. 377–387, 1970.
- DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. **The Semantic Web; A guide to the future of XML, Web Services and Knowledge Management**. Indianapolis: Wiley, 2005. 312p.
- DAHLBERG, I. Teoria do conceito. **Ciência da informação**. Rio de Janeiro, v. 7, n. 2, p. 101–107, jul./dez. 1978.
- DUMMETT, M. 1981 *Frege; Philosophy of Language*. New York: Harper & Row, 1973. 698p.
- FIKES, R.; MCGUINNESS, D. (2001). **An Axiomatic Semantics for RDF, RDF-S, and DAML+OIL**. Disponível em: <<http://www.w3.org/TR/daml+oil-axioms>>. Acesso em: 27 jan. 2006.
- FOX, E.; SULEMAN, H.; LUO, M. **Building Digital Libraries Made Easy; Toward Open Digital Libraries**. Disponível em: <<http://www.springerlink.com/content/7x87gky9k5ggt5yq/>>. Acesso em: 31 mai. 2004.
- GÄRDENFORS, P. How to make the Semantic Web more semantic. In: VARZI, A. C.; VIEU, L. (Eds.). **Formal Ontology in Information Systems**. Amsterdam: IOS Press, 2004, p. 19-36.
- GARSON, J. **Modal Logic**. Disponível em: <<http://plato.stanford.edu/entries/logic-modal/>>. Acesso em: 13 jan 2009.

HORRIDGE, M. et al. **A Practical Guide To Building OWL Ontologies Using The Protégé-OWL Plugin and CO-ODE Tools.** (2004). Disponível em: <<http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>>. Acesso em: 03 jun. 2008.

HODGES, W. **Model Theory.** (2005). Disponível em: <<http://plato.stanford.edu/entries/model-theory/#Def>>. Acesso: 28 fev. 2008.

KROENKE, D. M. **Bancos de Dados: Fundamentos, Projeto e Implementação.** 6ed. Rio de Janeiro: LTC, 1998. 382p.

MORRIS, M. **An Introduction to the Philosophy of Language.** Edimburgh: Cambridge, 2007. 326p.

NETO, J. B. Semantica dos modelos. In: MULLER, A. L.; NEGRÃO, E. V.; FOLTRAN, M. J. (Eds.). **Semantica formal.** São Paulo: Contexto. 2003, p. 47-60.

NIRENBERG, S.; RASKIN, V. **Ontological Semantics: Language, Speech, and Communication.** Cambridge: MIT Press, 2004. 442p.

NOTH, W. **Handbook of Semiotics.** Indianapolis: Indiana University Press, 1995. 576p.

OBRST, L. **Ontologies & the Semantic Web for Semantic Interoperability.** (2004). Disponível em: <<http://www.webservices.gov/OntologiesSemanticWebSemInteropSICOP909-Obrst.ppt>>. Acesso em: 21 ago. 2005.

OGDEN, C. K. et al. **O significado de significado: um estudo da influência da linguagem sobre o pensamento e sobre a ciência do simbolismo.** Rio de Janeiro: Zahar, 1972. 349 p.

OLIVÉ, A. **Conceptual Modeling of Information Systems.** Berlin: Springer, 2007. 455p.

OLIVEIRA R. P. **Semântica formal: uma breve introdução.** Campinas: Mercado de Letras, 2001. 264p.

OMG. **Unified Modeling Language (OMG UML), Infrastructure, V2.1.2. (2007).** Disponível em: <<http://www.omg.org/docs/formal/07-11-04.pdf>>. Acesso em: 28 dec. 2008.

_____. **Object Constraint Language OMG Available Specification Version 2.0 (2006).** Disponível em: <<http://www.omg.org/docs/formal/06-05-01.pdf>>. Acesso em: 27 dec. 2008.

PECKHAM, J.; MARYANSKI, F. **Semantic Data Models.** (1988). Disponível em: <<http://portal.acm.org/citation.cfm?id=62062>>. Acesso em: 30 jul. 2007.

PORTNER, P.H. **What is meaning? Fundamentals of Formal Semantics.** Oxford: Blackwell, 2005. 235p.

SAYÃO, L. F. **Modelos teóricos em ciência da informação: abstração e método científico.** (2001). Disponível em: <<http://www.scielo.br/pdf/ci/v30n1/a10v30n1.pdf>>. Acesso em: 2 dez. 2002.

SEDLEY, D. **Plato's Cratylus.** (2006). Disponível em: <<http://plato.stanford.edu/entries/plato-cratylus/>>. Acesso em: 21 dez. 2007.

SHETH, A.; RAMAKRISHNAN, C.; THOMAS, C. (2005). **Semantics for the Semantic Web: The Implicit, the Formal and the Powerful.** Disponível em: <<http://lstdis.cs.uga.edu/library/download/SRT05-IJ-SW-IS.pdf>>. Acesso em: 3 nov. 2008.

SILBERCHATZ, A.; KORTH, H. e SUDARSHAN, S. **Sistema de Bancos de Dados.** 3ed. São Paulo: Makron Books, 1999. 778p.

SOWA, J. **Ontology, Metadata, and Semiotics.** (2000). Disponível em: <<http://www.jfsowa.com/ontology/ontometa.htm>>. Acesso em: 27 fev. 2002.

SOWA, J. **A Dynamic Theory of Ontology.** (2004). Disponível em: <<http://www.jfsowa.com/pubs/dynonto.htm>>. Acesso em: 5 mai. 2006.

W3C. **RDF Semantics.** (2004). Disponível em: <<http://www.w3.org/TR/rdf-mt/>>. Acesso 3 set. 2006.

_____. **OWL; Web Ontology Language Overview.** (2004). Disponível em: <<http://www.w3.org/TR/2004/REC-owl-features-20040210/>>. Acesso em: 4 set. 2006.

Agradecimentos

Bolsa CAPES - Processo no BEX 4071/09-3 (primeiro autor).

Abstract

The term “semantics” has been used over the past years in several areas of research and, more specifically, in those related to information technology. One reason for that appropriation is the Semantic Web vision, which originated from the expansion of the web and the inherent limitations of the syntax-based search engines. However, it is reasonable to demand an explanation about which is the semantics addressed in the Semantic Web since the study of the subject is complex and controversial. The present article aims to make a contribution towards this sense, analyzing what semantic is, discussing it within the scope of information technology and proposing explanatory alternatives. We intend to critically evaluate a spectrum that suggests a classification of instruments (models, languages, taxonomic structures, etc) in a semantic-based scale of graduated rates. We hope to provide a better understanding about the use of the term in different contexts, as well as offer a new spectrum from the considerations of the present article.

Keywords: Semantic Web. Ontologies. Semantics. Knowledge representation.

Originais recebidos em: 11/11/2009

Aceito para publicação em: 24/10/2010