

# A VIDA SECRETA DOS METADADOS NO WIKIDATA: um enfoque sobre o sentido na (Web) Semântica Formal

Silvana Drumond Monteiro\*

## RESUMO

Desde o anúncio da *Web Semântica*, até o presente momento, muitas tecnologias e inovações foram desenvolvidas para dar corpo e efetividade ao conceito. Essa teia de significações é composta por linguagens e tecnologias semânticas, como o XML (*eXtensible Markup Language*), RDF (*Resource Description Framework*), Declarações, entre outras que estruturam os dados que perambulam pelo ciberespaço, formando grandes bases de conhecimento, como o Wikidata (*Wikimedia Foundation*). A lógica principal da *Web Semântica* está em descrever as entidades (coisas) do mundo real e vinculá-las a outras, para que essa rede de significações possa ser ativada, recuperada e interpretada pelas máquinas e humanos. A pesquisa de delineamento documental teve como *corpus* de pesquisa o Wikidata. A investigação visou perscrutar essa base para apreensão da estruturação de metadados para responder a questão: que tipo de Semântica estaria sendo produzida em uma base de conhecimento? Como resultado, observou-se que a Semântica Formal é o fundamento para a geração de metadados, entretanto, no paradigma pós-moderno, o sentido – e a própria Semântica – são um fenômeno sociotécnico no ciberespaço.

**Palavras-chaves:** Wikidata. Metadados. Semântica Formal. Web Semântica

\* Doutora em Comunicação e Semiótica pela Pontifícia Universidade Católica de São Paulo, Brasil. Professora associada do Departamento de Ciência da Informação da Universidade Estadual de Londrina, Brasil. Pós-Doutora em Ciência da Informação pela Universidade Federal de Minas Gerais, Brasil. Líder do Grupo de Pesquisa CNPq Informação e Conhecimento no Ciberespaço. Pesquisadora Produtividade CNPq.  
E-mail: silvanadrumond@gmail.com.

## I INTRODUÇÃO

*Somos uma máquina semântica inteligente.*  
(OLIVEIRA, 1995)

A semantização do ciberespaço, preconizada pela Web Semântica (WB) por Tim Berners-Lee *et al.* (2001), ganha escalabilidade, no ciberespaço, graças às bases de conhecimento (*knowledge bases*) e a visualização dos metadados nos mecanismos de busca. Nessa direção, o *Knowledge Graph* (KG) - ou grafo do conhecimento - associado ao Wikidata - base de dados de conhecimento - estão transformando os índices em relações semânticas enciclopédicas, já que para Eco (2013), a enciclopédia é uma representação semântica do saber, e a Wikipédia,

na matriz das entidades e do conhecimento enciclopédico, na atualidade.

De acordo com essa reflexão, mas com foco no Wikidata, algumas questões de investigação emergem: estaria o conceito de Semântica no sentido tecnológico, de certa forma, contemplado nas concepções (mais clássicas) de Semântica? Ou ainda, estaríamos vislumbrando um novo paradigma de Semântica, a partir dos engendramentos híbridos de significações no ciberespaço?

Considerando a língua natural como uma maquinaria, provida de criatividade, referencialidade e redes de sentenças, como transpor esses atributos para agentes computacionais e assim emular a interpretabilidade do mundo? Para compreender algumas dessas facetas, o campo da Semântica,

que pretende realizar uma descrição do significado, foi sumarizado, neste artigo, especialmente por intermédio de autores como Eco (2013); Oliveira (2012) e Bréal (1992).

Destarte, a Semântica Formal (SF) foi objeto principal de reflexão, pois ela visa construir uma metalinguagem, uma forma semântica que espelhe a interpretação de uma sentença, assim como a Web Semântica. Lévy (2014, p. 168, grifo do autor) considera a metalinguagem “[...] uma espécie de escrita de segundo grau.”, fundada pela computação semântica e raciocínios automáticos.

O objetivo do estudo, além de responder as questões supracitadas, foi fazer a aproximação da SF com a epistemologia contemporânea para entender os fenômenos sociotécnicos e assim tornar inteligível, a partir da teoria e dos metadados, a Semântica que ocorre no Wikidata. Isso posto, o artigo enfoca somente a camada de descrição dos dados no Wikidata, não abordando todas as camadas e tecnologias semânticas da WS.

Já os metadados, tidos como “dados sobre os dados”, definição clássica na Ciência da Informação (CI), no meio digital “[...] referem-se à estrutura descritiva da informação sobre outro dado [...]” (PINHEIRO, 2004, p. 25), isto é, ganham *status semântico*, descrevem não somente elementos identificadores de obras, mas formam uma teia de significações em torno de um item.

Assim, a propósito do título, parafraseando o filme “A vida secreta das palavras”, de Isabel Coixet (2005), é uma alusão ao modo de existência da WS, que nas camadas invisíveis do ciberespaço, o sentido encontra-se amalgamado nos metadados.

## 2 SOBRE “AS SEMÂNTICAS”

Bréal (1992, p. 20), considerado o pai da Semântica, em seu livro seminal (publicado em 1897) alerta para que se “[...] veja este livro como uma simples Introdução à ciência que me proponho a chamar de Semântica. [...] a ciência das significações, do verbo, ‘significar’, por oposição à fonética, a ciência dos sons.” Seu livro propõe pensar o sentido das palavras como uma evolução humana, ou seja, as leis intelectuais da linguagem tem sua morada na inteligência humana.

Ciência abrangente, muitas vezes abstrata, com várias abordagens, tem como objetivo o

estudo do significado. De acordo com Tamba (2006, p. 12), a “Semântica se furta a toda apreensão teórica homogênea.” Assim, propõe-se, nesta seção, uma visão de conjunto sobre “as semânticas”, apenas para localizar a SF, que é, de fato, o objeto deste artigo. Para a autora, é possível distinguir quatro períodos dessa disciplina, a saber:

- 1) o período evolucionista da Linguística comparada, no qual domina uma Semântica lexical e histórica (a exemplo de Bréal);
- 2) o período estrutural;
- 3) o período das gramáticas formais;
- 4) o período cognitivista.

Para Eco (2013), a partir de Bréal, a Semântica pode ser categorizada em cinco tipos específicos, a saber:

- 1) Semântica como estudo do significado dos termos, fora de qualquer contexto (conjunto textual): diz respeito às **definições** e à discussão dicionário-enciclopédia;
- 2) Semântica estrutural ou como estudo dos sistemas de conteúdo;
- 3) Semântica como estudo da relação entre o termo (ou enunciado) e referente ou como estudo da referência;
- 4) Semântica como o estudo das condições de verdade dos enunciados: Eco (2013) a denomina como semântica verifuncional;
- 5) Semântica como o estudo do sentido específico que termos ou enunciados assumem no contexto ou no conjunto textual.

O estudo do significado, na dimensão 3, foi reforçado com a apresentação visual do famoso triângulo semiótico publicado em 1923, por Ogden e Richards (1972), e reiterado pelos linguistas sucessores, embora sem contemplar a captura do referente pela linguagem, reforçando o funcionamento lógico-ideal do significantesignificado (BLIKSTEIN,1995).

Já na Semântica 4, a questão mais complexa é precisar o conceito de verdade. Nessa dimensão, a relação da linguagem com o mundo (realidade) torna-se mais fácil à inteligibilidade, ou seja, com a referência. Nessa direção, a verdade de um enunciado é avaliada sob as condições e circunstâncias em que ele pode ser verdadeiro.

Assim, das definições de Semântica que Eco (2013) apresenta, adota a defesa do número 5, inspeção de sistemas do conteúdo que se apoia em uma vasta cognição de textos precedentes e seu sentido, ou da desambiguação de textos complexos “[...] são o texto e o contexto que fixam definitivamente o significado dos termos.”, como explica Eco (2013, p. 543).

Já Oliveira (2012) resume a Semântica contemporânea em três abordagens:

- 1) Semântica Argumentativa (SA): elaborada por Oswald Ducrot (1984), na França na década de 70. Seu postulado consiste em afirmar que a linguagem é usada para convencer o interlocutor e não exclusivamente para falar algo sobre o mundo. Seu propósito é o convencimento, e nessa perspectiva, aproxima-se da Pragmática;
- 2) Semântica Cognitiva (SC): surge na década de 80 com George Lakoff e Mark Johnson (1980), EUA, e se contrapõe à Semântica Formal por acreditar que o pensamento não é proposicional, mas um mapa, isso é, estruturado por esquemas, como as metáforas e imagens;
- 3) Semântica Formal (SF): tem início na Grécia antiga com os estudos sobre os silogismos, de Aristóteles. A partir da junção da Gramática gerativa, de Chomsky e da Gramática categorial de Montague, Barbara Partee (1973), na década de 70 nos EUA, desenvolve a Semântica Formal das línguas naturais. As relações de significado podem ser descritas formalmente.

Enquanto a SA pauta-se no sentido a partir do convencimento do interlocutor, diminuindo os limites entre Semântica e Pragmática, a SF centra-se no sentido a partir da referência. Já a SC,

Nessa perspectiva, a linguagem ordinária, nossas ações e nosso pensamento são estruturados, em sua maior parte, por metáforas, mapeamentos entre domínios conceituais distintos. Uma sentença como ‘João gastou dois dias neste trabalho’ é metafórica, porque ela é uma manifestação de um mapa cognitivo que entende o tempo como algo que consumimos, que gastamos. (OLIVEIRA, 2012, p.27).

Percebe-se que, a partir da exposição breve das vertentes da Semântica, os modos de descrição de um significado variam de acordo com as teorias linguísticas. Nesse sentido, este texto adota a classificação de Oliveira (1995) por focar a SF, objeto de inteligibilidade das bases de conhecimento que povoam o ciberespaço, portanto dos metadados e da Web Semântica, a partir do Wikidata. Para tanto, faz-se necessário explicitar a SF para então perscrutar a “vida secreta dos metadados”, ou seja, dos dados que são estruturados, a partir de declarações ou sentenças no Wikidata.

## 2.1 Semântica formal

A SF deve seu desenvolvimento aos estudos das gramáticas formais e, em especial, de Frege sobre o conceito de significado. A abordagem verifuncional entende que a referencialidade é fundamental para entender o fenômeno do significado, isto é, a premissa básica é que ao usar a linguagem, fala-se sobre o mundo e “[...] alcança um objeto particular no mundo.” (OLIVEIRA, 2012, p.89):

A reflexão de Frege não contribuiu apenas porque esclarece o conceito de significado, mostrando que nele estão embutidas as noções de sentido e referência [...]; a partir dela abre-se uma tradição de análise da estrutura semântica das sentenças, que é a base da semântica formal. Esta estrutura é muitas vezes denominada de forma lógica. (OLIVEIRA, 2012, p.121).

Assim como a abordagem verifuncional, a SF pressupõe uma forma semântica composta de predicados e argumentos para descrição dos conceitos. Compreende também que o significado engloba o sentido e a referência “O sentido é o caminho que nos leva à referência.” (OLIVEIRA, 2012, p. 95). Para exemplificar essa abordagem, a autora apresenta a seguinte reflexão. “Quando falamos em significado, ora pensamos na referência, ora no sentido, daí a ambiguidade da pergunta: qual o significado do azul?” Pode-se pensar a amostra de cor azul ou no sentido de azul (p. 95).

A partir dessas premissas, a função de um semanticista é construir uma metalinguagem que descreva a língua-objeto e também forneça as condições em que a sentença é verdadeira.

Para desempenhar essas funções, a SF deve ser capaz, em sua estrutura, de traduzir as relações de sentenças por meio de cálculos (ou notações) dos predicados. Em seu livro, Oliveira (2012, p.16-17) apresenta a questão: “[...] até que ponto o cálculo dos predicados traduz as línguas naturais?” O objetivo é desenvolver uma máquina que traduza e reproduza o conhecimento que um falante tem. A resposta para tal indagação poderia muito bem ser utilizada para descrever a Web Semântica.

Para Lyons, “Por predicado entende-se um termo que é usado em combinação com um nome a fim de fornecer uma determinada informação acerca do indivíduo a que o nome se refere: i.e., a fim de atribuir uma propriedade.” (apud OLIVEIRA, 2012, p.128), conforme modelo abaixo:

ARGUMENTO [ser] brasileiro ----- PREDICADO  
(Lugar vazio)



Os predicados são expressões insaturadas e ao preenchê-las, tornam-se saturadas e completas, com sentido e referência. Ou ainda, um predicado é uma estrutura recorrente com espaços vazios que podem ser preenchidos por expressões e por nomes próprios (estes últimos podem referir-se a indivíduos no mundo).

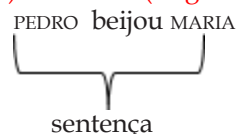
Traduzindo o exemplo para uma metalinguagem lógica e cálculo de predicados, ele pode ser representado da seguinte maneira:

(Ser brasileiro (João)) ou “Bj” em que: “ser brasileiro” é representado pela letra maiúscula “B” e “João” por letra minúscula “j”.

Oliveira alerta que a escolha das letras é arbitrária e outros símbolos são acrescentados conforme as sentenças fiquem mais complexas, como os operadores e os conectivos, como: “não” (~), “e” (&), “ou” (v), “se e somente se” (v), “se... então” (→).

No entanto, a classe de palavras que pode ser predicado difere da gramática: trata-se de uma estrutura com lacunas que indicam a possibilidade de preenchimentos alternativos, conforme o próximo exemplo (OLIVEIRA, 2012, p. 129):

ARGUMENTO beijar ARGUMENTO  
(Lugar vazio) PREDICADO (Lugar vazio)



Retornando à questão da referência: a sentença citada por Oliveira (2012, p. 42) “Este casaco é bonito.” será verdadeira se e somente se o objeto que o falante aponta é um casaco e se este possui a propriedade de ser bonito. As condições de verdade têm relação com uma propriedade.

A referência na SF é extremamente importante, pois ela é tomada como verdade e, portanto, esvaziada de questões filosóficas. Isso dito, Oliveira (2012, p.95) alerta que a tarefa da SF não é dizer se uma sentença é verdadeira ou falsa, mas estabelecer em que condições no mundo ela seria verdadeira e “O sentido é o caminho que nos leva à referência.”

O estudo da Semântica na Computação também é presente e conforme Uschold (2001) apud Almeida e Souza (2011), as principais vertentes dessa disciplina nos estudos de sistemas são: 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 dos objetos do mundo para um modelo computacional.” (ALMEIDA; SOUZA, 2011, p. 36).

Nessa direção, sintetizam as abordagens encontradas na literatura, de acordo com o Quadro 1:

QUADRO 1: Quadro sinótico das abordagens da Semântica para sistemas computacionais e para a WS

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 web	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

Fonte: ALMEIDA; SOUZA (2011, p. 39)

É interessante observar que a maioria dos estudos supracitados é **derivada** da SF, com as exceções em que não é possível rastrear a sua origem (indeterminada). O Quadro 1, embora não seja o objeto discorrer todas as vertentes da Computação, ilustra a importância da SF nos estudos da Semântica, em especial na WS.

### 3 O WIKIDATA À PERSPECTIVA DA SEMÂNTICA FORMAL: os metadados

As construções colaborativas técnicas, sociais e linguísticas (semióticas) no ciberespaço vêm aumentando consideravelmente, de forma que Lévy (2015, p. 156) alerta que devemos repensar as formas de organização dos arquivos e da memória digitais, ou seja, pensar em uma nova esfera semântica, computável, aberta, universal e democrática. “[...] os saberes implícitos são transformados em *dados* (via as entradas de *blogs*, de *wikis* e *artigos*) e em *metadados*, por intermédio de uma atividade de categorização participativa.” (grifos do autor). Dessa forma, admite que os metadados semânticos permitem a organização e a pesquisa de informações.

De acordo com Rocha (2004), no meio digital, o conceito de metadado é estendido, sendo possível utilizá-lo para descrever as propriedades dos recursos<sup>1</sup>, condições de uso,

estratégias de preservação, sua localização, entre outros. Já

Na Web Semântica, metadados são usados para descrever o significado dos recursos Web, e agentes inteligentes, desenvolvidos através de técnicas de Inteligência Artificial, usam essas descrições para auxiliar usuários na web na localização e manipulação. (p. 117).

Importante salientar que metadados, no contexto da web, são informações compreensíveis por **máquinas** (W3C, 2017). Daí a analogia a sua vida secreta, bem como com um novo paradigma do sentido (agenciado também por máquinas), neste artigo.

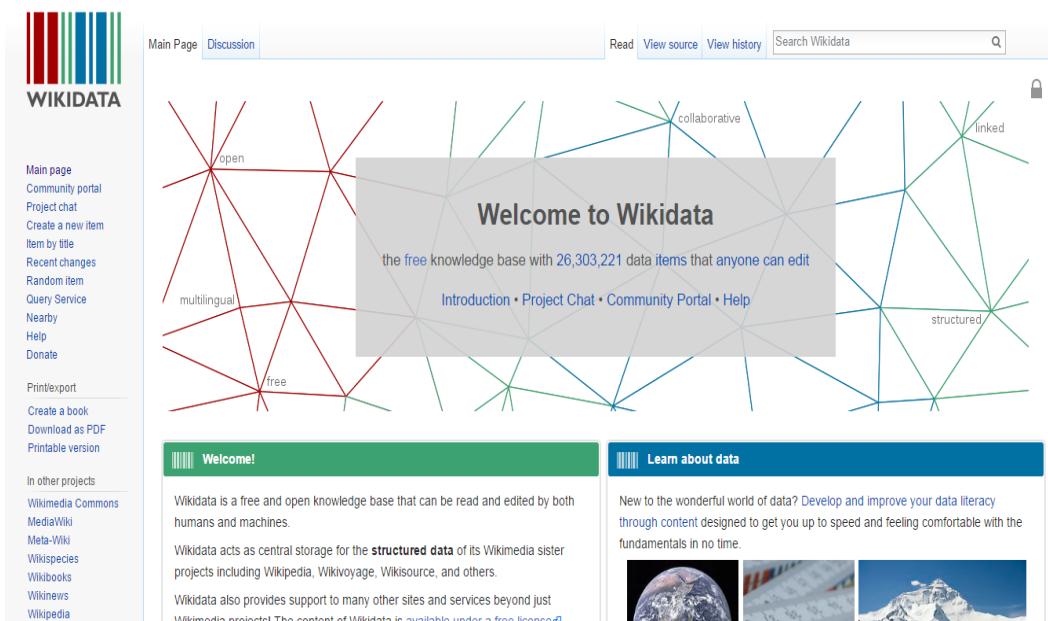
Transpondo essa realidade para o Wikidata, esta seção intenta demonstrar a criação de metadados, a partir da predicação (propriedades) das entidades (itens), ou seja, atribuindo referências “às coisas” para instauração de uma teia de sentidos.

O Wikidata é um projeto, desde 2006, da Fundação Wikimedia, considerada “Wikipédia para dados”, por Vrandečić e Krötzsch (2014). Trata-se de uma base de dados (secundário) livre, colaborativo e multilíngue de dados estruturados para fornecer suporte para a Wikipédia, Wikimedia Commons e outros projetos da Wikimedia; também denominado, na literatura internacional, *knowledge bases* ou *knowledgedatabases*.

<sup>1</sup> Recurso, na WS, é qualquer objeto da web identificado por meio de uma URI (*Uniform Resource Identifier*) (ROCHA, 2004).



Figura 1: Wikidata



Fonte: Wikidata (2017)

De acordo com Perez (2012), o projeto é financiado pelo Instituto Allen (para a Inteligência Artificial, EUA), pela Fundação Gordon e Betty Moore (EUA), pelo Google e desenvolvido pelo braço alemão da Wikimedia Foundation, especialmente pelos professores Marcus Krötzsch (Universidade de Oxford, Inglaterra) e Denny Vrandečić (Instituto de Tecnologia de Karlsruhe, Alemanha).

O atributo “secundário” diz respeito ao tipo de fonte de informação em que se constitui o Wikidata, isto é, não publica investigações, então é importante salientar que um sujeito que tiver acesso ao Wikidata não terá um conteúdo referencial de chofre, mas a um repositório wiki de itens (extraídos da Wikipédia) para estruturar (descrever) dados, constituindo os “metadados”, editáveis por humanos e máquinas.

Para Vrandečić e Krötzsch (2014), o Wikidata tem as seguintes abordagens: edição aberta; controle comunitário, pluralidade; dados secundários; dados multilíngues; fácil acesso; evolução contínua.

No entanto, é uma base de dados em que as informações são direcionadas para aqueles que tenham conhecimento sobre WS e são legíveis por tecnologias de captura e visualização,

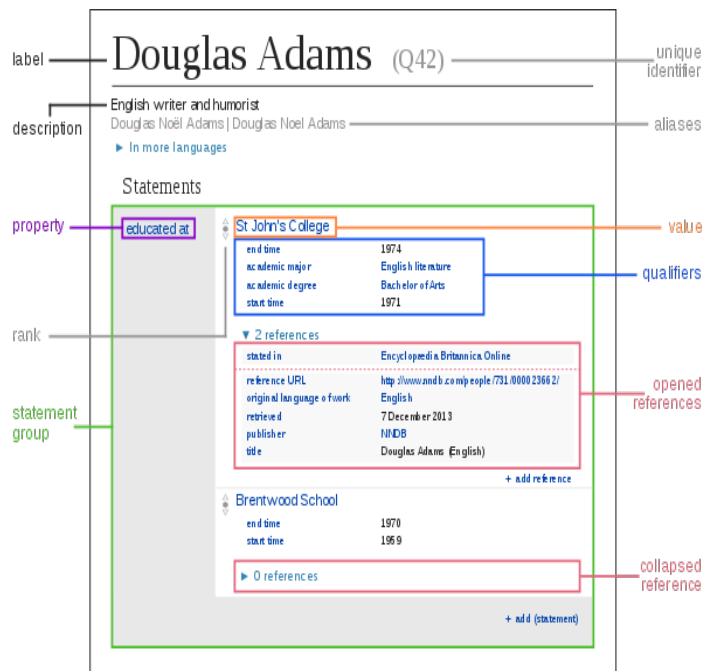
desenvolvidas atualmente pelos grandes serviços de busca, como o Google, mas também para ajudar a própria Wikipédia, tornando seus verbetes contextualizados. “Seguindo essa filosofia, agentes automatizados podem realizar tarefas para os usuários, usando metadados semânticos.” (PINHEIRO, p. 23, 2009).

Observa-se que, quando se menciona Web Semântica, neste artigo, contempla-se somente a camada “esquema”, ou seja, a parte da descrição dos dados, e não as camadas ontológica e lógica, e suas respectivas tecnologias, conforme a famosa representação visual do *layer cake* (bolo de camadas) da WS.

De acordo com o *site*, as informações constituem-se em declarações e não em estruturas RDF (*Resource Description Framework*<sup>2</sup>), mas é capaz de exportar os dados nesse formato (WIKIDATA, 2017). O exemplo da Figura 2 ilustra o modo de estruturação de um item.

2 “O RDF é um padrão de metadados que evoluiu a partir do XML, definindo um modelo baseado em recursos, propriedades e valores.” (ALMEIDA; SOUZA, 2011, p. 42).

Figura 2: Exemplo de estruturação de dados no Wikidata



Fonte: KRITSCHMAR, Charlie, 2017

A importância da SF no ambiente da WS é que ela promove o deslocamento do significado do nível da palavra para o nível da sentença, uma vez que permite a aplicação da noção de proposição (ALMEIDA; SOUZA, 2011), e assim declarações podem ser construídas.

Cada declaração (*statement*) consiste de uma propriedade (ou mais) atribuída ao item (*label*) para o estabelecimento do valor (*value*), construindo assim, uma declaração. Os metadados são constituídos por:

- 1) **itens:** são representações do conhecimento e são flexíveis o bastante para representar conceitos como: infância, fome, peso, bem como objetos do mundo real como: “televisão”, “kayak” e “vulcão”. Os itens são unidades básicas que representam o conhecimento. Cada item tem uma página própria e um identificador; no caso do Diagrama 1, Q42;
- 2) **rótulos:** os nomes dos itens são chamados de etiquetas (*label*) e devem ser adicionados a todas as páginas de itens;
- 3) **descrição:** descrição não é declaração, mas uma breve explicação sobre o item e deve ser curta: no “caso escritor e humorista inglês”;

- 4) **nomes alternativos:** denominados *aliases*, sem uma tradução fidedigna para o português, confere ao item de conhecimento a desambiguação: “Douglas Noël Adams ou Douglas Noel Adams”;
- 5) **propriedades:** propriedade tem um único nome, distinto, como cor, cidadania, parentesco, acidente geográfico, entre outros. A propriedade limita e define que tipo de valores de dados que podem ser adicionados a um item;
- 6) **valor:** é o preenchimento da propriedade. Por exemplo, a propriedade “data de nascimento” requer um valor numérico e não verbal. Esses parâmetros impedem o vandalismo no Wikidata;
- 7) **qualificadores:** são outras informações (ou mesmo valores) que servem como parâmetros para a especificação de um dado;
- 8) **referências:** o Wikidata armazena não só suas instruções, mas suas fontes, o que reflete a diversidade do conhecimento disponível apoiando a noção de verificabilidade;  
Pode-se, a partir do exemplo da Figura 2, depreender a seguinte declaração:

Douglas Adams (ITEM)

- educação (PROPRIEDADE) – St. John’s College (VALOR) 1971-1974 (qualificadores).

“Douglas Adams [ITEM] estudou [PROPRIEDADE] no St. John’s College [VALOR] de 1971 A 1974 [qualificadores]”

A chave da compreensão dos metadados do Wikidata é a “ligação” ou comparando com a SF, os preenchimentos dos espaços vazios, ou seja, pode-se preencher qualquer propriedade [cor, parentesco...] ao item e assim definir o valor [que é determinado pela propriedade]. As propriedades são, na verdade, relações (estudou em, marido de, filho de, pai de, nascido em, e assim por diante).

Se na SF os preenchimentos do predicado se dão pelos “argumentos”, no Wikidata faz-se pela “propriedade” e “valor”. No primeiro caso, formam-se as sentenças, no segundo, as declarações, como a supracitada. Em ambos os casos, várias são as propriedades ou relações que podem ser atribuídas ao sujeito (diferentemente da Lógica Formal).

Os itens são representações do conhecimento flexíveis (ou expressões insaturadas) o suficientemente para designar conceitos abstratos ou reais: pode-se ligar (relacionar) uma coisa à outra. Nesse sentido, a pertença à Lógica está na vinculação, na atribuição de uma propriedade para formulação de uma definição ou declaração, a partir de um valor.

A fim de confirmar como as propriedades são espaços vazios e flexíveis, os modos de predicação do exemplo, supracitado na Figura 1, “Douglas Adams”, podem ser preenchidos por: imagem; sexo ou gênero; país de cidadania; nome de nascimento; nome no idioma nativo; prenome; sobrenome; data de nascimento; local de nascimento; local de falecimento; circunstância da morte; causa da morte; local do enterro; imagem do túmulo; pai; mãe; irmã ou irmão; cônjuge; descendência; língua materna; línguas faladas; ocupação; empregador; nominada a (prêmios); *alma mater*; residência; grupo étnico; religião; altura; instrumento; obra; *website* oficial; categoria *commons*; identificadores.

Esses metadados são editáveis, por humanos e máquinas, capturados e visualizados,

em forma de expressão ou *query*, por agentes computacionais, a partir de pesquisas dos mecanismos de busca. Assim, esses itens (que estão circundados por uma rede metadados, ou seja, de **referências e relações**) ficam “prenhes” de **sentidos**.

Nesse sentido, Bax (2012, p. 11), em seu denso artigo, informa que “A Web Semântica representa um avanço por permitir explicitar a distinção existente entre sentido e referência. A referência passa ser o objeto em si, presente no mundo, e o sentido é a descrição RDF desse objeto.”

No entanto, o autor problematiza a própria natureza do sentido na WS. Há posições teóricas antagônicas, dentro da WC3<sup>3</sup> (*World Wide Web Consortium*): primeira, como descrição lógica (RDF e a teoria das descrições); segunda, como apontamento direto à coisa a partir da URI (e a teoria causal da referência).

Bax (2012, p.22) apresenta e também se posiciona a favor de uma terceira posição teórica, o sentido como noção objetiva e pública, à esteira de Wittgenstein (linguagem como uso):

[...] o significado de qualquer termo é dado pelo seu uso social. Assim o significado da URI é ‘seu uso’ por uma comunidade de agentes (humanos e máquinas) e não apenas por uma referência direta, nem apenas por uma teoria formal de modelos (razão pura).

Como o foco deste artigo são os metadados do Wikidata, a abordagem recaiu sobre a posição formal, ou seja, o sentido estaria nas declarações lógicas, naquilo que as relações (propriedades) designam sobre a coisa (referente), embora se considere que a Web Semiótica (ou Pragmática) é o plano ideal da WS, isto é, permite a operacionalização da dimensão pragmática da linguagem e a compreensão do sentido no ciberespaço (MONTEIRO; MOURA, 2016).

<sup>3</sup> A primeira posição é defendida por Patrick Hayes, a segunda, por Berners-Lee e a terceira, por H. Halpin e Bax.



## 4 CONSIDERAÇÕES FINAIS

Buscar responder as questões da abertura deste artigo implica em pensar no paradigma pós-moderno da Semântica.

Para respondê-las, o artigo buscou, mesmo que brevemente, trazer uma síntese das vertentes da Semântica, enfocando a SF, na Linguística, mas também uma importante síntese, constante no Quadro 1, de Almeida e Souza (2011), das abordagens da Semântica nos sistemas computacionais e na WS. Verificou-se que nesse ambiente, a SF é prevalente.

No sentido tecnológico da WS, a Semântica visa à construção de um modelo que reproduza a capacidade interpretativa que um interpretante tem. Na SF, o objetivo é reproduzir a capacidade semântica que um falante tem de sua língua.

Na primeira acepção foi usado o termo “interpretante” e na segunda, “falante” para destacar as duas diferenças paradigmáticas: 1ª) interpretante, no sentido semiótico, não significa somente a mente humana, mas qualquer uma que possa, logicamente, interpretar (no caso, a WS, os agentes computacionais); 2ª) falante, nas concepções clássicas da Semântica, contempla apenas a interpretação humana, embora os modelos clássicos da SF já apontavam para a possibilidade de construção de metalinguagens computacionais.

No entanto, ambas têm como objetivo a construção de uma máquina Semântica.

Para compreender os metadados, o Wikidata foi estudado, visando ilustrar o processo estruturação da uma sentença, a partir da predicação. Em sua vida secreta, eles ganham significação (sentido e referência) nas camadas invisíveis (ou não visíveis para o sujeito comum) existentes nas bases de dados ou conhecimento. Eles são importantes para a construção do sentido, por agentes humanos e máqunicos, em busca de informação nos mecanismos de recuperação.

Lévy (2014) reconhece que os metadados são uma metalinguagem que reflete a mutação epistemológica das ciências. Ela permite a explicitação e a interconexão das ideias e dos dados e deve ter duas características: a computabilidade; o poder de exprimir as relações de significação complexas. Enfim, são fenômenos sociotécnicos de significação que ocorrem no ciberespaço.

Para concluir, Oliveira (2012) alerta que a SF é um campo vasto e heterogêneo, ou seja, há muitas maneiras de descrever o dado linguístico, no entanto, o que deve ser aceito, no estudo em questão, é que as relações de significado (sentido e referência) devem ser descritos formalmente, e esse é o modelo de Semântica na WS.

---

Artigo recebido em 08/06/2017 e aceito para publicação em 29/01/2018

---

### **THE SECRET LIFE OF METADATA ON WIKIDATA: a focus on meaning in (Web) Formal Semantics**

#### **ABSTRACT**

*Since the Semantic Web announcement, to date, many technologies and innovations have been developed to give body and effectiveness to the concept. This web of meanings consists of languages and semantic technologies, such as XML (eXtensible Markup Language), RDF (Resource Description Framework), Declarations, among others that structure the data that roam through cyberspace, forming great knowledge bases, such as Wikidata (Wikimedia Foundation). The main logic of the Semantic Web is to describe the entities (things) of the real world and link them to others, so that this network of meanings can be activated, retrieved and interpreted by machines and humans. The research of documentary delineation had the Wikidata as corpus of research. The research aimed to examine this bases for apprehending the structuring of metadata to answer the question: what kind of Semantics is being produced in a knowledge bases? As a result, it has been observed that Formal Semantics is the foundation for the generation of metadata, however, in the postmodern paradigm, meaning - and Semantics itself - are considered a sociotechnical phenomenon in cyberspace.*

**Keywords:** Wikidata. Metadata. Formal Semantics. Web Semantics

## REFERÊNCIAS

- ALMEIDA, Maurício Barcellos; SOUZA, Renato Souza. Avaliação do espectro semântico de instrumentos para organização da informação. **Enc. Bibli.** R. Eletr. Bibliotecon. Ci. Inf., v. 16, n. 31, p. 25-50, 2011. Disponível em: < <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2011v16n31p25/19441> >. Acesso em: 11 jan. 2017.
- BAX, Marcelo. A evolução da Web rumo à Web Semântica. **Prisma.com**. Revista de Ciências e Tecnologias de Informação e Comunicação do CETAC.MEDIA, n. 19, p. 1-27, 2012. Disponível em: <<http://revistas.ua.pt/index.php/prismacom/article/view/2411>>. Acesso em: 26 maio 2014.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. **The Semantic Web**. 2001. Disponível em:<[http://isel2918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01\\_theSemanticWeb.pdf](http://isel2918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01_theSemanticWeb.pdf)>.
- Acesso em: 18 mar. 2017.
- BLIKSTEIN, Isidoro. **Kaspar Hauser ou a fabricação da realidade**. São Paulo: Cultrix, 1995.
- BRÉAL, Michel. **Ensaio de Semântica**. São Paulo: Educ; Pontes, 1992.
- DUCROT, Oswald. **O dito e o não dito**. Campinas, S.P: Pontes, 1987.
- GREIMAS, A. J. **Sémantique structurale**. Paris: Larousse, 1966.
- ECO, Umberto. **Da árvore ao labirinto: estudos históricos sobre o signo e a interpretação**. Rio de Janeiro: Record, 2013.
- KRITSCHMAR, Charlie (WMDE). **Como funciona o Wikidata?** Wikidata. Disponível em: <<https://commons.wikimedia.org/w/index.php?curid=49616867>> Acesso em: 08 fev. 2017.
- ODGEN, C. K.; RICHARDAS, I. A. **O significado do significado**. Rio de: Zahar, 1972.
- OLIVEIRA, Roberta Pires de. **Semântica formal**. 3.ed. Mercado de Letras, 2012.
- LÉVY, Pierre. **A esfera semântica: computação, cognição, economia da informação**. São Paulo: Annablume, 2014.
- LAKOFF, G.; JOHNSON, M. **Metaphors we live by**. Chicago/London: The University of Chicago Press, 1980.
- MONTEIRO, Silvana Drumond; MOURA, Maria Aparecida. Knowledge Graph e Peirce: uma abordagem semiótica dos índices contemporâneos no ciberespaço. In: TOMAÉL, Maria Inês; ALCARÁ, Adriana R. (Org.). **Fontes de informação digital**. Londrina: EDUEL, 2016. p. 119-146.
- PARTEE, B.H. Some structural analogies between tenses and pronouns in English. **Journal of Philosophy**, n. 70, p. 601-609, 1973.
- PEREZ, Sarah. Wikipedia's next big thing: Wikidata, a machine-readable, user editable database funded by Google, Paul Allen And Others. Disponível em: <<https://techcrunch.com/2012/03/30/wikipedias-next-big-thing-wikidata-a-machine-readable-user-editable-database-funded-by-google-paul-allen-and-others/>>. Acesso em: 08 fev. 2017.
- PINHEIRO, José Maurício dos Santos. Web Semântica: uma rede de conceitos. **Cadernos UniFOA**, n. 9, p. 23-27, abr. 2009. Disponível em: <<http://web.unifoa.edu.br/cadernos/edicao/09/23.pdf> >. Acesso em 31 jun. 2017.
- ROCHA, Rafael Port da. Metadados, Web Semântica, categorização automática: combinando esforços e computacionais para a descoberta e uso dos recursos da web. **Em questão**, v. 10, n. 1, p. 109-121, jan./jun. 2004. Disponível em: < <http://seer.ufrgs.br/index.php/EmQuestao/article/view/86/1127>>. Acesso em: 26 maio 2017.
- TABA, Irène. **A Semântica**. 2.ed. São Paulo: Parábola, 2006.
- A VIDA secreta das palavras. Direção de Isabel COIXET. Espanha, 2005. 125 min
- VLANDEČIĆ, Denny; KRÖTZSCH, Marcus. Wikidata: a free collaborative knowledgebase. **Communications of the ACM**, v. 257, p. 78-85, October, 2014.
- W3C. **About W3C**. Disponível em: <<https://www.w3.org/Consortium/>> . Acesso em 28 maio 2017.
- WIKIDATA. **Welcome to Wikidata**. Disponível em: <[https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page) >. Acesso em: 26 abr. 2017.