DE GRUYTER MOUTON
DOI 10.1515/sem-2013-0007 — Semiotica 2013; 193: 119–143

Catherine Legg

Peirce, meaning, and the Semantic Web

Abstract: This paper seeks an explanation for the challenges faced by Semantic Web developers in achieving their vision, compared to the staggering near-instantaneous success of the World Wide Web. To this end it contrasts two broad philosophical understandings of meaning and argues that the choice between them carries real consequences for how developers attempt to engineer the Semantic Web. The first is Rene Descartes' "private," static account of meaning (arguably dominant for the last four-hundred years in Western thought), which understands the meanings of signs as whatever their producers intend them to mean. The second is Charles Peirce's still relatively unknown "public," evolutionary account of meaning, according to which the meaning of signs just is the way they are interpreted and used to produce further signs. It is argued that only the latter approach can avoid the unmanageable attempts to "preprocess" interpretation of signs on the Web that have dogged the project in its many stages, and thereby do justice to the scale, rapid changeability, and exciting possibilities of online information today.

Keywords: pragmatism; Semantic Web; Peirce; Descartes; meaning; interpretant

Catherine Legg: University of Waikato. E-mail: clegg@waikato.ac.nz

1 Introduction

When the World Wide Web (henceforth: WWW) first emerged, semioticians were quick to flag it as an amazing semiotic development. Examples include Shank and Cunningham (1996), Nadin (1998), who speaks of the internet as "semiotic engineering," Neumüller (2000), who writes that "[s]emiotics promises a broader understanding of Hypertext as a structure of signs" (2000: 57), and Santaella-Braga (2004), who claims that the internet has made possible "an entirely new idea of electronic online publication" that renders the computer well on the way to becoming the semiotic "medium of all media."

Less often commented on from a semiotic perspective are the specific details of the Web's development since then, in particular the ambitious project known tantalizingly as *The Semantic Web*. It is little known outside IT circles that when

Tim Berners-Lee first conceived the WWW with its structure of URLs, hyperlinks, and the HTML language, this was only the first stage of his vision for the Web. He claimed that it would shortly thereafter evolve into a much more sophisticated entity, replacing a mere *web of links* with a *web of meaning* (Berners-Lee et al. 2001; Berners-Lee 2002). Vast software engineering work has been devoted to realizing this latter vision, on the assumption that the first stage was so spectacularly successful that the second was worth betting on. Thus there are now at least three major international conferences,[1] and two special-purpose journals[2] devoted entirely to the Semantic Web. The World Wide Web Consortium (W3C)[3] currently has over 350 member organizations, from academia, government, and private industry, and developing the Semantic Web is one of its three long-term goals (the other two being *universal access* and "*a web of trust*" [Halpin 2004]). But despite this generous investment, so far the Semantic Web seems to have lacked the unstoppable growth and momentum of the early WWW. Even a 2006 paper co-authored by Berners-Lee himself acknowledges, "It's still apparent that the Semantic Web isn't yet with us on any scale" (Shadbolt et al 2006: 97).

The contrast between the staggering near-instantaneous success of the WWW and the Semantic Web's sluggish start raises intriguing questions for the IT-literate semiotician. One notable question is: *What does it mean to* "*give*" *Web pages meaning?* The question is general and philosophical, but the IT context renders investigation of it fascinatingly concrete. It thus supplies an ideal opportunity to apply Charles Peirce's *Pragmatic Maxim* (Peirce 1878), which urges that to better understand abstract concepts, such as *meaning*, it is most helpful to think about their specific effects in specific applications.

Many complex and technical discussions of meaning have taken place in Western philosophy over the past four-hundred years. I will argue however that such debates have almost all shared a basic set of assumptions that is most unhelpful from an engineering perspective. I will call these assumptions *The Cartesian Framework for Understanding Meaning*, and argue that the terrain lying beyond them is only just being glimpsed (and understood as inviting) with the help of Peirce and his unique and still underappreciated theory of signs.

The paper will begin by outlining a Cartesian "private and static" conception of meaning (section 2), then sketch a Peircean "public and dynamic" alternative (section 3). These two understandings of meaning will then be traced to different forms of semiotic engineering which are actually steering current Semantic Web

1 *ISWC, ESWC, SWWS.*

2 The *Journal of Web Semantics*, the *International Journal on Semantic Web and Information Systems*.

3 http://www.w3c.org.

research (sections 4 and 5), and it will be argued that the Peircean approach is producing far superior results.

2 The Cartesian framework for understanding meaning

The influence of Descartes on modern thinking about thought and language cannot be overstated, whether or not he himself always adhered clearly and distinctly to the new paradigm he inspired. The key Cartesian idea for our purposes is that *the meaning of a sign is fully determined by the intention of its producer.* This "intention" has two main features:

1. It is *private*. It has a location somehow "in" a person's mind. Despite the spatial terminology, the key issue is that only the sign's producer has true epistemic access to its meaning. For Descartes, the intention was so inaccessible to external observation as to constitute a non-physical substance – hence the famous "Cartesian dualism."
2. It is *incorrigible*. The signs I produce mean all and only what I intend them to mean. This is sometimes referred to as a claim of "first-person authority" with respect to meaning.

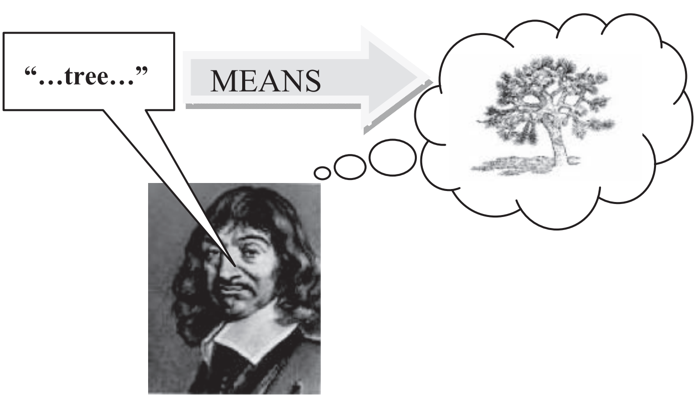

Fig. 1: Cartesian dyadic theory of meaning

Although Descartes does not discuss meaning explicitly in his *Meditations* (1996 [1647]), these claims derive from what he says about *ideas*, which for him are the basic building-blocks of thought and meaning. For instance, in Meditation III he claims that we only have direct access to our ideas, that things in the world are ontologically separate from the ideas that accurately or falsely represent them. In other words, methodologically the sign-user's mind is separate from the "external" world, so much so that he claimed to coherently doubt whether the latter even exists. The sign-user's mind must therefore be the ultimate authority on what its ideas mean, on pain of complete ignorance. Error is possible, but not about what one's ideas *mean*, only about the way they are put together, or insofar as we assume that our simple ideas resemble reality:

> When ideas are considered solely in themselves and not taken to be connected to anything else, they can't be false; for whether it is a goat that I am imagining or a chimera, either way it is true that I do imagine it . . . the only kind of thought where I must watch out for mistakes – are judgments. And the mistake they most commonly involve is to judge that my ideas resemble things outside me. (Descartes 1996 [1647])

Of course later thinkers in the so-called "early modern" period such as Locke and Hume embraced a naturalistic empiricism, and abandoned Descartes' dualistic understanding of mind as a separate substance from matter. However they (arguably inconsistently) retained his concept of the private and incorrigible idea as the basic unit of meaning. Thus Locke famously stated: "[W]ords, in their primary or immediate signification, stand for nothing but the ideas in the mind of him that uses them . . ." (Locke 1994: 3, II, ii).

In the nineteenth century, Gottlob Frege rejected the early modern understanding of meaning in terms of ideas. He pointed out that any word can be associated with many different ideas in the minds of different people. Frege wanted to distinguish between how people actually *do* understand the meaning of a sign and how they *should* understand it in order to grasp the truth. He therefore claimed (1952 [1892]) that associated with every unambiguous term was not only ideas, but also a sense (*Sinn*). This sense was an abstract object, understood by everyone who grasps the term's meaning. He sometimes referred to it as the "mode of presentation" of the sign's reference. Thus Frege gave up the *privacy* of the Cartesian model of meaning. However the internalism of the phrase "mode of presentation" (i.e., to the mind) seems to suggest that he retained the *incorrigibility*.[4]

Frege dreamed that his new "concept-script," which introduced the predicate calculus to the world, might enable a newly clear and objective understanding of the meaning of all signs. He hoped that it would then be possible to build human

4 This point is made at greater length in Legg (2005).

knowledge, starting with arithmetic, into an integrated system that was *deductively complete*. This dream was shattered by what is known as "Russell's Paradox."[5] Nevertheless Frege's insights helped to shape twentieth century philosophy's so-called "linguistic turn," which (to put it crudely) shifted from seeing meaning as an *idea-world relationship* to seeing it as a *word-world relationship* (Hacking 1975). A series of new "word-world" theories of meaning were explored, with many variations. For instance, Quine (who tellingly titled a key book *Word and Object* [1960]) attempted to behavioristically reduce the concept of meaning to the "stimuli" that would make someone assent to a given sentence. The later Wittgenstein (1953), while strongly opposed to such reductionism, also reacted against Frege's view of meaning as abstract object, giving birth to the view of meaning as "use," which has been influential in neo-pragmatist circles, among others. Although complaining, "how much the general concept of the meaning of a word surrounds the working of language with a haze which makes clear vision impossible," Wittgenstein claimed that the answer to this problem is to "study the phenomena of language in primitive kinds of use in which one can clearly survey the purpose and functioning of the words" (1953: §5). Later, Donald Davidson developed an account of the meaning of propositions in terms of their *truth-conditions*, which was then vastly complicated and sophisticated *via* the technical concept of possible worlds, spawning a plethora of descendent theories of meaning in contemporary analytic philosophy.

Meanwhile on the so-called "Continental" side, Saussure developed a theory of signs also structured around a key dyadic relation – between "signifier" and "signified" – *albeit* mitigated by a structuralism that taught that the meaning of any term (for instance "happy") greatly depends on its contrast with salient siblings (for instance "blissful," "distressed," and so on . . .). This opening up of the sign's meaning to influences beyond the signifier-signified pair led on to post-structuralism's challenge to the signified – its existence and its theoretical utility, which went so far as to, if not abolish it, at least defer it indefinitely (Derrida 1976, 1978). To sum up twentieth century developments, then, although they largely abandon the privacy of the Cartesian model, with the possible exception of Quine's behaviorism they offer little challenge to the first-person authority. It is also worth pausing to note the overwhelming focus in these theories on *words* as the sole repository of meaning. We will have cause to critique this later.

5 Misleadingly so known, according to Peirce, who claimed that the paradox's key idea was already to be found in the work of Cantor. Some backing for Peirce's claim is to be found in Kleen (1952: 37).

3 A Peircean alternative framework for understanding meaning

The key Peircean idea for our purposes is that *the meaning of a sign is the process of interpretation that occurs as the sign is used*.[6] Peirce denied both privacy *and* incorrigibility of meaning.

By understanding meaning as a relationship between a sign and a thing in the world (whether the sign is thought of as an idea as in the early modern period, or word(s) as in the twentieth century), the Cartesian framework can be attributed an essentially *dyadic structure*.[7] Peirce's *triadic* model of the sign, by contrast, consists in an irreducible relationship between three elements: i) the actual signifying item (*representamen*), ii) an *object*, what the sign refers to in the world, iii) an *interpretant,* which is just further uses of the *same representamen* to represent the *same object*. This is just to say that a sign must represent an object in such a way that it (the sign) is able to be understood and used again. If I decide to name a new tree species, this baptism will not create a new living sign unless other people learn my name and use it. Otherwise my act is literally meaningless, whatever my intentions.

The arrows in Figure 2 represent an "order of semiotic determination": the object brings about the sign through some interaction with a mind, and the sign brings about the interpretant in its shaping of the mind's further thoughts. In this way, the sign may be said to *mediate* between the object and the interpretant, although it is important to note that, as Santaella-Braga has neatly put it, "[a] lthough the sign is determined by the object, the latter is logically accessible only through the mediation of the sign" (2004: 16).

This formal model has been thoroughly explicated by Peirce scholars (e.g., Liszka 1996; Short 2007). However for our purposes it is useful to highlight that although downstream interpretants pick out the same object as the original sign, they can interpret it differently. They may not just continue but also augment or

---

6 Some Peirce scholars would argue that this formulation misses a key modal dimension, particularly to Peirce's mature semiotics. In other words, a sign's meaning should be thought to include *possible* as well as *actual* interpretations of it. Against this I would argue that Peirce's pragmatist theory of meaning draws the possible back into the actual, insofar as if a sign is *never* used in a particular way in *any* circumstance the claim that it is us*able* begins to lack meaning.

7 It is worth noting that Frege *nearly* escapes this dyadicity by postulating a sense as well as a reference for every sign. However given that sense for him is an abstract object that is *sui generis*, logically speaking he has arguably replaced a single dyadic relationship with two dyadic relationships.

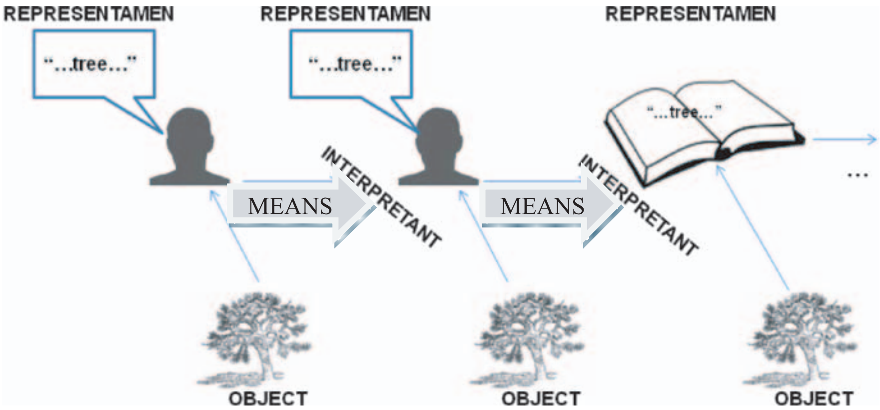

Fig. 2: Peircean triadic theory of meaning

even *shift* the object to which the sign refers. A useful example is the word *atom* as used by Democritus, and by us. Etymologically, in ancient Greek "a-tom" meant something that cannot be decomposed, but of course we have now "split the atom." Yet in *some* sense today's physicists are arguably still talking about the same things Democritus was, and the transition from the ancient to the present meaning was a slow, gradual shift. Thus by contrast to the Cartesian framework, we now have *corrigibility* with respect to meaning – when Democritus talked of atoms, he meant more than he knew.

One might ask at this point: So what is the *real* meaning of the sign? The original or the ultimate interpretation (interpretations that pick out what Peirce termed an "immediate object" and a "dynamic object" respectively)? However, is it necessary to choose? Peirce's theory raises the possibility that it is not. Arguably now it is more helpful to understand meaning not as an *object* (whose "properties" can be argued over and had better not be contradictory), but rather as a *process*. In some real sense the meaning of a sign is what that sign *does* – how it spreads and grows (if, indeed, it does). Thus, Peirce famously wrote: "[N]o present actual thought (which is a mere feeling) has any meaning, any intellectual value; for this lies not in what is actually thought, but in what this thought may be connected with in representation by subsequent thoughts; so that the meaning of a thought is altogether something virtual" (Peirce 1868).

In other words, not only do signs mediate between objects and interpretants, such mediation is the life-blood of meaning-creation. Without it all signification would cease. (In that sense, there is no "going straight to the point.")

Note how this account renders meaning *public*. In the Cartesian framework, to really know what a sign means, one would need to enter the head of its producer and “read” their ideas, an impossible task. In the Peircean framework, to know what a sign means, one merely needs to look at what people are *doing* with it, how it is actually being interpreted. In this way the responsibility for the meaning of a sign resides in the entire sign-using community, and this community, it should be noted, is open-ended as to the future. Relatedly, Peirce derived his account of *truth* by idealizing over the process of developing and using signs in a *community of inquiry*, writing, “The opinion which is fated to be ultimately agreed to by all who investigate, is what we mean by the truth” (Peirce 1878). Such pragmatist accounts of truth have been widely criticized as insufficiently objective, and excessively optimistic about human epistemic virtues.[8] Yet Peirce denied this, claiming that over the long-run, within a broad enough community, sign-use was intrinsically self-correcting. He also claimed (in a Kantian spirit) that the existence of a public community of inquiry was a necessary precondition for any truth-seeking (*CP* 2.654).

It is also important to note that what the community is “doing with” a given sign is not just what they are “doing with it in their heads” by thinking about it, but what kinds of *practical activities* they are scaffolding with its help. Consider for instance our term *tree*. For Peirce it is part of the very meaning of this term that people plant trees, chop them down, decorate and display them at Christmas, and so on. For this reason, as the other (independent) co-founder of predicate logic, Peirce pursued a vision interestingly different from Frege’s regarding how this powerful new system should be used to advance human knowledge. *Qua* pragmatist, Peirce thought Frege’s attempt to explicitly formalize the entire meaning of signs impossible, for an irreducible dimension of the meaning of any sign is the effects that an agent situated in the world would experience in relevant situations, such as tree-climbing and botanical investigations, not all of which can be anticipated in advance. In summary, then, Peirce replaces a static model of meaning (*meaning-as-object*) with a dynamic model (*meaning-as-process*), where what a sign means is open to view (public), able to shift and change over time (corrigible), and inextricably entwined with actual tasks and projects.

As noted, the goal of the Semantic Web is to allow computers to “become much better able to process and ‘understand’ the data that they merely display at present” (Berners-Lee et al. 2001). How is this to be achieved? Semantic Web researchers all agree that this will somehow involve marking up web pages with metadata to enable Web searches to access “semantically enriched” material. (“The goal is to create annotations with well-defined semantics, however those semantics may be defined” [Uren et al. 2006: 14]). Beyond that however, many

8 See for instance, the influential Quine (1960: 23).

questions remain. For instance: How exactly are these annotations to be used? And most importantly, how are they to be given a *well-defined semantics*? We will now examine this second question through the lenses of the Cartesian and the Peircean frameworks respectively.

4 A Cartesian approach to web semantics

If one believes that the meaning of a sign is determined by the (private, incorrigible) *intention* of its producer, then it stands to reason that if one wishes to "give" webpages meaning, one will *try to define an authoritative, unambiguous, machine-readable intention* for them, as fully and determinately as possible. Almost all early and some current Semantic Web research can be read as a fascinating series of just such attempts. For example, consider a website dedicated to the preservation of New Zealand kauri trees. The Semantic Web aims to create metadata to somehow capture what this website "is about," and serve it up in machine-readable form. But how is this to be done? We will now look at some specific attempts, and their limitations. The example is useful as it already foregrounds one significant pitfall in moving from syntax to semantics – *homonymy*. For there is a concept of "tree" in both biology and mathematics and in this case we want to focus cleanly on the former.

4.1 XML

The first and most obvious method to attempt to create semantic metadata was to generalize the tags of HTML, taking a cue from Berners-Lee's statement, "The Semantic Web is what we will get if we perform the same globalization process to Knowledge representation that the Web initially did to Hypertext" (1998). Whereas HTML users are restricted to a few simple predefined tags (e.g., <b>: bold, <hr>: line return), why not allow users to name their own tags after key human concepts (e.g., <tree>, <person>)? This was one of the defining ideas of XML: the eXtensible Markup Language (Bray et al. 2000).[9]

How are these user-generated tags to be defined and shared? It was decided that each new tag would be indexed to a defining *namespace*, identified by a Universal Resource Indicator (URI), a generalization of the URL. One could then

9 Although it should be noted that XML also grew out of SGML, a language predating the Semantic Web effort and designed to enable users sharing documents to design and transmit their own *document format*.

define <tree> in a biological namespace using the following kind of statement (the URI is invented):

xmlns:bio="http://www.biologyXML.org/definitions"

At the same time one could define it in a mathematical namespace as follows:

xmlns:math="http://www.MathKnowledgeOnSemanticWeb.org/trees"

The tags would be prefixed in use with the appropriate namespaces (<bio:tree> <math:tree> respectively), leaving no chance of confusing them. One could then scatter tags in our web page on kauri preservation such as <bio:tree>, <geo:New Zealand flora> and so on, and this, it was envisaged would provide some automatically accessible and indexable clue as to its meaning.

Early Semantic Web developers were enthusiastic that XML would allow them to define meaning in machine-understandable terms:

> [C]an we really build intelligence into the 3 billion or 10 billion documents that make up the Web? . . . The first step is to get a fulcrum under the mountain and lift it, and it is well under way. That fulcrum is the extensible markup language (XML). A sort of HTML-on-steroids, this coding system isolates, under the hood, the dozens or even hundreds of data elements a Web page might contain. Right now, HTML coding serves mostly to control the appearance and arrangement of the text and images on a Web page . . . With new XML tags, <price>, for instance, a software agent might be able to, for example, comparison shop across different Web sites . . . (Cherry 2002)

However it is important to note that a tag's namespace does not have to include a *definition* for it. In fact, the URI does not even have to point to a webpage – even if it does, the page may be blank. In practice namespaces often just serve as a means of *indexing* tags using unique prefixes. As a means of semantic differentiation, this is a fairly blunt instrument. It does distinguish homonyms, but not by virtue of explaining their different meanings. Meanwhile a converse, equally grave challenge as homynymy is *hidden synonymy*, e.g., <tree> and <arbre>, which as far as XML is concerned might as well be <tree> and <&*^HTY9>. In order to address this problem some way of translating between namespaces is required, and this is an unsolved problem in XML: "Additional effort is required in . . . generating mapping procedures for XML documents based on established domain mappings. Using a more suitable formalism than pure XML for data transfer can save much of this additional effort" (Decker et al. 2000: 69).

To some developers it seemed obvious that the way around such problems was to try to build the means to "validate" XML documents as conforming with *a* *shared set of meanings*. Surely there must be some way of codifying the biological meaning of tree, then checking whether an XML document conformed to it and not the mathematical meaning? Surely if only the right semantic standard could be specified, this would solve the problem? A popular approach to this task was to build structured entities that became known as *schemas*.

4.2 XML schemas

An XML schema is a further XML document that purports to encode a set of definitions or "grammar."[10] A given XML document is checked by applying a schema to it, which produces a further file (the "Post-Schema Validation Infoset"). This file lists the document's vocabulary (i.e., element and attribute names), content model (i.e., relationships, data structure and types), and checks that these are all present in allowable combinations and nestings. Thus an XML schema for botany might specify allowable tags, such as lists of known tree species and genera, along with rules that species must be subordinated to genera and not vice versa. But what if new botanical terms are developed, for instance as new species are discovered? If the schema is updated, then in a reprise of the referential opacity we saw when defining terms in different namespaces, the system cannot tell the difference between this and a whole new schema whose meaning is unknown. In the words of the W3C itself, "XML data is very sensitive to the XML Schema it refers to. If the XML Schema changes, the *same* XML data may become invalid, i.e., being rejected by Schema-aware parsers."[11]

Thus despite much early effort thrown into developing XML schemas by Semantic Web developers, it was realized that this would not solve the issues with using XML to build the Semantic Web. Essentially XML is hampered in defining Web semantics by its origins in describing document *format*. This is an essentially syntactic matter, including features such as the document's number of elements and the tree structure they fall into. Thus even the most sophisticated and complex XML schema can make no distinction between document *structure* and *content* (Decker et al. 2000: 66). Therefore developers tempered their expectations for what XML could achieve on the Semantic Web: "[XML is] useful for data interchange between parties that both know what the data is, but not for situations where new communications partners are frequently added" (Decker et al. 2000: 68).

10 The somewhat confusingly named XML Schema was the most popular XML schema, but there were others.
11 http://www.w3.org/RDF/FAQ

How then *was* a document's semantics to be isolated and defined? Some researchers suggested that what was needed to cleanly distinguish mere formatting information from genuine content was a markup system explicitly designed to model *propositional form*. This led to the development of RDF.

4.3 RDF

Around 2000, the hopes of Semantic Web developers turned to RDF, or "Resource Description Framework" (Lassila and Swick 1999; Beckett 2004). RDF is strictly speaking not a language but a data-model. The model consists of propositions each represented by a *triple* of *subject, predicate,* and *object*. Thus in the proposition, The Kauri is a kind of pine tree, the subject is Kauri, the predicate is kind of, and the object is pine tree. Triples may also have further triples as members, thereby allowing assertions about assertions. A formal semantics was defined for RDF (Hayes 2004) thereby aligning it with the resources of formal logic, most notably deductive inference, conveniently realizable on the essentially Boolean architecture of computers in the form of sophisticated theorem provers.

Developers became enthusiastic that RDF would enable them to define meaning in machine-understandable form and thereby implement the Semantic Web:

> It's wonderful that we can create URIs and talk about them with our web pages. However, it'd be even better if we could talk about them in a way that computers could begin to process what we're saying . . . RDF gives you a way to make statements that are machine-processable . . . For example, I could search the Web for all book reviews and create an average rating for each book. Then, I could put that information back on the Web. Another website could take that information (the list of book rating averages) and create a "Top Ten Highest Rated Books" page. (Swartz 2002)

When it comes to semantic interoperability, RDF has significant advantages over XML: The object-attribute structure provides natural semantic units because all objects are independent entities (Decker et al. 2000: 69).

Berners-Lee and other core developers formalized the role of RDF on a famous "layer-cake" diagram of the Semantic Web, which envisages an XML layer at which applications will exchange tags understood purely as character strings, and an RDF layer which will define their meaning.[12]

12 For the formal syntax of the relationship between the two, see Beckett (2004). The layer-cake diagram derives from a power-point slide used by Tim Berners-Lee in many talks, with no clear publication event.

However, a number of challenges for RDF soon emerged. First, although we can now explicitly express a great many propositions, the triple structure is too logically simple to express *all* propositions. For one thing, it allows only binary relations, which is unfortunate if one wants to express more complex relations (e.g., Bill and Melissa Gates were married in Las Vegas). It is not possible to express negations (e.g., The Capital of France is not London). Second, once again there is an issue with unresolved indices. Even where a propositional structure exists, with URIs assigned to all three components, RDF does not determine what the URIs are indexed to. If one website refers to *War and Peace* using the tag <book> and another by <title>, the putative Semantic Web book review aggregator invoked above will miss the connection without some *further resource to identify that these tags have the same meaning*.

Therefore once again some Semantic Web researchers tempered their expectations: "By standardizing the notations, XML and RDF take an important first step, but that step is insufficient for data sharing without some way of comparing, relating, and translating the vocabularies" (Sowa 2000). Meanwhile other developers suggested a markup system that was more detailed and specific than RDF, arguing that what was needed was not just propositional structure but also the use of it to express some shared set of definitions and background knowledge. Surely if only the right semantic standard could be specified this would solve the problem? Thus the Semantic Web researchers turned to developing an RDF *schema*.

4.4 RDF schema

RDF Schema (Brickley and Guha 2003) drew deeper on the resources of formal logic. It created the means to define classes and properties (e.g., A tree is a kind of plant), populate classes with instances (My Christmas tree is a pine), and organize them into a subsumption hierarchy (All pines are conifers, therefore all kauri are conifers), as subclass relationships provide an easy means of grounding a large proportion of deductive reasoning. As such this work began to overlap with a large body of previous research in a field known as Knowledge Representation. This field arose from AI after the latter encountered difficulties that led to a humbling scaling-back of goals and enthusiasm (and funding). In the 1980s, influential researchers decided that in order to parse and understand natural language machines needed an infusion of general knowledge (a.k.a. "common-sense") and strenuous attempts were made to capture such knowledge in formal languages amenable to theorem provers. The field was thus characterized by careful axiomatized definitions and complex deductive proofs of simple facts such as: If a cup is in a room, its handle is in the room, and: If Bruce is having lunch, he is not dead.

In building the upper levels of the relevant subsumption hierarchies, these AI researchers gradually came to realize that they were recapitulating the history of metaphysics. For instance, what is a person? Is it an entity that exists, or *per*dures wholly at a particular time (and thus has one address)? Or is it an entity that *en*dures a whole "person-lifetime" (and thus can have multiple addresses)? The researchers thus began to refer to their work as *ontology*. To some degree a convergence with this work was anticipated from the beginning of the Semantic Web project: "For the Semantic Web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning" (Berners-Lee et al. 2001).[13] Thus RDF Schema developers began to see themselves also as ontologists. *This* richer approach, it was thought, would finally enable semantics to be captured and standardized:

> RDF itself is a handy way to describe resources . . . But RDF by itself only gets us part way toward realizing the Semantic Web, in which agents can infer relationships and act on them. Classification is extremely important on the Semantic Web. Each community of related interests defines categories for the matters that it discusses. For instance, the snowboarding community defines items such as snowboards, parks, tricks, and manufacturers. The definition of a manufacturer in snowboarding terms is related to the definition of a manufacturer in the general business sense. The snowboarding community can enshrine these definitions by creating a schema for its RDF models. (Ogbuji 2002: 32)

However, despite this new logical sophistication, RDF Schema turned out to be *still* too logically simple to express a great deal of what one might wish to say if one were seeking to authoritatively define the meanings of terms. For instance, although one can declare new classes and populate them with instances, one cannot say anything further about these classes and instances, e.g., that two names denote the same person. One still can't negate or state disjunctions, for which full first-order logic is needed. And once more RDF Schema terms are indexed *via* namespaces whose further meaning is opaque.

Therefore the Semantic Web developer community decided to add a new, even more expressive and explicit ontological layer, above both RDF and RDFS, onto their Web "cake." This layer, it was thought, would add the expressivity which all lower layers lacked, and enable them to translate terms, resolve syn-

13 For a stringent critique of Semantic Web developers for not heading off this convergence at the pass, see Halpin (2004).

onyms and homonyms, facilitate inference, and authoritatively determine the meaning of what went below.

4.5 OWL

OWL (“Web Ontology Language”), which became a W3C Recommendation in February 2004, went beyond RDFS by providing more vocabulary and greater logical expressivity. This included abilities such as defining new classes in terms of logical relationships between known classes (e.g., A herbivore is an animal that is not a carnivore) and declaring class cardinality (e.g., Obama has two daughters) and equality (e.g., The Unabomber is Ted Kaczynski). This language has the expressivity of full first-order logic. Its greater expressivity has costs in inferential tractability, however, so following dispute on the development committee OWL was released in three versions, progressively simpler: *OWL Full*, *OWL DL, and OWL Lite*.

Its developers announced it with enthusiasm: “The Semantic Web will build on XML’s ability to define customized tagging schemes and RDF’s flexible approach to representing data. The first level above RDF required for the Semantic Web is an ontology language that can formally describe the meaning of terminology used in Web documents” (McGuinness and van Harmelen 2004). The W3C envisaged that once they provided OWL the world would respond by defining and contributing ontologies, and a number of ontology libraries/clearing-houses were initially set up for this purpose. However at present coverage is patchy at best, and it would appear that OWL is not currently widely used.[14]

A number of criticisms have been levelled at OWL. First, there are technical concerns regarding complexity and scalability.[15] Possibly more damningly, it has been charged that all of its complexity *still* doesn’t do the job of capturing basic human knowledge of meaning:

14 For instance in the DAML ontology library: http://www.daml.org/ontologies/ *nothing has been* added since 2003. The Protégé ontology library: http://protege.cim3.net/cgi-bin/wiki.pl?ProtegeOntologiesLibrary and Ontology Design Patterns (http://ontologydesignpatterns.org) have more recent additions, but still pale into insignificance compared to the size, vibrancy, and, most importantly, coverage of other user-supplied content on the Web – e.g., Wikipedia, on which more below.

15 Bergman (2006) argues that OWL cannot express the information in a typical business spreadsheet; McGuinness and van Harmelen (2004) argue that the exact formal relationship between OWL and RDF is delicate and underspecified.

> It is an unfortunate dogma of computer science in general, and the Semantic Web in particular, that all semantic contents are reducible to first order logic or to set theory. Berners-Lee, Hendler and Lassila claim that "fortunately, a large majority of the information we want to express is along the lines of "a hex-head bolt is a form of machine bolt." Unfortunately, this is not true. If one considers how humans handle concepts, the class relation structures of the Semantic Web capture only a minute part of our information about concepts. (Gardenförs 2004: 18)

As examples of what is missing, Gardenförs cites: i) *similarity* between concepts, as opposed to the generalization and specialization of subsumption hierarchies (e.g., tree and bush), and ii) *concept-combinations*, which in logical languages such as OWL can only be modelled as property conjunction and thus class intersection. This won't do as for instance stone lion does not refer to all lions that are also stones, nor honey bee to all bees that are honey.

Once again, the familiar complaint arose that what was being captured was syntax not semantics: "One way to pinpoint the symbol grounding problem for the Semantic Web is to note that there is no general way to resolve a conflict between two ontologies, since all information is syntactic, that is, expressed in symbolic structures . . ." (Gardenförs 2004: 4). More radically, Gardenförs argues that even if it is possible to translate between ontologies, this does not solve their lack of overall meaning, since if these semantic islands really are free-floating, "it does not help to tie [them] together" (Gardenförs 2004: 4).

Finally, the Semantic Web's restriction of inferencing to deductive logic has also received heavy criticism:

> The people working on the Semantic Web greatly overestimate the value of deductive reasoning (a persistent theme in Artificial Intelligence projects generally.). . . In the real world, we are usually operating with partial, inconclusive or context-sensitive information. When we have to make a decision based on this information, we guess, extrapolate, intuit, we do what we did last time, we do what we think our friends would do or what Jesus or Joan Jett would have done . . . but we almost never use actual deductive logic. (Shirky 2003)

4.6 Analysis

The above summarizes development on the Semantic Web so far. Surely a larger pattern may be discerned here. What we have seen is a series of attempts to create *ex nihilo* the meaning of signs on the Web *via* a set of antecedent definitions. Arguably this misunderstands what it is for something to have meaning. What is being attempted is essentially to "preprocess" interpretation. For instance with our key term 'tree', an attempt is being made to specify that it *must* be understood to refer to the biological organism under conditions X,Y. . . , resulting in possible inferences A,B. . . , and that it *must* be understood mathematically under conditions W,Z. . . , resulting in possible inferences C,D. . . This essentially seeks to reverse the order of semiotic determination that we saw in section 3: from (object → sign → interpretant), to (interpretant → sign → object). However, at the risk of waxing melodramatic, to attempt this is in Peircean terms literally to kill meaning. For we saw that the lifeblood of meaning-creation is continued mediation of the sign's object to minds *via* specific uses of the sign in specific contexts for specific purposes. Cartesian dualism, with its idealized pregiven meaning postulated in the sign-user's head, misses this.

From a Peircean perspective, the mere fact that languages such as OWL are not widely used *is* the key argument against their having real significance. And this is not a purely academic problem given the vast amount of research money at stake. Calling the Semantic Web "a machine for creating syllogisms," Shirky writes: "This is the promise of the Semantic Web – that it will improve all the areas of your life where you currently use syllogisms. Which is to say, almost nowhere" (Shirky 2003). He seems to suggest that the Semantic Web vision is entirely misguided and should be given up. But should it?

5 Peircean approaches to web semantics

If it is not possible to "preprocess" interpretation, does this mean that it is not possible for computers to understand the meaning of documents on the Web more deeply than via syntactic string searches? Is it impossible to replace a "web of links" with a "web of meaning"? I believe such pessimism is premature, and will offer two related suggestions. First, one can build applications that allow interpretants to freely *grow*, within whatever communities choose to use them. Second, one can *harvest* those interpretants to produce further interpretants that are possessed of genuine added semiotic value. Moreover this is an added value whose power lies – by contrast to the Cartesian framework – precisely in the fact that the further interpretants were *never* intended by the producers of the original signs.

Web 2.0 is an umbrella term for a new paradigm of open collaborative activity on the Web. Coined in 1999 (DiNucci 1999), the term became common currency around 2004. Although Tim Berners-Lee has (interestingly) called the term "a piece of jargon," it is now widely accepted that some such large paradigm-shift has taken place. It has been characterized by the spontaneous emergence of a number of new entirely Web-born *genres*. These include:

1. *Blogs*. Originally, each entry of a "web log" presented a hyperlink to a webpage the author found interesting, with some personal commentary. In Peircean terms, the incoming link and comment may be viewed as an *interpretant* of the page in question. For by linking to the page it also refers to its content, encouraging others to view and interpret it, and thus in some suitably general sense it shares the page's *object*. What is interesting here is that a hyperlink may be regarded as itself a sign, despite not being a word.[16] It was claimed earlier that purely word-based models of signification would be seen as too limited, here now is some evidence for that.

   The genre quickly evolved into full-blown online journals, frequently embellished with hyperlinks, as well as photos and videos that constitute further non-verbal interpretants of a blog's linked pages and its own content. Where a blog author develops a distinctive and widely-known perspective on the world, this perspective can become its own high-level sign, and an item's merely being noticed and blogged about by that author can become an interpretant of that item in itself. Blogs are also frequently organized around some special interest (e.g., restaurant reviews, local politics), thus creating a further structural interpretant of sites so collected that once again is irreducible to purely word-based meaning.
2. *Tags*. Tags are labels added voluntarily by users to information objects on the Web. Labelling of course is a form of interpretation. If a photo is tagged "cute," this means that the tagger is passing some further comment (whether straight or ironic) on what it represents. Once again, the practice began as an individual's means of labelling web-pages with words or phrases in order to rediscover them quickly, but it has spread to embrace a number of other much more public uses. A variety of websites has emerged to serve as tag clearing-houses.[17]
3. *Collaborative websites*. These sites provide a shared semiotic space in which users can define, describe, and discuss topics of mutual interest. Sites exist for an enormous variety of specialized topics, however one of the original and most impressive is the online encyclopedia *Wikipedia*. It offers over three million articles written by thousands of contributors, plus a wealth of further useful structural features, such as "infoboxes" and a category network (Medelyan et al. 2009). Its motto is "The free encyclopedia anyone

16 This is leaving aside "anchor text" within a link, which may or may not be present.
17 Examples include del.icio.us. for sharing tagged bookmarks to web-sites (http://del.icio.us/) and Flickr for tagged photographs (http://www.flickr.com/).

> can edit." Who knew such a social arrangement could bear such fruit? This project is arguably a remarkable and unanticipated realization of Peirce's "community of inquiry." Its ever-increasing level of accuracy (Giles 2005) has caused considerable surprise in those who do not hold to Peirce's theory of truth, but not in those who do.

As of current writing further new genres are rapidly germinating in the realm of so-called social software (among which one might mention the *Status Update*, the *Wall*, *Friending* – not to mention *de-Friending*). Any attempt to definitively describe this new semiotic landscape is doomed to be quickly outdated, but what is clear is that ordinary people are pouring signification into the public domain for free. In fact, a fascinating (to some, alarming) trend is apparent towards public performance of significations that would have once been considered private (such as photos of oneself, messages to friends, party invitations). This in itself would constitute interesting material for a future study, but for our purposes what is notable is the sheer volume and variety of human sign-activity now available worldwide for arbitrary interpretative purposes.

5.1 True added semantic value: The lesson of Google

Google became the most successful web search engine (as of present writing at least) due to its *PageRank* algorithm, named after company founder Lawrence Page. Its key idea is to gauge how relevant and useful a page is likely to be by calculating the number of links to the page from the Web as a whole, taking into account in turn how highly ranked the linking pages are (Brin and Page 1998). Once again we can say that Google realized that hyperlinks constitute *interpretants* of the web-pages they link to, in their indication that the creator of the page linked from thought it was in some sense relevant to, and thus once again in a broad sense *designating the same object as*, the page linked to. However, remarkably, the PageRank algorithm goes even further than this. It derives from the links an entirely automated judgment regarding the page's *quality*. For web-authors are now so many and so diverse that quality pages will be found and linked to more often on average. By ranking such pages higher the algorithm merely distils these millions of individual choices: "Google's founding philosophy is that we don't know why this page is better than that one: If the statistics of incoming links say it is, that's good enough. No semantic or causal analysis is required" (Anderson 2008). The PageRank algorithm can thus be understood as deriving a meaning, namely, the relative significance of a given website, *which no individual* *hyperlink creator ever intended.*[18] The importance of this cannot be overstated – it is genuine added semiotic value.

Google have pointed out that one source of their success is their ability to draw on an absolutely unprecedented quantity of data, and have suggested that more general morals might be drawn from this by IT researchers. In an article entitled "The Unreasonable Effectiveness of Data," three Google scientists note that if one only accumulates enough data, answers may appear to problems that previously appeared insoluble:

> In many cases there appears to be a threshold of sufficient data. For example, James Hays and Alexei A. Efros addressed the task of *scene completion:* removing an unwanted, unsightly automobile or ex-spouse from a photograph and filling in the background with pixels taken from a large corpus of other photos. With a corpus of thousands of photos, the results were poor. But once they accumulated millions of photos, the same algorithm performed quite well. (Halevy et al 2009: 9)

It is intriguing from a semiotic perspective to think about why this strategy, which substitutes orders of magnitude of data for authorial intention, is so successful. Does this constitute the discovery of a new informational "law of large numbers"? (A "law of large corpora"?)

This key idea of harvesting interpretants "in the wild" to create further automated interpretants has been built on by IT researchers. For instance, Milne and Witten (2008) gather and analyze hyperlinks in and out of Wikipedia pages in order to disambiguate terms on the Web. Kiwi could refer to bird or fruit, but the fact that the fruit's Wikipedia page links to Berry and Woody_plant, while the bird's page links to Flightless_Bird, and Ratite distinguishes two concepts. Others use Wikipedia for language learning without a teacher, or machine translation (Sorg and Cimiano 2008). Gabay et al. (2008) use Wikipedia hyperlinks and anchor text to automatically detect word boundaries in Hebrew.

Meanwhile Wikipedia's stored history of edits is a gold-mine of judgments of relative correctness. It can be mined to automatically learn common textual errors (Nelken and Yamangil 2008). Others (Druck et al. 2008) attempt to predict the quality of Wikipedia contributions by whether the Wikipedia community "reverts" them and if not, how long they last. Kittur et al. (2007) use the same information as a measure of how controversial a given article is. Here once again it is important to note that what is being mined to construct these further interpre-

18 Spam attempts aside-Google strenuously attempts to discount spammed links as a distortion of the meaning that they are endeavouring to capture.

tants of Wikipedia is structural features, not just words.[19] Finally, Chena et al. (2007) apply the PageRank algorithm to scientific citations in physics to derive a measure of the scientific quality of papers that they claim is superior to mere citation counting. Favoring papers cited by papers that are themselves highly cited allows the identification of so-called "gem" papers whose influence militates *against* their being always explicitly cited, as they are so widely known.

One might wonder: what are the limits to this kind of automated interpretation? Surely the coherence of intentional thought in the mind of a single conscious being (the Cartesian model's greatest strength) must still be required somewhere, in order to say that a given sign is truly understood? It now may be outmoded to understand such coherence as some form of Kantian conscious "transcendental apperception" – arguably the most sophisticated working out of the Cartesian model in the Western philosophical tradition. Nevertheless, shouldn't it at least approximate the operationally integrated understanding of a subject matter attained, say, by a working scientist? *Even here*, controversially, it has been suggested that massive quantities of data and the ability to harvest it might even begin to substitute for the scientific method itself. In a "blue-sky" piece in *Wired* magazine, Anderson speculates:

> Scientists are trained to recognize that . . . no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you must understand the underlying mechanisms that connect the two . . . Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot . . . (Anderson 2008)

As an example of this he cites "shotgun gene sequencing by J. Craig Venter," who trained gene-sequencing computers onto large quantities of material taken at random from sea and air. Anderson (2008) claims that this random process revealed thousands of previously unknown species, albeit only as, "a statistical blip – a unique sequence that, being unlike any other sequence in the database, must represent a new species."

The skeptic will protest: but who will understand (i.e., interpret) these patterns? The ruthless Peircean answer comes back that someone must interpret the patterns or they will die (cease to reproduce themselves). Or, as we might more accurately say, "some-one or some-thing . . ."

19 This suggests possible fruitful convergences with the well-developed field of machine learning that are however beyond the scope of this journal.

6 Conclusion

Cartesians assume that in order to make the Semantic Web happen it is necessary to make a huge defining effort to somehow encode in machine-readable form the private intentions with which we produce signs. The Peircean approach by contrast involves realizing that, in the words of Google scientists, “The first lesson of Web-scale learning is to use available large-scale data rather than hoping for annotated data that isn’t available” (Halevy et al. 2009: 8). We have seen that a wealth of interpretants exists on the Web – in hyperlinks, tags, blogs, Wikipedia edits, and even raw text. The kinds of inferencing that will harvest such interpretants and generate new understanding of what they truly *mean* is not the neat, deductivist, rule-based reasoning of “good old fashioned AI.” We need new models.

Having mentioned AI, it is worth noting that in this field also philosophical theories of meaning are not mere abstract speculation but directly influence what researchers envision and attempt to build. This is not surprising since as we have seen the Semantic Web consists of many old AI goals in 1990s dress. The classic 1950s-era model of AI – something like a digital encyclopedia in the head of a robot – may now be seen as a poignant attempt to make concrete the Cartesian picture of meaning as ideas *in* the head. By contrast, Peirce’s account of a sign’s meaning in terms of its interpretants led him to write, “just as we say that a body is in motion, and not that motion is in a body, we ought to say that we are in thought, and not that thoughts are in us” (Peirce 1868). In this sense, perhaps as with “Web semantics” also with “Web intelligence” we already have more at our disposal than we realize. Further exploration of the current state and ambitions of Artificial Intelligence from a semiotic perspective would be most fruitful.[20]

References

Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired* 16(7). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory (accessed 9 September 2012).

20 Honourable exceptions to this claimed lack of engagement between semiotics and AI include Santaella-Braga (2004) and Sowa (2001, 2000). Big thanks are due to Ian Witten, Olena Medelyan, David Milne, and David Nichols for discussions and joint research without which this paper would not have been possible. Thanks are due also to participants in a symposium at the 2007 University of Helsinki *Applying Peirce* conference, at which a much earlier version of this paper was presented, in particular Vincent Colapietro.

Beckett, D. 2004. RDF/XML syntax specification (Revised). http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/ (accessed 9 September 2012).
Berners-Lee, T. 1998. What the Semantic Web can represent. http://www.w3.org/DesignIssues/RDFnot.html (accessed 9 September 2012).
Berners-Lee, T. 2002. Foreword. In D. Fensel, J. A. Hendler, H. Lieberman & W. Wahlster (eds.), *Spinning the Semantic Web*: *Bringing the World Wide Web to its full potential*, i–xxiii. Cambridge: MIT Press.
Berners-Lee, T., J. Hendler & O. Lassila. 2001. The Semantic Web. *Scientific American* 284(5). 34–43.
Bray, T., J. Paoli, C. M. Sperberg-McQueen, E. Maler, & F. Yergeau. 2000. Extensible Markup Language (XML) 1.0 (Third Edition). http://www.w3.org/TR/REC-xml/ (accessed 9 September 2012).
Brickley, D. & R. V. Guha. 2003. RDF Vocabulary Description Language 1.0: RDF Schema. http://www.w3.org/TR/rdf-schema/ (accessed 9 September 2012).
Brin, S. & S. Page. 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30. 107–117.
Chena, P., H. Xieb, S. Maslovc & S. Rednera. 2007. Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics* 1. 8–15.
Cherry, Steven M. 2002. Weaving a web of ideas. *IEEE Spectrum Online*. http://spectrum.ieee.org/telecom/internet/weaving-a-web-of-ideas (accessed 9 September 2012).
Decker, S., F. van Harmelen, J. Broekstra, M. Erdmann, D. Fensel, I. Horrocks, M. Klein & S. Melnik. 2000. The semantic Web – On the respective roles of XML and RDF. *IEEE Internet Computing* 4. 63–74.
Derrida, J. 1976. *Of grammatology*, Gayatri Chakravorty Spivak (trans.). Baltimore & London: Johns Hopkins University Press.
Derrida, J. 1978. *Writing and difference*, Alan Bass (trans.). Chicago: University of Chicago Press.
Descartes, R. 1996 [1647]. *Meditations on first philosophy: With selections from the objections and replies,* Jonathan Bennett (trans.). http://www.earlymoderntexts.com/de.html (accessed 9 September 2012).
DiNucci, D. 1999. Fragmented future. *Print* 53(4). 32, 221–222.
Druck, G., G. Miklau & A. McCallum. 2008. Learning to predict the quality of contributions to Wikipedia. In *Proceedings, workshop on Wikipedia and AI, AAAI08*, Chicago, 7–12. Palo Alto: AAAI Press.
Frege, G. 1952 [1892]. On sense and reference. In Peter Geach & Max Black (eds.), *Translations from the philosophical writings of Gottlob Frege*, 56–78. Oxford: Blackwell.
Gabay, D., Z. Ben-Eliahu & M. Elhadad. 2008. Using Wikipedia links to construct word segmentation corpora. In *Proceedings, workshop on Wikipedia and AI, AAAI08*, Chicago, 61–63. Palo Alto: AAAI Press.
Gärdenfors, P. 2004. How to make the Semantic Web more semantic. In A. Varzi & L. Vieu (eds.), *Formal Ontology in Information Systems: Proceedings of the Third International Conference (FOIS 2004)*, 17–34. Torino: IOS Press.
Giles, J. 2005. Internet encyclopedias go head to head. *Nature* 438. 900–901.
Hacking, I. 1975. *Why does language matter to philosophy?* Cambridge: Cambridge University Press.
Halevy, A., P. Norvig & F. Pereira. 2009. The unreasonable effectiveness of data. *IEEE Intelligent Systems* (March/April). 8–12.

Halpin, H. 2004. The Semantic Web: The origins of AI redux. In *Proceedings, the fourth international workshop on the philosophy and history of mathematics and computation*. San Sebastien.
Hayes, P. 2004. RDF semantics. http://www.w3.org/TR/rdf-mt/ (accessed 9 September 2012).
Kittur, A., B. Suh, B. Pendleton & E. Chi. 2007. He says, she says: Conflict and coordination in Wikipedia. In *Proceedings, the SIGCHI conference on human factors in computing systems*, San Jose, 453–462. New York: ACM.
Kleene, S. C. 1952. *Introduction to metamathematics*. Amsterdam: North-Holland.
Lassila, O. & R. Swick. 1999. Resource Description Framework (RDF) model and syntax. http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/ (accessed 9 September 2012).
Legg, C. 2005. The meaning of meaning-fallibilism. *Axiomathes* 15(2). 293–318.
Liszka, J. J. 1996. *A general introduction to the semiotic of Charles Sanders Peirce*. Bloomington: Indiana University Press.
Locke, J. 1994. *An essay concerning human understanding*. New York: Prometheus.
McGuinness, D. & F. van Harmelen. 2004. OWL Web Ontology Language: Overview. http://www.w3.org/TR/owl-features/ (accessed 9 September 2012).
Medelyan, O., D. Milne, C. Legg, & I. H. Witten. 2009. Mining meaning from Wikipedia. *International Journal of Human-Computer Interaction* 67(9). 716–754.
Milne, D. & I. H. Witten. 2008. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proceedings, workshop on Wikipedia and AI, AAAI08,* Chicago, 25–30. Palo Alto: AAAI Press.
Nadin, M. 1998. Computer. In Paul Bouissac (ed.), *The encyclopedia of semiotics*, 65–68. Oxford: Oxford University Press.
Nelkin, R. & E. Yamangil. 2008. Mining Wikipedia's article revision history for training computational linguistics algorithms. In *Proceedings, workshop on Wikipedia and AI, AAAI08*, Chicago, 31–36. Palo Alto: AAAI Press.
Neumuller, M. 2000. In applying computer semiotics to hypertext theory and the World Wide Web. In S. Reich & K. M. Anderson (eds.), *OHS/SC 2000, LNCS 1903*, 57–65. Berlin: Springer.
Ogbuji, U. 2002. The languages of the Semantic Web. *New Architect* 7(6). 30–33.
Peirce, Charles S. 1868. Some consequences of four incapacities. *Journal of Speculative Philosophy* 2. 140–157.
Peirce, Charles S. 1878. How to make our ideas clear. *Popular Science Monthly* 12. 286–302.
Peirce, Charles S. 1931–1966. *The collected papers of Charles S. Peirce*, 8 vols., C. Hartshorne, P. Weiss & A. W. Burks (eds.). Cambridge: Harvard University Press. [Reference to Peirce's papers will be designated *CP* followed by volume and paragraph number.]
Quine, W. V. O. 1960. *Word and object*. Cambridge, MA: Harvard University Press.
Santaella-Braga, L. 2004. *Cultura das mídias* [The computer as a semiotic medium], 4th edn. São Paulo: Experimento.
Shadbolt, N., W. Hall & T. Berners-Lee. 2006. The Semantic Web revisited. *IEEE Intelligent Systems* 21(3). 96–101.
Shank, G. & D. Cunningham. 1996. Mediated phosphor dots: Toward a post-Cartesian model of computer-mediated communication via the semiotic superhighway. In C. Ess (ed.), *Philosophical perspectives on computer-mediated communication*, 29–44. New York: SUNY Press.
Shirky, C. 2003. The Semantic Web, syllogism and worldview. http://www.shirky.com/writings/semantic_syllogism.html (accessed 9 September 2012).

Short, T. L. 2007. *Peirce's theory of signs*. Cambridge: Cambridge University Press.
Sorg, P. & P. Cimiano. 2008. Enriching the crosslingual link structure of Wikipedia – A classification-based approach. In *Proceedings, workshop on Wikipedia and AI, AAAI08*, Chicago, 49–54. Palo Alto: AAAI Press.
Sowa, J. 2000. Ontology, metadata, and semiotics. In B. Ganter & G. W. Mineau (eds.), *Conceptual structures: Logical, linguistic, and computational issues* (Lecture Notes in AI 1867), 55–81. Berlin: Springer-Verlag.
Sowa, J. 2001. Signs, processes, and language games. http://www.jfsowa.com/pubs/signproc.htm (accessed 9 September 2012).
Swartz, A. 2002. The Semantic Web in breadth. http://logicerror.com/semanticWeb-long (accessed 9 September 2012).
Uren, V., P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, M. Motta & F. Ciravegna. 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics* 4: 14–28.
Wittgenstein, L. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.

Bionote

Catherine Legg (b. 1967) is a senior lecturer at University of Waikato ⟨clegg@waikato.ac.nz⟩. Her research interests include philosophy of language, metaphysics, epistemology, and American philosophy. Her publications include "Argument-forms which turn invalid over infinite domains: Physicalism as supertask?" (2008); "The problem of the essential icon" (2008); "Letting reality bite" (2008); and "The hardness of the iconic must: Can Peirce's existential graphs assist modal epistemology?" (2011).