

# Semantics, Hermeneutics, Statistics: Some Reflections on the Semantic Web

Graham White  
Electronic Engineering and Computer Science  
Queen Mary University of London  
Mile End Road  
London E1 4NS  
UK  
www.dcs.qmul.ac.uk/~graham  
graham@dcs.qmul.ac.uk

**We start with the ambition – dating back to the early days of the semantic web – of assembling a significant portion human knowledge into a contradiction-free form using semantic web technology. We argue that this would not be desirable, because there are concepts, known as essentially contested concepts, whose definitions are contentious due to deep-seated ethical disagreements. Further, we argue that the nineteenth century hermeneutical tradition has a great deal to say, both about the ambition, and about why it fails. We conclude with some remarks about statistics.**

*semantic web, ontology creation, consistency, hermeneutic circle, Schleiermacher, scale-free distributions*

## 1. THE AMBITIONS OF THE SEMANTIC WEB

The Semantic Web has, from its inception, had impressive ambitions: it was intended to be “a web of machine-readable information whose meaning is well-defined by standards”. (Berners-Lee 2005, p. xi) What is ambitious in this statement is the idea of defining meaning by standards: the benefits of machine-readable information are not contentious, and neither are the benefits of definition by standards, but the idea that one could extend these benefits to the realm of meaning is somewhat more ambitious.

Berners-Lee was, of course, somewhat more modest:

There will be information on the Web that has a clearly defined meaning and can be analysed and traced by computer programs: there will be information, such as poetry and art, that requires the whole human intellect for an understanding that will always be subjective. (Berners-Lee 2005, p. xiii)

Despite this, he still has sizeable ambitions: computers should be able to manage us

the exciting possibility of letting programs run over [data about the projects we are engaged in] and help us analyse and manage what we are doing. The computer renders the scene visibly as a software

agent, doing anything it can to help us with the bulk of data, to take over the tedium of anything that can be reduced to a rational process, and to manage the scale of our human systems. (Berners-Lee 2005, p. xiv)

“We are forming cells within a global brain”, he writes, “and we are excited that we might start to think collectively”. (Berners-Lee 2005, p. xxiii)

There are two issues that arise here. The first is whether the fulfilment of these ambitions would be desirable: the second is whether it might be possible. (Undesirable but impossible might not be a benign state of affairs, because the effort to implement this sort of thing might be still very damaging.) We will examine the first issue in the remainder of this section, and the second one in Sections 2 and 3.

First, some disclaimers. We are not arguing that the Semantic Web is a bad thing: in the applications developed so far, such as those described in the demos at <http://iswc2010.semanticweb.org/accepted-poster-demo>, seem to be quite benign. We are merely arguing against certain exaggerated ambitions. We are also not conducting a transcendental, impossible-in-principle argument, perhaps along the lines of Harnad (2003): on the contrary, our argument will focus on the practicalities. It will be an argument

against the possibility, or desirability, of attaining consensus algorithmically.

### 1.1. Essentially Contested Concepts

Essentially contested concepts (Gallie 1956) are those “the proper use of which inevitably involves endless disputes about their proper uses on the part of their users”. These concepts are typically evaluative or in some sense politically loaded: for an example, see the discussion of power in Allen (2008). These concepts are essentially contested because any attempt to define them would involve attaining, or imposing, consensus on political or evaluative norms on which there is (and likely cannot be) consensus: a typical case seems to be that of a tradition in which the authenticity, or otherwise, of temporal continuations of that tradition is always under the possibility of re-evaluation.<sup>1</sup>

Suppose that we had an automatic processing machine for extracting meaning from text corpora: it would have, for example, text parsers, and then there would be software to turn the parse trees into some sort of semantic representation. So we could automatically generate a lot of semantic representations.

But that would not be enough, because we have to resolve contradictions between these texts: we have to say when terms in the semantic representation co-refer, when different texts make contradictory assertions about the same individuals, and so on: and, when we find such contradictions, we have to resolve them in some way. There seems to be some expectation that statistics will help here: that, when we can scan entire libraries (Foster 2011) and extract semantics from them, we can use the brute power of statistics to resolve contradictions.

I shall argue, in the remainder of this paper, that such convergence is unlikely to help us: that statistics, though it may be an indispensable tool in a number of other ways, is unlikely to help us towards a consistent global state of knowledge. There is, however, one point which should be made now: that, if this did work, we would not have any essentially contested concepts left, and we would have arrived at that state without discussion and without any conscious, rational resolution of the issues which make these concepts essentially contested. One could, perhaps, imagine how those who represented the losing side might feel: that, perhaps, this scanning of entire libraries had merely expressed the majority view, and that being the majority view did not make it right; or, perhaps, that this process of scanning entire

<sup>1</sup>Compare MacIntyre (1988, p. 12): “A tradition is an argument extended through time in which certain fundamental agreements are defined and redefined.”

libraries followed by statistical resolution did not have the sort of reflective, deliberative capabilities which any serious attempt to resolve these issues should appeal to. One need not be a Luddite to feel a good deal of sympathy with these positions.

## 2. HERMENEUTICS

These are questions which have a long history. In particular, there is a long tradition of reflection on the process of interpretation of texts, on the problems of using texts which originated in a thought world radically different from our own, of deciphering texts which may use an obscure vocabulary, of reconciling contradictions between or within texts. There had been reflection on these matters since the ancient world, but it was particularly in the modern era (from Spinoza onwards) that this subject flourished, and, particularly, acquired a name: it has, since the mid eighteenth century, been known as hermeneutics (Ramberg and Gjesdal 2009).

We will summarise an argument from Schleiermacher’s *Hermeneutics* (1977):<sup>2</sup> it is an argument that the process of interpretation of a text may well not terminate, even if one has a great deal of information about the context the text was written in. In this respect it is analogous to modern arguments for the indeterminacy of translation such as Quine’s (1969). However, it is probably more interesting (and certainly more realistic): Schleiermacher, as well as being a philosopher and theologian (Frank 1977; Forster 2008; Mariña 2005), also did a great deal of translation (he was responsible for most of what was, for a very long time, the standard German translation of Plato). So in this respect he was talking about something that he knew well from the practical point of view, and this is a comparative rarity in discussions of this sort.

There are two methodological decisions that Schleiermacher makes. Firstly, he argues that

[t]he meaning of any word in any given place must be determined according to its relationship with those words which surround it. (Schleiermacher 1977, p. 116), (1998, p. 44)

Words are ambiguous (a fact which was obvious to Schleiermacher and which has been made obvious again by bitter experience of computational linguistics): so, in practice, their senses are determined by the constraints placed on them by their grammatical and semantic relations with other words in their neighbourhood. These relations can

<sup>2</sup>I shall use my own translations of (Schleiermacher 1977): the English translation (Schleiermacher 1998) is over-literal and quite hard to understand.

extend quite far – for example, if a noun is a subject of a series of verbs (a series which, because of anaphora, can extend beyond the boundaries of a single sentence), then the constraints on that noun come from all of the verbs, and, reciprocally, the verbs constrain each other because of the requirement to have a common subject. So, as Schleiermacher remarks, in order to apply this criterion, one must draw a boundary (1977, p. 135) (1998, p. 61), and it is far from obvious where this boundary goes.

In this context, Schleiermacher introduces a distinction between what he calls *primary* and *secondary* thoughts: primary thoughts (Hauptgedanken) are those which are said for their own sake, secondary thoughts (Nebengedanken) those which there because they serve to articulate the primary thoughts (Schleiermacher 1977, p. 133) (1998, pp. 59f.). This distinction is important, because the criterion above only strictly applies to primary thoughts: communicating those is the author's main project, and it is that project which leads to the desire for consistency and so the mutual constraints. (Schleiermacher 1977, p. 136) (1998, p. 62) This leads to several reasons why particular phrases may not yield relevant constraints. For example, an author may use contradictory assertions in order to illuminate their main argument: as Schleiermacher remarks, "Such contradictions often bring about a definite interpretation more effectively than analogies, because a contradiction is far more striking than either an analogy or simply a distinction". (1977, p. 137) (1998, p. 63) So one has to be able to recognise when these contradictions occur (because otherwise the constraints one produces will be very strange indeed). This may, of course, not be easy.

In a similar way an author can introduce a digression, or some sort of parenthetical insertion. (Schleiermacher 1977, p. 141) (1998, p. 66) These interrupt the series of primary thoughts (and, consequently, the construction of semantic constraints) but they do not end it. After the end of the parenthesis, one resumes the constraints. However, beginnings and ends of parentheses may not be marked by explicit punctuation (punctuation is generally very rare in ancient texts), and so one might not actually know whether the context has ended or not: the series of primary thoughts may just end, and one may read on, and then suddenly it may begin again. (Schleiermacher 1977, p. 141) (1998, p. 66) So here again we have a situation in which recognising the boundary of the context is not easy.

Difficulties are compounded if one wants (as one frequently does) to extend the context to parallel locations in other texts: the text that you have in

front of you may simply not give enough information to determine its own meaning. And there is also a theoretical reason why we may want to do this: in general, no utterance can be understood without taking into account its relation, on the one hand, to the thought of the person who produced it, and, on the other hand, to the totality of the language that it belongs to (Schleiermacher 1977, p. 77) (1998, p. 8). And so we again require judgement in order to determine the extent of the relevant context. (Schleiermacher 1977, p. 137) (1998, p. 63)

Now these distinctions are not only rather subtle and not obviously algorithmic, but they are also circular: each one of them depends on some sort of understanding of the texts that you are applying it to. Schleiermacher was, of course, conscious of this, and does say that, before you start any serious interpretation, you should get an overview of the text in question by reading it through (1977, p. 134) (1998, p. 61). And it is for these reasons that he says that the task of hermeneutics is infinite, only to be achieved by approximation (1977, p. 168) (1998, p. 91). This is not, I think, merely a rhetorical gesture for him: as he writes,

One should check the semantic value of all elements of the sentence [in question] and not only the one which one has run up against, since it can often happen that we run up against one because of ignorance of others. Of course, this has an exception if, because of earlier usage and practice on other texts in the language one has attained certainty that only this one element is unknown to one. But one should carefully test oneself on this, so as not to end up in confusion which could easily have been avoided by working more exactly. (Schleiermacher 1977, pp. 134f.) (1998, p. 61)

So we have our argument: because of the circular nature of the interpretation process, and because of the difficulty of some of the judgements we have to make while performing it, we have no guarantee that the process may terminate, or, that if it does, it will always terminate with the same result.

## 2.1. Politics

How does this fit with the example we started with, that of essentially contested concepts? Consider one of these – maybe, for example, a definition of power – and consider a scholar writing a polemic against one of the meanings, arguing, let us say, that power should be seen as a resource rather than as a relation of domination. This scholar may, for example, outline her own conception as well as quoting, or even simply parodying, her opponent's conception; may break off the argument to develop a narrative illustration, and then resume the argument; may state opponent's views in the middle of her

text; may simply refer to technical terms without explanation, in such a way that the reader is taken to know the literature in which these terms are introduced and explained. Consider the task of a future scholar trying to interpret this text: they would face all of the difficulties which Schleiermacher describes, and it is not obvious how, removed from the context in which the polemic was written and which gives it its meaning, these difficulties could be easily solved. So these problems are very real.

### 3. STATISTICS

So how would a supposed library-scanning machine deal with the example of the polemical scholar? What could statistical methods tell us? To an extent this depends on the methods used: standard methods which apply to entire documents could not really disentangle the two competing meanings, simply because, to distinguish them, they would have to look at contexts smaller than entire documents. So, one would think, a successful statistical technique would have to filter out the polemicist's argument from her opponent's: but the use of parody seems to show that any suitable criterion could be spoofed by a suitably faithful parody.

To an extent, these are problems which both hermeneutics and statistical methods suffer from, although the hermeneutic tradition is, I would argue, more conscious of the difficulties. Consider the case of a scholar with revolutionary insights, expressed either in a new terminology or in old terminology expressed in new ways; this is a case which Schleiermacher analyses (1977, pp. 139f.) (1998, p. 65). At the time where the new concepts are introduced, or the old concepts subverted, there will be no relevant context of other documents from which to establish the meaning (either hermeneutically or statistically). And so the result will be either that the texts cannot be understood in their period, or that they have alien meanings imposed on them. This is, again, a real difficulty: there is a large historical literature on the medieval precursors of Galileo, but these medieval thinkers remain, despite the best efforts of good scholars, very difficult to interpret, more or less for the reasons that Schleiermacher describes. It would be hard to imagine statistics doing any better.

There are two things to be said in summing up. One is positive: statistical methods do seem very capable of performing certain things – of detecting gross similarities between documents, for example, or of doing effective search for queries of certain sorts. But they do this at a certain cost: that of neglecting certain sorts of data.

Let us start with an example, one which was very familiar to Schleiermacher. The New Testament can be considered as a corpus, and quite a small corpus at that: it has 140315 words (word *tokens*, that is). These words are instances of 19886 word type: of these type, 11,588 (that is, over half) occur only once (they are what are called *hapax legomena*). So we only have these words in single contexts, if we look solely at the New Testament: in practice we can get enough information about them by looking at etymology and other Greek documents from the same period and milieu. In part, this statistic is easy to explain: the New Testament is, after all, a comparatively small sample.

But if we go up to a larger sample – we can construct a larger corpus of 2247850 words from Greek literature roughly of the period – we find that it still has a non-negligible number of *hapax legomena*: 9851, to be precise. The number has decreased, but really not very much in proportion to the increase in size of the corpus (which has become over two orders of magnitude larger). Furthermore, 68 of the *hapax legomena* in the New Testament are still *hapax legomena* in the larger corpus. This is because the distribution of word frequencies in text has what is called a *scale-free* distribution: however many individuals one samples, the distribution of frequencies in the sample always looks very much the same, *hapax legomena* and all. Statistical approaches to language usually deal with this difficulty as follows:

In practical systems, it is usual to not actually calculate  $n$ -gram models for all words. Rather, the  $n$ -grams are calculated as usual only for the most common  $k$  words, and all other words are regarded as Out-Of-Vocabulary items . . . Commonly, this will be done for all words that have been encountered only once in the training corpus (*hapax legomena*). (Manning and Schütze 2003, p. 199)

This approach is, as I have said, successful for a lot of things. But it amounts to a deliberate neglect of rare words, and there are circumstances in which one might not want to do that. Consider the example of revolutionary terminology: this may well, statistically considered, be quite rare at its inception. But it may well be *significant*: significant, that is, in the human, ethical sense, rather than in the statistical sense. And this refers us back to the argument at the beginning: how do we know that the users of such a minority vocabulary might be comfortable with being swamped in statistical averages of word types? And, on a suitably weighted average, they will be swamped: we cannot assume that, if we have a minority vocabulary in a particular sample, it will automatically be found again in a larger sample. The corpus statistics that we

have seen do not support that assumption. The New Testament, of course, is something of an exception here: we have a minority vocabulary which went on to become mainstream (and, if I were to include, in my larger Greek corpus, much later literature, in particular commentaries on the New Testament, then the *hapax legomean* would disappear, because they would, of course, all be quoted in the commentaries). But most minority vocabularies, one would hypothesise, are not like that: they correspond to social or intellectual practices that briefly flourished and then died out. This is probably the typical case: this is what a scale-free distribution of phenomena like this gives you.

#### 4. EVALUATION

This is a rather technical argument, quoting freely from philosophy and statistics, and using historical examples that came to hand. What lessons can we draw from it?

**Consensus and Politics** Many of the tools of which Berners-Lee describes as computers “render[ing] the scene visibly as a software agent, doing anything it can to help us with the bulk of data, ... tak[ing] over the tedium of anything that can be reduced to a rational process, ... manag[ing] the scale of our human systems” can, in some contexts, be redescribed as the imposition of a consensus. There are politically loaded contexts in which such processes carry a considerable political weight: one should be aware, despite the shiny optimistic vocabulary, of just what the dangers are.

**Statistics and Complexity** In many of these cases the statistics does not support what one tries to do with it: the data can be irreducibly complex and can resist summarisation. Traditional statistical methods rely on the absence of long-range correlations, which are pervasive in data like this. If the methods don't work, they don't work, but they still give you numbers. In a culture like ours, where numbers are simultaneously revered and avoided, this is extremely dangerous.

**Humanism** The sort of intellectual landscape that I have been describing is pervasive in the culture of humanism: the engagement with texts and the circumstances in which they were produced, the distrust of averages, the caution about the automatic assumption of superiority. It is worth talking to these people.

#### 5. REFERENCES

- Amy Allen. Feminist perspectives on power. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition, 2008.
- Tim Berners-Lee. Foreword. In Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, editors, *Spinning the Semantic Web: Bringing the Worldwide Web to its Full Potential*. MIT Press, 2005.
- Michael Forster. Friedrich Daniel Ernst Schleiermacher. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition, 2008.
- Ian Foster. How computation changes research. In Thomas Bartscherer and Roderick Coover, editors, *Switching Codes: Thinking Through Digital Technology in the Humanities and the Arts*. Chicago University Press, 2011.
- Manfred Frank. Einleitung. In *Hermeneutik und Kritik* Schleiermacher (1977). ed. Manfred Frank.
- W.B. Gallie. Essentially contested concepts. *Proceedings of the Aristotelean Society*, 56:167–198, 1956.
- Stevan Harnad. Symbol-grounding problem. *Encyclopedia of Cognitive Science*, LXVII(4), 2003. URL <http://cogprints.org/3018/>.
- Alasdair MacIntyre. *Whose Justice? Which Rationality?* University of Notre Dame Press, 1988.
- Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 2003.
- Jacqueline Mariña, editor. *The Cambridge Companion to Friedrich Schleiermacher*. Cambridge University Press, 2005.
- Bjrn Ramberg and Kristin Gjesdal. Hermeneutics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2009 edition, 2009.
- Friedrich Daniel Ernst Schleiermacher. *Hermeneutik und Kritik*, volume 211 of *Suhrkamp Taschenbuch Wissenschaft*. Suhrkamp, 1977. ed. Manfred Frank.
- Friedrich Daniel Ernst Schleiermacher. *Hermeneutics and Criticism And Other Writings*. Cambridge Texts in the History of Philosophy. Cambridge University Press, 1998. ed. Andrew Bowie.
- Willard van Ormond Quine. Ontological relativity. In *Ontological Relativity and Other Essays*, pages 26–68. Columbia, New York NY, 1969.