Computational Semantics
Requires Computation

Yorick Wilks

Florida Institute of Human and Machine Cognition
ywilks@ihmc.us

Abstract

The paper argues, briefly, that much work in formal Computational Semantics (alias **CompSem**) is not computational at all, and does not attempt to be; there is some mis-description going on here on a large and long-term scale. Moreover, the examples used to support its value for the representation of the meaning of language strings have no place in normal English usage, or their corpora, and this should be better understood. The recent large-scale developments in Natural Language Processing (NLP), such as machine translation or question answering, which are quite successful and undeniably semantic and computational, have made no use of such techniques. Most importantly, the Semantic Web (and Information Extraction techniques generally) now offer the possibility of large scale use of language data so as to achieve concrete results achieved by methods deemed impossible in formal semantics, namely annotation methods that are fundamentally forms of Lewis' (1970) "markerese".

Introduction

As one of the earliest users of the phrase Computational Semantics, I feel a certain historical disappointment about the way the term has come to be used for work having no real relation to computation at all. The book long ago with that title (Charniak and Wilks 1976) included my tutorial on the work of Wittgenstein and Montague and their possible relevance to CompSem. Linguists had then only recently discovered Montague and his influence was rising, whereas Wittgenstein, who had never had much, was losing even the tight grip his disciples then held on philosophy or at least British philosophy. I remain as convinced now as then of both the relevance of Wittgenstein (Wilks 2008a) to language processing and the lack of relevance of Montague to computational methods. I am certain time has proved that view right, and my aim in this brief paper is to justify that position, not only as regards Montague himself

but also those in the tradition he founded, and I will try to specify loosely what the features of that tradition now are. But the paper is not negative and suggests that CompSem is at last coming of age, on the improbable coat-tails of the Semantic Web (Antoniou and van Harmelen 2008) movement, which is itself related to several forms of earlier CompSem in ways that are not yet fully appreciated. When I refer to, and criticize, formal CompSem in this paper, I mean representation methodologies for language meaning descending from Montague, even if not explicitly, including attempts to use model theory semantics and other methods beyond the mere use of predicate calculus as a descriptive language; I do this in order to avoid criticism of the work of contemporary authors.

Computation is essential for CompSem

The thesis of this brief note is that CompSem must involve computation, or it is mis-described; yet a considerable amount of what passes for CompSem makes no pretence at computation; computation is not for its own sake---it must serve some purpose. If we look at the successes over the last ten years of machine translation (MT), freely available on the web, as well as to a limited degree, Information Extraction, Question Answering and dialogue processing, one surely cannot deny that, in reaching the levels they have, they must have been processing some form of meaning, which is to say doing CompSem. MT is, almost by definition, the origin of all NLP and Computational Linguistics and is by definition the *transfer of meaning by computation* from one et of language strings to another. So it is surely not possible to deny that the reasonably effective MT we now have is doing CompSem in some form.

But, and the next move is obvious, does anyone believe that is being done or will be done by any of the formal methods now current? Given the success of current MT, can one imagine catching up to that level now by formal description methods of CompSem, given their total lack of

performance so far? One must concede here that there has been an MT system based on Montague Grammar (Landsbergen 1982) and one on Situation Semantics (Rupp 1982) but that hardly proves the point, not only because of the failure of those systems to leave any mark whatever on the field, but because of the old rule (updated in Wilks 2008b) that "Any method whatever can be the basis of SOME MT", even word for word substitution, but nothing at all follows from that, since those methods were never subjected to any large-scale evaluation, and only methods that can be have advanced the field of MT to its current state.

Let me restate some form of what I shall call the AI hypothesis about language processing (and indeed any other subfield of AI) ---and more sophisticated versions of this principle can be found in (Wilks 1990a) along with discussion of what kind of model and AI theory is in philosophy of science terms------ "our best bet for understanding human functioning, until the Jubilee Day of neurophysiology comes, is via those computational processes that reproduce that behavior at some adequate level". If one accepts any form of that, then the consequence is clear: CompSem is actually being investigated in the plainly semantics-bearing areas of language processing listed above with some success but by a mix of statistical and pattern-matching methods (see Wilks 2008b) for a general account of the current successful methods in MT).

On possible move here, to save the traditional formal semantics disciplines above, is to say: we do not accept the AI hypothesis, because we are scientists and we therefore wish to understand, not engineer, and above all to predict. That has been the position in academic linguistics descending from Chomsky for some decades. But, alas, now in the era of big data, whether statistical, Semantic Web or some other form of NLP implementation, that argument can no longer be used. Prediction is indeed the heart of science, but prediction requires objectively obtained data, in statistically significant quantities, and the use of such data is effectively the monopoly of statistical and Semantic Web approaches. Formal linguistics has never used data in any systematic way and Chomsky argued explicitly against it on the grounds that it could only lead to performance models (Chomsky 1957, p.17). From that the whole farrago of competence and examples judged in the light of intuition followed. We cannot go back to that and no "science hypothesis" can now be placed in plausible contrast to the AI one.

Let us stay with data for a moment. I believe that the approach to data and evidence in formal/computational semantics (as with syntax in the same formalist tradition) has been fundamentally dishonest; I do not of course mean individual researchers, but the tradition itself. It has been said, for example, of Discourse Representation Theory (van Eijck and Kamp 1997) that its advantage is that it can deal with donkey sentences, like the reference of "it" in the immortal "Every farmer who owns a donkey beats it" (see e.g. Kamp and Reyle 1993).

But there ARE no donkey sentences, aside of course from the corpus of linguistic discussions. They do not exist, any more than anyone ever says, "John wants to marry a Norwegian" to mean a particular Norwegian—you simply do not put it that way; hence that class of sentences are not ambiguous in the way the theory requires. I conducted a quick and informal web search for "wants to marry a German" on the web---because of the linguistic corpus itself I could not of course use "Norwegian"-----and the result was that every one in the first twenty results from Google was the generic reading. There are attested forms with the "particular German" reading such as "He wants to marry a German girl", but they are not in the standard form of the example. The whole of the classic discussion in the literature (e.g. Kadmon 2001) rests on a tradition of making no contact with data at any point.

Here is another piece of data---which can of course never be more than suggestive--- but from a good author: "From one point of view it's perfectly reasonable for me to hate my work, lots of other people, all of them probably perfectly reasonable, also hate it" (Gray 2006, p.230)

What kind of beast have we snared here? Is it a genuine donkey? Which is to say, is it:

A. ((There are (many people) such that)(each person has his work))+(that person hates it)

Which would be the genuine distributed article, or is it is the non-donkey:

B. ((There are (many people) such that)(each person hates Simon-Gray's work))

Context makes clear it is B. and so it is no donkey. Of course not; had Gray wanted to say A. he would have ended the sentence "also hate theirs", which is natural and although a donkey in spirit it is not one in form (as the game requires) because there is not, and cannot be, the crucial "it".

This is not an issue of access to limited data, but of a whole tradition of making up and discussing examples that do not correspond to anything anyone would or could say in real life. It is as glaring an artificial construction of a subject as Russell's attributed observation that no one had ever in fact proved anything with a syllogism.

I am opposing two traditions here, though with added complexities in both camps. On the formal side, much of what is now discussed and rehashed in formal linguistic CompSem was actually set out on the AI side before Montague, and quite independently, in the work of McCarthy and Hayes (1969) and in the tradition they started: namely that the representation of the world should be done through

some form of first order predicate calculus, and its associated semantics. They were not adopting the particular form of logic that Montague later did, but the spirit is exactly the same: to take individual sentences and display their complex representations as expressing particular aspects of the sentence's content. But McCarthy and Hayes were not interested in language as such, and therefore their thrust seems different from that of the CompSem researchers who are interested in the representation of language.

But this difference is more superficial than real. In (Wilks 1990b) I discussed this formal tradition in AI, of which McCarthy and Hayes were the originators, in opposing it to what I then called "Common-sense semantics", and that difference has now re-emerged more or less unchanged within the Semantic Web (SW) community: as between those who believe in strong representations like OWL (e.g. Horrocks, Sattler and Tobies 1997), and those content to work with something like the simple triples of RDF and with little logical baggage (Wilks 2008c). However, this dispute (within SW) is now an empirical one because the stakes are high, the available data is very large, and something must be settled soon: neither group is prone to the use of artificial examples and all participants believe in real data.

One must concede here that there has also been a tradition of AI researchers interested in language, such as Hobbs (Hobbs, Stickelt, Appelt and Martin 1990), who have attempted to make the assignment of logical-semantic structures to language sentences on a large scale an empirical reality, and even promoted DARPA competitions to do that. But that activity not prospered, and the practical goals envisaged for such a technique have been achieved by other means, such as very large-scale fact extraction from the web (e.g. Pantel et al. 2004). The situation is reminiscent of automatic WSD, which has a long and honorable tradition in NLP (e.g. Stevenson and Wilks 2001) but can now seem pointless—even when empirically successful, as it was, and it is undoubtedly CompSem, it has not in fact been used to improve MT, even though it had always been assumed WSD was essential to cure the defects of earlier MT.

The point of this brief paper has been the detachment from reality of formal CompSem activities: their detachment from real technologies that proceed to success without them, the fact that computation is now no more than a metaphor in their writings—even though they cling to the label *Computational* (see for example much of the work presented under that title at IWCS2011), and computation is not even applied to the world of artificial sentences within which its discussions move.

Consider the unreality of the following introduction of the notion of reference: "Reference happens in communication whenever a speaker points out an object to a hearer, for example by saying 'The President of the US' " (van Deemter, personal communication). This is clearly intended as setting out the very basis of CompSem, but its claim has no possible connection to computation, or even the basis of representation itself. Pointing out an object is one thing, sometimes called ostension, which also has its limits as Quine (1960) famously pointed out with the word "Gavagai". But using phrases like that non-ostensively is quite another. Of course, I could point out a man and say, truly or falsely, and as the quotation suggests, "that is the President of the US". But that is a very rare linguistic event; yet the quotation above is written as if Wittgenstein had not discussed all these difficulties in great detail eighty years ago (1953).

In using the phrase above I do not in general point anyone out: in the cases like "the President of the US has to be 35" or "The President is always in the wrong after one term" I not only do not point anyone out, I do not refer to anyone at all, in any sense of that vexed word. If on the other hand, I say "The President of the US was wrong to start the Iraq war", I still do not point anyone out—there is no ostension---but what I do is to provide enough information for an informed hearer to link that phrase to whatever representation they have for one or both of the George Bushes, and await more information to know which. Machine Translation of a document containing that phrase about the President, or its embedding in a machine conversation, could well require those processes. But, and this is the principal point of this paper, none of that computation is aided or even described at all well by the quoted account above, or the introduction of a notion of "reference", which a technical philosophical notion derived ultimately from work by Russell (1905) and his contemporaries whose concern was the truth or otherwise of statements about non-existent rulers of France.

Conclusion

The problem is, I am afraid, that complex and sophisticated theories like those described have contributed nothing to concrete systems of computational linguistics, dialogue understanding or machine translation—that is simple fact. A paradox here is that large-scale computation was brought into CompSem by those, like Jelinek, who called themselves speech engineers and who had little time for notions of meaning (e.g. Jelinek and Lafferty 1991). But their successes, or at least those of the researchers who refined Jelinek's methods for modern MT and other technologies, must have been manipulating meaning if they were doing effective MT; as we argued above, that must have been CompSem, no matter what their own disclaimers about it.

Earlier attempts to link semantics and statistics (e.g. Carnap and Bar Hillel, 1964) were met with derision in their time as a bizarre outgrowth of "Information Theory". Now that derision looks less tenable. In addition, such

large-scale semantic experiments we now have tend to refute all linguistically-based claims about imaginary data (see the works of Hanks 1996)

In conclusion, there is a final irony in our present situation that is worth noting: in some sense the Semantic Web is coming about, in the work of researchers like (Ciravegna et al.,2004) by means of large-scale annotation systems, which is to say, algorithms that go through texts and annotate all company names, for example, by a feature like COMPANY. Such annotations allow fairly accurate answering of questions from the web, an activity that, if successful, cannot be denied to be CompSem. I have argued (Wilks 2008c) that this kind of annotation is precisely what Lewis (1970) rejected as semantics at all by calling it "markerese" and deeming it inherently circular, and that stigmatization gave rise to much of formal CompSem as we now have it.

The role of the SW as the apotheosis of annotation will also be the paradoxical triumph of markerese. It will be a final irony if the success of the SW on a large scale shows that "markerese" was both possible and necessary for CompSem. One could have guessed this from the continuing success of (monolingual) dictionaries; if markerese is circular, and language meaning cannot be given in language, then a dictionary entry can tell one nothing.

References

Antoniou, G. and van Harmelen, 2008. *A Semantic Web Primer,* 2nd Edition. The MIT Press, Cambridge, MA.

Carnap, R. and Bar-Hillel, Y. 1964. Semantic Information Processing, reprinted in Bar-Hillel, *Language and Information,* London: Academic Press: 29-39.

Charniak, E. and Wilks, Y. eds. 1976. *Computational Semantics*, Amsterdam: North-Holland.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Ciravegna, F., Chapman, S., Dingli, A and Wilks, Y. 2004. Learning to harvest the semantic web, In Proc. European Semantic Web Symposium (ESWS04) : 111-132.

Hanks, P. 1996. Contextual dependency and lexical sets. *International Journal of Corpus Linguistics*, 1(1): 22-31.

Hobbs, JR., Stickel, M, Appelt, D and Martin, P. 1990. Interpretation as Abduction, SRI International Artificial Intelligence Center Technical Note 499.

Horrocks, I., Sattler, U., and Tobies, S. 2000. Practical reasoning for very expressive description logics. *Journal of the Interest Group in Pure and Applied Logic*, 8(3): 67-64.

IWCS2011: International Conference on Computational Semantics: http://www.meaningfactory.com/iwcs2011/

Jelinek, F. and Lafferty, J.D.. 1991. Computation of the Probability of Initial Substring Generation by Stochastic Context Free Grammars. *Computational Linguistics* :315-323.

Kadmon, N. 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*. Oxford: Blackwell Publishers.

Kamp, H.and Reyle, U. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Landsbergen, J. 1982. Machine Translation based on logically isomorphic Montague grammars. In *Proc. 9th. Conference on Computational Linguistics*, Prague: 451-476.

Lewis, D. 1970. General semantics. *Synthese:*18–67.

McCarthy, J. and Hayes, P.J. 1969. Some philosophical problems from the point of view of Artificial Intelligence. In D. Michie (ed) *Machine Intelligence 4*. Edinburgh: Edinburgh U.P.

Pantel, P., Ravichandran, D., and Hovy, E. 2004. Towards Terascale Knowledge Acquisition. In *Proceedings of Conference on Computational Linguistics (COLING-04)*. Geneva, Switzerland: 771-777.

Quine, W.V.O. 1960. *Word and Object*. Cambridge, Mass.: Harvard UP.

Rupp, C. J. 1989. Situation semantics and Machine Translation. In Proc. EACL. 89-110.

Russell, B. 1905. On denoting, *Mind*. 67-85.

van Deemter, K. 2010. personal communication.

van Eijck, J. and Kamp, H. 1997. Representing discourse in context. *Handbook of Logic and Language*. Amsterdam: Elsevier.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

Wilks, Y. 1990a. One small head: models and theories. In Partridge, D. and Wilks, Y. eds. *The Foundations of Artificial Intelligence: a sourcebook.* Cambridge: Cambridge University Press: 121-134.

Stevenson, M. and Wilks, Y. 2001. The Interaction of Knowledge Sources in Word Sense Disambiguation. *Computational Linguistics*: 69-89.

Wilks, Y. 1990b. Form and Content in Semantics. In *Synthese*, 81: 373-389.

Wilks, Y. 2008a. What would a Wittgensteinian Computational Linguistics be like? In Proceedings of AISB'08, Workshop on Computers and Philosophy, Aberdeen: 23-32.

Wilks, Y. 2008b. *Machine Translation: Its Scope and Limits*. New York: Springer.

Wilks, Y. 2008c. The Semantic Web and the Apotheosis of annotation. In *Proceedings of IEEE Information Systems*: 121-123.