

What would a Wittgensteinian computational linguistics be like?

Yorick Wilks¹

Abstract. The paper tries to relate Wittgenstein's later writings about language with the history and content of Artificial Intelligence (AI), and in particular, its sub-area normally called Computational Linguistics, or Natural Language Processing. It argues that the shift, since 1990, from rule-driven approaches to computational language and logic, associated with traditional AI and the linguistics of Chomsky, to more statistical models of language have made those connections more plausible, in particular because there is good reason to think the latter is a better model of use than the former. What statistical language models are not, of course, are immediately plausible models of meaning. Moreover, a statistical model seeking a model of a whole language, one can now look at the World Wide Web (WWW) as an encapsulation of the usage of a whole language, open to computational exploration, and of a kind never before available. I describe a recent empirical effort to give sense to the notion of a model of a whole language derived from the web, but whose disadvantage is that that model could never be available to a language user because of the sheer size of the WWW. The problematic issue in such an analogy (Wittgenstein and NLP) is how one can go beyond the anti-rule aspect of both to some view of how concepts can even appear to exist, whatever their true status.

“A main source of our failure to understand is that we do not command a clear view of the use of our words – our grammar is lacking in this sort of perspicuity. A perspicuous representation produces just that understanding which consists in “seeing connexions”. Hence the importance of finding and inventing intermediate cases. The concept of a perspicuous representation is of fundamental significance for us. It earmarks the form of account we give, the way we look at things.” Wittgenstein: *Philosophical Investigations* §122. (My emphasis)

1 INTRODUCTION

Seeking out its intellectual roots or scholarly ancestors is not an activity popular or respected in the technology called Natural Language Processing (NLP, alias Computational Linguistics [1]). Many of its researchers have some vague notion that logical predicate representation, now almost a form of shorthand in NLP, owes a lot to Frege and Russell, but few know or care that, long before Chomsky ([2], if we agree to allow him by courtesy into the history of NLP) Carnap, Chomsky's teacher, set up in the 1930s what he called *The Logical Syntax of Language* ([3]) with formation and transformation rules whose function was to

separate meaningful from meaningless expressions by means of rules. Carnap's driving role behind all that has been utterly forgotten and Chomsky's own work has now simply filled in all the intellectual space in formal linguistics.

Another contemporary of Carnap, also now largely lost to view, is Wittgenstein, whose long campaign against simple-minded notions of linguistic rules was largely provoked by Carnap. He predated Chomsky and NLP, of course, although his influence lived on as a source of Anglo-Saxon linguistic philosophy for many decades, whose practitioners mostly had little time or patience for what they saw as Chomsky's simplicities and certainties.

An attempt to connect Wittgenstein to linguistics thirty years ago was Brown's "Wittgensteinian Linguistics" [4], but his main concern was to contrast Wittgenstein with Chomsky's views, which were more central to language studies then than they are now. Brown noted that Wittgenstein had much in common with Chomsky's anthropological predecessors, from whom he separated himself so clearly with his rule-driven, Carnap-inspired linguistics. Malinowski's observation ([5]:287ff) that language is "a mode of action, rather than a counter-sign of thought" is a sentiment that Wittgenstein could have expressed, and the latter's notion of communities of use who share assumptions and language forms, however bizarre, is not far from anthropological views (often associated with Whorf and Sapir) on the language and belief systems valid in their own terms. Quine [6] later took up the same scenario, that of remote languages, unknown to the observer, and the non-veridical nature of any communication based on translation or supposed meaning equivalence: how could we ever know definitively, he asked, what "Gavagai" meant simply from the utterances (and pointings) we observed?

Wittgenstein seemed less sceptical about translation than Quine; perhaps living in two languages and cultures, as he did, made it seem more natural to him: classic sentiments like "the limits of my language mean the limits of my world" ([7]) do not imply that one cannot be in two or more such worlds. He listed (PI [8] pp.11-12) translation as among normal human activities, and he seemed sceptical about the nature and function of none of his list. It also seems clear that Wittgenstein did believe in some conceptual world over and above surface use, but the problem is knowing what that was, and how it was grounded within usage. In his early work, what he called forms of facts [7] were separate from language and identified with "pictures of fact" and it is not clear that he ever rejected the explanatory power of diagrams and pictures: he continued to use them, even though he was unsure how they "worked" (cf. The problem of knowing why the arrow so obviously points the way it does [8] PI: (129). Pictures and drawings remained important to Wittgenstein because they expressed intention in a way that natural objects in the world do not.

¹ Oxford Internet Institute, University of Oxford.
yorick.wilks@oii.ox.ac.uk

In spite of many things he says that appear to be classic behaviourism –e.g. the apparent denial of the possibility of a private language – Wittgenstein was not an empiricist in the sense that Chomsky intended by that word, as is someone like Sampson [9] who insists that we have no evidence that anything more is innate in humans than a learning mechanism. Wittgenstein could never have written "It is conceivable....that all the processes of understanding, obeying, etc. should have happened without the person ever having been taught the language" (PI §12) had that been his position. Chomsky himself seems to have no understanding whatever of Wittgenstein's overall position, given remarks like: (Chomsky [10]:p60) "[For Wittgenstein] meanings of words must not only be learned, but also taught (the only means being drill, explanation, or the supplying of rules...." . Chomsky has no feeling at all for Wittgenstein's investigation of how we could know that someone was following [a linguistic] rule, and for the simple reason that Chomsky always claims to know that we are following rules, and when, and to see no problem about a statement that a rule is being followed by a speaker.

These arguments, that effectively separate Wittgenstein in every way from the Chomskyan enterprise, can be found in Brown's work, but one must add here that Chomsky and classic Artificial Intelligence (AI, e.g. [11])—with its emphasis on the role of logic as a "mental representation" – -are not very different positions when contrasted to Wittgenstein. However, our focus here will be to contrast and compare Wittgenstein with developments specifically in NLP and computational linguistics, which has become more central within linguistics as a whole, as Chomsky's influence has declined, and not with the rule-driven paradigm (of Chomsky and in its different way, classical Artificial Intelligence) but with the more statistical paradigm that has replaced it.

Since Brown, one can hear new echoes in NLP of Wittgenstein's influence, as when Veronis called recently for looking "not for the meaning but the use" [12], thus reviving one of the best known Wittgensteinian slogans. One could hear it, too, in Sinclair's call to let a corpus "speak to one" [13], without the use of analytical devices and in Hanks' claim [14] that a dictionary could be written consisting only of use citations. This last may well be false, for it is hard to see the function of a dictionary that did not explain, but it does contain the authentic Wittgensteinian demand to look at language data, even if not in the way a linguist would mean who gave the same exhortation (i.e. to form a generalization from it, in the linguist's case).

Wittgenstein, of course, knew nothing of computers in the modern sense, although he trained as an engineer. All I can do in this brief paper is to more or less assume his views on language known to the reader, and to note what movements in modern NLP those are closer to and farther from, and why his arguments and insights should still be taken account of by those concerned to process language by machine. This paper will not be about scholarly claims of direct influence, for there are probably few to be found. Margaret Masterman [15] and perhaps the present author are two of the very few NLP researchers who acknowledged his influence and referred to him often. One thinks here, too, of Graeme Hirst's immortal and not wholly serious: "Most Artificial Intelligence programs are in Wittgenstein and only the degree of implementation varies", which only serves to show how much remedial work there is to do.

2 THE WORLD WIDE WEB AS A CORPUS OF USE

Wittgenstein's appeal to look for the use rather than the meaning is not, on its face, a clear injunction: elsewhere he writes of giving meanings by means of explanations (Blue Book [16] p27) and one may reasonably infer that the meanings NOT to look for are pointings at objects, and that when meanings are to be given they are in terms of more words, paraphrases (and not, he makes clear elsewhere, definitions) rather than an artificial coded language for meaning expression, such as that traditionally offered by logic, and later by linguistics and AI.

All this suggests an approach to actual language use more sympathetic than that usually associated with philosophers, and that was indeed the movement he created. Later, Quine, who made many of the same assumptions as Wittgenstein, explicitly linked looking at language use with the methods of structural (i.e. pre-Chomskyan or anthropological) linguistics, seeking data in languages not understood by the researcher, and drew a range of conclusions [6] very close to those of Wittgenstein, in particular that it was not mere language data that would do the trick but data in a language that was understood, by whatever process.

This also shows how wary one must be of trying, as Brown did, to place Wittgenstein somehow closer to the anthropological-empirical tradition than to Chomsky. It is true that Wittgenstein had something in common with the earlier writers, as Brown noted, but his emphasis on seeing language "from the inside", as something already understood and distinctively human, rather than as an object for scientific observation, brings him closer to Chomsky's emphasis on the native speaker and intuition. The truth is that, while Chomsky was a committed anti-behaviourist, Wittgenstein maintained an ambiguous position, one which declined to give the speaker veridicality on what he meant, so that he could not be wrong, a certainty Wittgenstein considered vacuous.

Among those who traditionally drew the attention of NLP researchers to data in large quantities were lexicographers, of linguistic or computational bent, as the remarks of Sinclair and Hanks above show. Since the return of machine learning and statistical methods to NLP, applied to large corpus data bases since the early 1990s, and following their proven success in speech recognition, NLP has taken large collections of text seriously as its databases. Recently, Kilgariff and Grefenstette [17] based a journal issue on the notion of "web as corpus": the use of the whole web in a given language as a corpus for NLP and, given Grefenstette's estimates [18], it is now clear the total of pages in English is up to forty times the number indexed by Google (currently more than 10 billion).

A corpus of that size is of course a data base of use/usage, one far greater than any human could encounter in a lifetime, and it is not structured in the way any human would encounter language, e.g. as dialogue, rather than prose, and graded appropriately for age on encountering it. But, of course, that is just a search problem, too, for there must be, in those 300 billion pages of English, a great deal of dialogue and child language at all levels. We must give up any idea that such a vast corpus could be a cognitive model of any kind: it would take a reader, reading constantly, at least 60,000 years to train on the current English web corpus. One can compare this with Roger Moore's observation that [19] if a baby learned to speak using the best

models of speech acquisition currently available, it would take 100 years to learn to talk.

The question we can now ask is, does that access to the whole web as a corpus by NLP research bring us closer to an ability to compute over usage in a language as a whole, to language surveyed in its full variety, rather than the examples an individual might think up, or generate from rules, or whatever? The odd answer seems to be that, although a web corpus, even now, only fifteen years after its inception, is so vast in human-life (i.e. of reading) terms, it is still no kind of full survey of language possibilities and never can be. And the reason for that lies not any kind of Chomskyan notion of novelty to do with the infinite number of sentences that can be generated from a finite base of rules.

For there is no finite base in any straightforward sense: as far as words (what some call unigrams) are concerned, it is clear they will continue to occur at a steady rate no matter how large the corpus [20]. This fact also holds for all forms of combinations of words. These are only examples of what is known as “data sparseness”, and maybe no more than a statistical/combinatoric updating of Chomsky’s point: as Jelinek has put it from a statistical point of view: “language is as system of rare events”. But is it vital to emphasise (since this whole discussion will have to be brought back to the notion of rules in due course) how wrong that finite base assumption of Chomsky’s was. Krotov induced all possible phrase-structure rules explicitly from the large passed corpus called the Penn Tree Bank (PTB) and plotted them against the length of the (part of) the corpus that gave rise to them. What was clear and astonishing was that at the end of the process –i.e. training on the whole of the PTB – the number of rules found (over 18K) was still rising linearly with the length of the corpus! It is quite unclear that there is any empirical justification for the idea of a finite syntactic base, at least for English, for that would require that that graph flattened at some point. This suggests that any rule base will continue to grow indefinitely with a new corpus, just as the (unigram) vocabulary does. There is no reason to think this tendency will change with much longer corpora; given that fact, assuming it is one, it is one hard to grasp within the history of modern formal linguistics. Chomsky took it simply as an article of faith that there must be a finite set of rules underlying a language, if only they could be written or found [21] suggests this is simply not so.

We are approaching a paradox here: there is an opposition, clear in Wittgenstein, to the notion of boundedness of a language implied by the rule-driven approach to a natural language, one found in Carnap, and which continued in Chomsky’s work. Wittgenstein wanted to question both that we could be said to be using any such rules and that any set of them could bound a language and determine well-formedness. Goedel’s results on undecidability in mathematics [22] must have seemed to him analogues from that world, and this is explicit in the Remarks on the Foundations of Mathematics [23] see also [24]).

However, just as it may be the case that the rule set for a language, like its sentence set, is not finite at all, so it may be the case that a corpus, for a language itself cannot be bounded, no matter how large it grows; or, rather, there is no corpus that captures the whole language, and so usage/use itself it not something finite that can be appealed to. One could, presumably, restrict oneself to all the sentences of English up to, say, 15 words long and bound that by permutations, but the problem

remains that the word set itself is shifting all the time: e.g. more than 900 words a year are being added to non-scientific English (The Times, 9/10/03).

Can Wittgenstein’s appeal to use be related to the fact that NLP over the whole web now surveys enormously more use than it did? It is clear that there can now be real experiments that appeal to use in a very satisfying way: Grefenstette, for example, (18) has described a novel algorithm for machine translation – following an earlier suggestion due to Dagan – in which a (two word) Spanish bigram XY is translated into, say, English by taking the n senses of a Spanish word X in a Spanish-English bilingual dictionary, and making a Cartesian product with the m senses of Spanish word Y, and then seeking the n x m resulting English bigrams in an English corpus and ranking them by frequency of occurrence. One may be confident that the most frequent one is always the correct translation.

This algorithm is in fact quite hard to explain and justify a priori: it feels exactly like “Asking the audience” in the popular quiz show “Who wants to be a Millionaire?” where, again, the most frequent answer from the audience is usually, but not always, correct, a phenomenon very close to what some would call the Google-view-of-truth, or what is now referred to as the “Wisdom of Crowds” But whatever is the case about that, there is no doubt this algorithm is precisely an appeal to use rather than meaning and a model for the future deployment of the web-as-corpus to solve linguistic problems.

3 BACK TO THE PRESENT STATE OF CL/NLP

Let us turn back now to the state of computational linguistics and NLP by computer. One could generalize very rapidly as follows: in the 1970’s, there arose movements such as Schank’s conceptual dependency or preference semantics (Wilks, [26]) which could be described as attempting to map a “deep grammar” of concepts and what I would call the preferential relations between concepts. This theme was closely allied with various forms of Fillmore’s [27] case grammar in linguistics, and his more recent work [28] could certainly be described as a continuing search for local, but deep, grammatical relations – based on systematic substitution relations in semi-fixed phrases in English – outside the concerns of the main thrust of work in computational syntax, which is little concerned with words themselves or local effects in language. Fillmore’s hand-coded lexicography just mentioned has been a survivor, but virtually all other attempts at conceptual mapping have been overtaken by one of the two separate movements to introduce empiricism into CL and NLP: the connectionist movement of the early 1980s, and the statistical corpus movement, driven by Jelinek’s successes in speech and then translation in the late 1980s. [29] The first was not a success but the second is still continuing: a classic of the first movement would be Waltz and Pollack’s [30] neural networks showing how concepts attracted and repelled each other in terms of contexts supplied to the network, from corpora or from dialogue. The work was exciting but such networks were never able to process more than tiny fragments of language. There were more radical (or “localist”) connectionists. such as [31] who went further and declined to start from explicit language symbols at all, in an attempt to show how symbols could have been reached from simpler associationist algorithms that built, rather than assumed, the symbols we use. If this had

been done it might have broken through the impasse that the title of this paper suggests, namely how can one have a theory of language which does not build in from the very start all that one seeks to explain, as intuition-based theories in linguistics, logic and AI always seem to. Connectionist theories could never give a clear account of the theory-free “simples” from which to begin, and in any case they also failed to “scale up” to any reasonable sample of language use, or to confirm any strong claims about human cognition of language.

The second movement, that followed connectionism, the one we are still within, at the time of writing, was statistical associationism, driven by Jelinek with his translation work derived from trigram models of speech [29], and which had some success and undoubtedly used language on a very large scale indeed, too large as we noted earlier, to be cognitively plausible for human beings. This movement has been committed to an “empiricism of use” but can such approaches ever build back to reconstruct concepts empirically? This movement, as we noted earlier, shares many assumptions with the technology of Information Retrieval (IR) (see e.g. [32]): a view that language consists only of words without the meta-codings that concepts and linguistic features” claim to provide, and that all the decorations and annotations that intuitive theories add are unexplained and unacceptable as explanatory theory. IR, it must always be remembered, underlies the successful search theories that have given us the World Wide Web search tools.

After his surprisingly successful machine translation project at IBM, done only by statistics, Jelinek became disillusioned with his first set of statistical functions and came to the view that language data is too sparse to allow the derivation of what he called a full trigram model of a language, which is to say, one that would have to be derived from a corpus so large that one could expect to have seen, when training on it, every trigram one could find in any text being tested subsequently—every possible sequence of three words the language allows (three here being an arbitrary number, cut off at a level where computation is unfeasible for larger numbers).

If I present this paper at the AISB, I will at this point briefly describe some recent experimental work with colleagues at Sheffield that suggests that Jelinek may have been too pessimistic, and that a full trigram model might now be within reach, using a device called a “skipgram”. These are “trigrams with gaps” or discontinuous trigrams, and one can expect to locate these in a smaller corpus than full trigrams with the same three elements. We have shown [33] that the whole corpus that would give all 3-slot skip grams is much smaller than that for true trigrams, can probably be computed without loss of generality, and from a corpus not much larger than the web now is. But first, let us ask what would be the point in a fuller associationist model like this, one that covered a language, English, say: how could having that get us closer to rebuilding concepts from all this data, on the assumption that that is what we really want to do, and is the key challenge of machine language understanding?

Let me give two simple examples of this, one from Jelinek’s own laboratory (REF) where they showed that simple association criteria could determine semantically coherent classes of objects far more easily than had been thought, provided one had enough data. One can see this most easily now on Google, where what was a research discovery fifteen years ago IBM is now a toy. On labs google.com/sets one can input

any small set of objects one likes and ask Google to find more, in response to this request, from the more than 8 billion English pages it indexes. So, if one types in “Scots, Bavarian, American, German”, Google replies with something like “French, Chinese, Japanese, etc”. In other words, it has “grasped the concept” of nationality words from context and is, as Wittgenstein would put it, able to go on. This is most certainly a derivation of something clearly semantic from nothing but word data, the problem being that the system does not know what the name of the class is!

A second notion is that of ontologies, forms of knowledge representation that have now become the standard way of looking at formalised knowledge in a wide range of AI, science, medicine and web applications: they contain technical and everyday information structured by set inclusion and membership as well as functional, causal etc. information about sets and objects, and they may or may not have additional strong underlying logical structure. The problem about such structures has always been, as with other forms of knowledge representation discussed here, that they are traditionally written down by human intuition. So what are we to make of the meanings of the terms they contain: are they referential or causal in meaning and can we gather anything from looking at their place in an ordered ontological hierarchy?

This is a straightforwardly Wittgensteinian question and the only proper answer is his own: namely that we cannot tell any term’s meaning by looking at it, only by seeing it deployed in use. It is a corollary of that view, assumed in this paper, that all such terms are terms in a language, the language they appear to be in (usually English), and that is so no matter how much their designers protest to the contrary. This is an issue discussed in detail in (Nirenburg and Wilks, [34]).

Ontologies, then, pose something of the problem here that logic traditionally does, or do formal features in linguistics (such as Fodor & Katz’ semantic markers [35]): they are claimed to be formal objects, kept apart from language and its vagaries, and with only the meanings assigned to them by scientists. But this isolation cannot in fact be maintained (see Mellor’s [36] dispute with Putnam on this issue [37]), and a more reasonable position is that ontologies will have justifiable meanings when they can be linked directly to language corpora, chiefly by being built automatically from them, and subsequently maintained automatically consistently with future corpora. An example of such a current project is ABRAXAS [38], one of a number of projects that claims to do exactly that.

Elements of an ontology can be thought of as triples (e.g. hand – PART OF – body) and our earlier references to skipgrams of trigrams hinted at a large amount of empirical research on using such apparently superficial methods to capture large volumes of such “facts” automatically from large corpora. Such methods go back to the very earliest days of NLP (e.g. [39]). Most recently, a new use for structures of this general type has appeared, namely the subject-relation-object triples (called RDF) that are to carry basic knowledge at the bottom level of the Semantic Web [40], the proposed structure intended to encapsulate human knowledge, based on the world wide web we now have, but annotated in a form to display something of a text’s meaning so that computers can use the web themselves. This is too large a vision to discuss here, but one last historical association may be worth making.

Bar Hillel [41] famously attacked the very possibility of machine translation (MT) on the ground that the kinds of

interpretation that translators make require knowledge of vast numbers of facts about the world, and machine translation would therefore need them too. So, you cannot interpret (and so translate) “carbon and sodium chloride” unless you know whether or not there is such a thing as carbon chloride, and so know the inner structure of that phrase (i.e. as carbon+sodium chloride versus carbon chloride + sodium chloride – and it is of course the first of those in this universe). Bar Hillel went on to argue that machines could not have such extensive knowledge of the facts of the world, and so MT was demonstrably impossible.

It was from exactly that point, conceptually if not historically, that AI set out on its long journey to develop mechanisms for representing all the facts in the world (of which the CyC project, [42], is the longest running example.) All this was done in a practical spirit, of course, with no thought or memory of Wittgenstein’s declaration that the world was the totality of facts (REF), and what if anything that could possibly mean. It was all practical, energetic computation rather than philosophical thinking, but still, in some sense, fell under Longuet-Higgins’ famous adaptation of Clausewitz, that AI was the pursuit of metaphysics by other means. It is interesting that empirically-based NLP has now brought back concepts like the derivation of a totality of facts, not painfully hand-constructed as in CyC, but extracted perhaps by relative simple means from the vast resources of the web’s corpora.

4 CONCLUSION

The new vision of the Semantic Web (SW)[40] is in part a revival of the traditional AI project to formalise knowledge, but now also a scientific reality in that so much of science and medicine is already encoded, indispensably so, in structures of this general sort. The process of its construction requires giving meaning progressively to the “upper level” concepts in its ontologies. These upper level concepts are still written down by intuition, which may have validity in scientific area if done by experts – who else can write a map of biology? – But is as much at risk as all the knowledge structures in AI if not grounded in something firmer. I want to argue, in conclusion, that the future SW may offer the best place to see the core of a Wittgensteinian computational linguistics coming into being, as a way of grounding high-level concepts, such as the primitives at the tops of ontologies (e.g. neutrinos, Higgs boson, genes), in real usage of the sort we see in the web-as-corpus.

What I think we are seeing in the SW is a growing together of these upper conceptual levels based on the name spaces and RDF triples derived from texts by skip grams or richer techniques like Information Extraction [43], a successful shallow technology for extracting items and facts that now rests wholly on the success of automated annotation. My belief is that the top and bottom levels will grow together and that interpretation or meaning will “trickle up” from the lower levels to the higher: this is the only way one can imagine the higher conceptual labels being justified on an empirical base. It is a process reminiscent of the concept of “semantic ascent” pioneered by Braithwaite [44] as a description of the way in which interpretation “trickled up” scientific theories from observables like cloud-chamber tracks to unobservables like neutrinos. It is hard to imagine any other route from the distributional analysis on which the revolution in language processing rests up to the interpretation of serious scientific concepts. It is also a process reminiscent of

Kant’s dictum synthesising Rationalism and Empiricism: “Concepts without percepts are empty; percepts without concepts are blind.”

I would argue that the SW is a development of great importance to AI as a whole, even though we still dispute about what it means, and how it can come into being. Many seem to believe that it means Good Old Fashioned AI (GOF AI) is back in a new form, a rebranding of the old tasks of logic, inference, agents and knowledge representation. Core AI tasks have come to something of an impasse: we do not see them marketed much in products after fifty years of research. But a key feature of the SW is that its delivery must be gradual, coming into being at points on the World Wide Web (WWW), probably starting with the modelling of biology and medicine. One cannot easily imagine how it could start somewhere completely new, and without being piggy-backed in on the WWW, yet it will be much more than those same texts “annotated with their meanings”, as some would put it.

The key possibility I think the SW offers to traditional AI is to deliver some of its value in a depleted form initially, by trading representational expressiveness for tractability, as some have put it. The model here could be search technology and machine translation on the WWW (or even speech technology): each is available now in forms that are not perfect but we cannot imagine living without them. This may all seem obvious, but machine translation has only recently crossed the border from impossible (or failed) to commonplace. It is far better for a field to be thought useful, if a little dim at times, than impossible or failed. It will be important that web services using the Semantic Web are chosen so as not to be crucial, but merely a nuisance, should they fail. My own current interests are in lifelong personal agents, or Companions, conversationalists as well as agents, where it should not matter if they are sometimes wrong or misleading, any more than it does for people.

This view of the future of the SW is personal and partial; many researchers do not see the need to justify the meanings of logical predicates or ontological terms any more now than they did when they set out in AI and representation in the Sixties. But the history of the CyC project is a good demonstration, if one were needed, of why that cannot be a foundation for AI in the long term: in that project it has not proved possible to keep the interpretation of logical predicates stable over the decades of the system’s development; this is a highly significant long-term experimental result for those who believe in the immutability of the meanings of formal items. There is a related view, also current in the SW, that meanings will be saved or preserved by trusted data bases of objects (URIs), referential items in the world, rather in the way digit strings “ground” personal phone numbers in a data base. But this way out will not protect knowledge structures from the changes and vagueness of real words in use by human beings. Putnam considered this problem in the Sixties and declared that scientists should therefore be the ultimate “guardians of meaning”. As long as they knew what “heavy water” really meant, it did not matter whether the public knew and perhaps better if they did not. But people call heavy water “water” because it is – because it is indistinguishable from water – otherwise it would have been called “deuterium dioxide”. We, the people, are the guardians of meaning and “getting meaning into the machine”, probably via the SW, should entail doing it our way, and what could be more in the spirit of Wittgenstein than that?

REFERENCES

- [1] Y. Wilks. The History of Natural Language Processing and Machine Translation. In *Encyclopedia of Language and Linguistics*, Kluwer: Amsterdam. (2005).
- [2] N. Chomsky. *Syntactic Structures*, Mouton: The Hague. (1957)
- [3] R. Carnap, *Logische Syntax der Sprache*. English translation 1937, The Logical Syntax of Language. Kegan Paul, London. (1936)
- [4] C. H. Brown, *Wittgensteinian Linguistics*. The Hague: Mouton & Co., (1974)
- [5] B. Malinowski, The problem of meaning in primitive languages. In: C.K. Ogden & I.A. Richards (Eds.), *The meaning of meaning*, pp. 296-346. London: Routledge & Kegan Paul. (1923)
- [6] W. V. O. Quine, *Word and Object*, Cambridge, Cambridge UP (1960).
- [7] L. Wittgenstein, *Tractatus Logico-philosophicus*, Routledge: London. (1961)
- [8] Wittgenstein, L. *Philosophische Untersuchungen, Philosophical Investigations*, 2nd ed. Oxford: Basil Blackwell, (1958)
- [9] G. Sampson, G: *The 'Language Instinct' Debate*, Continuum, (2004).N.
- [10] N. Chomsky, (1985) Aspects of the Theory of Syntax. Cambridge: The MIT Press.
- [11] J. McCarthy and P.J. Hayes. (1969) Some philosophical problems from the point of view of Artificial intelligence. In *Machine Intelligence 4*, (Eds.) Michie and Meltzer, Edinburgh, Edinburgh UP.
- [12] J. Veronis, (1993) Sense tagging, does it make sense? <http://citeseer.ist.psu.edu/685898.html>
- [13] R. Moon, (2007) Sinclair, lexicography, and the Cobuild Project: The application of theory. *International Journal of Corpus Linguistics*, Volume 12, Number 2, 2007.
- [14] K. Church, W. Gale, P. Hanks, D. Hindle, (1989) Parsing, word associations and typical predicate-argument relations, Proceedings of the workshop on Speech and Natural Language, October 15-18, 1989, Cape Cod, Massachusetts.
- [15] M. Masterman, (2006) Language, Cohesion and Form: selected papers, (Ed.) Y. Wilks, Cambridge UP, Cambridge.
- [16] L. Wittgenstein, (1958) The Blue and Brown Books, Oxford: Basil Blackwell.
- [17] A. Kilgarriff, G. Grefenstette (2003). Introduction to the Special Issue on Web as Corpus. *International Journal of Corpus Linguistics* 6 (1).
- [18] G. Grefenstette, (2002) Lecture, Sheffield University.
- [19] R. K. Moore, (2007) Spoken language processing: Piecing together the puzzle, *Speech Communication*, 49.
- [20] T. Dunning. (1993) Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*.
- [21] A. Krotov, R. Gaizauskas, and Y. Wilks. (2001) Acquiring a stochastic context-free grammar from the Penn Treebank. In Proceedings of Third Conference on the Cognitive Science of Natural Language Processing.
- [22] K. Goedel. Ueber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. In Solomon Feferman, (Ed.) *Kurt Goedel: Collected Works, volume 1*, pages 144–195. Oxford University Press, German text, parallel English translation. (1986)
- [23] Wittgenstein, L. *Remarks on the Foundations of Mathematics*, rev. edn, ed. G. H. von Wright, R. Rhees, and G. M. Anscombe, trans. G. E. M. Anscombe, Cambridge, MA: MIT Press. (1978)
- [24] Y. Wilks, Y. Decidability and Natural Language, *Mind* LXXX (1971).
- [26] Y. Wilks, Y. Preference Semantics, In *The formal semantics of natural language* (Ed.) E. Keenan, Cambridge, Cambridge UP (1975)
- [27] C. Fillmore, The Case for Case. In Bach and Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 1-88. (1968)
- [28] C. Fillmore, Frame semantics and the nature of language, In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*. Volume 280: 20-32. (1976).
- [29] P.F. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, J. Lafferty, R.L. Mercer, P. Roossin. A Statistical Approach to Machine Translation. *Computational Linguistics* 16:2: 79-85, (1990)
- [30] D. L. Waltz, J. B. Pollack: Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation. *Cognitive Science* 9(1): 51-74 (1985)
- [31] T. Sejnowski and C. Rosenberg. Parallel networks that learn to pronounce English text. *Complex Systems*, 1:145–168, (1987).
- [32] A. Singhal. Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (4), p 35-43. (2001)
- [33] D. Guthrie, B. Allison, W. Liu, L. Guthrie, Y. Wilks, Y. A Closer Look at Skip-gram Modelling. In *Proc. Fifth International Conference on Language, Resources and Evaluation (LREC'06)*, pp. 1222-1225, (2006).
- [34] S. Nirenburg and Y. Wilks. What's in symbol. In *Journal of Theoretical and Experimental AI (JETAI)* (2000)
- [35] J.J. Katz and J. Fodor. The structure of a semantic theory, *Language* (1963).
- [36] D. H. Mellor, Natural Kinds, *British Journal for the Philosophy of Science* 28 (1977).
- [37] H. Putnam, Is Semantics Possible? *Metaphilosophy* 1 : p187-201. (1970)
- [38] C. Brewster, H. Alani, S. Dasmahapatra, and Y. Wilks. Data-driven Ontology Evaluation. In *Proc. of 4th International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. (2004)
- [39] Y. Wilks, *Text Searching with Templates*. Cambridge Language Research Unit Memo, ML.156. (1964)
- [40] T. Berners-Lee, J. Hender, and O. Lassila, The Semantic Web, *Scientific American*, May 2001, p. 29-37. (2001)
- [41] Y. Bar Hillel, *Language and Information*. Reading, MA: Addison Wesley. (1964)
- [42] D. Lenat and R. V. Guha. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley. (1990)
- [43] http://en.wikipedia.org/wiki/Information_extraction
- [44] Braithwaite, R. *Scientific Explanation*, Cambridge UP, Cambridge. (1956)