Homer, the Author of *The Iliad*
and the Computational-Linguistic Turn

Sergei Nirenburg

*Institute for Language and Information Technologies, University of Maryland,
Baltimore County, MD, USA*

**Abstract:** This paper analyzes two sets of opposing opinions about the nature of meaning representations and knowledge resources. The first of these axes of disagreement is the opposition between an ineffable, "revealed" language of thought in the Fodor tradition and Wilks' position that (using its strongest formulation) elements of the language of knowledge representation are essentially elements of a natural language. The second opposition is between a "scientifically" defined ontology, in Guarino's sense, and human-oriented resources of knowledge about language, such as MRDs or WordNet. An attempt will be made to clarify some of the motivation behind these differing opinions. I will try to formulate my own positions on the above issues and will use as illustrations some modules of ontological semantics, a computationally-tractable theory of meaning, as implemented in the OntoSem text analyzer and the knowledge resources that support it.

This paper continues an ongoing discussion of the nature of natural language processing (NLP), specifically, AI/NLP, that strain within NLP that is closest to the concerns of artificial intelligence (AI). AI/NLP studies issues of how to build machines that understand and generate language in a coherent manner approaching that of people. I happen to adhere to a representationalist approach that is concerned with explicit modeling of the processes of understanding and generating text.¹

¹ Such modeling relies on knowledge that might include but is not reducible to stochastic methods, a form of case-based text processing carried out essentially by reference to human performance recorded in large text corpora. In all essentials, the questions that such case-based reasoning asks are best formulated in terms of text generation, e.g., "What did people say/write next given what has already been generated up to the current point in the current text/sentence?" In this paradigm, it is not necessary (and not possible, without major modifications to the supporting knowledge) to seek the meaning of the text (for analysis) or the intention of the speaker (for generation). In fact, complex statistics of co-occurrence of natural language elements in documents is sometimes (as in latent semantic analysis, Landauer et al. 1998) declared to constitute the meaning of natural language texts. Case-based approaches may be viable, at least, in conception, for tasks that can be interpreted as relying, at the core, on comparisons of textual strings. This is why the successes of such approaches include such applications as detecting cheating in student essays, establishing text authorship or (possibly, more controversially) machine translation.
The representationalist approach involves modeling memory and language behavior using methods that potentially can help to provide answers to questions like *What does this text mean? What goals does the author pursue? What new knowledge do I gain from it about my model of the world, including knowledge about types and about tokens of elements of that model?* How does this text alter my agenda of goals and plans? etc. This approach requires a knowledge apparatus that is a combination of knowledge about the world, knowledge about language and knowledge about the speech situation, in the broad sense that applies also to reading and writing, not only dialog. Knowledge about the world includes knowledge about imaginary, hypothetical and non-existent entities as well as those that really exist. It is also multi-faceted and includes knowledge about types of object and events and their instances. The knowledge is used to support text understanding and generation in a variety of NLP systems that rely on methods beyond sophisticated comparisons of textual strings.

My points of departure in this paper are some of Wilks’ (2001) arguments against Fodor and Lepore’s (1998) criticism of Pustejovsky’s (1995) generative lexicon and his criticism (Wilks, 2002) of Guarino’s (1997, 1998; Gangemi et al. 2001) program of making the nature and format of ontologies for NLP sufficiently precise. Fodor, Lepore and Guarino are representationalists, just like Wilks (and Pustejovsky), so in this sense all the protagonists belong to the same large camp that stands in opposition to various forms of asemanticism – behaviorism, connectionism, statistical NLP, etc. But within the representationalist camp, further divisions appear. Thus, Wilks and, to a large degree, Pustejovsky are what in AI circles is informally known as “scruffies,” on account of their preference for dealing with the messy subject matter first and with developing clean metalanguages for the description of this subject matter later. Within the representationalist camp, the “scruffies” stand in opposition to the “neats” who start with idealized, almost always logic-based, languages and only having honed these ideal tools by studying their properties, turn to using them as metalanguages for the description of messy reality. Guarino is one of many “neats.” For the purposes of this paper, we will induct Fodor and Lepore into this cohort, too (though they might be offended to be so cavalierly bunched with AI types). Of course, the delimitation of camps or factions is always relative and unstable – consider, e.g., Sowa’s (2000) demonstration of similarities between the theories propounded by such apparent opposites as the formal philosopher Montague and the quintessential “scruffy” Schank.

I would like right away to make a disclosure: my work in the area of Ontological Semantics (Nirenburg and Raskin 2004) makes me squarely a “scruffy.” It is not surprising, therefore, that I broadly agree with the opinions and especially the methodological and paradigmatic positions of Wilks (and Pustejovsky). But there are a few, as it were, intra-factional disagreements on which I would like to comment. Some of my arguments will revisit the topics of the dialog with Wilks in Nirenburg and Wilks (2001) concerning the nature of symbols in meaning representations. I will comment on some issues related to the nature of the representation language, the purpose of procedures in largely declarative knowledge representations and ambiguity and paraphrasing in static knowledge resources for NLP. In doing so,
I will make use of some examples of representational solutions in OntoSem, the most recent implementation of ontological semantics. I will also suggest that, while a century ago analytical philosophy turned to the study of language as a safer and more productive way of studying reality (or, at least, as a necessary precursor to studying reality), one way to bring philosophy into the twenty-first century is for it to describe a computational-linguistic turn and start discussing the use of natural language and languages of meaning representation by intelligent agents—either computer programs or people who build, study or use them.

9.1 Language ↔ LANGUAGE

Here is what Wilks has recently had to say about the nature of symbols in knowledge representations.

"Meaning in the end is best thought of as other words and that is the only position that can underpin a lexicon and procedure-based approach to meaning and one should accept this—whatever its drawbacks—because the alternative is untenable and not decisively supported by claims about ostension." (Wilks F "F")

"The persistent, and ultimately ineradicable, language-likeness of purported ontological terms... means that we cannot ever have purely logical representations, purged of all language-like qualities." (Wilks Ontotherapy).

So, for Wilks, language itself supplies the lexical stock of the metalanguage for the description of meaning. This is, incidentally, not a claim that all metalanguage is a natural language, it is a claim about the nature of its symbols. Can the lexical elements of a metalanguage really be just "other words?" If so, there seem to exist a number of peculiarities that make these "metawords" quite dissimilar from the words that appear in regular texts and those whose meanings are defined in human-oriented dictionaries (let's call them simply words).

First, metawords would have to be learned by native speakers of the language in question, or else they would not be able to make simple inferences. This is true if we assume that, as is traditionally held, the metalanguage for representing meaning should not be ambiguous. Indeed, the actual interpretation of table (elements of the metalanguage are presented here in small caps) will fail if this metaword is not somehow disambiguated at least between the furniture and diagram senses. At the same time, something like table-furniture and table-diagram, if used as metawords, are already not quite elements of the object language—while I have not checked it, I do not think that such words will be attested in many corpora of English. Moreover, the meaning of even such complex metawords would often be only approximately understood by the native speaker who does not have access to their definitions (cf., e.g., the symbols for-profit-corporation or involuntary-olfactory-event in the OntoSem ontology). These difficulties have been observed over and over again during the process of ontology acquisition in the OntoSem environment.
One can, of course, argue that the symbols of the metalanguage need not be unambiguous and support this claim by observing that meaning of a text could be generated dynamically—just like people do it!—from a collection of ambiguous symbols appearing in it, and, therefore, the symbols of the metalanguage need not be anything but words in a language. Of course, this position concentrates on word senses and their disambiguation; under this approach, as under others, there will still be a need for extracting and representing elements of meaning that do not, in the general case, reduce to the meanings of words and phrases (e.g., causality, modality, and many other phenomena). But even within that rather narrow purview, there are problems with this approach. It simply passes the buck to the processing component of the system from the static knowledge component. As a corollary, any indexing, co-referring or use of such “potential” meaning representations in reasoning becomes inefficient and cumbersome. Indeed, to use such a representation as a source of heuristic knowledge in the left-hand side of an inference rule, one will first have to run the dynamic disambiguation procedure, and do it every time. This model may be an acceptable hypothesis in studying human language understanding. However, when one must deal with the constraints of computer hardware, one is best advised to follow the precepts of dynamic programming and to store any intermediate results overtly so that they can be later accessed and used efficiently.

In general, it seems to me that those who insist on word-like character of symbols tend to take for granted the amazing meaning extraction ability of people and do not make sufficient allowances for the major constrains under which computer modeling of these processes must operate: the unfortunate need to make all knowledge used by computers explicit and unambiguous.²

Second, metawords can be treated as regular terminology (e.g., chemical nomenclature; indeed, a lay person cannot be expected to know the names of chemical substances), and ontological concepts can be explained using natural language in a manner similar to the content of human-oriented lexicons (as is advocated, e.g., in Brewster et al. 2005). This will result in lexicons that are primarily compiled to support the operation of computer systems but are also intended for people. To underscore the special status of basic semantic primitives, Wilks (Wilks et al. 1996: 19) uses the example of instructions on how to use the phone in a foreign

² Given these premises, it is not surprising that over the last 50 years or so suggestions, plans and promises were repeatedly made concerning teaching computers to understand just like people. However logically simple and appropriate this desire may be, it is only attainable if the machine can indeed learn, and the easiest way toward that is for the machine to be able to understand language. Attempts to bootstrap this ultimate learning from very partial means may lead to some local advances but do not seem to hold much promise for the main goal of building an intelligent machine that can manipulate language with a facility similar to that of people. It does not seem that this task can be finessed in any serious way: one must create much more knowledge in machine-tractable form before the machine is able to understand and use this ability—among other things—to enhance its own understanding ability autonomously.
country. He says that the “text assumes you already understand all the ‘manipulative primitives’ such as HOLD-RECEIVER-TO-EAR or INSERT-COIN.” A lexicon/ontology intended for a computer program would have to find a way of explaining, in a formal metalanguage, these and other notions that are assumed to be known by humans. However, since the entries cannot presuppose human background knowledge, they will end up irritatingly detailed, inelegant and difficult for people to read.

Third, if metawords are viewed as regular terminology, then, because of their use by the linguistic community, sense shifts, sense splits and other modifications will be attested in time. Such changes will have to be recorded in human-oriented lexicons. But people, unlike machines, have other means of learning new senses of the words – from “live” text understanding. As machines lack this ability, there is no nontrivial sense in which metawords used in a machine-oriented knowledge resource will change their meaning over time. Semantic interpretations might “take a life of their own” in applications and lead to unexpected closures and inferences but this would be due to human error (or, more likely, inconsistency and lack of anticipation of the modes of use of the knowledge). In the case of machine-oriented resources such errors can be amplified much more than in human-oriented lexicons or thesauri where the most such errors may lead to would be a situation where the reader may gloat in his or her self-perceived superiority to the lexicographer on seeing an inconsistency or a definition that is too tightly circular.

So, if meaning is indeed other words, as Wilks suggests, then these metawords are really different from “regular” words in their use, they must be learned anew by the native speakers and they should resist any semantic shifts or other “unplanned” fluctuations. This, plus the needs of an NLP system, really impart a special status to the metawords and make them in reality much closer to the status they are assigned in NLP-oriented ontologies.

How far should the metaphor of representation languages being natural languages be stretched? There is some circumstantial evidence that people separate word senses from meanings. Consider the verbatim recall experiments most of which show that people are able to recall the meaning of some texts much better than the exact way that meaning was formulated in the text. For example, Chafe (1977) reports results of delayed-recall experiments that demonstrate that people not only forget the actual words but also habitually report not the actual meaning of the utterance they are asked to recall but rather a presupposition or an entailment of it. This suggests that results of text (or speech) understanding are recorded in a language-independent way after (usually subconscious) disambiguation and when they are used as components of input to the process of text (speech) production, the (lexical, syntactic and other) means of their realization in language are not directly recalled from the results of earlier analysis but are rather selected from the typically broad inventory of synonymic realization means available to the language producer/consumer. Recall experiments have even been used in forensics (e.g., Johnston 2004). They suggest that people store knowledge in terms other than natural language word senses.
People who suffered temporary aphasia reported a disconnect between language of thought and language of communication. Thus, the nineteenth century French clinician Jacques Lordat went through a period of aphasia and recorded his observations as follows: “Within twenty-four hours all but a few words eluded my grasp. Those that did remain proved to be nearly useless, for I could no longer recall the way in which they had to be coordinated for the communication of ideas... I was no longer able to grasp the ideas of others, for the very amnesia that prevented me from speaking made me incapable of understanding the sounds I heard quickly enough to understand their meaning... Inwardly, I felt the same as ever... I used to discuss within myself my life work and the studies I loved. Thinking caused me no difficulty whatever... My memory for facts, principles, dogmas, abstract ideas, was the same as when I enjoyed good health... I had to realize that the inner workings of the mind could dispense with words” (quoted in Kapur 1997). Sacks (2005) reports other similar cases. While it is possible that the language of thought used by Lordat was the same as the one he could not temporarily use for communication, it is equally possible that the two functions of thinking and communication are actually carried out using different languages.

The Iliad was written not by Homer but by another man with the same name. Why is this funny? A plausible answer is, because this sentence violates the hearer’s expectation of unambiguity of representational symbols (or structures represented by symbols) in a felicitous dialogue.3 (It is probably beside the point that this answer also casts doubts on the degrees to which designators in this most real of the possible worlds are really rigid; moreover, I am not sure that Kripke (1982) and others after him considered this rigidity to be a scalar value.) Analysis of this sentence is not just a matter of simple reference resolution and certainly not ostension – few of us can boast to be able or to have ever been able to point to a person and felicitously say that this is Homer (or even “This is Homer Simpson”). The language symbol is ambiguous (it can refer to any person named Homer). However, if we substitute, say, John, for Homer (or “This novel” for “The Iliad”?) in the input text, it stops being funny. Moreover, in the absence of conversational context, the sentence with John instead of Homer should elicit a clarification question: “John who?”

It is only because we have a separate, non-language related representation for our knowledge about Homer the bard that the above sentence works as a joke. In this representation, the name itself is just one of its properties. Others will identify this individual, among other things, as the author of the Iliad, as having been blind and – even – as possibly never having existed. This representational structure is, incidentally, not a part of the ontology (or the lexicon or the thesaurus). It is a part of a model of the human long-term memory of instances of ontological concepts, with their own values of ontological properties defined in the ontology. In ontological semantics this collection of assertions is called fact repository (FR). FR is an important source of heuristics for both general reasoning and semantic

3 On the semantic analysis and generation of humor see, e.g., Raskin (1986) and Hempelmann et al. (2006).
analysis itself. Thus, ontological semantics includes knowledge obtained from FR in determining the preferences of preference semantics.

9.2 Science and Metaphors

Experience shows that making an intelligent agent's knowledge sound and consistent, as Guarino recommends, and at the same time expecting it to be sufficiently deep and broad to support realistic reasoning applications is a tall order. The same argument goes for the agent's reasoning processes. If one has to decide which of these conditions should be relaxed, I side with Wilks in believing that breadth and depth of coverage are more important than formal discipline. Guarino's requirement is imported from formal logic and its extensions, such as model-theoretic semantics.

Is logic necessary for AI/NLP? I agree with Wilks that “[s]emantic well-formedness is not a property that can decidedly be assigned to utterances in the way that truth can to formulas in parts of the predicate calculus and as it was hoped for many years that syntactically well-formed structures would be assignable to sentences.” (Fodor “Fodor”).

Still, importing methods, biases and metaphors from other fields of science and technology is often a valid strategy and has a venerable history. For example, in the seventeenth century Harvey had the idea of blood circulation by making a mental analogy with pumps. In the late nineteenth century the Young Grammarians studied laws of language change because of the influence of the Origins of Species. In the twentieth century Sapir and Whorf overtly used the relativity metaphor from physics. It is therefore plausible that the requirements that Guarino would like to impose on an object that he would be happy to call an ontology are the result of metaphorical thinking and projection from another field, in this case, formal logic. (It is immaterial for this argument that formal logic, at its inception, was motivated by the desire to analyze natural language; this desire could be interpreted in the light of seeking formal ways to distinguish texts about texts from texts about the world, a concern associated with the linguistic turn mindset in analytical philosophy. Formal logic is a discipline unto itself and its methods are true imports into the study of language, as is corroborated by the title of McCawley's (1981) book: “Everything that linguists have always wanted to know about logic (but were ashamed to ask).”)

Sometimes such metaphorical thinking and the importation of ideas and methods is fruitful. In some other cases, it succeeds mostly to remind one of the well-known saying to the effect that if one has a hammer then everything looks suspiciously like a nail. When such cross-pollination succeeds, it almost always requires significant modification of the original metaphor: there are probably more differences than similarities between actual hearts and actual pumps. In the field of NLP systems, such modification practically always involves the introduction of large amounts of descriptive detail to the original statements of the imported theory. Often, the needs of such description lead to extensions of the original theory (e.g., to take the example application of text generation, extensions of systemic functional grammar
theory (e.g., Halliday 1985) by Mann and Matthiessen (1983) or of the Meaning-
Text theory (e.g., Mel’čuk 1995; Mel’čuk et al. 1995) by Iordanskaja et al. (1991)).
In the end, the underpinnings of the resulting systems may bear only a passing
resemblance to the original theoretical statements. In fact, it is this phenomenon that
explains Wilks’ seemingly paradoxical “theorem” stating that “there is no linguistic
theory, no matter how bad, that cannot support the development of a successful
NLP system.”

The above does not mean that metaphors should not be used to drive progress
in science, simply that some such metaphors are a better fit than others. Using
formal logic as the means of “shaping up” descriptions of world knowledge and
knowledge of language is a corollary of the desire to eliminate inconsistencies
and to make the descriptions usable by current formal reasoning systems. But this
position assumes the expressive power of representations to be a handmaiden to the
formal properties of the metalanguage. If we are to make progress in attaining a
human level of performance in intelligent software agents, such a paradigm is too
limiting. People operate reasonably well with knowledge bases that are incomplete
and unsound. So, instead of constraining the expressive power of representations
to gain immediate partial results, one may consider keeping expressive power
unchained and developing richer, if possible, less general and more knowledge-
dependent methods of reasoning, including reasoning for text understanding. (An
interesting example of a graduated attitude to the tension between expressivity and
computability is the decision to include several “dialects” in OWL, the knowledge
representation language for semantic web applications (Java et al. 2005, 2006).)

I think that there is a more apt metaphor for AI/NLP work than the formal logic
one. It is the psychological metaphor introduced into economics by Kahneman,
Twersky and their associates (e.g., Kahneman et al. 1982). It suggests, roughly,
that *homo economicus* should not, as was done previously, be viewed as a fully
rational agent operating using well-defined utility functions but rather as a complex
cluster of reason, habit and emotions. As a result, predicting behavior of people and
groups in economics-related areas is becoming even more complex, as it involves
even more variables than previously thought. The Nobel prize awarded in 2002 to
Kahneman and others for this work is an oblique endorsement of the primacy of
descriptive adequacy over available reasoning methods.

Indeed, it is difficult to expect that a realistic model of an intelligent agent
in general can be created solely on the basis of deductive reasoning over sound
and complete knowledge bases. This became clear to logicians in early twentieth
century, and they have been hard at work on alleviating this state of affairs. A
number of modifications of reasoning tools (in the form of specialized logics) have
been suggested. Thus, allowing abductive reasoning helps to partially alleviate the
brittleness of strong logical methods. Attaching probabilities to knowledge and
inference processes also helps with applications but excludes from the purview the
task of explaining and motivating the interpretation of utterances. All such tools still
presuppose some version of soundness and consistency of the data. Otherwise, it is
believed, the knowledge representation and reasoning enterprise would be judged
as not scientific.
Guarino’s concern is strictly scientific: enforcing formal constraints on ontologies (and, by extension, all formal representations of knowledge about the world and the language). Wilks’ polemical position is that this desire is misguided and futile, in part on the grounds that the complexity of the world and language makes it unlikely that any strict formalism will adequately describe it. And if this approach is called “unscientific”, so be it. It is not entirely clear whether it is under all circumstances that being labeled “unscientific” should be considered an insult. Newton and Galileo were acclaimed as the first scientists not because they knew more than their old-style contemporaries or predecessors but (at least, in part) because they changed the purview of their inquiries. While the “medieval scholars” attempted to catalog and reconcile a vast number of individual facts, thoughts and observations in a coherent whole, the scientists curtailed the purview of their pursuits and concentrated on a smaller set of phenomena about which they could make coherent statements (theories) corroborated by experimentation and successful applications. Thus, science should be associated, among other things, with relying on the notion of the “wastebasket” – a set of phenomena that it – consciously or not – chooses to ignore. In other words, science can be considered to be the art of the possible. If, as a result, the description of the phenomenon under investigation is at least sufficient to support applications, fine. If there are important applications that are not supportable, then the method should be somehow extended (or relaxed) to allow reformulation under less strenuous requirements.

9.3 Ontology ↔ ONTOLOGY

“Items in ontologies and taxonomies are and remain words in natural languages – the very ones they seem to be, in fact – and this fact places strong constraints on the degree of formalisation that can ever be achieved by the use of such structures.” (Wilks, Ontotherapy). This statement argues against the plausibility of Guarino’s program of “firming up” existing NLP-oriented ontologies by applying formal constraints, such as identity criteria, or making the concept of ontologies more precisely defined, for instance, on the basis of set inclusion and eschewing the notion of properties. It is not clear that it is productive to talk about whether thesauri and lexicons are different from ontologies, because this does not help, beyond maybe establishing battle lines. In fact, I am quite prepared to admit that OntoSem represents meaning using not ontology but an entirely different construct referred to by the same name.

Traditional definitions tell us that ontologies must describe what is there in the real world. The world of a transparently intelligent software agent should include many very different things. It would be in some sense like my world, which includes, among a vast amount of other entities, knowing that violets are blue and that when people feel sudden pain they might say “Ouch” in English and that Horatio Nelson died in 1805 but also that Horatio was a friend of Hamlet’s and studied at Wittenberg and that I am glad that I remembered to send my mom flowers for her birthday the other day and that my friend Jack does not like swimming.
The agent’s world view, just like that of a human agent, would also accommodate contradictions. Indeed, people can and do operate rather well with incomplete and ambiguous knowledge... I would like a truly intelligent agent to attain the same capabilities. It is clear that the state of the art in 2006 cannot support this level of sophistication. However, it seems important to me that the ultimate goal of building such an agent remains central – especially because much of the recent and current work in AI and NLP eschews such “remote” objectives and concentrates on picking the “low-hanging fruit.”

In an argument about what kind of knowledge to impart to an intelligent software agent – an ontology or a lexicon/thesaurus – there is a third option. One can require the agent to have both, and then some. Semantic and pragmatic interpretation must rely not only on knowledge of lexical meaning, it also requires contextual knowledge and world knowledge. I will touch on processing context in connection with procedural semantics later. In this section, I will suggest how factual knowledge that was not directly mentioned in the text can and should be leveraged for extracting text meaning. To prepare the reader for this discussion, I must take a brief detour to introduce the terms and notions of ontological semantics.

The OntoSem system is the latest implementation of the theory of ontological semantics (Nirenburg and Raskin 2004). It is a text-processing environment that takes as input unrestricted raw text and carries out preprocessing, morphological analysis, syntactic analysis, and semantic analysis, with the results of semantic analysis represented as formal text-meaning representations (TMRs) that can then be used as the basis for many applications. Text analysis relies on:

- The OntoSem language-independent ontology, which is written using a metalanguage of description and currently contains around 8,500 concepts, each of which is described by an average of 16 properties.
- An OntoSem lexicon for each language processed, which contains syntactic and semantic zones (linked using variables) as well as calls for procedural semantic routines when necessary. The semantic zone most frequently refers to ontological concepts, either directly or with property-based modifications, but can also describe word meaning extra-ontologically, for example in terms of modality, aspect and time. The current English lexicon contains approximately 30,000 senses, including most closed-class items and many of the most frequent and polysemous verbs, as targeted by corpus analysis. The base lexicon is expanded at runtime using an inventory of lexical rules. (An extensive description of the lexicon, formatted as a tutorial, can be found at http://ilit.umbc.edu.)
- An onomasticon, or lexicon of proper names, which contains approximately 350,000 entries.

---

4 As concepts, specifically, ABSTRACT-OBJECTS, aspect, modality and time do belong to the ontology. But as features used in describing meanings of texts in ontological semantics, they don’t carry ontological status, because their values pertain to contextual meaning, not persistent knowledge about objects and events.
- A fact repository, which contains real-world facts represented as numbered "remembered instances" of ontological concepts (e.g., SPEECH-ACT-3366 is the 3366th instantiation of the concept SPEECH-ACT in the world model constructed during the processing of some given text(s)).

- The OntoSem syntactic-semantic analyzer, which covers preprocessing, syntactic analysis, semantic analysis, and the creation of TMRs. Instead of using a large, monolithic grammar of a language, which leads to ambiguity and inefficiency, we use a special lexicalized grammar created on the fly for each input sentence (Beale et al. 2003). Syntactic rules are generated from the lexicon entries of each of the words in the sentence, and are supplemented by a small inventory of generalized rules. We augment this basic grammar with transformations triggered by words or features present in the input sentence.

- The TMR language, which is the metalanguage for representing text meaning.

OntoSem knowledge resources are at this time acquired primarily manually (though note that the knowledge acquirers use a variety of efficiency-enhancing tools—graphical editors, enhanced search facilities, capabilities of automatically acquiring knowledge for classes of entities on the basis of manually acquired knowledge for a single representative of the class, and the like). The ontology has been under continuous development, with varying levels of effort, for around 20 years. It took approximately two and a half years of work by a PhD-level linguist to compile the current lexicon. (Although the OntoSem environment has always utilized an English lexicon, previous versions aimed for a coarser grain-size of description and did not reflect recent theoretical and practical advances). The ontomasticon was extracted automatically from corpora and structured sources. The fact repository is populated automatically from text-meaning representations. Knowledge acquisition is largely driven by lacunae found during the processing of actual texts; it is expedited using OntoSem's DEKADE environment (see McShane et al. 2005). We are currently working on developing a "push me pull you" knowledge acquisition strategy that incorporates machine learning (ML) of lexicon and ontology into our knowledge-rich environment: the more knowledge we learn with the help of ML, the more resources we will have to support the learning of still more knowledge. We do not consider the "knowledge bottleneck" to be anywhere near the impasse that many make it out to be: acquiring knowledge simply requires effort, no different from or more extensive than the effort currently being exerted in creating annotated corpora.

A high-level view of OntoSem text processing is illustrated in Figure 9.1.

TMRs represent propositions connected by causal, temporal, rhetorical and other relations (see Nirenburg and Raskin 2004, Chapter 6 for details). Propositions are headed by instances of ontological concepts, parameterized for modality, aspect, proposition time, overall TMR time, and style. Each proposition is related to other instantiated concepts using ontologically defined relations (which include case roles and many others) and attributes. Coreference links form an additional layer of linking between instantiated concepts.

Now we are ready to discuss the use of non-ontological world knowledge in text analysis. It might seem trivial to state that world knowledge (the extratextual
context) is needed for text meaning analysis. But while most people acknowledge this need, few actually propose operational models for doing so. The fact repository in OntoSem is a core component of such an operationalization. A module such as the OntoSem fact repository has not prominently featured in earlier semantic interpretation approaches, including Wilks’ (e.g., 1975, 1977) preference semantics. I will illustrate its use in text understanding by showing how it helps to treat reference resolution, a notoriously difficult NLP problem.

Within OntoSem, processing reference is understood as detecting all referring expressions in a text or a corpus and associating them with their anchors in the fact repository (FR), which is a collection of interlinked real-world instances of objects and events extracted from text after it has been interpreted by the OntoSem analyzer. The information in the FR both supports the processing of any given text (it is a substrate of computer-tractable knowledge) and is supplemented by information from that text. Under this conception of full resolution of reference, the text string Colin Powell is not resolved until it is linked to its anchor in the FR, if there is one, or until a new FR anchor is instantiated, if none yet exists. Thus, the OntoSem engine must try to link every pronoun, relative date (last week), relative time (later), definite description (that man), etc., not only to other co-referential elements in the given text, but to the actual anchor in the ever-growing world model. This is reference beyond co-reference.
The fact repository contains a list of "remembered instances" of ontological concepts. In other words, whereas the ontology has the concept CITY, the FR contain entries for London, Paris and Rome that are instances of the ontological concept CITY; whereas the ontology has the concept SPORTS-EVENT, the FR has an entry for the Salt Lake City Olympics. Facts from TMRs for the input texts are converted (possibly, after some filtering) into persistent objects in FR. FR, thus, becomes the assertion component of the overall knowledge base, just as the ontology forms its description component and the lexicon connects elements of both ontology and the fact repository with their realizations in a language.

Below is a pretty-printed version of the automatically generated TMR for two short sentences that illustrate a difficult case of reference resolution (most sentences we process are much longer):

Colin Powell met with Jack Straw. The American official asked for support.

Ontological concepts are written in small caps. For orientation, all references to Colin Powell are in boldface, while all references to Jack Straw are in italics. Colin Powell met with Jack Straw.

**HUMAN-79**

AGENT-OF MEETING-80
HAS-NAMESPACE [FIRST "Colin"] [LAST "Powell"]
FR-REFERENCE HUMAN-FR24
ROOT-WORDS "Colin Powell"

**MEETING-80**

TIME < FIND-ANCHOR-TIME
AGENT HUMAN-81
AGENT HUMAN-79
ROOT-WORDS "meet with"

**HUMAN-81**

HAS-NAMESPACE [FIRST "Jack"] [LAST "Straw"]
AGENT-OF MEETING-80
FR-REFERENCE HUMAN-FR40
ROOT-WORDS "Jack Straw"

The American official asked for support.

**NATION-104**

HAS-NAMESPACE "United States of America"
ROOT-WORDS "American"
FR-REFERENCE NATION-FR213

**SOCIAL-ROLE-105**

AGENT-OF PROPOSE-107
RELATION NATION-104
AUTHORITY-ATTRIBUTE 0.7
ROOT-WORDS "official"
These TMRs are read as follows. In the first sentence, the input "Colin Powell" instantiates the concept HUMAN, which is appended with the number 79 since this is the 79th time a HUMAN was instantiated during this run of the analyzer. When the system checked the FR for Colin Powell, it found a suitable match (what we call an anchor), which is called HUMAN-FR24 (the 24th HUMAN stored permanently in the FR). Below is an excerpt:

HUMAN-FR24
HAS-FIRST-NAME "Colin"
HAS-LAST-NAME "Powell"
HAS-MIDDLE-INITIAL "L"
HAS-MIDDLE-NAME "Luther"
SOCIAL-ROLE SECRETARY-OF-STATE
CITIZENSHIP NATION-FR213
COREFERENCE CABINET-MEMBER-FR2, MILITARY-OFFICER-FR1,
          MILITARY-OFFICER-FR3, etc.
HAS-CITY-OF-BIRTH CITY-FR1465
HAS-NATION-OF-BIRTH NATION-FR213
GENDER male
MARITAL-STATUS married
HAS-DATE-OF-BIRTH absolute-time (year 1937)
HAS-SPOUSE HUMAN-FR2134
HAS-CHILDREN HUMAN-FR2323, HUMAN-FR2324, HUMAN-FR2325
AGENT-OF ATTEND-ACADEMIC-INSTITUTION-FR3, EARN-DEGREE-FR356,
          ATTEND-ACADEMIC-INSTITUTION-FR4, EARN-DEGREE-FR400,
          SPEECH-ACT-FR151, SPEECH-ACT-FR152,
          REQUEST-ACTION-FR23, etc.

Thus, a given object or event has a fleeting number associated with it for the given text and a static FR number associated with it.

The concept MEETING is instantiated for the lexical string "meet with" because the syntax of the input matched the following verbal sense of meet in the lexicon:

meet-v3
def "to meet with s.o."
ex  “She met with her boss about the upcoming deadline.”

syn-struc subject root $var1 cat n
root $var0 cat v
pp root $var2 cat prep root with (obj root $var3 cat n)

sem-struc MEETING
AGENT value $var1
AGENT value $var3
$var2 null-sem +; shows the the meaning of the preposition;
has already been accounted for in the sem-struc

The time in the TMR for sentence 1 is specified as “< find-anchor-time”, which
means “before the anchor time of the given text”. OntoSem attempts to determine
the anchor time using procedural semantic routines that rely on the dateline of
the given article and other such heuristics. The AGENTs of MEETING are Colin
Powell and Jack Straw, respectively, as can be seen by tracing the concept numbers
to their specifications. It is hoped that this brief walk through the first TMR is
sufficient for general orientation.

The most interesting aspect of the sample pair of sentences is the reference
connection between the American official and Colin Powell, which the OntoSem
analyzer can automatically establish using lexical, ontological and world knowledge.
More specifically:

FR information
HUMAN-FR24 ; Colin Powell
SOCIAL-ROLE SECRETARY-OF-STATE
CITIZENSHIP NATION-FR21 ; USA

Lexical information
One of the meanings of “official” is described as SOCIAL-ROLE with AUTHORITY-
ATTRIBUTE = .7 (that is, a social role with a great deal of authority).

The adjectival form of a NATION (e.g., American) followed by a SOCIAL-
ROLE (e.g., official) means that the SOCIAL-ROLE has CITIZENSHIP in that
NATION (this is lexically encoded as a sense of American, Canadian, and other
such adjectives).

Ontological information

SECRETARY-OF-STATE is a kind of GOVERNMENTAL-OFFICIAL.

The analyzer exploits this knowledge roughly as follows. It searches for
antecedents that are semantically compatible with the meaning SOCIAL-ROLE with
AUTHORITY-ATTRIBUTE .7 (e.g., it would reject “peon”, which is a SOCIAL-
ROLE with AUTHORITY-ATTRIBUTE 0; it would reject any ABSTRACT-
OBJECTs; etc.). Once it has a list of potential antecedents, it attempts to narrow
it down to exactly one. One way of doing this – as in our example – is to match features about an entity noted in the text to features about it stored in the FR (the same way as humans do, by the way). Here, the information we have is that the official is American, which we know refers to citizenship based on the selected sense of American in our lexicon. So the analyzer checks the citizenship of each candidate entity in the FR and finds that Colin Powell matches whereas Jack Straw does not. This translates into a high confidence preference for Colin Powell as the antecedent for the American official. If we did not have the necessary disambiguating evidence in the FR, we would use a combination of other heuristics to favor one or another coreference link. Many of our heuristics are those used widely by the knowledge-lean approaches (see, e.g., Mitkov 1998), but we add semantic heuristics available only in our knowledge-rich environment.

Although it is not immediately evident from this one example, our approach to using FR information for reference resolution is actually generalized – that is, it extends beyond cases of citizenship and social roles. However, since we are less than a year into this particular type of work, we are still discovering new types of cases and ways of exploiting the FR information to its best advantage.

As our example above has illustrated, difficult cases of reference resolution require extra-textual knowledge: in fact, the need for extra-textual knowledge is precisely what makes certain cases of reference difficult for computers (the same reference resolution is not difficult for people because they bring all the necessary knowledge to the table; or, if they don’t, they seek it: “Is Jack Straw or Colin Powell an American?”). Therefore, even if one attempts only to establish textual co-reference links, the knowledge to do that will often derive from repositories of world knowledge about the given entity. Our previous work in reference resolution has only strengthened our belief that, when it comes to reference, there can be no hard line drawn between the text (including the knowledge therein) and the world knowledge brought to bear when interpreting it.

9.4 Being Happy With What You Have

I share Wilks’ concern about covering the actual language and world phenomena even at the expense of not conforming to the discipline desired by Guarino. I agree with Wilks that such discipline is both impossible to impose on realistic knowledge resources and may be, in the final analysis, immaterial. I do not share his pessimism about the attainability of broad and deep coverage. This pessimism and the understandable reluctance to undertake the complex task of creating resources with a depth of description sufficient for AI/NLP needs leads Wilks to two parallel conclusions: (a) acquisition of such knowledge resources is possible only if done automatically and (b) acceptance of available resources, such as MRDs and WordNet, stressing their utility over their obvious shortcomings (see comments on WordNet below). As MRDs and WordNet use natural language as their metalanguage (either entirely or, in the case of WordNet, predominantly), the theoretical position of allowing representation languages to be similar to natural languages gets another, teleological, boost in addition to its motivation by the linguistic turn.
Wilks stresses that ontological and lexical resources (e.g., WordNet or other MRDs) can themselves be objects of research. What is meant here is that useful information can be mined from such sources to support various NLP applications. This is indeed manifestly true for statistics-based processing and possibly in support of non-semantic analysis of text. But it is not so evident when such knowledge is supposed to serve AI/NLP applications. The survey of MRD research by Ide and Véronis (1993) covered the period before statistics-based NLP really took off. It described work devoted to creating machine-tractable lexicons from MRDs, not providing features to support classification. The survey had a suggestive title “Extracting knowledge bases from machine-readable dictionaries: Have we wasted our time?” and concluded: “The previous ten or fifteen years of work in the field has produced little more than a handful of limited and imperfect taxonomies.” I think that this statement is basically appropriate today, too: the utility of MRDs in developing computational-semantic lexicons has been truly limited, in spite of successes declared in the contributions in Guo (1995) or the upbeat picture in Wilks et al. (1996). It might be that this inability to seriously bootstrap the development of a computational-semantic lexicon from MRDs that led Wilks to take the position that “imperfect resources are better than no resources.” This position is reasonable enough. However, the level of imperfection in various resources is highly variable. And some of them, for example WordNet, fall seriously short of being able to support semantic interpretation. Here is a brief sampling of reasons why.

It has become common practice to consider as self-evident that the extensive citing of WordNet in the NLP literature is proof of its utility. That conclusion is, actually, unfounded: people are certainly trying their best to find good uses for it since it is available, but that does not imply that their attempts have shown great promise or that success will improve with better machine learning techniques. A common result of machine-learning efforts with and without WordNet is a small increase in results using WordNet and no indication of where the given work can proceed. Take as examples two experiments from the realm of word sense disambiguation (WSD): Stetina et al. (1998) achieve 75.2% accuracy by choosing the first lexical word sense in a dictionary, and 80.3% using WordNet, and Mihalcea and Moldovan (1998) reach 58% precision in WSD using semantic density in WordNet. However, here the experiments stop: the ML methods have been used, they do the best they can with the available resources but are still far from 100% or from human performance on natural tasks, such as understanding and disambiguation in vivo (as opposed to tasks that are unnatural for people, such as fitting a word sense into a Procrustean bed of a predefined set of senses). These relatively low ceilings of results are only to be expected if the difficult problems of NLP are approached using resources that do not target the difficult problems, and using procedures that – because they do not use sufficient amounts of deep knowledge – have to be satisfied with results that may be state-of-the-art but are unimpressive in absolute terms.

Naturally, the argument from the other side is that the field – not to mention society – needs results right away, and there is no time to build large knowledge resources. Our response is that time will be spent either way, and if time is spent
on developing the resources that the community really needs for higher-end applications, in the long run it will be well worth the effort. I believe that the amount of annotation work required to allow the statistical methods to tackle higher-end applications exceeds the amount of work needed to create broad and deep knowledge resources for knowledge-based NLP. I firmly believe that the best way to use statistical methods is to apply them to the task of aiding the knowledge acquisition efforts of knowledge-based NLP. When the results of statistical analysis of corpora are validated and used by humans, the overall efficiency of knowledge acquisition is significantly enhanced.\footnote{In the OntoSem environment, the use of statistical methods in knowledge acquisition is a central direction of work at the time of writing. Incidentally, OntoSem also uses statistical techniques in the processing itself, as a source of heuristic decisions for the cases when other heuristics are weak or non-existent.}

No available resources that claim to provide semantic support for NLP have proved directly applicable to the OntoSem analyzer, though some have been indirectly useful: e.g., WordNet is among the many on-line and paper sources of synonyms that OntoSem acquirers can use during manual acquisition. A comparison between the representation of verbs expressing change in WordNet and OntoSem will serve as an illustration of the difference in semantic richness of these two resources.

In describing the presentation of verbs of change in WordNet Fellbaum (1999: 252) writes: "...Verb phrases like change magnitude, change shape, and change surface were entered [as nodes in WordNet] on the basis of purely semantic considerations. These concepts were needed to distinguish three groups of verbs that were otherwise all daughters of one node containing the verb change. To have represented verbs like increase, dwindle, and wax as sisters of verbs like flatten, bend and twist as well as of verbs like buckle, fold, and smoothen just did not seem felicitous and seemed to result in a semantically non-homogenous class."

OntoSem takes the semantic specification of verbs denoting change a large step further, representing these notions beyond iconic listing in a hierarchy. All verbs of change in OntoSem are lexically mapped to the ontological concept CHANGE-EVENT but their respective lexicon entries specify their meaning in terms of preconditions and effects. Take, for example, the verb increase, whose meaning depends on the theme of the increase. E.g., if the THEME of the increase is mapped to a SCALAR-ATTRIBUTE – like price (mapped to COST) or height (mapped to HEIGHT) – then the PRECONDITION has a lower value on the given abstract scale (0–1) than the EFFECT does. A call to a meaning procedure that incorporates the correct scalar into the representation of the change event is listed in the lexical entry for all change events. So, a TMR for the price increased (in presentation format) will be:

\begin{verbatim}
CHANGE-EVENT
THEME COST
PRECONDITION.COST.VALUE < EFFECT.COST.VALUE
TIME < SPEECH-ACT.TIME
\end{verbatim}
To return to the broader picture, I will mention just four of the many coverage-related and organizational limitations of WordNet. First, it effectively (though apparently not by design) follows the methodology of splitting word senses, with all the concomitant problems that this poses for NLP. Of course, as WordNet was developed specifically for human use, this methodological choice cannot be held against it. Second, WordNet does not handle the semantics of adjectives well, as reported in Fellbaum (1999); compare this with OntoSem’s fundamental treatment of even the most polysemous of adjectives, as described in Raskin and Nirenburg 1999. Third, different diatheses of a given verb are presented in different parts of the hierarchy: e.g., active sell has a superordinate of exchange while middle sell has a superordinate of be – which Fellbaum describes as a result of the design of WordNet (Fellbaum 1999: 256–257). She further notes that “Researchers who have tried [to] find the semantic properties that are both necessary and sufficient to characterize the class of verbs that can undergo middle formation have not been completely successful…” (259) The search for such semantic overlap is, in our opinion, an invented problem: there need not be any such properties, and an environment for representing semantics should best start from the needs presented by the language rather than the restrictions of a given formalism. Fourth, complex expressions and complex notions (even if expressed succinctly) cannot be integrated, as reported by Fellbaum (1998) (who focuses on idioms, but the same issues arise with semantically compositional expressions). Among the types of excluded entities are: (a) idioms that do not fit into any of WordNet’s categories N(P), V(P), Adj(P) or Adverbial(P): e.g., the more the merrier; (b) structures that require negation like not give a hoot; (c) full sentences; (d) idioms that contain variables, like blow one’s stack; (e) idioms that express concepts that can’t be paraphrased by a single notion, like drown one’s sorrows; (f) idioms meaning become smith… as in hit the roof. OntoSem, for comparison, permits all of these types of entities, with their corresponding semantic representations, to be expressed in lexical entries that can include variables, optional elements, and expressions of any length or complexity. In short, OntoSem imposes no limits on the granularity of semantic (not to mention syntactic) expressiveness: semantics can be expressed by any combination of ontological mappings, preconditions and effects, property values, values of mood or aspect, etc.; and if a means of representation does not exist, we create it to fill a practical need.

But the main issue with WordNet is its weakness in supporting ambiguity resolution. Ambiguity is, in our view, the main challenge for NLP. It is, therefore, reasonable to say that if a lexicon and an ontology used for NLP do not support disambiguation, they cannot be sufficient for truly high-level applications. WordNet’s inability to support ambiguity resolution is understandable because ambiguity poses virtually no problem for humans, and WordNet seeks to depict how humans organize lexical knowledge. In other words, if WordNet accurately depicts how humans organize lexical knowledge, then use of the resource should presuppose all of the world knowledge, pragmatics, goals and general analytical skills possessed by humans. Machines, however, do not have these advantages.
A relevant comparison is the utility of a thesaurus to a native speaker versus its relative opaqueness to a language learner.

WordNet is used by many as a source of knowledge in NLP simply because it is there. Its actual efficacy varies among applications: e.g., Vieira and Poesio (2000) found it of little help in reference resolution, and its utility in query expansion for information retrieval has been mixed (see below). The widespread use of WordNet for NLP has spurred efforts to make it a better NLP resource, with version 2.0 including more noun-verb links and a topical organization for certain domains. However, the nature of this resource as a hierarchy of semantically undefined lexical items remains, we believe, an insurmountable disadvantage for machine processing.

In sum, the fact that WordNet combines some ontological knowledge with lexical knowledge does not, by itself, disqualify it from use in applications. What limits its utility is the lack of knowledge (e.g., about selectional constraints) that is required to support automatic extraction of meaning from text.

9.5 Ambiguous Symbols ↔ Unambiguous Expressions

If one did not know about Wilks’ track record in his “day job” as a practicing NLPer (e.g., Wilks 1975 on compositional procedural semantics using preferences or Wilks and Stevenson 1997 on word sense disambiguation), from the papers under discussion one could get an impression that he does not consider the issue of ambiguity resolution central to the enterprise of AI NLP. His position is that atoms (symbols) in the representation language can be (and cannot be but) ambiguous while representation language expressions can be made to be unambiguous and that this resolution of ambiguity “would have to be resolved by the processor that used them” (Nirenburg and Wilks 2001). If so, the processor must have access to sufficient knowledge to make disambiguation decisions. But simply saying that something will be the responsibility of the processor does not absolve Wilks from the responsibility of describing how this is to be done. This looks suspiciously like passing the buck.

What does the position that in a representation language atoms are (ambiguous) natural language words and expressions are unambiguous, entail? It entails that there is no representation that is declaratively unambiguous and therefore whenever a representation language statement is required to support some processing (either related to text meaning extraction or to application-oriented reasoning), it must first be disambiguated by running some procedure to understand what specific inferences it supports. Note that this disambiguation will have to be carried out every time an element of representation is called upon for reasoning – because in its static form the representation is not disambiguated. Thus, faced with a representation language statement of the kind “car consume gasoline,” any system will first have to understand that in this context car does not mean railroad car and consume does not mean intake of foodstuffs through the mouth by higher animals (supposing that gasoline is not ambiguous). It is only after such disambiguation that the inference maker will be able to use this knowledge (together with other knowledge elements,
certainly) to establish (among a number of other entailments) that when somebody says "It's a long drive, we must buy gasoline," the reason for buying gasoline is to make sure that the car has enough gasoline to consume in order to complete this drive. Note that if such a finding is recorded, once again, using ambiguous words as representation atoms, then every time these inference results are accessed, another quantum of disambiguation will have to be run. Even without going any deeper, this is surely a less-than-efficient proposition.

Another practical consideration at this point would involve supplying outputs that are amenable for processing using the currently available reasoning systems. These systems typically require absence of ambiguity in representations as well as completeness and soundness of the knowledge before any discussion of the possible utility of the reasoning systems themselves. One example of such a procedure was the AQUA project within the AQUAINT R&D program devoted to question answering. AQUA used the JTP reasoning system (Fikes et al. 2003). The knowledge over which this system reasoned in this project was automatically produced by the OntoSem semantic analyzer (Beale et al. 2003), whose results were then automatically converted into the representation language used by JTP. Of course, the conversion between the two representation language formats was largely semantics-free. Another example is the SemNews system (Java et al. 2005), which uses OntoSem to annotate RSS news feeds with TMRs and make them available on the internet. The format conversion here was between the custom metalanguage of OntoSem and a dialect of OWL, the representation language of the Semantic Web.

9.6 Ambiguity in Representations and its Causes

"[T]he key distinction between Wordnet and an ontology is this: Wordnet has lexical items in different senses (i.e. that multiple appearance in Wordnet in fact defines the notion of different senses) which is the clear mark of a thesaurus. An ontology, by contrast, is normally associated with the claim that its symbols are not words but interlingual or language-free concept names with unique interpretation within the ontology. However, and given that this issue is one of great antiquity, the position argued here is that, outside the most abstract domains, there is no effective mechanism for ensuring, or even knowing it to be the case, that the terms in an ontology are meaning unique." (Wilks, Ontotherapy).

This statement combines the issue of natural language as metalanguage with the issue of whether ontologies are different from thesauri. Wilks motivates the opinion using the example of Lisp, a programming language, instead of a natural language: "For example, it is generally agreed that in the basic original forms of the Lisp programming language the symbol ‘NIL’ meant at least false and the empty list, though this ambiguity was not thought fatal by all observers. But it is exactly this possibility that is denied within an ontology (e.g. by Nirenburg) though there is no way, beyond referring to human effort and care, of knowing that it is the case or
not.” (Ontotherapy). The fact that NIL started to be used in Lisp in several different meanings might indeed have happened initially by chance. But this situation was then reviewed by the developers of the language (and apparently the observers mentioned by Wilks), who adjudged this situation benign. It was recognized that these uses were in a complementary distribution and that the interpreter and the compiler would not face conflicts in evaluating expressions with NIL. As a result of this judgment, the ambiguity of NIL was retained for the sake of convenience, tradition and parsimony. It is beyond question that if this ambiguity interfered with or hampered processing, it would have been eliminated.

The situation with ambiguity in knowledge resources used in AI/NLP is more complex. Indeed, there are several kinds of ambiguity that are apparently impossible to eliminate in building ontologies, lexicons and fact repositories capable of supporting realistic applications. This is, again, a case of a conflict between the desire to have “clean” knowledge resources and the need to achieve coverage and depth. In discussions of scientific method, the two most important desiderata concern fidelity to empirical evidence and simplicity and consistency of logical formulation. But fidelity to evidence takes precedence in cases of conflict (e.g., Caws 1967; see also Nirenburg and Raskin 2004: 73–74). In other words, it is a good idea to strive for “clean” knowledge resources, but when the needs of description make this goal impossible or too difficult in building an application (though not necessarily if the project is purely theoretical), it should be waived.

In the rest of this section, I will present a few examples of synonymy and potential ambiguity in OntoSem TMRs, ontology and fact repository.

Practice has shown that the relation between texts and text meaning representations is not one-to-one. It can clearly be one-to-many, as text meaning representations are subject to paraphrasing (that is, exhibit synonymy). One possible cause is the possibility of expressing certain meanings with different levels of specificity (or, alternatively, vagueness). For example, in the OntoSem ontology, the event EMBARGO is a leaf node on the path:

\[
\text{EMBARGO} \rightarrow \text{SANCTION} \rightarrow \text{COMMERCE-EVENT} \rightarrow \text{FINANCIAL-EVENT} \rightarrow \text{SOCIAL-EVENT} \rightarrow \text{EVENT}
\]

Each of these concepts is used to explain the meaning of different words in the lexicon, and these links are primary clues for the construction of TMRs. So, for the input

The US has imposed an arms embargo on Somalia.

the TMR will contain an instance of the event concept EMBARGO. For the input

The US has imposed sanctions on Somalia.

the TMR would contain an instance of SANCTION (note that the concept SANCTION expresses only one of the two senses of the English sanction, the other being explained in terms of the ontological concept APPROVE). SANCTION will also anchor the TMR for
The US has sanctioned Somalia by prohibiting arms imports into the country.

Since sanctioning imports is, in fact, embargoing, the meaning of this past sentence could be paraphrased by a TMR using EMBARGO, a more specific concept than SANCTION. The connection between paraphrases in TMR may even be more remote; for example, the input can be paraphrased as:

The US has prohibited arms imports into Somalia.

The TMR for this sentence will feature the concept IMPORT (which is not a parent of EMBARGO but its ontological first cousin once removed: it is a part of the path IMPORT → COMMERCE-EVENT → ...).

The question is, how to resolve and use these paraphrases. It is a practical question because such resolution would facilitate a correct answer to the question “What sanctions were imposed on Somalia?” when the fact repository contains only the statement that US imposed an arms embargo on Somalia.

One algorithm that can be applied to the resolution of paraphrases is described by Mahesh et al. (1997), where it is used to provide additional support for basic disambiguation algorithms in OntoSem. For example, if the system could not disambiguate SANCTION between SANCTION and APPROVE, due to the corresponding constraints being insufficiently strong, an attempt was made to specialize the TMR by moving down the ontological hierarchy and using ontological descendants of both SANCTION and APPROVE to see whether only one of the candidates conformed to the selectional and other constraints. In the above examples, EMBARGO would work fine, while APPROVE would not.

The algorithm in question has been used not only in cases of ambiguity but also when the basic analysis fails to find a single coherent interpretation because no solution conforms to all the required constraints. In such cases, the algorithm essentially relaxes constraints by moving up the ontological hierarchy. Thus, it might involve using COMMERCE-EVENT instead of either IMPORT or EMBARGO. If a successful TMR “anchor” is found using this method, then the resulting TMR will be effectively ambiguous between several candidate meanings. This state of affairs is accepted because OntoSem prefers to output vague (but not necessarily inconsistent) solutions rather than producing no solution at all.

Now to a few brief comments about ambiguity in static knowledge resources. In an ontology, ambiguity can result from multiple inheritance, which causes problems for the interpretation of the [admittedly, clumsy] sentence “Don’t hit me with the newspaper you work for” if the lexicon has just one sense of newspaper. A particular fact repository could (and almost certainly will!) have multiple entities that the system failed to co-refer. It is entirely plausible and to be expected that Walter Scott could be stored there as HUMAN-FR334 and “the author of Waverley” could be stored as HUMAN-FR5298. Moreover, some knowledge about Walter Scott may be also stored under AUTHOR-FR94, with some of the information about him appearing
in the human-FR334 frame and some other, in author-FR94. This is another illustration of why establishing co-reference relations among FR entities is a crucially important task.

9.7 Procedures in Representations

Procedural attachments appear in a broad variety of declarative knowledge representation schemata and are traditional enough to be described in introductory textbooks (e.g., Brachman and Levesque 2004). In AI/NLP, one of the reasons to use them is to compensate for descriptive difficulties that result from the desire to help text analysis by allowing an analyzer to choose from among a small set of senses for each word. To make the sets of senses small, it is often necessary to “bunch” them, that is to group somewhat distinct senses into coarser entities. Difficulties in processing arise here because, while it is easier to select one of three rather than one of sixteen during processing, the actual interpretation of that sense can become unfocused. Pinpointing the actual meaning is then delegated to procedures that take into account additional (usually from the immediate context) information that cannot be readily stipulated in static knowledge resources.

In this section I discuss connections between sense discrimination, lexicon content and the use of procedures to enhance the static knowledge resources.

A major point of disagreement between Fodor and Lepore on the one hand and Wilks and Pustejovsky on the other hand is their attitude to procedures as components of lexical semantics. Fodor and Lepore do not believe in them. Pustejovsky uses them essentially because he thinks that it could be done, with all the theoretical and practical benefits accruing. Wilks believes that lexical specification cannot be carried out without procedural knowledge, in part because he does not believe that a purely declarative metalanguage could be devised. Moving from the language side to the metalinguage side, Guarino criticizes WordNet and OntoSem ontologies because, among other issues, they do not provide identity criteria for all concepts. Wilks argues that this requirement (among other restrictions) is only tangentially useful for reasoning in NLP; moreover, he doubts the possibility of constructing a realistic and useful ontology conforming to Guarino’s discipline. The decision to include procedures in the representational apparatus can render some of Guarino’s criticism moot because there may remain fewer word senses (and concepts in an ontology) and, therefore, fewer cases of unspecified identity conditions. But to compensate for that, Guarino’s task will be made more complicated because now “cleaning” the knowledge resources would involve entities (namely, procedures) of a different kind from declarative concepts.

The fact that the English words window, book or newspaper (“Don’t hit me with the newspaper you work for!”) exhibit regular polysemy does not necessarily require that their meanings be represented by a single ontological concept, a fact that attracted Guarino’s critical attention. Also, in general, lexical elements (atoms) of the ontological metalanguage (ontological concepts) do not necessarily correspond to word senses in natural language. Moreover, as mentioned above, many word
senses are interpreted not by pointing to an ontological concept but through the use in the text meaning representation language of parametric, extra-ontological features, such as speaker attitudes (called modalities in OntoSem), co-reference relations, rhetorical relations or by procedural means – for example, by calling a meaning procedure attached to a lexicon entry to determine the contextual interpretation of a word sense.

With respect to delimiting the number of word senses to be encoded in the computational-semantic lexicon, there are three options: bunching, splitting and passing the buck. Bunching would imply positing one sense for window (to be interpreted as both an aperture and as a frame), expressed using one ontological concept with multiple inheritance; this is a widely recommended strategy (e.g., Palmer et al. forthcoming; Ide and Wilks forthcoming; Nirenburg and Raskin 2004: 331–344) for use in computational linguistic applications. Splitting involves trying to capture as many lexical-semantic distinctions as possible and therefore positing as many different senses for a word as such distinctions would require. This is the approach taken by many lexicographers in developing human-oriented dictionaries. Finely split senses, if used in a semantic analyzer, lead to the necessity of describing each of them in sufficient detail for an analyzer to be able automatically to select the appropriate one. This approach leads to obvious logistical difficulties. It also is exposed to the same criticism to which Weinreich (1966) subjected the semantic theory of Katz and Fodor (1963), namely, that there is no good criterion to limit polysemy and thus stop splitting senses. Nirenburg and Raskin 2004: 331–332, comment: “...[H]aving determined that one of the senses of eat is ‘ingest by mouth’, should we subdivide this sense of eat into eating with a spoon and eating with a fork, which are rather different operations? Existing human-oriented dictionaries...do not have theoretically sound criteria for limiting polysemy of the sort Weinreich talked about. It might be simply not possible to formulate such criteria at any but the coarsest level of accuracy.” Finally, “passing the buck” means, in the case of window, positing one sense but not committing to a particular ontological concept or concepts as explanation in the representation itself; instead, encoding the decision as a procedure that decides on the appropriate meaning when called during actual text analysis (that is, when textual context is present). The assumed existence of such procedures is a major prerequisite of the generative lexicon approach of Pustejovsky (1995). The main claim of this approach is twofold: (a) word senses of a lexeme are systematically related by general relations from a specified inventory; and (b) once this inventory is established, it becomes possible to split a lexeme into fewer senses because all the other senses will be derivable using lexical rules based on these senses. Nirenburg and Raskin (2004: 115–133) examine Pustejovsky’s claims in some detail and conclude, among other things, that in many cases there is no effective difference between the amount of work necessary to compile a generative lexicon and that required for a more traditional enumerative lexicon. Indeed, while some lexical rules – for example, rules capturing phenomena from derivational morphology – have wide applicability and have been shown (e.g., Viegas et al. 1996) to be useful in practice, the utility of many other rules, in particular, those reflecting regular polysemy – most famously, the “grinding” rule (Briscoe and
Copestake 1991) – are much less obvious because of the number of “exceptions” that must be individually encoded and for which the application of the rule must be explicitly blocked. In these cases, the purported advantage over enumerative lexicons does not materialize.

Another important class of phenomena – in OntoSem referred to as semantic ellipsis (e.g., McShane 2005), cf. I forgot his name vs. I forgot (to bring) the umbrella – is equally difficult for both enumerative and generative lexicons. Fodor and Lepore contend that a rule like:

X wants Y → X wants to do with Y whatever is normally done with Y

cannot be formulated because its application to particular lexemes, e.g., beer in John wants some beer cannot be safely recorded anywhere but with the appropriate sense of beer. They advocate a totally individual and separate element of the language of thought for each of the word senses. But this criticism applies equally to the case when drink is marked as the typical purpose of beer in the lexicon and to the case when determining this constraint is delegated to an inference rule operating over the ontology and the lexicon.

My criticism is from the opposite side – it seems that one can never guarantee that sufficient disambiguating constraints can be obtained from the content of the lexicon entry (no matter whether the lexicon is enumerative or generative) because one can normally do things with beer that are not normally done with, say, more generic drink or liquid – for example, John might want to make Carbonnades a la Flamande. So, the disambiguation process must take into account not only the knowledge in the ontology and the lexicon but also the results of processing the broader context. This realization is ultimately a vote for the use of scripts, typical event sequences, in automatic text understanding. While a number of scripts have been acquired for the OntoSem ontology, the semantic ellipsis resolution algorithms described in McShane et al. 2004 do not yet use script-based knowledge in the left hand sides of their inference rules. Incidentally, in OntoSem the approach to resolving semantic ellipsis is mixed – some cases invoke the use of meaning procedures (see below) while others are treated through lexicalization.

To illustrate this lexicalization method with just one example from the above paper, consider the verb invite. When followed by a direct object indicating a human (or, by extension, an organization) and a prepositional phrase or adverb indicating location (or destination) directly or metonymically, it actually means “invite someone to come/go to that place”; the verb of motion is semantically elided. Examples include the following:

- Civilians invited into the prison by the administration to help keep the peace were unable to stanch the bloodshed.
- “If they invited us back tomorrow to govern the mainland, frankly we would hesitate,” Vice Foreign Minister John H. Chang told a US governor’s delegation.
- All 13 OPEC oil ministers were invited to the meeting.
• He often is one of a handful of top aides invited into the Oval Office for the informal sessions at which President Bush likes to make sensitive foreign-policy decisions.

The OntoSem lexicon entry that covers this use of invite is as follows, in presentation format:

```
invite-v2
def "+direct object (human) + pp of destination, implies 'invite to come'"
ex "She invited him to the meeting/to Paris"
syn-struc
  subject  root $var1  cat n
  v
  directobject  root $var2  cat n
  pp-adjunct  root $var3  cat prep
  root (or to onto into on)
  obj root $var4  cat n
sem-struc
  INVITE
  AGENT value ^$var1
  THEME COME
  DESTINATION value ^$var4
  AGENT value ^$var2
  ^$var3 null-sem+
```

The syntactic structure (syn-struc) says that this sense of invite requires a subject, direct object and PP, and that the PP must be headed by the word to, onto, into or on. The semantic structure (sem-struc) is headed by an INVITE event, whose AGENT is the subject of the clause (note the linked variables) and whose theme is COME. The AGENT and DESTINATION of COME are the meanings of the direct object and prepositional object, respectively, of the input clause. (We gloss over formal aspects of the entry that are tangential to the current discussion.) Note that there is no verb of motion in the input text: COME is lexically specified since it is a predictable semantically elided aspect of meaning in the given configuration. For a description of procedural treatment of ellipsis in OntoSem, see McShane (2005).

In summary, inclusion of procedural knowledge in semantic analysis and in computational lexicography may not lead to economies in knowledge acquisition. In this, I am in agreement with Fodor and Lepore. A choice of granularity of sense distinctions in the lexicon, however, does not influence the decision about the nature of representational symbols in any way. On the other hand, the position that the representation language has, unlike Fodor’s, some explanatory power in terms of a core of undefined properties, does not necessarily mean that its elements belong to a natural language.

In the OntoSem ontology, objects and events are given denotation in terms of unique sets of property-value pairs that describe them. Thus, the question of the nature of symbols relates only to the meta-meta-language of semantic description,
that is, to the inventory of ontological properties. In OntoSem currently only about 350 concepts out of the total of over 8,500 concepts (with, on average, 16 properties specified for each) belong to meta-meta-language. The properties are weakly and circularly defined in terms of constraints on their domains and ranges. For example, the domain of the property \textit{mass} is constrained to the ontological subtree with the root \textit{physical-object} and the range of \textit{mass} is constrained to a positive real number or interval. Based only on the formal comparison of the constraints on their domains and ranges, the property \textit{mass} is indistinguishable from the property \textit{length} (and a few other properties). These properties are, however, distinguished by the inventories of inference rules (including meaning procedures within OntoSem proper and external inferences in any reasoning system that uses OntoSem knowledge resources) in which they occur. But it is still incorrect to contend that \textit{length} or \textit{range} as elements of ontology are also senses of natural language words. Within the semantic analysis and reasoning environment, they are simply labels, names of variables used to formulate various constraints and inference rules. The system does not have an understanding of what their meaning in natural language may be. It is the knowledge acquirer who must take into account the intended meanings of these properties. So, it is only to people that these primitives can be considered elements of natural language, not from the point of view of the system.

Incidentally, the preference for using meaningful (for people) names for variables in computer programs does not impinge on the success or failure of the program, though their denotational semantics may be formulated, informally, using the appropriate senses of words. This is the use of denotational semantics intended by McDermodt (1978): “The semantics is for our use, as a tool for analyzing knowledge representations,” and is not intended to be an integral part of the representational or reasoning system itself. So, a compromise on the issue of the nature of symbolic representations may be reached around the statement: “The defining vocabulary of the metalanguage of symbolic representations, its ‘meta-meta-language,’ can receive its denotational semantics in terms of word senses in a natural language.” Of course, some of these word senses will be quite different from what we usually think of as word senses. Indeed, in many cases the knowledge acquirer will have actually to learn from the acquirer-oriented definitions (which are similar in form, content and intent to definitions in regular human-oriented dictionaries) the meaning of properties in the ontology. Thus, for example, the definition for the OntoSem property \textit{measures-property} is given the following definition in the OntoSem ontology: “this property is used to specify what property is measured when one measures something: e.g., if one measures a table, it might be its length, width, height; with verbs other than ‘measure’, there can be even more ambiguity: ‘evaluate the race conditions’ can mean measuring the air temperature, slipperiness of the surface, etc. Thus, \textit{measures-property} is generally used in conjunction with an indication of the thing measured: e.g., measure theme: table, property-measured: height.”

The meaning of text elements and combinations thereof depends to varying extents on the meanings of the elements of the context in which they appear.
Whereas outside of context many nouns (e.g., sheep) and verbs (e.g., dance) conjure a relatively stable image, the adverbials approximately, very and nearly depend crucially on their context to concretize their meaning. In the Ontological Semantic (OntoSem) text processing environment, context-dependent meaning is arrived at by a combination of static lexical descriptions (which link to a language-independent ontology) and procedures that attempt to specify text meaning based on ontological, contextual and other information. In OntoSem, the procedural elements are encoded in meaning procedures (MPs). Meaning procedures are used to account for reference resolution, final assignment of case roles, resolution of absolute time and location, specification of approximations, etc.

As an example, the meaning procedure DELIMIT-SCALE is used in the lexicon entries for words such as very, quite, somewhat, etc. The meaning of a word like very modifies the value of the property that is the meaning of the word that very modifies. For example, in very big, the property will be SIZE, a scalar attribute in OntoSem nomenclature. The meaning of big in OntoSem is represented as a value (an interval) on the abstract scale \{0, 1\} that is defined as the legal filler of the range slot of the scalar attribute SIZE. The meaning of very is procedural: it is a command to make the scale interval representing big narrower and to move it toward the nearest extreme of the scale (the actual numerical calculation will not be detailed here). If, suppose, the meaning of big is the interval \{.8, 1\}, then the meaning of very big may be \{.8, 1\}.

More formally, here is (in a simplified presentation format) the OntoSem lexicon entry for very:

```
very-l

cat adv

def "toward the more extreme end of the given scale"

ex "very big, very late, very small"

syn-struc

mods $var0 (cat adv)

root $var1 (cat (or adj adv))

meaning-procedure

delimit-scale (value ^$var1) extreme .1
```

The syn-struc (syntactic structure) says that very ($var0$) is an adverb that modifies an adjective or an adverb ($var1$). Unlike typical lexicon entries, this one has no static sem-struc (semantic structure) zone, since the meaning of very relies on its composition with what it modifies. Instead, it has a meaning-procedure zone that calls the delimit-scale MP with three arguments:

- the value of the meaning of $var1$ (indicated by ^$var1$), which is a value between 0 and 1;
- whether the scalar value is shifted toward the extreme or the mean of the given scale;
- the amount by which the value is augmented.
So, very small would be calculated by taking the value of small (size \{0, .3\}), and shifting it to the extreme by .1, returning a value of (size \{0, .2\}). Analogously, moderately small would be calculated by taking the value of small (size \{0, .3\}) and shifting it toward the mean by .1, returning a value of (size \{.1, .4\}). (The MP for moderately has the 2nd argument as “mean” rather than “extreme”.)

An interesting situation occurs if one modifies a scalar such that its value is off the scale. For example, extremely is defined as shifting the scalar value by .2 toward the extreme, so an extremely extremely expensive car will be calculated as follows:

extremely + extremely + expensive  
.2+.2+.8=1.2

The value 1.2 lies outside of the \{0, 1\} scale; however, this is exactly what we want as a semantic interpretation of extremely extremely: a value that exceeds any expectation for the property given the specific object in context. OntoSem-style meaning procedures seem to be a natural way of encoding the meaning of context-dependent elements that are too often excluded from the purview of both formal semantics and work on the semantic annotation of corpora.

9.8 Computational-linguistic Turn

Wilks’ positions on the interaction of language and knowledge – which he has held with remarkable consistency at least since his dissertation work (Wilks 1968) – have been repeatedly expressed in his polemics with Fodor, Guarino, myself and many others. These positions seem to be strongly influenced by Wittgenstein’s (1953) views on the treatment of meaning using the metaphor of “language games,” by ordinary language philosophy (e.g., Ryle 1953; Austin, 1962) and in general by what has been called by Bergmann (1964) – and popularized by Rorty (1967) as – “linguistic turn.” When linguistic turn is applied to philosophical deliberations, this means, in simple terms, that studying meaning in language is essentially studying the world, and that the world can be described only through the use of elements of language.

“To execute the linguistic turn... one would have to subscribe to a substantial metaphilosophical thesis namely, as Rorty (1967) puts it, ‘that philosophical problems are problems which may be solved (or dissolved) either by reforming language or by understanding more about the language we presently use’... Rorty and Bergmann take the view that philosophical problems are problems of language as the ‘least common denominator’ of the metaphilosophical positions of both camps in analytic philosophy, Ideal Language Philosophy and Ordinary Language Philosophy.” – Watzka (2002).

Could it be that Fodor’s position ascends to ideal language philosophy even as Wilks’ reflects a form of ordinary language philosophy views?
One of the motivations for the linguistic turn and the development of ordinary language philosophy was to get away from the obscurative, esoteric and overly technical and terminological language used by earlier philosophers. Another was to distinguish ordinary language from formal or conceptual language. It is not easy, however, to pinpoint precisely what language is ordinary and what language is not ordinary. And it does not help that Ryle (1971), as illustrated by Hacker (1996: 160), distinguishes the “use of ordinary language” from the “ordinary use of language” and “ordinary linguistic usage.”

The distinction between ideal language and ordinary language was formulated essentially before the age of computation. While it is difficult for people to maintain the distinction between the ideal and the ordinary, it seems that for a computer system this distinction is essentially moot because no language is ordinary for a computer program. It cannot rely on the powerful reasoning mechanisms that humans possess (and that allow them, for example, to make sense of difficult philosophical discourse – whether it was conducted in ordinary or ideal language or – as seems more plausible – in some mixture of well-defined and ordinary, that is, ambiguous, language).

Maybe it is time to suggest a modification of the linguistic turn for the age of the computer: a computational-linguistic turn, whereby philosophers will formulate their arguments about the world and language in a metalanguage that will be suitable for a software agent in a society that includes both other software agents and people. Software agents can be programmed to communicate among themselves in the metalanguage and with people in language. Philosophers then would be able to observe transcripts of both kinds of communication and judge whether the messages are meaningful and appropriate using, if they so choose, their favorite version of denotational semantics.

9.9 Stand Up (And Be Counted) Philosophy

Yorick Wilks is as close to being a polymath as anybody in the fields of computational linguistics and AI. He has developed, alone or leading research teams, a number of computer systems in machine translation, cognitive modeling, information retrieval and extraction, dialog and other application areas. He has formulated important AI theories, notably that of preference semantics. He has contributed to the conception and development of a number of NLP-oriented resources, such as machine-tractable dictionaries, and researchers’ toolkits, such as GATE. What makes him truly stand out among AI/NLP practitioners is his remarkable ability to assess developments in the field, their importance and their promise for solving long-standing problems in applications. This is the mark of a philosopher, which is what Yorick is and has been all along, first and foremost. What makes him different from most other philosophers interested in AI, cognitive science, computational linguistics and related disciplines is that he does not shy away from standing up and being counted among the creators, not only assessors and critics of systems and theories. Add to this the high culture of argument and the sense of timing and style
worthy of the best public performers in any field, and it becomes clear that Yorick is in a class by himself among colleagues. The field should be looking forward to many more of his contributions.

Acknowledgement

Many thanks to Marge McShane, Patrick Hanks and an anonymous reviewer for fine-grained and very constructive criticism.

References


McCawley, J. 1981. Everything that Linguists have Always Wanted to know About Logic (but were ashamed to ask). Chicago: University of Chicago Press, and Oxford: Blackwell.


