Operative Strategies in Ontological Semantics

Sergei Nirenburg   **Marjorie McShane**   **Stephen Beale**

Institute for Language Information Technologies
University of Maryland, Baltimore County
{sergei,marge,sbeale}@umbc.edu

Abstract

In this paper, we briefly and informally illustrate, using a few annotated examples, the static and dynamic knowledge resources of ontological semantics. We then present the main motivations and desiderata of our approach and then discuss issues related to making ontological-semantic applications feasible through the judicious stepwise enhancement of static and dynamic knowledge sources while at all times maintaining a working system.

1. Introduction

This paper discusses selected issues in ontological semantics (OS), an implemented computational-semantic theory that deals with the extraction, representation and use of meaning in natural language texts. Unlike practically all other work in computational semantics, OS makes itself responsible for all the necessary components and stages in automatic text meaning analysis: it addresses lexical and compositional meaning as well as pragmatics and discourse issues. Its processing heuristics are derived from syntax, morphology and other "preprocessing," non-semantic analysis stages that are still incorporated in the system, as well as from detailed underlying world models that include specifications not only of basic events, objects and properties but also of complex events, or scripts.

The goal of OS is the extraction, representation and manipulation of meaning in natural language texts with a view toward supporting applications such as MT or question answering. Text meaning is represented in *text meaning representations* (TMRs) that are derived compositionally, primarily from meanings of words and phrases in the text. Word and phrase meaning is encoded in the *ontological-semantic lexicon*. The underlying *ontology* is the main metalanguage of lexical meaning specification.[1] As a result, TMRs largely consist of

instances of ontological concepts. Some of these instances are remembered (as "facts") and stored in the *fact repository*, FR, a knowledge base of remembered ontological instances. Some facts in the fact repository are referred to by proper names in texts–personal names, toponyms, names of organizations, specific artifacts ("the statue of Liberty"), etc. These proper names are stored in the *onomasticon*, the semantic zones of whose entries contain a pointer to a corresponding FR element.[2]

The following example illustrates our ontological-semantic knowledge resources (for a more detailed description see Nirenburg and Raskin 2003, Chapters 6-7). The example is much simpler than the sentences from real texts with which the ontological-semantic analyzer typically works and is used here for pedagogical reasons and to save space. Consider the input sentence *Alex Patrick makes tools*. Morphological and syntactic analysis of this input will yield the following structure (throughout this paper we use a presentation notation that is simplified for readability):

```
root          make
cat           verb
tense         present
subject
    root          'Alex Patrick'
    cat           noun-proper
direct-object
    root          tool
    cat           noun
    number        plural
```

The relevant zones of the ontological-semantic entry for the appropriate (first verbal) sense of *make* are as follows:

```
make-v1
    syn-struc
        root      make
```

```
        cat      verb
        subject
            root     $var1
            cat      noun
        direct-object
            root     $var2
            cat      noun
    sem-struc
        CREATE-ARTIFACT
            AGENT   ^$var1
            THEME   ^$var2
```

In the above entry, the variables are used for linking; the caret means 'the meaning of'; CREATE-ARTIFACT is an ontological concept (ontological concepts are in SMALL-CAPS); AGENT and THEME are among its case roles and are used in this lexicon entry to specify selectional restrictions on this sense of *make*. As it happens, the selectional restrictions listed in the ontological definition of the concept CREATE-ARTIFACT are sufficient for this case of *make*, so that no further specialization or generalization of these restrictions in the lexicon is required.

```
CREATE-ARTIFACT

    …
    AGENT   sem      HUMAN
    THEME   sem      ARTIFACT
    …
```

In general, constraints in lexicon entries can modify and supplant those listed in the ontology; this is a means of controlling the proliferation of ontological concepts in the system. Lexicon entries for the appropriate senses of the other elements of the input sentence are as follows (we assume that the FR does not include an Alex Patrick, so that regular lexicon entries will have to be used):

```
alex-n1
    syn-struc
        root     alex
        cat      noun-proper
    sem-struc
        HUMAN
            FIRST-NAME value    alex
            GENDER     sem      male
```

In the above, the filler of the property GENDER is introduced through the *sem* facet since it is defeasible (the lexicon acquirer judged it plausible that there may be females called Alex).

```
patrick-n2      ;this is the last name sense
    syn-struc
        root     patrick
        cat      noun-proper
    sem-struc
```

```
        HUMAN
            LAST-NAME value    patrick

tool-n1
    syn-struc
        root     tool
        cat      noun
    sem-struc
            TOOL
```

In the entry for *tool-n1*, the semantics is simply a univocal mapping to an ontological concept.

On the basis of the knowledge from the ontology, the lexicon and syntactic analysis, the ontological-semantic analyzer will produce the following semantic dependency that will serve as the basis of the TMR (numbers appended to ontological concept names mark them as instances of these concepts):

```
CREATE-ARTIFACT-73
    AGENT  human-209
                FIRST-NAME     alex
                LAST-NAME      patrick
    THEME  set-46
                element-type   tool
                cardinality    > 1
```

1.1 Disambiguation

A number of important representational and processing issues have been omitted in this presentation, notably, a discussion of disambiguation heuristics. The first-line mechanism of ambiguity resolution in OS is matching selectional restrictions. In our example, selectional restrictions on the theme of the proposition head matched successfully: indeed, tools are artifacts. As to the restrictions on the agent, they have been found to be too weak to resolve the ambiguity completely: both the last-name and the first-name (not shown) sense of Patrick fit the selectional restrictions on the proposition head (indeed, *Alex Patrick* may be also be a double first name). Additional disambiguation means are required in this case.

We have developed two general methods for additional sense disambiguation: dynamic tightening of selectional restrictions (Mahesh et al., 1997) and determining weighted distances among ontological concepts activated in the input (using the Ontosearch procedure, e.g., Onyshkevych, 1997). None of these methods will, incidentally, help in our example, so that additional heuristic procedures will have to be built for this type of ambiguity. Incidentally, such heuristic procedures could include evidence from a wide variety of sources, including text corpora. Supporting semantic analysis in this way should become an important direction of work in corpus-oriented computational linguistics (see further discussion below).

Residual ambiguity is one of several possible initial outcomes of the analysis process. Figure 1 illustrates the entire set of possible outcomes of the analyzer operation. Note that, in principle, the single-candidate outcome may prove to be less than ideal–it might simply reflect errors and omissions in static knowledge resources. Still, at this time deriving a single output serves as the halting condition of the analysis process.
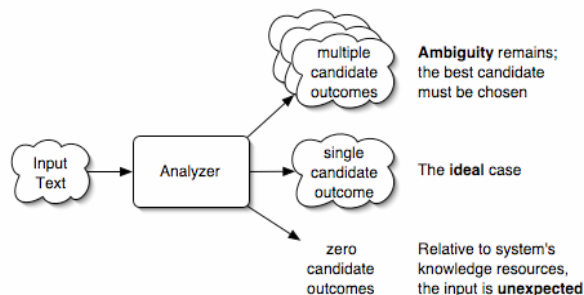

Figure 1. Eventualities in semantic analysis.

If *Alex Patrick* were the name of a company (cf. Merrill Lynch) and were not listed in the FR and there were no textual clues (e.g., *Inc*.) to recognize it as such, the system would fail to produce the correct analysis (that is, we would end up with zero candidate outcomes). Currently, the only way to rectify this state of affairs is to (manually) add to the fact repository an instance of CORPORATION with the name *Alex Patrick* and all its known property values. However, we are working on coercion rules that in the above example would favor the corporation reading if the event in question had the selectional restriction CORPORATION in the appropriate case role.

1.2 Multivalued Selectional Restrictions

Returning to our original example, if *Alex Patrick* is a corporation, then the selectional restrictions on CREATE-ARTIFACT will be violated (a company is not a human). As a result, the corporation sense will always lose to the human sense because in the latter case selectional restrictions do match! We have, naturally, noted that organizations are often used in texts in positions that are "officially" occupied by people; indeed, this is one of the most widespread types of metonymy. We therefore decided to relax the selectional restrictions in such cases by introducing a *relaxable-to* facet for property fillers, in addition to the rigid *value* and abductively overridable *sem* facets. This amounted to the introduction of multivalued selectional restrictions. Matches on fillers of *sem* facets are preferred but matches on fillers of *relaxable-to* facets are not discarded as wrong. If, as in our example, ambiguity results, additional means of its resolution are used.

Multivalued selectional restrictions are also used to treat metonymy: e.g., the THEME of PLAY-MUSICAL-INSTRUMENT is constrained to MUSIC-PIECE but the standard metonymy of composer name will be also noted in the same lexicon entry, together with its expansion to MUSIC-PIECE with the property AUTHORED-BY filled by the metonym.

Processing unexpected input (including lexically unaccounted for metonymies) can be done dynamically using Ontosearch. To treat metonymies, Ontosearch calculates whether the ontological distance between the metonym and the selectional restriction that it must match is below a preset threshold (in which case dynamic relaxation is allowed and the case is declared true metonymy).

1.3 More Expressive Means in OS

The *Alex Patrick* example illustrated some of the issues involved in building a basic semantic dependency. Ontological semantics, however, goes beyond that limit and into what is traditionally covered in pragmatics and discourse – on the assumption that, since these types of meanings are expressed in language they ought to be represented and manipulated. The following example illustrates both the treatment of additional types of meaning and additional expressive means in OS.

The English verb *prevent* has two senses in the ontological-semantic lexicon. The first sense is used, for example, in *The union organizer prevented a strike*. The basic meaning of *prevent-v1* can be roughly glossed as "the union organizer carried out an unnamed action or actions as a result of which a strike that was looming did not materialize."

In the *sem-struc* zone of *prevent-v1*, the meaning of the syntactic subject (^\$var1) has selectional restrictions on two facets (*sem* and *relaxable-to*). The precondition for the preventive event is the potential of the thing that is being prevented. The effect is that the thing that was prevented did not materialize. This is encoded using value ranges of certain *modalities* of the event. Modalities in OS are means of expressing speaker attitudes toward various elements of meaning (or even entire TMRs). *Potential* modality reflects the probability of a certain event or state of affairs taking place; *epistemic* modality measures the degree of factivity of the elements in its scope. Modalities are attributed to specific sources and take their values from abstract value ranges between 0 and 1. Epistemic modality 0 means that the event did not take place.[3]

---

[3] Ontological semantics recognizes modalities other than epistemic and potential. The inventory of modalities continues to grow and develop – witness the differences between the accounts in Nirenburg and Raskin 2003, Section 8.5.3. and in Nirenburg et al. (in preparation).

```
prevent-v1
   syn-struc
        root    $var0
        cat     verb
        subject
              root    $var1
              cat     noun
        direct-object
              root    $var2
              cat     noun
   sem-struc
     EVENT
        AGENT       ^$var1
                         sem           HUMAN
                         relaxable-to  ORGANIZATION

        PRECONDITION    ^$var2
                              sem      EVENT
                              modality  potential > .5
                                        epistemic < 1

        EFFECT          ^$var2
                              sem      EVENT
                              modality  epistemic 0
```

The second sense of *prevent* is the one used, for example, in *Negotiations prevented a strike*. Here the meaning is expressed as a CHANGE-EVENT: the event that is the meaning of the subject (^$var1) caused this change event; there was a potential for ^$var2 to take place; as a result of this change, ^$var2 has not taken place (its epistemic modality is 0). There is no need for overt specification of time dependencies, as causality determines temporal ordering. (Note that the syntactic analyzer currently used in OS recognizes certain –*ing* forms, e.g., *striking,* as nouns.)

```
prevent-v2
   syn-struc
        root    $var0
        cat     verb
        subject
              root    $var1
              cat     noun
        direct-object
              root    $var2
              cat     noun
   sem-struc
     CHANGE-EVENT
        PRECONDITION    ^$var2
                              sem      EVENT
                              modality  potential > .5
                                        epistemic < 1

        EFFECT              ^$var2
```

```
                              sem      EVENT
                              modality  epistemic 0

        CAUSED-BY
                         ^$var1
                              sem    EVENT
```

The above entries are simplified. For instance, we did not show the treatment of generic, timeless statements like *Being well prepared prevents unpleasant surprises,* whose meaning specification does not involve creating a *specific* instance of either of the two events but rather a so-called generic ontological instance representing any element of the class (e.g., *tigers* in *tigers are ferocious*). A special heuristic rule will be used in this case to determine whether generic ontological instances are appropriate. One clue that casts a vote for the generic reading is the present simple form of the verb, which is often associated with timelessness. Another is the indefiniteness of the direct object. However, these clues do not always guarantee that the statement is indeed generalized. One way to ascertain this fact is to call a special procedure to check whether ^$var2 co-refers with a specific concept instance either in the TMR or in the fact repository.

Procedures such as the above are called meaning procedures. For example, the meaning of intensifiers (e.g., *very*) is best expressed in terms of a meaning procedure – take the meaning of the adjective modified by *very*; it will be expressed as a range on some scale, e.g., the meaning of dim will be the range ($<>$ 0.1 0.4) on the scale of BRIGHTNESS. Relative values of such properties are expressed as ranges on abstract scales between 0 and 1. The expected absolute boundaries for such properties will be defined in the corresponding ontological concepts. Thus, while the HEIGHT property of HUMAN-ADULT may be marked as the range between 145 and 200 cm (these values are, in fact, overridable), the corresponding values for PROFESSIONAL-BASKETBALL-PLAYER may be between 180 and 230 cm. Values on relative scales can, thus, be resolved to ranges on absolute scales. The meaning of *very* roughly amounts to narrowing the range toward its extreme.

The presence of meaning procedures demonstrates that OS combines declarative and procedural semantics: on the one hand, knowledge in OS is separated from the processing engines; on the other hand, meaning is calculated dynamically.

2. Some Distinguishing Features of OS

The above, we hope, has presented an informative glimpse into the representational, descriptive and processing concerns of OS. Space restrictions prevent us from describing the many remaining elements of and decisions taken in OS. At this point, we'll attempt to

summarize important preferences and tenets of our "operational philosophy" that distinguish our work from other work and generally characterize OS.

The overall emphasis in the development of OS is given to:

- maintaining and improving a comprehensive meaning extraction and representation system;
- gradual enhancement of both the breadth and the depth of description (which stand in a trade-off relation relative to a given amount of resources) through a large-scale descriptive effort;
- formulation of heuristics for extracting intended meanings and supporting task-oriented processing;
- attaining adequate descriptive coverage of the mass of "uninteresting" linguistic phenomena that account for a huge majority of occurrences in texts.

Relatively lower (though not zero) priorities are accorded to the study of:

- capabilities of the representation medium;
- computational complexity of the processes;
- consistency of the knowledge base;
- individual "interesting" linguistic cases that are difficult for people to describe and judge (e.g., complex issues in quantification).

Ontological semantics is a comprehensive approach that does not relinquish responsibility for attaining its goal to other areas of study or rely on prerequisites that are either beyond the reach of the current state of the art or are expected as a result of large amounts of research by others. It is for this reason that OS includes in its purview all the preprocessing stages in text analysis: tokenization, morphology and syntax. Moreover, its purview is broader than that of traditional formal or lexical semantics: for example, it does not see a need to separate pragmatics and discourse into separate disciplines (allowing, for example, formal semanticists to bypass many kinds of meaning description on account of their being outside the purview of their theories).

Ontological semantics seeks to incorporate all the relevant topics often studied in relative isolation. These topics include language- and knowledge-related issues such as aspect, modality, time, causality, quantification, text-level relations, non-literal meaning, style, deductive and abductive presuppositions and entailments, nominal compounds, prepositional phrase attachment, the meaning of adjectives that do not semantically modify their syntactic governors, etc. They also include architecture and control issues connected with optimizing the search for the best semantic analysis in the potentially very large candidate space. We strive to develop our own accounts (we call them "microtheories") for

each such topic exploiting the research findings of others and modifying and expanding them in accordance with the practical goals and needs of OS. Unfortunately, no feasible solutions have been proposed in the literature for most microtheories that we need: much of the reported work is devoted to rather narrow topics (for example, meanings of single words at a grain size beyond the capabilities – and often the needs – of the analyzer) and still other proposals invent very complex formal notations without any emphasis on heuristics for the assignment of values to language phenomena. (Indeed, one of the knowledge acquisition rules of thumb in OS is not to include in the descriptions everything that can be said, only those parts that are actually useful.)

The purpose of all the microtheories is to help derive text meaning. Some microtheories are driven by the needs of the meaning representation (its content, not format). For example, all TMRs must at least attempt to resolve the deictic indices (speaker, hearer, place, time) and, more broadly, all references. Other microtheories are forward-chaining – they are triggered not by the needs of the output but by the occurrences of certain lexical and syntactic configurations in the input – for example, nominal compounds. Developing microtheories means acquiring heuristic rules for the treatment of specific phenomena. In OS, the strategy is to use any and all potentially useful kinds of knowledge in the left-hand sides of such rules. Some of this knowledge is encoded in the static knowledge sources of the system. Other sources of heuristics include syntactic or morphological information in the input and general properties of the input document right down to its formatting properties. Finally, the heuristic rules can use evidence from various co-occurrence measures for subsets of input in a corpus. We believe that supporting heuristic processing in a computational-semantic system is one of the more useful applications of corpus linguistics.

3. Balancing Desiderata and Practical Constraints

Ontological-semantic descriptive work is guided by the desiderata of breadth and depth of coverage. Constraining these, however, are the twin limitations of human resource availability and the computability of knowledge. In this section we present a number of resource acquisition choices, each of which is driven by practical needs and represents a whole class of phenomena whose continued development occupies the day-to-day work in OS. We fully understand that it is inappropriate to talk about a particular ontology or even a particular lexicon as "the correct one." While some constraints on ontology construction may be considered universal, there are many other equally acceptable choices in

specifying the world model to be used both as the meta-language for text meaning description and as a major source for knowledge to support reasoning.

Grain size of static knowledge specification. High precision in semantic description always incurs increased acquisition time and often increased processing challenges as well. For this reason, OS explicitly prefers well selected simplifications, defined as those not expected to noticeably impoverish the usefulness of the resulting TMR. For example, *about* in the meaning *travel about Europe* is mapped to the spatial relation INSIDE-OF rather than a) forcing the inclusion of a new concept with precisely this meaning or b) adding to the lexicon entry complex inferences or effects – e.g., seeing numerous different places. Similarly, the meaning *along* as in *trees grew along the road* is mapped to BESIDE, despite the loss of the precondition that the trees cover some unspecified stretch of the road interpreted by the speaker as significant.

Such conscious simplifications – which, we must emphasize, are always open to finer representation given the needs of an application and the resulting decision to thus deploy resources – apply to all aspects of knowledge acquisition. For example, due to the quagmire of semantic and pragmatic issues involved in interpreting conjunctions like *and* and *but* when they serve to link clauses, we level their semantics, creating a separate TMR for each clause linked by the functional binder 'conjoined'. Contrast this with our treatment of *since, because, so that,* etc., which are less ambiguous, more important for reasoning, and are thus rendered by the concepts CAUSED-BY, EFFECT, and ENABLEMENT, as applicable.

Another example of methodical simplification concerns the expression of time. Temporal properties of TMR elements are expressed using a very "lean" set of parameters: the ontological relations BEFORE (<) and LONG-BEFORE (<<) and their inverses AFTER (>) and LONG-AFTER (>>), the concepts START-TIME, END-TIME, SPEAKER-TIME, DURATION, and time measurement units. We also define two meaning procedures – *find-anchor-time* and *combine-time*. *find-anchor-time* attempts to determine the deictic index, or anchor, relative to which other times are defined in the text. It uses a variety of heuristics, including times of events mentioned in the text as well as the dateline of the text, if available. *find-anchor-time* may fail to determine the anchor, in which case, at present, the time statements in TMRs are made relative to an unknown time of speech, t_0. *combine-time* takes as input an anchor time and an expression denoting a time period, one of whose ends is the anchor, and returns the (actual or relative) time of the combination.

As an illustration, below are abbreviated meaning representations of some time-related word and phrase senses:

after-p4 (as in "ten hours after the operation")
 time (combine-time ^$var2.time ^$var3 after)

for-p5 (as in "dribbling for thirty seconds")
 time (duration $var2 (sem TEMPORAL-UNIT))

for-p6 (as in "for the time being")
 time (find-anchor-time)

Always open to reconsideration, these and other such decisions in favor of a coarse grain size, at a minimum, help to drive forward the work of supporting our broad-coverage text processing system.

Economy of expressive means. One tenet of OS is to avoid a proliferation of ontological concepts, in line with the recommendation by Hayes (1979) that the ratio of knowledge elements used to describe a set of elements of the world to the number of these latter elements must be kept as low as possible. The factors involved in deciding when and when not to introduce a new concept constitute something of a cognitive art too subtle to capture in a thumbnail sketch of a short article; however, a contrastive example might provide some insight.

Consider spatial *after*, as in *the first house after the stop sign*. Previous examples showed cases in which an ontologically available spatial relation was deemed close enough to describe given lexical entities. In this case, though, there is no such relation: NEXT-TO, the best candidate, is inappropriate since the house could be miles after the stop sign. One option would be to describe the semantics as a script: the first house one sees after passing the stop sign during a motion event – but this was judged too complex. Instead, we went ahead and added a new spatial relation, AFTER-SPATIAL, to accommodate this and synonymous lexical items from all the languages that express such a relation (we must emphasize that the ontology is language independent). One reason we did not expend the effort to more fully describe the semantics of *after* is that we have not yet developed a sophisticated microtheory of spatial relations. Therefore, we are not currently poised to make complex inferences about spatial relations in texts. However, when we do ultimately embark upon a microtheory of spatial relations – either stimulated by the demands of an application or through the natural overall progression of ontological-semantic research – we may reconsider whether a descriptive rather than an ontological representation of *after* would be most beneficial.

The opposite decision was taken in the case of *about* and its synonyms when they convey approximation. Approximation is too important for reasoning for us to make an ontological concept and consider all approximated entities thus resolved. Instead, we need meaning procedures to fix ranges on scales. Preliminary analysis suggested that a 7% expansion either way

around a magnitude works reasonably well in most cases: e.g., about 5 gallons is 4.65 to 5.35 gallons; about 150 lbs. is 139.5 to 160.5 pounds. However the 7% rule produces bad results in some instances: e.g., resolving about 6 feet tall to 5'7''-6'5'' is bad. What we need, instead, is 7% of the amount by which people's heights can vary, which is about 2 feet. The 7% rule applied to 2 feet yields 5'10.5'' to 6'1.5'' – a much better reflection of reality. Another failure of the straightforward 7% rule involves clock time. For pragmatic reasons, saying *around 10* or *around 5:30* implies 10 minutes in either direction, whereas *around 5:15* permits a smaller range and *around 7:07* a smaller range still. This example of clock time – work on which is still under way – shows that some issues considered to be of crucial import to TMRs and the reasoning they support are immediate priorities in OS despite the overhead of research and implementation they impose.

Including results of abductive reasoning in TMRs. To offset any impression that the current state of development of OS always opts for shallower analysis, we offer an example relating to our evolving treatment of verbs that that have a relatively general meaning despite the fact that, in a given context, speakers understand them to carry more meaning.

Consider once again, for example, *make* in the sense of *create an artifact* (i.e., a man-made physical object). This sense of *make*, which is only one of many, can contextually imply different specific types of events including but not limited to baking (*make a cake*), sculpting (*make a sculpture*), recording (*make a CD*), filming (*make a film*) and knitting (*make a sweater*). The verbal sense of *make* that covers all of these examples says that *make* is a CREATE-ARTIFACT event whose subject is HUMAN and whose object is an ARTIFACT.

Representing this meaning of *make* as CREATE-ARTIFACT captures all the meaning that is explicitly conveyed by this lexical item. In that sense, our analysis should be considered complete. However, we as users of language know much more precisely what activity is actually carried out in making a cake: by default, it's baking. When we set about to reflect this knowledge in the ontological-semantic knowledge resources, we list CAKE as a default THEME of BAKE. Whereas inheritance in the ontology allows the match of CAKE with CREATE-ARTIFACT, a more exact, constrained match obtains between CAKE and BAKE. (Note that CAKE matches CREATE-ARTIFACT because the THEME of CREATE-ARTIFACT includes a union of the fillers of THEME of all its descendants, which include, among many others, BUILD, PREPARE-FOOD, MAKE-LAW, AUTHOR-EVENT, FILM-EVENT, RECORD-SOUND, RECORD-TEXT, CREATE-FABRIC-MATERIAL-ARTIFACT, etc., as well as *their* descendants.)

An important question at this point is whether we should act upon the extra knowledge that cakes are usually baked or simply allow the correct, though more vague direct TMR of the input text to stand. The answer depends on the circumstances of an application. For example, in machine translation, all other things being equal, if the direct TMR can be used as the input to successful target text generation, there is no need for any specialization procedure. There are, however, many cases in which such a procedure would improve the performance of various analysis tasks. For example, such a procedure helps to resolve certain ambiguities, e.g., by preferring the *oven* to the *stove* sense of *range* in *John prepared the cake using the range*—see a detailed description in Mahesh et al., 1997. It also helps in the selection and resolution of referring expressions. For example, in *John made a good cake though the oven was not hot enough*, the definite referring expression *the oven* is legitimately used for a discourse-initial reference—and, therefore, no co-reference resolution procedure should be called—because the ontology lists OVEN as the default instrument of BAKE. This type of inference is ultimately made possible by the availability of ontological scripts.

In principle, if we decide to carry out proactive abductive meaning specialization whenever we come across constructions whose verbs have rather general meanings (note, incidentally, that many such verbs have meanings that are even vaguer and less specific than that of *make*–consider, for example, such verbs as *use, do, have* or *get*), we can call the meaning procedure *seek-specification* in each such case. However, the analyzer currently triggers specialization exclusively on an as-needed basis, not proactively because proactive evocation of *seek-specification* will result in conclusions that might never be used.

Ontological semantics defines many types of meaning procedures, and their application results in both more precise and more fine-grain specification of text meaning. In fact, when all meaning procedures are applied, the resulting TMR will contain information that was both overtly present in the original text and information abductively inferred form the background knowledge of the system. We call the former type of TMR basic and the latter, extended.

4. Summary, Status and Future Work

In this paper, we first briefly and informally illustrated, using a few annotated examples, the static and dynamic knowledge resources of OS. We then presented the main motivations and desiderata of our approach and discussed issues relating to how to make ontological-semantic applications feasible through judicious stepwise enhancement of static and dynamic knowledge sources while at all times maintaining a working system.

The latest implementation of OS uses an ontology

of about 6,500 concepts each of which is a named collection of property-value pairs (on average, every concept in the current ontology has 16 properties defined for it), with the "meta-metalanguage" of properties numbering at the time of this writing about 350 relations and attributes. The current English lexicon is about 35,000 entries and growing. There are also Spanish and Chinese lexicons, and lexicons for other languages are under construction. We have also developed a collection of ever growing onomasticons. The English onomasticon at present contains over half a million entries. In addition to the lexicons and the ontology, the static knowledge sources in OS also include morphological and syntactic grammars to support text analysis and generation, as well as a growing fact repository.

We are actively developing a variety of ontological-semantic microtheories including microtheories of time, quantification, approximateness, issues relating to sets and ordered lists, several different aspects of reference, modality, discourse cohesion, script-based abductive reasoning. We are also developing pedagogical microtheories devoted to ontology, lexicon and script acquisition. Additional microtheories are planned for development as soon as the current ones reach minimum utility levels. We continue to augment the basic knowledge resources and are investigating methods of using current knowledge resources to speed up acquisition of knowledge resources for languages other than those already in the system.

We are also developing a system for the automatic extraction of fact repository elements from TMRs. With the help of this system, we plan to acquire a large fact repository that will be used as the search space in the ontological-semantic approach to information extraction and question answering.

Many difficulties still remain in the path toward high-quality, broad-coverage extraction and manipulation of meaning from texts. Still, we believe that the ontological-semantic approach is on the right path toward this goal – mainly because it is predicated on detailed and flexible descriptive work on language and world knowledge, does not rely on unattainable or uncomputable prerequisites and is driven by the requirements of specific high-end computational-linguistic applications.

References

Hayes, P. 1979. The Naive Physics Manifesto. In: Mitchie, D (ed.), *Expert Systems in the Microelectronic Age*. Edinburgh: Edinburgh University Press.

Mahesh, K., S. Nirenburg, and S. Beale 1997. If You Have It, Flaunt It: Using Full Ontological Knowledge for Word Sense Disambiguation. *Proceedings of TMI-97*, Santa Fe, NM, 1-9.

Nirenburg, S., M. McShane and S. Beale (in preparation). Beyond Basic Semantic Dependencies I: The Microtheory of Modality in Ontological Semantics.

Nirenburg, S. and V. Raskin. 2003. *Ontological Semantics*. MIT Press (forthcoming).

Onyshkevych, B. 1997. An Ontological-Semantic Framework for Text Analysis. Unpublished Ph.D. thesis, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA.