To appear in Proc. IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing, Aug. 19-21, 1995, Montreal.

A Situated Ontology for Practical NLP*

Kavi Mahesh and Sergei Nirenburg

Computing Research Laboratory
Box 30001, Dept. 3CRL
New Mexico State University
Las Cruces, NM 88003-8001 USA
mahesh@crl.nmsu.edu sergei@crl.nmsu.edu

Abstract

A situated ontology is a world model used as a computational resource for solving a particular set of problems. It is treated as neither a "natural" entity waiting to be discovered nor a purely theoretical construct. This paper describes how a semantico-pragmatic analyzer, Mikrokosmos, uses knowledge from a situated ontology as well as from language-specific knowledge sources (lexicons and microtheory rules). Also presented are some guidelines for acquiring ontological concepts and an overview of the technology developed in the Mikrokosmos project for large-scale acquisition and maintenance of ontological databases. Tools for acquiring, maintaining, and browsing ontologies can be shared more readily than ontologies themselves. Ontological knowledge bases can be shared as computational resources if such tools provide translators between different representation formats.

1 A Situated Ontology

World models (ontologies) in computational applications are artificially constructed entities. They are created, not discovered. This is why so many different world models were suggested. Many ontologies are developed for purely theoretical purposes or without the context of a practical situation (e.g., Lenat and Guha, 1990; Smith, 1993). Many practical knowledge-based systems, on the other hand, employ world or domain models without recognizing them as a separate knowledge source (e.g., Farwell, et al. 1993). In the field of natural language processing (NLP) there is now a consensus that all NLP systems that seek to represent and manipulate meanings of texts need an ontology (e.g., Bateman, 1993; Nirenburg, Raskin, and Onyshkevych, 1995). In our continued efforts to build a multilingual knowledge-based machine translation (KBMT) system using an interlingual meaning representation (e.g., Onyshkevych and Nirenburg, 1994), we have developed an ontology to facilitate natural language interpretation and generation. The central goal of the Mikrokosmos project is to develop a system that produces a comprehensive Text Meaning Representation (TMR) for an input text in any of a set of source languages.[1] Knowledge that supports this process is stored both in language-specific knowledge sources and in an independently motivated, language-neutral ontology (e.g., Carlson and Nirenburg, 1990; Mahesh, 1995).

An ontology for NLP purposes is a body of knowledge about the world (or a domain) that a) is a repository of primitive symbols used in meaning representation; b) organizes these symbols in a tangled subsumption hierarchy; and c) further interconnects these symbols using a rich system of semantic and discourse-pragmatic relations defined among the concepts. In order for such an ontology to become a computational resource for solving problems such as ambiguity and reference resolution, it must be actually constructed, not merely defined formally, as is the practice in the field of formal semantics. The ontology must be put into well-defined relations with other knowledge sources in the system. Figure 1 illustrates the Mikrokosmos architecture for analyzing input texts. In this application, the ontology supplies world knowledge to lexical, syntactic, and semantic processes.

*Research reported in this paper was supported in part by Contract MDA904-92-C-5189 from the U.S. Department of Defense.

[1] The first system prototype is an analyzer of Spanish.

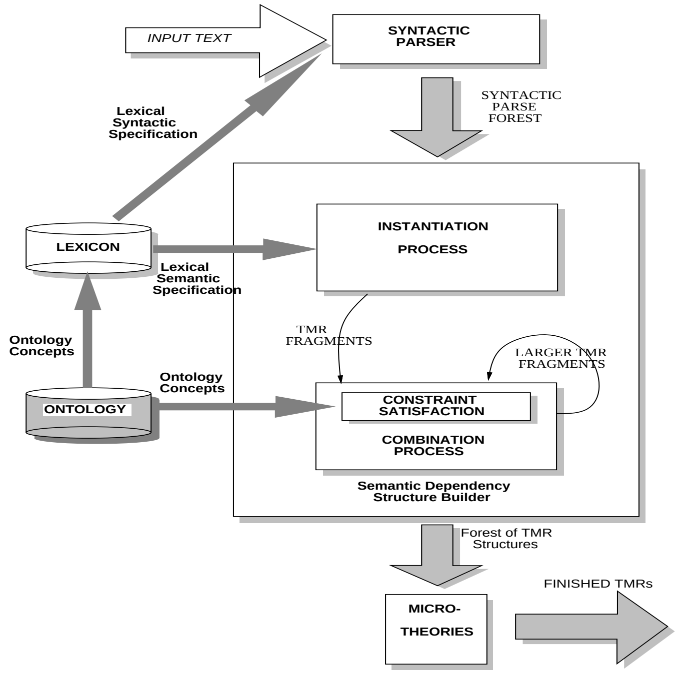

Figure 1: The Ontology Situated in the Mikrokosmos NLP Architecture. It Supplies Conceptual Knowledge both for Lexical Representation and for Constraining Semantic Interpretation.

In what follows we present the Mikrokosmos view of what an ontology is, how to develop it, and how it could be shared with other systems, projects, and people. We show how a language-independent ontology works in close collaboration with a language-specific lexicon in representing word meanings and mapping texts to TMRs.

1.1 Our Product: The Mikrokosmos Ontology

The Mikrokosmos project is supposed to process texts about mergers and acquisitions of companies. However, since the input language is unrestricted, the ontology must, in fact, cover a wide range of concepts outside this particular domain. The analyzer encounters a variety of well-known problems which require a significant amount of world knowledge.

We are currently in the process of a massive acquisition of objects and events related to the domain of company mergers and acquisitions.[2] Over the period of about three months, the Mikrokosmos ontology has acquired over 2000 concepts organized in a tangled hierarchy with ample interconnection across the branches. The ontology emphasizes depth in organizing concepts and reaches depth 10 or more along a number of paths. The branching factor is kept much less than 5 on an average. Each concept has, on average, 10 to 15 slots linking it to other concepts or literal constants. The top levels of the hierarchy have proved very stable as we are continuing to acquire new concepts at the lower levels.

Unlike many other ontologies with a narrow focus (e.g., Casati and Varzi, 1993; Hayes, 1985; Mars, 1993), our ontology must cover a wide variety of concepts in the world. In particular, our ontology cannot stop at organizing terminological nouns into a taxonomy of objects and their properties; it must also represent a taxonomy of (possibly, complex) events and include many interconnections between objects and events to support a variety of disambiguation tasks.

The Mikrokosmos ontology is a successor to ontologies developed in earlier projects at Carnegie Mellon University (e.g., Carlson and Nirenburg, 1990). However, it differs from its predecessors significantly in its content as well as in quality. Its domain of mergers and acquisitions of companies is one of considerable interest and significance. It has much richer concepts and a higher degree of connectivity between concepts, a tighter control of excessive

[2] In parallel, a Spanish lexicon that maps lexemes to concepts in this ontology is also being acquired on a massive scale.

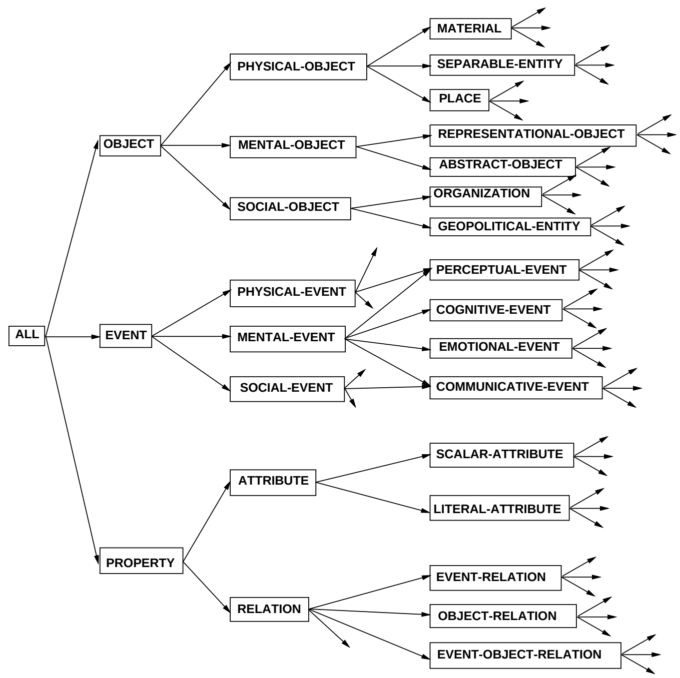

Figure 2: Top-Level Hierarchy of the Mikrokosmos Ontology Showing the First Three Levels of the Object, Event, and Property Taxonomies.

fanouts, and already has more than twice the number of concepts in previous ontologies. The concepts being acquired are also based not just on intuition but on reliable sources (e.g., SICM, 1987).

All entities in the Mikrokosmos ontology are classified into *objects, events,* and *properties.* Figure 2 shows the top-level hierarchy in the ontology. Objects, events, and properties constitute the *concepts* in the ontology which are represented as *frames.* Each frame is a collection of slots with one or more facets and fillers. The slots (including inherited ones) collectively define the concept by specifying how the concept is related to other concepts in the ontology (through *relations*) or to literal or numerical constants (through *attributes*). Lexicon entries represent word or phrase meanings by mapping them to concepts in the ontology.

A TMR is a result of instantiating concepts from the ontology that are referred to in a text and linking them together according to the constraints in the concepts as well as those listed in lexicon entries and special TMR building rules. A number of concepts in the domain of mergers and acquisitions are located under the ORGANIZATION subtree under SOCIAL-OBJECTs and the BUSINESS-ACTIVITY subtree under SOCIAL-EVENTs (see Figure 2). Figure 3 shows a sample frame in the ontology along with a lexical mapping to that concept.

There is an inherent duality between properties represented as concepts and slots in free-standing entities. Every slot is a property—a relation or an attribute—that is well defined in the ontology as a concept under the property hierarchy. Properties are defined as mathematical relations by specifying their domains and ranges. Formally, the Mikrokosmos ontology is a directed graph with only two patterns in its subgraphs corresponding to the two types of slots: relations and attributes. Further details of the representation of frames are not particularly relevant here and are described elsewhere (Mahesh, 1995).

2 What Does the Ontology Do for NLP?

The ontology aids natural language processing in the following ways:

- It represents selectional preferences for relations between concepts. This knowledge is invaluable for resolving ambiguities by means of the constraint satisfaction process shown in Figure 1. For example, if the theme of "adquirir" in a sentence is not an ABSTRACT-OBJECT, then the meaning of that verb is likely to be ACQUIRE

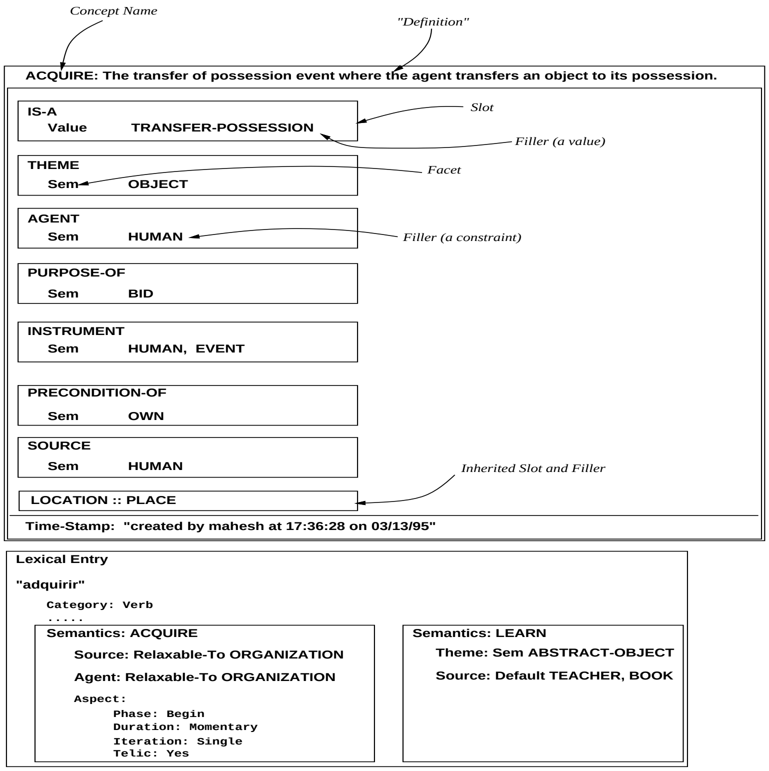

Figure 3: Frame Representation for Concept ACQUIRE. Also Shown is a Part of the Lexical Entry for the Spanish Verb "adquirir" with Semantic Mappings to ACQUIRE and LEARN Events. The Mappings Modify the Constraints in the Ontology and Add New Information such as Aspect.

and not LEARN (see Figure 3).

- It enables inferences to be made from the input text using knowledge contained in the concepts. This can help resolve ambiguities as well as fill gaps in the text meaning. A default value from the ontological concept can be filled in a slot, for example, when a text does not provide a specific value.

- It enables inferences to be made using the topology of the network, as in searching for the shortest path between two concepts. Such search-based inferences can support metonymy and metaphor processing, figuring out the meaning of a complex nominal or be used in constraint relaxation when the input cannot be treated with the available knowledge.

The ontology parallels Mikrokosmos lexicons both in the representations and in development. Word meanings are represented partly in the lexicon and partly in the ontology (see Figure 3 for an example). In principle, the separation between ontology and lexicon is as follows: language-neutral meanings are stored in the former; language-specific information, in the latter. In a multilingual situation, it is not easy, however, to determine this boundary. As a result, ontology and lexicon acquisition involves a process of daily negotiations between ontology and lexicon acquirers. The easiest solutions to many difficult problems in lexical semantic representation require the addition of a new concept to the ontology under certain "catch all" frames (this is the only solution in the "word sense approach" to ontology development where word senses are mapped one to one to concept names (e.g., Bateman, et al. 1990; Knight and Luk, 1994)). In Mikrokosmos, a set of guidelines was developed suggesting ways of finding solutions to lexical problems without adding "catch all" concepts (see below for a sample of these guidelines).

The ontology also aids meaning representation and, in particular, lexical representation as follows:

- It forms a substrate upon which word meanings in a particular language are constructed in the lexicon (see, e.g., Nirenburg and Levin, 1992). It guarantees that every symbol used in representing lexical semantics is defined as a concept, is well-formed, and has known relations to all other symbols. This is a big methodological advantage since it allows us to partition the task of developing multilingual machine translation (MT) systems into the independent development of analyzers and generators for the different languages. The ontology serves as a common ground both between an analyzer and a generator and between different languages. It provides a way for a clean separation of language independent world knowledge from linguistic knowledge.

- It makes lexical representations highly parsimonious. Meaning common to many words in a language can be factored out and represented in the ontology as a part of the concept to which the words map. Moreover, since concepts in the ontology can be modified further through the relations and attributes in their slots, the ontology allows the lexicon to capture word senses through far fewer entries than found in a typical dictionary. For example, Nirenburg, Raskin, and Onyshkevych (1995) have shown that the 54 meanings for the Spanish verb "dejar" listed in the Collins Spanish-English dictionary can be collapsed into just 7 different lexical mappings using the Mikrokosmos approach. Much of this power comes from the expressiveness of the representation that allows complex combinations and modifications of ontological concepts in meaning representations. See Onyshkevych and Nirenburg (1994) for detailed examples.
- It allows vague and incomplete lexical representations which are nevertheless well-formed. For example, in mapping certain adjectives and adverbs, the lexicon needs to refer to the agent slot (say) of an event without knowing which particular event it is. This can be done in Mikrokosmos by directly referring to the agent slot of the event concept (even though not every event has an agent) and using a variable-binding mechanism in lexical representation to map to the particular instance of an event in a given sentence.
- It allows variable-depth semantics where lexical semantics can range from a direct mapping to a concept without any modifications all the way to a complex combination of several concepts with various additional constraints, relaxations, and additional information such as time, aspect, quantities, and their relationships. It helps avoid unnecessarily deep decomposition of complex meanings by resorting to a controlled proliferation of primitive concepts (see below).
- It also enables sophisticated ways of reducing manual effort in lexicon acquisition. For example, it allows a derivational morphology engine to automatically suggest concepts for the lexical mappings of the derived words which can then be refined by the acquirer instead of entering from scratch.

3 Computational Ontologies for NLP

Just as there is no single grammar that is the "true" grammar of a natural language, it is reasonable to argue that there is no unique ontology for any domain. The Mikrokosmos ontology is one possible classification of concepts in its domain constructed manually according to a well developed set of guidelines. Its utility in NLP can only be evaluated by the quality of the translations produced by the overall system or through some other evaluation of the overall NLP system (such as in an information extraction or retrieval test).

This is not to say that the ontology is randomly constructed. It is not. Its construction has been constrained throughout by the guidelines (see Section 4.2 below) as well as by the requirements of lexical semantics and their acquisition.

In NLP work, the term "ontology" is sometimes also used to refer to a different kind of knowledge base which is essentially a strict hierarchical organization of a set of symbols with little or no internal structure to each node in the hierarchy (e.g., Farwell, et al. 1993; Knight and Luk, 1994). Frames in the Mikrokosmos ontology, however, have a rich internal structure through which are represented various types of relationships between concepts and the constraints, defaults, and values upon these relationships. It is from this rich structure and connectivity that one can derive most of the power of the ontology in producing a TMR from an input text. Mere subsumption relations between nearly atomic symbols do not afford the variety of ways listed above in which the Mikrokosmos ontology aids lexicon acquisition and disambiguation in language processing.

The above distinction between highly structured concepts and nearly atomic concepts can be traced to a difference in the grain size of decomposing meanings. A highly decompositional (or compositional) meaning representation relies on a very limited set of primitives (i.e., concept names). As a result, the representation of many basic concepts becomes too complex and convoluted. The other extreme is to map each word sense in a language to an atomic concept. As a result, the nature of interconnection among these concepts becomes unclear, to say nothing about the explanatory power of the system (cf. the argument about the size of the set of conceptual primitives in

Hayes, 1979). In Mikrokosmos, we take a hybrid approach and strive to contain the proliferation of concepts for a variety of methodological reasons, such as tradeoffs between the parsimony of ontological representation and that of lexical representation and the need for language independent meaning representations. Control over proliferation of concepts is achieved by a set of guidelines that tell the ontology acquirer when not to introduce a new concept (see Figure 5).

The Mikrokosmos ontology also makes a clear distinction between conceptual and episodic knowledge and includes only conceptual knowledge. Instances and episodes are acquired in a separate knowledge base called the *onomasticon.* The methodology for acquiring the onomasticon includes a significant amount of automation and is very different from ontology acquisition which is done manually via continual interactions with lexicographers.

3.1 Characteristics of the Mikrokosmos Ontology

Some of the key characteristics of the Mikrokosmos approach to ontology development for practical NLP can be summarized as follows:

1. Language independence: The ontology is language independent in two ways:
   (a) It is not specific to any particular language such as English or Spanish, though, for convenience, concepts are given English names.
   (b) The concepts in the ontology do not have a one-to-one mapping to word senses in natural languages. Many concepts may not map to any single word in a language; other concepts may map to more than one word in the same language and vice versa.
2. Independent motivation: Concept acquisition is not dictated by the lexicon. Ontology development and lexicography are sister processes that aid each other and at the same time constrain each other in significant ways.
3. Well-formedness, according to well defined slot, facet, and filler representations. Each concept has a rich structure to it. The ontology is internally consistent in structure, naming, and content throughout as per well developed guidelines.
4. Consistency and compatibility with the lexicon, the semantic analyzer, the language of the TMR, and other components of the system in which it is situated.
5. Comprehensibility and simplicity: In addition to being a computational entity, the ontology must be easy to browse and present; it should facilitate acquirer training. For example, the ontology does not use And-Or trees with disjunctive inheritance because such inheritance is rather difficult to comprehend and figure out for both ontology and lexicon acquirers.
6. Utility: It must ultimately aid language processing in resolving a variety of ambiguities and making necessary inferences. The situated ontology is language independent but is built for the specific purpose of natural language processing.
7. Limited proliferation: Situated development limits the size of the ontology though presumably any piece of knowledge could be useful for the task in question. The ontology is not limited to its domain but is more developed in the chosen domain.
8. Separation of episodic knowledge: The knowledge we are acquiring is conceptual, not episodic.
9. Reliance on technology. Acquisition is made more tractable by the deployment of latest technologies: faster machines, color graphical user interfaces, graphical browsers and editors, on-line lexicons, corpora, and other ontologies, as well as semi-automated tools for consistency maintenance and interfaces for lexicographer interactions.

4 Ontology Acquisition: Situated Development

A situated ontology is best developed incrementally, relying on continuous interactions with other knowledge sources. In practice, this translates into the concurrent development of both the ontology and the lexicon through a continual negotiation. This negotiation to meet the constraints on both a lexical entry and a concept in the ontology leads to the best choice in each case. It also ensures that every entry in each knowledge base is consistent, compatible with its counterparts, and has a purpose towards the ultimate objective of producing quality TMRs. The ideal method of situated development of knowledge sources for multilingual NLP is one where an ontology and at least two lexicons for different languages are developed concurrently. This ensures that the ontology is truly language independent and that representational needs of more than one language are taken into account.

Ontology acquisition is a very expensive empirical task. Situated development is a good way to constrain the process and make it attainable. For example, in the NLP situation, the acquirer must focus on concepts in the

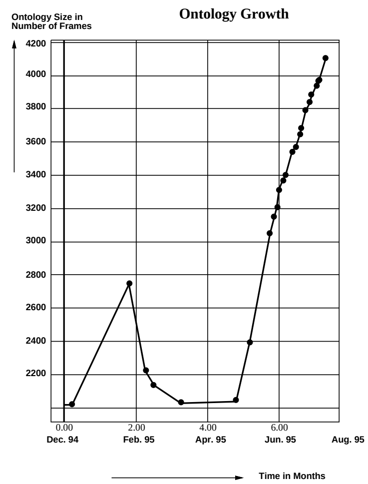

Figure 4: Rate of Growth of the Mikrokosmos Ontology.

domain of the input texts and thereby increase the ratio of the number of concepts (or their slots) that are actually used in processing a set of texts to the total number of concepts present in the ontology. The best example of a large ontological database acquired with enormous efforts but entirely out of any situation is CyC (Lenat and Guha, 1990). While the utility of CyC in a particular situation such as large scale NLP is yet to be demonstrated, it is also true that most projects cannot afford to spend as many resources as it has taken to develop CyC and must strive to constrain acquisition significantly or share existing ontologies.

In addition to situated development, ontology acquisition can be made more tractable by partial automation and by the use of advanced tools. However, the kind of language-independent ontology described above, where each concept has a rich structure and relationships to other concepts, is almost impossible to acquire automatically from a corpus of texts that has been tagged on mere syntactic or superficial semantic features. It is conceivable, however, that an ontology can be acquired incrementally from a corpus where each text is tagged with an entire meaning representation (TMR) for the text.

Figure 4 shows the rate of growth of the ontology over the last eight months. This graph shows our initial acquisition phase starting from an older ontology developed at Carnegie Mellon University (Carlson and Nirenburg, 1990), an intermediate clean up phase when we deleted hundreds of questionable and unrelated concepts and the current phase of massive acquisition.

4.1 Technology for Ontology Development

In order to aid ontology acquisition and maintenance, to check its consistency, and to support interactions with lexicographers, a variety of semi-automated tools have been developed and deployed in the Mikrokosmos project. Tools are in use for:

- browsing the hierarchies and the internals of concepts in the ontology;
- graphical editing support: the "Mikrokarat" tool[3] supports complete functionality for editing the graph structures in an ontology;
- translating between two different representations: the object oriented one suitable for computational purposes and the plain text representation that is more suitable for certain other programs and manual search and maintenance purposes;
- various types of consistency checking both within the ontology and with the lexicon, and for conformance with the guidelines;

[3] Developed by Ralf Brown at the Center for Machine Translation, Carnegie Mellon University.

---

1. Do not add instances as concepts in the ontology. Rules of thumb for distinguishing an instance from a concept are:
   - **Instance-Rule1:** See if the entity can have its own instance. Instances do not have their own instances; concepts do.
   - **Instance-Rule2:** See if the thing has a fixed position in time and/or space in the world. If yes, it is an instance. If not, it is a concept. For example, SUNDAY is a concept, not an instance, because it is not a fixed position in time ("last Sunday," "the first Sunday of the month," etc.).
2. Do not decompose concepts further into other concepts merely because you can. It is important to focus on building those parts that are needed immediately for the Mikrokosmos task. For example, though EVENTs like BUY or MARKETING can be decomposed to a great extent, unless there is an indication that detailed decompositions are needed for the task, do not decompose such EVENTs.
3. Do not add a concept if there is already one "close" to it or slightly more general than the one being considered. Consider the expressiveness of the representation provided by gradations (i.e., attribute values) before adding separate concepts. For example, we do not need separate concepts for "suggest," "urge," and "order." They are all gradations of the same concept, a DIRECTIVE-ACT, with various degrees of force which can be captured in an appropriate attribute.
4. Do not add specialized EVENTs with particular arguments as new concepts. For example, we do not need separate concepts for "walk to airport terminal" and "walk to parking lot."
5. Certain elements of text meaning such as aspect, temporal relations, attitudes, and so on, that are instance-specific belong only in the TMRs. For example, BREAKFAST is probably a concept in the ontology (and a subclass of MEAL, say) but a meal that happened at 3 O'clock on a particular day is not a separate concept in the ontology.
6. One must also remember that ontologies are supposed to be language independent. As such, if any part of a meaning representation is specific to a particular language that part does not belong in the ontology.
7. Mikrokosmos representations have a very expressive `set` and `subset` notation. Hence, there is no need to create ontological concepts for collections of different types of things in the world.

Figure 5: Guidelines for Deciding What Concepts to Add.

- supporting interactions with lexicon acquirers through an interface for submitting requests for changes or additions.

This set of tools is being shared across geographical, disciplinary, and project group boundaries on a daily basis. For example, on a typical day, 8-10 people browse the Mikrokosmos ontology using the Mikrokarat tool.

4.2 Acquisition Methodology and Guidelines

The basic methodology for concept acquisition employed in the Mikrokosmos project involves a fine-grained cycle of requests for concepts from the lexicon acquisition team and the resulting responses which may involve pointing out an existing concept, adding a new concept, enhancing the internal structure of one or more concepts, or suggesting a different lexical mapping for the word in question. If it is determined that a word sense requires a new concept in the ontology, the "algorithm" applied for adding the new concept hinges on viewing the ontology as a discrimination tree. The acquirer discriminates from the top down until at some point there is no child that subsumes the meaning in question. A new concept is added as a child at that point. In the Mikrokosmos project, sets of guidelines have emerged for making various kinds of decisions in ontology acquisition. These guidelines, some of which are shown in Figures 5 and 6, collectively define the methodology for ontology building.

5 Discussion and Conclusion

We are agnostic about the existence of a truly general ontology of the world. In the Mikrokosmos project, we are concerned about an ontology only as an empirically constructed tool that aids NLP. In spite of being based on an underlying theory of knowledge representation and being situated in the NLP task, the Mikrokosmos ontology is ultimately an artifact constructed according to our needs, biases, abilities, and tools.

———————————————————————————————————

1. Whenever possible, use "scientific" rather than lay terms.
2. Use the English word whose word sense maps to the concept when possible.
3. Use only alphabetic characters and '-'.
4. Do not use plurals in concept names.
5. Consistency across concepts is more important than conformance with a dictionary. Since there is no single word in English for "forprofit" but there is one for "nonprofit," we have no choice but to hyphenate both for-profit and non-profit.
6. Do not use names longer than three words (and the two '-'s between them).
7. Avoid compound nouns in concept names unless the relations between the meanings of the nouns are unambiguous and obvious. For example, do not use TIME-UNIT; use UNIT-OF-TIME instead.
8. When there is a large discrepancy in frequencies between different word senses of a word, name the most frequent one with just the word and add hyphens to others. For example, the concept BANK will stand for the "money holding place" sense of bank; BANK-RIVER and BANK-DEVICE will stand for a river bank and a bank of generators or disk drives.
9. Keep in mind that no two frames can have the same name in the ontology. Many properties and objects tend to suggest the same name. We must use different names for them. For example, if "employee" is both an object and a property, name the object EMPLOYEE and the property EMPLOYED-BY (and its inverse EMPLOYER-OF).
10. For relation names, append typical prepositions to distinguish them from objects as well as to indicate the direction of the relation (and hence distinguish it from its inverse relation) as per the following guidelines:
    - Use hyphenated names for relationss in both directions or use simple names in both directions.
    - If possible, do not use the same preposition in both a relation and its inverse.
11. Try to be consistent in the names of ontological concepts while going up or down a subtree. For example, EVENT has subclasses MENTAL-EVENT, PHYSICAL-EVENT, and SOCIAL-EVENT.
12. Whenever possible, try to include an indication of some distinguishing characteristic of the concept in its name. It is especially useful to include a characteristic that distinguishes the concept from its immediate siblings. For example, VOLUNTARY-VISUAL-EVENT and INVOLUNTARY-VISUAL-EVENT indicate events that involve vision, with voluntary or involuntary participation (perhaps corresponding to the English verbs "look" and "see").
13. English words must be used consistently in only one sense throughout the ontology. For example, we should not have a GROCERY-STORE as well as a STORE-MEDICINE (the former an object and the latter an event). Perhaps we should name them GROCERY-STORE and PRESERVE-MEDICINE (but then we must rule out PEACH-PRESERVE for a jam made of peaches).

———————————————————————————————————

Figure 6: Guidelines for Naming a Concept.

5.1 Ontology as a Sharable Resource

Only ontologies that are constructed as computational entities can be shared effectively. The informational content of a computational ontology is much more important in solving practical problems than the form in which it is represented. We do not place much emphasis on the representational formalism used in an ontological database. We are open to converting the Mikrokosmos ontology into another format if there is a standard that emerges. Until such time, we must live with different choices of primitives, conceptual relations and configurations, and their representation.

Given the diversity in ontological designs, a good way to share them as computational resources may be to share a common set of tools which provide a substrate upon which various translators can be built for converting an ontology from one representation to another so that they can be shared between modules and subgroups of a project, with other projects, and with the community at large. We already have such a system in operation within the Mikrokosmos project where the ontology is routinely translated between two different representational forms and is shared among the large group of people working for the project. We also have the beginnings of other translators such as one that converts TMRs to a standard template of the TIPSTER information retrieval initiative.

5.2 Conclusion

The Mikrokosmos ontology is not only a situated ontology as such. Its acquisition is also situated in the context of close interactions with lexicon acquisition and semantic analysis. We have shown how such a state of affairs can help limit the proliferation of concepts to make ontology development more tractable. Knowledge contained in such an empirically acquired computational ontology can be shared with other programs and projects by adopting a common set of tools and building translators between different idiosyncratic forms of representation.

References

Bateman, J. A. (1993). Ontology construction and natural language. In Proc. International Workshop on Formal Ontology. Padua, Italy, pp. 83-93.

Bateman, J. A., Kasper, R. T., Moore, J. D., and Whitney, R. A. (1990). A general organization of knowledge for NLP: the PENMAN upper model. Technical Report, USC/Information Sciences Institute, 1990.

Carlson, L. and Nirenburg, S. (1990). World Modeling for NLP. Technical Report CMU-CMT-90-121, Center for Machine Translation, Carnegie Mellon University, Pittsburgh, PA.

Casati, R. and Varzi, A. (1993). An Ontology for Superficial Entities I: Holes. In Proc. International Workshop on Formal Ontology. Padua, Italy, pp. 127-148.

Farwell, D., Guthrie, L., and Wilks, Y. (1993). Automatically creating lexical entries for ULTRA, a multilingual MT system, *Machine Translation,* vol. 8:3, pp. 127-146.

Hayes, P. J. (1979). Naive physics manifesto. *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press.

Hayes, P. J. (1985). Naive Physics I: Ontology for liquids. In *Formal theories of the common sense world,* ed. J. Hobbs and R. C. Moore, pp. 71-107. Norwood, NJ: Ablex.

Knight, K. and Luk, S. K. (1994). Building a Large-Scale Knowledge Base for Machine Translation. In *Proc. Twelfth National Conf. on Artificial Intelligence,* (AAAI-94).

Lenat, D. B. and Guha, R. V. (1990). *Building Large Knowledge-Based Systems.* Reading, MA: Addison-Wesley.

Mahesh, K. (1995), Ontology Development: Ideology and Methodology. Technical Report (under preparation), Computing Research Laboratory, New Mexico State University.

Mars, N. (1993). An ontology of measurement units. In Proc. International Workshop on Formal Ontology. Padua, Italy, pp. 297-303.

Nirenburg, S. and Levin, L. (1992). Syntax-driven and ontology-driven lexical semantics. In *Lexical semantics and knowledge representation,* Pustejovsky, J. and Bergler, S. Eds. Heidelberg: Springer Verlag.

Nirenburg, S., Raskin, V. and Onyshkevych, B. (1995). Apologiae Ontologiae. Proceedings of the Conference on Theoretical and Methodological Issues in Machine Translation. Leuven, Belgium, July.

Onyshkevych, B. and Nirenburg, S. (1994). The lexicon in the scheme of KBMT things. Technical Report MCCS-94-277, Computing Research Laboratory, New Mexico State University.

Smith, B. (1993). Ontology and the logistic analysis of reality. In Proc. International Workshop on Formal Ontology. Padua, Italy, pp. 51-68.

Standard Industrial Classification Manual. (1987). Executive Office of the President, Office of Management and Budget, US Government. National Technical Information Service.