

Practical World Modeling for NLP Applications

Lynn Carlson

U.S. Department of Defense
Ft. George G. Meade, MD 20755
lcarlson@a.nl.cs.cmu.edu

Sergei Nirenburg

Center for Machine Translation
Carnegie Mellon University
Pittsburgh, PA 15213
sergei@nl.cs.cmu.edu

1 Why Does One Need a World Model?

Practical NLP applications requiring semantic and pragmatic analysis of texts necessitate the construction of a world model, an ontology, to support interpretation of text elements. Constraints on world model elements serve as heuristics on the cooccurrence of lexical and other meanings in the text, facilitating both natural language understanding and generation. Propositional meanings (defined in the lexicon in terms of links to the world model) trickle down to the text meaning representation as *instances* of world model entities. Our primary objective in world modeling is to support multilingual applications, so constructing a language-independent ontology is crucial. The word sense view of ontology building leads to proliferation of concepts whenever words in different languages do not “line-up” (see EDR 1990), while using a core set of “primitives” is limited for large-scale applications, if shades of meaning are to be captured. In our environment, concept acquisition is guided by examining cross-linguistic evidence and representational trade-offs. In other large-scale ontology projects, the separation of lexical from conceptual knowledge is not always clear, as in the Cyc project at MCC, a knowledge base containing millions of facts about the world (Lenat and Guha, 1990), or the KT system (Dahlgren, 1988), which classifies commonsense knowledge for English words.

In the DIONYSUS project at CMU, the world model, the lexicon and the text meaning representation are closely interconnected, in terms of their content and format. World modeling is supported by the ONTOS system, which consists of a) a *constraint language*, b) an *ontology*, or set of general concepts, c) a set of *domain models* and d) an intelligent knowledge acquisition *interface*. The basic features of the ONTOS constraint language are as follows (see Carlson & Nirenburg, 1990, for details). A **world model** is a collection of frames. A **frame** is a named set of *slots*, interpreted as an ontological concept (*voluntary-olfactory-event*, *geopolitical-entity*). A **slot** represents an ontological property (*temperature*, *caused-by*) and consists of a named set of facets. A **facet** is a named set of fillers. Facets refer to the status of property values, e.g.:

<i>value</i>	actual values of property (e.g., for concept <i>instances</i>)
<i>default</i>	typical value of a property
<i>sem</i>	set of “legal” values; akin to selectional restrictions

A **filler** is a symbol, number, range, etc. A symbolic filler (prefixed by “*”) names an ontological concept: (ALL (SUB-

CLASSES (value *property *object *event))).

2 Ontology Building in Context: Scalar Attributes

In the ONTOS system, a mechanism relating scalar attributes (AGE, TEMPERATURE) to measuring units (TEMPORAL-UNIT, THERMOMETRIC-UNIT) allows scalar information to be converted into a standard format for interlingua representation. The DOMAIN slot of a scalar attribute defines the types of concepts the attribute can describe. In the ATTRIBUTE-RANGE slot, the *sem* facet specifies an absolute constraint on the range of numerical values the attribute can have, while the *measuring-unit* facet designates a standard unit for interpreting the constraint:

```
(AGE
 (DOMAIN (sem *object))
 (ATTRIBUTE-RANGE (sem (> 0))
 (measuring-unit *second)))
```

Ontology building has never been a totally independent project in our environment. World knowledge in DIONYSUS has been acquired with the express purpose of using it in a natural language processing system. Knowledge representation requirements of such a system include, in addition to ontology specification, a representation for a lexicon entry and a language for recording the meaning of input text. The interaction between these static knowledge sources and a natural language analyzer is illustrated in Figure 1.

World modeling decisions about scalar attributes are influenced by the way the lexicon is built and vice versa. In the DIONYSUS system, relative scalar terms like *very old*, or *high/low* are not given a separate ontological status. Instead, lexical entries for such words are associated with ontologically motivated constraints on scalar attributes. The language-specific relationship between word-modifier and the language-independent relationship between concept-property is illustrated using the example of *fresh-brewed coffee*. In the ontology, the concept COFFEE appears as follows:

```
(COFFEE
 (IS-A (value *beverage))
 (AGE (sem (> 0))
 (default (<> 0 4))
 (measuring-unit *hour)))
```

The *default* facet of the AGE slot expresses a typical range of values for the age of COFFEE, which can be overridden, as

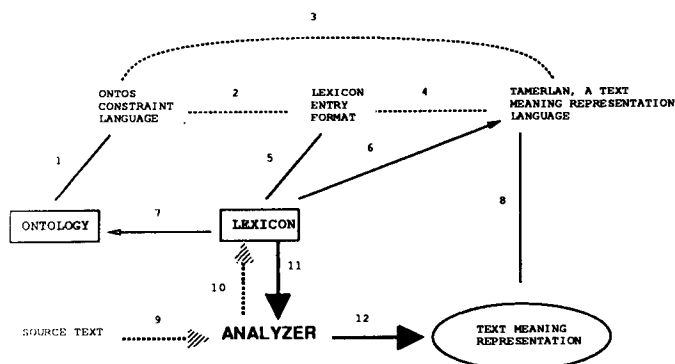


Figure 1: Interrelationship of Ontology, Lexicon, Text Meaning Representation and Processor in a Natural Language Understanding System. The numbered links in the figure are interpreted in the following manner. 1: ontological concepts are represented in the ONTOS constraint language. 2, 3, 4: portions of the ONTOS constraint language, the lexicon entry format, and TAMERLAN are shared. 5: lexical entries are represented in the lexicon entry format. 6, 7: lexical entries have pointers to information in TAMERLAN and/or information in the ontology. 8: TAMERLAN is a formal language for representing the meaning of NL texts. 9: source text is supplied to the analyzer for processing. 10, 11: part of the analysis process involves accessing and retrieving syntactic, semantic and pragmatic information stored in the lexicon. 12: the output of semantic analysis is the representation of text meaning (interlingua text), expressed in TAMERLAN.

long as the absolute constraint is not violated. The *measuring-unit* facet selects an appropriate measuring unit from the class TEMPORAL-UNIT.

In the lexicon, the link between a word sense and an ontological concept is established in the SEM zone of an entry (see Meyer et al. 1990). In the simplest case, there is a direct link, e.g., between the lexeme +coffee-n2 (the sense of *coffee*, *the beverage*), and the concept COFFEE. However, adjectives like *fresh-brewed*, *old*, etc., which represent relative information about age, are linked *indirectly* to an ontological concept, via a constraint on the default range of the AGE property for a given class of objects. For example, the SEM zone of lexical entry +fresh-brewed-n1 establishes the following linkage:

```
instance-of: ``COFFEE or TEA''
age:         (range < 0.1)
```

The range function calculates a default value for *fresh-brewed coffee*, namely, less than 10% of the ontologically specified default range for the AGE of COFFEE.

Lastly, we demonstrate informally the interdependence of lexicon, ontology and text meaning representation in DIONYSUS for the sentence *I smelled the fresh-brewed coffee*. First, we identify the predicate-argument structure, and record the syntactic pattern information in the SYN-STRUC zone of the lexical entry for *smell*:

```
smell (the sense of 'voluntary perception')
  entity that performs a voluntary
  perceptual event: I (SUBJ)
  entity that is the target of a
  voluntary perceptual event:
coffee (OBJ)
```

Next, links between word sense and ontological concept are created for the open-class lexical items in the sentence. These links are recorded in the SEM zone of the lexical entry, where a correspondence is established between semantic and syntactic roles:

```
smell →
VOLUNTARY-OLFACTORY-EVENT
AGENT: ANIMAL (^$var1)
      ;links to the SUBJ role
THEME: PHYSICAL-OBJECT (^$var2)
      ;links to the OBJ role
```

The ^\$var1 retrieves the meaning of the lexeme bound to the variable \$var1 during syntactic parsing, places it in the AGENT role of VOLUNTARY-OLFACTORY-EVENT, and checks to make sure that the ontologically specified constraint (AGENT: ANIMAL) is satisfied.

Finally, we illustrate the key components of the text meaning representation for a sentence:

```
clause
  head: voluntary-olfactory-event_1
  aspect:
    phase: end
    duration: prolonged
    iteration: single
voluntary-olfactory-event_1
  agent: speaker
  theme: coffee_1
coffee_1
  age: < 0.4 hour
```

3 Conclusions

We have addressed some of the issues that arise when world modeling is viewed as a component of knowledge support for natural language processing. In DIONYSUS, the interdependence of ontology, lexicon and text meaning representation is such that ontology acquisition never proceeds in an isolated fashion. The discussion presented here regarding scalar attributes is just one example of ontological decision making in context. We would like to thank the members of the DIONYSUS project—Ralf Brown, Ted Gibson, Todd Kaufmann, John Leavitt, Ingrid Meyer, Eric Nyberg and Boyan Onyshkevych. Thanks also to Irene Nirenburg and Ken Goodman.

References

- Carlson, Lynn and Sergei Nirenburg. 1990. World Modeling for NLP. TR CMU-CMT-90-121. Carnegie Mellon University.
- Dahlgren, Kathleen. 1988. *Naive Semantics for Natural Language Understanding*. Boston: Kluwer Academic Press.
- EDR. 1990. *Concept Dictionary*. TR 027. Japan Electronic Dictionary Research Institute.
- Lenat, Douglas and R.V. Guha. 1990. *Building Large Knowledge-Based Systems*. Reading, MA: Addison-Wesley.
- Meyer, Ingrid, Boyan Onyshkevych, and Lynn Carlson. 1990. *Lexicographic Principles and Design for Knowledge-Based Machine Translation*. TR CMU-CMT-90-118. Carnegie Mellon University.