Lexicon, Ontology and Text Meaning

Boyan A. Onyshkevych

*U.S. Department of Defense and
Computational Linguistics Program,
Carnegie Mellon University
bao@a.nl.cs.cmu.edu*

Sergei Nirenburg

*Center for Machine Translation
School of Computer Science
Carnegie Mellon University
sergei@nl.cs.cmu.edu*

Abstract

A computationally relevant theory of lexical semantics must take into consideration both the form and the content of three different static knowledge sources — the *lexicon*, the *ontological domain model* and a *text meaning representation language*. Meanings of lexical units are interpreted in terms of their mappings into the ontology and/or their contributions to the text meaning representation. We briefly describe one such theory. Its precepts have been used in DIONYSUS, a machine translation system prototype.

1 Introduction

A comprehensive natural language processing system (NLP) must be able to analyze natural language texts, and, having understood their content, generate other natural language texts, depending on the nature of the underlying reasoning system. Thus, in a machine translation system this "other" NL text will be a text in a different language but with the same meaning; in a NL front end to a database, it will be the response to a query. In order to produce an appropriate NL output, the NLP system must "understand" the input text. In some systems this understanding is modeled as an acceptable match for a pre-existing structural or literal pattern. In knowledge-based systems it is customary to model understanding through representing the results of input text analysis as a set of well-formed structures in a specially-designed formal artificial language. Such languages come in a number of varieties and we are not arguing here about which ones are more appropriate or promising. The crucial point is that in order to have any explanatory power, the atoms of this meaning representation language must be interpreted in terms of an independently motivated model of the world. Moreover, if any realistic experiments have to be performed with such an NLP system, this world model (sometimes called *ontology*) must be actually built, not only defined algebraically. The process of NL analysis is then interpreted as putting lexical, syntactic, and prosodic units of the source text in correspondence with elements of the text meaning representation language. During generation, the representation language units serve as input and the target language units, as output. (This view of language processing is, of course, highly schematic — as will be seen from the more detailed discussion which follows; but this is exactly the level of detail needed at this point.)

In our understanding, the purpose of the lexicon is to facilitate the above linking. Of course, the lexicon contains not only mappings into the ontology. It also contains information about morphological, syntactic, pragmatic, and collocational properties of the lexical unit — since all and any of these components serve as clues at various stages of the process of mapping text into representation or vice versa. But the semantics of most of the open-class lexical items is described in terms of their mappings into instances of ontological concepts.

A generic view of the data flow in a system of the above class is given in Figure 1. The experiences described in this paper relate to our work on DIONYSUS, a machine translation system. The interaction between the ontology, the text meaning representation, and the lexicon in this system was the central point of our study. This interaction takes place in two spheres — both during actual processing of text and during knowledge acquisition phase. In what follows, we first briefly describe these three knowledge sources and their interactions, as they are implemented in DIONYSUS. Next, we discuss some issues of distinguishing lexical and non-lexical knowledge and their co-existence in a comprehensive NLP application.

Justifying particular representation choices, details of coverage of specific phenomena and comparisons with other semantic and pragmatic NLP processors are beyond the scope of this paper. See Nirenburg and Defrise, forthcoming, a and b; Carlson and Nirenburg, 1990; Meyer et al., 1990; Nirenburg, 1989; and Nirenburg and Goodman, 1990 for additional details.

Each of the three knowledge sources have both content and form — that is, a special knowledge representation language has been developed for each of the knowledge sources (TAMERLAN for text meaning representation, a constraint language for representing ontology, and the structure of the lexicon entry for the lexicon). These languages are necessarily connected in many ways (and in the implementation of DIONYSUS all are based on the single underlying knowledge representation system FRAMEKIT (Nyberg, 1988). Our comments will touch upon features from both the content and the form facets of the knowledge sources.

2 Ontology

In formal semantics, one of the most widely accepted methodologies is that of model-theoretic semantics in which syntactically correct utterances in a language are given semantic interpretation in terms of truth values with respect to a certain model (in Montague semantics, a "possible world") of reality. Such models are in practice never constructed in detail but rather delineated through a typically underspecifying set of constraints. In order to build large and useful natural language processing systems one has to go further and actually commit oneself to a detailed version of a "constructed reality" (Jackendoff's term). Interpreting the meanings of textual units is really feasible only in the presence of a detailed world model; elements of this model may be linked to the various textual units (either directly or indirectly, individually or in combinations) via is-a meaning-of links. World model elements will be densely interconnected through a large set of well-defined ontological links – properties – which will enable the world modeler to build descriptions of complex objects and processes in a compositional fashion, using as few basic primitive concepts as possible.

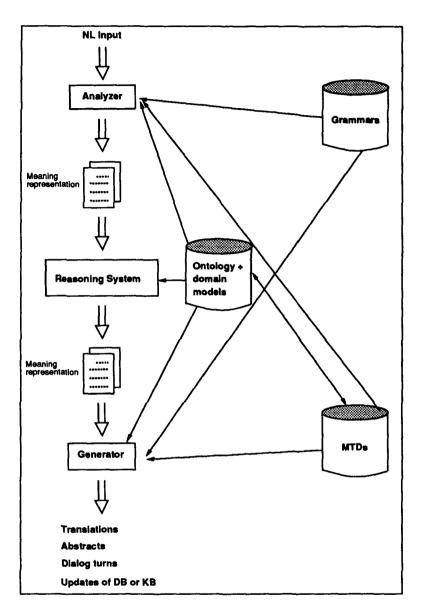Knowledge bases in FRAMEKIT take the form of a collection of frames. A frame is

Figure 1: **A Knowledge-Based NLP System.**

a named set of *slots*. A slot is a named set of *facets*. A facet is a named set of *views* and a view is a named set of *fillers*. A filler can be any symbol or a Lisp function call. The above is the basic set of constraints and features of FRAMEKIT. Though FRAMEKITactually specifies some extensions to this basic expressive power, it remains, by design, quite general and semantically underspecified. The actual interpretation and typing of its basic entities is supposed to take place in a particular application, in our case, world modeling. The ONTOS constraint language is built on top of FRAMEKIT. This strategy is deliberate and is different from many knowledge representation languages and environments, in which the basic representation language is made much more expressive at the expense of its relative unwieldiness, difficulty in learning and some format-related constraints on application development.

In the ontology module of DIONYSUS FRAMEKIT frames are used to represent *concepts*. A concept is the basic building block of our ontology. Examples of concepts are *house*, *four-wheel-drive-car*, *voluntary-olfactory-event* or *specific-gravity*. FRAMEKIT slots are interpreted as a subset of concepts called *properties*. In practice some properties recorded in concepts do not relate to the domain model but rather serve as administrative and explanatory material for the human users (see examples below). The FRAMEKIT *views* are used to mask parts of the information from a certain users. This mechanism is useful when working with very large domain models serving more than one domain so that some information will be different depending on whether a data element belongs to the domain of, say, chemistry or the domain of software engineering. FRAMEKIT fillers are constrained to be names of atomic elements of the ontology, expressions referring to modified elements of the ontology, collections of the above, or special-purpose symbols and strings used by the knowledge acquirers.

The example concepts from the ontology serve to illustrate some aspects of the constraint language used.

```
(MAKE-FRAME
   MACINTOSH
      (IS-A
         (VALUE
            (COMMON PERSONAL-COMPUTER)))
      (PRODUCED-BY
         (SEM
            (COMMON APPLE)))
      (SUBCLASSES
         (SEM
            (COMMON PLUS SE II IIFX)))
      (HAS-AS-PART
         (SEM
            (COMMON DISK-DRIVE SYSTEM-UNIT MONITOR CPU MEMORY
             EXTERNAL-DISKETTE-DRIVE INTERNAL-DISKETTE-DRIVE))))
```

Other slot values inherit from PERSONAL-COMPUTER (e.g., MAXIMUM-NUMBER-OF-USERS), or from COMPUTER, INDEPENDENT-DEVICE, DEVICE, ARTIFACT, etc. In the above example, MACINTOSH is a frame name, IS-A is a slot name, VALUE and SEM are examples of facet names, COMMON is the view name, and APPLE and MEMORY are fillers, for example.

```
(MAKE-FRAME
   VOLUNTARY-OLFACTORY-EVENT
      (IS-A
         (VALUE
```

241

```
          (COMMON VOLUNTARY-PERCEPTUAL-EVENT)))
    (AGENT
        (SEM
            (COMMON MAMMAL BIRD REPTILE AMPHIBIAN)))
    (INSTRUMENT
        (SEM
            (COMMON OLFACTORY-ORGAN))))
```

Additional slots such as EXPERIENCER and BENEFICIARY would be inherited from among the ancestors of this concept, e.g., VOLUNTARY-PERCEPTUAL-EVENT or EVENT.

3   Representation of Text Meaning

Results of source text analysis in DIONYSUS are represented in a formal, frame-based language, TAMERLAN (Nirenburg and Defrise, forthcoming). The types of knowledge which are represented in TAMERLAN include the meanings of natural language units and knowledge about the speech situation, including speaker attitudes. Meaning proper is represented through instantiating, combining, and constraining concepts available in the ontology. These processes crucially rely on lexical-semantic information in the appropriate lexicon entries. This information relates not only to ambiguity resolution in "regular" compositional semantics (i.e., straightforward monotonic dependency structure building), but also the identification of idioms, treatment of metonymy and metaphor, and resolution of reference ambiguities.

It is well known, however, that the intent of a text is typically not capturable by just instantiating ontological concepts; what is additionally needed is the representation of pragmatic and discourse-related aspects of language — speech acts, deictic references, speaker attitudes and intentions, relations among text units, the prior context, the physical context, etc. As most of the knowledge underlying realizations of these phenomena is not society-general but is rather dependent on a particular cognitive agent (a particular speaker/hearer in a particular speech situation), the pragmatic and discourse knowledge units were not included in the ontology (which is supposed to reflect, with some variability, a societal view of the world — indeed, if every speaker/hearer had a totally idiosyncratic world view, communication would not be possible). The representation of this "meta-ontological" information is thus added to the representation of meaning proper to yield a representation of text meaning; a resulting representation thus reflects both lexically-triggered ontologically-based semantic information and non-ontological information reflecting pragmatic and discourse factors.

The propositional part of text meaning is obtained by instantiating relevant ontological concepts, accessed through corresponding lexicon entries, which list mappings between word senses and ontological concepts. In order to carry out the mappings of complex structures with dependencies, information about argument structure of lexical units (also stored in the lexicon) is used.

However, many lexical units in the source text contribute to the nonpropositional meaning of the text. The structures for the nonpropositional components of text meaning are also derived from the lexicon, where appropriate patterns are stored (see example lexicon entries below). The nonpropositional information is not stored in the ontology for reasons explained above. In what follows we motivate the most important non-ontological components of our text meaning representation formalism (for more detailed discussion

see Nirenburg and Defrise, forthcoming, a) — speaker attitudes, stylistic features and rhetorical relations.

A critical facet of capturing the intent of a speaker in a meaning representation is rendering of the attitudes that the speaker holds toward objects or situations represented in the propositional (compositional-semantic) component of text meaning representation. Indeed, the speaker may convey attitudes about producing the text in the first place (the speech act), elements of the speech context, or even other attitudes. An attitude is represented by a structure which includes the type of attitude (from the list below), a value (a decimal value taken from the interval [0, 1] indicating the polarity or strength of the attitude), an indication of to whom the attitude can be attributed (typically the speaker), a scope (identifying the entity towards which the attitude is expressed), and a time (representing the absolute time at which the attitude was held). Types of attitudes recognized in TAMERLAN are: epistemic, evaluative, deontic, expectation, volition and saliency.

The stylistic overtones of a lexical entry, even if not contributing directly to the compositional semantics of a text, serve a crucial role in conveying the speaker attitudes and in achieving his intent in producing the text. Thus we identify that the stylistics of a lexeme needs to be encoded in the lexicon entry in addition to the lexical semantic information. In encoding lexicons for languages with rich social deictics, such as Japanese, the issue of stylistics becomes even more acute.

We are utilizing a set of style indicators which is a modification of the set of pragmatic goals from Hovy 1988. This set consists of six stylistic indicators: formality, simplicity, color, force, directness, and respect. In our implementation, we need to label lexical entries (including idioms, collocations, conventional utterances, etc.) with appropriate representations of values for these stylistic indicators; values for these factors are represented as points on the interval [0,1], where 0.0 is low, 1.0 is high, and 0.5 represents a default, neutral value. In NL generation, these factors are used in lexical selection; in NL analysis, the values are available for assisting in disambiguation (relying on expected values for the factors and utilizing the heuristic that typically the stylistics will be consistent across words in an utterance).

Some examples of English lexical entries that would need to include style features are:

```
upside:        formality - low
               color - high
delicious:     formality - somewhat high
               color - high
drop by:       formality - low
great:         formality - low
one:           personal-reference - low (pronominal sense)
               formality - high
               force - low
```

Relations in TAMERLAN are a mixed bag – some of them refer to real-world connections and are, therefore, defined in the ontology while some others (such as rhetorical relations, e.g., conjunction and contrast) refer to properties of text itself and are, therefore, not a part of the world model. At present, the major classes of relations in TAMERLAN include causal, conjunction, alternation, coreference, temporal, spatial, textual, reformulation and domain-text relations.

243

4 The Lexicon

An entry in the DIONYSUS lexicon is comprised of a number of zones (each possibly having multiple fields), integrating various levels of lexical information. The zones are CAT (syntactic category), ORTH (orthography – abbreviations and variants), PHON (phonology), MORPH (morphological irregular forms or class information), SYN (syntactic features such as attributive), SYN-STRUC (indication of sentence- or phrase-level syntactic inter-dependencies such as subcategorization), SEM (lexical semantics / meaning representation), LEXICAL-RELATIONS (collocations, etc.), PRAGM (pragmatics hooks for deictics, for example, and stylistic factors), and STUFF (user information such as modification audit trail, example sentences, etc.)

In this paper we concentrate on the SEM zone of the lexicon, since it is the locus of lexical-semantic analysis. This zone contains two fields — MEAN-PROC and LEX-MAP.

The MEAN-PROC field invokes procedural semantic functions. One such function is "intensify-range" used to realize the meaning of intensifiers such as *very*. This function operates on scalar attribute ranges as input, returning compressions of the input ranges. For example, the English word *old* may involve a range (> 0.8) of the AGE attribute. The intensify-range function would be invoked in the phrase *very old*, for example, returning (> 0.9). Similarly, for *very young* this function would return (< 0.1), when given (< 0.2). For *very average* (perhaps ($0.44 < x < 0.6$)), this function would return ($0.45 < x < 0.55$).

The LEX-MAP field contains mappings indicating the links between lexical units and ontological concepts. The simplest version of such a mapping is the straightforward link between a concept in the ontology and the meaning of the lexical unit in question. In analysis the link identifies which ontological concept to instantiate as the representation of a lexical unit. So, for example, the lexical semantics for the primary sense of the English word *dog* would be a pointer or link to the concept of "dog" in the ontology: (marked, by convention, *dog). In DIONYSUS, the semantic analyzer instantiates an instance of the dog concept (e.g., *%dog34*) when processing a sentence containing the word "dog" or a reference to it. This mechanism works only in the case of one-to-one mapping between words and ontological concepts, so that the ontology becomes essentially a collection of word senses.

There is no consensus in the field with respect to size and composition of the ontological world model. The "word-sense" view of the ontology, in addition to the obvious disadvantage of ontology size, leads to a problems in multilingual applications — often roughly comparable words in different languages do not "line up" the same way, which leads to proliferation of new concepts with every new language. This makes the translation process difficult. Another well-known approach prefers a small restricted set of primitives which are combined or constrained in an attempt to render the meaning of any lexical unit. This decision has been amply criticized as it leads to difficulty in building large-scale world models and capturing shades of meaning; additionally, this approach can yield enormous, unmaintainable lexicon entries for complex concepts. Further discussion of this issue (and pointers to the literature on this subject) can be found in Meyer et al., 1990.

The approach taken in DIONYSUS lies somewhere in between the two extremist positions outlined above. Therefore, in some cases, it is possible to find one-to-one correspondences between lexeme and concept. Unfortunately, in the majority of cases, this is not possible without excessive proliferation of concepts. Therefore, mappings are allowed to concepts with "similar" but not identical meanings – to the most specific concept that is still more

general than (i.e., that subsumes) the meaning of the lexeme in question. Once the most directly corresponding concept is determined, additional information is specified (thus constraining the general concept) in the LEX-MAP field in the lexicon. This modified mapping may either add information to the concept as it is specified in the ontology, override certain constraints (e.g., a selectional restriction), or indicate relationships with other concepts expected in the sentence.

Before proceeding to illustrate the constrained mappings, we first introduce the knowledge representation tools used to express them. In the DIONYSUS formalism, ontological concepts are interpreted as frames, and their properties, as slots (see the discussion of FRAMEKITabove). It is natural, then, to use the *facet* mechanism to encode constraints introduced in the LEX-MAP mappings. The constraining or "specialization" facets used in the DIONYSUS lexicon representations are as follows:

1. VALUE - a specific value (e.g., number of sides for a triangle = 3, sex of a man = male). This is the facet where actual information is represented. When this meaning pattern is instantiated by use in a sentence, semantic dependency structure links will be indicated typically in a VALUE facet.

2. DEFAULT - usual (i.e., typical, expected) value (e.g., color of diapers = white).

3. SEM - akin to a traditional selectional restriction (e.g., color of diapers has to be a COLOR), this is essentially a constraint on what VALUE may be.

4. RELAXABLE-TO - maximum relaxability, if any, of SEM restrictions.

5. SALIENCE - a scalar value in the range [0.0, 1.0] designating the significance of a specific attribute slot or role (partly reflecting the notion of "defining properties" vs. "incidental properties").

The structure below, for the lexeme "eat-v1" (the primary transitive verb sense) illustrates a simple case of lexical semantic mapping. The SYN-STRUC zone of the lexical entry establishes the subcategorization of the verb, and the structure is used in parsing or generation. In the course of parsing, the variables $var1 and $var2 are bound to a "placeholder" for the lexical semantics of the subject and object of the verb, respectively. Once the lexical semantics of the syntactic constituents in those syntactic roles is determined, it becomes available for the compositional purposes — to build a semantic representation for a higher-level text component (e.g., a sentence).

```
(%ingest
    (AGENT  (value ~$var1)   ; subject of lexeme will get bound to
            (sem *animal))   ; $var1, which corresponds to
                             ; AGENT slot of the ontological concept
                             ; INGEST
    (THEME  (value ~$var2)   ; object of lexeme will get bound to $var2,
            (sem *ingestible) ; which must correspond to the THEME slot
                             ; of INGEST
            (relaxable-to *physical-object)))) ; indication of
                             ;extent of allowable semantic constraint
                             ; relaxation
```

As an illustration of the process of semantic dependency structure building, consider the analysis of the sentence *The man visited the store.* This analysis will involve the establishing (among others) of TAMERLAN entities *%visit132* and *%human135* (the numbers are

245

added for unique reference, the percent sign marks an instance of an ontological concept), which are instances of the ontological concepts *visit and *human. This is made possible through the content of the LEX-MAP fields of lexicon entries for (the corresponding senses of) the English words "visit" and "man." The semantics of the entire sentence will include the instantiated concepts, connected by using %human135 to fill the AGENT slot in %visit132. The knowledge necessary for establishing this dependency is found in the syntactic component of the lexical entry for "visit" (where the $var variables would be shown as corresponding to subject and object positions); this information is then linked to the semantic sphere using the following notation in the LEX-MAP field of the appropriate sense of "visit:"

```
(%visit (AGENT (VALUE ^$var1))...)
```

The caret is an operator (akin to an intension operator) which retrieves the meaning of the syntactic constituent bound to the variable. The resulting TAMERLAN structure in will, thus, include the following:

```
(%visit132 (AGENT (VALUE %human135)) ...)
```

At this point we are ready to illustrate constrained mapping between word senses and ontological concepts. An example of constraining the sense of an ontological concept is the lexeme taxi-v1, identifying the sense of taxi as an instance of the ontological concept *move-on-surface, but with important further constraints — the theme can be only an aircraft, (e.g., The plane taxied to the terminal). The LEX-MAP field for taxi-v1 is, then, as follows:

```
(%move-on-surface
        (SURFACE (SEM *water *ground))
        (THEME (SEM *aircraft)
                (RELAXABLE-TO *vehicle)))))
```

In this example the thing that is moving-on-the-surface is being constrained to be an aircraft, possibly relaxable to include any vehicle (perhaps in more florid speech). Note that there may be second-order constraints (i.e., constraints on constraints) in addition to the above: jet-v1 (the literal sense of 'to travel by jet', e.g., The presidential candidate spent most of the year jetting across the country from one campaign rally to another.)

```
(%move
    (INSTRUMENT (SEM *aircraft
                    (PROPELLED-BY
                        (value %jet-engine)))))))
```

This constrains the vehicle used in jetting to be a jet-propelled vehicle.

There does not need to be a regular correspondence between the syntactic category (part of speech) of the lexeme and the general high-level partition into which the corresponding ontological concept falls. The highest node in the ontology, ALL, is partitioned into the subclasses EVENT, OBJECT and PROPERTY (the latter being partitioned into ATTRIBUTE and RELATION), which one might expect to correspond to verbs, nouns, and adjectives or prepositions, respectively. However, there is a great deal of variability in the correspondence between ontological partition and linguistic part of speech. See Meyer et al., 1990 for further discussion of this issue.

In some cases, the constraint of an ontological concept may be achieved through the use of scalar attributes: SHOP ==> STORE of a certain size; CHILD ==> HUMAN of a certain age; FRESH-BREWED ==> modifies COFFEE or TEA of a certain age. The RANGE facet is used to access particular information about scalar attributes in the ontology, such as AGE, TEMPERATURE, LUMINANCE, and SIZE. Any attribute which can be measured on a scale can also be described in relative terms. For example, we can say *That house is 4000 square feet* or *That house is very large*; *That water is 200 degrees (F)* or *That water is very hot*. By means of the RANGE facet, we can establish a correspondence between measured and relative values for various ontological concepts.

5 Integration of Multiple Knowledge Sources in Lexicon Entries

As is evident from the brief discussion above, the lexicon can be seen as the nexus through which a number of knowledge sources are integrated in order to produce an appropriate representation of the "meaning" of the input. The clearest case of this is the triggering of instantiation of ontological concepts by a "meaning template" in the LEX-MAP field of the lexicon entry. Furthermore, the additional information (augmenting and constraining information) presented in the LEX-MAP field provides a mechanism of tailoring the cross-linguistic ontological (world model) information to language-specific, lexeme-specific, representation of a conceptual notion.

In some lexicon entries LEX-MAP contains a mapping not into an ontological concept but rather into the representation of a pragmatic or discourse entity, such as a relation or an attitude. In other cases, lexical units have mixed ontological and non-ontological meaning components. In both cases, this is a mechanism for providing for the integration of non-ontological (hence "non-semantic") information with ontological, as triggered by lexical entries. The following examples will illustrate such phenomena.

The following is an example of how an evaluative attitude can be used for describing the meaning of a lexeme:

```
excellent-adj1
(ATTITUDE
        (type (value evaluative))
        (attitude-value (value 1))
        (scope ^$var1)) ; $var1 corresponds to the object for which the
                        ; attitude is held, as identified syntactically
        (attributed-to (value [$SS].deictic-indices.speaker))
                        ; the holder of the attitude is the speaker (i.e., the
                        ; producer of the text)
```

The entry for smell-v5 (defined as perceive something negative intuitively, as in *I could smell trouble brewing* or *Everyone could smell the hostility in the air*) demonstrates the integration of ontological and non-ontological structures:

```
(%perceptual-event
        (EXPERIENCER (value ^$var1) ; link to the syntactic subject
                     (sem *human))
        (THEME (value ^$var2)  ; link to the syntactic object
               (sem *mental-object)))
(ATTITUDE
        (type (value evaluative))
```

```
(value (value (< 0.2))) ; less than .5, so negative feelings
(scope (%perceptual-event.theme)) ;link to the contents of
                        ; the indicated slot in the ontological
                        ; concept %perceptual-event referenced above.
(attributed-to
        (value [$SS].deictic-indices.speaker))))) ; a pointer
                        ; to the deictic index in the context frame
```

Here there is a restriction on the EXPERIENCER, who must be a human, and on the THEME, which must be a MENTAL-OBJECT. A further item of information, however, is that there is a negative attitude towards the theme (e.g., one does not normally say things like *I could smell happiness in the air, Everyone could smell the positive attitude she had,* etc.) The attribution of the attitude is a hook into the deictic indices of the speech situation of the utterance – the formula in the attribution identifies that the speaker of the current speech situation be identified, and that the speaker be linked in as the holder of attitude.

The semantic mapping mechanism as a representation language in the LEX-MAP field of the lexicon (outlined above) allows for a fluent integration of encyclopedic knowledge into the framework without "corrupting" the linguistic lexicon with this type of information. The approach in question is the formation of an "episodic memory" knowledge source, consisting of a set of data structures identical to lexicon entries. However, the person Ronald Reagan would be represented in the episodic memory as a named instance of the concept **human,* or some more specific concept, along with other information known about him rendered in various slots of the concept instantiation, not by a trigger to produce a new instantion of the concept as would be the case by a linguistic lexicon entry.

6   Status and Future Work

All three static knowledge sources described briefly above are still under development. In fact, only when the ontology and the lexicon reach nontrivial size, can one put a computationally relevant lexical-semantic theory to a real test in a large-scale application. A complete theory of lexical semantics must, of course, include the description of the actual processes which lead from a natural language text to its meaning representation (NL analysis) and in the opposite direction (NL generation). We continue to develop these processes, in the tradition of knowledge-based NLP.

Acknowledgements

References

Carlson, Lynn and Sergei Nirenburg. 1990. World Modeling for NLP. Technical Report CMU-CMT-90-121. Center for Machine Translation. Carnegie Mellon University.
Goodman, Kenneth and Sergei Nirenburg (eds.) 1989. KBMT-89. CMU-CMT Project Report.

Hirst, Graeme. 1986. Semantic Interpretation and the Resolution of Ambiguity. Cambridge: Cambridge University Press.

Hirst, Graeme. 1989. Ontological Assumptions in Knowledge Representation. In Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning (Toronto, May 1989).

Hovy, Edward. 1988. Generating Natural Language Under Pragmatic Constraints. Ph.D. Yale University.

Meyer, Ingrid and James Steele. 1990. The Presentation of an Entry and of a Super-entry in an Explanatory Combinatorial Dictionary. In James Steele (ed.), The Meaning-Text Theory of Language: Linguistics, Lexicography, and Practical Implications. Ottawa, Canada: University of Ottawa Press.

Meyer, Ingrid, Boyan Onyshkevych, and Lynn Carlson. 1990. Lexicographic Principles and Design for Knowledge-Based Machine Translation. Technical Report CMU-CMT-90-118. Center for Machine Translation. Carnegie Mellon University.

Monarch, I. and S. Nirenburg. 1988. ONTOS: An Ontology-Based Knowledge Acquisition and Maintenance System. Proceedings of the Second Workshop on Knowledge Acquisition. Banff, Canada.

Nirenburg, Sergei, and Christine Defrise (forthcoming, a). Practical Computational Linguistics. In R. Johnson and M. Rosner (eds.), Computational Linguistics and Formal Semantics. Cambridge: Cambridge University Press.

Nirenburg, Sergei, and Christine Defrise (forthcoming, b). Lexical and Conceptual Structure for Knowledge-Based Machine Translation. In: J. Pustejovsky (ed.), Semantics and the Lexicon. Dordrecht, Holland: Kluwer.

Nirenburg, Sergei and Kenneth Goodman. 1990. Treatment of Meaning in MT Systems. Proceedings of the Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language. Linguistic Research Center, University of Texas at Austin. June 1990.

Nyberg, Eric. 1988. The FrameKit User's Guide, Version 2.0. CMU-CMT-MEMO. Center for Machine Translation. Carnegie Mellon University.