FrameNet meets the Semantic Web: Lexical Semantics for the Web

SriN Narayanan, Collin Baker, Charles Fillmore, and Miriam Petruck

International Computer Science Institute (ICSI),
1947 Center Street, Berkeley, CA 94704, USA,
smarayan@icsi.berkeley.edu,
www home page http://www.icsi.berkeley.edu/~framenet/

Abstract. This paper describes FrameNet [9,1,3], an online lexical resource for English based on the principles of frame semantics [5,7,2]. We provide a data category specification for frame semantics and FrameNet annotations in an RDF-based language. More specifically, we provide an RDF markup for lexical units, defined as a relation between a lemma and a semantic frame, and frame-to-frame relations, namely Inheritance and Subframes. The paper includes simple examples of FrameNet annotated sentences in an XML/RDF format that references the project-specific data category specification.

Frame Semantics and the FrameNet Project

FrameNet’s goal is to provide, for a significant portion of the vocabulary of contemporary English, a body of semantically and syntactically annotated sentences from which reliable information can be reported on the valences or combinatorial possibilities of each item included.

A semantic frame is a script-like structure of inferences, which are linked to the meanings of linguistic units (lexical items). Each frame identifies a set of frame elements (FEs), which are frame-specific semantic roles (participants, props, phases of a state of affairs). Our description of each lexical item identifies the frames which underlie a given meaning and the ways in which the FEs are realized in structures headed by the word. The FrameNet database documents the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its senses, through manual annotation of example sentences and automatic summarization of the resulting annotations. FrameNet I focused on governors, meaning that for the most part, annotation was done in respect to verbs; in FrameNet II, we have been annotating in respect to governed words as well. The FrameNet database is available in XML, and can be displayed and queried via the web and other interfaces. FrameNet data has also been translated into the DAML+OIL extension to XML and the Resource Description Framework (RDF). This paper will explain the theory behind FrameNet, briefly discuss the annotation process, and then describe how the FrameNet data can
be represented in RDF, using DAML+OIL, so that researchers on the semantic web can use the data.

Frame Semantic Background

In Frame Semantics [4,6,2,11], a linguistic unit, in our case, a word (in just one of its senses), evokes a particular frame and may profile some element or aspect of that frame.\footnote{The term profile (used here as a verb) is borrowed from [8], esp. pp. 183ff.} An “evoked” frame is the structure of knowledge required for the understanding of a given lexical or phrasal item; a “profiled” entity is the component of a frame that integrates directly into the semantic structure of the surrounding text or sentence. The frames in question can be simple – small static scenes or states of affairs, simple patterns of contrast, relations between entities and the roles they serve, or possibly quite complex event types that provide the background for words that profile one or more of their phases or participants.

For example, the word bartender evokes a scene of service in a setting where alcoholic beverages are consumed, and profiles the person whose role is to prepare and serve these beverages. In a sentence like The bartender asked for my ID, it is the individual who occupies that role that we understand as making the request, and the request for identification is understood against the set of assumptions and practices of that frame.

Replacement: An Example Frame A schematic description of the replacement frame will include an agent effecting a change in the relationship between a place (which can be a role, a function, a location, a job, a status, etc.) and a theme. For example, in the sentence Sal replaced his cap on his bald head, Sal fills the role of agent, his cap instantiates the fe theme, and on his bald head is the place. The words defined in terms of this frame include, among others, exchange.v, interchange.v, replace.v, replacement.n, substitute.v, substitution.n, succeed.v, supplant.v, swap.v, switch.v, trade.v.

The replacement frame involves states of affairs and transitions between them such that other situations are covered: an “old theme”, which we refer to as old, starts out at the place and ends up not at the place, while a “new theme”, which we call new, starts out not at the place and ends up at the place (as in Factory owners replaced workers by machines).

Syntactically, the role of agent can be expressed by a simple NP (as in Margot switched her gase to the floor, a conjoined NP (as in Margot and her admirer exchanged glances), or two separate constituents, an NP and a PP (as in Margot exchanged glances with her admirer). Similarly, place may be expressed as one PP or two. Compare Ginny switched the phone between hands and Ginny switched the phone from one hand to the other. And, if old and new are of the same type, they are expressed as a single FE (as in The photographer switched lenses).
The FrameNet Process

Using attested instances of contemporary English, FrameNet documents the manner in which frame elements (for given words in given meanings) are grammatically instantiated in English sentences and organizes and exhibits the results of such findings in a systematic way. For example, in causative uses of the words, an expression about replacing NP with NP takes the direct object as the OLD and the oblique object as the NEW, whereas substituting NP for NP does it the other way around. A commitment to basing such generalizations on attestations from a large corpus, however, has revealed that in both UK and US English, the verb substitute also participates in the valence pattern found with replace, i.e. we find examples of substituting the OLD with the NEW.

In their daily work, FrameNet lexicographers record the variety of combinatorial patterns found in the corpus for each word in the FrameNet lexicon, present the results as the valences of the words, create software capable of deriving as much other information about the words as possible, from the annotations, and add manually only that information which cannot—or cannot easily—be derived automatically from the corpus or from the set of annotated examples.

FrameNet has been using the British National Corpus, more than 100,000,000 running words of contemporary British English. In the current phase, we have begun to incorporate into our work the North American newswire corpora from the Linguistic Data Consortium (http://www.ldc.upenn.edu), and eventually we hope to be able to add the full resources of the American National Corpus (http://www.cs.vassar.edu/ide/anc/).

Frame-to-Frame Relations

The FrameNet database records information about several different kinds of semantic relations, consisting mostly of frame-to-frame relations which indicate semantic relationships between collections of concepts. The two that we consider here are inheritance and subframes.

Inheritance Frame inheritance is a relationship by which a single frame can be seen as an elaboration of one or more other parent frames, with bindings between the inherited semantic roles. In such cases, all of the frame elements, subframes, and semantic types of the parent have equally or more specific correspondents in the child frame. Consider for example, the change_of_leadership frame, which characterizes the appointment of a new leader or removal from office of an old one, and whose FEs include: selector, the being or entity that brings about the change in leadership (in the case of a democratic process, the electorate); OLD leader, the person removed from office; OLD order, the political order that existed before

---

2 Our use of the BNC is by courtesy of Oxford University Press, through Timothy Benbow. The version of the corpus we use was tokenized at Oxford, lemmatized and POS-tagged at the Institut für Maschinelle Sprachverarbeitung at the University of Stuttgart. Information about the BNC can be found at http://info.ox.ac.uk/bnc.
the change; NEW LEADER, the person appointed to office; and ROLE, the
position occupied by the new or old leader. Some of the words that belong
to this frame describe the successful removal from office of a leader (e.g.
overthrow, oust, depose), others only the attempt (e.g. uprising, rebellion).
This frame inherits from the more abstract REPLACEMENT frame described
above, with the following FEs further specified in the child: OLD and NEW
are narrowed to humans beings or political entities, i.e. OLD_LEADER and
NEW_LEADER, respectively; and PLACE is an (abstract) position of political
power, i.e. ROLE.

Subframes The other type of relation between frames which is currently rep-
resented in the FN database is between a complex frame and several simpler
frames which constitute it. We call this relationship (subframes). In such
cases, frame elements of the complex frame may be identified (mapped) to
the frame elements of the subparts, although not all frame elements of one
need have any relation to the other. Also, the ordering and other temporal
relationships of the subframes can be specified using binary precedence rela-
tions. To illustrate, consider the complex CRIMINAL_PROCESS frame, defined
as follows: A Suspect is arrested by an AUTHORITY on certain CHARGES,
then is arraigned as a DEFENDANT. If at any time the DEFENDANT pleads
guilty, then the DEFENDANT is sentenced, otherwise the DEFENDANT first
goes to trial. If the VERDICT after the trial is guilty, then the DEFENDANT
is sentenced. In the end, the DEFENDANT is either released or is given a
SENTENCE by a JUDGE at the sentencing. For each step in the process,
there is a separate frame in the database, including ARREST, ARRAI-
GNMENT, TRIAL, SENTENCING, and so on. Each of these frames is related to
the CRIMINAL_PROCESS frame via the SubFrame relation in the frame ed-
tor. Moreover, subframes (of the same complex frame) are related to each
other through their ordering.

We have recognized the need to deal with other types of relations among
frames, and, so far, have identified two, SeeAlso, and Using, but have not yet
represented them in our database.

The FrameNet Product

The FrameNet database contains descriptions of more than 7,000 lexical units
based on more than 130,000 annotated sentences. This information is available
for a wide range of natural language processing applications, including question
answering, machine translation, and information extraction.

The FrameNet database differs from existing lexical resources in the speci-
ficity of the frames and semantic roles it defines, the information it provides
about relations between frames, and the degree of detail with which it describes,
for each lexical item it covers, the possible syntactic realizations of semantic
roles.

While Ide, et al., (2000) offers a representation scheme for dictionaries and
other lexical data, the kind of information in the FrameNet database is not ex-
pressed in the same level of depth in any existing print dictionary or computational lexical resource. For instance, while WordNet describes semantic relations between words, it does not recognize conceptual schemas, i.e. frames, that mediate in these relations, and therefore does not have the means to link arguments of predicing words with the semantic roles they express. FrameNet also differs from WordNet in showing semantic relations across parts of speech, and in providing contextual information enriched with semantics (beyond the “Someone _s something” format of WordNet argument-structure representations).

The FN database can be seen both as a dictionary and a thesaurus. As a dictionary, each lexical unit (lemma in a given sense) is provided with (1) the name of the frame it belongs to and access to a description of the frame, (2) a definition (either original or from the Concise Oxford Dictionary, courtesy of Oxford University Press), (3) a valence description which summarizes the attested combinatorial possibilities in respect to both semantic roles and the syntactic form and function of the phrases that instantiate those roles (generated as a report from the database), and (4) access to annotated examples illustrating each syntactic pattern found in the corpus and the kinds of semantic information they contain.

The FN database can also be seen as a thesaurus by noting that lemmas are linked to the semantic frames in which they participate, and frames, in turn, are linked both to the full set of words which instantiate them and to related frames.

1 Criminal Process: A Full Example of Annotation and Representation

![Diagram](image-url)

**Fig. 1.** The Criminal Process frame and some of its subframes
The FrameNet project is currently working on a very complex background frame called CRIMINAL,PROCESS, which contains many subframes, as part of our analysis of texts in the crime domain.

The CRIMINAL,PROCESS frame is defined as follows: a SUSPECT is arrested by an AUTHORITY on certain CHARGES, then is arraigned as a DEFENDANT. If at any time the DEFENDANT pleads guilty, then the DEFENDANT is sentenced, otherwise the DEFENDANT first goes to trial. If the VERDICT after the trial is guilty, then the DEFENDANT is sentenced. In the end, the DEFENDANT is either released or is given a SENTENCE by a JUDGE at the sentencing.

In addition to having a fairly large number of FEs, the CRIMINAL,PROCESS frame, as its name suggests, represents a complex process with many parts. These are treated in FrameNet as subframes. Fig. 1 shows part of the subframe structure. The stages such as arrest, arraignment, and trial are subframes of CRIMINAL,PROCESS and some of them (such as ARRAIGNMENT) have subframes of their own. The FE-to-FE links are not shown, but it should be clear that the person referred to as the SUSPECT in the Arrest frame is identified with the DEFENDANT in the TRIAL frame, that (if justice is done) this will also be the OFFENDER in the OFFENSE frame, etc.

One step in the process depicted here is the arrest of a suspect. We have a frame called ARREST, which includes the verb apprehend. So part of the work of FrameNet on the ARREST frame will be to collect corpus sentences containing apprehend and annotate them with FEs such as AUTHORITIES (those with the power to arrest), SUSPECT (the person(s) who are arrested), CHARGES (the name of the offense which the Suspect is to be charged with), and more general FEs such as PLACE and TIME.

![Fig. 2. Annotation of a sentence in the ARREST frame.](image)

Fig. 2 shows the FrameNet annotation software in use. The user has selected the sentence "In July last year, a German border guard apprehended two Irishmen with Kalashnikov assault rifles, and has marked the FEs TIME, AUTHORITIES,
A Data Category Specification for Frame Semantics in RDF

The World Wide Web (WWW) contains a large amount of information which is expanding at a rapid rate. Most of that information is currently being represented using the Hypertext Markup Language (HTML), which is designed to allow web developers to display information in a way that is accessible to humans for viewing via web browsers. While HTML allows us to visualize the information on the web, it doesn’t provide much capability to describe the information in ways that facilitate the use of software programs to find or interpret it. The World Wide Web Consortium (W3C) has developed the Extensible Markup Language (XML) which allows information to be more accurately described using tags. As an example, the word crawl on a web site might represent an offline search process (as in web crawling) or an exposition of a type of animate motion. The use of XML to provide metadata markup, such as for crawl, makes the meaning of the word unambiguous. However, XML has a limited capability to describe the relationships (schemas or ontologies) with respect to objects. The use of ontologies provides a very powerful way to describe objects and their relationships to other objects. The DAML language was developed as an extension to XML and the Resource Description Framework (RDF). The latest release of the language (DAML+OIL) (http://www.daml.org) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable.

Framenet-1 has been translated into DAML+OIL. We developed an automatic translator from FrameNet to DAML+OIL which is being updated to reflect FrameNet2 data. With periodic updates as the FrameNet data increases, we expect it to become useful for various applications on the Semantic Web. DAML+OIL is written in RDF (http://www.w3.org/TR/daml+owl-walkthru/#RDF1), i.e., DAML+OIL markup is a specific kind of RDF markup. RDF, in turn, is written in XML, using XML Namespaces (http://www.w3.org/TR/daml+owl-walkthru/#XMLNS), and URIs. Thus, our framenet declaration begins with an RDF start tag including several namespace declarations of the form:

```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE rdf [ 
  <!ENTITY rdfs "http://www.w3.org/1999/02/22-rdf-syntax-ns"> 
  <!ENTITY xsd "http://www.w3.org/2000/01/XMLSchema"> 
  <!ENTITY daml "http://www.daml.org/2001/03/daml+owl"> 
]>
```
<http://www.w3.org/1999/02/22-rdf-syntax-ns#>. This is a conventional RDF declaration appearing verbatim at the beginning of almost every RDF document. The second and third declarations make similar statements about the RDF Schema and XML Schema datatype namespaces. The fourth declaration says that in this document, elements prefixed with @daml: should be understood as referring to things drawn from the namespace called http://www.w3.org/2001/03/daml+owl#. This again is a conventional DAML+OIL declaration. We use the XML entity model to use shortcuts with referring to the URIs. The other DAML+OIL ontologies used in the FrameNet description include the DAML-S (http://www.daml.org/services) service ontologies, the OpenCYC DAML ontology (http://www.cyc.com/2002/04/04/cyc.daml), and the SRI time ontology (http://www.ai.sri.com/daml/ontologies/sr-basis/1-0/Time.daml) which is currently being revised with the new DAML+OIL time ontology effort. http://www.icsi.berkeley.edu/snarayan/frame-2.daml has a complete namespace and imported ontology list.

The most general object of interest is a frame. We define the FRAME class as a @daml:Class We then define a bunch of bookkeeping properties on the FRAME class. An example of the name property is shown below.

```xml
<daml:Class rdf:ID="Frame">
  <rdfs:comment>The most general class</rdfs:comment>
</daml:Class>
```

In FrameNet, the basic relation between a word (Lemma) and a frame is the Lexical Unit (LU). The domain of the Lexical Unit is a Lemma or word and its range is a Frame. An LU is defined in DAML as a property.

```xml
<daml:ObjectProperty rdf:ID="LU">
  <rdfs:domain rdf:resource="#Frame"/>
  <rdfs:range rdf:resource="#Lexeme"/>
</daml:ObjectProperty>
```

Roles are relations defined on frames ranging over the specific type of the filler. We use @daml:objectProperty to define the roles of a frame. The domain of

Note that all URIs are globally scoped, so without this the entire path has to be specified.
a role is its frame. We leave the type of the filler unrestricted at this level, allowing specific roles to specialize this further. Note that we use the `daml:isPropertyAs` relation to specify synonyms. The fragment below specifies that Frame Element, Role, and FE are synonyms.

```
<owl:ObjectProperty rdf:ID= "role">
  <rdfs:domain rdfs:resource="#Frame"/>
  <rdfs:range rdfs:resource="#daml:Thing"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID= "frameElement">
  <owl:samePropertyAs rdfs:resource="#role"/>
</owl:ObjectProperty>

<owl:ObjectProperty rdf:ID= "FE">
  <owl:samePropertyAs rdfs:resource="#role"/>
</owl:ObjectProperty>
```

We use the various constructs `daml:maxCardinality`, `daml:minCardinality`, `daml:cardinalityQ`, etc. from DAML to specify cardinality restrictions on the fillers of a role property. The markup fragment below shows the specification of a single valued role.

```
<owl:ObjectProperty rdf:ID= "singleValueRole">
  <rdfs:range rdfs:resource="#Frame"/>
  <rdfs:subClassOf>
    <owl:Restriction owl:maxCardinality="1"/>
    <owl:onProperty rdfs:resource="#role"/>
  </owl:Restriction>
</owl:Class>
```

The relation between frames (such as ARREST) and CRIMINAL PROCESS is often captured by a set of bindings between frame elements (such as the arrested person is the same individual as the person charged who is the same individual as the defendant in a criminal process).

By far the most important binding relation is the identification of roles (i.e. they refer to the same value (object)). This can be specified through the relation `identify` which is a `subProperty` of `bindingRelation`. Note that in order to do this, we have to extend the DAML+OIL language which does not allow properties to be defined over other properties. We use the DAML-S ontology (http://www.daml.org/services/daml-s/0.7/) primitive `daml:s:sameValues` to specify the `identify` relations.

```
<owl:ObjectProperty rdf:ID= "identify">
  <rdfs:subPropertyOf rdfs:resource="&daml-s;sameValues"/>
  <rdfs:domain rdfs:resource="#role"/>
  <rdfs:range rdfs:resource="#role"/>
</owl:ObjectProperty>
```

In FrameNet, a frame may inherit (A ISA B) from other frames or be composed of a set of subframes (which are frames themselves). For instance, the frame CRIMINAL PROCESS has subframes that correspond to various stages (ARREST,
ARRAIGNMENT, CHARGE, etc.). Subframe relations are represented using the `daml:objectProperty`.\(^4\)

```xml
<daml:ObjectProperty rdf:ID="subFrameOf">  
  <dфи:domain rdf:resource="#Frame"/>
  <dфи:range rdf:resource="#Frame"/>
</daml:ObjectProperty>
```

A central relation between subframes is one of temporal ordering. We use `precedes` (in the sense of immediately precedes) to encode this relation between subframes.

```xml
<daml:ObjectProperty rdf:ID="precedes">  
  <dфи:domain rdf:resource="#subFrame"/>
  <dфи:range rdf:resource="#subFrame"/>
</daml:ObjectProperty>
```

We can define a property `temporalOrdering` that is the transitive version of `precedes`.

```xml
<daml:TransitiveProperty rdf:ID="temporalOrdering">  
  <dфи:label>TemporalOrdering</dфи:label>
</daml:TransitiveProperty>
```

Note that the `temporalOrdering` property only says it is transitive, not that it is a transitive version of `precedes`. DAML+OIL does not currently allow us to express this relation. (See http://www.daml.org/2001/03/daml+oil-walkthru#properties).

Frame Elements may also inherit from each other. We use the `rdf:subPropertyOf` to specify this dependences. For example, the following markup in DAML+OIL specifies that the role (Frame Element) MOTHER inherits from the role (Frame Element) PARENT. Note we can add further restrictions to the new role. For instance, we may want to restrict the filler of the MOTHER to be female (as opposed to animal for PARENT).

```xml
<daml:ObjectProperty rdf:ID="mother">  
  <dфи:subPropertyOf rdf:resource="#parent"/>  
  <dфи:range rdf:resource="#Female"/>
</daml:ObjectProperty>
```

With these basic frame primitives defined, we are ready to look at an example using the Criminal Process frames.

**An Example: The Criminal Process Frame**

The basic frame is the CRIMINAL PROCESS Frame. It is a type of background frame. CP is used as a shorthand for this frame.

---

\(^4\) The `subFrameOf` relation has a direct translation to a richer semantic representation that is able to model and reason about complex processes (such as buying, selling, reserving tickets) and services on the web. While the details of the representation are outside the scope of the this paper, the interested reader can look at [10] for an exposition of the markup language and its operational semantics.
The CRIMINALPROCES frame has a set of associated roles. These roles include that of COURT, DEFENDANT, PROSECUTION, DEFENSE, JURY, and CHARGES. Each of these roles may have a filler with a specific semantic type restriction. FrameNet does not specify the world knowledge and ontology required to reason about Frame Element filler types. We believe that one of the possible advantages in encoding FrameNet data in DAML+OIL is that as and when ontologies become available on the web (such as OpenCYC), we can link to them for this purpose.

In the example fragment below we use the CYC Court-Judicial collection to specify the type of the COURT and the CYC Lawyer definition to specify the type restriction on the frame element DEFENSE. For illustrative purposes, the DAML+OIL markup below shows the use of a different ontology (from CYC) to restrict the defendant to be of type PERSON as defined in the example ontology. This restriction uses the DAML+OIL example from http://www.daml.org/2001/03/daml-oil-ex

The set of binding relations involves a set of role identification statements that specify that a role of a frame (subframe) has the same value (bound to the same object) as the role of a subframe (frame). We could specify these constraints either a) as anonymous subclass restrictions on the criminal process class (see http://www.daml.org/2001/03/daml-oil-ex for examples) or b) we could name each individual constraint (and thus obtain a handle onto that property). We chose the later method in our DAML+OIL encoding of FrameNet to allow users/programs to query any specific constraint (or modify it). Note also that the use of the dotting notation (A.b) to specify paths through simple and complex frames and is not fully supported in DAML+OIL (see http://www.daml.org/services/daml-s/2001/10/rationale.html and also [10] for more info).
Subframes of the **CriminalProcess** frame are defined by their type (LexicalFrame or a BackgroundFrame). For example, **Arrest** and **Arraignment** are Lexical Frames while **Trial** is a BackgroundFrame (all are subframes of **CriminalProcess**). We subtype the `subFrameOf` property to specify the individual subframe relations (shown below for the relation `subFrameOf(CriminalProcess, Arraignment)`).

To specify the the relation `precedes(Arrest, Arraignment)` we restrict the property `precedes` within (the domain of) the **Arrest** frame to have as one of its range values the frame (class) **Arraignment**. This is done using the property restriction feature with DAML+OIL as follows.

With this markup of the ontology, we can create annotation instances for examples with targets that belong to the **CriminalProcess** (or its associated) frames.
Basic Annotation of Verb Arguments and Complements as Triplets

Consider the following sentence, which is annotated for the target nab, a verb in the ARREST frame; the frame elements represented are the arresting AUTHORITIES, the SUSPECT and the TIME when the event took place:

\[
\text{[Authorities Police nabbed [Suspect the man], who was out on licence from prison, [Time when he returned home].}
\]

The phrase *who was out on licence from prison* provides additional information about the SUSPECT, but it is not syntactically an argument or complement of the target verb, nor semantically an element of the ARREST frame, so it is not annotated.

How do we propose to represent this in XML conforming to the proposed standards? The header of the file will refer to the FrameNet Data Category specification discussed in the last section:

\[
1 \text{<xml version="1.0" encoding="UTF-8">}
2 \text{<DDTYPE definitions like those shown in the preceding section go here}>}
3 \text{<lexunit-annotation name="nab" frame="arrest" pos="v">}
4 \text{<definition>COP: catch (someone) doing something wrong. </definition>}
5 \text{<subcorpus name="y-001-all">}
\]

The entity <lexunit-annotation>, which comprises the rest of the file includes attributes giving the name of the lexical unit (nab), the name of the frame (ARREST), and the part of speech of the lemma (verb). The first included element is a definition of the lemma within the frame, seen on line 4.

The entities contained within the lexunit-annotation are called subcorpora; we select sentences for annotation, according to their syntactic patterns, collocations, etc. Each such selectional pattern results in a subcorpus which is made available to the annotators; in our representation of the annotated sentences we want to preserve these groupings. In the case of nab, there are so few instances of the word that we have lumped them all into one subcorpus as shown on line 5.

It might seem logical that the entities within the subcorpus should be sentences, but in fact, we recognize the possibility that one sentence might be annotated several times, for several targets. There might even be several instances of the same target lemma in the same sentence in the same frame (e.g. *The FBI nabbed Jones in NYC, while the Mounties nabbed Smith in Toronto*), each with its own set of FEs. Therefore, the next smaller entity is the annotation set (line 6).

The annotation set, shown below, consists of the <sentence>, which thus far contains only the <text> of the sentence, and a set of layers, each consisting of a set of labels. Each label has attributes start and end, giving the starting and ending position in the text to which it is applied. This sentence is typical of the basic FrameNet annotation style, in that there are three main layers, one for frame elements (“FE”, line 8), one for the phrase type (PT) of each FE (line
22), and one for the grammatical function (GF) of each FE (line 15). In each case, there are three coextensive labels; thus the word Police, in text positions 0-5 expresses the FE AUTHORITIES (line 10), has the phrase type “NP” (line 24) and is the subject of the verb nab, which we refer to as external argument “Ext” (line 17). The other two frame elements are shown by similar triplets, SUSPECT-NP-Obj and TIME-Sw-Comp, the latter meaning a complement of the verb consisting of a clause (S-node) introduced by a WH-relative.

There are three other layers shown in the example, none of which contain labels, called Sentence, Verb, and Other. The layer Target contains the single label Target; the fact that nab is the target word is indicated in the same way as the information about FEs.

As the basic unit of annotation is the label, which can be applied to anything ranging from a single character to an entire sentence, and there are no a priori constraints on labels overlapping, a great variety of information can be represented in this way. We will not be able to demonstrate all the possibilities here, but we will give a some representative examples.

In FrameNet, event nouns are annotated in the same frame (and hence with the same FEs) as the corresponding verbs; the main differences are that the syntactic patterns for the FEs of nouns are more varied, and (with rare exceptions),
no FEs of nouns are required to be expressed. Consider the noun arrest, also in the ARREST frame, in the sentence:

Two witnesses have come forward with information that could lead to $\text{[Suspect the killer ’s] arrest}$. 

In this case the SUSPECT is expressed as a possessive (the killer’s; it could equally well have been in a PP headed by of (the arrest of the killer)).

In addition to marking the FE SUSPECT from ARREST, we could also annotate the same sentence again in the CAUSATION frame with the target lead, which would create an annotation set listed under the the LU lead_to:

Two witnesses have come forward with $\text{[Cause information that]}$ could lead $\text{[Effect to the killer’s arrest]}$.

The same sentence would be annotated in two different frames, and the semantics of the two frames could (in theory) be combined compositionally to get the semantics of the phrase information that could lead to the killer’s arrest. Similar processes of annotating in multiple frames with targets come_forward (and possibly witness as well) should yield a full semantics of the sentence.

At the current stage, we have converted all of FrameNet 1 data (annotations and frame descriptions) to DAML+OIL. The translator has also been updated to handle the more complex semantic relations (both frame and frame element based) in FrameNet 2. We plan to release both the XML and the RDF-based DAML+OIL versions of all FrameNet 2 releases.

**Conclusion**

The World Wide Web (WWW) contains a large amount of information which is expanding at a rapid rate. The information contained encompasses diverse ranges and types of data, from structured databases to text. While XML allows for the specification of metadata as a means to specify the structure and syntax of a URL, it does not provide much capability to describe the semantic relations between the different pieces of information in a manner that can be exploited by software programs for automation or interpretation. DAML+OIL is a widely used language related to the Semantic Web initiative that is poised to remedy this situation. The DAML+OIL language is being developed as an extension to XML and the Resource Description Framework (RDF). The latest release of the language (DAML+OIL) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable.

This rapid growth of the web also engenders a need for sophisticated techniques to represent lexical and sense distinctions in a machine readable and interpretable manner. The FrameNet database documents the range of semantic and syntactic combinatory possibilities (valences) of each word in each of its
senses, through manual annotation of example sentences and automatic summarization of the resulting annotations. We believe that FrameNet offers promise as a potential resource to aid in the automatic identification and disambiguation of word meanings on the semantic web.

This paper described an encoding of FrameNet data in the DAML+OIL language. FrameNet 1 (both frame descriptions and annotations) has already been translated into DAML+OIL, and we reported on the updated translator which translates the richer set of frame and frame element relations in FrameNet 2 to DAML+OIL.

With periodic updates as the FrameNet data increases, we expect the DAML+OIL encoding of FrameNet to become useful for various applications on the semantic web. Conversely, as the set of DAML+OIL ontologies matures, FrameNet can directly link to the semantic web to incorporate domain-specific information including semantic typing of frame elements.

References