

FrameNet Meets the Semantic Web: A DAML+OIL Frame Representation*

Srini Narayanan, Charles J. Fillmore, Collin F. Baker, Miriam R. L. Petruck
{snarayan,fillmore,collinb,miriamp}@icsi.berkeley.edu

Abstract

The Berkeley FrameNet Project (Baker, Fillmore, & Lowe 1998; Fillmore & Baker 2001) (URL: <http://framenet.icsi.berkeley.edu/framenet>) is creating an online lexical resource for English, based on the principles of **Frame Semantics** and supported by corpus evidence. 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. The FrameNet database is available in XML, and can be displayed and queried via the web and other interfaces. The FrameNet I data has also been translated into the DAML+OIL extension to XML and the Resource Description Framework (RDF), which can represent our ontologies and to make FrameNet information machine readable and understandable. We have developed an automatic translator from FrameNet I data in XML to DAML+OIL; this paper reports on our representation of this data in DAML+OIL.

Introduction

It is generally assumed that the goals of computational linguistics and those of ordinary linguistic analysis are unlikely to coincide in the area of lexicon building. The full range of the knowledge speakers have about the words in their language appears to be out of reach of practical efforts to acquire and organize such knowledge for efficient use in NLP applications.

FrameNet, as a computational linguistics project, aims at a compromise between two extremes. At the one end it is

*We are grateful to the National Science Foundation for funding the work of the FrameNet project through two grants, IRI #9618838 “Tools for Lexicon Building” March 1997–February 2000, and ITR/HCI #0086132 “FrameNet++: An On-Line Lexical Semantic Resource and its Application to Speech and Language Technology” September 2000–August 2003. Copyright © 2002, American Association for Artificial Intelligence (www.aai.org). All rights reserved.

clearly possible to do careful and subtle conceptual analysis of the items in a small lexicon designed for restricted purposes in a narrow domain, where the goal of achieving reliable language understanding and accurate inference generation seems attainable. At the other end, one can aim at fairly superficial information retrieval or other NLP applications by using statistical methods for acquiring very large lexicons or by creating clever programs for making maximal use of the information available in machine-tractable dictionaries. The latter contain large amounts of data and are capable of yielding important facts not anticipated by their compilers (Wilks, Slator, & Guthrie 1996) but they are limited to what somebody believed would be relevant to human users.

The FrameNet team are convinced that at a certain level, the depth and subtlety of careful linguistic analysis can be attained using the wisely exploited judgments of linguistically trained researchers who annotate syntactically sorted sentences¹ taken from a large natural language corpus. Some of the information made available by such research will be recorded directly by the annotators, and some will be derivable—partly manually, partly automatically—from the results of such annotations. This process is very labor intensive, but we are also moving toward a system in which a priori rules are used to pre-mark most of the frame elements, and humans can approve or disapprove of these algorithmic assignments.

The FrameNet database now contains information about more than 4,000 lexical units (senses of words) based on the annotation of more than 100,000 sentences. We want to make this information available for a wide range of uses, from students trying to understand how a word is used in English to machine translation and information extraction programs. 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 it will become a resource for the Semantic Web.

¹It may be misleading to refer to this as simply “annotating sentences”: each of the sentences is seen as a concordance line illustrating a context for a single word, and the annotations are limited to recording the relevant features of the context of that word.

Background

The basic assumption of Frame Semantics (Fillmore 1976; 1977; Fillmore & Atkins 1992; Petruck 1996) as it applies to the description of lexical meanings is that each word (in a given meaning) **evokes** a particular frame and possibly **profiles** some element or aspect of that frame.² 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—which we can call **scenarios**—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 to be understood against the set of assumptions and practices of that frame.

Replacement: An Example Frame

FrameNet has tried to concentrate on frames which help explicate the meanings of groups of words, rather than frames that cover just one word. One such frame is the REPLACE-MENT frame. An abstract description of this frame requires the positing of a PLACE (which can be a role, a function, a location, a job, a status, etc.), and two distinct entities (X_1 and X_2) whose histories include being in this PLACE. We can then line up the separate pieces of the REPLACEMENT Scenario as follows:

Scenario 1

- Scene 1. X_1 is at Place.
- Scene 2. X_2 is not at Place.
- Transition 1. X_1 moves away from Place.
- Transition 2. X_2 moves to Place.
- Scene 3. X_1 is not at Place.
- Scene 4. X_2 is at Place.

Ordering Constraints

- Constraint: Scene 1 precedes Transition 1.
- Constraint: Scene 2 precedes Transition 2.
- Constraint: Transition 1 precedes Transition 2.
- Constraint: Transition 1 precedes Scene 3.
- Constraint: Transition 2 precedes Scene 4.

This Scenario 1 is common to (and is “inherited by”) two other scenarios which add certain elaborations. The temporal precedence relation holding between Transition 1 and Transition 2 is general enough to allow anything from immediate replacement of X_1 by X_2 , as in the game “King-of-the-Hill” (where X_2 pushes X_1 off the hill) to a temporal gap of considerable length, as when the role occupied by a

²The term *profile* (used here as a verb) is borrowed from (Lan-gacker 1987), esp. pp. 183ff.

retired or deceased company officer remains unoccupied for many years.

In terms of their separate historical association with the Place in Scenario 1, we can refer to the entities X_1 and X_2 as the Old Theme and the New Theme respectively, or, simply, Old and New. We refer to such names as OLD, NEW, and PLACE as **frame elements** or **FEs**.³

Now we can look at some of the vocabulary that fits the description of this frame.

Limiting ourselves to the scenario just sketched out (which contains no suggestion of agency or causation), we can see that the noun *replacement* has two uses: in one of its senses it can profile the whole event, as with the complex subject NP in the sentence [*Sylvia’s replacement by Adam on the committee*] was a surprise to everyone; or it can refer to (i.e., “profile”) only the NEW entity in the final scene, as identified in the subject of the sentence [*Sylvia’s replacement*] turned out to be completely unprepared for the job. The verb *replace* temporally profiles Transition 2, as in *Adam replaced Sylvia on the committee*. This means that if a temporal qualification is added to this clause, e.g. *last week*, it identifies the time when Adam showed up in the PLACE, and that does not necessarily include the time when Sylvia left it.

The expressions we have examined so far have only two FEs, the NEW and the OLD. With these sentences we intend only the interpretation that Adam came to have the position on the committee which Sylvia used to have. But the verb *replace* actually has two more possible profilings, fitting two slightly different elaborations of our initial scenario, both involving an agency that brings about one or both of these transitions, i.e. two types of **causative**.

In the first (and more restricted) of these elaborations, there is a new entity X_3 that brings about Transition 2 but not Transition 1. Thus:

Scenario 2

- Scene 1. X_1 is at Place.
- Scene 2. X_2 is not at Place.
- Transition 1. X_1 moves away from Place.
- Transition 2. X_2 moves to Place.
- Cause: X_3 causes Transition 2.
- Scene 3. X_1 is not at Place.
- Scene 4. X_2 is at Place.

In an interpretation supported by this second scenario, the entity X_3 (which we will refer to as AGENT) can participate in a profiling of the “Cause” clause, as in *Lydia replaced Sylvia (with Adam) on the committee*. In this use of *replace* it is possible that the AGENT, Lydia, had nothing to

³The FEs are in many ways analogous to the **case roles** or **thematic roles** of various frameworks. For a small number of frames, the FEs could be labeled by familiar terms from such traditions, like **Agent**, **Theme**, **Instrument**, etc., but we do not require ourselves to accommodate our FE labels to such lists in cases where there is no obvious fit. The FE names are understood in terms of roles in specific frames and do not need to be thought of as selected from a restricted universal set. These labels are *reusable*, to be sure—which is why we seek to include as many words as possible within individual frames—but our work can proceed without deciding on an initial complete inventory a priori.

do with Sylvia's move in Transition 1. That may have happened years ago. Given this second scenario, even though a NP mentioning the OLD is grammatically **required**, Sylvia might not be a participant in the profiled event at all. A temporal qualifier would only identify the time of Transition 2.

In Scenario 3, the second elaboration on the simple REPLACEMENT frame, we have a situation in which the AGENT is responsible for both of the transitions. Thus:

Scenario 3

- Scene 1. X₁ is at Place.
- Scene 2. X₂ is not at Place.
- Transition 1. X₁ moves away from Place.
- Transition 2. X₂ moves to Place.
- Cause: X₃ causes Transition 1 & Transition 2.
- Scene 3. X₁ is not at Place.
- Scene 4. X₂ is at Place.

On the interpretation of causative *replace* in this third scenario, we see Lydia as performing the two-part act of (1) removing Sylvia from the position and (2) placing Adam in it instead. As noted, the interpretation of temporal modifiers goes along with profiling: in Scenario 3, the sentence *Lydia replaced Sylvia with Adam last week* places both transitions within the same period.

Whether the existence of two interpretations corresponding to the two causative scenarios requires the positing of two **senses** for *replace* may be a problem for theories of polysemy, but there is no doubt of the difference in the implications. If you are told that *you can be replaced*, you have been warned that you might lose your job: you could become an unwilling participant in Transition 1. But if you are told that *you can never be replaced*, or equivalently that you are *irreplaceable*, then you will be pleased: they will never find anyone who can fill your shoes.

There are many other words that evoke one or more of these scenarios. The verbs *succeed* and *supersede* are for the most part limited to Scenario 1. The verb *supplant* has more or less the same possibilities, syntactic and semantic, as *replace* in all three scenarios. The verb *substitute* fits all three scenarios, but has a different syntax and subtly different semantic implications. In Scenario 1, where the verb is used intransitively, the NEW *substitutes for* the OLD, differing from *replace* in requiring the oblique rendering of the NEW with the preposition *for*. In the causative scenarios, the two verbs exhibit an alternation between direct and oblique objects, and a difference in prepositions for the oblique: *the AGENT substitutes the NEW for the OLD*, but *the AGENT replaces the OLD with the NEW*.

The word *successor* is a relational noun profiling the NEW in respect to the OLD, whereas the noun *predecessor* profiles the OLD in respect to the NEW, at a time period at the end of the transitions. The phrasal preposition *instead of* also fits this frame: *You can use raw sugar instead of refined sugar* is a paraphrase of *You can substitute raw sugar for refined sugar*. The adjective *existing* marks something as destined for replacement: if you see a campus map with one of the buildings identified as *the existing stadium*, you know that this structure is destined to be replaced. The time adjective *previous* belongs to this frame. If you tell me about your *former wife* or your *ex-wife*, I will know that she is some-

one who was once married to you; but if you speak of your *previous wife*, I will also know that she has in the meantime been replaced.

Numerous nouns are understood with reference to a specialization of the replacement frame: *crown prince*, *heir apparent*, *standby*, *understudy*, and so on. Thus, by defining the REPLACEMENT frame, we are able to concisely represent the semantics of at least twenty lexical items, including nouns, verbs, adjectives and a preposition.

The FrameNet Process

The job of FrameNet is to document from attested instances of contemporary English the manner in which frame elements (for given words in given meanings) are grammatically instantiated in English sentences and to organize and exhibit the results of such findings in a systematic way. We have seen, for example, that 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*.

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

The corpus used in FrameNet so far has been the British National Corpus, more than 100,000,000 running words of contemporary British English.⁴ In the current phase, we are beginning 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/>).

Lexicographic Relevance

The part of FrameNet descriptions concerned with the combined semantic and syntactic valence of lexical items requires consideration of both frame structure and syntactic structure. Our objective is to record exactly how the language relates frame elements to those syntactic constituents that are syntactically dependent on the lexical items that evoke the frame.

Clear instances of lexicographically relevant positions are those accepted in narrow conceptions of *subcategorization*

⁴Our 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>.

frames. If the target word is the verb *demonstrate* and the sentence is [*The instructor*] *demonstrated* [*the process*] [*to the students*], the direct object of the verb (*the process*) is clearly relevant. Since the act of demonstration requires an audience, most analysts would include the *to*-phrase (*to the students*). In other words, a lexical description of the verb must say that the subject and these two syntactic positions function in identifying frame elements that are necessary parts of the concept of the frame evoked by it. For the deverbal noun *demonstration*, clearly relevant positions for annotation would be a preceding possessive determiner or following prepositional complements, as in the NP [*the instructor's*] *demonstration* [*of the process*] [*to the class*].

An example of a context providing information that could be relevant to completing information about the frame evoked by a particular lexical item but which is out of reach of lexically determined contexts is to be seen in the reference to *the students* and *the process* in the following sentence, *The students complained that they couldn't really understand the process without a demonstration*. It is clear from a natural interpretation of the sentence as a whole that the audience of the demonstration is supposed to be *the students*, and that *the process* is what needs to be demonstrated, but such information does not come from noticing how the combinatorial possibilities of the noun *demonstration* are realized in this sentence. The process of understanding the sentence as a whole needs to call on some means of recognizing such information in the surrounding discourse. But here our lexicographic task is more restricted: to characterize the combinatorial requirements and possibilities that are specifically associated with individual words rather than with the possibilities provided by the whole grammar of the language. On lexicographic relevance, see (Fillmore & Atkins 1998).

Relations between Frames and FEs

Inheritance

We have mentioned one relation between frames, inheritance, without fully defining it. Frame inheritance is an IS-A relation. If frame B inherits from frame A, then B elaborates A, and is a subtype of A. Furthermore, all the frame elements of A are inherited by B. They may appear in B with exactly the same names and semantic types, or they may be subtypes (elaborations) of the FEs in A.

Since FE names are relative to the frames in which they occur, an FE called *SPEAKER* in B is distinct from an FE called *SPEAKER* in A, whether or not B is an elaboration of A, and whether or not its semantic type is further restricted in B. However, it is customary in FrameNet to use relatively frame-specific names when possible. Where we do choose the same name for FEs in two different frames, we have some sort of semantic similarity in mind, but this must ultimately be made specific through semantic typing and FE inheritance.⁵

⁵There is a certain similarity - and certainly a personal-history connection - between the "case frames" plus linking rules in Case Grammar (Fillmore 1968) on the one hand and the valence patterns derived from the FrameNet annotations on the other. Case Grammar assumed a fixed relevant-across-the-board collection of under-

The picture is further complicated by situations of multiple inheritance; an FE in one frame may inherit from FEs in two parent frames, combining their semantics. For example, the sentence ... *Simpson boldly accused him to his face of treason*... belongs in the *JUDGEMENT_COMMUNICATION* frame, which inherits from both *JUDGEMENT* and *COMMUNICATION*. Thus, *Simpson* is both a *JUDGE* by virtue of inheritance from *JUDGEMENT* and a *SPEAKER* by virtue of inheritance from *COMMUNICATION*. *Of treason* is the *REASON* for the judgement, and *to his face* is the *MANNER* of speaking, both of which are FEs in the *JUDGEMENT_COMMUNICATION* frame.

Subframes

The other type of relation between frames which is currently represented in the FN database is between a complex frame and several simpler frames (**subframes**) which constitute it. We call this frame composition, by which we mean the situation in which a complex frame is made up of parts which are also frames, as with the scenarios mentioned earlier in connection with the *REPLACEMENT* frame. This is true of a large family of *EXCHANGE* frames, in which ownership relations in one stage are reversed in a transition of transferring ownership. Verbs incorporating caused events also have subframes for the causing event and the caused event and are thus further instances of frame composition.

The structure of frames and subframes posited for a complete picture of the Criminal Process, as we will suggest, can be quite complex. Subframes of the entire process include *Arraignment* and *Trial*, but each of these in turn contains multiple subframes, often with different constellations of participants. We are open to the possibility that we will need to deal with other types of relations among frames, but have not yet represented them in our database.

The FrameNet Product

The FN database can be seen both as a dictionary and a thesaurus. It is a dictionary in that each **lexical unit** is provided with (1) the name of its frame, (2) a definition, (3) a valence description which summarizes the attested combinatorial possibilities, and (4) access to annotated examples illustrating each syntactic pattern found in the corpus and the kinds of semantic information instanced with such patterns. (Several more technical types of dictionary information are missing, such as etymology and various word-subclass details.) It is a thesaurus in that, by being linked to frames, each word is directly connected with other words in its frame(s), and further extensions are provided by working out the ways in which a word's basic frames are connected with other frames through relations of inheritance (possibly multiple inheritance) and composition.

The FrameNet database provides a great deal of information not included in familiar dictionaries. Consider the *CHANGE_OF_LEADERSHIP* frame, which characterizes the lying "cases" (a.k.a. thematic roles), whereas the accumulation of semantic relations in FrameNet is more deliberately inductive - if only to avoid wasting time trying to decide which of some standard list should be distorted to cover some new relationship.

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 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 discussed above, with the PLACE further specified as an (abstract) position of political power, and the X_1 and X_2 narrowed to humans or political entities.

Compare this with the minimal information provided in the online edition of the Merriam-Webster Dictionary and Thesaurus (<http://www.m-w.com>) for the verb *overthrow*, where the user learns that the definition of the word is *to cause the downfall of: BRING DOWN, DEFEAT*, and that some related words are *depose, dethrone, oust, remove, unseat*. No example sentences or semantic information are provided.

The resulting annotation database has been provided with a powerful search engine, called FrameSQL, developed by Hiroaki Sato, which gives a web interface to the MySQL database. FrameSQL has also been extended in certain ways by the recognition of certain kinds of words (e.g. support verbs, transparent nouns); it allows subtle queries about lexical collocations, grammatical tagging, frame-semantic differences among related words, and selection of support verbs for event nouns.

Criminal Process: A Full Example of Annotation and Representation

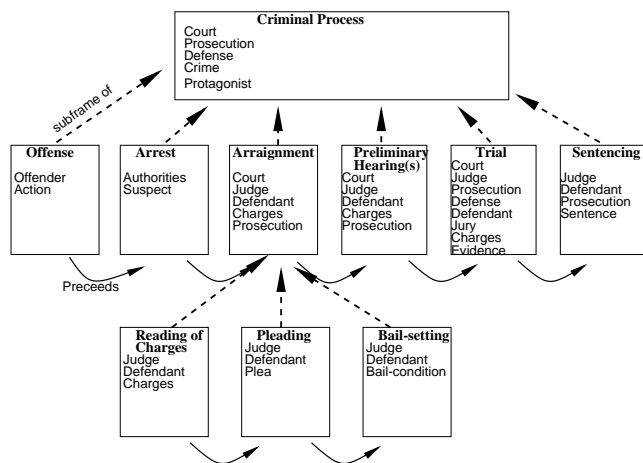


Figure 1: The CRIMINAL PROCESS frame and some of its subframes

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 definition of the frame is as follows:

CRIMINAL_PROCESS

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.

The frame elements are defined like this:

COURT	This FE identifies the court involved in a trial.
PLACE	This FE identifies the place where the event occurs.
TIME	This FE identifies the time when the event occurs.
DEFENDANT	The DEFENDANT is charged with an offense.
JUDGE	The JUDGE heads the court where arraignment occurs and the case is tried.
PROSECUTION	This FE identifies the attorney(s) prosecuting the Defendant.
CHARGES	This FE identifies the Charges brought against the Defendant.
DEFENSE	The DEFENSE represents the interests of the DEFENDANT.
JURY	This FE identifies the people who are charged by the Court to listen to the testimony, evidence, and arguments, and come to a consensus about the guilt or innocence of the DEFENDANT.
OFFENSE	This FE identifies the OFFENSE which the DEFENDANT is accused of committing.

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 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, and SUSPECT. The resulting FE annotations can be displayed in a bracketed notation like this:

[*Time* In July last year] [*Authorities* a German border guard] apprehended^{*Target*} [*Suspect* two Irishmen with Kalashnikov assault rifles].

A DAML+OIL representation of FrameNet

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 *crawl*, makes the meaning of the work 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) provides a rich set of constructs with which to create ontologies and to markup information so that it is machine readable and understandable.

Version 1 of Framenet has been translated into DAML+OIL (<http://www.daml.org>). 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+oil-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+oil-walkthru/#XMLNS>), and URIs. Thus, our framenet declaration begins with an RDF start tag including several namespace declarations of the form:

```
<?Xml version='1.0' encoding='ISO-8859-1'?>
<!DOCTYPE uridef[
  <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns">
  <!ENTITY rdfs"http://www.w3.org/2000/01/rdf-schema">
```

```
<!ENTITY xsd "http://www.w3.org/2000/10/XMLSchema">
<!ENTITY daml"http://www.daml.org/2001/03/daml+oil">
]>
<rdf:RDF
  xmlns:rdf = "&rdf;#"
  xmlns:rdfs = "&rdfs;#"
  xmlns:xsd = "&xsd;#"
  xmlns:daml = "&daml;#"
>
```

So in this document, the rdf: prefix should be understood as referring to things drawn from the namespace called <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+oil#>. This again is a conventional DAML+OIL declaration. We use the XML entity model to use shortcuts with referring to the URIs.⁶ 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/2002/04/08/cyc.daml>), and the SRI time ontology (<http://www.ai.sri.com/daml/ontologies/sri-basic/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:unionOf* a BACKGROUNDFRAME class (such as CRIMINAL PROCESS frame) and a LEXICALFRAME class (such as the ARREST frame). We then define a bunch of bookkeeping properties on the FRAME class. An example of the name property is shown below.

```
<daml:Class rdf:ID="Frame">
  <rdfs:comment> The most general class </rdfs:comment>
  <daml:unionOf rdf:parseType="daml:collection">
    <daml:Class rdf:about="#BackgroundFrame"/>
    <daml:Class rdf:about="#LexicalFrame"/>
  </daml:unionOf>
</daml:Class>

<daml:ObjectProperty rdf:ID="Name">
  <rdfs:domain rdf:resource="#Frame"/>
  <rdfs:range rdf:resource="&rdf-schema;#Literal"/>
</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 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

⁶Note that all URIs are globally scoped, so without this the entire path has to be specified.

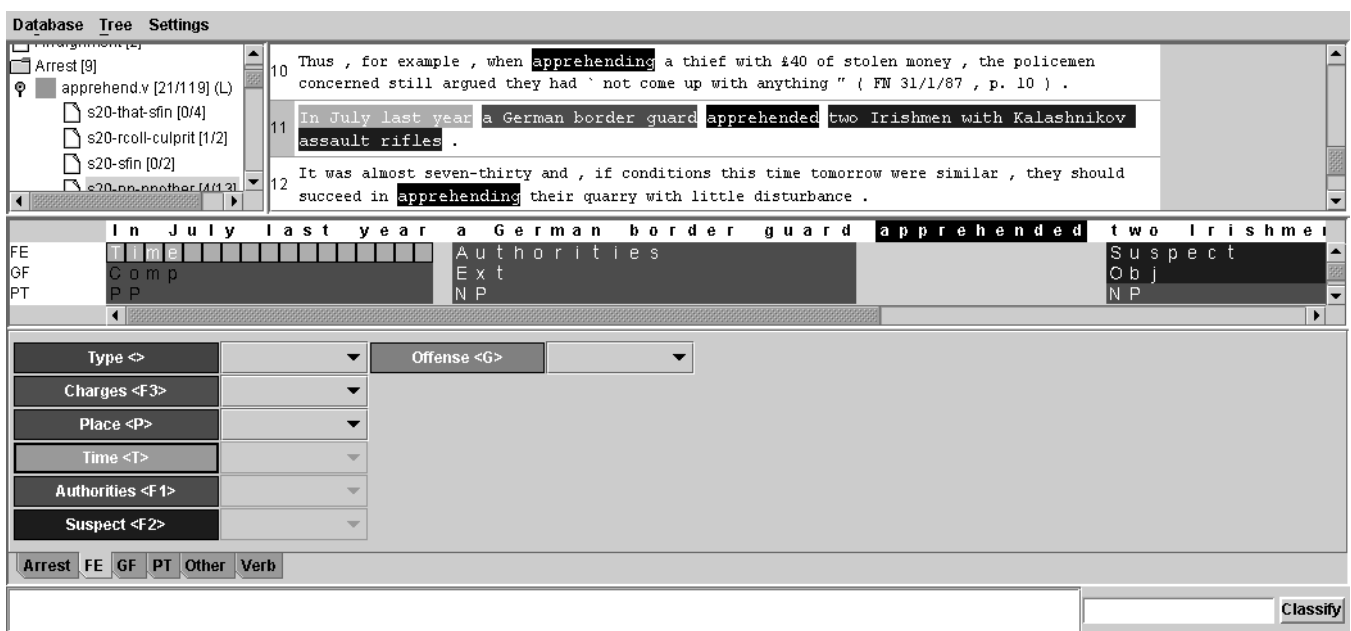


Figure 2: Annotation of a sentence in the ARREST frame.

use the *daml:samePropertyAs* relation to specify synonyms. The fragment below specifies that Frame Element, Role, and FE are synonyms.

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

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

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

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

```
<daml:ObjectProperty rdf:ID= "singleValuedRole">
  <rdfs:domain rdf:resource="#Frame"/>
  <rdfs:range>
    <rdfs:subClassOf>
      <daml:Restriction daml:maxCardinality="1">
        <daml:onProperty rdf:resource="#Role"/>
      </daml:Restriction>
    </rdfs:subClassOf>
  </rdfs:range>
</daml:Class>
```

The relation between lexicalized frames (such as ARREST) and background frames (such as CRIMINAL PRO-

CESS 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). To capture such bindings, we introduce a special relation called *bindingRelation* whose domain and range are roles (from the same or different frames).

```
<daml:ObjectProperty rdf:ID="bindingRelation">
  <rdfs:domain rdf:resource="#Role"/>
  <rdfs:range rdf:resource="#Role"/>
</daml:ObjectProperty>
```

By far the most important binding relation is the identification of roles (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 service ontology (<http://www.daml.org/services>) primitive *daml-s:sameValuesAS* to specify the *identify* relations.

```
<daml:ObjectProperty rdf:ID="identify">
  <rdfs:subPropertyOf rdf:resource="#bindingRelation"/>
  <rdfs:domain rdf:resource="#Role"/>
  <daml-s:sameValuesAs rdf:resource="#rdfs:range"/>
</daml:ObjectProperty>
```

In FrameNet, a frame may inherit (A ISA B) from other frames or be a *composition* of a set of subframes (which are frames themselves). For instance, an event frame CRIMINAL PROCESS may have subframes that correspond to various stages (ARREST, ARRAIGNMENT, CHARGE, etc.).⁷

⁷The *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 ser-

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

A central relation between subframes is one of temporal ordering. We use *precedes* (in fact we have both *temporalOrdering* (transitive *precedes*) and *precedes* (immediate *precedes*)) to encode this relation between subframes.

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

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

```
daml:TransitiveProperty rdf:ID="TemporalOrdering">
  <rdfs:label>TemporalOrdering</rdfs: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 *rdfs:subPropertyOf* (*rdfs* is the reference to *rdfs-schema*) 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*).

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

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

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.

```
<daml:Class rdf:ID="CriminalProcess">
  <rdfs:subClassOf rdf:resource="#BackgroundFrame"/>
</daml:Class>

<daml:Class rdf:ID="CP">
  <daml:sameClassAs rdf:resource="#CriminalProcess"/>
</daml:Class>
```

The *CRIMINALPROCESS* 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 at <http://www.cyc.com/2002/04/08/cyc.daml>), we can link to them for this purpose.

In the example fragment below we show the 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>

```
<daml:ObjectProperty rdf:ID="court">
  <rdfs:subPropertyOf rdf:resource="#FE"/>
  <rdfs:domain rdf:resource="#CriminalProcess"/>
  <rdfs:range rdf:resource="&CYC;#Court-Judicial"/>
</daml:ObjectProperty>
```

```
<daml:ObjectProperty rdf:ID="defense">
  <rdfs:subPropertyOf rdf:resource="#FE"/>
  <rdfs:domain rdf:resource="#CriminalProcess"/>
  <rdfs:range rdf:resource="&CYC;#Lawyer"/>
</daml:ObjectProperty>
```

```
<daml:ObjectProperty rdf:ID="defendant">
  <rdfs:subPropertyOf rdf:resource="#FE"/>
  <rdfs:domain rdf:resource="#CriminalProcess"/>
  <rdfs:range rdf:resource="&daml-ex;Person"/>
</daml:ObjectProperty>
```

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/damls/2001/10/rationale.html> and also (Narayanan & McIlraith 2002) for more info).

```
<daml:ObjectProperty rdf:ID="prosecutionConstraint">
  <rdfs:subPropertyOf rdf:resource="#identify"/>
  <rdfs:domain rdf:resource="#CP.prosecution"/>
  <rdfs:range rdf:resource="#Trial.prosecution"/>
</daml:ObjectProperty>
```

```
<daml:ObjectProperty rdf:ID="defendantConstraint">
  <rdfs:subPropertyOf rdf:resource="#identify"/>
```



```

<rdfs:domain rdf:resource="#CP.defendant"/>
<rdfs:range rdf:resource="#Arrest.suspect"/>
</daml:ObjectProperty>

```

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*(Criminal Process, Arraignment)).

```

<daml:Class rdf:ID="Arrest">
<rdfs:comment> A subframe </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#LexicalFrame"/>
</daml:Class>

<daml:Class rdf:ID="Arraignment">
<rdfs:comment> A subframe </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#LexicalFrame"/>
</daml:Class>

<daml:Class rdf:ID="Trial">
<rdfs:comment> A subframe </rdfs:comment>
  <rdfs:subClassOf rdf:resource="#BackgroundFrame"/>
</daml:Class>

<daml:ObjectProperty rdf:ID="arraignSubFrame">
  <rdfs:subPropertyOf rdf:resource="#subFrameOf"/>
  <rdfs:domain rdf:resource="#CP"/>
  <rdfs:range rdf:resource="#Arraignment"/>
</daml:ObjectProperty>

```

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.

```

<daml:Class rdf:about="#Arrest">
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#precedes"/>
      <daml:hasClass rdf:resource="#Arraignment"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>

```

With this markup of the ontology, we can create annotation instances for examples with targets that belong to the CRIMINALPROCESS (or its associated) frames. Consider the annotation example in the previous section.

[*Time* In July last year] [*Authorities* a German border guard] apprehended^{*Target*} [*Suspect* two Irishmen with Kalashnikov assault rifles].

Shown below is our DAML+OIL representation for the sentence, viewed as an instantiation that uses the FrameNet frame for ARREST.⁸

⁸The Annotation object has additional properties that specify text indices, and allow for arbitrary text tags to be placed on various parts of the annotated sentence. For ease of exposition, we only show the frame element relevant annotation; syntactic information

```

<fn:Annotation>
  <tpos> "36352897" </tpos>
  <frame rdf:about ="&fn:Arrest">
    <time> In July last year </time>
    <authorities> a German border guard </authorities>
    <target> apprehended </protagonist>
    <suspect>
      two Irishmen with Kalashnikov assault rifles.
    </suspect>
  </frame>
</fn:Annotation>

```

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 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 (<http://www.semanticWeb.org>) 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,

is tagged in a similar manner. Partial annotations of the sentence are possible with other frame-bearing words taken as target.

FrameNet can directly link to the semantic web to incorporate domain-specific information including semantic typing of frame elements.

References

- Baker, C. F.; Fillmore, C. J.; and Lowe, J. B. 1998. The Berkeley FrameNet Project. In *COLING-ACL '98: Proceedings of the Conference*, 86–90.
- Fillmore, C. J., and Atkins, B. 1992. Towards a frame-based lexicon: The semantics of RISK and its neighbors. In Lehrer, A., and Kittay, E. F., eds., *Frames, Fields and Contrasts*. Lawrence Erlbaum Associates.
- Fillmore, C. J., and Atkins, B. 1998. Framenet and lexicographic relevance. In *Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain*.
- Fillmore, C. J., and Baker, C. F. 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop*. Held in conjunction with the NAACL Annual Meeting.
- Fillmore, C. J. 1968. The case for case. In Bach, E., and Harms, R., eds., *Universals in Linguistic Theory*. New York: Holt, Rinehart & Winston.
- Fillmore, C. J. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, 20–32.
- Fillmore, C. J. 1977. Scenes-and-frames semantics. In Zampolli, A., ed., *Linguistic Structures Processing*, number 59 in Fundamental Studies in Computer Science. North Holland Publishing.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar, Volume I: Theoretical Prerequisites*. Stanford, California: Stanford University Press.
- Narayanan, S., and McIlraith, S. 2002. Simulation, verification and automatic composition of web services. In *Proceedings of the Eleventh International World Wide Web Conference (WWW2002), Honolulu, Hawaii, May 7-10, 2002*.
- Petruck, M. 1996. Frame semantics. In Verschueren, J.; Östman, J.-O.; Blommaert, J.; ; and Bulcaen, C., eds., *Handbook of Pragmatics*. Philadelphia: John Benjamins.
- Wilks, Y.; Slator, B. M.; and Guthrie, L. M. 1996. *Electric Words: Dictionaries, Computers, and Meanings*. ACL-MIT Press series in natural-language processing. Cambridge, Mass.: MIT Press.