The FrameNet Frame Element Taxonomy

Ken Litkowski
CL Research
9208 Gue Road
Damascus, MD 20872 USA
ken@clres.com

Abstract

The identification of semantic roles has been a prominent topic within linguistics in general and within the computational linguistics community for decades. To a great extent, these efforts have been a matter of opinion and theoretical discussions. Frame semantics has followed in this tradition. FrameNet, the embodiment of frame semantics, has identified over 1,100 distinctly named frame elements as its lexicographers have developed this resource. While this proliferation of semantic roles may seem even more confusing, FrameNet, with its frame-to-frame relations, provides a basis by which a definitive, data-driven taxonomy can be identified. We describe the methods by which these relations have been analyzed to develop this taxonomy and identify the resultant primitives. This data is being made available online for perusal and suggested changes and for download in the hope that this taxonomy may prove useful for those involved in semantic role labeling.

1 Introduction

Semantic role labeling has become a prominent task in computational linguistics, particularly since the groundbreaking study of Gildea and Jurafsky (2002), in which the authors conflated FrameNet frame-specific roles into 18 semantic roles. O'Hara and Wiebe (2009) provide an inventory of semantic relations derived from the frame elements of FrameNet, along with a mapping identifying what was done for semantic roles occurring at least 50 times. They also discuss a number of other schemes for identifying a set of basic semantic relations or roles. Marquez et al. (2008) also discuss the nature of semantic roles, particularly noting the lack of agreement among linguistic researchers. Dowty (1991) proposed only two roles: proto-agent and proto-patient. However, as Baker et al. (1998) point out, the role names (frame elements) are considered local to particular conceptual structures (frames). Thus, each FrameNet frame element in each frame constitutes a distinct semantic role (now over 10,000).

In the cases where frame elements have been collapsed, a common thread is that researcher's judgment has been used to develop the inventories. While these inventories appear to be reasonable, a data-driven approach may be useful in furthering agreement and providing a mechanism for improvements. The frame-to-frame relations in the FrameNet database can be used as the basis for a more principled identification of abstract semantic roles (primitive frame elements). The method to accomplish this uses a digraph analysis of a dictionary of frame elements constructed out of individual mappings developed in the FrameNet project itself.

In section 2, we motivate our approach by describing other research that coarsens the FrameNet data. In section 3, we describe FrameNet's frame-to-frame relations and how they constitute an appropriate basis upon which to perform the analysis. Section 4 provides an initial characterization of the dictionary and its digraph. Section 5 details the methods and set of operations used to identify and resolve inconsistencies. Section 6 describes the resultant taxonomy and frame element primitives. Finally, section 7 identifies some potential uses and further development of the taxonomy and section 8 draws conclusions from this analysis.

2 Motivation for Collapsing Frame Elements

With over 1100 frame elements, FrameNet data provide very fine-grained semantic roles, with perhaps few instances upon which to base semantic parsers. Several efforts have been more specifically designed to collapse the frame elements based on the FrameNet frame hierarchy.

Matsubayashi et al. (2009) used the frame to frame hierarchy, the frame element names (as human-understandable descriptors), the FrameNet semantic types, and VerbNet thematic roles to achieve a 19 percent improvement in the semantic role classification task. They suggested the need for further analyzing the weakness of the FrameNet hierarchy in the hopes of improving their results.

McConville and Dzikovska (2008) note that FrameNet's level of role name granularity creates problems for parsing and demonstrate that role inheritance can reduce the size of the role set without losing information. They focus on the core subcategorization frames of verbs, where deep parsing generally uses only a small number of semantic roles. They used the inheritance hierarchy to link fine-grained roles of child types with the more generic roles of their parent types. Their methodology limited the number of cases to which their rules applied, but they concluded that this process helped consolidate subcategorization frames with individual frames.

Ruppenhofer et al. (2010) present a method and a tool for creating customized versions of the FrameNet database by coarsening the sense inventory, merging entire frames and/or word senses. The tool, called FrameNet Transformer, does this by using a user-specifiable set of frame relations, selected from the frame-to-frame relations included in the FrameNet data. Thus, child frames are merged with parent frames, producing a new "release" of FrameNet. The authors show that the resultant coarsened FrameNet does not affect parser performance or certain task-specific results (such as recognizing textual entailment).

Ovchinnikova et al. (2010) focus on improving the conceptual structure of FrameNet for the sake of using this resource for such reasoning tasks as question answering and recognizing textual entailment. Their methodology involves the use of ontological analysis of frame relations to identify the similarity between frames. They cluster frames based on an overlap of frame elements and the commonality of lexemes evoking frames. In the context of recognizing textual entailment, the authors plan to develop further methods for mapping frame elements of related frames.

3 FrameNet Frame-to-Frame Relations

FrameNet uses frame-to-frame relations to map frame elements of one frame into those of a more primitive frame.[1] These can be exploited for developing a frame element hierarchy, and ultimately a strict taxonomy. Specifically, we used the following seven frame-to-frame relations as sources of child to parent relations: INHERITS (Inheritance), USES (Using), PRECEDES (Precedes), IS_SUB_OF (Subframe), PERSP_ON (Perspective On), INCH_OF (Inchoative Of), and CAUSE_OF (Causative Of).[2] For example, Table 1 shows the mapping for **Achieving_first** INHERITS **Intentionally_create**. Notice that six frame elements are identical, whereas the bolded items are different, and four frame elements are not involved in the mapping. Based on this mapping, we say that *Created_entity* is a hypernym of *New_idea*.

Achieving_first	Intentionally_create
Basis	**Components**
Cognizer	**Creator**
Field	**Depictive**
Instrument	Instrument
Location_of_appearance	
Manner	Manner
Means	Means
New_idea	**Created_entity**
Place	Place
Purpose	Purpose
Reason	
Time	Time
	Co_participant
	Role

Table 1. **Achieving_first** INHERITS **Intentionally_Create**

With this basic idea, we created a dictionary with the 1170 distinct frame elements from FrameNet 1.5 as main entries or headwords.[3] In

[1] FrameNet's FrameGrapher provides a visualization tool to examine the relations between frames and their frame elements.

[2] We did not use See_also or ReFraming_Mapping, since they do not provide a child to parent relationship.

[3] We first developed this methodology with FrameNet 1.3, which contained 1015 frame element names. We will sometimes refer to this earlier effort and will indicate differences,

creating this dictionary with CL Research's Dictionary Maintenance Program (DIMAP)[4], we used all the frame-to-frame mappings identified above as the basis for creating hypernymic relations between the frame element entries (following the pattern in Table 1). In addition, we counted the number of frames in which the frame element is used and put this in the "definition" field for each frame element entry; thus, the **Agent** frame element entry has the definition "180". We then used DIMAP to perform a digraph analysis to create and analyze the network of frame elements, i.e., to find its primitives and lay out its hierarchical structure.[5] The steps involved in this analysis are described below.

There were many problems in the initial network and hierarchy induced from the frame-to-frame relations, including

- Circularity,
- Frame elements without hypernyms, thus resulting in a large number of so-called primitive elements,
- Questionable identification of hypernyms, and
- Editorial inconsistencies in FrameNet.

We developed a set of procedures for resolving these problems, making changes to individual entries to arrive at a consistent set.

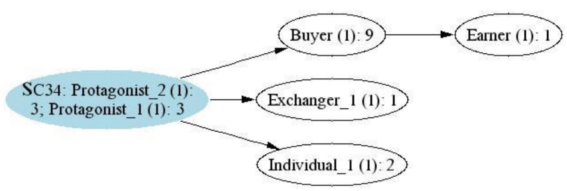

Figure 1. Portion of Initial Digraph

also showing that the additional frame elements fold into the earlier version.

[4] DIMAP is a Windows program that provides a generalized structure for creating entries with multiple senses, with fields for identifying hypernyms, hyponyms, features, and roles, intended for use in NLP applications. DIMAP dictionaries are available for WordNet and FrameNet, among others.

[5] DIMAP includes a capability for analyzing a dictionary's digraph via its hypernym links using Tarjan's strongly connected components algorithm. This is linear in the number of edges in the graph, i.e., O(|V| + |E|), where V are the vertexes and E are the edges. For this frame element dictionary, the running time is typically about 10 seconds.

Figure 1 shows a portion of the initial digraph, where two frame elements (**Protagonist_1** and **Protagonist_2**) are in a strong component node (i.e., they are hypernyms of each other) and the other nodes have this node as a hypernym. In this figure, the number in parentheses after the frame element name is the sense number for the entry and the numbers after the colon identify the number of frames in which the given frame element appears. In addition, neither of the two frame elements in the strong component has a hypernym, so this node is also a primitive.

4 Initial Results

The main objectives of analyzing the frame element digraph are to identify the primitive frame elements and to show the derivational hierarchy of each of the other frame elements. The initial frame element dictionary contains 1150 frame elements, 20 fewer than the number of distinct names used in FrameNet (explained below). Based on the hypernym relationships, the resultant digraph contains exactly 1000 nodes with exactly 500 primitives. The primitives fall into two classes:

- 425 frame elements with no hypernyms and not used as the hypernym with for any other frame element
- 75 frame elements with no hypernyms and that are used as the primitives in defining other frame elements

These 75 base nodes in the digraph are used as hypernymic links for the remaining 500 frame elements. The digraph includes 82 strong components (i.e., circularities with multiple frame elements in one node), generally consisting of 1 to 3 frame elements (but with one of 133) (constituting circular definitional paths) and that are used in defining other frame elements. These strong components reduce the number of nodes in the digraph from the expected 1170 and require elimination, as described below.

Figure 1 shows a portion of the initial digraph, where two frame elements (**Protagonist_1** and **Protagonist_2**) are in a strong component node (implying that they are hypernyms of each other) and the other nodes have this node as a hypernym. In this figure, the numbers after the colon identify the number of frames in which the given frame element appears. In addition, neither of the two frame elements in the strong component has a hypernym, so this node is also a primitive.

5 Identifying and Resolving Inconsistencies

While at first glance, the digraph may appear hopelessly complex, it is possible to articulate a set of operations (**Move**, **Merge**, **Delink**, **Delete**, and **Split**) that can be used to transform this digraph into a strict taxonomy. The following subsections detail these operations. While they are presented sequentially, they were actually applied opportunistically on individual frame elements, followed by rerunning the digraph analysis. Typically, changes to individual frame elements would have implications for the overall digraph, such as breaking apart the large strong component into several smaller strong components.

5.1 Overarching Principles of Analysis

In making changes to the frame element dictionary, a primary rule of thumb is that the list of frame elements should not change. That is, we want to retain the exact set of frame elements, so that the dictionary reflects whatever is present in FrameNet. When our analysis suggests that some change is needed in the underlying data, it will be recorded for consideration by the FrameNet lexicographers. Only the DIMAP dictionary will be changed.

As mentioned above, there is a basic inconsistency between the number of frame elements (1170) and the number of nodes in the digraph (1150). This inconsistency is the result of different capitalization in the frame element names in the frame-to-frame relations from which the dictionary was developed. For example, the frame-to-frame relations use both **Legal_Basis** and **Legal_basis** (giving rise to two senses, rather than two entries, in the frame element dictionary). This capitalization difference occurs for 20 frame element names, explaining the difference between the number of frame elements and the number of nodes.

Beyond this, there are several other editorial differences that affect the total number of frame elements as well as the paths between them, including misspellings (**depictive** and **depicitive**) and editorial variation (**Duration_of_end_state** and **Duration_of_endstate**, **Entity_1** and **Entity1**).

After considering editorial variations, the substance of analyzing and modifying the digraph begins. Essentially, this consists in making changes to the hypernymic link within each frame element. It is important to observe that such a change is a local decision, as opposed to making some global design change. Making a local change immediately reverberates throughout the full digraph. Rerunning the digraph analysis within the DIMAP frame element dictionary only takes a few seconds. Thus, the primary task is to clearly lay out steps and rationales for making changes to the hypernymic links.

While it might seem that the problem with the entry **Depicitve** can be solved by merging it with **Depictive**, this would violate the primary rule of thumb.[6] Instead, the problem is solved by creating a hypernymic link from it to **Depictive**. When the underlying data from the FrameNet frame-to-frame relations is corrected, a new frame element dictionary will not contain the misspelled frame element. That is, the digraph analysis will highlight where changes to the underlying data are needed.

As indicated, every hypernymic link change made to the frame element dictionary needs to be clearly documented, so that changes can be reapplied if changes are made to the underlying data and so that the validity of the changes can be assessed (by others). The following general methods were used:

- Analyze the circular strong components to eliminate them (e.g., **Purpose** and **Reason** need to be separated into two nodes to eliminate the circularity),
- For entries with editorial variations, create appropriate links that tie these frame elements to a base form,
- There are 54 frame elements with a "1" or "2" in their names. Most of these have a corresponding frame element without a number; the ones with numbers can be linked to those without the number, e.g., **Entity_1** and **Entity_2** can be linked to **Entity**,

[6] In analyzing FrameNet 1.3, the frame element **Depicitive** was present. This was corrected in FrameNet 1.5, but with this other misspelling introduced.

- Many frame elements are plurals (e.g., **Entities**) and there is another frame element (e.g., **Entity**) in a singular form. Create a link from the plural form to the singular form (with an implicit "singular_of" relation from the plural to the singular),
- Examine the definitions of the frame elements within each of their frames to identify hypernyms (e.g., there are 180 frame element definitions for **Agent**), particularly for the 400 frame elements that initially had no hypernymic links. This analysis can include a simple examination of the frame element names (e.g., **Focal_entity** can be considered in relation to **Entity**).
- When FrameNet frame element definitions are not informative, make use of outside dictionaries to identify an appropriate hypernym, ensuring that there is an existing frame element with that name.

Each of these methods is described in detail in the following sections.

5.2 Regularizing Hypernymic Links Based on Frame Element Names

(Start here) The first steps in refining the frame element hierarchy are primarily editorial in nature and essentially involve examination of the frame element names. In what follows, we describe the application of each change proposed, with the approximate number of affected entries in the frame element dictionary. While each change was performed for all cases to which it applied, some changes give rise to others of a type that had already been completed. So, the number of cases for each type of change is only approximate.

Removal of multiple senses: Among the 1170 frame element names, 20 differ only in capitalization, such as **Affected_Party** and **Affected_party**. In DIMAP, these are treated as different senses. In these cases, the two senses were combined, a sense was deleted, any entries using these alternate capitializations as hypernyms were made consistent to a single spelling, and the number of occurrences for the two senses in frames was combined. Such cases are handled with a **Merge** operation (e.g., **Merge Legal_Basis with Legal_basis**), where the capitalization variant with the larger number of uses is selected.

Misspellings and editorial variation: These changes were applied to misspellings (**depictive** and **depicitive**) and editorial variation (**Duration_of_end_state** and **Duration_of_endstate**). For cases involving editorial variations, a choice was made as to the “more correct” underlying form. These changes were applied to five frame elements with a **Move** operation (e.g., **Move Depicitve under Depictive**), which merely establishes the hypernymic link. Seven frame elements were deleted (such as **Re-encoding**, **Sub-region**, and **Hot/Cold_source**) after combining them with more standard forms (**Re_encoding**, **Sub_region**, and **Hot_cold_source**). These were deleted using a **Merge** operation because the program used to create the digraph image was separating the dashed forms into two nodes.[7]

Frame elements with a number: Although 54 frame elements had been identified with a number “1″ or “2″ in their names (e.g., **Entity_1** and **Entity_2**), hypernymic link changes were made to only 28, i.e., using a base form without a number (using a **Move** operation to place it under the desired hypernymic link). In the remaining cases, the entries already contained a hypernym link, so these were not changed.

Plural forms: There are 52 frame elements in a plural form (ending in “s” or “a”). Of these, several did not have a singular form (e.g., **Tools**). Several had a singular form, but already had a hypernym induced from the frame-to-frame relations (e.g., **Members** was already linked to **Individuals**). Only 17 frame elements were linked to a singular frame element (e.g., **Recipients** linked to **Recipient**). These links also were made using the **Move** operation.

Underscore hypernyms: Approximately 460 frame elements have underscores in their names (e.g., **Dangerous_entity** and **Location_of_protagonist**). For many of these frame elements, the underscore can be interpreted as indicating a hypernymic link (e.g., **Dangerous_entity** is a kind of **Entity** and **Location_of_protagonist** is a kind of **Location**). Approximately 150 distinct potential hypernyms were identified and each was examined. Two crite-

[7] DIMAP generates a JPG image using the graph visualization software Graphviz.

ria were used for the addition of a hypernym link: (a) the frame element had no existing hypernym (induced from the frame-to-frame relations) and (b) the putative hypernym is a frame element. For example, of 28 frame elements ending in **_entity**, **Entity** is the hypernym link in 16. In all, 220 frame elements have hypernymic links generated in this way. These links also were made using the **Move** operation.

Many of the rules described in this section were applicable to more than one frame element, and the numbers may not reflect precisely the final results. Moreover, the rules were applied opportunistically, generally keeping in mind the general principle of keeping the hypernymic link induced from the frame-to-frame relations. We could follow the progress of the changes by frequently rerunning the digraph analysis, reducing the number of primitives, the number of frame elements without hypernymic links, and the number of cycles (strong components) in the digraph.

5.3 Examining the Frame Element Definitions for Hypernyms

A major step in laying out the frame element hierarchy involves an examination of the frame element definitions included in the characterizations of each frame in FrameNet. The examination of frame element definitions has been both productive and instructive.

As indicated earlier, frame elements with the same name may appear in many frames; the rigor of the FrameNet lexicographers in laying out a frame has resulted in a definition for each frame element. Thus, for example, there are 180 definitions of the **Agent** frame element. The FrameNet lexicographers are careful to point out that just because frame elements have the same name in different frames, the meaning of the frame element is not guaranteed to be the same. Nonetheless, as a working hypothesis, it is assumed that the meanings are the same. The practical effect of this hypothesis at the moment was that no detailed examination was made of the 180 definitions of the **Agent** frame element.

After applying the initial steps in analyzing the frame element digraph, there were many frame elements (identified as primitives) without hypernymic links. These frame elements are the ones whose definitions were examined. Frame element names are all nouns or noun phrases. In addition, many of the FE definitions use a typical noun hypernym that is present in the FE dictionary, e.g., **Required** is a **State_of_Affairs**. Since the FEs not used as hypernyms occur only infrequently in the frames, usually once or twice, these received the initial focus of our efforts. However, since many FEs used as hypernyms (i.e., as identified in previous steps) also occur infrequently, these were also analyzed. The general goal of this analysis of FE definitions was **to place the frame elements at an appropriate position in the frame element hierarchy**.

In carrying out this analysis,[8] i.e., changing the dictionary, we were able to immediately perform a new digraph analysis to determine the effect of any hypernym assignments, particularly to identify any new circularities that may have arisen as a result. When making hypernym assignments based on the FE definition, a general principle was to choose the most specific rather than the most general hypernym. This has the effect of spreading out the image, rather than having a large number of nodes directly adjacent to the primitive.

Approximately 250 hypernym assignments were based on the FE definitions. Many of these were quite straightforward, as in the following examples:

- **Route**: "The Route is the usual path that the Vehicle travels" (Hypernym **Path**)
- **Honoree**: "The person for whom the Social_event is held" (Hypernym **Person**)
- **Vividness**: "The degree of detail and/or immediacy of a remembered Experience (Hypernym **Degree**)

In making these assignments, the hypernym was immediately identified as being the name of another frame element in the dictionary, so there was little ambiguity in making the selection.

In examining the definitions, there were several frame elements that had no definitions. These frame elements came from frames that were themselves incomplete, i.e., they appeared to be frames in the process of development and not yet completely instantiated in FrameNet. These are handled with the **Delete** operation (e.g., **Delete**

[8] The tool used here will be identified in the final paper, with a link for its download.

Place_holder). This operation was applied to eight frame elements in FrameNet 1.3, but only one in FrameNet 1.5. Their deletion at this time does not appear to be problematic.

After about one-third of the FEs had been assigned hypernyms, the process became somewhat more difficult. Many FEs have definitions which are not really definitional, such as

- **Accuracy**: “This FE indicates the Accuracy of the prediction”
- **Trigger**: “The Trigger most commonly occurs in a PP Complement headed by _to_”
- **Case**: “The Case which a Trial is convened to decide”

Clearly, these definitions are not useful in identifying an appropriate hypernym. To deal with such situations, we used an ordinary dictionary, where the noun frame element names could be looked up and where their definitions would identify a noun hypernym. We used an online version of the *Oxford Dictionary of English* (ODE), in which each noun definition is also placed into a noun hierarchy. This was used rather than WordNet because its noun hierarchy was based on lexicographic principles (although guided by the tops of the WordNet nouns). For the most part, an appropriate hypernym was identified directly, but where necessary, we could traverse the hierarchy to identify an ancestor hypernym. Thus, for example, **Accuracy** is defined as "the degree to which … ", where **Degree** was used as the hypernym (via a **Move** operation). Similarly, **Hiatus** is defined as "a pause or break …", with the superordinate **interval** in the ODE noun hierarchy, linked immediately to **Period** (which was chosen as the appropriate hypernym.

5.4 Breaking Circularities in the Frame Element Hierarchy

In a digraph analysis, links between nodes may give rise to **strong components**, i.e., nodes that are mutually reachable. In the frame element hierarchy, 82 strong components (or circularities) were present in the initial digraph. To provide a strict hierarchy for frame elements, these circularities need to be broken. There are two primary methods that were used.

The first method uses the **Delink** operation. This operation essentially deletes one or more of multiple hypernyms induced by the frame-to-frame relations. Each frame element with more than one hypernym was examined to determine if one of the links could be deleted. In general, such a decision was based on whether a link (1) was clearly to a more specific frame element (e.g., **Entity2** as a hypernym of **Entity**) or (2) seemed completely unrelated to the essence of the frame element (e.g., **Entity** as a hypernym of **Goal**). Thus, we would have, e.g., the operation **Delink Entity2 from Entity**. After making several such changes, the structure of the digraph became increasingly more taxonomic, e.g., the largest strong component decreased in size from 133 down to 7.

The second method involves a more detailed examination of the circularities. In this method, it is necessary to identify the circularity and the relations that lead to it and then describe the rationale and changes to eliminate it (via a **Delink** operation). This generally involves examining the frame-to-frame relations (or mappings) between individual frame elements of two frames. However, as indicated earlier, some have been introduced as the result of other editing actions. These latter require a decision as to which of two or three frame elements is better used as the more primitive. A few examples will illustrate the type of reasoning involved.

(Protagonist_1, Protagonist_2): This cycle arises from the relation **Reciprocality_subordinate_event** PERSP_ON **Reciprocality**. While showing a valid type of inheritance, this relation indicates a sort of interchangeability between the two protagonists. We believe the mapping should preserve the order of the protagonists, and in any event, in the dictionary, both frame elements are simply given the hypernym **Protagonist**.

(Firearm, Instrument): In the relations **Shoot_projectiles** INHERITS **Intentionally_affect** and **Shoot_projectiles** USES **Cause_motion**, **Instrument** is the hypernym of **Firearm**. However, in **Hit_target** USES **Shooting_scenario**, **Firearm** is the hypernym of **Instrument**. While **Hit_target** clearly does not require that the instrument be a firearm, it is a slight error to have it use **Shooting_scenario** in which the instrument is always a firearm. The solution in this case is simply to eliminate **Firearm** as a hypernym of **Instrument**.

(Cause, Formational_cause): The link from **Formational_cause** to **Cause** was added in analyzing frame element names. The link from **Cause** to **Formational_cause** arose from the frame-to-frame relation **Body_mark** INHERITS **Entity** (which contains the Extra-thematic frame element **Formational_cause**). To break this circularity, **Formational_cause** was removed as the hypernym of **Cause**.

As the digraph moved closer to a complete taxonomy, it was necessary to make some judgments about the remaining circularities. The following cycles were among those involved: (**Type, Category**), (**Information, Message, Communication**), (**Theme, Patient, Entity**), (**Reason, Purpose**), and (**Act, Event, Action, Behavior**).

5.5 The Final Taxonomic Digraph

After making all the local decisions specified in the preceding sections, the frame element dictionary contains 1145 entries. The digraph analysis of this dictionary identifies 12 primitives: **Act**, **Cause**, **Degree**, **Entity**, **Path**, **Place**, **Purpose**, **Reason**, **Role**, **State**, **Topic**, and **Type**.[9] These are shown in Table 1, along with the number of descendants (including the primitive) in the trees rooted at these primitives. As the number of descendants suggests, this taxonomy is quite unbalanced, with over half the frame elements rooted at **Entity**, and several having very few descendants. For such primitives as **Degree**, **Path**, **Purpose**, **Reason**, and **Role**, there are very few ways in which nuanced characterizations of these frame elements can be stated. These primitives may be said to have face validity, i.e., they look elemental. However, they don't include many syntactic roles that are common in linguistics, such as **Agent**, **Theme**, **Experiencer**, **Instrument**, **Goal**, and **Time**. While these roles are present in the taxonomy, their absence as primitives warrants further discussion (see next section).

Primitive	Descendants
Act	112
Cause	29
Degree	16
Entity	605
Path	11
Place	95
Purpose	22
Reason	7
Role	2
State	86
Topic	130
Type	30
	1145

Table 2. Frame Element Primitives

[9] The digraph analysis based on FrameNet 1.3, containing 1004 entries, also had 12 primitives. It included **Instrument**, **Phenomenon**, and **Time** instead of **Act**, **Place**, and **Type**.

In comparing this taxonomy with the one generated from FrameNet 1.3, the additional 142 frame elements all folded into an existing structure. They represented elaborations and refinements of existing nodes. This suggests that, as FrameNet expands to cover more areas, any additional frame elements will be located under existing frame elements.

(Move these 2 sents below) One such final path is **(Act, Event, Action, Behavior, Misdeed, Crime, Offense, Wrong)**. The maximum depth of any path in the final digraph is 8 nodes.

6 The Nature of the Frame Element Taxonomy

To begin with, the frame element taxonomy does not constitute an ontology, i.e., it does not correspond to a characterization of the world, but rather only to the way in which frame elements are characterized within frame semantics (see Fillmore (1986)). Thus, frame elements are intended to capture some aspect of predicate-argument structure, so the taxonomy may be viewed as representing a putative comprehensive view of this structure.

Since the taxonomy is essentially a reflection of hypernymic links generated automatically based on the frame-to-frame relations, it provides a mechanism that can be used for assessing how well this comprehensive view of predicate-argument structure is captured and what changes may be warranted. This can begin with correcting the simple editorial inconsistencies noted in the analysis above. The next, more difficult step would involve a closer examination of the FrameNet frame element naming and the frame-to-frame relations. The taxonomy can facilitate these kinds of analyses.

As indicated above, we did not perform an exhaustive examination of all the frame element definitions. Such an examination can aid in ensuring whatever level of consistency is intended in FrameNet. One important aspect of such an analysis, however, is a need to make subtle distinctions in the frame elements, via a **Split** operation. As indicated above, the FrameNet lexicographers clearly state that just because frame elements from different frames have the same name, they do not necessarily have the same meaning. A simple example is the frame element **Score**, used in seven frames, with two meanings: a quantity for competitions and a musical score. By splitting **Score** into two senses, it would be possible to place them into distinct subtrees of the taxonomy.

One final issue concerning the taxonomy is the issue of what is used as its backbone. We have focused on the frame element names and used their noun-like status (with definitions consisting of a genus term and differentiae) to make decisions about hypernymic links using the genus term alone. We have not taken the differentiae into account. Further development of the taxonomy might consider the significance of these differentiae in characterizing predicate-argument structure.

7 Use of the Frame Element Taxonomy

We expect that a first use of the taxonomy will be an ability to make use of all FrameNet data in semantic role analyses, rather than restricting any analyses to smaller subsets because of the proliferation of frame element names. Secondly, we hope that the taxonomy will facilitate more rigorous analyses of predicate argument structure, rather than depending on *a priori* reasoning. Finally, the taxonomy may facilitate analyses of frame element realizations. Thus, for example, it may be possible to use the FrameNet data on valence realizations (phrase types, grammatical functions, and the actual strings) to characterize these realizations more fully.

To facilitate the use of the taxonomy, all of the data involved in this analysis will be freely available.[10] The data includes the raw data from which the dictionary was created and the frame element dictionary itself, with an entry for each frame element containing its hypernymic link, its number of instantiations in FrameNet frames, and a list of the frames where it is used. The data also includes a MySQL database consisting of three tables: the frame element hierarchy, the definitions of each frame element in all its frames, and all the operations used in moving from the initial digraph to the final taxonomy. Accompanying the data is an image (4.5 MB) of the final digraph, i.e., the taxonomy tree rooted in the 12 primitives.

As indicated above, the decisions made about the placement of specific nodes can be questioned. By providing all the data involved in the taxonomy's creation, it should be possible to make modifications to suit specific needs. An online version of the taxonomy will portray its current status and also provide viewers to suggest changes using the five operators described above.

8 Conclusions

We have presented methods for creating a taxonomy of FrameNet frame elements based on an analysis of FrameNet frame-to-frame relations. We suggest that this taxonomy provides a vehicle for improved analyses of semantic roles. Unlike many other studies of semantic roles, we have made no *a priori* judgments about these roles, but have used a data-driven approach rooted in the corpus-based data of FrameNet. We believe that the successful ability to create a taxonomy is based on making local decisions, first from the decisions of the FrameNet lexicographers in tagging corpus instances and second from specific examination of the hypernymic links induced from the frame-to-frame relations.

By making the data available for download and viewing online, we believe that we have provided a capability for further *a posteriori* analyses, not only on the taxonomy but in studies that make use of the taxonomy. We suggest that one such use may be a detailed examination of the valence realizations of frame elements using the FrameNet data.

Acknowledgments

To be included in the final version.

References

[10] Appropriate links will be provided in the final version of the paper.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLING-ACL 1998*, pp. 86-90.

David Dowty. 1991. Thematic Proto-roles and Argument Selection. *Language,* 67(3), pp. 547-619.

Charles J. Fillmore. 1982. Frame Semantics. In *Linguistics in the Morning Calm***.** Hanshin Publishing Co., Seoul, South Korea, pp. 111-137.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of semantic roles. *Computational Linguistics*, 28(3):245-288.

Y. Matsubayashi, N. Okazaki, and J. Tsujii. 2009. A Comparative Study on Generalization of Semantic Roles in FrameNet. In *Proceedings of the 47th Annual Meeting of the ACL*, pp. 19–27, Suntec, Singapore.

M. McConville and M.O. Dzikovska. 2008. Using inheritance and coreness sets to improve a verb lexicon harvested from FrameNet. In *Proceedings of the Linguistic Annotation Workshop*, pp. 33–40.

Tom O'Hara and Janice Wiebe. 2009. Exploiting Semantic Role Resources for Preposition Disambiguation. *Computational Linguistics*, 35(2):151-184.

Ekaterina Ovchinnikova, Laure Vieu, Alessandro Oltramari, Stefano Borgo, and Theodore Alexandrov. 2010. Data-driven and Ontological Analysis of FrameNet for Natural Language Processing. In *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC)*, 3157-3164.

Josef Ruppenhofer, Manfred Pinkal, and Jonas Sunde. 2010. Generating FrameNets of various granularities. In *Proceedings of the seventh international conference on Language Resources and Evaluation (LREC)*, 2736-2743.