FrameNet DIMAP Dictionary

The FrameNet 1.5 data have been converted into an alphabetic dictionary that can be viewed and manipulated with the CL Research Dictionary Maintenance Program (DIMAP). (A free demonstration version of DIMAP is available from CL Research.) This dictionary contains 11053 entries, with 8568 entries for lexical items (many having multiple senses with different parts of speech) and 2485 entries that encode the frames and frame relations. This dictionary was created with FrameNet Explorer for Windows, also available at CL Research. Details about the FrameNet data can be found through the main FrameNet site.

For details on each type of lexical entry, see:
- FrameNet Lexical Items
- FrameNet Frame Entries
- FrameNet Frame Relation Entries

This dictionary and earlier versions based on FrameNet 1.3 and 1.4a have been used in three Senseval and SemEval tasks in 2004, 2007, and 2010:
- Explorations in Disambiguation Using XML Text Representation (Senseval-3)
- CLR: Integration of FrameNet in a Text Representation System (SemEval 2007)
- CLR: Linking Events and Their Participants in Discourse Using a Comprehensive FrameNet Dictionary (SemEval 2010)

The use of FrameNet in these evaluation exercises was somewhat preliminary. Further work on the exploitation of FrameNet for each of these exercises is continuing and has reached levels comparable to those of other participants, demonstrating the utility of the DIMAP FrameNet dictionary.

FrameNet Lexical Items

DIMAP provides sufficient capability to capture all aspects of the FrameNet data in various types of built-in data structures. First, it is necessary to capture each lexical unit and to create a distinct sense for each frame in which a lexeme is used. For FrameNet 1.5, the DIMAP dictionary contains 8568 entries, with many entries having multiple senses. For each sense, the FrameNet part of speech, definition, frame name, ID number, definition source, and associated frame element realizations (if available) are captured from the FrameNet files. The DIMAP dictionary is created using **FrameNet Explorer** (freely available from CL Research).

FrameNet Explorer uses the FrameNet file **frameIndex.xml** to identify the lexical units and their associated **lu*.xml** files. The lexical unit ID is also identified in this process and is used to load the appropriate lexical unit files. When uploaded into DIMAP, a lexical unit entry is an ordinary DIMAP entry, consisting of an entry name and one or more senses, each with its own part of speech. The definition is extracted from the **lu*.xml** file. Most definitions in the FrameNet data are preceded by a source code ("COD" for *Concise Oxford Dictionary* and "FN" for *FrameNet*). This code is removed from the definition. The definition is entered in the DIMAP definition field. The frame, the source code, and the ID number are entered as DIMAP features. If the lexical unit has an incorporated frame element (e.g., Garment for *abaya*), this is captured in a DIMAP feature **INC** with a value equal to the frame element name.

If the FrameNet lexical entry contains frame element realizations, this information is also captured in the appropriate sense. In DIMAP, this is done in an attribute-value feature structure. Each non-empty feature element realization in the FrameNet data is captured. A DIMAP feature **attribute** is constructed as a conflation of the phrase type and the grammatical function, e.g. "NP(Dep)". The feature **value** is a conflation of the valence unit frame element name and the number of annotations in the FrameNet corpus, e.g., "Cognizer (28)". This manner of capturing FrameNet information is done to facilitate processing; the DIMAP feature structure is frequently used to access information

about lexical items. Experience in using this format in Senseval and SemEval is described in the papers listed under [FrameNet DIMAP Dictionary](#). A DIMAP sense may have several of these features.

The lexical units are encountered in conjunction with the traversal of the frames. The same lexical unit may be encountered in another frame, with different properties. When this data is uploaded into a DIMAP dictionary and an existing entry has already been created, the new data is added as another sense.

FrameNet Frame Entries

A DIMAP frame entry is a lexical entry distinguished by having a leading **#** followed by the frame name, e.g., **#Manufacturing**. The frame name is taken exactly as given in the "name" attribute of the <u>frame</u> node in **frameIndex.xml**, i.e., it is capitalized and may contain underscores. Each frame entry consists of one sense, which is given "none" as the part of speech. The frames identified in **frameIndex.xml** are used to extract information from the ***.xml** files in the **./frame** directory of the standard FrameNet distribution.

Frames are captured in the [same dictionary](#) as lexical units. However, they are not treated as lexical units, but rather as "meta-entries". In the DIMAP dictionary, frame names are entered as dictionary entries beginning with the symbol "#". In these entries, different data structures of a DIMAP entry are used to capture the different kinds of information about the frames.

Each frame entry has a set of DIMAP features that identify the frame's frame elements and their core type ("Core", "Core-Unexpressed", "Peripheral", or "Extra-thematic"). The frame element is the feature attribute and the core type is the feature value. These are extracted from the "name" and "coreType" attributes of the <u>FE</u> nodes from the ***.xml** files in the **./frame** directory.

Each frame entry has a set of DIMAP instances that provide the lexical units for the frame. The lexical units are identified from the <u>lexUnit</u> children of the frame. The "name" attribute of each child is dissected to remove its part of speech code (a period and part of speech at the end of the "name" attribute). The "ID" attribute is identified as the sense number in DIMAP instances; this number corresponds to the lexical unit file for this entry, i.e., "lu**n**.xml", where **n** is the ID number. The existence of a number does not imply that the FrameNet data contain any annotations for the lexical unit.

If a frame participates in any frame-to-frame relations, as identified in the file **frRelation.xml**, these are captured in DIMAP roles. Each DIMAP role consists of a name and one or more links. For example, the "#Aggregate" frame has two roles, IS_INHERITED_BY with links to #Store and #Organization, and HAS_INCHOATIVE with a link to #Come_together. To obtain the data for these roles, an XPath expression is used to identify the relevant nodes. This XPath consists of a concatenation of "//frameRelationType[@name=\"", the relation name, "\"]//frameRelation[@", the relation direction ("super" or "sub"), "FrameName=\"", the frame name, and "\"]". The relation name is one of the frame relation types ("Inheritance", "Subframe", "Using", "Inchoative_of", "Causative_of", "Precedes", or "Perspective_on"); the frame relations "See_also" and "ReFraming_Mapping" are not further analyzed. Each relation type, except "Perspective_on", gives rise to two entries: the direct relation and its inverse, as follows:

- For frame relation type **Inheritance**, the role names are INHERITS and IS_INHERITED_BY (e.g., #Change_of_consistency INHERITS #Event and #Event IS_INHERITED_BY by #Change_of_consistency)
- For frame relation type **Subframe**, the role names are IS_SUBFRAME_OF and HAS_SUBFRAMES (e.g., #Precipitation IS_SUBFRAME_OF of #Weather and #Weather HAS_SUBFRAMES #Precipitation)
- For frame relation type **Using**, the role names are USES and IS_USED_BY (e.g., #Time_vector USES #Direction and #Direction IS_USED_BY #Time_vector)
- For frame relation type **Inchoative_of**, the role names are INCHOF and HAS_INCHOATIVE (e.g., #Change_position_on_a_scale INCHOF #Position_on_a_scale and #Change_position_on_a_scale

HAS_INCHOATIVE #Position_on_a_scale)

- For frame relation type **Causative_of**, the role names are CAUSOF and IS_CAUSED_BY (e.g., #Cause_to_move_in_place CAUSOF #Moving_in_place and #Moving_in_place IS_CAUSED_BY #Cause_to_move_in_place)
- For frame relation type **Precedes**, the role names are PRECEDES and IS_PRECEDED_BY (e.g., #Process_end PRECEDES #Process_completed_state and #Process_completed_state IS_PRECEDED_BY #Process_end)
- For frame relation type **Perspective_on**, the role name is PERSPON (e.g., #Obligation_scenario PERSPON #Being_obligated)

The relations **Distinguishes**, **Has CoreSet**, **Has Excludes**, and **Has Requires** do not occur in the FrameNet frame-to-frame relation set in "frRelation.xml" and hence, are not presently captured in the DIMAP FrameNet dictionary.

Frame-to-Frame Relation Entries

A DIMAP frame-to-frame relation entry is a lexical entry distinguished by having a leading **@** followed by a frame name, a relation name, and another frame name, e.g., **@PrecipitationIS_SUB_OFWeather**. These entries are derived from the file **frRelation.xml**. The frame-to-frame relation entries capture the mapping of frame elements between two frames, as specified in the FrameNet data. In particular, the entries capture a hypernymic relationship between the first frame and the second frame.

To obtain the data for these roles, an XPath expression is used to identify the relevant nodes. This XPath consists of a concatenation of "//frameRelationType[@name=\"", the relation name, "\"]//frameRelation. The relation name is one of the frame relation types ("Inheritance", "Subframe", "Using", "Inchoative_of", "Causative_of", "Precedes", or "Perspective_on"); the frame relations "See_also" and "ReFraming_Mapping" are not further analyzed. Unlike the **frame entries**, the relations are analyzed in one direction only; the inverse relations are not considered in constructing these entries. The relation names captured in these entries is one of the following: INHERITS, IS_SUB_OF, USES, INCH_OF, CAUSE_OF, PRECEDES, or PERSP_ON.

Frame-to-frame relation entries are captured in the **same dictionary** as lexical units. However, they are not treated as lexical units, but rather as "meta-entries". In the DIMAP dictionary, frame-to-frame relation entries are entered as dictionary entries beginning with the symbol "@". Each frame-to-frame relation entry consists of one sense, which is given "none" as the part of speech. The mapping is captured in two DIMAP features. The first feature has the name **0** (corresponding to *subframe*) and has a feature value consisting of the **frame element** names of the first frame in the entry name. The second feature has the name **1** (corresponding to *superframe*) and has a feature value consisting of the **frame element** names of the second frame in the entry name. These frame element names are derived from the FERelation children of each frameRelation node under each frameRelationType node of **frRelation.xml**. As a result, the number of frame elements associated with each DIMAP feature **0** and **1** is the same.

The number of frame elements included in the mapping may not correspond to the number of frame elements for a given frame. This simply indicates that the FrameNet lexicographers mapped only a subset of the frame elements. For those frame elements that were mapped, the mapping establishes a hypernymic relation for frame elements positionally the same. That is, a frame element in feature **1** may be construed as the hypernym of the frame element in the same position in feature **0**. In the majority of cases, the frame elements are identically named and hence are equivalent. When this is not the case, the mapping may be construed as hypernymic in nature. For example, #Locative_relation INHERITS #State (in the DIMAP entry **@Locative_relationINHERITSState**), the subordinate frame element **Figure** (included under feature **0**) inherits the superordinate frame element **Entity** (included under feature **1**). Note that #Locative_relation has 6 frame elements and #State has 2 frame elements, so only one frame element has been mapped in this instance.

The creation of these entries is intended to provide a capability for users who may wish to carry out mappings between frames. CL Research used these entries as the basis for creating a frame element taxonomy. A detailed description of the steps followed in the analysis and construction of the dictionary is available at CL Research's blog (**The Clog**). The blog entries were posted over a two month period, with the final entry describing the end result. To view the steps in the order in which they were performed, use the following links:

- Frame element hierarchy
- Analyzing the frame element digraph: Initial steps
- Circularities in the frame element hierarchy
- Refining the frame element hierarchy: Application of initial steps
- Analyzing frame element definitions

FrameNet Dictionary Feedback

To report any bugs, request new or enhanced features, obtain product help or documentation, ask a question, make a comment, or request further information from CL Research, send feedback to CL Research (http://www.clres.com/db/feedback.php?clrdict=fndict).