SemLink

1 BASIC INFORMATION

1.1 Resource composition

SemLink (Palmer, 2009) provides a mapping between complementary lexical resources. In the current release, two mappings are available:
- a mapping between VerbNet (Kipper et al, 2008) and PropBank (Palmer et al, 2005)
- a mapping between VerbNet and FrameNet (Fillmore et al, 2008). The version of FrameNet used is v1.2

1.2 Representation of the resource (flat files, database, markup)

The mappings are encoded within a set of XML and plain text files

1.3 Character encoding

The characters are UTF8 encoded.

2 ADMINISTRATIVE INFORMATION

2.1 Contact person

Name: Martha Palmer
Address: Department of Linguistics, University of Colorado at Boulder
295 UCB Boulder, Colorado 80309-0295
Affiliation: Department of Linguistics, University of Colorado at Boulder
Position: Professor
Telephone: + 1 (303) 492-1300
Fax: + 1 (303) 492-4416
e-mail: martha.palmer@colorado.edu

2.2 Delivery medium (if relevant; description of the content of each piece of medium)
The resource is available on the MetaShare platform as an archive.

2.3 Copyright statement and information on IPR
The resource is freely available for research purposes, according to the terms of the licence included within the archive.

3 TECHNICAL INFORMATION

3.1 Directories and files

The archive contains the directory *semlink1.1*. It contains the following sub-directories and files:

- *vn-fn* – directory containing mappings between VerbNet and FrameNet. The following files are contained within this directory:
  - *VNclass-FNframeMappings.xml* – provides mappings between VerbNet verbs and FrameNet lexical entries
  - *VN-FN_roleMapping.xml* – contains the possible role correspondences between VerbNet thematic roles and FrameNet frame elements.
  - *README.txt* - a file providing more details about the mappings contained within the above two files and their XML encoding
- *vn-pb* - directory containing mappings between VerbNet and PropBank. The following files are contained within this directory:
  - *type_map.xml* – specifies potential mappings between PropBank rolesets and VerbNet classes for a given lemma
  - *vnprop.txt* – specifies the correct mapping between PropBank roleset and VerbNet class for each predicate in the PropBank corpus. Contains only VerbNet role labels.
  - *vnpbprop.txt* – same as *vnprop.txt* except that PropBank role labels are also included
  - *mapping_stats.txt* – Provides general statistics regarding mappings between PropBank and VerbNet argument types
  - *README.TXT* – Provides information about the VerbNet-PropBank mapping files.
- LICENSE.TXT – provides details of the licence that must be followed when using SemLink
- README.TXT – Provides general information about SemLink

3.2 Data structure of an entry

This is not relevant as the corpus is a set of text files.

3.3 Resource size (nmb. of tokens, MB occupied on disk)

The SemLink directory requires 27.2MB of disk space. The following statistics provide information about the size of the resource:

- The mapping between VerbNet verbs and FrameNet frames considers 4755 VerbNet verbs.
  - Mappings between 2168 of these VerbNet verbs and FrameNet lexical units have been found.
- According to VerbNet/FrameNet mappings identified, 598 mappings are provided between sets of thematic roles in VerbNet classes and FrameNet frame elements

- The VerbNet-PropBank mappings consider 1881 lemmas from the PropBank corpus, leading to a total of possible 2665 mappings between PropBank rolesets and VerbNet classes.
- 112,917 occurrences of predicates appearing in the PropBank corpus have links to appropriate VerbNet classes.

4 CONTENT INFORMATION

4.1 Type of the resource (language (in)dependent)

The resource is language-dependent

4.2 The natural language(s) of the corpus

The language of the corpus is English.

4. 3 Domain(s)/register(s) of the resource

The resource is concerned with language behavior in general English texts.

4.4 Annotations in the resource (if an annotated resource)

4.4.1 Types of annotations

Mappings exist at several levels, i.e. classes of verbs/predicates, individual members of these classes, and at the level of their arguments. See next section for further details.

4.4.2 Tags (if POS/WSD/TIME/discourse/etc –tagged or parsed),

The VerbNet-FrameNet mapping consists of two types of information, in separate files.
1. Possible mappings between individual verbs belonging to VerbNet classes and individual lexical entries belonging to FrameNet frames (*VNclass-FNframeMappings.xml*). The mapping is many-many: VerbNet verbs can map to multiple lexical entries in FrameNet, and vice versa.
2. For all VerbNet classes and FrameNet frames that are linked according to the mappings above, mappings are provided between the thematic roles in the VerbNet classes and the frame elements in the FrameNet frames (*VN-FN_roleMapping.xml*)

The VerbNet-PropBank mapping also consists of 2 types of information, in separate files.

1. For each lemma that occurs in PropBank, each possible roleset is listed, together a mapping to an appropriate class in VerbNet (*type_map.xml*) Appropriate mappings between the arguments of the PropBank roleset and corresponding thematic roles in VerbNet are also shown.

2. For each verbal occurrence in the PropBank corpus, the appropriate PropBank roleset-VerbNet class mapping, according to the possible mappings listed in *type_map.xml*. This information is contained in *vnprop.txt* and *vnpbprop.txt*. These files also include mappings between PropBank argument labels and VerbNet thematic role labels.

4.4.3 Alignment information (if the resource contains aligned documents: level of alignment, how it was achieved)

Not applicable – this is a monolingual resource.

4.4.4 Attributes and their values (if annotated)

VerbNet-FrameNet mapping

VNclass-FNframeMappings.xml

This file lists elements that include the following attributes:

  class   -- VerbNet class ID (numeric)
  vnmember -- VerbNet class member (string, the verb lemma)
  fnframe -- FrameNet Frame (string)
  fnlexent -- FrameNet lexical entry ID (numeric)
  versionID -- VerbNet version ID (either 1.5 or 2.0)

The structure of an element can be demonstrated by this example:

  <vncls   class='9.1-2'   vnmember='put'   fnframe='Placing' fnlexent='5355' versionID='vn2.0' />

Note that the string values for 'fnframe' and 'vnmember' attributes can include hyphens and underscores.

There are two special values possible for the 'fnframe' attribute:

  fnframe='DS' -- Different Sense
  fnframe='NA' -- Not Available

"Different Sense" covers cases where a particular VerbNet lemma exists as the word form of one or more Lexical Units in FrameNet, but none that share the lexical semantics of the VerbNet Class member closely enough.

"Not Available" covers cases where a VerbNet lemma doesn't exist as a word form at all in FrameNet.

The 'fnlexent' attribute provides the lexical entry ID number FrameNet assigned to the verb.

VN-FN_roleMapping.xml

The thematic role / frame element mapping file includes the possible role correspondences for the VerbNet Classes and FrameNet Frames that have been mapped. The number of role mappings depends on the particular Class and Frame, and so will vary in number.

The two new attributes included here are the following:

  fnrole   -- FrameNet frame element (string)
  vnrole   -- VerbNet thematic role (string)

The structure of an element can be demonstrated by this example:

```
<vncls class='9.1' fnframe='Placing'>
 <roles>
   <role fnrole='Agent' vnrole='Agent'/>
   <role fnrole='Cause' vnrole='Agent'/>
   <role fnrole='Goal' vnrole='Destination'/>
   <role fnrole='Theme' vnrole='Theme'/>
 </roles>
</vncls>
```

Note that the string values for the 'fnframe' attribute can include hyphens and underscores.

VerbNet-PropBank mapping

type_map.xml

The type mapping is provided as a single xml file, containing entries of the form:

```
<predicate lemma="muzzle">
 <argmap pb-roleset="muzzle.01" vn-class="9.9">
   <role pb-arg="1" vn-theta="Destination" />
```

```
      <role pb-arg="0" vn-theta="Agent" />
      <role pb-arg="2" vn-theta="Theme" />
    </argmap>
    <argmap pb-roleset="muzzle.01" vn-class="22.4">
      <role pb-arg="1" vn-theta="Patient1" />
      <role pb-arg="0" vn-theta="Agent" />
      <role pb-arg="2" vn-theta="Patient2" />
    </argmap>
  </predicate>
```

Each <predicate> entry describes a single verb lemma, and contains one or more <argmap> entries. Each <argmap> entry describes the mapping between arguments for a specific (PropBank roleset, VerbNet class) pair, using one or more <role> entries. Each <role> entry describes the mapping between PropBank ARGn labels and VerbNet thematic roles for a single argument role.

vnprop.txt / vnpbprob.txt

These files provide the mappings between Propbank rolesets and VerbNet classes at the level of individual predicates occurring within the PropBank corpus.

Both files use the same format as PropBank's prop.txt file. In particular, each line describes a single predicate and its arguments. The columns are as follows:

 wsj-filename sentence terminal tagger verb inflection arguments...

Where:

- 'wsj-filename' is the name of the file in merged penn treebank, wsj section

- 'sentence' is the number of the sentence in the file (starting with 0)

- 'terminal' is the number of the terminal in the sentence that the location of the verb. Note that the terminal number counts empty constituents as terminals and starts with 0. This will hold for all references to terminal number in this description.

- 'tagger' is the name of the annotator who performed the mapping.

- 'verb' is a token identifying the verb's PropBank roleset and VerbNet class. It has the form <roleset>;VN=<vncls> where <roleset> is a PropBank roleset and <vncls> is a VerbNet class number.

- 'inflection' consists of 5 characters representing person, tense, aspect, voice, and form of the verb, respectively. See the PropBank documentation for details.

- 'arguments...' is a string representing the annotation associated with a particular argument or adjunct of the proposition. Each proplabel is dash '-' delimited and has the following columns

  1. column for the 'syntactic relation'. See the PropBank documentation for details.

  2. column for the 'label'. In vnprop.txt, this will consist of a VerbNet thematic role label (Agent, Patient, etc); or a PropBank role (ARG0, ARG1, etc) if the role does not have an appropriate mapping target in VerbNet. In vnpbprop.txt, this will have the form "ARG<n>[<theta>]", where <n> is a PropBank role number and <theta> is a VerbNet thematic role; or simply "ARG<n>" if there is no appropriate mapping target. The label "rel" is used to mark the position of the relation word (i.e., the verb).

  3. column for feature. See the PropBank documentation for details.

4.5 Intended application of the resource

The SemLink mappings provide a means to make future improvements to semantic role labeling systems, and will benefit Question-Answering, Information Extraction, inferencing and other NLP applications.

4.6 Reliability of the annotations (automatically/manually assigned) – if any

The type-to-type mappings are manually assigned. Token to token mappings were created automatically using the type-to-type mappings and are in the process of being hand-corrected.

5 RELEVANT REFERENCES AND OTHER INFORMATION

Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck (2003) Background to FrameNet. *International Journal of Lexicography*, 16:235– 250.

Kipper, Karin, Anna Korhonen, Neville Ryant, and Martha Palmer (2008) A large-scale classification of English verbs. *Language Resources and Evaluation Journal*, 42(1):21–40.

Palmer, Martha (2009) Semlink: Linking PropBank, VerbNet and FrameNet In *Proceedings of the 5th International Conference on Generative Approaches to the Lexicon*. Sept. 2009, Pisa, Italy: GenLex-09, 2009

Palmer, Martha, Daniel Gildea, and Paul Kingsbury (2005). The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Further information about FrameNet is available at: https://framenet.icsi.berkeley.edu/fndrupal/

Further information about SemLink is available at: http://verbs.colorado.edu/semlink/. The unified verb index makes use of these mappings: http://verbs.colorado.edu/verb-index/

Further information about VerbNet is available at: http://verbs.colorado.edu/~mpalmer/projects/verbnet.html