Introduction to Semantic Role Labelling

Behrang QasemiZadeh
zadeh@phil.hhu.de
Computational Linguistics Department, HHU – DRAFT

October 2018–January 2019

Language Resources for Semantic Role Labeling

Language resources for Semantic Role Labeling:

- FrameNet
- PropBank
- VerbNet
- ...

VerbNet

VerbNet is another language resource that implements a hierarchy of English verbs similar to Levin's verb classes.

Compared to Levin's work, firstly, VerbNet is a machine readable resource (for most part).

Additionally, VerbNet has a better coverage of lexical items (more than 4000 lemmas), which are classified to a set of more refined classes (VerbNet doubles Levin's classes to more than 400) in a hierarchy of depth 4.

VerbNet (contd.)

Palmer states several motivations for building VerbNet (when it is compared to Levin classes)"

- Levin's classes are not *semantically* homogeneous: unrelated verbs can be found in one category (something you can investigate);
- A verb can have multiple class listings in which the cause is not known (e.g., polysemy or homonymy?);
- Sometimes alternation patterns are contradictory (Carry verbs disallow Conative but it contains verbs that license Conative alternation);

VerbNet (contd.)

By contrast, VerbNet aims for a more coherent classification of verbs based on their syntactic and semantic properties, mainly by explicitly encoding information about

- verb's syntactic patterns;
- their semantic roles;
- and, relations between senses.

The main purpose is to have a coherent classification based on observable syntactic and semantic behaviors which can facilitate acquisition of new class members.

VerbNet (contd.)

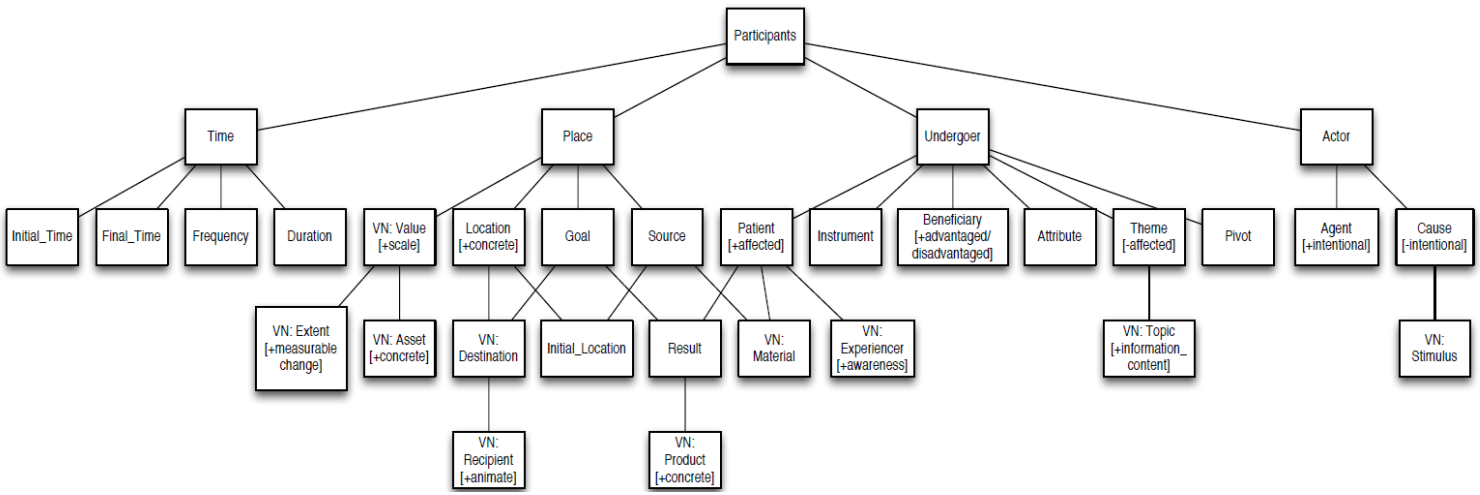
Most importantly, VerbNet assigns to each class, alongside its syntactic frames, an argument-structure pattern consists of a set of semantic role labels, e.g., Agent REL Patient, or Patient REL (for verb *break*).

To do so, VerbNet started from the Levin's 47 top/coarse categories, but additional specification of verbs resulted in a hierarchy that does not necessarily fit Levin's classes (e.g., one class in Levin hierarchy can be represented as two different sub-classes in VerbNet).

As implied, the specification of classes are based on syntactic and semantic properties of verbs and their arguments.

VerbNet (contd.)

Our textbook (Palmer et al., 2010) states that 24 semantic role categories are used in VerbNet: It is slightly out-of-dated. VerbNet 3.2 uses 33 thematic roles.



Inside the VerbNet

Similar to FrameNet, VerbNet can be seen as a dictionary, in which verb classes are the entries.

The main component of this dictionary are verb-classes, their members (which are sense-tagged verbs), a number of thematic roles, VerbNet frames (syntactic and semantics).

These all have been made available in a set of XML files, one per verb class (download from `https://verbs.colorado.edu/verbnet/`).

Following is a more detailed description.

Inside the VerbNet (contd.)

Each entry in VerbNet has a *class id*, (e.g., `try-61`, which indicates the verb class) and it consists of several components:

MEMBERS: are the verbs that belong to this class. Each verb, is linked to *WordNet* and it has a sense grouping identifier (from the OntoNote project).

For instance, the verb *try* in `try-61` is mapped to WordNet's entry "try2:41:00" and OntoNote grouping of "try.01".

```
<MEMBER name="try" wn="try2:41:00" grouping="try.01" />
<MEMBER name="intend" wn="intend2:31:00" grouping="intend.01" />
```

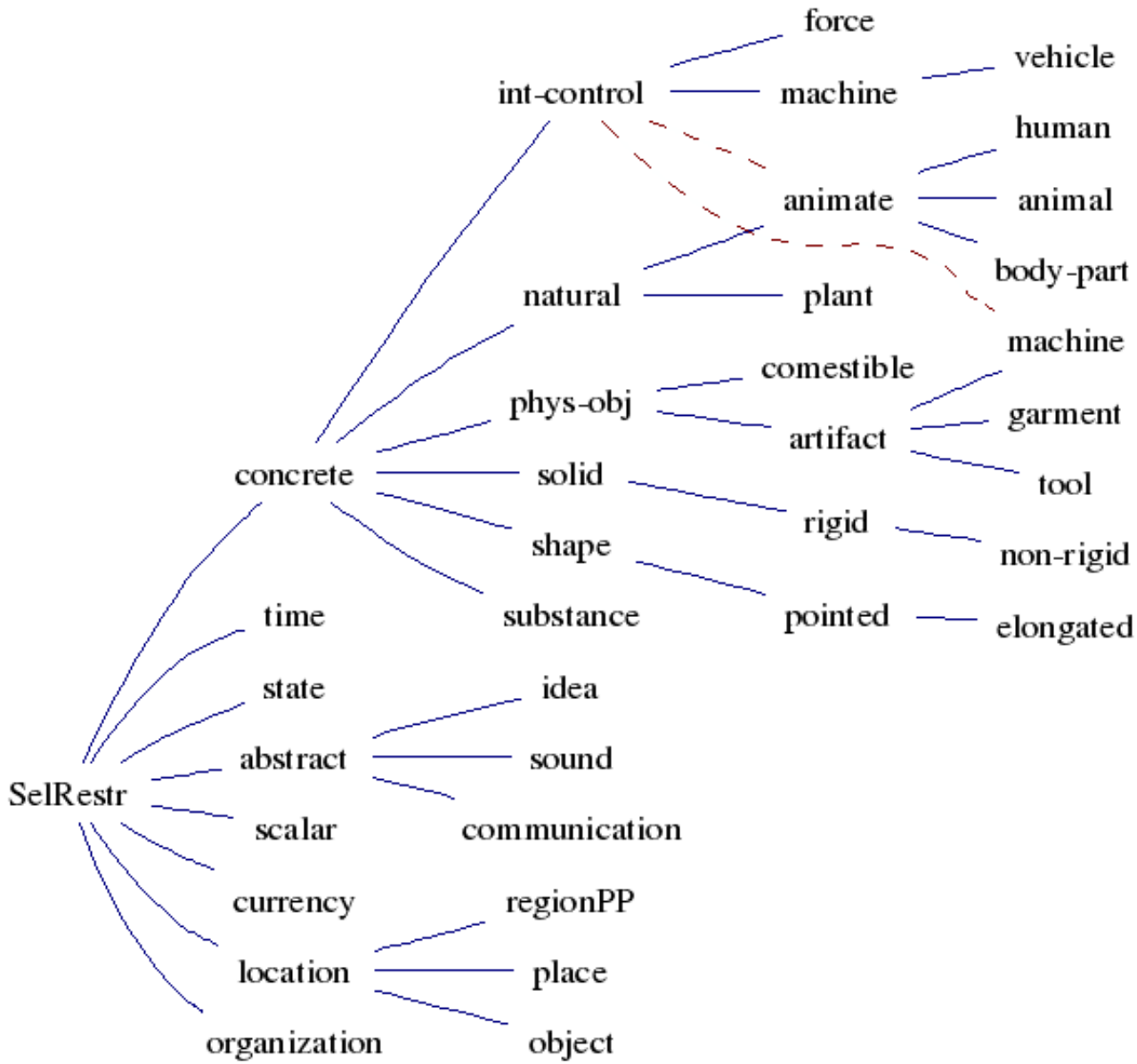
Inside the VerbNet (contd.)

Thereafter, thematic roles and their associated selectional restrictions are listed (from the EuroWordNet Project, 39 of them). (These roles are used in frames that are associated to the verb class.)

For instance, for `try-61`, the thematic roles are `Agent`, and `Theme`:

```
<THEMROLES>
        <THEMROLE type="Agent">
            <SELRESTRS logic="or">
                <SELRESTR Value="+" type="animate" />
                <SELRESTR Value="+" type="organization" />
            </SELRESTRS>
        </THEMROLE>
        <THEMROLE type="Theme">
            <SELRESTRS/>
        </THEMROLE>
</THEMROLES>
```

Inside the VerbNet (contd.)



Inside the VerbNet

VerbNet defines a set of frames per verb class (i.e., each entry contains one or more frames).

Each frame provide a rich syntactic and semantic characterization of the verb class and its members.

Each frame consist of a *description*, *example(s)*, a syntactic frame, and a semantic predicates.

Inside the VerbNet (contd.)

DESCRIPTION: Contains some basic information regarding the surface syntactic realization of the verb class.

The primary description uses a "phrasal-part-of-speech-tag-based" pattern representation while the secondary description is more similar to Levin's alternations (e.g., transitive, intransitive, resultatives, etc.) (and sometimes XTAG trees).

```
<FRAME>
    <DESCRIPTION descriptionNumber="8.1"
    primary="NP V S_ING" secondary=" ING-NP-OMIT" xtag="0.2" />
    <EXAMPLES>
        <EXAMPLE>I tried exercising.</EXAMPLE>
    </EXAMPLES>
    ...
```

Inside the VerbNet (contd.)

Syntactic Frames are the surface realizations of the argument structure which are encoded in terms of thematic roles (which are sometimes accompanied by syntactic restriction information (constraints on the form) :

```
<SYNTAX>
    <NP value="Agent">
        <SYNRESTRS/>
    </NP>
    <VERB/>
    <NP value="Theme">
        <SYNRESTRS>
            <SYNRESTR Value="+" type="np_omit_ing" />
        </SYNRESTRS>
    </NP>
</SYNTAX>
```

Inside the VerbNet (contd.)

Syntactic frames are accompanied by semantic predicates.

Each semantic predicate has a name and a number of arguments (from the thematic roles inventory). Additionally, to assert temporal relations, a special variable *e* (denoting the underlying Event) is added to predicates. Simple example:
Attampt(**During**(e), AGENT, THEME):

```
<SEMANTICS>
    <PRED value="attempt">
        <ARGS>
            <ARG type="Event" value="during(E)" />
            <ARG type="ThemRole" value="Agent" />
            <ARG type="ThemRole" value="Theme" />
        </ARGS>
    </PRED>
</SEMANTICS>
```

Inside the VerbNet (contd.)

However, semantic frames can contain more than one predicate in the form of conjunctive predicates. E.g. (guess what):

CAUSE(Agent, e) ∧ CONTACT(DURING(e), ?Instrument, Patient) ∧ PHYSICALFORM(RESULT(e), Form, Patient)

```
<SEMANTICS>
    <PRED value="cause">
        <ARGS>
            <ARG type="ThemRole" value="Agent" />
            <ARG type="Event" value="E" />
        </ARGS>
    </PRED>
    <PRED value="contact">
        <ARGS>
            <ARG type="Event" value="during(E)" />
            <ARG type="ThemRole" value="?Instrument" />
            <ARG type="ThemRole" value="Patient" />
        </ARGS>
    </PRED>
    <PRED value="physical_form">
        <ARGS>
            <ARG type="Event" value="result(E)" />
            <ARG type="VerbSpecific" value="Form" />
            <ARG type="ThemRole" value="Patient" />
        </ARGS>
    </PRED>
</SEMANTICS>
```

Inside the VerbNet (contd.)

(1) Tony bent the rod.

CAUSE(Tony, BENDING-EVENT) ∧
CONTACT(DURING(BENDING-EVENT), _, rod) ∧
PHYSICALFORM(RESULT(BENDING-EVENT), bend, rod)

For an overview of the current VerbNet's controlled vocab please see `http://verbs.colorado.edu/vn3.2.3-test-uvi/vn/reference.php`.

VerbNet can be browsed online:
`http://verbs.colorado.edu/verb-index/index.php`

The class hierarchy can be viewed here: `http://verbs.colorado.edu/vn3.2.3-test-uvi/vn/class-h.php`

SemLink: A wish yet to fulfill

We have all these valuable resources (FrameNet, PropBank, VerbNet, WordNet, NomBank, and so on) but they are like islands with no bridges between them.

SemLink is an effort to address this need: Making a unified index over all these resources (I guess it is a wish yet to be fulfilled).

SemLink for PropBank–VerbNet Mappings: According to Palmer et al. (2010), SemLink builds two types of mapping, one at a lexical level, another at an instance level.

Lexical mapping gives potential PropBank–VerbNet mappings for a given verb *out-of-context* and not for their occurrences in sentences.

SemLink: A wish yet to fulfill (contd.)

These mappings can be obtained easily for most verbs (thanks to PropBank and VerbNet).

The instance-level mapping, unfortunately, demands certain amount of corpus annotation effort. In reality, we have a lot of manually annotated data for PropBank and that must be mapped to VerbNet-style annotations.

The problem is that the mapping between two resources are, for most part, *many-to-many*. In turn, this means that automatic or semi-automatic methods do not produce reliable annotations.

Each PropBank-style annotated verb has several candidates in VerbNet and for that SemLink becomes a real project. For several verbs, this has been done SEMI-manually (be careful if you ever use it, expect some level of noise).

SemLink: A wish yet to fulfill (contd.)

FrameNet mappings in SemLink are not as reliable as PropBank and VerbNet mappings and, for certain, not as informative.

In most cases, the mappins are limited to **the most frequent sense** of the verbs (and their corresponding FrameNet class) and the remaining senses and frames have been ignored (finding examples is not that difficult).

The above mentioned problem is further intensified by shortcomings at the FrameNet side of the work. A simple example is the verb *build*, which can evoke the MANUFACTURING (among many other frames) but still not listed in FrameNet as LU: A simple cause that leads to SemLink gives wrong mappings for the verb build.

SemLink: A wish yet to fulfill (contd.)

To these, one can add trivial matters (for instance tokenization, or which version of what corpus has been used during the experiments, and so on) that, at the end, become nontrivial problems.

These are nontrivial in the sense that solving them demands a considerable amount of time (sometimes).

SemLink 1.2 is available online and free for analyses and manipulation!

SemLink: A wish yet to fulfill (contd.)

Palmer et al. (2010) emphasizes that SemLink mappings (as features) can improve Semantic Role Labelling: This can be a topic for us to investigate.

Other resource to consider

There are a number of other valuable resources that you may want to consider to use:

- SDP 2015: Broad-Coverage Semantic Dependency Parsing: It contains full text annotations for several languages in different formalism (similar to PropBank, mostly syntactic that semantics)
- OntoNotes: A large dataset, covers several languages and genre, PropBank style annotation.
- CoNLL 2005 dataset: Usually used as a reference for performance comparison.
- For German, there is SALSA dataset but more.
- For French, the French treebank has semantic annotations.

Other resource to consider (contd.)

- Any corpus for AMR parsing (most possibly this will be a derivation from OntoNotes).

* Semantic-Semantic Labeling: FrameNet!

. . . making feasible the creation of a substantial training corpus annotated with VN thematic role labels and class membership assignments, to be released in 2007. This will finally enable . . .

Bibliography

Palmer, M., Gildea, D., and Xue, N. (2010). *Semantic Role Labeling*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.