Introduction/Problem

Supervised semantic parsers (semantic role labelers) are trained on manually annotated example sentences that illustrate syntax/semantics mappings. Sparsity of this training data limits parser performance:

- Unknown pairs of words and frames, and pairs without examples limit parser performance on unseen data to 50%. [3]
- Lexicographic examples do not allow statistical learning. FrameNet ‘fulltext’ corpus is small (5946 sentences) and scarcely annotated (50.5%).
- Low syntactic coverage: Non-illustrated syntax/semantics patterns. (this work)

Contribution

- We analyze the extent of the syntactic coverage problem of FrameNet lexicographic annotations on FrameNet fulltext corpora. The lexicographic annotations contain the correct syntax/semantics mapping for only 53.4% of all annotations.
- We propose a simple method to apply existing annotations to new verbs within the same frame. Our algorithm uses Verbnet to ensure syntactic compatibility of annotations.
- As the result of our method we release a comprehensive dictionary of syntax/semantics mappings that covers most verb/frame pairs. The new mappings are syntactically correct (93.6%). The corresponding example sentences and mostly semantically well-formed (78.7%).

The dictionary is available at [http://www.cs.columbia.edu/~speech/text2scene/resources.html](http://www.cs.columbia.edu/~speech/text2scene/resources.html)

Sample Subcategorization Patterns and Sentences Applied to New Verbs

<table>
<thead>
<tr>
<th>Frame</th>
<th>Subcategorization</th>
<th>Example Text (new verbs in bold)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attaching</td>
<td>[ subj/Agent obj/Item dep(into)/Goal ]</td>
<td>Data can also be pasted / pinned into word processing documents.</td>
</tr>
<tr>
<td>Appointment</td>
<td>[ subj/Selector obj/Official dep(vp)/Function ]</td>
<td>In 93 , Tsar Simson appointed / designated Clement to be the first Slav bishop of the diocese.</td>
</tr>
<tr>
<td>Grooming</td>
<td>[ subj/Agent dep(with)/Medium ]</td>
<td>Iosif gave Sue’s hair a good trim before shampooing / soaping with Natural Styling Perm Hair Bath.</td>
</tr>
<tr>
<td>Categorization</td>
<td>[ subj/Cognizer obj/Item dep(onto)/Category obj/Cognizer ]</td>
<td>Rosa interpreted / understood this behavior as a desire to upset her.</td>
</tr>
<tr>
<td>Cause_to_be_wet</td>
<td>[ subj/Agent dep(aw)/Manner obj/Undergo ]</td>
<td>He sucked at his cigarette and then wet / humidified his lips distastefully.</td>
</tr>
<tr>
<td>Sedating</td>
<td>[ subj/Cognizer obj, Ground dep(ing)/Phenomenon ]</td>
<td>I scanned / surveyed the street for lurking strangers as I came near , and no one was there.</td>
</tr>
<tr>
<td>Experience</td>
<td>[ subj/Stimulus obj/Experience ]</td>
<td>The soundlessness of nature impressed and soothed / beguiled her.</td>
</tr>
<tr>
<td>Cooking, creation</td>
<td>[ subj/Cook obj/Repaint dep(onto)/Produced_food dep(into)/Place ]</td>
<td>Instead she set about cooking / baking herself a suitable supper ....</td>
</tr>
</tbody>
</table>

Verbs in one class behave syntactically equivalent.

- Tony broke the window (with a hammer).
- NP V NP (PP,instrument)
- The window broke.
- NP,patient V
- The hammer broke the window.
- NP,Instrument V NP
- Tony broke the window to pieces.
- NP V NP POblique :

Contribution

We analyze the extent of the syntactic coverage problem of FrameNet lexicographic annotations on FrameNet fulltext corpora. The lexicographic annotations contain the correct syntax/semantics mapping for only 53.4% of all annotations.

We propose a simple method to apply existing annotations to new verbs within the same frame. Our algorithm uses Verbnet to ensure syntactic compatibility of annotations.

As the result of our method we release a comprehensive dictionary of syntax/semantics mappings that covers most verb/frame pairs. The new mappings are syntactically correct (93.6%). The corresponding example sentences and mostly semantically well-formed (78.7%).

The dictionary is available at [http://www.cs.columbia.edu/~speech/text2scene/resources.html](http://www.cs.columbia.edu/~speech/text2scene/resources.html)

FrameNet

FrameNet[1] groups lexical items into frames which share contextual structure, i.e. they have the same set of frame elements (syntactic roles; color-coded in the following sample definition).

Dissimilar patterns are collapsed into a Duplication frame. The frame involves a Copy, making a duplicate, the Original of some Replica entity. Arguments like the location of the Window and those of the Window may be expressed.

Lexical items form lexical units with frames. Verbs in one frame are semantically similar. Example annotations illustrate syntax/semantics mappings.

Verbs in one class behave syntactically equivalent. For instance, we wish to prevent:

- John sprayed/*covered paint on the wall.
- John painted/*buttered wall plaster on the wall.

Results

- 209,475 new example sentences (see sample).
- Precision of new examples evaluated by independent judge: Syntactically adequate 93.8%
- Obey semantic selectional restrictions 78.7%
- Recall for subcat patterns on 10% leave-out lexicographic annotations: 99.6%
- Repeated coverage analysis on fulltext annotations:

<table>
<thead>
<tr>
<th>FrameNet</th>
<th>Verbnet</th>
</tr>
</thead>
<tbody>
<tr>
<td>butter-9.9</td>
<td>butter-9.9</td>
</tr>
<tr>
<td>spray-9.7</td>
<td>spray-9.7</td>
</tr>
<tr>
<td>fill-9.8</td>
<td>fill-9.8</td>
</tr>
<tr>
<td>coloring-24</td>
<td>coloring-24</td>
</tr>
<tr>
<td>spray</td>
<td>spray</td>
</tr>
<tr>
<td>cover</td>
<td>cover</td>
</tr>
<tr>
<td>contigious</td>
<td>location-47.8</td>
</tr>
</tbody>
</table>

Increasing Coverage - Algorithm

The algorithm partitions verbs in each FrameNet frame into syntactic equivalence classes (equal set of Verbnet classes). Annotations can be shared safely between verbs in an equivalence class.

For instance, we wish to prevent:

- John sprayed/*covered paint on the wall.
- John painted/*buttered wall plaster on the wall.

References


Acknowledgments

Supported by the National Science Foundation under Grant No. IIS-0904361.