Increasing Coverage of Syntactic Subcategorization Patterns in FrameNet Using Verbnet

Daniel Bauer and Owen Rambow
Columbia University, New York
{bauer, rambow}@cs.columbia.edu

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)

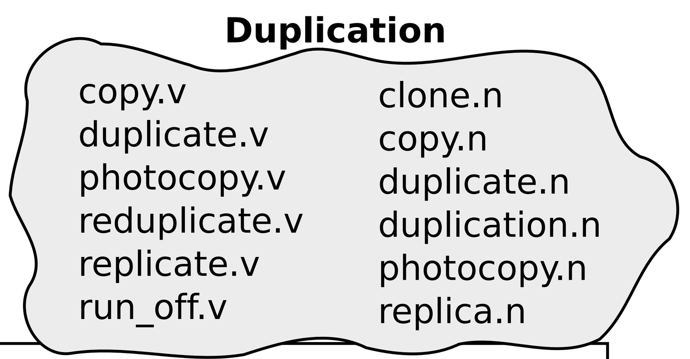
Contribution

- 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 ensures 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 (93.8%). 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`

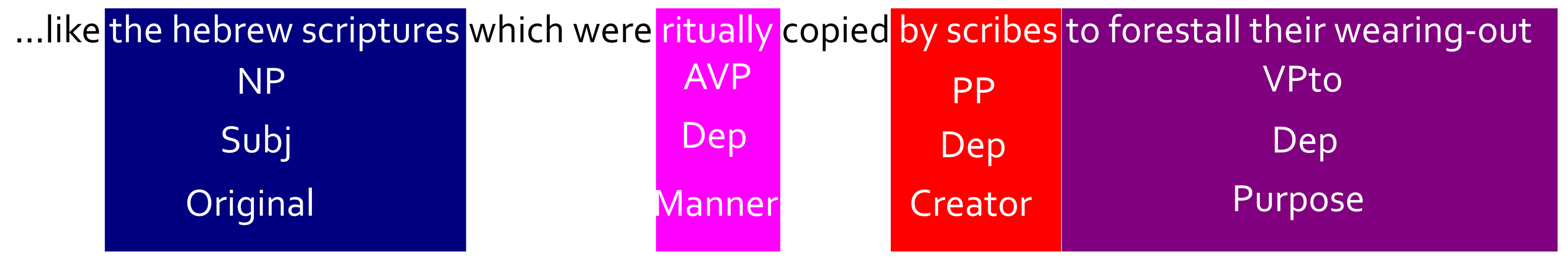
FrameNet

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



Duplication
This frame involves a creator making a duplicate, the Copy, of some Original entity. A Source, the location of the Original, and Goal, the location of the Copy, may be expressed.
None-core frame elements: Manner Purpose...

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



Subcategorization Mappings and Coverage Analysis

For each example annotation from the FrameNet *lexicographic annotations* we extract subcategorization patterns and their mapping to frame elements (normalized to active voice).
⟨copy.v, Duplication⟩:

{subj/Creator, obj/Original, dep(AVP)/Manner, dep(VPto)/Purpose}.

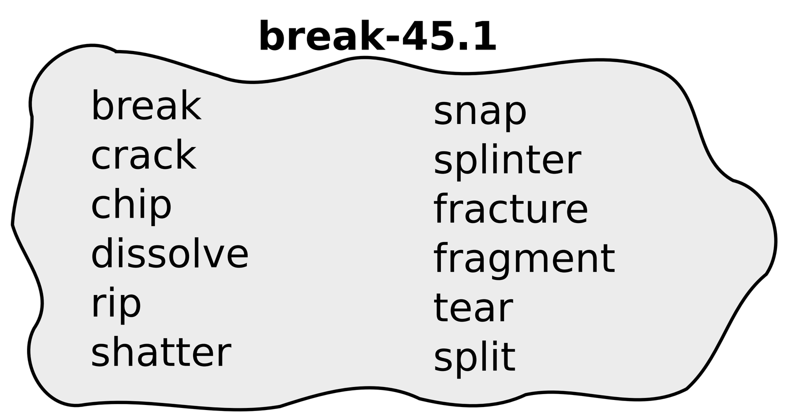
We evaluate coverage of these patterns on FrameNet 1.5 *fulltext annotations*.

#verbs	#anno	%TR_LU	%seen scats	
			per anno	per TR_LU
13510	6828(50.5%)	77.6	53.4	**68.9**

#verbs = number of verbs in the corpus. #anno = number of annotated verbs.
%TR_LU = perc. of annotations with any subcat pattern for lexical unit.
%seen_scats = perc. of annotations with correct subcat pattern for lexical unit.

Verbnet

Verbnet[2] groups verbs into classes, according to diathesis alternations, which are sets of possible subcategorization patterns a verb can occur in.
Verbs in one class behave *syntactically equivalent*.

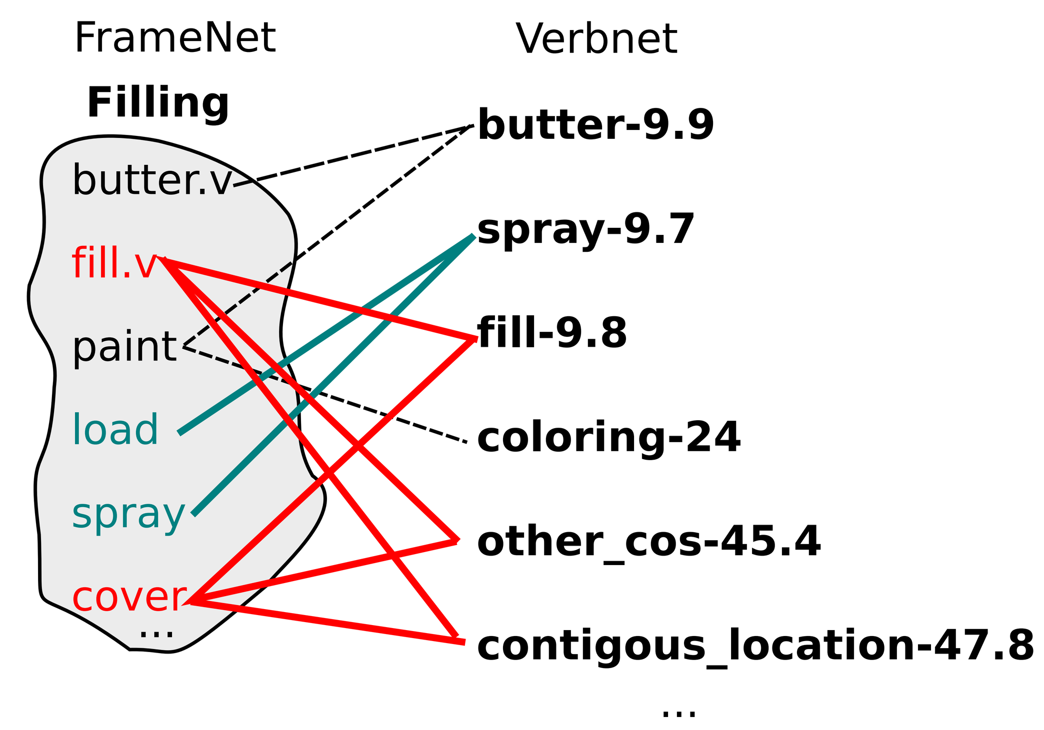


- Tony broke the window (with a hammer).
  NP V NP (PP.instrument)
- The window broke.
  NP.patient V
- The hammer broke the window.
  NP.Instrument V NP
- Tony broke the window to pieces.
  NP V NP PP.oblique

⋮

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.

Sample Subcategorization Patterns and Sentences Applied to New Verbs

Frame	Subcat Mapping	Example Text (new verbs in **bold**)
Attaching	{ subj/Agent obj/Item dep(into)/Goal }	Data can also be pasted/**pinned** into word processing documents.
Appointing	{ subj/Selector obj/Official dep(vpto)/Function}	In 893 , Tsar Simeon appointed/**designated** Clement to be the first Slav bishop of the diocese...
Grooming	{ subj/Agent dep(with)/Medium}	Ian gave Sue 's hair a good trim before shampooing/**soaping** with Natural Styling Perm Hair Bath.
Categorization	{ subj/Cognizer obj/Item dep(as)/Category obj/Cognizer }	Rosa interpreted/**stereotyped** this behavior as a desire to upset her.
Cause_to_be_wet	{ subj/Agent dep(avp)/Manner obj/Undergoer }	He sucked at his cigarette and then wet/**humidified** his lips distastefully.
Scrutiny	{ subj/Cognizer obj/Ground dep(pping)/Phenomenon}	I scanned/**surveyed** the street for lurking strangers as I came near , and no one was there.
Experiencer_obj	{ subj/Stimulus obj/Experiencer }	The soundlessness of nature impressed and solaced/**beguiled** her .
Cooking_creation	{ subj/Cook obj/Rcpient dep(np)/Produced_food dep(in)/Place }	Instead she set about cooking/**baking** herself a suitable supper ...

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*:

%TR_LU	%seen scats	
	per anno	per TR_LU
79.1	78.7	**99.5 (+30.6)**

References

[1] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The Berkeley FrameNet Project. In *COLING-ACL 1998*, pages 86–90, Montréal, 1998.
[2] Karin Kipper Schuler. *Verbnet: A Broad-Coverage, Comprehensive Verb Lexicon*. PhD thesis, University of Pennsylvania, 2005.
[3] Alexis Palmer and Caroline Sporleder. Evaluating FrameNet-style semantic parsing : the role of coverage gaps in FrameNet. In *COLING 2010*, Beijing, 2010.

Acknowledgments

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