FrameNet+: Fast Paraphrastic Tripling of FrameNet

Ellie Pavlick[1]   Travis Wolfe[2,3]   Pushpendre Rastogi[2]
Chris Callison-Burch[1]   Mark Dredze[2,3]   Benjamin Van Durme[2,3]
[1]Computer and Information Science Department, University of Pennsylvania
[2]Center for Language and Speech Processing, Johns Hopkins University
[3]Human Language Technology Center of Excellence, Johns Hopkins University

Abstract

We increase the lexical coverage of FrameNet through automatic paraphrasing. We use crowdsourcing to manually filter out bad paraphrases in order to ensure a high-precision resource. Our expanded FrameNet contains an additional 22K lexical units, a 3-fold increase over the current FrameNet, and achieves 40% better coverage when evaluated in a practical setting on New York Times data.

1 Introduction

Frame semantics describes a word in relation to real-world events, entities, and activities. Frame semantic analysis can improve natural language understanding (Fillmore and Baker, 2001), and has been applied to tasks like question answering (Shen and Lapata, 2007) and recognizing textual entailment (Burchardt and Frank, 2006; Aharon et al., 2010). FrameNet (Fillmore, 1982; Baker et al., 1998) is a widely-used lexical-semantic resource embodying frame semantics. It contains close to 1,000 manually defined *frames*, i.e. representations of concepts and their semantic properties, covering a wide array of concepts from Expensiveness to Obviousness.

Frames in FrameNet are characterized by a set of semantic roles and a set of lexical units (LUs), which are word/POS pairs that "evoke" the frame. For example, the following sentence contains a mention (i.e. *target*) of the Obviousness frame: *In late July, it was barely visible to the unaided eye.* This particular target instantiates several semantic roles of the Obviousness frame, including a Phenomenon (*it*) and a Perceiver (*the unaided eye*). Here, the LU `visible.a` evokes the frame. In total, the Obviousness frame has 13 LUs including `clarity.n`, `obvious.a`, and `show.v`.

---

[1]*well* received a rating of 3.67 as a paraphrase of *clearly* in the context *the intention to do so is clearly present.*

accurate, **ambiguous**, **apparent**, **apparently**, audible, **axiomatic**, **blatant**, **blatantly**, **blurred**, **blurry**, **certainly**, **clarify**, clarity, clear, clearly, **confused**, **confusing**, **conspicuous**, **crystal-clear**, **dark**, **definite**, **definitely**, **demonstrably**, **discernible**, **distinct**, evident, **evidently**, **explicit**, **explicitly**, **flagrant**, **fuzzy**, **glaring**, **imprecise**, **inaccurate**, **lucid**, manifest, **manifestly**, **markedly**, **naturally**, **notable**, **noticeable**, **obscure**, **observable**, obvious, obviously, **opaque**, **openly**, **overt**, **patently**, **perceptible**, **plain**, **precise**, **prominent**, **self-evident**, show, show up, **significantly**, **soberly**, **specific**, **straightforward**, **strong**, **sure**, **tangible**, **transparent**, **unambiguous**, **unambiguously**, **uncertain**, unclear, **undoubtedly**, **unequivocal**, **unequivocally**, **unspecific**, **vague**, **viewable**, **visibility**, visible, **visibly**, **visual**, **vividly**, **well**,[1] **woolly**

Table 1: 81 LUs invoking the Obviousness frame according to the new FrameNet+. New LUs (bold) have been added using the method of paraphrasing and human-vetting described in Section 4.

The semantic information in FrameNet (FN) is broadly useful for problems such as entailment (Ellsworth and Janin, 2007; Aharon et al., 2010) and knowledge base population (Mohit and Narayanan, 2003; Christensen et al., 2010; Gregory et al., 2011), and is of general enough interest to language understanding that substantial effort has focused on building parsers to map natural language onto FrameNet frames (Gildea and Jurafsky, 2002; Das and Smith, 2012). In practice, however, FrameNet's usefulness is limited by its size. FN was built entirely manually by linguistic experts. As a result, despite many years of work, most of the words that one confronts in naturally occurring text do not appear at all in FN. For example, the word *blatant* is likely to evoke the Obviousness frame, but is not present in FN's list of LUs (Table 1). In fact, out of the targets we sample in this work (described in Section 4), fewer than 50% could be mapped to a correct frame using the LUs in FrameNet. This finding is consistent with what has been reported by Palmer and Sporleder (2010). Such low lexical coverage prevents FN from applying to many real-world applications.

Frame	Original	Paraphrase	Frame-annotated sentence
Quantity	amount	figure	It is not clear if this **figure** includes the munitions...
Expertise	expertise	specialization	...the technology, **specialization**, and infrastructure...
Labeling	called	dubbed	...eliminate who he **dubbed** Sheiks of sodomite...
Importance	significant	noteworthy	...assistance provided since the 1990s is **noteworthy**...
Mental_property	crazy	berserk	You know it's **berserk**.

Table 2: Examples paraphrases from FrameNet's annotated fulltext. The bolded words are automatically proposed rewrites from PPDB.

In this work, we triple the lexical coverage of FrameNet quickly and with high precision. We do this in two stages: 1) we use rules from the Paraphrase Database (Ganitkevitch et al., 2013) to automatically paraphrase FN sentences and 2) we apply crowdsourcing to manually verify that the automatic paraphrases are of high quality. While prior efforts have entertained the idea of expanding FN's coverage (Ferrández et al., 2010; Das and Smith, 2012; Fossati et al., 2013), none have resulted in a publicly available resource that can be easily used. As our main contribution, we release FrameNet+, a huge, manually-vetted extension to the current FrameNet. FrameNet+ provides over 22,000 new frame/LU mappings in a format that can be readily incorporated into existing systems. We demonstrate that the expanded resource provides a 40% improvement in lexical coverage in a practical setting.

2 Expanding FrameNet Automatically

The Paraphrase Database (PPDB) (Ganitkevitch et al., 2013) is an enormous collection of lexical, phrasal, and syntactic paraphrases. The database is released in six sizes (S to XXXL) ranging from highest precision/lowest recall to lowest average precision/highest recall. We focus on lexical (single word) paraphrases from the XL distribution, of which there are over 370K.

Our aim is to increase the type-level coverage of FN. We use the rules in PPDB along with a 5-gram Kneser-Ney smoothed language model (Heafield et al., 2013) to paraphrase FN's full frame-annotated sentences (called *fulltext*). We ignore paraphrase rules which are redundant with LUs already covered by FN. This method for automatic paraphrasing has been discussed previously by Rastogi and Van Durme (2014). However, whereas their work only discussed the idea as a hypothetical way of augmenting FN, we apply the method, vet the results, and release it as a public resource.

In total, we generate 188,061 paraphrased sentences, covering 686 frames. Table 2 shows some of the paraphrases produced.

3 Manual Refining with Crowdsourcing

Our automatic process produces a large number of good paraphrases, but does not address issues like word sense, and many of the paraphrased LUs alter the sentence so that it no longer evokes the intended frame. For example, PPDB proposes *free* as a paraphrase of *open*. This is a good paraphrase in the Secrecy_status frame but does not hold for the Openness frame (Table 3).

✓	Secrecy_status
	The facilities are **open** to public scrutiny
	The facilities are **free** to public scrutiny
✗	Openness
	Museum (**open** Wednesday and Friday.)
	Museum (**free** Wednesday and Friday.)

Table 3: Turkers approved *free* as a paraphrase of *open* for the Secrecy_status frame (rating of 4.3) but rejected it in the Openness frame (rating of 1.6).

We therefore refine the automatic paraphrases manually to remove paraphrased LUs which do not evoke the same frame as the original LU. We show each sentence to three unique workers on Amazon Mechanical Turk (MTurk) and ask each to judge how well the paraphrase retains the meaning of the original phrase. We use the 5-point grading scale for paraphrase proposed by Callison-Burch (2008).

To ensure that annotators perform our task conscientiously, we embed gold-standard control sentences taken from WordNet synsets. Overall, workers were 76% accurate on our controls and showed good levels of agreement– the average correlation between two annotators' ratings was ρ = 0.49.

Figure 1 shows the distribution of Turkers' ratings for the 188K automatically paraphrased targets. In 44% of cases, the new LU was judged to retain the meaning of the original LU given the frame-specific context. These 85K sentences contain 22K unique frame/LU mappings which we are

able to confidently add to FN, tripling the total number in the resource. Figure 1 shows 69 new LUs added to the Obviousness frame.
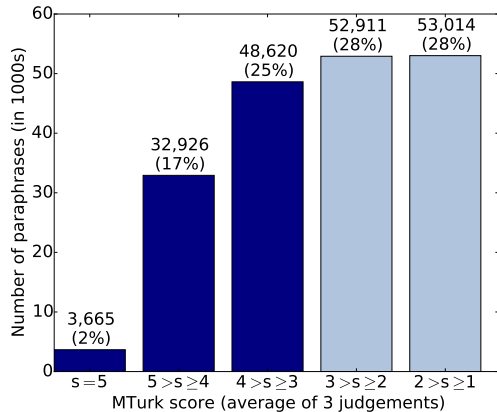

Figure 1: Distribution of MTurk ratings for paraphrased full-text sentences. 44% received an average rating ≥ 3, indicating the paraphrased LU was a good fit for the frame-specific context.

4 Evaluation

We aim to measure the type-level coverage improvements provided by our expanded FrameNet in a practical setting. Ideally, one would like to identify frames evoked by arbitrary sentences from natural text. To emulate this setting, we consider potentially frame-evoking LUs sampled from the New York Times. The question we ask is: does the resource contain an entry associating this LU with the frame that is actually evoked by this target?

FrameNet+ We refer to the expanded FrameNet, which contains the current FN's LUs as well as the proposed paraphrased LUs, as FrameNet+. The size and precision of FrameNet+ can be tuned by setting a threshold t and only including LU/frame mappings for which the average MTurk rating was at least t. Setting $t = 0$ includes all paraphrases, even those which human's judged to be incorrect, while setting $t > 5$ includes no paraphrases, and is equal to the current FN. Unless otherwise specified, we set $t = 3$. This includes all paraphrases which were judged minimally as "retaining the meaning of the original."

Sampling LUs We consider a word to be "potentially frame-evoking" if FN+ ($t = 0$) contains some entry for the word, i.e. the word is either an LU in the current FN or appears in PPDB-XL as a paraphrase of some LU in the current FN. We

sample 300 potentially frame-evoking word types from the New York Times: 100 each nouns, verbs, and adjectives. We take a stratified sample: within each POS, types are divided into buckets based on their frequency, and we sample uniformly from each bucket.

Annotation For each of the potentially frame-evoking words in our sample, we have expert (non-MTurk) annotators determine the frame evoked. The annotator is given the candidate LU in the context of the NYT sentence in which it occurred, and is shown the list of frames which are potentially evoked by this LU according to FrameNet+. The annotator then chooses which of the proposed frames fits the target, or determines that none do. We measure agreement by having two experts label each target. On average, agreement was good ($\kappa = 0.56$). In cases where they disagreed, the annotators discussed and came to a final consensus.

Results We compute the *coverage* of a resource as the percent of targets for which the resource contained a correct LU/frame mapping. Figure 2 shows the coverage computed for the current FN compared to FN+. By including the human-vetted paraphrases, FN+ is able to return a correct LU/frame mapping for 60% of the targets in our sample, 40% more targets than were covered by the current FN. Table 4 shows some sentences covered by FN+ that are missed by the current FN.
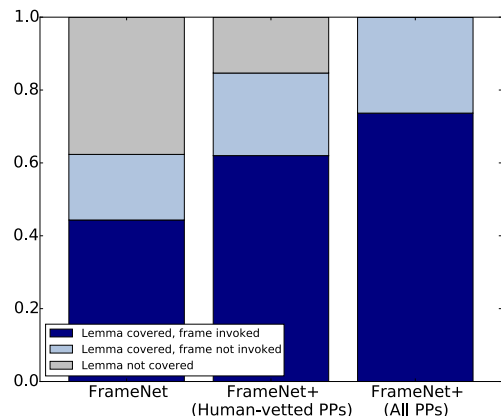

Figure 2: Number of LUs covered by the current FrameNet vs. two versions of FrameNet+: one including manually-approved paraphrases ($t = 3$), and one including all paraphrases ($t = 0$).

Figure 3 compares FN+'s coverage and number of LUs per frame using different paraphrase quality thresholds t. FN+ provides an average of more than 40 LUs per frame, compared to just over 10 LUs per frame in the current FN. Adding un-vetted

410

LU	Frame	NYT Sentence
outsider	Indigenous_ origin	. . . I get more than my fair share because I 'm the ultimate **outsider**. . .
mini	Size	. . . a **mini** version of "The King and I " . . .
prod	Attempt_ suasion	He gently **prods** his patient to step out of his private world. . .
precious	Expensiveness	Keeping **precious** artwork safe.
sudden	Expectation	. . . on the **sudden** passing of David .

Table 4: Example sentences from the New York Times. The frame-invoking LUs in these sentences are not currently covered by FrameNet but are covered by the proposed FrameNet+.

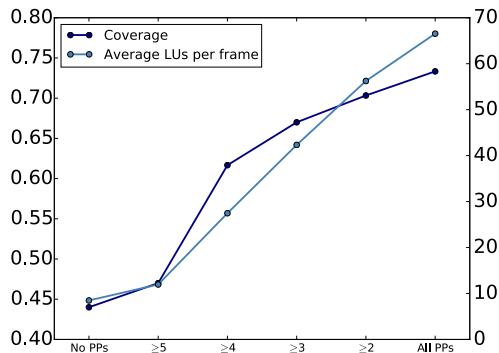LU paraphrases (setting $t = 0$) provides nearly 70 LUs per frame and offers 71% coverage.

Figure 3: Overall coverage and average number of LUs per frame for varying values of t.

5 Data Release

The augmented FrameNet+ is available to download at http://www.seas.upenn.edu/~nlp/resources/FN+.zip. The resource contains over 22K new manually-verified LU/frame pairs, making it three times larger than the currently available FrameNet. Table 5 shows the distribution of FN+'s full set of LUs by part of speech.

Noun	12,786	Prep.	455	Conj.	14
Verb	10,862	Number	163	Wh-adv.	12
Adj.	6,195	Article	43	Particle	6
Adv.	749	Modal	22	Other	19

Table 5: Part of speech distribution for 31K LUs in FrameNet+.

The release also contains 85K human-approved paraphrases of FN's fulltext. This is a huge increase over the 4K fulltext sentences currently in FN, and the new data can be easily used to retrain existing frame semantic parsers, improving their coverage at application time.

6 Related Work

Several efforts have worked on expanding FN coverage. Most approaches align FrameNet's LUs to WordNet or other lexical resources (Shi and Mihalcea, 2005; Johansson and Nugues, 2007; Pennacchiotti et al., 2008; Ferrández et al., 2010).

Das and Smith (2011) and Das and Smith (2012) used graph based semi-supervised methods to improve frame coverage and Hermann et al. (2014) used word and frame embeddings to improve generalization. All of these improvements are restricted to their respective tool rather than a general-use resource. In principle one of these tools could be used to annotate a large corpus in search of new LUs, but their precision on unseen predicates/LUs (our focus here) is still below 50%, considerably lower than this work.

Fossati et al. (2013) added new frames to FN by collecting full frame annotations through crowdsourcing, a more complicated task that again did not result in a useable resource. Buzek et al. (2010) applied crowdsourced paraphrasing to expand training data for machine translation. Our approach differs in that we expand the number of LUs directly using automatic paraphrasing and use crowdsourcing to verify that the new LUs are correct. We apply our method in full, resulting in a large resource can be easily incorporated into existing systems.

7 Conclusion

We have applied automatic paraphrasing to greatly increase the type-level lexical coverage of FrameNet, a widely used resource embodying the theory of frame semantics. We use crowdsourcing to manually verify that the newly added lexical units are correct, resulting in FrameNet+, a high-precision resource that is three times as large as the existing resource. We demonstrate that in a practical setting, the expanded resource provides a 40% increase in the number of sentences for which FN is able to identify the correct frame. The data released will improve the applicability of FN to end-use applications with diverse vocabularies.

References

Roni Ben Aharon, Idan Szpektor, and Ido Dagan. 2010. Generating entailment rules from framenet. In *ACL*, pages 241–246.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The Berkeley FrameNet Project. In *COLING*.

Aljoscha Burchardt and Anette Frank. 2006. Approaching textual entailment with LFG and FrameNet frames. In *Proceedings of the Second PASCAL RTE Challenge Workshop*. Citeseer.

Olivia Buzek, Philip Resnik, and Benjamin B Bederson. 2010. Error driven paraphrase annotation using mechanical turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pages 217–221. Association for Computational Linguistics.

Chris Callison-Burch. 2008. Syntactic constraints on paraphrases extracted from parallel corpora. In *EMNLP*, pages 196–205. Association for Computational Linguistics.

Janara Christensen, Stephen Soderland, Oren Etzioni, et al. 2010. Semantic role labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60. Association for Computational Linguistics.

Dipanjan Das and Noah A Smith. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *ACL*, pages 1435–1444.

Dipanjan Das and Noah A Smith. 2012. Graph-based lexicon expansion with sparsity-inducing penalties. In *NAACL*.

Michael Ellsworth and Adam Janin. 2007. Mutaphrase: Paraphrasing with framenet. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 143–150.

Oscar Ferrández, Michael Ellsworth, Rafael Munoz, and Collin F Baker. 2010. Aligning FrameNet and WordNet based on semantic neighborhoods. In *LREC*.

Charles J Fillmore and Collin F Baker. 2001. Frame semantics for text understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*. Association for Computational Linguistics.

Charles Fillmore. 1982. Frame semantics. *Linguistics in the morning calm.*

Marco Fossati, Claudio Giuliano, and Sara Tonelli. 2013. Outsourcing FrameNet to the crowd. In *ACL*, pages 742–747.

Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. Ppdb: The paraphrase database. In *HLT-NAACL*.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational linguistics*, 28(3):245–288.

Michelle L Gregory, Liam McGrath, Eric Belanga Bell, Kelly O'Hara, and Kelly Domico. 2011. Domain independent knowledge base population from structured and unstructured data sources. In *FLAIRS Conference*.

Kenneth Heafield, Ivan Pouzyrevsky, Jonathan H Clark, and Philipp Koehn. 2013. Scalable modified kneser-ney language model estimation. In *ACL*.

Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In *ACL*.

Richard Johansson and Pierre Nugues. 2007. Using WordNet to extend FrameNet coverage. In *Building Frame Semantics Resources for Scandinavian and Baltic Languages*, pages 27–30. Department of Computer Science, Lund University.

Behrang Mohit and Srini Narayanan. 2003. Semantic extraction with wide-coverage lexical resources. In *NAACL-HLT*, pages 64–66.

Alexis Palmer and Caroline Sporleder. 2010. Evaluating FrameNet-style semantic parsing: The role of coverage gaps in framenet. In *COLING*. Association for Computational Linguistics.

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *EMNLP*.

Pushpendre Rastogi and Benjamin Van Durme. 2014. Augmenting FrameNet via PPDB. In *Proceedings of the 2nd Workshop on Events: Definition, Detection, Coreference, and Representation*. Association of Computational Linguistics.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *EMNLP-CoNLL*.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. In *Computational linguistics and intelligent text processing*. Springer.