

Sister Help: Data Augmentation for Frame-Semantic Role Labeling

Ayush Pancholy[♦] Miriam R. L. Petruck[♥] Swabha Swayamdipta[♣]

[♦] University of California, Berkeley

[♥] International Computer Science Institute, Berkeley

[♣] Allen Institute for Artificial Intelligence

{ayush.pancholy@,miriamp@icsi.}berkeley.edu

swabhas@allenai.org

Abstract

While FrameNet is widely regarded as a rich resource of semantics in natural language processing, a major criticism concerns its lack of coverage and the relative paucity of its labeled data compared to other commonly used lexical resources such as PropBank and VerbNet. This paper reports on a pilot study to address these gaps. We propose a data augmentation approach, which uses existing frame-specific annotation to automatically annotate other lexical units of the same frame which are unannotated. Our rule-based approach defines the notion of a **sister lexical unit** and generates frame-specific augmented data for training. We present experiments on frame-semantic role labeling which demonstrate the importance of this data augmentation: we obtain a large improvement to prior results on frame identification and argument identification for FrameNet, utilizing both full-text and lexicographic annotations under FrameNet. Our findings on data augmentation highlight the value of automatic resource creation for improved models in frame-semantic parsing.

1 Introduction

Among the challenges to Natural Language Processing (NLP) systems is access to sufficient and accurate information about the mapping between form and meaning in language (Bender and Koller, 2020). While the number of resources that provide such information has increased in recent decades, producing these resources remains labor-intensive and costly. This is particularly true for FrameNet (FN; Ruppenhofer et al., 2016), a rich resource of semantic annotations, popularized for semantic role labeling (SRL) since the pioneering work of Gildea and Jurafsky (2002).

Nevertheless, FrameNet still relies (almost exclusively) on expert manual linguistic annotation, resulting in far fewer annotations than other resources that are able to scale up the annotation process easily (FitzGerald et al., 2018). Moreover, the



Figure 1: Illustration of Sister Lexical Units (**Sister LUs**) with FrameNet annotation for the verbs *buy* and *purchase*, which are defined in terms of the *Commerce_buy* frame. The arguments of both LUs are identical, both realizing the frame elements (FEs) **BUYER**, **GOODS**, **SELLER**, and **MONEY**. Targets in the sentence that trigger the frame are the words *bought* and *purchased*, respectively.

fine-grained annotations in FrameNet make them harder to produce compared to other lexical resources such as PropBank (Palmer et al., 2005) and VerbNet (Kipper et al., 2000). As a result, while SRL has continued to be a mainstay of core NLP, with applications in event extraction, participant tracking, machine translation, and question-answering, far fewer SRL systems use FrameNet for training, in contrast to PropBank.¹

Our work seeks to address this low resource problem in FrameNet in English by automatic data augmentation. We leverage the FrameNet hierarchy where multiple lexical units are associated with the same frame, but might have a different number of annotations associated with them (§2). We propose a rule-based approach to transfer annotations among sister LUs; see Fig. 1 for an illustration. We implement and extend the proposal by Petruck and Swayamdipta (2019) to generate new annotation for previously unannotated LUs in the same frame, by exploiting the already identified target and frame of a given lemma. We hypothesize that our rule-based approach to transfer *frame-specific* annotation to generate semantic role labels for example

¹PropBank annotations were also included in CoNLL shared tasks in 2005 (Carreras and Màrquez, 2005) and 2012 (Pradhan et al., 2012), and in OntoNotes (Pradhan et al., 2013).

sentences of lexical units (LUs) for which none exist, can indeed result in a FrameNet with higher coverage annotations, a state of affairs that can help in training frame-semantic role labeling systems (§3). Our experiments show that this is indeed the case; models trained on augmented data are able to surpass a strong baseline (OPEN-SESAME; Swayamdipta et al., 2017) for frame and argument identification (§4). We also provide a summary of the types of linguistic errors in the augmented data and the SRL task.

While our work provides a proof of concept, its novelty lies in exploiting frame-specific annotation to produce augmented data yielding new annotation. Our implementation and augmented data are publicly available.²

2 Background

2.1 FrameNet

FrameNet (Ruppenhofer et al., 2016) is a research and resource development project in corpus-based computational lexicography grounded in the principles of **frame semantics** (Fillmore, 1985). One of the goals of this effort is documenting the **valences**, i.e., the syntactic and semantic combinatorial possibilities of each item analyzed. These valence descriptions provide critical information on the mapping between form and meaning; NLP and more broadly, natural language understanding (NLU) require such mapping.

At the heart of the work is the **semantic frame**, a script-like knowledge structure that facilitates inference within and across events, situations, relations, etc. FrameNet defines a semantic frame in terms of its **frame elements** (FEs), or participants in the scene that the frame captures; a **lexical unit** (LU) is a pairing of a lemma and a frame, thus characterizing that LU in terms of the frame that it evokes. Valence descriptions derive from the annotation of FEs, i.e. **semantic roles**, on example sentences that illustrate the linguistic manifestation of the participants in a scene for the target of analysis. The first sentence in Figure 1 illustrates annotation with respect to the verb *buy*, which FN defines in terms of the `Commerce_buy` frame, whose FEs are a **BUYER**, **SELLER**, **GOODS**, and **MONEY**.³

²<https://github.com/ayush-pancholy/sister-help>

³Frame names appear in Typewriter font; FE names are in SMALL CAPS; and LUs are in *italics*.

The FrameNet⁴ database (version 1.7) holds 1,224 frames, 10,478 frame-specific FEs (in lexical frames),⁵ over 13,500 LUs. The primary annotations are in the form of nearly 203,000 manually annotated sentences providing information about the mapping between form and meaning in English; these are called the **lexicographic** annotations. A subset of the sentences with lexicographic annotations contain annotations for all possible frames of meaning in them; FrameNet calls this subset full-text annotations. Despite FrameNet’s costly annotation efforts, 38% of the over 13,500 LUs in the database remain without annotation. That is, a significant amount of LUs in the database have no associated annotated sentences to serve as training data for downstream NLP applications.

2.2 Frame Semantic Role Labeling

Semantic role labeling (SRL) involves automatically labeling who did what to whom in a text document. SRL systems facilitate developing NLP applications, like question answering (Shen and Lapata, 2007), machine translation (Pedersen, 2001), and text summarization (Han et al., 2016), to name but a few. Frame semantic role labeling (**frame-SRL**) is a special type of SRL where the task is to identify target tokens (which evoke a frame of meaning), the frame itself (of target LUs) and FEs, or semantic roles, in text. The relationship between frames and their respective FEs requires that the system first identify the correct frame (FrameID) to identify the correct FEs (ArgID). Identifying the frame incorrectly necessarily means incorrect FE identification.⁶ In practice, frame-SRL is usually implemented as a pipeline of structured prediction (Smith, 2011) tasks: identifying targets and LUs, frames, and finally the FEs (Das et al., 2010).

Using an early version of FN data (Johnson et al., 2001), Gildea and Jurafsky (2002) developed the first SRL system, which also initiated SRL as a now well-recognized task in the field. Recent years have seen the development of several (relatively) high-performing SRL systems. SEMAFOR (Das et al., 2014) uses a pipeline of discrete, manually designed feature-based linear clas-

⁴https://framenet.icsi.berkeley.edu/fndrupal/framenet_request_data

⁵FrameNet distinguishes between lexical and non-lexical frames, where the latter do not have associated annotation sets.

⁶Baker et al. (2007) also considered correct identification of the spans of a dependent of a lexical unit, even if the frame and by definition the FEs were identified incorrectly.

sifiers for target identification, frame identification (FrameID), and argument identification (ArgID). The core constraints of frame-semantic analyses, e.g. not repeating core FEs, are satisfied via integer linear programming. PathLSTM (Roth and Lapata, 2016) uses neural features to embed path relationships between frames and FEs, in a pipeline similar to that of SEMAFOR. Yang and Mitchell (2017) use a joint model to leverage Propbank SRL annotations to improve frame-SRL. OPEN-SESAME (Swayamdipta et al., 2017, 2018) uses an unconstrained, neural approach in a pipeline like SEMAFOR and PathLSTM. Continuous representations and a sophisticated global model based on semi-Markov conditional random fields (Sarawagi and Cohen, 2004) improve identifying arguments. We employ OPEN-SESAME⁷ as the baseline of choice in our experiments (§4). With the exception of a few approaches, like that of Yang and Mitchell (2017), most prior work uses only the full-text annotations for frame-SRL; we present experiments considering the larger set of lexicographic annotations as well.

3 Method: Data Augmentation with Sister LUs

The relationship among LUs in a frame motivates examining their use for paraphrasing, where one LU replaces another resulting in an alternative phrasing of a sentence. Ellsworth and Janin (2007) produced paraphrases based on the relationship among LUs of a frame or closely related frames, augmenting a meager data set for use in speech recognition.

Most importantly, LUs in a frame tend overwhelmingly to follow the same annotation structure: the arguments of a lexical unit in a frame consist of the same FEs, regardless of the LU.⁸ Note the parallel annotation for the two sentences in Figure 1 for the verbs, *buy* and *purchase*. We use this critical insight in our work, hypothesizing that existing annotation would inform the automated generation of annotation for unannotated LUs in a given frame.

To this end, Petruck and Swayamdipta (2019) defined **Sister Lexical Unit** as any LU for which annotation exists in a frame, and whose annotation can serve as a model to generate new annotation. A

⁷<https://github.com/swabhs/open-sesame>

⁸Two or more FEs might be part of a *core set*, yet only one will be realized for valid annotation. For a given frame, different LUs may not realize all the same FEs in all examples.

Sister is the LU with annotation and an **EmptyLU** lacks annotation; the Sister LU and the Empty LU must be of the same part of speech.

To generate augmented labeled data, we first identified all EmptyLUs. For each such LU, we identified a corresponding Sister, specifically, the LU with the greatest number of annotation sets and of the same POS as the EmptyLU in a given frame. We replaced each occurrence of a Sister in an annotated sentence with the EmptyLU. The replacement process included steps to ensure that the newly generated instance included the correct word forms, i.e., singular and plural forms of nouns or conjugated verbs of a Sister LU, following the strategy that Rastogi and Van Durme (2014) employed, which augmented FN using a paraphrase database (Ganitkevitch et al., 2013) to add new lemmas and then rewrote existing annotated sentences with those lemmas.

We transferred each sentence in the annotation set of the Sister into that of each EmptyLU, and replaced instances of the Sister with those of the EmptyLU. The replacement process ensured that newly generated data included correct word forms, i.e., singular and plural forms of nouns, conjugated verbs, or those with tense marking.

Ensuring agreement of tense, person, and number between instances of a sister and an EmptyLU required replacing the original word forms with those of the same grammatical number for nouns and tense, person, and number for verbs. However, this does not always yield correct results, e.g., for lemmas with irregular plural or past tense forms, as in *ox*, *oxen* and *bring*, *brought*.

3.1 Candidates for Augmentation

The aforementioned process identified 2,805 previously unannotated LUs, of which approximately 500 either were multiword expressions (MWEs) or had an MWE as a potential Sister. Thus, the process added at least one sentence to each previously empty annotation set, yielding the potential to provide new annotation for (approximately) 45% of the unannotated LUs in FrameNet 1.7. See Table 2 in the Appendix A for a few pairs of sister and empty LU pairs. The replacement process did not always yield grammatically correct results. Sometimes FrameNet (arbitrarily) treats MWEs as single LUs, and using a MWE (as a Sister or an EmptyLU) often resulted in ungrammatical sentences. Thus, we also eliminated MWEs from the data in this

work, resulting in 2,300 previously unannotated LUs, eligible for augmentation.

4 Experiments

To test the idea of leveraging frame-specific annotation to generate new annotation for LUs lacking any, we wish to compare the performance of our approach with a baseline model. This facilitates determining the extent to which our augmentation algorithm would improve SRL performance, and thus (potentially) contribute to the FrameNet corpus. Our augmentation approach is applied to FrameNet version 1.7.

4.1 Training Data

In principle, our data augmentation algorithm (§3) could provide new annotation for 2,300 LUs (§3.1), all of which could be used for training. However, we wish to determine whether using augmented data improves performance on the existing test sets. Additionally, care must be taken to avoid the logical flaw of using augmentation algorithm from producing labeled sentences specifically for the test set: the process would circularly assume that the algorithm offers accurate labelling.

We addressed this issue by adjusting the training data for the baseline: we removed annotation from 1,500 randomly selected LUs from the lexicographic portion of the training data for FN 1.7. Each of these 1,500 LUs functions as Empty-LUs when stripped of their annotation. Our baseline model was trained with the remaining lexicographic annotations. Our full model was trained on data given by the augmentation algorithm to re-annotate the 1,500 LUs, in addition to the data used for the baseline. Note that not all LUs are eligible for training in OPEN-SESAME, which involves a pre-processing step to remove those LUs incompatible with the parser; Table 1 provides the final training data statistics.

In addition to the lexicographic annotations, both the baseline and the augmented setup use all the full-text annotations in FrameNet 1.7. Our training data adjustments above allow using the standard validation and test sets for FrameNet 1.7, following Swayamdipta et al. (2017).

4.2 Model and Hyperparameters

All of our experiments were performed using the non-syntactic model provided in OPEN-SESAME (Swayamdipta et al., 2017, 2018). We follow the

basic setup in OPEN-SESAME which assumes gold targets for FrameID and gold frames for ArgID. We left most hyperparameters associated with OPEN-SESAME at their default values, but decreased the learning rate for training the ArgID stage to 0.0001 from its original value of 0.0005. The decrease allowed training ArgID on the entire dataset in three epochs without error, which serves reproducibility requirements.

4.3 Results

Table 1 presents test and validation set performances comparing two OPEN-SESAME models, trained with and without the augmented data for ArgID and FrameID. All measures in the experimental setting outperformed the baseline, highlighting the value of data augmentation. Although our main interest was testing the system’s performance for ArgID (i.e., identifying semantic roles / FEs for given frame-LU annotations) of the augmented data, we also observed improvement in the identification of frames. The results in the experimental setting show an absolute improvement of 2.7 in F1 on ArgID and 1.6 in F1 on FrameID.⁹

4.4 Error Analysis and Discussion

At times the augmentation algorithm produced ungrammatical sentences. We do not believe that errors occurred because of ungrammatical sentences. Still, FrameNet would not conjure producing data with ungrammatical sentences.¹⁰ A manual analysis of a random sample of LUs showed three categories of errors: (a) word form mismatch; (b) incorrect or missing marker; and (c) semantic mismatch, examples of which appear in (a) - (c):

- (a) *The moon was now *occlude* by clouds.
- (b) *And he *complained* the endless squeeze on cash.
- (c) ?The faint *flash* from a street light showed him the outline of a hedge.

The semantic mismatch in (c) holds between *faint.a* and *flash.n*, yielding a semantically odd sentence not an ungrammatical one. By definition, a *flash* would not be *faint*.¹¹

⁹We further experimented with cross validation settings for both FrameID and ArgID; however, the runtime of OPEN-SESAME was prohibitively slow in this setting.

¹⁰A complete manual analysis of the data remains to be done. In future extensions of this work, we will report such an analysis, including the percentage of grammatical sentences.

¹¹The results showed several types of semantic mismatch. Space limitations preclude discussing all of them here.

	Training Data		Dev		Test			
			ArgID	FrameID	ArgID			FrameID
	#LUs	#Lexicographic Annotations	F1	F1	P	R	F1	F1
OPEN-SESAME	6905 7944	164372 260292	59.3 61.9	84.7 85.3	61.8 63.6	56.9 60.3	59.2 61.9	81.1 82.7

Table 1: Results comparing the OPEN-SESAME models trained on the original FrameNet 1.7 data for frame identification (FrameID) and argument identification (ArgID) tasks, with OPEN-SESAME models trained on our augmented data. Data augmentation improves performance on both tasks across all metrics; boldface indicates best performance. Note that under both settings we consider full-text annotations, a subset of the lexicographic annotations. Gold-standard frames were used for ArgID, and gold-standard targets for FrameID, in both the baseline and our models. Development set results only include F1 scores; (by default) OPEN-SESAME just reports that score.

Recall that incorrectly identifying a frame necessarily yields incorrect FEs (ArgID). The analysis showed that errors occurred with highly polysemous lemmas, such as *make.v*, where FrameNet has 10 LUs; in (d), OPEN-SESAME misidentified *making.v* as *Manufacturing*, not *Earnings_and_losses*. Similarly, (e) is a misidentification of *gathering* as *Food_gathering*, not *Come_together*. Note the metaphorical use of the verb in (e), compared to literal senses of the lemma *gather.v*.

(d) I like working and *making* money.

(e) ...the debate on welfare reform is *gathering* like a storm at sea...

5 Related Work

Data augmentation in the context of FrameNet and SRL is not novel. Pavlick et al. (2015) created an expanded FN via automatic paraphrasing and crowdsourcing to confirm frame assignment of LUs. Hartmann et al. (2017) used automatically generated training data by linking FrameNet, PropBank, and VerbNet to study the differences among those resources in frame assignment and semantic role classification. The results showed that the augmented data performed as well as the training data for these tasks on German texts. With similar goals to those of this work, Carvalho et al. (2020) proposed augmenting FN annotation using notions of lexical, syntactic, or semantic equivalence, and FN’s frame-to-frame relations (Inheritance, Using, SubFrame, etc.) to *infer* annotations across related frames, producing a 13% increase in annotation.

6 Conclusion and Future Work

We outlined an approach to Frame-SRL that seeks to leverage existing annotation in a frame to generate new annotation for previously unannotated

LUs in the same frame, by exploiting the already identified target and frame of a given lemma. Further, we demonstrated that augmentation of data improves Frame-SRL performance substantially. While we show the advantages of this approach on the open-SESAME baseline, future work might involve studying if using this type of data augmentation will yield improvements on a state-of-the-art SRL system, e.g. Fei et al. (2021).

Several other interesting directions for future work exist, including the following: refining the augmentation algorithm to produce more grammatical sentences; streamlining open-SESAME’s code to run more efficiently; use active-learning to inform the data generation and to assist in the SRL task; and experimentation with other computational techniques, as suggested by Petrucci and Swayamdipta (2019). These plans will advance defining processes to include Frame-SRL into FrameNet’s development process, the ultimate goal of this work.¹²

Acknowledgements

The authors are grateful to the three anonymous reviewers for their feedback, and to Nathan Schneider for his input on various aspects of the experiment, including very early conversations about pursuing the idea. Collin Baker and Michael Ellsworth have contributed to the completion of the paper in countless ways.

References

Collin F. Baker, Michael Ellsworth, and Katrin Erk. 2007. *Semeval-2007 task 19: Frame semantic structure extraction*. In *Proceedings of the 4th Inter-*

¹²Rigorous testing of multiple factors will be required to evaluate the viability of including Frame-SRL into FrameNet’s existing primarily manual process.

- national Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, June 23-24, 2007*, pages 99–104. The Association for Computer Linguistics.
- Emily M. Bender and Alexander Koller. 2020. [Climbing towards NLU: On meaning, form, and understanding in the age of data](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.
- Xavier Carreras and Lluís Màrquez. 2005. [Introduction to the CoNLL-2005 shared task: Semantic role labeling](#). In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)*, pages 152–164, Ann Arbor, Michigan. Association for Computational Linguistics.
- Breno Carvalho, Aline Paes, and Bernardo Gonçalves. 2020. [Augmenting linguistic semi-structured data for machine learning: A Case study using FrameNet](#). In *Proceedings of the International Conference on Machine Learning Techniques and NLP*, volume 10.12, pages 1–13.
- Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. [Frame-semantic parsing](#). *Computational Linguistics*, 40(1):9–56.
- Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. 2010. [Probabilistic frame-semantic parsing](#). In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 948–956, Los Angeles, California. Association for Computational Linguistics.
- 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, Prague. Association for Computational Linguistics.
- Hao Fei, Shengqiong Wu, Yafeng Ren, and Donghong Ji. 2021. [Second-order semantic role labeling with global structural refinement](#). *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:1966–1976.
- Charles J. Fillmore. 1985. [Frames and the semantics of understanding](#). *Quaderni di Semantica*, 6(2):222–254.
- Nicholas FitzGerald, Julian Michael, Luheng He, and Luke Zettlemoyer. 2018. [Large-scale QA-SRL parsing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2051–2060, Melbourne, Australia. Association for Computational Linguistics.
- Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. [PPDB: The paraphrase database](#). In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 758–764, Atlanta, Georgia. Association for Computational Linguistics.
- Daniel Gildea and Daniel Jurafsky. 2002. [Automatic labeling of semantic roles](#). *Computational Linguistics*, 28(3):245–288.
- Xu Han, Tao Lv, Zhirui Hu, Xinyan Wang, and Cong Wang. 2016. [Text summarization using FrameNet-based semantic graph model](#). *Scientific Programming*, 2016:1–10.
- Silvana Hartmann, Éva Mújdricza-Maydt, Ilia Kuznetsov, Iryna Gurevych, and Anette Frank. 2017. [Assessing SRL frameworks with automatic training data expansion](#). In *Proceedings of the 11th Linguistic Annotation Workshop*, pages 115–121, Valencia, Spain. Association for Computational Linguistics.
- Christopher R. Johnson, Charles J. Fillmore, Esther J. Wood, Josef Ruppenhofer, Margaret Urban, Miriam R. L. Petruck, and Collin F. Baker. 2001. [The FrameNet Project: Tools for Lexicon Building](#). ICSI Online Publication.
- Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. [Class-based construction of a verb lexicon](#). In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 691–696. AAAI Press.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. [The proposition bank: An annotated corpus of semantic roles](#). *Computational Linguistics*, 31(1):71–106.
- Ellie Pavlick, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze, and Benjamin Van Durme. 2015. [FrameNet+: Fast paraphrastic tripling of FrameNet](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 408–413, Beijing, China. Association for Computational Linguistics.
- Bolette S. Pedersen. 2001. [Lexical ambiguity in machine translation: Using frame semantics for expressing systemacies in polysemy](#). In *Recent Advances in Natural Language Processing*.
- Miriam R L Petruck and Swabha Swayamdipta. 2019. [Automating FrameNet annotation](#). Pre-Proposal for NSF CISE EAGER Grant.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. [Towards robust linguistic analysis using OntoNotes](#). In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, Sofia, Bulgaria. Association for Computational Linguistics.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. *CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes*. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40, Jeju Island, Korea. Association for Computational Linguistics.

Pushpendre Rastogi and Benjamin Van Durme. 2014. *Augmenting FrameNet via PPDB*. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, Baltimore, Maryland, USA. Association for Computational Linguistics.

Michael Roth and Mirella Lapata. 2016. *Neural semantic role labeling with dependency path embeddings*. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. *FrameNet II: Extended Theory and Practice*. ICSI: Berkeley.

Sunita Sarawagi and William W. Cohen. 2004. *Semimarkov conditional random fields for information extraction*. In *Proc. of NeurIPS*, pages 1185–1192.

Dan Shen and Mirella Lapata. 2007. *Using semantic roles to improve question answering*. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 12–21, Prague, Czech Republic. Association for Computational Linguistics.

Noah A. Smith. 2011. *Linguistic Structure Prediction*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool.

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. *Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold*.

Swabha Swayamdipta, Sam Thomson, Kenton Lee, Luke Zettlemoyer, Chris Dyer, and Noah A. Smith. 2018. *Syntactic scaffolds for semantic structures*. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3772–3782, Brussels, Belgium. Association for Computational Linguistics.

Bishan Yang and Tom Mitchell. 2017. *A joint sequential and relational model for frame-semantic parsing*. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1247–1256, Copenhagen, Denmark. Association for Computational Linguistics.

A Qualitative Examples

Table 2 provides lexicographic annotations with sister LUs and their corresponding empty LUs, the latter being populated by our approach.

Sister Lexical Units	Empty Lexical Units
He _{AGENT} <i>stamped his foot</i> → BODY_PART into his flying-boot PERIPHERAL .	He _{AGENT} <i>bended his foot</i> → BODY_PART into his flying-boot PERIPHERAL .
John-William , who knew that he would have been a Chartist himself had he remained a <i>poor man</i> PERSON , felt sorry about that death .	John-William , who knew that he would have been a Chartist himself had he remained a <i>rich man</i> PERSON , felt sorry about that death .
Because of the censorship , and the obvious need to avoid <i>dangerously</i> DEGREE <i>critical</i> COMMENTS EXPRESSOR <i>about the regime and the war</i> EVALUÉE , the correspondence to and from the front provides no easy guide to political attitudes .	Because of the censorship , and the obvious need to avoid <i>dangerously</i> DEGREE <i>commendable</i> COMMENTS EXPRESSOR <i>about the regime and the war</i> EVALUÉE , the correspondence to and from the front provides no easy guide to political attitudes .
When I did eventually tell her <i>she</i> EXPERIENCER was <i>really</i> DEGREE <i>embarrassed</i> , and tried telling me that I was making it up !	When I did eventually tell her <i>she</i> EXPERIENCER was <i>really</i> DEGREE <i>tormented</i> , and tried telling me that I was making it up !
This regulation prevented <i>US banks</i> THEME <i>located in the US</i> LOCATION , but not abroad, from paying interest on deposits above a given rate .	This regulation prevented <i>US banks</i> THEME <i>situated in the US</i> LOCATION , but not abroad, from paying interest on deposits above a given rate .

Table 2: Parallel lexicographic annotations of sister LUs (expert-annotated) and empty LUs (augmented by our approach). Frame Element (FE) names are in SMALL CAPS and their instantiations are color-coordinated; and LUs or targets are in *italics*.