Generating Lexical Representations of Frames using Lexical Substitution

Saba Anwar
Universität Hamburg
Germany
anwar@informatik.uni-hamburg.de

Artem Shelmanov
Skolkovo Institute of Science and Technology
Russia
a.shelmanov@skoltech.ru

Alexander Panchenko
Skolkovo Institute of Science and Technology
Russia
a.panchenko@skoltech.ru

Chris Biemann
Universität Hamburg
Germany
biemann@informatik.uni-hamburg.de

Abstract

Semantic frames are formal linguistic structures describing situations/actions/events, e.g. *Commercial transfer of goods*. Each frame provides a set of roles corresponding to the situation participants, e.g. *Buyer* and *Goods*, and lexical units (LUs) – words and phrases that can evoke this particular frame in texts, e.g. *Sell*. The scarcity of annotated resources hinders wider adoption of frame semantics across languages and domains. We investigate a simple yet effective method, lexical substitution with word representation models, to automatically expand a small set of frame-annotated sentences with new words for their respective roles and LUs. We evaluate the expansion quality using FrameNet. Contextualized models demonstrate overall superior performance compared to the non-contextualized ones on roles. However, the latter show comparable performance on the task of LU expansion.

1 Introduction

The goal of lexical substitution (McCarthy and Navigli, 2009) is to replace a given target word in its context with meaning-preserving alternatives. In this paper, we show how lexical substitution can be used for semantic frame expansion. A semantic frame is a linguistic structure used to describe the formal meaning of a situation or event (Fillmore, 1982). Semantic frames have witnessed a wide range of applications; such as question answering (Shen and Lapata, 2007; Berant and Liang, 2014; Khashabi et al., 2018), machine translation (Gao and Vogel, 2011; Zhai et al., 2013), and semantic role labelling (Do et al., 2017; Swayamdipta et al., 2018). The impact, however, is limited by the scarce availability of annotated resources. Some publicly available resources are FrameNet (Baker et al., 1998) and PropBank (Palmer et al., 2005), yet for many languages and domains, specialized resources do not exist. Besides, due to the inherent vagueness of frame definitions, the annotation task is challenging and requires semanticists or very complex crowd-sourcing setups (Fossati et al., 2013).

Seed sentence: I hope **Patti**$_{\text{Helper}}$ can **help**$_{\text{Assistance}}$ **you**$_{\text{Benefited\_party}}$ **soon**$_{\text{Time}}$.
Substitutes for Assistance: assist, aid **Substitutes for Helper:** she, I, he, you, we, someone, they, it, lori, hannah, paul, sarah, melanie, pam, riley **Substitutes for Benefited_party:** me, him, folk, her, everyone, people **Substitutes for Time:** tomorrow, now, shortly, sooner, tonight, today, later

Table 1: An example of the induced lexical representation (roles and LUs) of the Assistance FrameNet frame using lexical substitutes from a single seed sentence.

We suggest a different perspective on the problem: expanding the FrameNet resource automatically by using lexical substitution. Given a small set of seed sentences with their frame annotations, we can expand it by substituting the *targets* (words corresponding to lexical units of the respective frame) and *arguments* (words corresponding to roles of the respective frame) of those sentences and aggregating possible substitutions into an induced frame-semantic resource. Table 1 shows one such induced example. For this purpose, we have experimented with state-of-the-art non-contextualized (static) word representation models including neural word embeddings, i.e. fastText (Bojanowski et al., 2017), GloVe (Pennington et al., 2014), and word2vec (Mikolov et al., 2013); and distributional thesaurus, i.e. JoBimText (Bie-

Proceedings of the Probability and Meaning Conference (PaM 2020), pages 95–103

mann and Riedl, 2013); and compared their results with contextualized word representations of the state-of-the-art BERT model (Devlin et al., 2019), which has set a new benchmark performance on many downstream NLP applications. To complete the comparison, we also include the lexical substitution model of Melamud et al. (2015), which uses dependency-based word and context embeddings and produces context-sensitive lexical substitutes.

To generate substitutes, we decompose the problem into two sub-tasks: **Lexical unit expansion**: Given a sentence and its *target* word, the task is to generate frame preserving substitutes for this word. **Frame role expansion**: Given a sentence and an *argument*, the task is to generate meaning-preserving substitutes for this argument.

Contributions of our work are (i) a *method* for inducing frame-semantic resources based on a few frame-annotated sentences using lexical substitution, and (ii) an *evaluation* of various distributional semantic models and lexical substitution methods on the ground truth from FrameNet.

2 Related Work

Approaches to semantic frame parsing with respect to a pre-defined semantic frame resource, such as FrameNet, have received much attention in the literature (Das et al., 2010; Oepen et al., 2016; Yang and Mitchell, 2017; Peng et al., 2018), with SEMAFOR (Das et al., 2014) being a most widely known system to extract complete frame structure including target identification. Some works focus on identifying partial structures such as frame identification (Hartmann et al., 2017; Hermann et al., 2014), role labelling with frame identification (Swayamdipta et al., 2017; Yang and Mitchell, 2017), and simple role labelling (Kshirsagar et al., 2015; Roth and Lapata, 2015; Swayamdipta et al., 2018), which is considered very similar to standard PropBank (Palmer et al., 2005) style semantic role labelling, albeit more challenging because of the high granularity of frame roles. These supervised models rely on a dataset of frame-annotated sentences such as FrameNet. FrameNet-like resources are available only for very few languages and cover only a few domains. In this paper, we venture into the inverse problem, the case where the number of annotations is insufficient, similar to the idea of Pennacchiotti et al. (2008) who investigated the utility of semantic spaces and WordNet-based methods to automatically induce new LUs and reported their results on FrameNet.

Our method is inspired by the recent work of Amrami and Goldberg (2018). They suggest to predict the substitutes vectors for target words using pre-trained ELMo (Peters et al., 2018) and dynamic symmetric patterns, then induced the word senses using clustering. Arefyev et al. (2019) takes the idea of substitute vectors from (Amrami and Goldberg, 2018) for the SemEval 2019 (QasemiZadeh et al., 2019) frame induction task and replaces ELMo with BERT (Devlin et al., 2019) for improved performance. Zhou et al. (2019) show the utility of BERT for the lexical substitution task. Lexical substitution has been used for a range of NLP tasks such as paraphrasing or text simplification, but here, we are employing it, as far as we are aware, for the first time to perform expansion of frame-semantic resources.

3 Inducing Lexical Representations of Frames via Lexical Substitution

We experimented with two groups of lexical substitution methods. The first one use no context: non-contextualized neural word embedding models, i.e. fastText (Bojanowski et al., 2017), GloVe (Pennington et al., 2014), and word2vec (Mikolov et al., 2013), as well as distributional thesaurus based models in the form of JoBimText (Biemann and Riedl, 2013). The second group of methods does use the context: here, we tried contextualized word embedding model BERT (Devlin et al., 2019) and the lexical substitution model of Melamud et al. (2015).

3.1 Static Word Representations

These word representations models are inherently non-contextualized as they learn one representation of a word regardless of its context.

Neural Word Embeddings Neural word embeddings represent words as vectors of continuous numbers, where words with similar meanings are expected to have similar vectors. Thus, to produce substitutes, we extracted the k nearest neighbors using a cosine similarity measure. We use pre-trained embeddings by authors models: fastText trained on the Common Crawl corpus, GloVe trained on Common Crawl corpus with 840 billion words, word2vec trained on Google News. All these models produce 300-dimension vectors.

Distributional Thesaurus (DT) In this approach, word similarities are computed using complex linguistic features such as dependency relations (Lin, 1998). The representations provided by DTs are sparser, but similarity scores based on them can be better. JoBimText (Biemann and Riedl, 2013) is a framework that offers many DTs computed on a range of different corpora. Context features for each word are ranked using the lexicographer's mutual information (LMI) score and used to compute word similarity by feature overlap. We extract the k nearest neighbors for the target word. We use two JoBimText DTs: (i) DT built on Wikipedia with n-grams as contexts and (ii) DT built on a 59G corpus (Wikipedia, Gigaword, ukWaC, and LCC corpora combined) using dependency relations as context.

3.2 Contextualized Models

Static word representations fail to handle polysemic words. This paves the way for context-aware word representation models, which can generate diverse word-probability distributions for a target word based on its context.

Melamud et al. (2015) This simple model uses syntax-based skip-gram embeddings (Levy and Goldberg, 2014) of a word and its context to produce context-sensitive lexical substitutes, where the context of the word is represented by the dependency relations of the word. We use the original word and context embeddings of Melamud et al. (2015), trained on the ukWaC (Ferraresi et al., 2008) corpus. To find dependency relations, we use Stanford Parser (Chen and Manning, 2014) and collapsed the dependencies that include prepositions. Top k substitutes are produced if both the word and its context are present in the model's vocabulary. Melamud et al. (2015) proposed four measures of contextual similarity which rely on cosine similarity between context and target words, of which we report the two best performing on our task (BalAdd and BalMult).

BERT Although BERT was originally trained to restore masked tokens, it can produce a word distribution even without masking the target word. In this case, it will consider both the context and the semantics of the target word, leading to a more accurate probability distribution. For experiments, we choose one of the largest pre-trained models presented in Devlin et al. (2019), which is bert-large-cased (340M parameters) from the PyTorch implementation by Wolf et al. (2019). We produce a substitute word distribution without masking and selected substitutes with top k probabilities.

4 Experimental Setup

4.1 Datasets

We experimented with FrameNet (Baker et al., 1998) version 1.7. It contains around 170k sentences annotated with $1,014$ frames, $7,878$ types of frame roles, and $10,340$ lexical units. Frame roles and LUs can consist of a single token or multiple tokens. For this work, we have only considered a single-token substitution. The datasets for evaluation were derived automatically from FrameNet. To create a gold standard for LU expansion task, for each sentence containing an annotated LU, we consider other LUs of the corresponding semantic frame as ground truth substitutes. We keep only LUs marked as *verbs* in FrameNet. To make a gold standard for the role expansion task, for each of the sentences that contain an annotation of a given frame role, we consider all the single-word annotations from the rest of the corpus marked with the same role and related to the same frame as ground truth substitutes. The final datasets for experiments contain $79,584$ records for lexical unit expansion and $191,252$ records for role expansion (cf. Tables 4 and 5).

4.2 Evaluation Measures

To evaluate the quality of generated substitutes for a given target word, we use precision at k ($p@k$) top substitutes. To evaluate the quality of the entire list of generated substitutes, we use mean average precision at level k ($MAP@k$):

$$AP^i@k = \frac{1}{min(k, R^i)} \sum_{l=1}^{k} r_l^i \cdot p^i@l,$$

where $MAP@k = \frac{1}{N} \sum_{i=1}^{N} AP^i@k$. Here, N is a total number of examples in the dataset; R^i is a number of possible correct answers for an example i; r_l^i equals 1 if the model output at the level l is correct and 0 if not. We present $p@k$ at levels: 1, 5, 10, as well as $MAP@50$. Sometimes, the post-processing procedure leads to the generation of a list of substitutes shorter than k; we consider the absence of a substitute for a position as a wrong answer of a model.

Lexical Unit Expansion Task				
Algorithm	**p@1**	**p@5**	**p@10**	**MAP@50**
GloVe	0.359	0.243	0.195	0.127
fastText	0.374	**0.273**	**0.222**	0.151
word2vec	0.375	0.263	0.212	0.146
DT wiki	0.301	0.199	0.161	0.102
DT 59g	0.339	0.246	0.202	0.136
BalAdd	**0.380**	0.271	0.220	**0.152**
BalMult	0.379	0.270	0.220	0.151
BERT cased	0.378	0.258	0.203	0.136

Table 2: Evaluation of LU expansion.

Frame Role Expansion Task				
Algorithm	**p@1**	**p@5**	**p@10**	**MAP@50**
GloVe	0.301	0.249	0.200	0.069
fastText	0.182	0.134	0.102	0.028
word2vec	0.319	0.224	0.165	0.051
DT wiki	0.336	0.250	0.211	0.079
DT 59G	0.322	0.247	0.200	0.075
BalAdd	0.381	0.288	0.213	0.073
BalMult	0.379	0.282	0.209	0.073
BERT cased	**0.384**	**0.313**	**0.271**	**0.105**

Table 3: Evaluation of frame role expansion.

4.3 Post-processing

In post-processing, we remove numbers, symbols, special tokens from the generated list. There may also be multiple examples of the same word in different forms, especially word embeddings often produce multiple words with a shared root form. Therefore, we lemmatize the generated substitutes using the Pattern library (Smedt and Daelemans, 2012). The duplicates and the target words are dropped. For the lexical unit expansion task, as we just experiment with verbs, we drop the substitutes that cannot be verbs. We used a dictionary of verbs that aggregates verb lists taken from Pattern, WordNet (Miller, 1995), and FreeLing (Padró and Stanilovsky, 2012).

5 Results

5.1 Lexical Units Expansion Task

The results for the LU expansion task are presented in Table 2. The best performance was achieved by the BalAdd measure of Melamud et al. (2015) with $p@1 = 0.380$ and $MAP@50 = 0.152$. The fastText model achieves a comparable performance and even shows slightly better results for $p@5$ and $p@10$. The DTs considered in our experiments perform worse than word2vec, fastText, and models of Melamud et al. (2015). That is expected since the DTs need much larger datasets for training as compared to embedding-based models. Even though BERT performed comparably to fastText and word2vec, it could not outperform them except for $p@1$. However, a close examination of some examples shows that it does make a difference when the target word is polysemic.

Table 4 in the appendix contains example sentences with highlighted target words and top 5 substitutes generated by all models (along with the ground truth FrameNet annotations). The first example presents an LU that is associated with only one frame in FrameNet. Being unambiguous in meaning, all models produced many matching substitutes. The other two examples present an LU with multiple associated frames, which leads to different senses of the LU. All non-contextualized models could not produce any substitute for the *Abandonment* frame except fastText, and failed completely for the *Causation* frame, whereas BERT has successfully generated a sufficient number of matching substitutes for both examples.

5.2 Frame Role Expansion Task

The evaluation results of the methods for the frame roles expansion task are presented in Table 3. In this experiment, the non-contextualized models were outperformed by BERT with a significant margin with $p@1 = 0.384$ and $MAP@50 = 0.105$. The performance of fastText is worst compared to all models, in contrast to the previous experiment. The DTs perform substantially better than neural word embedding models. The better score is achieved by the DT trained on Wikipedia. The models of Melamud et al. (2015) achieve slightly worse results for $p@1$ and $p@5$ than BERT, but significantly lose in terms of $p@10$ and $MAP@50$.

Table 5 in the appendix enlists several substitutes for semantic roles in a hand-labelled seed sentence. The first example demonstrates several valid matching substitutes, because *Vehicle* is the most common sense of "car". Whereas, the other two examples present an argument with multiple roles. Again, BERT was able to distinguish both senses and produced valid substitutes.

6 Conclusion

We presented a simple practical technique for the generation of lexical representations of semantic frames using lexical substitution with several contextualized and static word representation models demonstrating that a single frame annotated example can be used to bootstrap a fully-fledged lexical representation of the FrameNet-style linguistic structures. Non-contextualized baseline models proved to be strong baselines, but failed to produce good substitutes for polysemic words (same word but different semantic frame), whereas BERT for such cases produced competitive substitutes. A prominent direction for future work is testing the proposed technology for building frame representations of low-resource languages and domains.

Acknowledgements

We thank the anonymous reviewers for valuable feedback and acknowledge the support of the Deutsche Forschungsgemeinschaft (DFG) under the "JOIN-T 2" project (BI 1544/4-2), the German Academic Exchange Service (DAAD) and the Higher Education Commission (HEC), Pakistan. The work of Artem Shelmanov in writing and experiments with BERT model was supported by the Russian Science Foundation, project #20-11-20166 "Cross-lingual Knowledge Base Construction and Maintenance".

References

Asaf Amrami and Yoav Goldberg. 2018. Word sense induction with neural biLM and symmetric patterns. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4860–4867, Brussels, Belgium. Association for Computational Linguistics.

Nikolay Arefyev, Boris Sheludko, Adis Davletov, Dmitry Kharchev, Alex Nevidomsky, and Alexander Panchenko. 2019. Neural GRANNy at SemEval-2019 task 2: A combined approach for better modeling of semantic relationships in semantic frame induction. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 31–38, Minneapolis, MN, USA. Association for Computational Linguistics.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Montréal, QC, Canada. Association for Computational Linguistics.

Jonathan Berant and Percy Liang. 2014. Semantic parsing via paraphrasing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1415–1425, Baltimore, MD, USA. Association for Computational Linguistics.

Chris Biemann and Martin Riedl. 2013. Text: now in 2D! A framework for lexical expansion with contextual similarity. *J. Language Modelling*, 1(1):55–95.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Danqi Chen and Christopher Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 740–750, Doha, Qatar. Association for Computational Linguistics.

Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. 2014. Frame-semantic parsing. *Computational Linguistics*, 40:9–56.

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, CA, USA. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, MN, USA. Association for Computational Linguistics.

Quynh Ngoc Thi Do, Steven Bethard, and Marie-Francine Moens. 2017. Improving implicit semantic role labeling by predicting semantic frame arguments. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 90–99, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Adriano Ferraresi, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. 2008. Introducing and evaluating ukwac, a very large web-derived corpus of english. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4)*.

Charles J. Fillmore. 1982. Frame Semantics. *In The Linguistic Society of Korea, eds. Linguistics in the Morning Calm. Seoul: Hanshin*, pages 111–137.

Marco Fossati, Claudio Giuliano, and Sara Tonelli. 2013. Outsourcing FrameNet to the crowd. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 742–747, Sofia, Bulgaria. Association for Computational Linguistics.

Qin Gao and Stephan Vogel. 2011. Utilizing target-side semantic role labels to assist hierarchical phrase-based machine translation. In *Proceedings of Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 107–115, Portland, OR, USA. Association for Computational Linguistics.

Silvana Hartmann, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. Out-of-domain FrameNet semantic role labeling. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 471–482, Valencia, Spain. Association for Computational Linguistics.

Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1448–1458, Baltimore, MD, USA. Association for Computational Linguistics.

Daniel Khashabi, Tushar Khot, Ashish Sabharwal, and Dan Roth. 2018. Question answering as global reasoning over semantic abstractions. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, (AAAI-18)*, pages 1905–1914, New Orleans, LA, USA. Association for the Advancement of Artificial Intelligence.

Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith, and Chris Dyer. 2015. Frame-semantic role labeling with heterogeneous annotations. 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 218–224, Beijing, China. Association for Computational Linguistics.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308, Baltimore, MD, USA. Association for Computational Linguistics.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL '98/COLING '98, pages 768–774, Montreal, QC, Canada. Association for Computational Linguistics.

Diana McCarthy and Roberto Navigli. 2009. The English lexical substitution task. *Language Resources and Evaluation*, 43(2):139–159.

Oren Melamud, Omer Levy, and Ido Dagan. 2015. A simple word embedding model for lexical substitution. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 1–7, Denver, CO, USA. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., Harrahs and Harveys, NV, USA.

George A. Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajic, Angelina Ivanova, and Zdenka Uresova. 2016. Towards comparability of linguistic graph banks for semantic parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3991–3995, Portorož, Slovenia. ELDA.

Lluís Padró and Evgeny Stanilovsky. 2012. FreeLing 3.0: Towards wider multilinguality. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2473–2479, Istanbul, Turkey. European Language Resources Association (ELRA).

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106.

Hao Peng, Sam Thomson, Swabha Swayamdipta, and Noah A. Smith. 2018. Learning joint semantic parsers from disjoint data. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1492–1502, New Orleans, LA, USA. Association for Computational Linguistics.

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 457–465, Honolulu, Hawaii. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies, Volume 1 (Long Papers), NAACL-HLT 2018, pages 2227–2237, New Orleans, LA, USA. Association for Computational Linguistics.

Behrang QasemiZadeh, Miriam R. L. Petruck, Regina Stodden, Laura Kallmeyer, and Marie Candito. 2019. SemEval-2019 task 2: Unsupervised lexical frame induction. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 16–30, Minneapolis, MN, USA. Association for Computational Linguistics.

Michael Roth and Mirella Lapata. 2015. Context-aware frame-semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:449–460.

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.

Tom De Smedt and Walter Daelemans. 2012. Pattern for python. *Journal of Machine Learning Research*, 13(Jun):2063–2067.

Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. 2017. Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold. *arXiv preprint arXiv:1706.09528*.

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.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*, abs/1910.03771.

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.

Feifei Zhai, Jiajun Zhang, Yu Zhou, and Chengqing Zong. 2013. Handling ambiguities of bilingual predicate-argument structures for statistical machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1127–1136, Sofia, Bulgaria. Association for Computational Linguistics.

Wangchunshu Zhou, Tao Ge, Ke Xu, Furu Wei, and Ming Zhou. 2019. BERT-based lexical substitution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3368–3373, Florence, Italy. Association for Computational Linguistics.

A Examples of Induced Lexical Semantic Frame Representations

This appendix contains additional examples of lexical substitutions of lexical units (LUs) and roles of the semantic frames resource along with the ground truth from FrameNet. Examples of the LU expansions are presented in Table 4 while roles are presented in Table 5.

Frame: Statement
Seed sentence: The report **stated**$_{\text{state}}$, however , that some problems needed to be solved , principally that of lack of encouragement of cadres and individuals to exercise their democratic right of freedom of expression .
GloVe: explain, note, agree, acknowledge, mention **fastText:** note, explain, indicate, reiterate, opine **word2vec:** comment, note, assert, remark, explain **DT wiki:** say, note, claim, comment, suggest **DT 59g:** note, say, claim, comment, think **BalAdd:** indicate, stipulate, assert, reiterate, say **BalMult:** indicate, stipulate, assert, say, aver **BERT:** say, find, conclude, note, declare **FrameNet gold:** proclaim, mention, claim, detail, profess, tell, caution, allow, propose, comment, preach, reaffirm, avow, challenge, recount, reiterate, pronounce, relate, remark, report, say, speak, state, allege, suggest, conjecture, talk, write, contend, venture, declare, add, hazard, pout, announce, exclaim, smirk, address, confirm, explain, assert, gloat, acknowledge, insist, maintain, note, observe, aver, refute, attest, describe
Frame: Abandonment
Sentence: When their changes are completed , and after they have worked up a sweat , ringers often skip off to the local pub , **leaving**$_{\text{leave}}$ worship for others below .
GloVe: return, back, left, rest, stay **fastText:** left, abandon, return, rejoin, exit **word2vec:** left, return, depart, exit, enter **DT wiki:** visit, enter, join, reach, represent **DT 59g:** visit, enter, occupy, beat, represent **BalAdd:** abandon, quit, allow, depart, prefer **BalMult:** abandon, allow, quit, prefer, cause **BERT:** give, abandon, do, let, left **FrameNet gold:** leave, abandon, forget
Frame: Causation
Seed sentence: Older kids , like Tracy and Kerry , **left**$_{\text{leave}}$ homeless after a recent murder - suicide in Indianapolis claimed Mom and Dad.
GloVe: right, back, left, off, rest **fastText:** left, right, return, lurch, move **word2vec:** return, right, depart, limp, go **DT wiki:** left, right, break, curve, rear **DT 59g:** left, right, break, swell, enlarge **BalAdd:** left, gash, vacate, depart, jolt **BalMult:** left, vacate, gash, jolt, depart **BERT:** left, send, raise, make, help **FrameNet gold:** cause, leave, mean, render, wreak, bring, dictate, sway, force, make, precipitate, send, raise, motivate, induce, put, see

Table 4: LU expansion examples. Green highlighting indicates matches with the gold annotations.

Frame: Vehicle
Seed sentence: I noticed the **car**$_{\text{Vehicle}}$ was bouncing up and down as if someone were jumping on it.
GloVe: vehicle, automobile, truck, auto, drive **fastText:** vehicle, automobile, car–and, car.but, car.it **word2vec:** vehicle, suv, minivan, truck, ford_focu **DT wiki:** vehicle, automobile, truck, sedan, bus **DT 59g:** vehicle, truck, automobile, sedan, jeep **BalAdd:** vehicle, bike, minivan, land-rover, horsebox **BalMult:** vehicle, bike, minivan, land-rover, passat **BERT:** thing, convertible, vehicle, sedan, cruiser **FrameNet gold:** helicopter, airplane, ship, vessel, subway, boat, vehicle, stryker, tank, truck, aircraft, bike, bus, car, train, plane, cab, carriage, automobile, buse, ferry, tram, sedan, taxi, tricycle, submarine, yacht, aeroplane, chopper
Frame: Part_orientational
Seed sentence: Repton was an Anglo-Saxon town, on the south **bank**$_{\text{Part}}$ of the River Trent, and was at one time a chief city of the Kingdom of Mercia.
GloVe: draft, financial, credit, lender, loan **fastText:** bank.the, bank.it, bank.thi, bank.so, bank. **word2vec:** draft, lender, banker, depositor, mortgage_lender **DT wiki:** shore, company, draft, lender, embankment **DT 59g:** lender, company, insurer, draft, brokerage **BalAdd:** aib, citibank, hsbc, bundesbank, riksbank **BalMult:** citibank, aib, hsbc, tsb, bundesbank **BERT:** side, shore, river, west, fork **FrameNet gold:** bottom, rear, north, north-south, northwest, west, side, territory, western, end, south, acquifer, back, left, window, top, heart, face, dynasty, tip, front, coast, southern, northernmost, northern, part, eastern, aegean, base, peak, area, portion, island, edge, sliver, strip, region, east, bank, fork, aisle, wall, shore, feet, leg, paw, quarter, wing, femora, half, halve, reach, slope, sea-board, borderland, ring, step, drawer, lip, realm, claw, border, ridge, foot, summit, door, gate, apse, façade, hemisphere, boundary, section, entrance, province, point, apex, corner, axle, page, pocket, seat, stair, underbelly, crest, layer, floor, button, shelf, flank, frontier, peninsula, hill, underside, coastline, spoiler, tailcone, panel, wheel
Frame: Abounding_with
Seed sentence: For their sledging trick, they love a steep, snow covered **bank**$_{\text{Location}}$ and will lie on the top, facing downhill, then tuck up their front paws so that they slide along upon their chests.
GloVe: draft, financial, credit, lender, loan **fastText:** bank.the, bank.it, bank.thi, bank.so, bank. **word2vec:** draft, lender, banker, depositor, mortgage_lender **DT wiki:** shore, company, draft, lender, embankment **DT 59g:** lender, company, insurer, draft, brokerage **BalAdd:** cahoot, citibank, hsbc, tsb, draft **BalMult:** cahoot, draft, citibank, hsbc, natwest **BERT:** slope, hill, ditch, mountain, river **FrameNet gold:** ringer, it, kitchen, hill, equipment, island, street, nut, place, which, plimsoll, paper, bread, roll, egg, scone, tin, salmon, dish, potatoe, kavo, hillside, fiord, sea, pottery, cuff-link, porcelain, bowl, room, somethe, that, pocket, hand, gorget, finger, office, bookshelve, stall, animal, bird, mushroom, olive, folder, fish, pepper, pension, panel, door, donut, stoneware, tile, window, eye, veal, walnut, i, jeep, collection, frame, mirror, everythe, bedroom, barge, easel, desk, arbour, bank, bar, cinema, appearance, raspberry, ful, glass, mug, tankard, river, goblet, pew, skin, ceil, bookcase, figure, face, plaster, wall, wood, buse, fishing-boat, sign, poplar, curtain, promenade, avenue, pasture, land, another, weapon, bottle, ditch, everywhere, meadow, pasta, depression, church, sandbag, sofa, bubble, car, countryside, closet, hallway, pond, train, road, home, accommodation, dwelling, fireplace, floor, roof, corridor, uniform, bed, oak, bath, dump, nylon, chalet, balcony, machinery, reef, overhead, belt, path, roadway, area, courtyard, terrace, entrance, character, liverpool, toenail, shaft, object, neck, fingerboard, they, unit, table, pot, fingernail, moccasin, tray, goldie, peach, inn, ingushetia, sidewalk, mast, nail, floorboard, rail, plywood, launch, cabin-top, toy, she, anglo-saxon

Table 5: Role expansion examples: Green highlighting indicates matches with the gold annotations.