

AUGMENTING LINGUISTIC SEMI -STRUCTURED DATA FOR MACHINE LEARNING - A CASE STUDY USING FRAMENET

Breno W. S. R. Carvalho¹, Aline Paes² and Bernardo Gonçalves³

¹IBM Research, Brazil. Institute of Computing, Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil.

²Institute of Computing, Universidade Federal Fluminense (UFF), Niterói, RJ, Brazil.

³IBM Research, Brazil

ABSTRACT

Semantic Role Labelling (SRL) is the process of automatically finding the semantic roles of terms in a sentence. It is an essential task towards creating a machine-meaningful representation of textual information. One public linguistic resource commonly used for this task is the FrameNet Project. FrameNet is a human and machine-readable lexical database containing a considerable number of annotated sentences, those annotations link sentence fragments to semantic frames. However, while the annotations across all the documents covered in the dataset link to most of the frames, a large group of frames lack annotations in the documents pointing to them. In this paper, we present a data augmentation method for FrameNet documents that increases by over 13% the total number of annotations. Our approach relies on lexical, syntactic, and semantic aspects of the sentences to provide additional annotations. We evaluate the proposed augmentation method by comparing the performance of a state-of-the-art semantic-role-labelling system, trained using a dataset with and without augmentation.

KEYWORDS

FrameNet, Frame Semantic Parsing, Semantic Role Labelling, Data Augmentation.

1. INTRODUCTION

A large proportion of humankind's knowledge is stored in textual form. Nevertheless, such unstructured information is hard to search, catalog, and query. To circumvent this difficulty, one needs to automate the extraction of information from texts, making it amenable for querying. It relates to the emerging area of Machine Reading [1], a task within the broader area of Natural Language Processing, NLP. Machine Reading is concerned explicitly with creating machine-friendly, yet nuanced, representations of text. A crucial task in Machine Reading is the Semantic Role Labeling task, SRL [2]. SRL consists of mapping elements of a given sentence to predefined sets of semantic roles. There are two main kinds of labeling: (i) deep labeling, i.e., the mapping of tokens of the sentence to somewhat complex semantic structures by building a composable representation of the utterance meaning; and (ii) shallow labeling, that consists of mapping the tokens to an abstract semantic role. For instance, figure 1 shows two shallow roles, namely, Content and Paradigm, which provide meaning to two subsets of tokens in the sentence. The present work is concerned with shallow labeling, which is itself far from a trivial computational

task and is hardly feasible without a good set of labeled sentences, whereby “good” we mean a set of sentences whose tokens are annotated with their expected deep roles in relatively good coverage.

The formation of black holes should be understood in astrophysic terms.
Content Paradigm

Figure 1: An example of shallow semantic roles assigned to tokens in a sentence.

One popular source of annotated sentences to support Machine Reading is FrameNet, a publicly available electronic language resource [3]. It consists of a network of concepts (called frames) such as Run, Motive and Location. Each frame is composed of frame elements, which define semantic roles in the (thereby semi-structured) domains. A key technical challenge, however, is that FrameNet’s set distribution of examples forms a long tail — a few frame elements have several examples over their related frames. In contrast, most of them have only one or none example at all — making it difficult to tackle less popular frame elements. This need gets even more pressing when we target specific domains within FrameNet.

In this paper, we propose a data augmentation method to enlarge the set of annotations and its distribution in FrameNet. The technique leverages on partial structure present in the annotation of frame elements in the sentences. That is, we carry out matching of frame elements over different frames — relying on notions of lexical, syntactic, or semantic equivalence — so that sentences receive new (inferred) annotations. We take advantage of the inter-frame connections to enrich the information available in the resource.

In the next section, we describe the analyzers that enable us to process natural language sentences, the SRL method that supports our evaluation, and we provide a more detailed view on FrameNet. Then we also introduce background aspects, preparing towards our research problem. In section 3 we present the augmentation method we propose in this paper. In section 4 we report its evaluation, based on comparing the performance of a state-of-the-art semantic-role-labeling method, with and without augmentation. In section 5, we situate this work within the literature through a discussion of related work. In section 6, we conclude the paper and point challenges and future work.

2. BACKGROUND

There are three core materials used in our work: the sentence analyzers, the semantic-role-labeling method, and FrameNet itself. Boxer and spaCy are, respectively, the semantic and syntactic analyzers. Open-Sesame is the semantic-role-labeling method that supports the evaluation of our proposed method. FrameNet provides us with the annotated sentences that can support machine-reading and that we want to augment.

2.1. Boxer and Spacy: Semantic and Syntactic Analyzers

Boxer is an open-domain semantic analyzer [5] based on Combinatorial Categorical Grammars and Discourse Representation Theory. It generates a neo-Davidsonian representation of sentences. We also use it as a syntactic analyzer, the dependence tree parser, and the part-of-speech tagging system provided by the spaCy NLP library (version 2.0.11).

To process the different representations that we generate, we convert them all to a standard logical form. The Boxer analysis result is a bit tricky to normalize. Although it is already provided in first-order logic, we still need to do variable grounding, followed by Skolemization.

We also remove any negated terms and unbound variables left in order to have a simple graph structure. Figures 2 and 3 show examples of those analyzers in action.

Alice and **Bob** were **arrested**
yesterday in the beach.

```
pernambob(b),  
pernamalice(a),  
r1Time(v,y),  
n1yesterday(y),  
r1Theme(v,s),  
v1arrest(v),  
r1subset_of(b,s),  
r1subset_of(a,s),  
n1beach(p),  
r1in(v,p).
```

Figure 2: Semantic Analysis by Boxer. Predicates (e.g., ‘v1arrest’) define the so-called thematic roles such as agent, theme, action etc., other semantic roles such as person name (pernam) and even nouns like beach.

Every predicate (except for the person name one) is prefixed by its syntactic role as well.

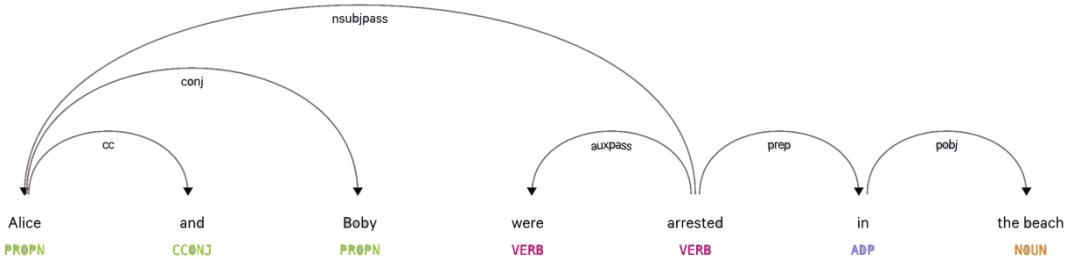


Figure 3: Syntactic Analysis by spaCy. The node labels (associated with the sentence tokens, e.g., ‘VERB’) give the part-of-speech tags, and the edge labels (associated with the tokens relationships, e.g., ‘conj’) are universal dependence labels.

2.2. Open-Sesame: the Supportive Semantic-Role-Labeling Method

Open-Sesame [6] is a state-of-the-art method for frame semantic parsing. This method is based on a segmental recurrent neural network [7], that supports its aim argument identification. It does not rely on syntactic representations during the testing phase, only during training. This way, this system presents itself as a cheaper alternative — regarding computational resources and human effort — to develop the syntactic parsers, while stays a competitive approach to the traditional pipeline that we follow in our work.

2.3. Framenet

We provide in this section a more detailed overview of FrameNet that suffices for the purpose of this paper. For a rigorous and comprehensive description of the FrameNet project, we refer the reader to Fillmore et al. [3]. In this work, we use FrameNet version 1.5.

FrameNet is an interconnected network of frames which provides the grounding for a cross-domain semantic representation. In this context, frames represent concepts like Arrest, Coming to Believe and Event. Those concepts also describe semantic roles that entities might have related to those concepts. For instance, some of the semantic roles described in the frame Arrest are Authority, Suspect and Place. Those semantic roles are called frame elements. Each frame

element occurring in a frame has its definition, written in a human-friendly form. Those definitions usually carry an example sentence where the frame elements are annotated as well as the frame itself. This way, we have both frame annotations, also called targets, and frame-element annotations together. For simplicity, we are going to refer to frame-element annotations just as annotations for the rest of the paper.

2.4. The Semantic Role Labeling (SRL) task from FrameNet's point of view

Here, we revisit the semantic role labeling (SRL) task, focusing on how FrameNet supports it as a resource. In doing so, we prepare for our specific research problem of augmenting FrameNet's semi-structured data in the next section.

In FrameNet, the sentences are annotated by humans. The general task of automatically generating those annotations is called frame-semantic parsing, which has SRL as one of its three components. Given a sentence, (i) target identification is the task of finding which token in the sentence should be matched to a frame; (ii) frame identification means to take a given token and assign it to a specific frame, and (iii) argument identification (SRL) is the task of matching frame elements that are members of the selected frames to the correct tokens in the sentence.

The SRL task induces our semi-structured data augmentation problem since SRL relies on a good set of annotated sentences as examples.

As discussed in the previous sections, FrameNet is a widely used resource supporting several NLP tasks. However, as a manually-built resource, it is error-prone and incomplete. For instance, fig. 7a shows that the frame coverage in FrameNet, that is, the number of frames that appear in at least one annotated sentence divided by the total number of frames, is only 70%.

In this work, we intend to increase this coverage so that NLP tasks in general — and SRL in particular — benefit from more frame annotations available. If we can achieve some increase in frame annotations coverage, even if it is not very large, it is bound to provide a relevant contribution to the machine reading community. That is because annotated sentences feed in all machine reading pipelines.

3. AUGMENTATION OF FRAMENET EXAMPLES

We start to state the data augmentation problem by introducing an example and follow it with our proposed methodology.

3.1. The Data Augmentation Problem

Consider the sentence “Most of us know where we took a photo but have a harder time remembering the time we took it.”, and assume that Create physical artwork be one correct frame identified with this sentence. The annotation of this sentence concerning the structure of frame Create physical artwork is depicted in Fig. 4. There are three frame elements of that frame, namely, Creator, Representation, and Location of representation, which are mapped to subsets of tokens in the sentence.

From a general point of view, the data augmentation problem in this context is to ask how we could create a new annotation of this sentence using the tokens already mapped to frame elements of the frame Create physical artwork. The goal is to use the already marked tokens to annotate the sentence for another frame.

Target	Representation	Location of representation
I DREW	his picture	on a sheet of paper first.

Figure 4: Create physical artwork annotation with respect to the frame Intentionally create.

Now consider Intentionally create, another frame which is related to Create physical artwork by the ‘has sub-frame of’ relation, as shown in Fig. 5. We exploit such inter-frame relations and then model the data augmentation problem accordingly. In our running example, the problem is reduced to whether or not we could build a new annotation of the sentence in terms of the structure of frame Create physical artwork. The new annotation must comprise not only the frame itself using the target token, but also its frame elements, namely, Creator, Created entity, and Place. It is quite intuitive that Creator from Create physical artwork should map to the frame element of same name from Intentionally create. The frame elements Created entity and Place from Create physical artwork should map to Created entity, and Place from Intentionally create, respectively.

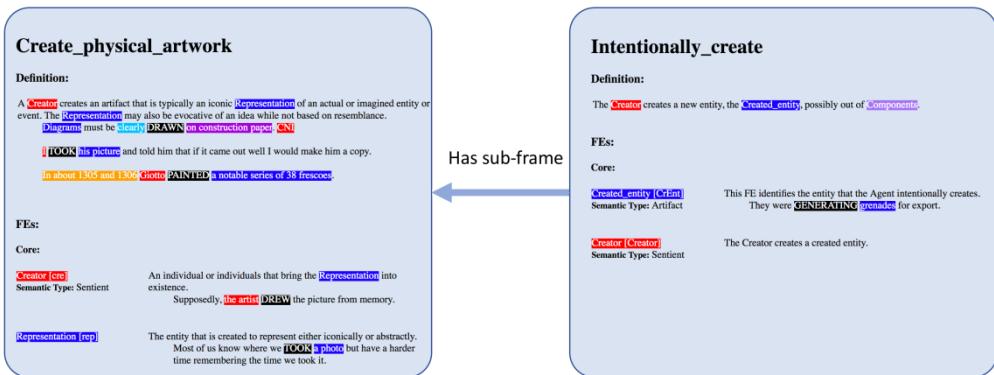


Figure 5: Intentionally create and Create physical artwork frames

3.2. The Notion of Frame Elements Equivalence

Frame elements equivalence is a rather vague concept. We model it in terms of three different notions of equivalence: lexical, semantic, and syntactic. We say that two frame elements from X and Y, respectively, are lexically equivalent if they have the same name. Two frame elements are said syntactically equivalent if there is at least one pair of examples from X and Y where these frame elements appear, and they have the same path of syntactic roles to the target in a syntactic representation. The semantic similarity follows the same concept of the syntactic equivalence, but, instead, we require a path of semantic roles turned into a semantic representation.

Consider the frames X and Y and an annotated sentence x with annotations of frame elements in X. Given that X is related to Y through one of the possible inter-frame relations (e.g., ‘is sub-frame of’), we want to find what annotations we could extend to Y. That is, we want to know if there can be a new annotation of the sentence regarding the frame elements belonging to Y. So, we will say that x is transferable from X to Y if all the frame element annotations in x are transferable to Y. Recall from section 2 that there are two kinds of annotations in an annotated sentence, namely: targets and frame element annotations. The second one we call annotations. An annotation is transferable from X to Y if its frame element is equivalent to one frame element in Y. This assured, we can rewrite the sentence annotation using frame elements of Y, and we can add a new annotation to the sentence.

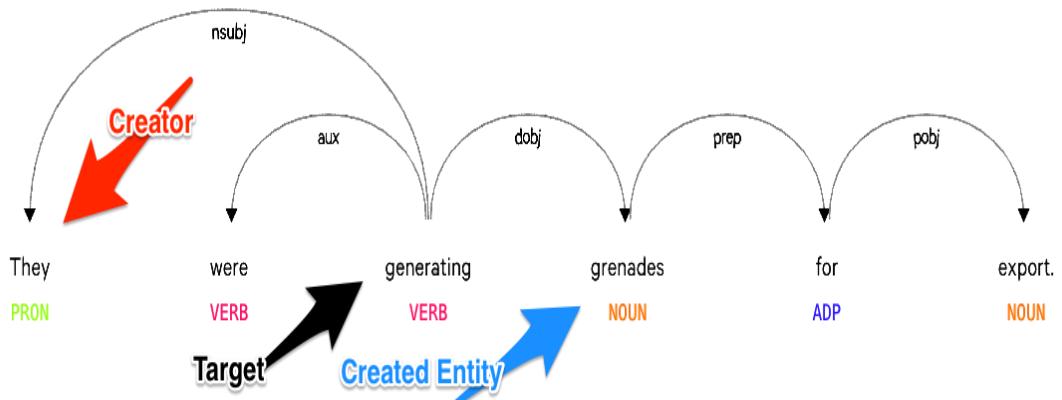
Let us recall the example depicted in figure 5. In order to know if this annotation can be adapted to another frame Create physical artwork, we first have to check if all frame elements of Intentionally create in the annotated sentence are equivalent to some frame element in Create physical artwork. Using the notion of lexical equivalence, we consider Creator to be the same as Creator in Create physical artwork as they both have the same name. Using the syntactic equivalence, we need to check if Created entity is equivalent to Representation. To do that, we take an example of Created entity from Intentionally create and one example of Representation from Create physical artwork and check if the syntactic path to the target is the same, as exhibited in figure 6. Since each frame element in the annotation is equivalent to some frame element in Create physical artwork, we can copy this example to Create physical artwork. If there were any frame elements left that have not an equivalent frame element in Create physical artwork, then the next step would be to check their semantic equivalence the same way we did for the syntactic equivalence.

The same method described before for expanding a frame example is used to expand annotated sentences from the FrameNet Project annotated documents. We show the results of this heuristic on whether we can borrow an annotated sentence in section 4.

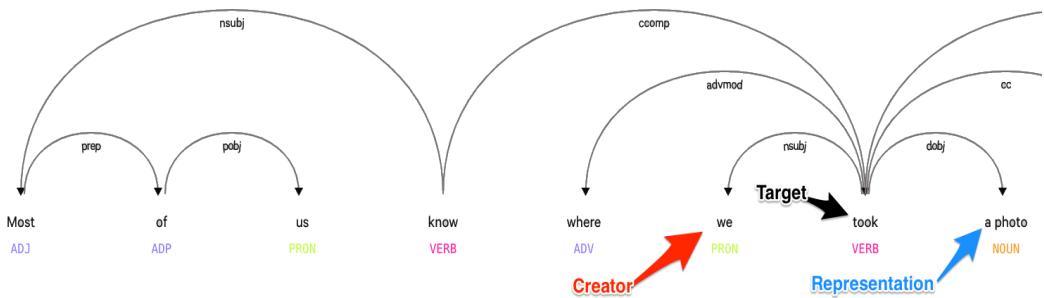
It is clear that ‘ways for people with disability to enter the workforce’ is not necessarily a piece of physical artwork as this augmented annotation suggests.

3.3. Frame Relations

To elaborate the proposed heuristics, we start by splitting the FrameNet inter-frame relations into two sets: (i) The set of hierarchical relations, depicted in the table 1, are the ones based in the inheritance and part-of concepts, and their reciprocal. (ii) The set of non-hierarchical relations comprises all the other relations and is depicted in table 2. This split is used to evaluate the effect of inheritance on the creation of new annotations. For instance, it is reasonable to think that annotations transferred from the frame Create physical artwork to its parent frame Intentionally create would be correct. Usually, the creation of an artwork is intentional, and all elements from the former frame have a corresponding element in the next frame.



Syntactic representation of example in Intentionally create



Syntactic representation of example in Create physical artwork

Figure 6: Syntactic representation of an example in the frame element descriptions

This way, when we say that the frame ‘Coming to believe’ inherits from ‘Event’, it means that ‘Coming to Believe’ is an ‘Event’. And when we say that a ‘Halt’ is a subframe of ‘Motion’ it means that the concept ‘halt’ is part of the concept of ‘motion’.

Table 1. Hierarchical relations

Relation	
Inherits from	is a frame of the same kind of the parent
Is Inherited by	the children frames have the same kind
Subframe of	is a part of the parent frame
Has Subframe(s)	is composed by those frames

Table 2. Non-hierarchical relations

Relation	
Perspective on	
Is Perspectivized in	
Uses	might be composed by those frames
Is Used by	might be part of the parent frame
Precedes	
Is Preceded by	
Is Inchoative of	the children are the cause of the root
Is Causative of	the root is the cause of the children
See also	Informational relation.

4. EXPERIMENTS

The purpose of the augmentation method we propose here is to increase the number of available training examples and expand the coverage over less popular frames. This augmentation is particularly useful once we consider the difficulty in manually expanding the FrameNet example set and also the difficulty of adding new documents.

4.1. Data

Our dataset consists of annotated sentences from the collection of annotated documents made available in FrameNet release 1.5. This collection consists of 78 documents annotated by FrameNet's staff; we use the same test set as [6, 8]. Those documents hold together almost 5946 annotated sentences. In those annotated sentences is a total of 23944 frame annotations and

48133 frame element annotations related to those frame annotations. The prefix, that is, the part of the document name before ‘ ’ refers to the source of the document, and the suffix is the document name. In total, there are more than 130000 sentences in the FrameNet project with some kind of annotation. More on the construction of this dataset and FrameNet, in general, is found in [9].

4.2. Evaluation Setting

We evaluate the augmentation strategies based on the improvement of the performance of a state-of-the art method in the literature, Open-Sesame. Each one of the multiple training instances is carried out until the same termination criterion is reached, for conformity and ease of comparison, the criterion is the same used in the Open-Sesame paper, we also used the default parameters reported in that paper [6]. This criterion is met when there were no updates in the best loss score reported after 28 validation epochs.

We used the same GloVe embedding [10] and optimized the model using ADAM [11], with a learning rate of 0.0005, and moving average parameter of 0.01. We also set the moving average variance to 0.9999, and we set the parameter (to prevent numerical instability) to 10^{-8} ; no learning rate decay is used, as done in the original Open-Sesame paper.

4.3. Results

We evaluated three kinds of augmentation in this project, namely lexical, syntactic, and semantic analysis (described in section 3). The overall gain on number of annotations from each one of those strategies is depicted in figures 7b, 7c, and 7d, respectively. We see a moderate increase of over roughly 13% of the original coverage using the different kinds of augmentations separately depicted in figure 8. This gain indicates that besides the noise addition, the augmentation strategy was beneficial to the semantic-role-labeling task.

The impact of the augmentation method on the performance of the SRL parser is expressed in table 3. Values in bold are the best values reported. We report precision, recall, and f1-score metrics micro-averaged. Our experimentation shows a small improvement in Open-Sesame’s performance when trained on datasets that undertook the augmentation strategies developed here. This improvement indicates that even with added noise, the use of the augmentation benefited the semantic parser. The annotations from the semantic and syntactic augmentation strategies did not perform better than the lexical strategy. Errors in the logical representations might cause it due to incorrect parsing of the sentences.





Figure 7: Augmentation frame coverage

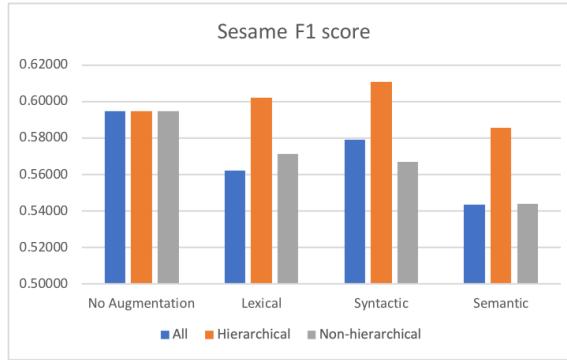


Figure 8: Comparison of Sesame F1 Score

Table 3. Performance of Sesame with the different augmentations

		Precision	Recall	F-1
Semantic	All	0.5946	0.5497	0.5712
	Hierarchical	0.5880	0.5060	0.5439
	Non-hierarchical	0.5975	0.5397	0.5671
Syntactic	All	0.5939	0.5337	0.5622
	Hierarchical	0.6041	0.4939	0.5434
	Non-hierarchical	0.6001	0.5595	0.5791
Lexical	All	0.6083	0.5955	0.6018
	Hierarchical	0.6136	0.5598	0.5854
	Non-hierarchical	0.6374	0.5865	0.6109
	No augmentation	0.5977	0.6030	0.6004

5. RELATED WORK

We considered the three main areas that we have built our contribution upon on, namely: Language resources augmentation, Sentence Representation, and Semantic Role Labeling.

5.1. Language Resources Augmentation

To the best of our knowledge, this is the first work that builds a data augmentation strategy relying only upon the data provided by FrameNet. Other venues of work combine additional language resources with FrameNet to produce SRL parsers. Shi and Mihalcea [12], Giuglea and Moschitti [13], Palmer [14], Laparra and Rigau [15], Tonelli et al. [16], and Green et al. [17] are

examples of work that combine other language resources, such as PropBank [18], VerbNet [19], and WordNet [20] with FrameNet Baker et al. [3], to complement each other or even to generate more frames. It is also possible to combine more than one of those resources; for example, the Predicate Matrix [21] is a new language resource created through the automatic combination of WordNet, Framenet, and Verbnet. Pavlick et al. [22] presents a FrameNet augmentation based on expanding the resources Lexical Units, LUs. They based their augmentation method on automatic paraphrasing using the Paraphrase Database (PPDB) [23] curated by manual crowd sourcing. The model proposed by Mousselly Sergieh and Gurevych [24] is based on word embedding to identify a mapping between Wikidata relations [25] and FrameNet frames and to annotate the arguments of each relationship with the semantic roles from the second resource. This is an example of a case where FrameNet is used to enrich other resources and is a clear contrast with our work that aims to enhance FrameNet without the use of external corpora, but only on parsing methods. This choice makes this approach flexible and agnostic of external data sources used to train those parsers.

5.2. Logical Form and Sentence Representation

Textual data is found in unstructured ways, as mentioned throughout this paper, and we want to make it as structured as possible, so it is machine-processable. Logical forms can be used to express both the syntactic and semantic aspects of the sentences of a textual document, and much work has been done on building such logical forms.

A usual step is to parse a sentence into a syntactic representation and use this intermediary representation to generate a semantic representation of the meaning covered in the sentence. In particular, [26] devise a system based on the lambda calculus for deriving neo-Davidsonian logical forms from dependency trees. They evaluate the quality of such logical forms derived from the dependency trees of the sentences by feeding those logical forms to a semantic parser. This semantic parser consists of a graph matching algorithm that matches the structure of the logical form to Freebase, a collaboratively created tuple-based knowledge base that later on was used to power Google's Knowledge Graph initiative, [27]. It generates a robust representation of the sentences and can be compared with our current approach in future work. Using this approach as our semantic parser would be a promising comparison since one of their claims is that this representation outperforms a CCG-based representation which composes the Boxer method, used in our work.

Similarly, to our work, [26] creates a new neo-Davidsonian representation of sentences that might improve our current method. [28] combine logical and distributional representations. They use similarity metrics to create weighted rules using Markov Logic Networks [29]. Beltagy et al. [28] show that besides estimating the similarity between sentences, this method can also recognize textual entailment. One can use this textual entailment as another feature for our augmentation purposes.

In the same way, we rely on Boxer to obtain a logic-based parsed output. Previous work has already started from this tool to extract and represent meaning in a structured, machine-processable format from text documents. In particular, [28, 30] combined the parsed logical representation with distributional semantics and Markov Logic Networks. The distributional semantics is used to construct a unified knowledge base from different sources, while MLN is used to perform inference. The neo-Davidsonian representation and MLN are also employed to solve the Science and Math challenge, an NLP competition that aims to produce systems that can answer fifth-grade science exams, as done in [31].

The difficulties of directly applying those methods without any tinkering to our problem are that we calculate if substructures in the sentence are similar, focusing on specific terms. It is not clear

how to apply this concept to most of those methods since they are not concerned with specific terms of the sentence, but the sentence as a whole.

5.3. Semantic Role Labeling

The Semantic Role Labeling, SRL, is the problem of finding semantic roles to entities located in textual documents. SRL is a fruitful area of research containing work that takes advantage of multiple language resources, including FrameNet. The most recent and state-of-the-art approaches are mostly based on statistical methods, in particular, machine learning methods.

The model presented in [4] uses latent variables and semi-supervised learning to improve frame disambiguation for targets unseen at training time. On the other hand, the work shown in [32] consists of a frame identification that is coupled into an argument parsing method to perform FSP. Sling, [33], is a framework for frame-semantic parsing that performs neural-network parsing with bidirectional LSTM input encoding and a transition based recurrent unit. It takes as input only the tokens of the sentence, skipping any previous syntactic or semantic parser. Both methods are machine-learning based.

The semantic parser developed in [13] connects VerbNet and FrameNet by mapping the FrameNet frames to the VerbNet Intersective Levin classes. To further increase the verb coverage, they use the lexicon contained in PropBank and the PropBank semantic annotations to evaluate their system.

6. CONCLUSION

Semantic Role Labeling (SRL) is an essential task towards creating a machine-meaningful representation of textual information. FrameNet is the main supportive resource for this task. However, as a manually-built resource, it is error-prone and incomplete. A large group of frames lacks useful annotations. In this work, we present a data augmentation method for FrameNet documents that increases by over 13% the total number of annotations. As a result, a new dataset is now available for SRL and frame semantic parsing in general. We also show that the annotations generated can improve the performance of a semantic-role-labeling method.

The augmentation methods present in the literature are usually methods for combining FrameNet with other linguistic resources. This work presents an approach to augment the data available in FrameNet using sentence examples in the resource's element descriptions themselves. This way, one can apply our method after (or before) applying some other method present in the literature for a more incisive expansion without necessarily adding redundant information.

A first line of future research is to investigate the impact of this data augmentation in combination with other methods present in the literature. Another possible investigation venture is the exploration of the inter-frame relationships. We suspect that it is possible to further explore the connections amongst frames to infer new relationships amongst frame elements. We also intend to test the method on other electronic (linguistic) resources. For example, WordNet seems a relatively close opportunity for short- to mid-term research.

Semantic Role Labeling (SRL) is an essential task towards creating a machine-meaningful representation of textual information. FrameNet is the primary supportive resource for this task. However, as a manually-built resource, it is error-prone and incomplete. A large group of frames lacks useful annotations. In this work, we present a data augmentation method for FrameNet documents that increases by over 13% the total number of annotations. As a result, a new dataset is now available for SRL and frame semantic parsing in general. We also show that the annotations generated can improve the performance of a semantic-role-labeling method.

The augmentation methods present in the literature are usually methods for combining FrameNet with other linguistic resources. This work presents an approach to augment the data available in FrameNet using sentence examples in the resource's element descriptions themselves. This way, one can apply our method after (or before) applying some other method present in the literature for a more incisive expansion without necessarily adding redundant information.

The first line of future research is to investigate the impact of this data augmentation in combination with other methods present in the literature. Another possible investigation venture is the exploration of inter-frame relationships. We suspect that it is possible to explore the connections amongst frames further to infer new relationships amongst frame elements. We also intend to test the method on other electronic (linguistic) resources. For example, WordNet seems a relatively close opportunity for short- to mid-term research.

REFERENCES

- [1] O. Etzioni, M. Banko, M. J. Cafarella, Machine reading, in: Proceedings, The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, AAAI Press, 2006, pp. 1517–1519.
- [2] O. Abend, A. Rappoport, The State of the Art in Semantic Representations, *Acl* 35 (2017) 23–24.
- [3] C. F. Baker, C. J. Fillmore, J. B. Lowe, The Berkeley FrameNet Project, in: Proceedings of the 17th International Conference on Computational Linguistics - Volume 1, COLING '98, Association for Computational Linguistics, Stroudsburg, PA, USA, 1998, pp. 86–90. URL: <https://doi.org/10.3115/980451.980860>. doi:10.3115/980451.980860.
- [4] D. Das, D. Chen, A. F. T. Martins, N. Schneider, N. Noah A. Smith, Frame-Semantic Parsing, *Computational linguistics* 40 (2014) 9–56.
- [5] J. Bos, Wide-coverage semantic analysis with Boxer, in: Proceedings of the 2008 Conference on Semantics in Text Processing, c, Association for Computational Linguistics, Venice, Italy, 2008, pp. 277–286. doi:10.3115/1626481.1626503.
- [6] S. Swayamdipta, S. Thomson, C. Dyer, N. A. Smith, Frame-Semantic Parsing with Softmax-Margin Segmental RNNs and a Syntactic Scaffold, *arXiv preprint arXiv:1706.09528* (2017).
- [7] L. Kong, C. Dyer, N. A. Smith, Segmental Recurrent Neural Networks, *arXiv preprint arXiv:1511.06018* (2015) 1–10. URL: <http://arxiv.org/abs/1511.06018>. doi:10.21437/Interspeech.2016-40.
- [8] D. Das, N. Schneider, D. Chen, N. A. Smith, Probabilistic Frame-Semantic Parsing, *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL)* 3 (2010) 948–956.
- [9] C. F. Baker, C. J. Fillmore, B. Cronin, The Structure of the FrameNet Database, *International Journal of Lexicography* 16 (2003) 281—296.
- [10] J. Pennington, R. Socher, C. Manning, Glove: Global Vectors for Word Representation, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014) 1532–1543. URL: <http://aclweb.org/anthology/D14-1162>. doi:10.3115/v1/D14-1162.
- [11] D. P. Kingma, J. Ba, Adam: A Method for Stochastic Optimization (2014). URL: <http://arxiv.org/abs/1412.6980>.
- [12] L. Shi, R. Mihalcea, Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing, *Computational Linguistics and Intelligent Text Processing* 34 (2005) 100–111. URL: http://link.springer.com/10.1007/978-3-540-30586-6_9. doi:10.1007/978-3-540-30586-6_9.
- [13] A.-M. Giuglea, A. Moschitti, Semantic Role Labeling via FrameNet, VerbNet and PropBank, in: *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, July, Association for Computational Linguistics, Sydney, Australia, 2006, pp. 929–936. doi:10.3115/1220175.1220292.
- [14] M. Palmer, SemLink-Linking PropBank, VerbNet, FrameNet, Technical Report, 2009. URL: http://www.flarenets.eu/sites/default/files/S3_01_Palmer.pdf.
- [15] E. Laparra, G. Rigau, Integrating WordNet and FrameNet using a Knowledge-based Word Sense Disambiguation Algorithm, *Proceedings of the International Conference RANLP-2009* (2009) 208–213. URL: <http://www.aclweb.org/anthology/R09-1039>.

- [16] S. Tonelli, C. Giuliano, K. Tymoshenko, Wikipedia-based WSD for multilingual frame annotation, *Artificial Intelligence* 194 (2013) 203–221. URL: <http://dx.doi.org/10.1016/j.artint.2012.06.002>. doi:10.1016/j.artint.2012.06.002.
- [17] R. Green, B. J. Dorr, P. Resnik, Inducing frame semantic verb classes from WordNet and LDOCE, *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics - ACL'04* (2004) 375–es. URL: <http://portal.acm.org/citation.cfm?doid=1218955.1219003>. doi:10.3115/1218955.1219003.
- [18] P. Kingsbury, M. Palmer, From Treebank to PropBank, *LREC* (2002) 1989–1993. doi:10.1007/s13398-014-0173-7.2.
- [19] K. Kipper, A. Korhonen, N. Ryant, M. Palmer, A large-scale classification of English verbs, *Language Resources and Evaluation* 42 (2008) 21–40. doi:10.1007/s10579-007-9048-2.
- [20] C. F. Baker, C. Fellbaum, Wordnet and framenet as complementary resources for annotation, in: *Proceedings of the Third Linguistic Annotation Workshop*, Association for Computational Linguistics, 2009, pp. 125–129.
- [21] M. Lopez De Lacalle, E. Laparra, I. Aldabe, G. Rigau, Predicate Matrix: automatically extending the semantic interoperability between predicate resources, *Language Resources and Evaluation* 50 (2016) 263–289. URL: <http://adimen.si.ehu.es/web/PredicateMatrix>. doi:10.1007/s10579-016-9348-5.
- [22] E. Pavlick, T. Wolfe, P. Rastogi, C. Callison-Burch, M. Dredze, B. Van Durme, FrameNet+: Fast paraphrastic tripling of framenet, *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference* 2 (2015) 408–413.
- [23] J. Ganitkevitch, B. V. Durme, C. Callison-Burch, PPDB: The Paraphrase Database, in: *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, Atlanta, Georgia, 2013, pp. 758–764. URL: <https://aclanthology.info/papers/N13-1092/n13-1092>.
- [24] H. Mousselly Sergieh, I. Gurevych, Enriching Wikidata with Frame Semantics, in: *Proceedings of the 5th Workshop on Automated Knowledge Base Construction*, 3, Association for Computational Linguistics, San Diego, CA, 2016, pp. 29–34. URL: <http://aclweb.org/anthology/W16-1306>. doi:10.18653/v1/W16-1306.
- [25] D. Vrandecic, M. Krotzsch, Wikidata: A Free Collaborative Knowledgebase, *Commun. ACM* 57 (2014) 78–85. doi:10.1145/2629489.
- [26] S. Reddy, O. Tackström, M. Collins, T. Kwiatkowski, D. Das, M. Steedman, M. Lapata, Transforming Dependency Structures to Logical Forms for Semantic Parsing, *Transactions of the ACL* 4 (2016) 127–140.
- [27] A. Singhal, Introducing the Knowledge Graph: things, not strings, 2012. URL: <http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>.
- [28] I. Beltagy, C. Chau, G. Boleda, D. Garrette, K. Erk, R. J. Mooney, Montague meets markov: Deep semantics with probabilistic logical form, in: *Proceedings of the Second Joint Conference on Lexical and Computational Semantics, *SEM 2013*, ACL, 2013, pp. 11–21.
- [29] M. Richardson, P. Domingos, M. Richardson, P. Domingos, Markov logic networks, *Machine Learning* 62 (2006) 107–136. doi:10.1007/s10994-006-5833-1.
- [30] I. Beltagy, S. Roller, P. Cheng, K. Erk, R. J. Mooney, Representing meaning with a combination of logical and distributional models, *Computational Linguistics* 42 (2016) 763–808.
- [31] T. Khot, N. Balasubramanian, E. Gribkoff, A. Sabharwal, P. Clark, O. Etzioni, Exploring markov logic networks for question answering, in: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015*, Lisbon, Portugal, September 17-21, 2015, ACL, 2015, pp. 685–694.
- [32] K. M. Hermann, D. Das, J. Weston, K. Ganchev, Semantic Frame Identification with Distributed Word Representations, *Proceedings of ACL* (2014) 1448–1458. URL: <http://www.aclweb.org/anthology/P14-1136>. doi:10.3115/v1/P14-1136.
- [33] M. Ringgaard, R. Gupta, F. C. Pereira, Sling: A framework for frame semantic parsing, *arXiv preprint arXiv:1710.07032* (2017).