Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 4183–4189
Marseille, 11–16 May 2020

WikiBank: Using Wikidata to Improve Multilingual Frame-Semantic Parsing

Cezar Sas[1], Meriem Beloucif[2], Anders Søgaard[1]
[1]Department of Computer Science, University of Copenhagen
[2]Department of Informatics, Universität Hamburg
cezar.sas@gmail.com, beloucif@informatik.uni-hamburg.de, soegaard@di.ku.dk

Abstract

Frame-semantic annotations exist for a tiny fraction of the world's languages, Wikidata, however, links knowledge base triples to texts in many languages, providing a common, distant supervision signal for semantic parsers. We present WIKIBANK, a multilingual resource of partial semantic structures that can be used to extend pre-existing resources rather than creating new man-made resources from scratch. We also integrate this form of supervision into an off-the-shelf frame-semantic parser and allow cross-lingual transfer. Using Google's SLING architecture, we show significant improvements on the English and Spanish CoNLL 2009 datasets, whether training on the full available datasets or small subsamples thereof.

Keywords: cross-lingual frame semantic parsing, multilinguality, data augmentation

1. Introduction

Shallow semantic parsing comes in many varieties, including frame-semantic parsing (Täckström et al., 2015; Ringgaard et al., 2017), semantic role labeling (SRL) (Surdeanu et al., 2008; Pradhan and Xue, 2009; Weischedel et al., 2013; Zhang et al., 2019), and semantic dependency parsing (Oepen et al., 2014). In this paper, we present WIKIBANK, a multilingual resource with partial semantic dependency structures projected from an existing knowledge base. WIKIBANK is created automatically, and used to augment pre-existing resources, or reduce the annotation effort for low-resource languages. We show how it can be used to improve an off-the-shelf frame-semantic parser, but this resource could also be equally useful for semantic role labeling systems and other semantic parsing frameworks. The frames in frame-semantic parsing present extended predicate-argument structures that determine "who did what to whom", "when", "where" and "why". For instance, in a sentence such as *John gives Mary a computer in the morning*, a frame-semantic parser may identify *John*, *Mary*, and *computer* as core arguments of the "giving" frame, and *in the morning* as a temporal, optional argument. This kind of information has been shown to improve multiple downstream tasks such as information extraction (Bastianelli et al., 2013), machine translation (Knight and Luk, 1994; Ueffing et al., 2007; Wu and Fung, 2009; Shi et al., 2016; Beloucif and Wu, 2018), discourse parsing (Mihaylov and Frank, 2016), question answering (Surdeanu et al., 2003; Moschitti et al., 2003; Shen and Lapata, 2007; Berant and Liang, 2014) and classifying reason and stance in online debates (Hasan and Ng, 2014). Frame-semantic parsers, however, are normally trained on manually annotated resources such as the FrameNet corpus (Baker et al., 1998) or the OntoNotes corpus (Pradhan and Xue, 2009; Weischedel et al., 2013). However, such annotations only exist for a small subset of the world's languages. Some recent work (Aminian et al., 2019) solves this problem using annotation projection. Moreover, in contrast with previous methods, the authors have no supervision from lemmas, part-of-speech (POS) tags or syntactic parse trees. However, they still require parallel corpora. This paper, in contrast, presents WIKIBANK, a multilingual resource of sentences partially annotated with semantic structures extracted from Wikidata, a knowledge base that contains decontextualized predicate-argument relations for hundreds of languages. We directly map relations from this knowledge base onto Wikipedia sentences to obtain partial semantic structures (see example in Table 1). We use WIKIBANK with pre-existing resources to train an off-the-shelf frame-semantic parser for (simulated) low-resource languages.

Language	Sentences
EN	**Angola** *borders* Namibia to the south, Zambia to the east, the **Democratic Republic of the Congo** to the north-east.
ES	**Angola**, es un país ubicado al sur de África que *tiene fronteras* con la **República Democrática del Congo**.
DE	**Angola** *grenzt* an Namibia, Sambia, die Republik Kongo, die **Demokratische Republik Kongo** und den Atlantischen Ozean.

Table 1: An example of a Wikipedia sentence in 3 different languages; in Wikidata, these are represented by **shares_border_with(Q916, Q971)**, where Q916 is the Wikidata-instance *Angola* and Q971 is the Wikidata-instance *Democratic Republic of Congo*. In WIKIBANK, *Angola* is labeled with *ARG0*, while *Democratic Republic of Congo* with *ARG1*.

2. Contributions

The main contribution of our work is WIKIBANK, a novel multilingual resource for semantic parsing obtained by aligning Wikidata knowledge base triples with Wikipedia sentences. This resource can be both used to augment current semantic parsing datasets, and also reduce the required annotations for new languages. Furthermore, we present experiments demonstrating how to use WIKIBANK to improve frame-semantic parsers for different languages. Using techniques from multi-task learning, we show how to train an off-the-shelf frame-semantic parser (Ringgaard et

Language	Sentences and annotations
EN	**Captain Nemo** [*ARG0*] is a fictional character **created** [*V*] by the French science fiction author **Jules Verne** [*ARG1*] (1828-1905).
ES	**Juan Antonio García Casquero** [*ARG0*] **nació** [*V*] en **Madrid** [*ARG1*] en 1961.
DE	Der **Satz von Weyl** [*ARG1*], **benannt** [*V*] nach **Hermann Weyl** [*ARG1*], ist ein wichtiger Satz aus der Theorie der Lie-Algebren.

Table 2: Examples for the 3 languages in WIKIBANK. The EN example represents the Wikidata relation **creator(Captain Nemo, Jules Verne)**, the entity is **Captain Nemo**, and the value is **Jules Verne**, the annotation are in square brackets.

al., 2017) on a mixture of labeled data and partial annotations automatically extracted from Wikidata, obtaining significant improvements across two metrics for three languages and different subsamples of the training data, simulating a low-resource training setup. Finally, we show how WIKIBANK could act as a bridge for cross-lingual transfer, training semantic parsers on partial structures projected from Wikidata to sentences in multiple languages, as well as a combination of annotated data from those languages.

3. WikiBank

WIKIBANK is our new partially annotated resource for the multilingual frame-semantic parsing task. It is based on a heuristics-driven extraction of mark-up from knowledge bases that has important similarities to linguistic structures, and which can therefore serve as auxiliary task data, enabling the leverage of potential synergies. WIKIBANK is composed of partially annotated sentences. Examples in WIKIBANK consists of semantically labelled sentences, where each sentence has partial-semantic annotation for the predicate and its semantic arguments (see example in Table 2). WIKIBANK is derived directly from Wikidata (Vrandečić and Krötzsch, 2014), and Wikipedia. Wikidata is a collaborative knowledge base, containing triples `(entity_id, property_id, value_id)` that define a type of relation holding between an entity and a value (which can also be an entity). Wikidata also contains labels, and aliases for the properties, entities, and values.

Following (Hewlett et al., 2016), for each Wikidata item, we replace the IDs in each statement with the text label for properties and values that are entities, and with a human readable version for numeric values (e.g., time-stamp is converted into readable date), obtaining triples `(entity, property, value)`. The extraction of sentences from Wikipedia is performed using distant supervision, as in (Levy et al., 2017). For each triple, we take the corresponding Wikipedia article for the entity then we extract the first sentence containing the entity, the property, and the value. To increase the amount of examples, we also use Wikidata aliases when matching sentences with triples. The matching is done using regular expressions; a sentence is extracted if the entity, property, and value or their aliases are contained in it. Moreover, we filter out the sentences where the alias of the property found in the sentence does not contain a verb. The last step is to automatically annotate the extracted sentences with frame-semantic labels. We assign the label *ARG0* to the entity, *V* for the target in the property, and *ARG1* to the value. If a sentence contains multiple verbs, we consider each verb annotation as a different example. Table 3 shows the number of sentences and examples extracted for each language.

Language	Sentences	Examples
DE	19K	24K
EN	795K	990K
ES	32K	38K

Table 3: The sizes of the sentences and examples contained in WIKIBANK for each languages. The full WIKIBANK is available at `https://github.com/SasCezar/WikiBank`

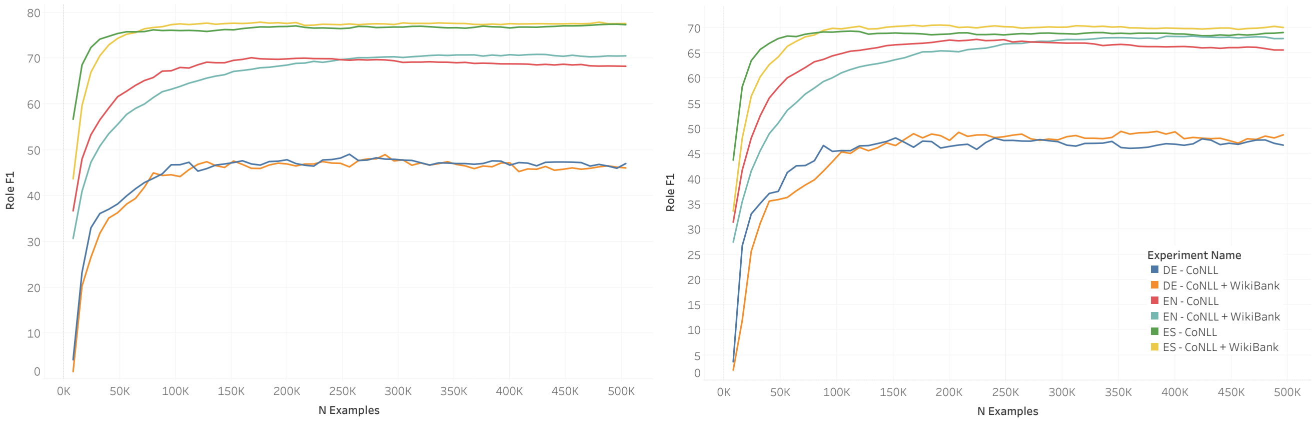
We used a pre-existing semantic role labeling model, DeepSRL (He et al., 2017), pretrained on CoNLL 2005, to see how well examples in WIKIBANK are annotated. For the evaluation we used a sample of 10K English sentences, and obtained the results shown in Table 4. We consider an argument label to be correct if the entity label from Wikidata is contained in the sequence extracted by DeepSRL. From Table 4, we note that the class *ARG0* has a lower accuracy compared to *ARG1*. After further data analysis, we realized that this is due to the *ARG0* containing names of locations, dates, and in general entities that are labeled as *ARG1* by the DeepSRL model.

Label	Correct	Total	Percent
ARG0	3,339	9,988	0.33
ARG1	14,288	14,555	0.98

Table 4: Number of correctly labeled arguments in WIKIBANK for 10K samples in English using a pretrained SRL model.

3.1. Baseline

SLING (Ringgaard et al., 2017) conceptualizes each frame as a list of slots. Each slot has a semantic role and a value. Each frame is categorized by one verb, for which the most likely meaning will be annotated. SLING is a transition-based parsing framework, not tied to any particular linguistic theory or knowledge ontology. Briefly put, SLING encodes the input text tokens using bidirectional Long Short Term Memories (LSTM). The encoding is then fed into a Transition Based Recurrent Unit (TBRU) in order to produce a sequence of transitions. The TBRU architecture is a single feed-forward unit, which takes the activations from the bidirectional-LSTM and combines them with the activations from the hidden layer from the previous step. The model then combines the transition system and the activation layer to create the input feature vector for the next step. The TBRU has multiple inputs such as the activations from both the left-to-right and right-to-left LSTMs, hidden layer activations of the transition steps, in order to have a continuous representation of the semantic context. SLING also in-



(a) *Simplified* label set (b) *Full* label set

Figure 1: Comparing the Role_F1 between the *simplified* label set and the *full* label set experiments.

corporates an attention mechanism based on neuro-science models of attention and awareness in (Nelson et al., 2017) and (Graziano, 2013). Specifically, the attention mechanism focuses on encoding the frame representation that the parser has created rather than encoding the tokens themselves.

3.2. Data and preprocessing

Our experiments are designed to provide a proof of concept for the usefulness of integrating WIKIBANK into training frame-semantic parsers. To that end, we train SLING with the full CoNLL 2009 German, English, and Spanish corpora (Hajič et al., 2009). The monolingual models trained only on the CoNLL corpora provide our baselines. We use the standard data splits: training, validation, and held-out test data; but we also purposely experiment with subsampling the training data to simulate low-resource scenarios. Using the CoNLL corpora rather than OntoNotes, as (Ringgaard et al., 2017) do, enables us to explore multilingual sharing and using WIKIBANK as a bridge for cross-lingual transfer. We facilitate cross-lingual transfer by using multilingual word embeddings (Lample et al., 2018) and, following (Johnson et al., 2017), by prepending all sentences with a language ID – as well as a task ID (CoNLL or WIKIBANK) for integrating the WIKIBANK sentences.
Furthermore, we experiment with reducing the *full* label sets in the corpora to a common, *simplified* label set, to facilitate cross-lingual transfer. The *simplified* labels disregard thematic role information. For both setups, we normalize labels to be conform with the PropBank (Palmer et al., 2005) notation (e.g., *A1* becomes *ARG1*). However, as shown in Figure 1, the experiments with the *full* label set have a slightly better accuracy than the ones with a *simplified* label set, so we will present only the results for the former.

3.3. Protocol

We experiment with three languages, and subsamples of the training data of size 0 (zero-shot), 100, 500, 1000, and 2000, as well as the full training set, and also with and without cross-lingual transfer. The cross-lingual training setup consists of various combination of source languages, with both CoNLL and WIKIBANK, and the target language WIKIBANK. The transfer is achieved using multilingual embeddings, the language as well as the previously-mentioned task IDs. In our experiments on subsamples of the training data, we only use 5,000 examples from WIKIBANK, not to swamp the supervision signal from the CoNLL 2009 corpora. The hyperparameters were set after performing a grid search on the OntoNotes (Pradhan and Xue, 2009) development set. We have $learning\_rate = 0.0005$, Adam (Kingma and Ba, 2015) as optimizer with $\beta_1 = 0.01, \beta_2 = 0.999$, $\epsilon = 1e-5$, no dropout, gradient clipping at 1.0, exponential moving average, no layer normalization, and training batch size of 8.
The goal of our experiments is to show that using easily accessible resources such as Wikidata helps improve the frame-semantics task. For that, we purposely fix the number of examples from CoNLL and WIKIBANK to exactly 5K, while varying the size of the target language CoNLL data. For instance, when training on ES, and having both EN and DE as source languages, the amount of data from WIKIBANK for ES, DE, and EN is 5K each, from CoNLL both EN, and DE have 5K examples, while for ES there would be one of 0, 100, 500, 1000, or 2000 examples. Additionally, we also experiment with the entire CoNLL data. The baselines are without the use of the auxiliary data and full CoNLL.

Training In our experiments we explore multiple strategies of training using different languages as source language. This cross-lingual setup is achieved using multilingual word representation that embeds words from all languages into a single semantic space so that words with similar meanings are close to each other despite of language. There are also specific tasks for CoNLL and WIKIBANK. We use the same hyperparameters described by (Ringgaard et al., 2017). For the embeddings we use MUSE multilingual embeddings (Lample et al., 2018). The number of steps is set to 100K.

3.4. Evaluation

We evaluate the quality of the model using Slot_F1 and Role_F1, following (Ringgaard et al., 2017). SLING compares the produced frames with the gold standard frames

Target	Source	Target CoNLL Size 0	100	500	1000	2000	ALL
DE	-	-	11.27	19.99	**29.82**	**34.33**	60.84
	DE W	**9.94**	11.25	**20.23**	27.28	33.56	**61.85**
	EN	7.28	11.90	18.08	26.76	28.03	-
	ES	6.45	**13.38**	20.10	28.45	33.02	-
	EN-ES	7.59	10.57	19.25	26.71	33.02	-
EN	-	-	44.15	52.36	56.91	61.83	80.48
	EN W	7.87	42.22	54.22	58.86	62.86	**80.72**
	DE	7.89	43.83	54.16	**60.41**	**63.20**	-
	ES	**18.59**	**43.91**	**56.08**	59.17	62.76	-
	DE-ES	15.42	43.41	46.47	41.49	46.96	-
ES	-	-	42.49	62.67	66.38	71.47	83.82
	ES W	8.74	45.57	**64.66**	68.55	**72.21**	**84.57**
	DE	1.20	**50.71**	63.90	**68.72**	71.56	-
	EN	**13.49**	44.57	60.18	66.14	69.98	-
	DE-EN	4.09	43.25	59.55	65.66	68.64	-

Table 5: Slot_F1 scores on the development data. We use these scores to find the best model, which we use to report the scores on the test data. Additionally, we only run all data with the same source language in order to contrast it to the full CoNLL data as a baseline.

from the evaluation corpus. The documents are matched using a graph where the document is the root. The document is connected to the spans, which are connected to the frames that they evoke. The graph is expanded using the span-to-span links defined by the roles. The graph of the produced output, and the gold standard one are aligned to produce the quality measures. The measures evaluate the performance for spans, frames, frame types, and roles that link to other frames ("roles"). They also define aggregated measures, the one used to define the best model is called *slot*, and is a combination of type and role. Note these metrics are not comparable to the F1-scores reported in the CoNLL 2009 shared task. Slot_F1 is similar in the spirit to the CoNLL F1-score, but because of how frames are represented internally in SLING, there is no straight-forward mapping into the CoNLL format.

We start by conducting experiments on all possible target-language/target-size combinations on the development dataset, which we consider as a tuning step; results are reported in Table 5. For instance, we note that for German as a target language, when the target CoNLL size is 100, the Spanish dataset helps significantly improve the results on both development and test sets. Moreover, we note that all the models without source language perform much worse than when adding WIKIBANK independently of the language.

4. Results

As described above, we experiment with different sizes, combination of languages and corpora. Results for the CoNLL development set are presented in Table 5.We can notice how all three languages have an improvement in performance, except in 3 cases, when adding the extra data in the same language from WIKIBANK. Additionally, when adding another languages as source, with data from both CoNLL and WIKIBANK, the models achieves even better results. However, when using 2 languages as source, there is a negative impact on the performances. The results for the test data are presented in Table 6, where we report our baseline performance on all training set sizes, as well as the performance of the models trained on a mixture of the CoNLL 2009 corpora and WIKIBANK; finally, we also report on our best cross-lingual models, which were trained on the combination of German, English, and Spanish CoNLL 2009 and WIKIBANK data that fared best on the target language validation data (WIKIBANK$^+$). For example, for ES trained with 1000 target examples, the best model can be found in Table 5, for this example it is the one trained using DE as extra source language.

Target	Metric		Target CoNLL Size 0	100	500	1000	2000	All
DE	Slot_F1	WIKIBANK	**09.73**	11.26	**22.03**	26.73	34.77	**59.88**
		WIKIBANK$^+$	**09.73**	**11.67**	**22.03**	24.83	34.77	-
		Baseline	-	10.25	20.93	**29.17**	**35.40**	58.81
	Role_F1	WIKIBANK	**06.20**	**07.23**	**16.03**	18.15	23.46	**48.58**
		WIKIBANK$^+$	**06.20**	06.49	**16.03**	15.32	23.46	-
		Baseline	-	07.07	11.98	**18.44**	**23.64**	47.07
EN	Slot_F1	WIKIBANK	08.34	43.27	55.29	60.96	64.41	**81.78**
		WIKIBANK$^+$	**18.53**	**44.25**	**56.58**	**61.50**	**64.89**	-
		Baseline	-	44.06	53.12	58.08	63.75	81.31
	Role_F1	WIKIBANK	**01.51**	16.18	30.14	37.97	42.67	69.86
		WIKIBANK$^+$	00.37	**18.00**	**31.96**	**38.71**	**43.61**	-
		Baseline	-	15.53	26.99	33.99	42.08	**71.06**
ES	Slot_F1	WIKIBANK	08.67	46.41	**64.66**	68.46	**71.75**	**84.42**
		WIKIBANK$^+$	**13.52**	**50.85**	**64.66**	**68.52**	**71.75**	-
		Baseline	-	42.18	62.54	66.30	70.97	84.28
	Role_F1	WIKIBANK	00.00	20.62	**38.71**	**44.44**	**49.92**	**70.55**
		WIKIBANK$^+$	**00.74**	**21.84**	**38.71**	44.04	**49.92**	-
		Baseline	-	12.40	35.06	40.81	47.89	70.13

Table 6: Results for the test data. Slot_F1 and Role_F1 scores, w/o WIKIBANK and w/o cross-lingual transfer. WIKIBANK$^+$ is results with cross-lingual transfer. We do not conduct full data experiments on WIKIBANK$^+$.

We contrast the baseline to two WIKIBANK models: a monolingual WIKIBANK and a cross-lingual transfer model, WIKIBANK$^+$. We note from Table 6 that our joint WIKIBANK outperforms the baseline in most of these cases. When training on the full datasets, training on WIKIBANK always leads to a better performance, except when evaluating English on Role_F1. On smaller subsamples of the training data, the English and Spanish WIKIBANK parsers are consistently best. For German, we see some fluctuation, and training with WIKIBANK only outperforms the baseline model when the sample size is 500 or all data. We also note from Table 6 that, WIKIBANK$^+$, the cross-transfer model is always better than the monolingual model WIKIBANK, except for 4 out 15 times with Slot_F1. Finally, our results also indicate that WIKIBANK provides a bridge enabling cross-language transfer, leading to small, but relatively consistent improvements across the board over just training on a mixture of the CoNLL 2009 corpora and WIKIBANK.

5. Background

Frame semantics (Fillmore, 1982; Baker et al., 1998; Swayamdipta et al., 2017) is the study of how linguistic forms affect frame knowledge, and how these frames thus could be integrated into an understanding of sentences and documents (Baker et al., 1998). Frame-semantic parsers typically rely on supervised learning to train complex models on top of a syntactic parser, induced from manually annotated resources, such as FrameNet (Baker et al., 1998) or PropBank (Palmer et al., 2005). One of the earliest works on frame-semantic parsing is by (Gildea and Jurafsky, 2000), who proposed a discriminative model for semantic role labeling using frame-semantics. (Thompson et al., 2003) proposed a generative model trained on FrameNet for shallow semantic parsing and frame identifications tasks. (Shi and Mihalcea, 2004) proposed to identify frames and their elements using a rule-based approach. (Johansson and Nugues, 2007) proposed a frame-semantic structure extraction model based on syntactic dependency parsing; they also propose a method to strengthen FrameNet by automatically adding new units to its lexical database.

More recent work on frame-semantic parsing includes SEMAFOR (Das et al., 2014; Kshirsagar et al., 2015), a frame-semantic parser for identifying and labeling the semantic arguments of a given predicate that evokes a specific FrameNet frame. (Ringgaard et al., 2017) use a modified version of OntoNotes (Pradhan and Xue, 2009; Weischedel et al., 2013) in their experiments, and we use a similar modification of the CoNLL 2009 corpora. That said, our parsing experiments are merely meant to illustrate the usefulness of the WIKIBANK resource.

To the best of our knowledge, there are two similar works regarding the creation of automatic labeled resources for semantic role labeling. In the first one (Exner et al., 2015), the authors create a dataset using loosely parallel sentences from Wikipedia and transfer the predicate-argument structure from the source language (English) to the target language. The alignment is achieved using the Wikidata IDs, extracted using a dictionary and a named entity linker. In the second one (Hartmann et al., 2016), the approach is to use the distant supervision paradigm to transfer the labels from a Linked Lexical Resource, a combination of several resource like WordNet, FrameNet and Wikitionary, to a large unlabeled corpus (e.g. web pages). Both of these solution require additional resource which are not always available for low-resource languages. (Exner et al., 2015) dictionary creation requires a POS tagger, language-dependent rules, and entity databases; while (Hartmann et al., 2016) use two Linked Lexical Resource, Uby and SemLink, which support very few languages.

6. Conclusion

We introduced WIKIBANK, a new multilingual resource for semantic parsing. WIKIBANK consists of partial semantic structures directly projected from Wikidata onto Wikipedia sentences. We presented a set of experiments meant to illustrate the usefulness of this resource. Specifically, we showed that when training Google's SLING frame-semantic parser on the CoNLL 2009 corpora, we can obtain significant improvements using WIKIBANK as auxiliary data. The multi-task learning method used in our experiments also facilitates cross-lingual transfer, and our experiments indicate that WIKIBANK acts as a bridge between languages, enabling joint training on multiple annotated corpora in different languages.

7. Bibliographical References

Aminian, M., Rasooli, M. S., and Diab, M. T. (2019). Cross-lingual transfer of semantic roles: From raw text to semantic roles. In Simon Dobnik, et al., editors, *Proceedings of the 13th International Conference on Computational Semantics, IWCS 2019, Long Papers, Gothenburg, Sweden, May 23-27 May, 2019*, pages 200–210. Association for Computational Linguistics.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The berkeley framenet project. In *Proceedings of the 17th International Conference on Computational Linguistics - Volume 1*, COLING '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bastianelli, E., Castellucci, G., Croce, D., and Basili, R. (2013). Textual inference and meaning representation in human robot interaction. In *Proceedings of the Joint Symposium on Semantic Processing. Textual Inference and Structures in Corpora, JSSP 2013, Trento, Italy, November 20-22, 2013*, pages 65–69. Association for Computational Linguistics.

Beloucif, M. and Wu, D. (2018). SRL for low resource languages isn't needed for semantic SMT. In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*.

Berant, J. and Liang, P. (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, Maryland, June. Association for Computational Linguistics.

Das, D., Chen, D., Martins, A. F. T., Schneider, N., and Smith, N. A. (2014). Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56.

Exner, P., Klang, M., and Nugues, P. (2015). A distant supervision approach to semantic role labeling. In Martha Palmer, et al., editors, *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics, *SEM 2015, June 4-5, 2015, Denver, Colorado, USA.*, pages 239–248. The *SEM 2015 Organizing Committee.

Fillmore, C. J., (1982). *Frame semantics*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Gildea, D. and Jurafsky, D. (2000). Automatic labeling of semantic roles. In *38th Annual Meeting of the Association for Computational Linguistics, Hong Kong, China, October 1-8, 2000.* ACL.

Graziano, M. S. A. (2013). *Consciousness and the Social Brain.* Oxford University Press, USA.

Hajič, J., Ciaramita, M., Johansson, R., Kawahara, D., Martí, M. A., Màrquez, L., Meyers, A., Nivre, J., Padó, S., Štěpánek, J., Straňák, P., Surdeanu, M., Xue, N., and Zhang, Y. (2009). The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL '09, pages 1–18, Stroudsburg, PA, USA. Association for Computational Linguistics.

Hartmann, S., Eckle-Kohler, J., and Gurevych, I. (2016). Generating training data for semantic role labeling based on label transfer from linked lexical resources. *Transac-*

tions of the Association for Computational Linguistics, 4:197–213.

Hasan, K. S. and Ng, V. (2014). Why are you taking this stance? Identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762. Association for Computational Linguistics.

He, L., Lee, K., Lewis, M., and Zettlemoyer, L. (2017). Deep semantic role labeling: What works and what's next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada, July. Association for Computational Linguistics.

Hewlett, D., Lacoste, A., Jones, L., Polosukhin, I., Fandrianto, A., Han, J., Kelcey, M., and Berthelot, D. (2016). WikiReading: A Novel Large-scale Language Understanding Task over Wikipedia. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1535–1545. Association for Computational Linguistics.

Johansson, R. and Nugues, P. (2007). LTH: Semantic structure extraction using nonprojective dependency trees. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 227–230, Prague, Czech Republic, June. Association for Computational Linguistics.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., and Dean, J. (2017). Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In Yoshua Bengio et al., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Knight, K. and Luk, S. K. (1994). Building a large-scale knowledge base for machine translation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence (Vol. 1)*, AAAI '94, pages 773–778, Menlo Park, CA, USA. American Association for Artificial Intelligence.

Kshirsagar, M., Thomson, S., Schneider, N., Carbonell, J., Smith, N. A., and Dyer, C. (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, July. Association for Computational Linguistics.

Lample, G., Conneau, A., Ranzato, M., Denoyer, L., and Jégou, H. (2018). Word translation without parallel data. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.

Levy, O., Seo, M., Choi, E., and Zettlemoyer, L. (2017). Zero-Shot Relation Extraction via Reading Comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342. Association for Computational Linguistics.

Mihaylov, T. and Frank, A. (2016). Discourse relation sense classification using cross-argument semantic similarity based on word embeddings. In Nianwen Xue, et al., editors, *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning: Shared Task, CoNLL 2016, Berlin, Germany, August 7-12, 2016*, pages 100–107. ACL.

Moschitti, A., Morarescu, P., and Harabagiu, S. M. (2003). Open domain information extraction via automatic semantic labeling. In Ingrid Russell et al., editors, *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference, May 12-14, 2003, St. Augustine, Florida, USA*, pages 397–401. AAAI Press.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., Cash, S. S., Naccache, L., Hale, J. T., Pallier, C., and Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114(18):E3669–E3678.

Oepen, S., Kuhlmann, M., Miyao, Y., Zeman, D., Flickinger, D., Hajic, J., Ivanova, A., and Zhang, Y. (2014). Semeval 2014 task 8: Broad-coverage semantic dependency parsing. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 63–72, Dublin, Ireland, August.

Palmer, M., Kingsbury, P., and Gildea, D. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Pradhan, S. S. and Xue, N. (2009). Ontonotes: The 90% solution. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Tutorial Abstracts*, pages 11–12. Association for Computational Linguistics.

Ringgaard, M., Gupta, R., and Pereira, F. C. N. (2017). SLING: A framework for frame semantic parsing. *CoRR*, abs/1710.07032.

Shen, D. and Lapata, M. (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)*, page 12–21.

Shi, L. and Mihalcea, R. (2004). Open text semantic parsing using FrameNet and WordNet. In *HLT-NAACL 2004: Demonstration Papers*, pages 19–22, Boston, Massachusetts, USA, May 2 - May 7. Association for Computational Linguistics.

Shi, C., Liu, S., Ren, S., Feng, S., Li, M., Zhou, M., Sun, X., and Wang, H. (2016). Knowledge-based semantic embedding for machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2245–2254. Association for Computational Linguistics.

Surdeanu, M., Harabagiu, S., Williams, J., and Aarseth, P. (2003). Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 8–15, Stroudsburg, PA, USA. Association for Computational Linguistics.

Surdeanu, M., Johansson, R., Meyers, A., Màrquez, L., and Nivre, J. (2008). The CoNLL-2008 shared task on joint parsing of syntactic and semantic dependencies. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, CoNLL '08, pages 159–177, Stroudsburg, PA, USA. Association for Computational Linguistics.

Swayamdipta, S., Thomson, S., Dyer, C., and Smith, N. A. (2017). Frame-semantic parsing with softmax-margin segmental rnns and a syntactic scaffold. *CoRR*, abs/1706.09528.

Täckström, O., Ganchev, K., and Das, D. (2015). Efficient inference and structured learning for semantic role labeling. *Transactions of the Association for Computational Linguistics*, 3:29–41.

Thompson, C. A., Levy, R., and Manning, C. D. (2003). A generative model for semantic role labeling. In Nada Lavrac, et al., editors, *Machine Learning: ECML 2003, 14th European Conference on Machine Learning, Cavtat-Dubrovnik, Croatia, September 22-26, 2003, Proceedings*, volume 2837 of *Lecture Notes in Computer Science*, pages 397–408. Springer.

Ueffing, N., Haffari, G., and Sarkar, A. (2007). Transductive learning for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 25–32. Association for Computational Linguistics.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85, September.

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., and Houston, A. (2013). OntoNotes Release 5.0 LDC2013T19. *Linguistic Data Consortium*.

Wu, D. and Fung, P. (2009). Semantic roles for smt: A hybrid two-pass model. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, NAACL-Short '09, pages 13–16, Stroudsburg, PA, USA. Association for Computational Linguistics.

Zhang, Y., Wang, R., and Si, L. (2019). Syntax-enhanced self-attention-based semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 616–626, Hong Kong, China, November. Association for Computational Linguistics.