Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 1976–1981
Marseille, 11–16 May 2020

Cross-lingual Structure Transfer for Zero-resource Event Extraction

Di Lu[1]*, Ananya Subburathinam[1], Heng Ji[2], Jonathan May[3], Shih-Fu Chang[4], Avirup Sil[5], and Clare Voss[6]
[1]Rensselaer Polytechnic Institute,[2]University of Illinois at Urbana-Champaign,
[3]University of Southern California Information Sciences Institute, [4]Columbia University,
[5]IBM T.J. Watson Research Center, [6]Army Research Laboratory
[1]dylanlu.dlu@gmail.com [2]hengji@illinois.edu

Abstract

Most current cross-lingual transfer learning methods for Information Extraction (IE) have been applied to local sequence labeling tasks. To tackle more complex tasks such as event extraction, we need to transfer graph structures (event trigger linked to multiple arguments with various roles) across languages. We develop a novel share-and-transfer framework to reach this goal with three steps: (1) Convert each sentence in any language to language-universal graph structures; in this paper we explore two approaches based on universal dependency parses and fully-connected graphs, respectively. (2) Represent each node in these graph structures with a cross-lingual word embedding so that all sentences, regardless of language, can be represented within one shared semantic space. (3) Using this common semantic space, train event extractors on English training data and apply them to languages that do not have any event annotations. Experimental results on three languages (Spanish, Russian and Ukrainian) without any annotations show this framework achieves comparable performance to a state-of-the-art supervised model trained on more than 1,500 manually annotated event mentions.

Keywords: Information Extraction, Less-Resourced Languages, Multilinguality

1. Introduction

Event Extraction is an important task in Information Extraction (IE) that aims to identify event triggers and arguments from unstructured texts and classify them into predefined categories. Compared to other IE tasks such as name tagging, the annotations for Event Extraction are more costly because they are structured and require a rich label space; full event structure annotation consists of its trigger span and type label as well as each of its one or more argument spans and role labels. Publicly available annotations for event extraction exist for only a few languages, such as English, Spanish, Chinese, and Arabic (Doddington et al., 2004; Getman et al., 2018). We propose a novel share-and-transfer framework to project training data for English only and test data for zero-event-resource languages into one common semantic space, so that we can train an event extractor on English annotations and apply it to target languages.

Currently most successful cross-lingual transfer approaches for IE are limited to sequence labeling (Feng et al., 2018a; Xie et al., 2018; Zhang et al., 2018a; Lin et al., 2018). In contrast, event extraction requires transferring complex graph structures that contain triggers and arguments. For example, in Figure 1, the words *fire/ fired* combined with different arguments indicate different event types. A transfer approach to IE with a typical sequence-based Long Short Term Memory (LSTM) encoder will incorporate language-specific characteristics, such as word order, into word representations, reducing its effectiveness in transfer between two languages with quite different word orders.

In this paper, we explore cross-lingual event transfer learning in a zero-resource setting where there is no annotation available for the target language. We propose to transfer language-universal *structures* instead of *surface features*. Specifically, we adopt two language-universal structures to represent sentences: dependency trees and fully connected graphs. Then, we use encoder mechanisms to generate word representations in a latent space: Tree-LSTM for dependency trees and Transformer (Vaswani et al., 2017) for complete graphs. Finally, we treat event trigger labeling and argument role labeling as mappings from the latent space to the event type and argument role respectively.

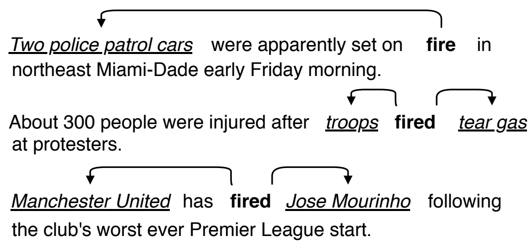

Figure 1: Disambiguating trigger ('fire/d' bolded) to obtain the correct type entails understanding argument structures.

By using English as the source language, experimental results on Spanish, Russian and Ukrainian show that the model based on the fully connected graph-Transformer encoder not only has better performance than the model based on the dependency tree-LSTM encoder, it also achieves performance comparable to a state-of-the-art supervised model trained on more than 1,500 manually annotated event mentions. In this work we make the following novel contributions:

- We show, for the first time, how cross-lingual event representations may be transferred between languages.
- We explore the use of structured representations for this transfer based on language-specific linguistic annotation and unsupervised structure discovery, and find a consistent benefit from the latter.

* Currently at Dataminr Inc.

2. Model

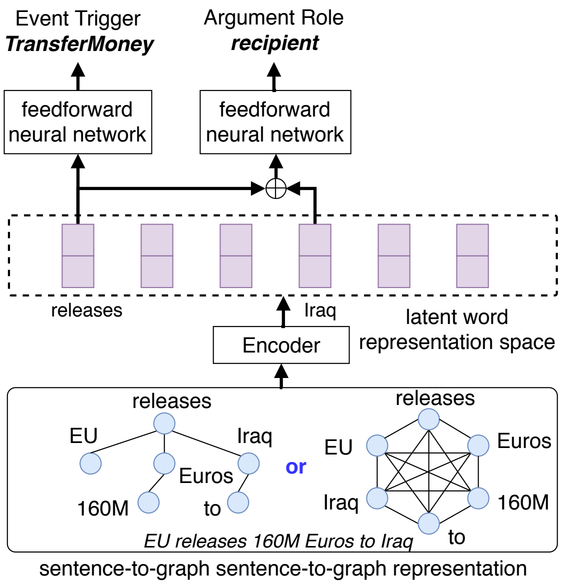

Figure 2: Overall Framework for Cross-lingual Event Structure Transfer.

In this section, we describe our approach to cross-lingual event structure transfer (Figure 2).

2.1. Encoder

We construct two language-universal graph structure representations for each input sentence: **universal dependency trees** and **fully connected graphs** for the edges, and cross-lingual word embeddings for the nodes (words). We train a BiAffine Dependency Parser (Dozat and Manning, 2016) for a particular language using the Universal Dependency treebanks (Nivre et al., 2016), and then apply the dependency parser to sentences to obtain universal dependency trees. For fully connected graphs, we regard each token in a sentence as a node in the graph and there's an edge between each pair of nodes. Then we apply **Tree-LSTM** encoder and **Transformer** encoder to generate word representations in the latent space, respectively.

Tree-LSTM Encoder. We exploit the Child-Sum Tree-LSTMs proposed by Tai et al. (2015). In contrast to the standard LSTM, here the memory cell updates of the Tree-LSTM unit are dependent on the states of all children units. The Tree-LSTM unit selectively incorporates information from each child.

Transformer Encoder. Our multi-layer bidirectional Transformer encoder is based on the architecture proposed in Vaswani et al. (2017), composed of a stack of N identical layers, where each layer has a multi-head self-attention sub-layer and a position-wise feed-forward sub-layer. Our implementation is identical to the original, except that here, crucially, we do not include positional encodings because word order is language-specific.

2.2. Event Trigger Labeling

Given an input sentence $\{w_1, w_2, ..., w_i\}$, where w_i is the i^{th} word, the encoder generates representations in the latent space for each word: $h_i = Encoder(w_i)$. We regard event trigger labeling as learning a mapping function from the latent space to event types. This model is composed of a linear layer (M_t) and a Softmax transformation σ. The objective function is as follows:

$$L_{trigger} = \sum_{i=1}^{N} y_i \ \log(\sigma(M_t \cdot h_i)) \qquad (1)$$

where y_i is a one-hot vector activated for w_i's correct label.

2.3. Event Argument Role Labeling

To predict the role of a candidate argument (w_j^{arg}) for an event trigger (w_i^{tri}), we first use the encoder to generate the trigger and argument representations in the latent space: $h_i^{tri} = Encoder(w_i^{tri}), h_j^{arg} = Encoder(w_j^{arg})$. The mapping function from the latent space to argument roles is composed of a concatenation operation ($[a_i; a_j]$), a linear layer (M_a) and a Softmax output layer. The objective function for argument role labeling is as follows:

$$L_{arg} = \sum_{i=1}^{N}\sum_{j=1}^{L_i} y_{ij} \ \log(\sigma(M_a \cdot [h_i^{tri}; h_j^{arg}])) \qquad (2)$$

Note that we do not fine-tune the parameters of the encoder during this step.

3. Experiments

3.1. Data

Language	Source	Train	Test
English	ERE	25,168	–
Spanish	ERE	–	5,164
Russian	AIDA	–	3,604
Ukrainian	AIDA	–	3,763

Table 1: Data statistics (Number of sentences).

We use the Entity, Relation and Event (ERE) (Getman et al., 2018) 2014-2016 English corpora for training the trigger labeling component, and only the 2016 English event corpus for training the argument labeling component because the 2014 and 2015 ERE corpora do not include annotations for arguments. We then test our approach to structure transfer on three target languages for which pre-existing, independently ground-truth annotations were available: Spanish, Russian and Ukrainian. The Spanish data is from ERE 2016 event corpus. The Russian and Ukrainian datasets are subsets[1] of the seedling corpus (LDC catalog: LDC2018E64) from the DARPA AIDA program[2], and they are annotated by a native speaker with the AIDA event ontology. We choose these languages because they are the only multilingual event datasets whose annotations can be mapped into the same schema. ACE (Walker et al., 2006) datasets include event annotations for English, Chinese and Arabic, but annotated following a different ontology. Table 1 shows the statistics of the datasets. Table 2 shows the distribution

[1] Refer to the Appendix for the list of document IDs

[2] https://www.darpa.mil/program/active-interpretation-of-disparate-alternatives

Event type / Language	Justice	Contact	Conflict	Transaction	Life	Movement	Personnel	Business	Manufacture
English	4,134	3,542	2,613	2,411	1,956	1,688	1,340	279	158
Spanish	65	824	346	301	207	275	146	0	58
Russian	114	549	539	53	116	651	58	14	9
Ukrainian	40	869	453	52	86	225	43	8	5

Table 2: Distribution of event types in various datasets (Number of event mentions). The statistics for English are from the training split, and the statistics for Spanish, Russian and Ukrainian are from testing splits.

of event types for each language. We follow the criteria in previous work (Ji and Grishman, 2008; Li et al., 2013) for evaluation.

3.2. Training Details

Treebanks. We use the Version 2.3 treebanks released by Universal Dependencies [3] to train the dependency parsers.
Tokenization. We use Spacy tokenization (Honnibal and Montani, 2017) for English and Spanish and the NLTK toktok tokenizer (Dehdari, 2014) for Russian and Ukrainian.
Word Embedding. We use multilingual word embeddings released by Facebook Research (Lample et al., 2017) [4]. The algorithm aligns word embeddings of various languages, which are pre-trained from Wikipedia articles (Joulin et al., 2016) [5], in a single vector space. It learns a mapping from the source space to the target space using Procrustes alignment by bilingual dictionaries. We do not fine-tune them during training.
Optimization. We use Adam optimizer (Kingma and Ba, 2014), which is based on adaptive estimates of lower-order moments, with warmup of 500, factor of 2.
Table 3 shows the hyperparameters to train the Tree-LSTM encoder and Table 4 shows the hyperparameters to train the Transformer encoder.

Hyperparameter	Value
word embedding size	300
hidden dimension size	150
dropout	0.5
learning rate	0.001
learning rate decay	0.001
batch size	1

Table 3: Hyperparameters of Tree-LSTM encoder.

3.3. Results

We compare several representations, each of which is used to train event recognition models as described in Sections 2.2 and 2.3. Each model is trained on annotated English examples but evaluated in Spanish, Russian, or Ukrainian, using no event annotation in these languages. The representations compared are:

[3]https://universaldependencies.org/
[4]https://github.com/facebookresearch/MUSE
[5]https://github.com/facebookresearch/fastText

Hyperparameter	Value
word embedding size	300
hidden dimension size	768
filter size	768
number of head	12
number of layer	12
dropout	0.2
learning rate	0.003
batch size	16

Table 4: Hyperparameters of Transformer encoder.

(1) *Bi-LSTM*, a baseline, using Bi-LSTM as the encoder to generate word representations based on flattened input sequences.

(2) *Tree-LSTM,* using Child-Sum Tree-LSTM as the encoder to generate word representations based on dependency trees.

(3) *Transformer,* using Transformer as the encoder to generate word representations based on fully-connected graph.

Table 5 shows the overall performance. We can see that both structure transfer methods significantly outperform the Bi-LSTM baseline, and the Transformer-based encoder generally outperforms Tree-LSTM, especially on trigger labeling, because universal dependency parsers are imperfect, with accuracy 57.2%-95.3% (Nivre et al., 2018). When the target language and source language are closer, such as Spanish and English, the gap between Bi-LSTM baseline and proposed models is smaller. For argument labeling, the Tree-LSTM-based encoder actually has better performance, which results in reduced gap with Transformer-based model. It is because that structural information is important for argument labeling and Tree-LSTM-based model benefits more from the explicit information derived from universal dependency parsing.

3.4. Comparison with Supervised Models

We also compare our approach with supervised event extractors trained from manual annotations. In the supervised setting, we use an LSTM-based sequence labeling model. Figure 3 shows the learning curves of these supervised models. We can see that without using any annotations for the target language, our best model achieves comparable performance with the supervised models trained on about 1,500 manually annotated event mentions.

Dataset	Model	Trigger Labeling (%) P	R	F1	Argument Labeling (%) P	R	F1
ERE Spanish	Bi-LSTM	45.16	25.83	32.87	**39.51**	10.70	16.84
	Tree-LSTM	**53.56**	28.44	37.15	32.12	11.89	**17.35**
	Transformer	44.75	**39.15**	**41.77**	23.09	**13.55**	17.08
AIDA Russian	Bi-LSTM	**60.64**	13.55	22.15	**36.65**	6.24	10.57
	Tree-LSTM	59.44	20.21	30.16	28.95	7.81	12.30
	Transformer	40.02	**50.00**	**44.46**	16.26	**12.09**	**13.87**
AIDA Ukrainian	Bi-LSTM	**86.81**	8.87	16.10	**41.69**	6.42	11.12
	Tree-LSTM	77.74	14.71	24.74	30.09	8.89	13.72
	Transformer	44.61	**39.34**	**41.81**	14.85	**12.87**	**13.79**

Table 5: Comparison on Various Representations for Cross-lingual Event Transfer Learning.

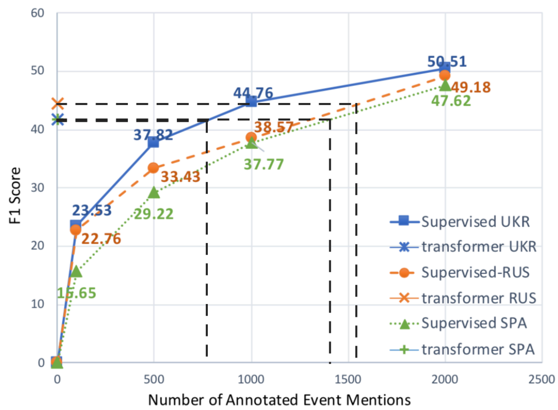

Figure 3: Comparison with Learning Curves for Supervised Bi-LSTM models on trigger labeling.

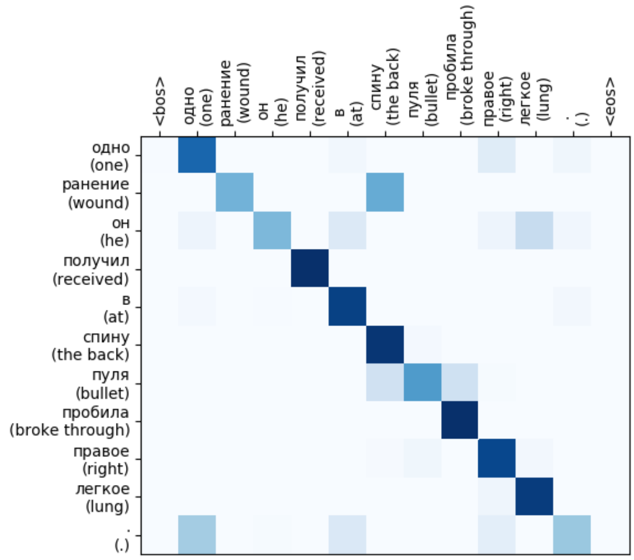

Figure 4: Visualization of Attention Weights for the First head of the Multi-head Attention Sublayer.

3.5. Visualization of Attentions

Figure 4 shows an example of visualized attention weights from the first head of the multi-head attention layer. In the Russian sentence *"одно ранение он получил в спину пуля пробила правое"* (he received one wound in the back, a bullet pierced his right lung), "ранение" (wound) is a trigger for an Life.Injure event. From the visualized attention weights, we can clearly see when the model generates the representation for the word *"ранение (wound)"*, *"спину (back)"* also contributes besides the word *"ранен"* itself. And *"спину (back)"* is an argument of the event mention triggered by *"ранение (wound)"* here. It indicates that the model successfully transfers structural information from the source language to the target language.

4. Related Work

A large number of supervised machine learning techniques have been used for event extraction, including traditional techniques based on symbolic features (Ji and Grishman, 2008; Liao and Grishman, 2011), joint inference models (Li et al., 2014; Yang and Mitchell, 2016), and recently with neural networks (Nguyen and Grishman, 2015; Nguyen et al., 2016; Chen et al., 2015; Nguyen and Grishman, 2018; Liu et al., 2018b). These approaches incorporate language-specific information, and thus require a substantial amount of annotations when adapted to a new language.
Traditional multilingual approaches (Li et al., 2012; Wei et al., 2017) to event extraction were all based on feature engineering. Recently, (Agerri et al., 2016; Danilova et al., 2014; Feng et al., 2018b) demonstrate methods for building multilingual event extraction systems. Hsi et al. (2016) have used language-independent features for event extraction for low-resource languages. (Lu and Nguyen, 2018) show that word sense disambiguation helps event detection via neural representation matching. (Liu et al., 2018a; Zhang et al., 2018b) propose event extraction by attention mechanism, e.g. the former use a gated multi-lingual attention technique. To the best of our knowledge, this is the first work to design a cross-lingual structure transfer framework to enable event extraction for a language without any event training data.

5. Conclusions and Future Work

In this paper, we propose a novel cross-lingual structure transfer framework for zero-resource event extraction. Experiments on three languages show promising results without using any annotation. In the future, we plan to conduct research on joint language-universal structure learning and event extraction.

Acknowledgement

This research is based upon work supported in part by U.S. DARPA KAIROS Program No. FA8750-19-2-1004, Air Force No. FA8650-17-C-7715, ARL NS-CTA No. W911NF-09-2-0053, DARPA LORELEI Program No. HR0011-15-C-0115, the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract No. FA8650-17-C-9116. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

6. Bibliographical References

Agerri, R., Aldabe, I., Laparra, E., Rigau Claramunt, G., Fokkens, A., Huijgen, P., Izquierdo Beviá, R., van Erp, M., Vossen, P., Minard, A.-L., et al. (2016). Multilingual event detection using the newsreader pipelines. In *Proceedings of LREC 2016*.

Chen, Y., Xu, L., Liu, K., Zeng, D., and Zhao, J. (2015). Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of ACL-IJCNLP 2015*.

Danilova, V., Alexandrov, M., and Blanco, X. (2014). A survey of multilingual event extraction from text. In *Proceedings of NLDB 2014*.

Dehdari, J. (2014). *A Neurophysiologically-Inspired Statistical Language Model*. Ph.D. thesis, The Ohio State University.

Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S. M., and Weischedel, R. M. (2004). The automatic content extraction (ace) program-tasks, data, and evaluation. In *Proceedings of LREC 2004*.

Dozat, T. and Manning, C. D. (2016). Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.

Feng, X., Feng, X., Qin, B., Feng, Z., and Liu, T. (2018a). Improving low resource named entity recognition using cross-lingual knowledge transfer. In *Proceedings of IJCAI 2018*.

Feng, X., Qin, B., and Liu, T. (2018b). A language-independent neural network for event detection. *Science China Information Sciences*.

Getman, J., Ellis, J., Strassel, S., Song, Z., and Tracey, J. (2018). Laying the groundwork for knowledge base population: Nine years of linguistic resources for tac kbp. In *Proceedings of LREC 2018*.

Honnibal, M. and Montani, I. (2017). spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing.

Hsi, A., Yang, Y., Carbonell, J., and Xu, R. (2016). Leveraging multilingual training for limited resource event extraction. In *Proceedings of COLING 2016*.

Ji, H. and Grishman, R. (2008). Refining event extraction through cross-document inference. In *Proceedings of ACL 2008*.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., and Mikolov, T. (2016). Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lample, G., Conneau, A., Denoyer, L., and Ranzato, M. (2017). Unsupervised machine translation using monolingual corpora only. *arXiv preprint arXiv:1711.00043*.

Li, P., Zhou, G., Zhu, Q., and Hou, L. (2012). Employing compositional semantics and discourse consistency in chinese event extraction. In *Proceedings of EMNLP-CoNLL 2012*.

Li, Q., Ji, H., and Huang, L. (2013). Joint event extraction via structured prediction with global features. In *Proceedings of ACL 2013*.

Li, Q., Ji, H., Yu, H., and Li, S. (2014). Constructing information networks using one single model. In *Proceedings of EMNLP 2014*.

Liao, S. and Grishman, R. (2011). Acquiring topic features to improve event extraction: in pre-selected and balanced collections. In *Proceedings of the RANLP 2011*.

Lin, Y., Yang, S., Stoyanov, V., and Ji, H. (2018). A multi-lingual multi-task architecture for low-resource sequence labeling. In *Proceedings of ACL 2018*.

Liu, J., Chen, Y., Liu, K., and Zhao, J. (2018a). Event detection via gated multilingual attention mechanism. In *Proceedings of AAAI 2018*.

Liu, X., Luo, Z., and Huang, H. (2018b). Jointly multiple events extraction via attention-based graph information aggregation. In *Proceedings of EMNLP 2018*.

Lu, W. and Nguyen, T. H. (2018). Similar but not the same: Word sense disambiguation improves event detection via neural representation matching. In *Proceedings of EMNLP 2018*.

Nguyen, T. H. and Grishman, R. (2015). Event detection and domain adaptation with convolutional neural networks. In *Proceedings of ACL-IJCNLP 2015*.

Nguyen, T. H. and Grishman, R. (2018). Graph convolutional networks with argument-aware pooling for event detection. In *Proceedings of AAAI 2018*.

Nguyen, T. H., Cho, K., and Grishman, R. (2016). Joint event extraction via recurrent neural networks. In *Proceedings of NAACL 2016*.

Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., McDonald, R. T., Petrov, S., Pyysalo, S., Silveira, N., et al. (2016). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of LREC 2016*.

Nivre, J., Abrams, M., Agić, Ž., Ahrenberg, L., Antonsen, L., Aplonova, K., Aranzabe, M. J., Arutie, G., Asahara, M., Ateyah, L., Attia, M., Atutxa, A., Augustinus, L., Badmaeva, E., Ballesteros, M., Banerjee, E., Bank, S., Barbu Mititelu, V., Basmov, V., Bauer, J., Bellato, S., Bengoetxea, K., Berzak, Y., Bhat, I. A., Bhat, R. A., Biagetti, E., Bick, E., Blokland, R., Bobicev, V., Börstell, C., Bosco, C., Bouma, G., Bowman, S., Boyd, A., Burchardt, A., Candito, M., Caron, B., Caron, G., Cebiroğlu Eryiğit, G., Cecchini, F. M., Celano, G. G. A., Čéplö, S., Cetin, S., Chalub, F., Choi, J., Cho, Y., Chun, J., Cinková, S., Collomb, A., Çöltekin, Ç., Connor, M., Courtin, M., Davidson, E., de Marneffe, M.-C., de Paiva, V., Diaz de Ilarraza, A., Dickerson, C., Dirix, P., Dobrovoljc, K., Dozat, T., Droganova, K., Dwivedi, P., Eli, M., Elkahky, A., Ephrem, B., Erjavec, T., Etienne, A., Farkas, R., Fernandez Alcalde, H., Foster, J., Freitas, C., Gajdošová, K., Galbraith, D., Garcia, M., Gärdenfors, M., Garza, S., Gerdes, K., Ginter, F., Goenaga, I., Gojenola, K., Gökırmak, M., Goldberg, Y., Gómez Guinovart, X., Gonzáles Saavedra, B., Grioni, M., Grūzītis, N., Guillaume, B., Guillot-Barbance, C., Habash, N., Hajič, J., Hajič jr., J., Hà Mỹ, L., Han, N.-R., Harris, K., Haug, D., Hladká, B., Hlaváčová, J., Hociung, F., Hohle, P., Hwang, J., Ion, R., Irimia, E., Ishola, Ọ., Jelínek, T., Johannsen, A., Jørgensen, F., Kaşıkara, H., Kahane, S., Kanayama, H., Kanerva, J., Katz, B., Kayadelen, T., Kenney, J., Kettnerová, V., Kirchner, J., Kopacewicz, K., Kotsyba, N., Krek, S., Kwak, S., Laippala, V., Lambertino, L., Lam, L., Lando, T., Larasati, S. D., Lavrentiev, A., Lee, J., Lê Hồng, P., Lenci, A., Lertpradit, S., Leung, H., Li, C. Y., Li, J., Li, K., Lim, K., Ljubešić, N., Loginova, O., Lyashevskaya, O., Lynn, T., Macketanz, V., Makazhanov, A., Mandl, M., Manning, C., Manurung, R., Mărănduc, C., Mareček, D., Marheinecke, K., Martínez Alonso, H., Martins, A., Mašek, J., Matsumoto, Y., McDonald, R., Mendonça, G., Miekka, N., Misirpashayeva, M., Missilä, A., Mititelu, C., Miyao, Y., Montemagni, S., More, A., Moreno Romero, L., Mori, K. S., Mori, S., Mortensen, B., Moskalevskyi, B., Muischnek, K., Murawaki, Y., Müürisep, K., Nainwani, P., Navarro Horñiacek, J. I., Nedoluzhko, A., Nešpore-Bērzkalne, G., Nguyễn Thị, L., Nguyễn Thị Minh, H., Nikolaev, V., Nitisaroj, R., Nurmi, H., Ojala, S., Olúòkun, A., Omura, M., Osenova, P., Östling, R., Øvrelid, L., Partanen, N., Pascual, E., Passarotti, M., Patejuk, A., Paulino-Passos, G., Peng, S., Perez, C.-A., Perrier, G., Petrov, S., Piitulainen, J., Pitler, E., Plank, B., Poibeau, T., Popel, M., Pretkalniņa, L., Prévost, S., Prokopidis, P., Przepiórkowski, A., Puolakainen, T., Pyysalo, S., Rääbis, A., Rademaker, A., Ramasamy, L., Rama, T., Ramisch, C., Ravishankar, V., Real, L., Reddy, S., Rehm, G., Rießler, M., Rinaldi, L., Rituma, L., Rocha, L., Romanenko, M., Rosa, R., Rovati, D., Ro□ca, V., Rudina, O., Rueter, J., Sadde, S., Sagot, B., Saleh, S., Samardžić, T., Samson, S., Sanguinetti, M., Saulīte, B., Sawanakunanon, Y., Schneider, N., Schuster, S., Seddah, D., Seeker, W., Seraji, M., Shen, M., Shimada, A., Shohibussirri, M., Sichinava, D., Silveira, N., Simi, M., Simionescu, R., Simkó, K., Šimková, M., Simov, K., Smith, A., Soares-Bastos, I., Spadine, C., Stella, A., Straka, M., Strnadová, J., Suhr, A., Sulubacak, U., Szántó, Z., Taji, D., Takahashi, Y., Tanaka, T., Tellier, I., Trosterud, T., Trukhina, A., Tsarfaty, R., Tyers, F., Uematsu, S., Urešová, Z., Uria, L., Uszkoreit, H., Vajjala, S., van Niekerk, D., van Noord, G., Varga, V., Villemonte de la Clergerie, E., Vincze, V., Wallin, L., Wang, J. X., Washington, J. N., Williams, S., Wirén, M., Woldemariam, T., Wong, T.-s., Yan, C., Yavrumyan, M. M., Yu, Z., Žabokrtský, Z., Zeldes, A., Zeman, D., Zhang, M., and Zhu, H. (2018). Universal dependencies 2.3. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Tai, K. S., Socher, R., and Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of NeurIPS 2017*.

Walker, C., Strassel, S., Medero, J., and Maeda, K. (2006). Ace 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*, 57.

Wei, S., Korostil, I., Nothman, J., and Hachey, B. (2017). English event detection with translated language features. In *Proceedings of ACL 2017*.

Xie, J., Yang, Z., Neubig, G., Smith, N. A., and Carbonell, J. (2018). Neural cross-lingual named entity recognition with minimal resources. *arXiv preprint arXiv:1808.09861*.

Yang, B. and Mitchell, T. (2016). Joint extraction of events and entities within a document context. *arXiv preprint arXiv:1609.03632*.

Zhang, B., Lin, Y., Pan, X., Lu, D., May, J., Knight, K., and Ji, H. (2018a). Elisa-edl: A cross-lingual entity extraction, linking and localization system. In *Proceedings of NAACL 2018*.

Zhang, J., Zhou, W., Hong, Y., Yao, J., and Zhang, M. (2018b). Using entity relation to improve event detection via attention mechanism. In *Proceedings of NLPCC 2018*.