Acquisition and Exploitation of Cross-Lingual Knowledge

Iker García-Ferrero[1]

[1]HiTZ Basque Center for Language Technologies - Ixa NLP Group, University of the Basque Country UPV/EHU

Abstract

Supervised neural networks have achieved great success in many Natural Language Processing tasks. However, for most of the more than 7000 spoken languages on Earth very limited or no resources are available for building NLP systems. Developing models and resources that allow us to perform NLP in multiple languages is an open challenge. We focus on the Zero-Resource Cross-lingual Sequence Labelling task. We propose a research project with the aim of developing high-quality sequence labelling models for languages for which no labelled data is available.

Keywords

Natural Language Processing Sequence Labelling, Multilingual, Cross-Lingual, Zero-shot

1. Introduction

This research project is framed within the area of Natural Language Processing (NLP). Natural Language Processing is a research field within artificial intelligence and linguistics, which studies how to computationally model human language.

Neural networks have become an indispensable resource in Natural Language Processing. Driven by the success of transformers [1], they have shown outstanding performance in very challenging NLP tasks such as General Language Understanding [2, 3], Question Answering [4], Text generation [5], Dialog [6], Text-Conditional Image Generation [7] among many others [8]. While all these models have been a breakthrough in the field, they are very expensive to train: They require huge computing capabilities, they come with a large carbon foot-print [9] and they require an enormous amount of data, that in many cases must be manually-annotated, which is very costly. The result is that most of the NLP systems cited above are limited to the English language. It is estimated that more than 7000 languages are spoken in the world today. For many of them, NLP resources are very limited or simply unavailable. Developing models and resources that allow us to perform NLP in multiple languages is an open challenge.

We focus our research on the Sequence Labelling task. Sequence labelling is the task of assigning a label to each token in a given input sequence. Figure 1 shows the example of Named Entity Recognition (NER). NER aims to locate and classify named entities in unstructured text

Doctoral Symposium on Natural Language Processing from the PLN.net network 2022 (RED2018-102418-T), 21-23 September 2022, A Coruña, Spain.

iker.garciaf@ehu.eus (I. García-Ferrero)
https://ikergarcia1996.github.io/Iker-Garcia-Ferrero/ (I. García-Ferrero)
0000-0001-9612-7134 (I. García-Ferrero)

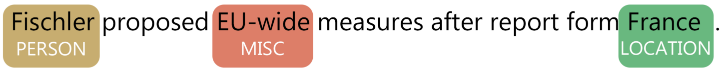


Figure 1: Illustration of the Named Entity Recognition Sequence Labelling Task

into a set of pre-defined categories such as organizations, locations, names of persons, dates... We choose to explore Cross-lingual Sequence Labelling because of the great challenge involved. Most successful approaches for sequence labelling involve supervised deep-neural networks. [10, 11, 12]. The difficulty of the task lies in the fact that model performance depends on the amount of manually annotated training data [13]. Moreover, models show a significant loss of performance when evaluated in out-of-domain data [14]. Thus, it would be necessary to develop annotated data for each language and domain of application. The cost of manual annotations makes this impossible. For most of the languages in the world, manually annotated corpora are simply nonexistent. The task of developing sequence labelling models for languages and domain-specific tasks, for which supervised data is not available, is a challenge of great interest. This task is known as *zero-resource cross-lingual sequence labelling*.

Our Main Research Question can be summarised as **"Which is the best technique to label a text in a language for which no labelled data is available?"**.

2. Related Work

Previous work has approached Cross-Lingual sequence labelling in two different directions: Data-based transfer and Model-based transfer methods.

2.1. Data transfer

Data transfer methods aim to automatically generate labelled data for a target language for which no labelled data is available. Ehrmann et al. [15] trains an English Sequence Labelling model using English gold labelled data. They use this model to label the English part of a multi-parallel corpus. The labels are then projected into all the other languages using statistical alignments of phrases. In this way, they generate annotated datasets in languages for which no data was initially available. Wang and Manning [16] projects models expectations instead of labels, this transfers the model uncertainty across languages. Ni et al. [17] improves previous works using a heuristic scheme that effectively selects good-quality projection-labelled data from noisy data. Instead of one-to-one projections, Agerri et al. [18] use labelled parallel data from multiple languages to project the labels to a single target language. The combination of multiple sources improves the equality of the projections. Li et al. [19] proposes to use the state-of-the-art XLM-R model [12] for labelling sequences in the source part of the parallel data and also for annotation projection.

Jain et al. [20] and Fei et al. [21] use machine translation instead of parallel data. A gold-labelled dataset in the source language is machine translated to the target languages. For this purpose, Jain et al. [20] first generates a list of projection candidates by orthographic and phonetic similarity. They use distributional statistics derived from the dataset to choose the best matching candidate. Fei et al. [21] leverages the word alignment probabilities calculated with FastAlign [22] and the POS tag distributions of the source and target words.

These methods assume that high-quality parallel data or machine translation systems are available for the source-target language pair. This is a strong assumption that is not true for many low-resource languages. Xie et al. [23] proposes to find word translations based on bilingual word embeddings trained on monolingual corpora from the source and target language. Guo and Roth [24] translates the source sentences to the target language word-by-word with a dictionary. Then, they generate high-quality annotated data in the target language using a constrained pre-trained language model.

2.2. Model transfer

Language models trained on monolingual corpora in many languages [11, 12] allow zero-shot cross-lingual model transfer. Using labelled data in one source language (usually English), we can fine-tune a pre-trained multilingual model and directly use it to make predictions in any of the languages included in the model [25]. The zero-shot cross-lingual capability can be improved for the sequence labelling task using different techniques. Wang et al. [26] and Ouyang et al. [27] use monolingual corpora in the source and the target language to improve the alignment of the language representations within a multilingual language model. [28] proposes to use many models from many source languages, they learn to infer which are the most reliable models in an unsupervised manner. The combination of the best models improves the zero-shot transfer to a new language. The approach of Wu et al. [29] take advantage of a Teacher-Student learning paradigm. Sequence Labelling models in the source languages are used as teachers to train a student model on unlabeled data in the target language. Bari et al. [30] propose an unsupervised data augmentation framework, using self-training they improve the cross-lingual adaptation of models. Hu et al. [31] use the minimum risk training framework to overcome the gap between the source and the target languages/domains. They propose a unified learning algorithm based on expectation-maximization.

Which one of the approaches produces the best results is unclear. Combinations of model-based and data-based transfer methods are also pending research. Some previous works claim contradictory results when using different language models. For example, Fei et al. [21] finds that their data transfer approach is superior to the zero-shot transfer method when using mBERT. On the other side, Li et al. [19] experiment with XLM-RoBERTa, a higher capacity multilingual model, and they obtain the best results for German and Chinese applying the data transfer approach, while the zero-shot approach is best for Spanish and Dutch. We seek to shed light on which is the best performing technique in each situation for Cross-Lingual Sequence Labelling and contribute with novel ideas to this line of research.

3. Proposed Experiments

3.1. Data-based transfer

RQ1: Can we automatically generate high quality data?

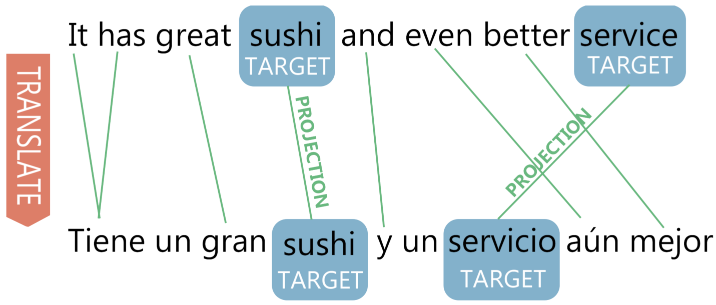


Figure 2: Illustration of the translation and annotation projection method for Opinion Target Extraction (OTE).

In Section 2.1 we have presented several previous works that successfully generate data for languages for which no labelled data is available. These methods rely on parallel data and annotation projection, which figure 2 illustratres. SimAlign [32] takes advantage of multilingual pre-trained language models to generate word alignments. SimAlign produces better results than previous statistical word alignment methods widely used in the field. AWESoME [33] improves the results even more by fine-tuning the language models on parallel text with unsupervised training objectives. In the machine translation field, M2M100 [34] can produce high quality translation between the 9,900 directions of 100 languages. These new systems have not been tested yet in the cross-lingual data transfer task. We expect that, since they are a qualitative leap over the systems used in previous research, they will generate improved data for languages for which no labelled data is available.

RQ2: Parallel Data vs Machine translation In Section 2.1 we present two main lines of research in data-transfer methods. On one side, some works take advantage of existing parallel data, while others use machine translation. The effect of using a parallel corpus or machine translation for data transfer is not well understood. We plan to explore both approaches to find out which type of data is better to use.

RQ3: Quality of the projections No in-deep study of the quality of the annotation projections produced by different systems and algorithms have been performed. Word alignment systems are evaluated with manually annotated word alignments, not in the annotation projection downstream task. Data-transfer methods are evaluated by training a model using the generated data. There is no evaluation of each step involved in the translation and annotation projection task. We plan to translate an English gold-labelled dataset and manually project the annotations. We will compare the results of the annotation projection systems with the manually annotated data. This will allow us to understand which are the errors produced in the annotation projection step. It will also allow us to decouple the translation and the annotation projection steps, to determine which of these most significantly affects the final performance of the models. We hope that the results of this experimentation will shed light on the errors made in each step of the data-transfer approach.

RQ4: Does the accumulation of automatically generated data for many languages yield to better results? Data-transfer methods allow to automatically data for a target language. Current translation [34] and multilingual pre-trained language models [12] support hundreds of languages. We can sequentially generate data for many languages. We want to leverage the accumulation of large amount of noisy data from many languages to produce high-quality data. This hypothesis has been successfully tested in the word alignment task [35].

RQ5: What is the impact of the amount of target language training data on prediction quality? Most cross-lingual sequence labelling methods assume a zero-shot setting, that is, no labelled data available in the target language. Manual annotations are very costly, however, labelling a small set of sentences in the target language can be feasible in many cases. We want to explore how a small amount of gold-labelled data in the target language affects the performance of the models. We expect that combining available gold labelled data in English with a small amount of target language labelled can yield good results.

3.2. Model-Transfer approaches

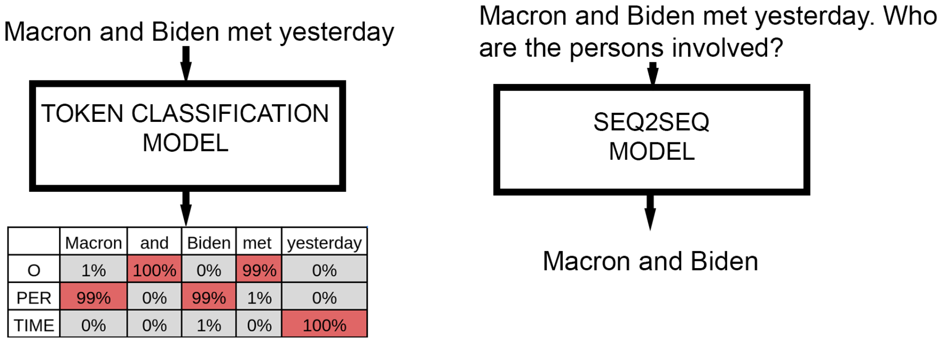
RQ6: How effective are state-of-the-art multi-lingual NLP models at cross-lingual sequence labelling There is a large number of pre-trained language models that can be fine-tuned for the Sequence Labelling Task. In Section 2.2 we also describe different works that aim to improve the cross-lingual capabilities of multilingual models. Most of these systems have not been evaluated against each other. It is not clear which one produces the best results. We plan to evaluate different models and systems.

RQ7: Model-transfer vs Data-transfer Fei et al. [21] finds that their data transfer approach is superior to the zero-shot transfer method when using Multilingual BERT. On the other side, Li et al. [19] experiment with XLM-RoBERTa and find opposite results, zero-shot model-transfer produce the best results for Spanish and Dutch. The cross-lingual capabilities of language models greatly differ between models with different capacities (number of parameters, training data...) and languages [25]. Which approach should be used given a target language, the available resources for the source and target languages and the available computer capacity? We want to empirically establish the required conditions for each of these two approaches, data-transfer and zero-shot model-transfer, to outperform the other.

3.3. Sequence Labelling as Text Generation

RQ8: Are seq2seq model a new paradigm for Cross-Lingual Sequence Labelling? Sequence classification is approached as a token classification task. Given a sequence, the probability scores for each word/token to belong to each predefined category are calculated. State-of-the-art models add a linear layer on top of each token representation of a transformer encoder [1] that has been pre-trained with a language modelling objective [11]. Recently a new trend for solving NLP tasks has emerged: The sequence to sequence (seq2seq or text2text) approach [3], taking text as input and producing new text as output. For example, we can input a text followed by the prompt "Who are the persons involved", and the model will produce a text enumerating the persons involved in the text. Figure 3 illustrates both, token classification and seq2seq approaches. This approach has already been tested with very promising results for Sequence Labelling in monolingual and cross-lingual zero-shot settings. [36].

Seq2Seq models can not only be trained to perform Sequence Labelling. They can be trained to generate new examples [37], which opens a new line of research in data-transfer methods.



	Macron	and	Biden	met	yesterday
O	1%	100%	0%	99%	0%
PER	99%	0%	99%	1%	0%
TIME	0%	0%	1%	0%	100%

Figure 3: Illustration of the Token Classification and Seq2Seq approaches for Sequence Labelling

We want to experiment with seq2seq models, such as the popular T5 [3] to find out if this new approach can improve previous work on zero-resource cross-lingual sequence labelling.

4. Conclusions

We present a research project in the field of Cross-Lingual Sequence Labelling in Zero-Resource Settings. We compile the most relevant previous research on the topic. We raise several research questions that will serve as the backbone of the experiments that we will carry out in the project.

Acknowledgments

Iker García-Ferrero is supported by a PhD grant from the Basque Government (PRE_2021_2_0219). I am grateful to my thesis supervisors German Rigau and Rodrigo Agerri for their guidance and help during the work done up to now.

References

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems 30, 2017, pp. 5998–6008. URL: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

[2] P. He, X. Liu, J. Gao, W. Chen, Deberta: decoding-enhanced bert with disentangled attention, in: 9th International Conference on Learning Representations, ICLR 2021,, OpenReview.net, 2021. URL: https://openreview.net/forum?id=XPZIaotutsD.

[3] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. J. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, J. Mach. Learn. Res. 21 (2020) 140:1–140:67. URL: http://jmlr.org/papers/v21/20-074.html.

[4] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, ALBERT: A lite BERT for

self-supervised learning of language representations, CoRR abs/1909.11942 (2019). URL: http://arxiv.org/abs/1909.11942. arXiv:1909.11942.

[5] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, CoRR abs/2005.14165 (2020). URL: https://arxiv.org/abs/2005.14165. arXiv:2005.14165.

[6] R. Thoppilan, D. D. Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H. Cheng, A. Jin, T. Bos, L. Baker, Y. Du, Y. Li, H. Lee, H. S. Zheng, A. Ghafouri, M. Menegali, Y. Huang, M. Krikun, D. Lepikhin, J. Qin, D. Chen, Y. Xu, Z. Chen, A. Roberts, M. Bosma, Y. Zhou, C. Chang, I. Krivokon, W. Rusch, M. Pickett, K. S. Meier-Hellstern, M. R. Morris, T. Doshi, R. D. Santos, T. Duke, J. Soraker, B. Zevenbergen, V. Prabhakaran, M. Diaz, B. Hutchinson, K. Olson, A. Molina, E. Hoffman-John, J. Lee, L. Aroyo, R. Rajakumar, A. Butryna, M. Lamm, V. Kuzmina, J. Fenton, A. Cohen, R. Bernstein, R. Kurzweil, B. Aguera-Arcas, C. Cui, M. Croak, E. H. Chi, Q. Le, Lamda: Language models for dialog applications, CoRR abs/2201.08239 (2022). URL: https://arxiv.org/abs/2201.08239. arXiv:2201.08239.

[7] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, M. Chen, Hierarchical text-conditional image generation with clip latents, 2022. URL: https://arxiv.org/abs/2204.06125. doi:10.48550/ARXIV.2204.06125.

[8] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heintz, D. Roth, Recent advances in natural language processing via large pre-trained language models: A survey, CoRR abs/2111.01243 (2021). URL: https://arxiv.org/abs/2111.01243. arXiv:2111.01243.

[9] E. Strubell, A. Ganesh, A. McCallum, Energy and policy considerations for deep learning in NLP, CoRR abs/1906.02243 (2019). URL: http://arxiv.org/abs/1906.02243. arXiv:1906.02243.

[10] A. Akbik, D. Blythe, R. Vollgraf, Contextual string embeddings for sequence labeling, in: Proceedings of the 27th International Conference on Computational Linguistics, 2018, pp. 1638–1649. URL: https://aclanthology.org/C18-1139.

[11] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 2019, pp. 4171–4186. URL: https://aclanthology.org/N19-1423. doi:10.18653/v1/N19-1423.

[12] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747. doi:10.18653/v1/2020.acl-main.747.

[13] R. Agerri, G. Rigau, Robust multilingual named entity recognition with shallow semi-supervised features, Artificial Intelligence 238 (2016) 63–82. URL: https://www.sciencedirect.com/science/article/pii/S0004370216300613. doi:https://doi.org/10.1016/j.artint.2016.05.003.

[14] Z. Liu, Y. Xu, T. Yu, W. Dai, Z. Ji, S. Cahyawijaya, A. Madotto, P. Fung, Crossner: Evaluating

cross-domain named entity recognition, in: The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, AAAI Press, 2021, pp. 13452–13460. URL: https://ojs.aaai.org/index.php/AAAI/article/view/17587.
[15] M. Ehrmann, M. Turchi, R. Steinberger, Building a multilingual named entity-annotated corpus using annotation projection, in: Proceedings of the International Conference Recent Advances in Natural Language Processing 2011, 2011, pp. 118–124. URL: https://aclanthology.org/R11-1017.
[16] M. Wang, C. D. Manning, Cross-lingual projected expectation regularization for weakly supervised learning, Transactions of the Association for Computational Linguistics 2 (2014) 55–66. URL: https://aclanthology.org/Q14-1005. doi:10.1162/tacl_a_00165.
[17] J. Ni, G. Dinu, R. Florian, Weakly supervised cross-lingual named entity recognition via effective annotation and representation projection, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, pp. 1470–1480. URL: https://aclanthology.org/P17-1135. doi:10.18653/v1/P17-1135.
[18] R. Agerri, Y. Chung, I. Aldabe, N. Aranberri, G. Labaka, G. Rigau, Building named entity recognition taggers via parallel corpora, in: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018. URL: https://aclanthology.org/L18-1557.
[19] B. Li, Y. He, W. Xu, Cross-lingual named entity recognition using parallel corpus: A new approach using xlm-roberta alignment, CoRR abs/2101.11112 (2021). URL: https://arxiv.org/abs/2101.11112. arXiv:2101.11112.
[20] A. Jain, B. Paranjape, Z. C. Lipton, Entity projection via machine translation for cross-lingual NER, 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), 2019, pp. 1083–1092. URL: https://aclanthology.org/D19-1100. doi:10.18653/v1/D19-1100.
[21] H. Fei, M. Zhang, D. Ji, Cross-lingual semantic role labeling with high-quality translated training corpus, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 7014–7026. URL: https://aclanthology.org/2020.acl-main.627. doi:10.18653/v1/2020.acl-main.627.
[22] C. Dyer, V. Chahuneau, N. A. Smith, A simple, fast, and effective reparameterization of IBM model 2, in: L. Vanderwende, H. D. III, K. Kirchhoff (Eds.), Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, The Association for Computational Linguistics, 2013, pp. 644–648. URL: https://aclanthology.org/N13-1073/.
[23] J. Xie, Z. Yang, G. Neubig, N. A. Smith, J. Carbonell, Neural cross-lingual named entity recognition with minimal resources, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018, pp. 369–379. URL: https://aclanthology.org/D18-1034. doi:10.18653/v1/D18-1034.
[24] R. Guo, D. Roth, Constrained labeled data generation for low-resource named entity recognition, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 4519–4533. URL: https://aclanthology.org/2021.findings-acl.396. doi:10.18653/v1/2021.findings-acl.396.
[25] T. Pires, E. Schlinger, D. Garrette, How multilingual is multilingual BERT?, in: Proceedings

of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 4996–5001. URL: https://aclanthology.org/P19-1493. doi:10.18653/v1/P19-1493.

[26] Z. Wang, J. Xie, R. Xu, Y. Yang, G. Neubig, J. Carbonell, Cross-lingual alignment vs joint training: A comparative study and a simple unified framework, 2019. URL: https://arxiv.org/abs/1910.04708. doi:10.48550/ARXIV.1910.04708.

[27] X. Ouyang, S. Wang, C. Pang, Y. Sun, H. Tian, H. Wu, H. Wang, Ernie-m: Enhanced multilingual representation by aligning cross-lingual semantics with monolingual corpora, 2021. arXiv:2012.15674.

[28] A. Rahimi, Y. Li, T. Cohn, Massively multilingual transfer for NER, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 151–164. URL: https://aclanthology.org/P19-1015. doi:10.18653/v1/P19-1015.

[29] Q. Wu, Z. Lin, B. Karlsson, J.-G. Lou, B. Huang, Single-/multi-source cross-lingual NER via teacher-student learning on unlabeled data in target language, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 6505–6514. URL: https://aclanthology.org/2020.acl-main.581. doi:10.18653/v1/2020.acl-main.581.

[30] M. S. Bari, T. Mohiuddin, S. Joty, Uxla: A robust unsupervised data augmentation framework for zero-resource cross-lingual nlp, 2021. arXiv:2004.13240.

[31] Z. Hu, Y. Jiang, N. Bach, T. Wang, Z. Huang, F. Huang, K. Tu, Risk minimization for zero-shot sequence labeling, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021, pp. 4909–4920. URL: https://aclanthology.org/2021.acl-long.380. doi:10.18653/v1/2021.acl-long.380.

[32] Z. Dou, G. Neubig, Word alignment by fine-tuning embeddings on parallel corpora, CoRR abs/2101.08231 (2021). URL: https://arxiv.org/abs/2101.08231. arXiv:2101.08231.

[33] M. Jalili Sabet, P. Dufter, F. Yvon, H. Schütze, SimAlign: High quality word alignments without parallel training data using static and contextualized embeddings, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, 2020, pp. 1627–1643. URL: https://www.aclweb.org/anthology/2020.findings-emnlp.147.

[34] A. Fan, S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, S. Goyal, M. Baines, O. Celebi, G. Wenzek, V. Chaudhary, N. Goyal, T. Birch, V. Liptchinsky, S. Edunov, E. Grave, M. Auli, A. Joulin, Beyond english-centric multilingual machine translation, CoRR abs/2010.11125 (2020). URL: https://arxiv.org/abs/2010.11125. arXiv:2010.11125.

[35] A. Imani, M. J. Sabet, L. K. Senel, P. Dufter, F. Yvon, H. Schütze, Graph algorithms for multiparallel word alignment, CoRR abs/2109.06283 (2021). URL: https://arxiv.org/abs/2109.06283. arXiv:2109.06283.

[36] K.-H. Huang, I.-H. Hsu, P. Natarajan, K.-W. Chang, N. Peng, Multilingual generative language models for zero-shot cross-lingual event argument extraction, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 2022, pp. 4633–4646. URL: https://aclanthology.org/2022.acl-long.317. doi:10.18653/v1/2022.acl-long.317.

[37] C. Qin, S. Joty, LFPT5: A unified framework for lifelong few-shot language learning based on prompt tuning of t5, in: International Conference on Learning Representations, 2022. URL: https://openreview.net/forum?id=HCRVf71PMF.