NECAT-CLWE: A SIMPLE BUT EFFICIENT PARALLEL DATA GENERATION APPROACH FOR LOW RESOURCE NEURAL MACHINE TRANSLATION

Rabiu Ibrahim Abdullahi *
Division of Agricultural Colleges (DAC)
Ahmadu Bello University, Zaria
Mando, Kaduna, Nigeria
raibrahim@abu.edu.ng

Idris Abdulmumin
Department of Computer Science
Ahmadu Bello University
Samaru, Zaria, Nigeria
iabdulmumin@abu.edu.ng

ABSTRACT

Many languages lack sufficient data to train qualitative translation systems, particularly those based on the cutting-edge neural machine translation architectures. Recently, it has been demonstrated that using an exact copy of the monolingual target data as the source data improves the quality of translation systems, allowing them to benefit from proper nouns and such similar words that do not require translation. However, using an exact copy of the target data contaminates the source data with terms in the target language that needs translation. As a result, we describe in this paper a similar but more effective parallel data generation approach for improving low-resource neural machine translation using named entity copying and approximate translations using cross-lingual word embedding (NECAT-CLWE). The work will be evaluated on the low resource English-Hausa neural machine translation.

1 INTRODUCTION

Since the dawn of time, language and communication have been important to human interactions. As a result, linguistic translation has played a crucial role in sociological and cultural achievements. To simplify and enable communication between different languages, an means is required. Because these means can be expensive or difficult to obtain, researchers have explored creating translations automatically. Machine Translation (MT) was one of the first applications that was designed to be achieved using a computer. Warren Weaver's "translation memorandum" coined this idea, and the IBM created a word-for-word translation system in 1954 (Hutchins, 1997). Since then, various strategies were created to handle the challenge of MT, the most notable of which is Statistical Machine Translation (SMT) (Koehn, 2009).

Neural Machine Translation (NMT) (Bahdanau et al., 2015; Vaswani et al., 2017), a deep learning-based approach to machine translation is currently being considered as a superior alternative to the prior state-of-the-art machine translation paradigm, the phrase-based SMT (PBSMT) (Koehn et al., 2003), due to a variety of factors, including its end-to-end learning flexibility and capacity to create superior translation (Toral & Sánchez-Cartagena, 2017). Because the quantity of parallel sentence pairs available for training has a direct impact on the performance of the NMT system and many languages lack sufficient amounts of this data, many research works have explored using other available resources, e.g., through Language Modelling (Lample et al., 2018) or Transfer Learning (Zoph et al., 2016). A strong emphasis has also been placed on producing parallel datasets (also known as bitext), in addition to research on novel MT approaches.

Recently, the back-translations (Sennrich et al., 2016a) and forward translations (Zhang & Zong, 2016) of the monolingual target and source data respectively are used to generate additional data. While the approaches have been shown to be successful at improving translation accuracy, they are very expensive to implement and the quality of the additional data is also not assured. A simple and cost-effective approach for utilizing the monolingual target data has been copying the target

*Corresponding Author.

data into the source language (Currey et al., 2017; Burlot & Yvon, 2018). This has been shown to benefit the translation model especially where the words are proper nouns or words that do not require translation. This approach, though, will litter the source language with words that are not in its vocabulary and require to be translated.

This research work, therefore, aims to improve upon the baseline work of Currey et al. (2017). While the named-entities will be shared between the source and target sentences, the unknown words that require translation will be replaced with the nearest words in the vocabulary of the source language using a pre-trained cross-lingual word embedding (JP et al., 2021). Cross-lingual word embedding have been shown to generate a similar vector representation for similar words across two or more languages (Zhang et al., 2019).

2 Problem Statement, Research Objectives and Significance

This section presents the motivation for conducting this research and the associated objectives or contributions.

2.1 Problem Statement (motivation)

Because of the diversity of languages and the never-ending need for inter-lingual communication among people, a means for automatic translation has never been so important. Neural Machine Translation (NMT) has shown a great potential for training excellent systems that can automatically translate between languages when the available manually-translated data between these languages are in abundance (Edunov et al., 2018). However, this method suffers from the scarcity of parallel data between the majority of languages, the Low Resource Languages (LRL).

Copying the target data into the source language has been shown as a straightforward and cost-effective method for utilizing the monolingual target data to improve low resource machine translation (Currey et al., 2017). The approach has been demonstrated to help the translation model, particularly with words that are proper nouns or words that do not need to be translated. However, using an exact copy of the target data will clog up the source data with words that are not part of its lexicon and, therefore, require translation.

2.2 Research Objectives (Contributions)

The aim of this study is to create an efficient parallel data generation technique based on named entity copying and approximate translation using cross-lingual word embedding. The following objectives are presented to achieve the stated aim.

1. To conduct an extensive review of current parallel data creation algorithms.
2. To develop a parallel data generation algorithm using named entity copying and approximate word translation with cross-lingual word embedding.
3. To evaluate the proposed approaches against current data generation algorithms on English-Hausa low resource NMT.

2.3 Significance of the Study

Due to the importance of having qualitative translation systems especially in languages that are categorized as low-resourced, the successful conduct of this research work will provide an efficient method for improving the performance of machine translation systems in low and even high resource languages. The work, being evaluated on English-Hausa NMT, will also help further develop research in the local Hausa language.

3 Related Literature

This section reviews literature that are related to the proposed approach: additional parallel data generation algorithm for low resourced languages.

3.1 Low Resource Languages

A low-resource problem in Natural Language Processing (NLP) can develop primarily because the investigated languages or domains are under-resourced (Hedderich et al., 2021). Researchers have attempted to define low-resourced languages (LRLs) by investigating several factors such as the number of mother-tongue speakers and the amount of unlabeled, labeled, or supplementary datasets available, as well as the availability of other NLP resources and tools (Hedderich et al., 2021). Over the years, there have been many initiatives to categorize languages according to the aforementioned different criteria.

Machine translation (MT) occurs between two languages. This means that in MT, the resourcefulness of a language pair is decided by the number of parallel corpora accessible between the target languages. When referring to the parallel corpora at hand, the terms extremely low-resource (or zero-resource), low-resource (LR) or high-resource (HR) have been used regularly (Hedderich et al., 2021). However, there is no minimum size requirement for parallel corpora to classify a language pair as high, low, or extremely low-resource. Some early research considered 1 million parallel sentences and below as LR (Zoph et al., 2016). Recent research appears to regard a language pair as LR or severely LR if the available parallel corpora for the studied pair for NMT trials is less than 0.5 million and 0.1 million respectively (Lakew et al., 2020; Liu et al., 2020; Platanios et al., 2018). However, these do not represent absolute numbers for the size of dataset.

3.2 Augmenting Low Resource Dataset

Translation into or out of under-resourced languages has always been viewed as a difficult endeavor (Durrani & Koehn, 2014; Haque et al., 2012; Nießen & Ney, 2004; Resnik & Smith, 2003; Utiyama & Isahara, 2007). Despite the ability of NMT to produce state-of-the-art outcomes in a wide variety of translation tasks, it continues to offer numerous challenges for language pairings with scarce resources (Cheng et al., 2019; Sennrich & Zhang, 2019). Researchers in MT have tried a variety of strategies to solve this challenge, such as using monolingual data from the source and/or target side for training (Cheng et al., 2019; Sennrich et al., 2016a), augmenting bilingual training data (Fadaee & Monz, 2018), and exploiting training data in other language pairs (Zoph et al., 2016).

As proposed in Currey et al. (2017), a novel method for incorporating target-side monolingual data entails copying a monolingual corpus and converting it into a bitext with identical source and target sides. This copied bitext is then combined with the parallel and/or back-translated data to train an NMT system with no distinction made between the copied, back-translated, and parallel data during training. They experimented a mixing ratio of 1:2:2 between the parallel, copied, and back-translated data and reused the same monolingual data for generating both the copied and back-translated corpora.

Burlot & Yvon (2018) undertook a comprehensive investigation of back-translation (BT), evaluating other uses of monolingual data and multiple data production processes. They also investigated several approaches to integrating monolingual data in an NMT framework, with an emphasis on the influence on quality and domain adaptability. While verifying the usefulness of BT, the researchers also recommended substantially less expensive alternatives to increase baseline performance, such as employing a slightly modified copy of the target rather than its complete BT. When no high-quality BT is available, they suggested doing domain adaptation by employing Generative Adversarial Networks (GANs) to get the pseudo-source sentences closer to natural source sentences.

As stated by Wu et al. (2019), back-translation has been widely employed in prior techniques to unsupervised neural machine translation, in which synthetic sentence pairings are constructed to train the models with a reconstruction loss. However, because translation errors accrue throughout training, these synthetic sentences are frequently of poor quality. To avoid this fundamental issue, they suggested a different but more effective approach, extract-edit, in which genuine phrases are extracted and subsequently edited from the target monolingual corpora. They also proposed a comparative translation loss for evaluating translated target sentences and thereby training unsupervised translation systems. Experiments show that their proposed approaches consistently outperforms previous state-of-the-art unsupervised machine translation systems.

Recently, Sennrich & Zhang (2019) adapted an NMT system to low-resource environments using a variety of strategies, including modifying the model's major hyper-parameters (learning rate, model depth, label smoothing rate), and lowering byte-pair encoding (BPE) (Sennrich et al., 2016b) Increasing the size of the vocabulary and decreasing the size of the batch. Their findings show that, contrary to earlier research, NMT can outperform PBSMT in low-data scenarios.

In the works of Abdulmumin et al. (2021a;b), the authors presented novel approaches that combine self-learning and back-translation to allow both backward and forward models to benefit from monolingual target data. Experimental results showed that the proposed strategy outperformed the classic back-translation method on English-German low resource neural machine translation. They also proposed an iterative self-learning strategy that outperforms iterative back-translation while requiring less model training and only monolingual target data.

Wu et al. (2021) discovered that the two methods can complement each other in the sense that the former can exploit context information via language-independent features but sees no task-specific information in the target language, whereas the latter generates pseudo target-language training data via translation but its exploitation of context information is hampered by inaccurate translations. Furthermore, previous studies have infrequently used unlabeled data in the target language, which is easily obtained and may provide useful information for improved findings. To address both issues, they offer Unifying Transfer, a novel approach that unifies both model and data transfer for cross-lingual named entity recognition, as well as better knowledge distillation to use the available information from unlabeled target-language data.

4 The Proposed NECAT-CLWE Parallel Data Generation Approach

We illustrate the proposed Named-Entity Copying and Approximate Translation using Cross-Lingual Words Embedding approach in Figure 1 and Algorithm 1.

Algorithm 1 Named-Entity Copying and Approximate Translation using Cross-Lingual Words Embedding

Input: Monolingual source and target data: $X = \{(x^{(u)})\}_{u=1}^{U}$ & $Y = \{(y^{(v)})\}_{v=1}^{V}$

```
procedure PARALLEL DATASET GENERATION
    Train a cross-lingual words embedding (CLWE) vector on a mixture of the source and target monolingual data, X ∪ Y;
    Build the vocabularies of each of the source and target monolingual datasets;
    repeat
        Tokenize the target sentence, Y;
        Extract all named-entities;
        Copy the named-entities into the same position in the source sentences, X′;
        For each word in the target sentence that require translation, select and insert into the same position in the source sentence the nearest word in the source vocabulary using the pretrained CLWE;
    until monolingual target sentences are exhausted;
end procedure
```

Output: pseudo-parallel data, $P^{'} = \{(x^{i(v)}, y^{(v)})\}_{v=1}^{V}$

Firstly, we get the monolingual data in each of the target and source languages, mix the two data and train a cross-lingual words embedding. Secondly, we copy all the named entity words into the generated source sentences at the same positions as in the target sentences. Then, for each word in the target sentences that require translation, we select and insert into the source sentences the nearest word in the source vocabulary using the pretrained cross-lingual words embedding. These steps are iterated until all the target sentences are paired with their pseudo-parallel, approximate translations.

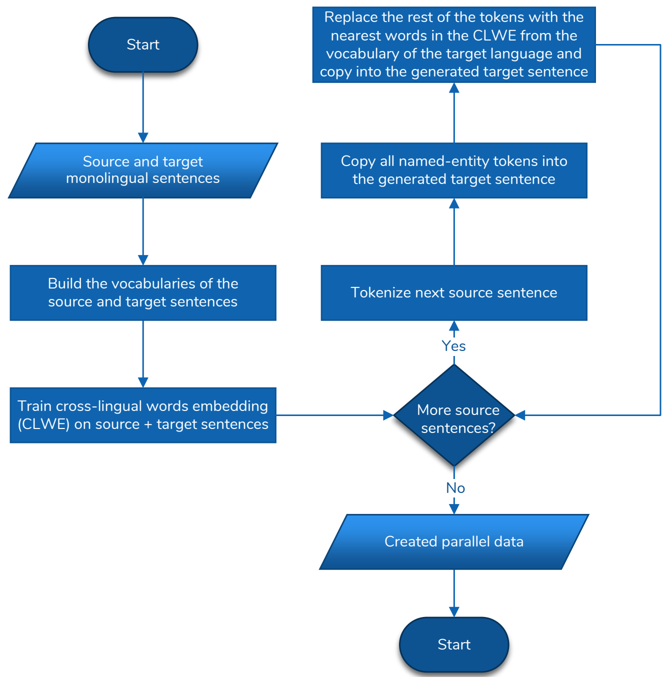

Figure 1: Named-Entity Copying and Approximate Translation using Cross-Lingual Words Embedding

We also experiment replacing the words that require translation with the special unknown token: *'<unk>'* – see Algorithm 2. Although the special token is another form of noising, it is still not a word that will confuse the decoder into thinking all the words in the target sentences do not require translation. Also, we hypothesize that the translation model will learn to ignore the special token.

Algorithm 2 Named-Entity Copying with Unknown Word Replacement

Input: Monolingual target data, $Y = \{(y^{(v)})\}_{v=1}^{V}$
1: **procedure** PARALLEL DATASET GENERATION
2: **repeat**
3: Tokenize the target sentence, Y;
4: Extract all named-entities;
5: Copy the named-entities into the same position in the source sentences;
6: Replace all other words that require translation with the unknown token: *'<unk>'*;
7: **until** monolingual target sentences are exhausted;
8: **end procedure**
Output: pseudo-parallel data, $P^{'} = \{(x^{(v)}, y^{(v)})\}_{v=1}^{V}$

5 CONCLUSION AND FUTURE WORK

Due to the absence of sufficient parallel datasets for training machine translation models, several research works are proposing efficient approaches for utilizing other resources to improve the accuracy of these models. In this work, we presented a conceptual framework, called NECAT-CLWE—Named-Entity Copying and Approximate Translation using Cross-Lingual Words Embedding, a work-in-progress Masters Dissertation study, that aims to utilize the abundant monolingual target data to improve low resource neural machine translation using named-entity copying and word substitution with cross-lingual words embedding. We aim to implement the proposed approach and evaluate it against current data generation algorithms.

REFERENCES

Idris Abdulmumin, Bashir Shehu Galadanci, Ibrahim Said Ahmad, and Rabiu Ibrahim Abdullahi. Data Selection as an Alternative to Quality Estimation in Self-Learning for Low Resource Neural Machine Translation. pp. 311–326. 2021a. doi: 10.1007/978-3-030-87013-3_24. URL https://link.springer.com/10.1007/978-3-030-87013-3_24.

Idris Abdulmumin, Bashir Shehu Galadanci, Abubakar Isa, Habeebah Adamu Kakudi, and Ismaila Idris Sinan. A Hybrid Approach for Improved Low Resource Neural Machine Translation using Monolingual Data. *Engineering Letters*, 29(4):1478–1493, nov 2021b. URL http://www.engineeringletters.com/issues_v29/issue_4/EL_29_4_21.pdf.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. In Yoshua Bengio and Yann LeCun (eds.), *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. URL http://arxiv.org/abs/1409.0473.

Franck Burlot and François Francois Yvon. Using Monolingual Data in Neural Machine Translation: a Systematic Study. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 144–155, Brussels, Belgium, oct 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-6315. URL https://www.aclweb.org/anthology/W18-6315.

Yong Cheng, Lu Jiang, and Wolfgang Macherey. Robust neural machine translation with doubly adversarial inputs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4324–4333, Florence, Italy, 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1425. URL https://aclanthology.org/P19-1425.

Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield. Copied Monolingual Data Improves Low-Resource Neural Machine Translation. In *Proceedings of the Second Conference on Machine Translation*, volume 1, pp. 148–156, Copenhagen, Denmark, 2017. Association for Computational Linguistics. doi: 10.18653/v1/W17-4715.

Nadir Durrani and Philipp Koehn. Improving machine translation via triangulation and transliteration. In *Proceedings of the 17th Annual conference of the European Association for Machine Translation*, pp. 71–78, Dubrovnik, Croatia, 2014. European Association for Machine Translation. URL https://aclanthology.org/2014.eamt-1.17.

Sergey Edunov, Myle Ott, Michael Auli, and David Grangier. Understanding Back-Translation at Scale. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 489–500, Stroudsburg, PA, USA, 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1045. URL http://aclweb.org/anthology/D18-1045.

Marzieh Fadaee and Christof Monz. Back-Translation Sampling by Targeting Difficult Words in Neural Machine Translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 436–446, Brussels, Belgium, 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1040. URL https://www.aclweb.org/anthology/D18-1040.

Rejwanul Haque, Sergio Penkale, Jie Jiang, and Andy Way. Source-side suffix stripping for bengali-to-english smt. In *Proceedings of the 2012 International Conference on Asian Language Processing*, IALP '12, pp. 193–196, USA, 2012. IEEE Computer Society. ISBN 9780769548869. doi: 10.1109/IALP.2012.61. URL https://doi.org/10.1109/IALP.2012.61.

Michael A. Hedderich, Lukas Lange, Heike Adel, Jannik Strötgen, and Dietrich Klakow. A survey on recent approaches for natural language processing in low-resource scenarios. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2545–2568, Online, 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.201. URL https://aclanthology.org/2021.naacl-main.201.

John Hutchins. From First Conception to First Demonstration: the Nascent Years of Machine Translation, 1947–1954. A Chronology. *Machine Translation*, 12(3):195–252, 1997. ISSN 1573-0573. doi: 10.1023/A:1007969630568. URL https://doi.org/10.1023/A:1007969630568.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6282–6293, Online, 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.560. URL https://aclanthology.org/2020.acl-main.560.

Sanjanasri JP, Vijay Krishna Menon, Soman KP, Rajendran S, and Agnieszka Wolk. Generation of Cross-Lingual Word Vectors for Low-Resourced Languages Using Deep Learning and Topological Metrics in a Data-Efficient Way. *Electronics*, 10(12), 2021. ISSN 2079-9292. doi: 10.3390/electronics10121372. URL https://www.mdpi.com/2079-9292/10/12/1372.

Philipp Koehn. *Statistical Machine Translation*. Cambridge University Press, 2009. ISBN 978-0-521-87415-1.

Philipp Koehn, Josef Franz Och, and Daniel Marcu. Statistical Phrase-Based Translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 127–133, 2003. URL https://www.aclweb.org/anthology/N031017.

Surafel M. Lakew, Matteo Negri, and Marco Turchi. Low resource neural machine translation: A benchmark for five african languages, 2020.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Unsupervised Machine Translation Using Monolingual Corpora Only. In *International Conference on Learning Representations*, 2018. URL http://arxiv.org/abs/1711.00043.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual Denoising Pre-training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, 8:726–742, 11 2020. ISSN 2307-387X. doi: 10.1162/tacl_a_00343. URL https://doi.org/10.1162/tacl_a_00343.

Sonja Nießen and Hermann Ney. Statistical machine translation with scarce resources using morpho-syntactic information. *Computational Linguistics*, 30(2):181–204, 2004. doi: 10.1162/089120104323093285. URL https://aclanthology.org/J04-2003.

Emmanouil Antonios Platanios, Mrinmaya Sachan, Graham Neubig, and Tom Mitchell. Contextual parameter generation for universal neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 425–435, Brussels, Belgium, 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1039. URL https://aclanthology.org/D18-1039.

Philip Resnik and Noah A. Smith. The web as a parallel corpus. *Computational Linguistics*, 29(3): 349–380, 2003. doi: 10.1162/089120103322711578. URL https://aclanthology.org/J03-3002.

Rico Sennrich and Biao Zhang. Revisiting Low-Resource Neural Machine Translation: A Case Study. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 211–221, Stroudsburg, PA, USA, 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1021. URL https://www.aclweb.org/anthology/P19-1021.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 86–96, Berlin, Germany, 2016a. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, Stroudsburg, PA, USA, 2016b. Association for Computational Linguistics. doi: 10.18653/v1/P16-1162. URL `http://aclweb.org/anthology/P16-1162`.

Antonio Toral and Víctor M. Sánchez-Cartagena. A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pp. 1063–1073, Valencia, Spain, 2017. Association for Computational Linguistics. URL `https://aclanthology.org/E17-1100`.

Masao Utiyama and Hitoshi Isahara. A comparison of pivot methods for phrase-based statistical machine translation. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pp. 484–491, Rochester, New York, 2007. Association for Computational Linguistics. URL `https://aclanthology.org/N07-1061`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In *31st Conference on Neural Information Processing Systems*, Long Beach, CA, USA, 2017.

Jiawei Wu, Xin Wang, and William Yang Wang. Extract and edit: An alternative to back-translation for unsupervised neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 1173–1183, Minneapolis, Minnesota, 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1120. URL `https://aclanthology.org/N19-1120`.

Qianhui Wu, Zijia Lin, Börje F. Karlsson, Biqing Huang, and Jian-Guang Lou. Unitrans: Unifying model transfer and data transfer for cross-lingual named entity recognition with unlabeled data. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, 2021. ISBN 9780999241165.

Jiajun Zhang and Chengqing Zong. Exploiting Source-side Monolingual Data in Neural Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1535–1545, Austin, Texas, 2016. Association for Computational Linguistics. doi: 10.18653/v1/d16-1160.

Mozhi Zhang, Keyulu Xu, Ken-ichi Kawarabayashi, Stefanie Jegelka, and Jordan Boyd-Graber. Are girls neko or shōjo? cross-lingual alignment of non-isomorphic embeddings with iterative normalization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3180–3189, Florence, Italy, 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1307. URL `https://aclanthology.org/P19-1307`.

Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. Transfer Learning for Low-Resource Neural Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1568–1575, Austin, Texas, 2016. Association for Computational Linguistics. doi: 10.18653/v1/D16-1163. URL `https://www.aclweb.org/anthology/D16-1163`.