

Content-Equivalent Translated Parallel News Corpus and Extension of Domain Adaptation for Neural Machine Translation

Hideya Mino^{1,3} Hideki Tanaka² Hitoshi Ito¹ Isao Goto¹

Ichiro Yamada¹ Takenobu Tokunaga³

¹ NHK Science & Technology Research Laboratories
1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan

² NHK Engineering System

1-10-11 Kinuta, Setagaya-ku, Tokyo 157-8510, Japan

³ Tokyo Institute of Technology

2-12-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

mino.h-gq@nhk.or.jp, tanaka.hideki@nes.or.jp,

{itou.h-ce, goto.i-es, yamada.i-hy}@nhk.or.jp, take@c.titech.ac.jp

Abstract

In this paper, we deal with two problems in Japanese-English machine translation of news articles. The first problem is the quality of parallel corpora. Neural machine translation (NMT) systems suffer degraded performance when trained with noisy data. Because there is no clean Japanese-English parallel data for news articles, we build a novel parallel news corpus consisting of Japanese news articles translated into English in a content-equivalent manner. This is the first content-equivalent Japanese-English news corpus translated specifically for training NMT systems. The second problem involves the domain-adaptation technique. NMT systems suffer degraded performance when trained with mixed data having different features, such as noisy data and clean data. Though existing domain-adaptation methods try to overcome this problem by using tags to distinguish the differences between corpora, it is not sufficient. We thus extend a domain-adaptation method by using multiple tags to train an NMT model effectively with both the clean corpus and existing parallel news corpora with some types of noise. Experimental results show that our corpus increases the translation quality, and that our domain-adaptation method is more effective for learning with multiple types of corpora than existing domain-adaptation methods are.

Keywords: Parallel News Corpus, Japanese-English, Machine Translation, Domain Adaptation, Back-Translation

1. Introduction

Recently, the number of foreigners visiting Japan has increased, and most of them cannot understand Japanese. The information gap between Japanese and these foreigners is a significant problem. Translation of Japanese news into English is one way to address this problem. Development of neural machine translation (NMT) systems requires a huge amount of clean parallel data. There is already work on Japanese-English parallel news corpora such as the Jiji Corpus¹ or the newspaper corpus of the Yomiuri Database Service². Though these parallel corpora are precious resources for developing news-focused NMT systems, they include much noise for NMT because of two main factors. The first is that some information is often omitted or added in the English sentences, as compared with the Japanese sentences, because the English news articles are generated for English-speaking readers through news writing, not just translating. The second is that these corpora have alignment errors because they are constructed with automatic sentence alignment. Figure 1 shows an example of data with noise due to both omission and addition. The noisier the training data becomes, the more the translation quality of NMT systems deteriorates. To alleviate the degradation due to existing corpora with noisy data, we constructed a new clean parallel corpus. It was made by manually translating Japanese news articles into English in a content-equivalent manner, and we thus

	Omitted	Added
Japanese sentence		臨時休校の間、友人宅を <u>行き来して遊んでいた</u> 孫は、授業再開が決まると「学校に行ける！」「 <u>給食何かな</u> 」と、登校を楽しみにしていたという。
English sentence	<u>According to the man,</u>	his grandchild was looking forward to the resumption of classes while the school was closed.

Content-equivalent translation of Japanese sentence:
His grandchild, who played with his friend while the school was closed, said “I can go to the school! What is the school lunch?” and was looking forward to going to the school when the resumption of classes was decided.

Figure 1: Example of a Japanese-English parallel sentence pair with noise. The underlined parts are not in the other language.

call it a “content-equivalent corpus.” Because the amount of content-equivalent corpus data is insufficient for developing a high-quality Japanese→English NMT system for news, we constructed two more Japanese-English parallel news corpora with different features. One was made with an automatic sentence alignment method, as in the case of existing parallel news corpora with noise, and we call this an “automatic-alignment corpus.” The other was made with a back-translation technique (Sennrich et al., 2016a) to leverage monolingual English news articles, and we call this a “back-translated corpus.” Thus, we used three different types of parallel data, from the content-equivalent,

¹<http://lotus.kuee.kyoto-u.ac.jp/WAT/jiji-corpus/>

²<https://database.yomiuri.co.jp/about/glossary/>

Table 1: Feature differences between corpora.

Corpus name	Source	Target
Content-equivalent corpus	Original news	Content-equivalent news
Automatic-alignment corpus	Original news	Original news (with noise between source and target sides)
Back-translated corpus	MT output	Original news

automatic-alignment, and back-translated corpora, to train NMT systems. The content-equivalent corpus was non-noisy. In contrast, the automatic-alignment corpus was noisy because of the two factors: first, that the target-side English sentences were generated through news writing, and second, that the Japanese and English sentences were aligned automatically. Furthermore, the back-translated corpus was noisy because the source-side Japanese sentences were generated by an English→Japanese NMT system. Table 1 lists the feature differences between the corpora.

To exploit corpora with these different features, a method of domain tagging (Chu et al., 2017; Kobus et al., 2017), which is a domain-adaptation technique, can be applied. Unfortunately, existing domain-tag methods cannot sufficiently express the differences between these corpora because of the lack of tag information. For example, though the target-side sentences in both the automatic-alignment corpus and the back-translated corpus come from the same domain of original news, as listed in Table 1, the existing methods cannot express this difference. Such features of the target-side sentences are significant for controlling the output sentences. To solve this problem, we developed a multi-tag method that can appropriately express the features of each corpus. Our experimental results showed that the multi-tag method improved the translation quality as compared with the existing domain-tag methods. In summary, we list the contributions of this paper as follows.

- For a Japanese-English parallel news corpus, we built a content-equivalent corpus consisting of Japanese news articles translated into English in a content-equivalent manner.
- We extended a domain-adaptation method with a multi-tag method to exploit multiple corpora with different features.

2. Building Parallel News Corpus Translated in Content-Equivalent Manner

This section describes the need for a content-equivalent parallel news corpus and the details of its construction method. The original news data used for corpus construction came from Japanese and English news articles by Jiji Press, a Japanese news agency. Jiji Press delivers Japanese news to its subscribers and also serves the general public on the internet. Selected Japanese news articles are translated into

Table 2: Numbers of news articles used for corpora construction

Japanese articles	1,561,143
Japanese-English articles	57,154

English by in-house translators of Jiji Press and also delivered to subscribers and served to the general public. Table 2 lists the numbers of Japanese and English news articles used for corpus construction. It shows that 3.7% of the Japanese articles were translated into English. It has links at the article level but not the sentence level. The linked articles can be applied for automatic sentence alignment.

2.1. Problems of Automatic Sentence Alignment

Automatic sentence alignment methods are quite cost effective for building parallel corpora: many bilingual corpora have been made with these methods, including some attempts mentioned in section 1.

Despite the cost effectiveness, however, application of automatic sentence alignment methods to our Japanese and English news data may invite noise into the results, as sentence pairs contain information that exists in either Japanese or English only. There are two reasons for this.

First, this problem is due to the content imbalance of the Japanese and English news articles. The Japanese articles are heavily edited by Jiji Press during translation for English-speaking readers. Overly detailed information is omitted, and necessary background information is supplemented. Such omission and addition occur at the levels of expressions, sentences, and paragraphs. Connecting such sentences will inevitably add noise.

Second, the noise is produced by alignment errors. Typical alignment methods such as Utiyama and Isahara (2007) use dynamic programming to align source and target sentences by maximizing the sum of the similarities of all linked pairs and also making the links not cross each other. Such methods may yield pairs sharing little common content, thus giving alignment errors. In particular, these errors often occur when the sentence order differs greatly between the source and target.

2.2. Construction Method of Content-Equivalent Corpus

Because of the two problems with automatic sentence alignment, we decided to create a high-quality parallel corpus (content-equivalent corpus) by asking professional translators to translate Japanese news articles into English.

2.2.1. Article Selection

We selected Japanese articles that did not have English translations and met the following requirements, as translation of all the Japanese sentences listed in Table 2 would have been too costly.

- Period
We selected Japanese articles dated between 2016 and 2018.

Table 3: Numbers of sentences and words (on the English side) in the training corpora.

Corpus name	# original		Japanese		English	
	data	# data	# sentences	# characters	# sentences	# words
Content-equivalent corpus	-	220,180	220,180	10,427,732	235,407	6,052,647
Automatic-alignment corpus	582,572	286,247	286,247	17,521,620	286,247	9,090,966
Back-translated corpus (CE-NMT)	-	533,581	533,581	22,798,841	533,581	14,348,304
Back-translated corpus (AA-NMT)	-	533,581	533,581	21,861,647	533,581	14,348,304

- **Format**

Japanese news articles like stock price reports have only numbers arranged in a tabular form, while articles on personnel transfers in government agencies have lists of names. We did not have such articles containing few sentences translated. We did select Japanese articles with 5 to 15 sentences, as 80% of the articles translated into English fell in that range.

- **Content**

As listed in Table 2, 3.7% of the Japanese news articles were translated into English by Jiji Press, and these articles were specially selected for English-speaking readers. We wanted to exclude Japanese articles whose topics differed greatly from those of the translated articles. For this purpose, we measured the similarity between the keyword lists of the articles in Japanese and those translated into English, and we excluded Japanese scripts with extremely low similarity.

2.2.2. Translation Policies

We asked the translators to follow the translation policies below.

- **Content equivalency**

We asked the translators not to imitate the Jiji approach of heavy editing in translation, but instead to preserve all Japanese information in the resulting English articles.

- **Translation unit and order**

One Japanese sentence was translated into one or more English sentences. We did not allow translation of multiple Japanese sentences into one English sentence. The order of the sentences in Japanese and English had to be the same.

- **Use of context**

To create natural English news articles, we allowed positive use of contextual information. For example, Japanese subjects are often omitted when they are clear from the context, but English subjects are essential. We thus asked the translators to refer to the context and provide correct subjects in English. Also, unlike in Japanese, inanimate subjects are used quite often in English, and we allowed the translators to use them with changes in voice when necessary.

- **Use of style guides**

Jiji Press uses in-house style guides that define term

usage and translation. Because the manually translated and aligned corpora were mixed together, it was desirable for the surface styles of the two corpora to be much alike. We thus provided the Jiji style guides to the translators, who might not have been well versed in Jiji news. We also supplied a supplementary style guide that we compiled through observation of past news. It covered the norms for translating titles and the characteristics of the Jiji news structure.

We hired four Japanese translation agencies to perform news translation according to the policies above. Although their translators were professionals, they were not necessarily well versed in news translation, especially according to the Jiji style guides.

To ensure and equalize quality across the companies, we asked another translation agency to review sample translations submitted by each agency once or twice during the work period. The reviewer checked for errors and discrepancies from the style guides and fed error reports back to all of the agencies to share the problematic parts among them. The agencies then corrected erroneous translations and also shared the error reports with their translators so as not to reproduce the same errors.

The review was conducted six times during the full corpus construction period. We consider the process to have greatly contributed to ensuring the high quality of our corpus.

2.3. Current Status and Future Plans

The first row in Table 3 lists the numerical specifications of our current corpus. Here we report some observations on it. Our translation policies stipulated that all Japanese sentences had to be translated, and that each Japanese sentence had to be translated into one or more English sentences. The number of English sentences was 235,407, corresponding to the 220,180 Japanese sentences, which means that only about 7% of the Japanese sentences in the training data were translated into multiple English translations. Our corpus thus has one-to-one sentence correspondence at a pretty high level.

We are now expanding the size of the corpus and trying different approaches besides human translation from scratch. Concerning its availability, we are in the process of negotiating with the content holders and public agencies to release the corpus.

Table 4: Attached tags in the four corpora for each method.

Domain-tag method for domain adaptation	Content-equivalent corpus	Automatic-alignment corpus	Back-translated corpus (CE-NMT)	Back-translated corpus (AA-NMT)
Single tag (Kobus et al., 2017)	<CE>	<AA>	<CE>	<AA>
Single tag with back-translation (Vaibhav et al., 2019)	<CE>	<AA>	<BT-CE>	<BT-AA>
Two types of tags (Berard et al., 2019a)	<CE>	<AA>	<BT> <CE>	<BT> <AA>
Multi-tag method (proposed)	<NS-S> <CE-T> <NO-BT> <NO-AN>	<NS-S> <NS-T> <NO-BT> <AN>	<CE-S> <NS-T> <BT> <NO-AN>	<NS-S> <NS-T> <BT> <AN>

3. Extension of Domain Adaptation for NMT

In this section, we describe the corpora we used in our experiments. We then discuss problems in training NMT models with existing methods and propose a new method to solve those problems.

3.1. Training Corpora

We used four corpora for training NMT models. Table 3 lists the numbers of sentences and words in each corpus. In addition to the content-equivalent corpus described in section 2, we augmented the training data with an “automatic-alignment corpus”³ and a “back-translated corpus.”

The automatic-alignment corpus that contains 286,247 sentence pairs, which were extracted as those with a similarity score above 0.3 from the aligned data containing 582,572 sentence pairs. Therefore, only 49% of the original data was used for the parallel corpus.

The back-translated corpus was constructed by a back-translation technique (Sennrich et al., 2016a). Back-translation is one of the most popular techniques to increase the size of parallel data when a large amount of target-side monolingual data is available. Because we also had a large amount of English monolingual news data, with a size of 533,581 sentences, we used it for back-translation. We trained two NMT models, CE-NMT and AA-NMT, in the English→Japanese direction, adapted to the content-equivalent and automatic-alignment corpora, respectively. Then, we translated all the English monolingual news data into Japanese with each of the two NMT models. We call these corpora the “back-translated corpus (CE-NMT)” and the “back-translated corpus (AA-NMT).” Though the amount of parallel data was increased through the back-translation, the augmented data included noise on the source side (Japanese) because of the use of imperfect translation results by the NMT system.

3.2. Existing Domain-Adaptation Methods Using Tags

The naive approach for NMT training is to use the entire corpora. NMT systems have been shown, how-

ever, to have degraded performance when trained with out-of-domain or noisy data (Luong and Manning, 2015; Belinkov and Bisk, 2018). Domain adaptation, which refers to the domain shift between the in-domain (same domain as the test set) and out-of-domain data, is one technique to address this problem. Kobus et al. (2017) and Chu et al. (2017) proposed a domain-adaptation technique for multi-domain NMT, which consists of inserting a domain tag into each source-side data entry to specify the domain of the corpus. Furthermore, Berard et al. (2019b) proposed two types of tags, a corpus tag and a type tag, also called a noise tag. The corpus tag indicates the domain of each corpus, as in Kobus et al. (2017). The noise tag indicates the type of noises, like back-translation noise or a lack of noise. Use of the noise-tag (i.e., back-translation tag) has been observed to be effective (Vaibhav et al., 2019; Caswell et al., 2019; Berard et al., 2019a).

For this paper, we implemented the existing domain-adaptation methods using tags as follows.

- A method with a single tag (Kobus et al., 2017) used the following domain tags:
 <CE> : for the content-equivalent corpus and the back-translated corpus (CE-NMT)
 <AA> : for the automatic-alignment corpus and the back-translated corpus (AA-NMT)
- A method with a single tag with back-translation (Vaibhav et al., 2019) used the following tags:
 <CE> : for the content-equivalent corpus
 <AA> : for the automatic-alignment corpus
 <BT-CE> : for the back-translated corpus (CE-NMT)
 <BT-AA> : for the back-translated corpus (AA-NMT)
- A method with two types of tags (Berard et al., 2019b) used the following tags:
 <BT> : for the back-translated corpora (noise tag)
 <CE> : for the content-equivalent corpus (corpus tag)
 <AA> : for the automatic-alignment corpus (corpus tag)

A tag was attached to the top of a source-side sentence. The first three rows in Table 4 list the attached tags for the three

³ One project member (c.f. Acknowledgments) applied an automatic sentence alignment method (Utiyama and Isahara, 2007) to the bilingual data in Table 2. Thus, we could fully leverage the results to make an automatic-alignment corpus.

existing domain-adaptation methods when implemented in the four corpora.

3.3. Problems of Existing Domain-Adaptation Methods Using Tag

The existing methods using tags were proposed to distinguish corpora with different features for training. An NMT model using tags can learn the particularities of each corpus such as writing styles that differ depending on the domain. Because the tag attached to the source-side sentence controls the output translation’s adaptation to the appropriate domain, it improves the translation quality. These methods cannot, however, completely distinguish the features of the training corpora because of the lack of information about the corpus and noise in the tags.

For information about the corpus, though these methods use only one kind of tag (<CE> and <AA>) to express the features of each corpus, as described in section 3.2, two kinds of tags are necessary in our case, because the features of the corpus are separated into the features of the source- and target-side data. In the case of the methods with two types of tags, we can see the lack of tag information as follows.

- Content-equivalent corpus and automatic-alignment corpus

The two corpora have the different corpus-tags (<CE> and <AA>), as listed in Table 4. Unfortunately, the source-side sentences of each corpus came from the same resource of original news.

- Content-equivalent corpus and back-translated corpus (CE-NMT)

The two corpora have the same corpus-tag (<CE>). Unfortunately, the target-side sentences of each corpus came from different resources. Specifically, the target-side sentences of the content-equivalent corpus came from the content-equivalent news, while those of the back-translated corpus (CE-NMT) came from the original news, as listed in Table 1.

For information about noise, though Berard et al. (2019b) proposed a noise tag, only one noise tag is attached to each data. The back-translated corpus (AA-NMT) needs two kinds of tags for each data, because it includes not only the back-translation noise but also nonequivalent noise, as listed in Table 1. Unfortunately, the existing method does not simultaneously express both kinds of noise.

3.4. Proposed Method

To solve the problems of the existing methods, we propose using multiple tags to express both the features of each corpus for the source- and the target-side sentences and the noise features, which consist of the nonequivalent noise and back-translation noise. This approach was inspired by Berard et al. (2019b) and Vaibhav et al. (2019). Thus, we implemented the proposed method, which uses four kinds of tags, as follows.

- Source-side tags:
<NS-S>: for corpora with source-side news sentences

<CE-S>: for corpora with source-side sentences back-translated with CE-NMT

- Target-side tags:
<CE-T>: for corpora with content-equivalent sentences on the target side
<NS-T>: for corpora with news sentences on the target side
- Back-translation noise tags:
<NO-BT> : for corpora without back-translation noise
<BT> : for corpora with back-translation noise
- Nonequivalent noise tags:
<NO-AN> : for corpora without nonequivalent noise
<AN> : for corpora with nonequivalent noise

The four kinds of tags are attached to the tops of the source-side sentences in the entire data. The fourth row in Table 4 lists the attached tags for the proposed method when implemented in the four corpora.

4. Experiment

In this study, we verified the effectiveness of both the content-equivalent corpus and the proposed domain-adaptation method using multiple tags. To show the effectiveness of the corpus, we trained multiple NMT models with different combinations of the four corpora, as listed in Table 4. To show the effectiveness of the proposed domain-adaptation method, we used both the existing methods and the proposed method to train NMT models with the four corpora.

4.1. Datasets and Setup

We used the parallel corpora described in section 3.1. Among these parallel corpora, we made a test set (size 2.0K) from the content-equivalent corpus because the automatic-alignment corpus includes noises, as shown in Figure 1. We used the remaining data in the content-equivalent corpus as training data. All of the datasets were preprocessed as follows. We used the Moses toolkit⁴ to clean and tokenize the English data and used KyTea (Neubig et al., 2011) to tokenize the Japanese data. Then, we used a vocabulary of 32K units based on a joint byte-pair encoding (BPE) (Sennrich et al., 2016b) for the source and target.

For the translation model, we used the encoder and decoder of the transformer model (Vaswani et al., 2017), which is a state-of-the-art NMT model. The transformer model uses a multi-headed attention mechanism, applied as self-attention, and a position-wise fully connected feed-forward network. The encoder converted the received source-language sentence into a sequence of continuous representations, and the decoder generated the target-language sentence. We implemented our systems with the Sockeye toolkit (Hieber et al., 2018) and trained them on an

⁴<https://github.com/moses-smt/mosesdecoder>

Table 5: Japanese→English translation results without domain-adaptation.

Training corpora	Data size	BLEU scores
Content-equivalent corpus	0.22M	20.93
Automatic-alignment corpus	0.29M	10.30
Content-equivalent corpus, automatic-alignment corpus	0.51M	20.68
Content-equivalent corpus, automatic-alignment corpus, back-translated corpora	1.57M	20.36

Table 6: Japanese→English translation results with domain-adaptation.

Domain-adaptation method	BLEU score
No tag	20.36
Single tag	22.41
Single tag with back-translation	24.19
Two types of tags	24.25
Proposed method	24.56

Nvidia P100 Tesla GPU. In training our models, we applied stochastic gradient descent (SGD) with Adam (Kingma and Ba, 2015) as the optimizer, using a learning rate of 0.0002, multiplied by 0.7 after every eight checkpoints. We set the batch size to 5000 tokens and the maximum sentence length to 99 BPE units. For the other hyperparameters of the models, we used the default Sockeye parameter values. We applied early stopping with a patience of 32. Decoding was performed through beam search with a beam size of 5, and we did not apply ensemble decoding with multiple models, although this could have improved the translation quality.

To evaluate the translation quality, we trained five models with different seeds, and we used the median BLEU score of the five translation results. We calculated case-sensitive BLEU (Papineni et al., 2002) scores by using `multi-bleu.perl`⁵. We used a statistical significance test with paired bootstrap resampling (Koehn, 2004), and the threshold was set to $p = 0.05$.

4.2. Results

4.2.1. Effectiveness of the Content-Equivalent Corpus

Table 5 summarizes the experimental results of NMT models with different combinations of the four corpora, as listed in the first column. The NMT model trained with only the content-equivalent corpus achieved a BLEU score of 20.93, which was the best score for Japanese→English news translation without using a domain-adaptation technique. Comparing the content-equivalent corpus and the automatic-alignment corpus, though the amount of data in the content-equivalent corpus was lower, the BLEU score of the NMT model trained with it was significantly improved. The results indicate that a clean corpus is a significant resource for improving translation quality. Furthermore, the translation quality was degraded by adding the other corpora

with noise, namely, the automatic-alignment corpus and the back-translated corpora.

4.2.2. Effectiveness of the Proposed Method

Next, Table 6 summarizes the experimental results of NMT models using the domain-adaptation method with tags. In this experiment, we used the whole training data.

First, we confirmed the effectiveness of the use of tags. Compared to the NMT model trained without tags, which had a BLEU score of 20.36, the four NMT models trained with tags achieved significantly higher BLEU scores.

Second, we compared the proposed method with the existing methods using tags. Use of a single tag had a BLEU scores of 22.41, while use of a single tag with back-translation had a score of 24.19. Meanwhile, the proposed method achieved a BLEU score of 24.56, which was significantly higher than those showed by the above mentioned two methods. Finally, use of two types of tags had a BLEU score of 24.25. The proposed method thus showed a slightly higher BLEU score, but the improvement was not statistically significant.

From Tables 4 and 6, we can recognize that the translation quality improves as the number of tags increases. This suggests that detailed corpus information contributes to the translation quality. The effect, however, depends on not only the number of tags but also the type of tag: the difference in BLEU scores between using a single tag and a single tag with back-translation was 1.78. In contrast, the difference between using a single tag with back-translation and using two types of tags was 0.06, and the difference between using two types of tags and the proposed method was 0.31. Investigation into tag information that contributes to translation quality is one of our future works.

5. Conclusion

In this work, we developed a Japanese-English news corpus, which was translated in a content-equivalent manner by humans. We described the corpus construction process and the different features between the proposed corpus and the other corpora. Furthermore, we proposed a domain-adaptation method using multiple tags to exploit our corpus and the other corpora. Then, we trained Japanese→English NMT systems with the proposed method and obtained quality improvement for Japanese→English news translation. For future work, we will apply other domain-adaptation methods such as fine tuning (Luong and Manning, 2015; Chu et al., 2017; van der Wees et al., 2017) and data selection (van der Wees et al., 2017; Wang et al., 2018) that can be implemented to the proposed method simultaneously.

⁵<https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu-detok.perl>

6. Acknowledgments

These research results were achieved through the project, “Research and Development of Deep Learning Technology for Advanced Multilingual Speech Translation,” as part of the Commissioned Research by the National Institute of Information and Communications Technology (NICT), Japan.

7. Bibliographical References

- Belinkov, Y. and Bisk, Y. (2018). Synthetic and natural noise both break neural machine translation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.
- Berard, A., Calapodescu, I., Dymetman, M., Roux, C., Meunier, J.-L., and Nikoulina, V. (2019a). Machine translation of restaurant reviews: New corpus for domain adaptation and robustness. In *Proceedings of the 3rd Workshop on Neural Generation and Translation*, pages 168–176, Hong Kong, November. Association for Computational Linguistics.
- Berard, A., Calapodescu, I., and Roux, C. (2019b). Naver labs Europe’s systems for the WMT19 machine translation robustness task. In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 526–532, Florence, Italy, August. Association for Computational Linguistics.
- Caswell, I., Chelba, C., and Grangier, D. (2019). Tagged back-translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 53–63, Florence, Italy, August. Association for Computational Linguistics.
- Chu, C., Dabre, R., and Kurohashi, S. (2017). An empirical comparison of domain adaptation methods for neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 385–391, Vancouver, Canada, July. Association for Computational Linguistics.
- Hieber, F., Domhan, T., Denkowski, M., Vilar, D., Sokolov, A., Clifton, A., and Post, M. (2018). The sockeye neural machine translation toolkit at AMTA 2018. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, pages 200–207, Boston, MA, March. Association for Machine Translation in the Americas.
- Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.
- Kobus, C., Crego, J., and Senellart, J. (2017). Domain control for neural machine translation. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 372–378, Varna, Bulgaria, September. INCOMA Ltd.
- Koehn, P. (2004). Statistical significance tests for machine translation evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain, July. Association for Computational Linguistics.
- Luong, M.-T. and Manning, C. D. (2015). Stanford neural machine translation systems for spoken language domains. In *Proceedings of the 12th International Workshop on Spoken Language Translation*, pages 76–79.
- Neubig, G., Nakata, Y., and Mori, S. (2011). Pointwise prediction for robust, adaptable japanese morphological analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 529–533, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.
- Sennrich, R., Haddow, B., and Birch, A. (2016a). Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany, August. Association for Computational Linguistics.
- Sennrich, R., Haddow, B., and Birch, A. (2016b). 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)*, pages 1715–1725, Berlin, Germany, August. Association for Computational Linguistics.
- Utiyama, M. and Isahara, H. (2007). A Japanese-English Patent Parallel Corpus. In *Proceedings of the MT Summit XI*, pages 474–482.
- Vaibhav, V., Singh, S., Stewart, C., and Neubig, G. (2019). Improving robustness of machine translation with synthetic noise. 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)*, pages 1916–1920, Minneapolis, Minnesota, June. Association for Computational Linguistics.
- van der Wees, M., Bisazza, A., and Monz, C. (2017). Dynamic data selection for neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1400–1410, Copenhagen, Denmark, September. Association for Computational Linguistics.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). Attention is all you need. In I. Guyon, et al., editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- Wang, R., Utiyama, M., Finch, A., Lemao, L., Kehai, C., and Sumita, E. (2018). Sentence selection and weighting for neural machine translation domain adaptation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.