

Neural Machine Translation Using Extracted Context Based on Deep Analysis for the Japanese–English Newswire Task at WAT 2020

Isao Goto[†] Hideya Mino[†] Hitoshi Ito[†] Kazutaka Kinugawa[†]
Ichiro Yamada[†] Hideki Tanaka[‡]

[†]NHK Science & Technology Research Laboratories

[‡]NHK Engineering System

{goto.i-es, mino.h-gg, itou.h-ce, kinugawa.k-jg,
yamada.i-hy}@nhk.or.jp
tanaka.hideki@nes.or.jp

Abstract

This paper describes the system of the NHK-NES team for the WAT 2020 Japanese–English newswire task. There are two main problems in Japanese–English news translation: translation of dropped subjects and compatibility between equivalent translations and English news-style outputs. We address these problems by extracting subjects from the context based on predicate-argument structures and using them as additional inputs, and constructing parallel Japanese–English news sentences equivalently translated from English news sentences. The evaluation results confirm the effectiveness of our context-utilization method.

1 Introduction

The NHK and NHK-ES team (NHK-NES) participated in the Japanese–English newswire translation task at WAT 2020 (Nakazawa et al., 2020). Currently, there are two main problems in the machine translation of news from Japanese to English.

One problem is that, in Japanese, some subjects that can easily be guessed tend to be omitted, because of the context. In contrast, in English, subjects are needed because grammatical rules do not allow them to be omitted. These phenomena are caused by differences between the characteristics of the two languages.

The other problem is the compatibility between equivalent translations and English news-style outputs. Parallel sentences extracted from Japanese news articles and English news articles are of low quality, in terms of bilingual equivalence, and contain substantial bilingual noise. This is because producing an English news article from the content of a Japanese news article is not sentence-level translation. Rather, it is

news writing, in which the structure of the article may be changed because of differences between the structures of news in Japanese and English, and the content is often summarized and supplemented because of differences in the background knowledge of the target readers. A neural machine translation (NMT) system trained on such data is supposed to output English sentences written in news style. However, the input articles tend not to be translated equivalently because omissions frequently occur unintentionally. At WAT 2019, to realize equivalent translation, we constructed parallel sentence pairs—Japanese news sentences that were equivalently translated to English by translators—and used these pairs for training (Mino et al., 2019). We confirmed that this method improved the performance of translating input sentences equivalently and that it was effective in terms of adequacy. However, the style of English produced by translators (“translated English”) was different from the style of English news (“news English”). The NMT system trained on these data output English sentences in the style of translated English, and the resulting outputs had low similarity (BLEU score) to the reference translations. We aim to output text in the style of news English rather than translated English. Therefore, our objective is to achieve both equivalent translation and output in the style of news English.

In this paper, we propose a method that analyzes the predicate-argument structures of the context of input sentences, extracts the topic in the context—based on the analyses—and uses it as the context of the input sentence to translate (§ 2). We show the difference between the translated-English and news-English styles, and construct Japanese–English parallel news sentences that are equivalently translated from English news to Japanese, rather than from Japanese

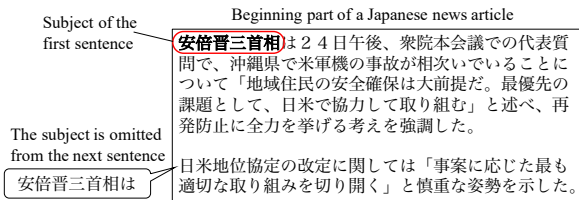


Figure 1: Example of a news sentence with an omitted subject in Japanese.

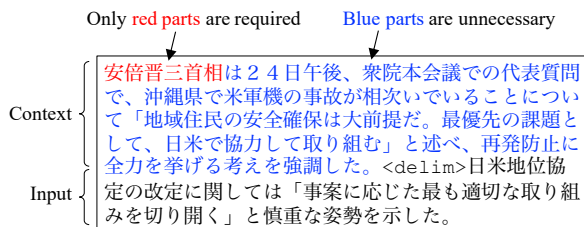


Figure 2: Context example of the 2-to-1 method. `<delim>` is used as the delimiter token.

news to English (§ 3). We provide a brief overview of the WAT 2020 Japanese-English newswire task (§ 4) and describe our system settings (§ 5). We then present the evaluation results and confirm the effectiveness of the proposed method (§ 6).

2 Using Extracted Context Based on Predicate-Argument Structure

Some subjects that can easily be guessed tend to be omitted in Japanese. In English, however, subjects cannot be omitted grammatically, except in the imperative form. For this reason, it is necessary to insert the implicit subjects when translating from Japanese to English. In news articles, the previous context often contains information that should be added. Figure 1 shows an example of a Japanese news sentence with an omitted subject, and its context.

Implicit subjects of sentences can often be easily guessed from the context, especially from topics and subjects in the context. Therefore, the preceding topics or subjects are thought to be useful as contextual information for translation.

An NMT method (2-to-1) uses context by concatenating the previous source sentence to the input sentence, with the insertion of a delimiter token (Tiedemann and Scherrer, 2017). However, as shown in Figure 2, a large amount of contextual information is input, and it is difficult to learn to identify and extract the necessary information by

backpropagation from the target sentence.¹

There are methods of complementing subjects of input sentences for machine translation by inferring zero subjects and pre-editing the input sentences (Taira et al., 2012; Russo et al., 2012; Kudo et al., 2014; Wang et al., 2016). However, these methods can also be problematic because pre-edit errors in the input sentences directly cause translation errors. Subject complement is not easy and is not sufficiently accurate. To alleviate the pre-edit error of input sentences, some methods have been proposed that use pre-edited source sentences as the targets of reconstruction (Wang et al., 2018a,b, 2019). However, these methods cannot solve the problem because choosing an output that reproduces incorrect pre-edited inputs degrades the output quality. Furthermore, in these methods, NMT—which is used to produce N-best hypotheses—does not use context. Therefore, NMT may not produce hypotheses that include appropriate contextual information.

We propose a method of extracting topics and subjects, which are sometimes useful as contextual information for translation, from context by analyzing syntactic and predicate-argument structure, and adding them to the input sentences as contextual information. This avoids the difficulty—in learning to find and extract the necessary information from the context—that arises when NMT is trained by backpropagation. NMT learns whether to utilize or ignore the extracted contextual information, and how to utilize it, using source and target language information. Because the proposed method does not pre-edit input sentences, it avoids the problem that the errors of subject completion are directly reflected in the translation when input sentences are pre-edited.

The proposed method is described in detail below. We use the following process to extract topics or subjects from context, up to the previous sentence, and add them as contextual information to the input sentence.

¹Extensions of the 2-to-1 method have also been proposed, in which the context used is increased or the neural network structure is changed (Wang et al., 2017; Voita et al., 2018; Müller et al., 2018; Kim et al., 2019; Scherrer et al., 2019; Maruf et al., 2019; Li et al., 2020b). However, from analysis of translation results, Kim et al. (2019) reported that 7.5% of the cases in which the translation edit rate (TER) was improved could be interpreted as utilizing document-level context and the other cases were mostly general improvements in adequacy or fluency that were not related to the given context. Li et al. (2020a) reported that the context affected training as a noise generator.

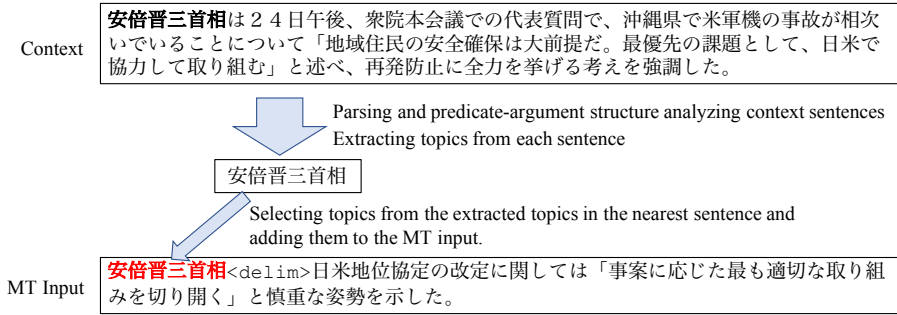


Figure 3: Proposed method: extraction of topic from context and addition to the input.

		Training	
		News	Translated
Test	News English	67.4	175.5
	Translated English	257.6	131.7

Table 1: Comparison between test set perplexities of news English and translated English.

1. The context before the input sentence(s) is parsed, and the predicate-argument structure is analyzed.
2. In each sentence of the context, the subject of the main clause is extracted. If it is not extracted, the subject of the sub-clause that depends on the main clause with the parallel relation is extracted. If it is still not extracted, the topic in the first *bunsetsu*² is extracted if the particle of the first *bunsetsu* is “*ha*”.
3. The topics and subjects that are extracted from the context sentence nearest to the input sentence are added to the input as the context information with the delimiter token.

An example of the process is shown in Figure 3.

The number of contextual sentences that can be used is not restricted. If any topic or subject is not extracted from the sentence immediately preceding the input sentence, we use a topic or subject extracted from an earlier sentence. For example, in the case that the sixth sentence of an article is an input sentence and any topic or subject is not extracted from the second to fifth sentences of the article, we use the topic or subject extracted from the first sentence.

²*Bunsetsu* is a Japanese chunk unit.

3 Equivalent Japanese-English News Parallel Sentences Translated from English News

At WAT 2019, we used 0.22 M pairs of Japanese news sentences and equivalently translated English sentences (translated English) (Mino et al., 2019). The evaluation confirmed that the translated sentences were highly adequate.

There were some differences in the characteristics (style) of the translated English and the English used in real news (news English). To examine the differences between the translated-English and news-English styles, we computed the test set perplexity of the two language models: one that was trained using translated-English sentences and one trained using news-English sentences. We used the CMU-Cambridge Statistical Language Modeling Toolkit v2³ with the settings of 4-gram and Good-Turing discounting. The models were trained with 6.0 M words of translated English and 7.3 M words of news English. A test set comprising 48 K words of translated English and a test set comprising 57 K words of news English were used for the test. The results are shown in Table 1.

The test set perplexity (175.5) of the news-English test data was worse than that (131.7) of the translated-English test data for the model trained on the translated-English data. Conversely, the test set perplexity (67.4) of the news-English test data was better than that (257.6) of the translated-English test data for the model trained on the news-English data. This inversion of these test set perplexities demonstrates that there is a difference between the translated-English style and the news-English style.

To obtain translations in news-English style, we constructed equivalently translated parallel news sentences whose English style was news En-

³<http://www.speech.cs.cmu.edu/SLM/toolkit.html>

Dataset type	Size
Training data	0.2 M sentence pairs
Development data with source context	479 sentence pairs
Test and reference data with source context	1851 sentence pairs
Source context data for development data	132 articles
Source context data for test data	546 articles

Table 2: Official datasets.

glish. The construction method was that translators rewrote the Japanese news sentences to equivalently match the corresponding English news sentences in Japanese–English news articles, such articles contain substantial noise in the bilingual relationship. We constructed the parallel news sentences using Jiji news (provided by Jiji Press). Our system used 98 K sentence pairs, constructed in this manner, for training. This strategy made it possible to learn how to perform equivalent Japanese–English news translation in the news-English style.

4 Translation Task with Test Sentence Context

In this section, we provide a brief overview of the Japanese–English newswire translation task. This task started at WAT 2017. Participants train the translation system on the training data and translate the test data. They then submit their translations to the workshop and the workshop organizers evaluate the submitted translations.

At WAT 2020, a new test set, which was extracted from Jiji news in Japanese and English, was added. The new test set consists of test sentences, their reference sentences, and context data of the test sentences. The context data are the articles from which the test sentences were extracted. Participants can use the context data to translate test sentences. The official dataset and their sizes are shown in Table 2. Training data other than that provided by the workshop can also be used for training.

5 System Setup

5.1 Data

The dataset used in the training are shown in Table 3.⁴ Here, all datasets, other than the official

⁴In this paper, a bilingual pair is referred to as a sentence pair. However, to be precise, each element in each language of a bilingual pair consists of one or more sentences.

training data, are external resources. As external resources, we used Jiji news in Japanese and English and Yomiuri news in Japanese and English. The English Jiji news sentences and their equivalent Japanese translations are described in Section 3. Other external resources are described in Mino et al. (2020). We increased the number of Japanese Jiji news sentences and their equivalent English translations, compared with those in Mino et al. (2020).

By inserting tags into the beginning of each source sentence, we controlled the equivalence and the English style in the output sentence. The <equivalent> and <noise> tags correspond to the feature of the amount of bilingual noise, and the <news.jiji>, <translation>, and <news.yomiuri> tags correspond to the feature of English style. The test sentences were tagged with “<equivalent> <news.jiji>” to specify equivalent translations and to output text in news-English style.

5.2 Tokenization

We used the Moses tokenizer and de-tokenizer for English, and jumanpp-2.0.0-rc3⁵ for Japanese. We used subword-nmt⁶ for subwording by the byte-pair-encoding method. The vocabulary size was 32 K for the concatenation of the source and target training sentences. We used knp-4.20⁷ for Japanese parsing and predicate-argument structure analysis.

5.3 NMT Configuration

We selected the Transformer model (Vaswani et al., 2017) for our NMT model, and used the sockeye-1.18.106 toolkit⁸. We used the following settings: max-seq-len was 200, weight-tying was trg_softmax, label_smoothing was 0.2, beam-size was 30, and four-ensemble used four models with different seeds. For the other hyperparameters, we used the default parameters of the toolkit.

5.4 Context Utilization

For the training data, we applied the context-utilization method described in Section 2 to the source side of “Japanese Jiji news sentences and their equivalent English translations” and “English Jiji news sentences and their equivalent Japanese

⁵<https://github.com/ku-nlp/jumanpp>

⁶<https://github.com/rsennrich/subword-nmt>

⁷<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

⁸<https://github.com/awslabs/sockeye>

Content	Sentence pairs	Bilingual noise	English characteristics	Tags added to input sentences
Official training data	0.2 M	many	news	<noise> <news_jiji>
JE parallel news sentences, automatically extracted from Japanese and English Jiji news wire articles	0.5 M	many	news	<noise> <news_jiji>
Japanese Jiji news sentences and their equivalent English translations	0.3 M	few	translation	<equivalent> <translation>
English Jiji news sentences and their equivalent Japanese translations	98 K	few	news	<equivalent> <news_jiji>
JE parallel news sentences, automatically extracted from Japanese and English Yomiuri news	0.59 M	many	news	<noise> <news_yomiuri>

Table 3: Our training data and tags for training.

	DataID	BLEU	RIBES	AMFM	JPO Adequacy
NMT _{with context}	3820	28.41	0.736924	0.62453	3.96
NMT _{without context}	3818	28.02	0.736564	0.61371	3.89
NMT _{official baseline}	3642	13.95	0.646680	0.47079	1.83

Table 4: Official scores.

Win	Loss	Tie
46	36	118

Table 5: Comparison of NMT_{with context} with NMT_{without context} based on the human evaluation.

translations” in Table 3. For the test data, we applied the context-utilization method to the context data and added the extracted contextual information to the test sentences.

6 Results and Discussion

We submitted two system outputs of the NMT system using context (NMT_{with context}) and the NMT system not using context (NMT_{without context}). Official results that were evaluated by the WAT 2020 organizers are shown in Table 4. For the JPO Adequacy evaluation, 200 sentences were evaluated by two human evaluators, based on the JPO Adequacy criterion. In Table 4, NMT_{official baseline} represents the baseline system, prepared by the organizers, which was trained on the official data. The BLEU scores and human scores of NMT_{without context} were higher than those of NMT_{official baseline}. This improvement is mainly due to the contribution of our additional training data. The BLEU scores and human scores of NMT_{with context} were higher than those of NMT_{without context}. This demonstrates the effectiveness of the context-utilization method.

Next, for the 200 sentences officially assessed by human evaluators, we compared the results with and without the use of context. For each out-

put sentence, two adequacy scores were awarded by each of two evaluators. We define win, loss, and tie by comparing the evaluators’ scores of NMT_{with context} to those of NMT_{without context}, as follows.

Win Both evaluators’ scores are better for NMT_{with context} than for NMT_{without context}, or one evaluator’s score is better and the other evaluator’s score is the same.

Loss Both evaluators’ scores are worse for NMT_{with context}, or one evaluator’s score is worse and the other evaluator’s score is the same.

Tie Any combination other than the above.

Table 5 shows the results. The number of wins was larger than that of losses. We checked the 46 “win” results, and found complements from contexts for nine input sentences could be used to translate the corresponding input sentences. For the other 37 input sentences, the complemented contexts were useless for translation. However, these useless complemented contexts did not directly degrade the translation quality. The number of remaining “win” (37) sentences was almost the same number of losses (36). We believe that, at this rate, the quality of the results depends not on the performance of the models, but on the variation of the model parameters.

Table 6 shows examples of inputs for NMT_{with context}, outputs of NMT_{with context},

Input (NMT _{with context})	<equivalent> <news_jiji> 首相 <delim> 3月末までに正式決定し、2019年秋の運用開始を目指す。
Reference	The government aims to finalize the policy by the end of March and start utilizing the funds in autumn 2019.
Output (NMT _{with context})	The government plans to formalize the plan by the end of March, aiming to put it into operation in autumn 2019.
Output (NMT _{without context})	The ministry plans to make an official decision by the end of March, aiming to put the system into operation in autumn 2019.
Input (NMT _{with context})	<equivalent> <news_jiji> 消防庁 <delim> 登録制メールのみの実施を予定している自治体には、ラジオのコミュニティー放送を併せて行うことなどを勧める。
Reference	The agency also plans to urge local governments that send only emails to registered residents to simultaneously air radio alerts, the sources said.
Output (NMT _{with context})	The agency will encourage local governments planning to use only registered emails to broadcast radio programs together with community broadcasts.
Output (NMT _{without context})	Local governments that plan to use only registered emails will be advised to broadcast radio programs together with community broadcasts.
Input (NMT _{with context})	<equivalent> <news_jiji> 金融庁 <delim> 仮想通貨取引所への立ち入り検査は初めて。
Reference	This is the first time for the FSA to conduct an inspection on a virtual currency exchange operator.
Output (NMT _{with context})	It was the first time for the FSA to conduct on-site inspections of a cryptocurrency exchange.
Output (NMT _{without context})	This is the first on-site inspection of a cryptocurrency exchange.
Input (NMT _{with context})	<equivalent> <news_jiji> 首相 <delim> トランプ氏は「関係者に協議させる」と応じた。
Reference	Trump replied that he will have relevant officials discuss the matter.
Output (NMT _{with context})	Trump replied that he will have relevant officials discuss the matter.
Output (NMT _{without context})	Trump replied that he will have related parties discuss the matter.

Table 6: Examples of inputs and outputs. The bold parts are thought to have been affected by the context added to the input. The first three examples show correct complements. The last example does not need a complement of subject.

	Rate (%)	(Number)
Sum: Subjects omitted	38.0	(57)
Correct complement	30.7	(46)
Incorrect complement	5.3	(8)
No complement	2.0	(3)
Sum: Subjects not omitted	62.0	(93)
Same complement	9.3	(14)
Different complement	48.7	(73)
No complement	4.0	(6)

Table 7: Analysis of subject complement for the second sentences of 150 Jiji news articles. “Same complement” means that the complement is the subject of the input sentence. “Different complement” means that the complement is different from the subject of the input sentence.

	DataID	BLEU
NMT _{with context}	3820	28.41
NMT _{without context}	3818	28.02
NMT _{2-to-1}	NA	28.22

Table 8: Comparison with 2-to-1 method.

and outputs of NMT_{without context}. Our complements are not restricted to pronouns. The first three examples require context to translate. The added contextual information affected the outputs of NMT_{with context}. In contrast, translation of the last input does not require context. In this case, the added contextual information seems to be

ignored in NMT_{with context}.

We evaluated the subject complement. In Jiji news in Japanese, the subject tends to be omitted in the second sentence of each article. Therefore, we examined the complements to the second sentences of each article. We used 150 articles from the context of the test data. Table 7 shows the results: Subjects were omitted for 38% of the sentences, whereas subjects were not omitted for the remaining 62% of the sentences. Of the 38% of sentences whose subjects were omitted, 30.7% of the subject complements were correct.

For comparison with the 2-to-1 method, we also conducted translation using that method. The results are shown in Table 8. The BLEU score of 2-to-1 was higher than that of NMT_{without context}, but lower than that of NMT_{with context}. The results confirmed that the proposed method was more effective in utilizing contextual information than the 2-to-1 method.

The submitted results are in news-English style. We believe that this is not the optimal style for adequacy because there is more training data for equivalent translations in translated English than for equivalent translations in news English. We added <equivalent> <translation> tags to the test

	DataID	Tag	BLEU	RIBES	AMFM	JPO Adequacy
NMT _{with context}	3820	<equivalent> <news_jiji>	28.41	0.736924	0.62453	3.96
NMT _{with context adequacy}	4175	<equivalent> <translation>	19.20	0.670908	0.59017	4.17

Table 9: Comparison with the translation results in the translated-English style.

sentences, instead of <equivalent> <news_jiji>, to evaluate the output in the translated-English style. We translated these test sentences using the NMT_{with context} model. NMT_{with context adequacy} represents the outputs in the translated English style. The results are shown in Tables 9. We submitted NMT_{with context adequacy} to the WAT automatic evaluation server and published it. The JPO adequacy in Table 9 was assessed by the same evaluators as the official evaluation, using the same criteria. However, the conditions were not exactly the same as that of the official evaluation because the evaluation was conducted independently after the official evaluation. Although the BLEU score of NMT_{with context adequacy} was lower than that of NMT_{with context}, the adequacy of NMT_{with context adequacy} was higher than that of NMT_{with context}. From the results, we confirmed that adequacy was improved by translating to the translated-English style, compared to translating to the news-English style, using the current training data.

7 Conclusion

We described our method using context, the resources used for training, and the system setup for the Japanese–English newswire translation task at WAT 2020. We proposed the extraction of topics and subjects from context, based on deep analysis, and their use as the context of input sentences. We also constructed a set of parallel Japanese–English news sentence pairs by equivalently translating English news sentences to Japanese sentences. This resource enabled us to directly learn equivalent translation with the news-English style. The evaluation results confirmed the effectiveness of our method.

Acknowledgments

The research results have been achieved by “Research and Development of Deep Learning Technology for Advanced Multilingual Speech Translation,” the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan. We thank Toshiaki

Nakazawa for introduction to related papers and for cooperation on the additional human evaluation.

References

- Yunsu Kim, Duc Thanh Tran, and Hermann Ney. 2019. [When and why is document-level context useful in neural machine translation?](#) In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 24–34, Hong Kong, China. Association for Computational Linguistics.
- Taku Kudo, Hiroshi Ichikawa, and Hideto Kazawa. 2014. [A joint inference of deep case analysis and zero subject generation for Japanese-to-English statistical machine translation.](#) In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 557–562, Baltimore, Maryland. Association for Computational Linguistics.
- Bei Li, Hui Liu, Ziyang Wang, Yufan Jiang, Tong Xiao, Jingbo Zhu, Tongran Liu, and changliang Li. 2020a. [Does multi-encoder help? a case study on context-aware neural machine translation.](#) In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3512–3518, Online. Association for Computational Linguistics.
- Xintong Li, Lemao Liu, Rui Wang, Guoping Huang, and Max Meng. 2020b. [Regularized context gates on transformer for machine translation.](#) In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8555–8562, Online. Association for Computational Linguistics.
- Sameen Maruf, André F. T. Martins, and Gholamreza Haffari. 2019. [Selective attention for context-aware 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)*, pages 3092–3102, Minneapolis, Minnesota. Association for Computational Linguistics.
- Hideya Mino, Hitoshi Ito, Isao Goto, Ichiro Yamada, Hideki Tanaka, and Takenobu Tokunaga. 2019. [Neural machine translation system using a content-equivalently translated parallel corpus for the newswire translation tasks at WAT 2019.](#) In *Proceedings of the 6th Workshop on Asian Translation*, pages 106–111, Hong Kong, China. Association for Computational Linguistics.

- Hideya Mino, Hideki Tanaka, Hitoshi Ito, Isao Goto, Ichiro Yamada, and Takenobu Tokunaga. 2020. [Content-equivalent translated parallel news corpus and extension of domain adaptation for NMT](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3616–3622, Marseille, France. European Language Resources Association.
- Mathias Müller, Annette Rios, Elena Voita, and Rico Sennrich. 2018. [A large-scale test set for the evaluation of context-aware pronoun translation in neural machine translation](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 61–72, Brussels, Belgium. Association for Computational Linguistics.
- Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar, and Sadao Kurohashi. 2020. Overview of the 7th workshop on Asian translation. In *Proceedings of the 7th Workshop on Asian Translation*, Suzhou, China. Association for Computational Linguistics.
- Lorenza Russo, Sharid Loáiciga, and Asheesh Gulati. 2012. [Improving machine translation of null subjects in Italian and Spanish](#). In *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 81–89, Avignon, France. Association for Computational Linguistics.
- Yves Scherrer, Jörg Tiedemann, and Sharid Loáiciga. 2019. [Analysing concatenation approaches to document-level NMT in two different domains](#). In *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, pages 51–61, Hong Kong, China. Association for Computational Linguistics.
- Hiroto Taira, Katsuhito Sudoh, and Masaaki Nagata. 2012. [Zero pronoun resolution can improve the quality of J-E translation](#). In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pages 111–118, Jeju, Republic of Korea. Association for Computational Linguistics.
- Jörg Tiedemann and Yves Scherrer. 2017. [Neural machine translation with extended context](#). In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pages 82–92, Copenhagen, Denmark. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc.
- Elena Voita, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. 2018. [Context-aware neural machine translation learns anaphora resolution](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1264–1274, Melbourne, Australia. Association for Computational Linguistics.
- Longyue Wang, Zhaopeng Tu, Shuming Shi, Tong Zhang, Yvette Graham, and Qun Liu. 2018a. [Translating pro-drop languages with reconstruction models](#). In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 4937–4945. AAAI Press.
- Longyue Wang, Zhaopeng Tu, Xing Wang, and Shuming Shi. 2019. [One model to learn both: Zero pronoun prediction and translation](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 921–930, Hong Kong, China. Association for Computational Linguistics.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. 2017. [Exploiting cross-sentence context for neural machine translation](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2826–2831, Copenhagen, Denmark. Association for Computational Linguistics.
- Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu. 2018b. [Learning to jointly translate and predict dropped pronouns with a shared reconstruction mechanism](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2997–3002, Brussels, Belgium. Association for Computational Linguistics.
- Longyue Wang, Zhaopeng Tu, Xiaojun Zhang, Hang Li, Andy Way, and Qun Liu. 2016. [A novel approach to dropped pronoun translation](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 983–993, San Diego, California. Association for Computational Linguistics.