Translation from Historical to Contemporary Japanese Using Japanese T5

Hisao Usui and **Kanako Komiya**
Tokyo University of Agriculture and Technology
h-usui@st.go.tuat.ac.jp
kkomiya@go.tuat.ac.jp

Abstract

This paper presents machine translation from historical Japanese to contemporary Japanese using a Text-to-Text Transfer Transformer (T5). The result of the previous study that used neural machine translation (NMT), Long Short Term Memory (LSTM), could not outperform that of the work that used statistical machine translation (SMT). Because an NMT model tends to require more training data than an SMT model, the lack of parallel data of historical and contemporary Japanese could be the reason. Therefore, we used Japanese T5, a kind of large language model to compensate for the lack of data. Our experiments show that the translation with T5 is slightly lower than SMT. In addition, we added the title of the literature book from which the example sentence was extracted at the beginning of the input. Japanese historical corpus consists of a variety of texts ranging in periods when the texts were written and the writing styles. Therefore, we expected that the title gives information about the period and style, to the translation model. Additional experiments revealed that, with title information, the translation from historical Japanese to contemporary Japanese with T5 surpassed that with SMT.

1 Introduction

This paper develops a translation system from historical Japanese to contemporary Japanese. In Japan, there are a large number of historical literature including untranslated ones. Although the experts are able to translate them, it is not only difficult but also time-consuming to translate them manually. Therefore, translation systems of historical Japanese are necessary.

To our knowledge, two works have been done for translation from historical Japanese to contemporary Japanese (see Section 2). Hoshino et al. (2014) used statistical machine translation (SMT), which presented a translation system whose BLEU score (Papineni et al., 2002) was 28.02, the highest in the previous study. Takaku et al. (2020) used neural machine translation (NMT), Long Short Term Memory (LSTM), and the BLEU score of their system was 19.95. We believe that the lack of translation data is why the NMT, specifically, translation with LSTM, could not outperform the SMT. An NMT model tends to require more training data than an SMT model does but the parallel data of historical and contemporary Japanese are limited; they are 86,684 sentence pairs. Therefore, we used Text-to-Text Transfer Transformer (T5) (Raffel et al., 2020), which is a pre-trained large language model (LLM), to compensate for the lack of data. After the release of the GPT (Radford et al., 2018) and the BERT (Devlin et al., 2019) in 2018, LLMs have achieved state-of-the-art results in various tasks of natural language processing. We hypothesise that the poor performance due to the limited training data could be complemented by the use of large pre-trained models (see Section 3).

In addition, we focused on the diversity of the historical texts. The Japanese historical corpus consists of a variety of texts ranging in periods when the texts were written and the writing styles. For example, The Complete Collection of Japanese Classical Literature published by Shogakukan(SHOGAKUKAN, 2010) [1] contains older texts written in the 800s and relatively new texts written in the 1800s. In addition, the collection consists of texts of various styles, including diaries, poems, folk stories, and so on (see Section 4). Therefore, it is not surprising if the diversity of the texts gives a negative influence on the performance of the translation system.

Li et al. (2016) reported that giving speakers' IDs to persona-based dialogue generation improves the dialogue quality when the training data consists of data from multiple speakers. We believe that this is because the information of the speakers alleviates the negative influence of the diversity of the train-

[1] https://japanknowledge.com/en/contents/koten/

Proceedings of the Joint 3rd NLP4DH and 8th IWCLUL, pages 27–35
December 1-3, 2023 ©2023 Association for Computational Linguistics

ing data. Instead of speakers' IDs, we utilized the titles of the literature book from which the example sentences were extracted.

We compared our methods with the previous studies using BLEU and BERT scores and analysed the quality of the translations (see Sections 5, 6 and 7).

The contributions of this paper are as follows:

1. We developed a translation system from historical Japanese to contemporary Japanese using T5;
2. We proposed giving the title information of the literature book from which the example sentence was extracted to the translation model;
3. The BLEU score of our system outperformed those of the previous research; and
4. We discussed the translation quality using some examples.

2 Related Work

Research on machine translation from historical Japanese into contemporary Japanese has been reported by Hoshino et al. (2014) who used SMT and Takaku et al. (2020) who used LSTM. The BLEU scores were 28.02 for SMT and 19.95 for LSTM. We used T5 to complement the lack of parallel data that is required for the training of the model.

An example of a previous study of translation using T5 is work reported by Emezue and Dossou (2021). They used multilingual T5 (mt5) to translate African languages such as Kosa, Yoruba, and Igbo into English and French. The number of data on these African languages is limited; Kosa, the largest corpus has 158,660 monolingual sentences and 137,000 parallel data with English. They reported that the BLEU score of the Kosa-English translation was 30.25. Their research has shown that T5 is effective for low-resource language translation. Agarwal et al. (2020) used T5 for machine translation to aid bilingual data-to-text generation and semantic parsing. They showed the machine translation using T5 improved performances of generation and parsing tasks.

We gave the title of the literature book from which the example sentence was extracted to the translation model. This is inspired by Li et al. (2016), who gave speakers' IDs to the persona-based dialogue generation system. The dialogue quality improved when the training data consisted of data from multiple speakers; the utterances output from the system became more speaker-specific. We expected that the information of the speakers would alleviate the negative influence of the diversity of the training data. Instead of speakers' IDs, we input the titles of the literature book to alleviate the negative influence of the diversity of the training sentences in period and style. In addition, Caswell et al. (2019) added an extra token to back-translation data for noising techniques [2]. They inform the model which data is back-translated to alleviate the noise's effect. Instead, we added the title of the literature book to inform the model the period and writing style information.

In addition, there is research on translation related to digital humanities such as (Gupta, 2022), (Zheng et al., 2022), and (Piper and Erlin, 2022).

3 Translation Using T5

T5 is a pre-trained model based on Transformer (Vaswani et al., 2017) trained with Colossal Clean Crawled Corpus (C4), which is a cleaned version of Common Crawl's web crawl corpus. Using such high-quality data and careful tuning of the model, T5 achieved state-of-the-art performance on 26 various tasks in 2019.

In this paper, we fine-tuned pre-trained Japanese T5 to translate historical Japanese into contemporary Japanese. We aim to compensate for the lack of data using a pre-trained LLM because the amount of parallel corpus of historical and contemporary Japanese is limited. We used sonoisa/t5-base-japanese[3] for Japanese T5. This model was trained with Japanese Wikipedia dump data, Japanese OSCAR corpus, and Japanese CC-100: Monolingual Datasets from Web Crawl Data. Although it is mostly trained with contemporary Japanese, we used it for transformation from historical Japanese to contemporary Japanese. Therefore, this method could be deemed as a diachronic domain adaptation. We employed this method because Komiya et al. (2022) reported that the performance of Japanese word sense disambiguation of historical Japanese significantly improved when they used diachronic domain adaptation using Japanese BERT trained with contemporary

[2]We attempted the back translation for this research but it did not work. We think that this is because the data we used, Aozora-bunko, is not enough similar to the original corpus of historical Japanese

[3]https://huggingface.co/sonoisa/t5-base-japanese

Parallel data	
Historical sentence Contemporary sentence	而るに、既に講の終る日に成て、道俗男女員不知ず参り集たり。 さて、いよいよ講の終る日になると、僧俗・男女を問わず、 数知れぬほどの人々が参詣してきた。
Title-added parallel data	
Historical sentence Contemporary sentence	今昔物語集**(Konjaku)02:**而るに、既に講の終る日に成て、道俗男女員不知ず参り集たり。 さて、いよいよ講の終る日になると、僧俗・男女を問わず、 数知れぬほどの人々が参詣してきた。

Table 1: Example of parallel data and title-added data. The only difference between the parallel data and the title-added data is the prefix of the historical sentence: the title and colon(:).

Japanese.

In addition, we gave the titles of the literature books from which the example sentences were extracted, to the translation model: the fine-tuned T5. We set a title and colon as a prefix of the historical Japanese sentence. Table 1 shows an example of the input: the prefix and text. The titles give information on the period and writing style of the original sentence to the model. The meanings of words sometimes vary depending on the writing style of the texts and the period when the texts were written. Therefore, we believe that this kind of information is useful for translation. It helps the model appropriately translate texts from various literature books written in a wide range of periods and in various writing styles. We were inspired by Li et al. (2016), who gave speakers' IDs to the persona-based dialogue generation system.

4 Data

We used a parallel corpus extracted from The Complete Collection of Japanese Classical Literature published by Shogakukan by Hoshino et al. (2014). This corpus consists of historical Japanese sentences paired with manually translated contemporary Japanese sentences. Table 2 shows the statistics of the parallel data of historical and contemporary Japanese. This is a diachronic corpus consisting of works from various periods.

We used 50 pieces of literature listed in Tables 3 and 4. However, they are 60 books because some literature has more than one volume and Hōgen monogatari and Heiji monogatari were compiled together in one book [4].

The corpus contains a total of 86,684 sentence pairs, some of which have multiple contemporary Japanese translations for a single historical Japanese. In the corpus, 221 historical Japanese data and 163 contemporary Japanese data were duplicated. Notably, these different translations of the same historical Japanese could decrease the BLEU score when they are split into the training and test data. The data was split into (training: development: test) = (82,591: 2,000: 2,093) sentence pairs, following (Takaku et al., 2020) To avoid data bias, the data were randomly split.

Historical Japanese	
Total Number of Sentences	86,684
Vocabulary Size	49,200
Number of Tokens	2,774,745
Contemporary Japanese	
Total Number of Sentences	86,684
Vocabulary Size	45,690
Number of Tokens	3,611,783

Table 2: Parallel corpus of historical and contemporary Japanese. The data for translation are extracted from the Complete Collection of Japanese Classical Literature.

We made another data where the titles of this sentence were added to the head of historical Japanese from The Complete Collection of Japanese Classical Literature. We added the title and the mark colon (:) to the head of every historical sentence. Table 1 shows examples from both data. We appended the title to the beginning of each sentence pair. The title of the literature book is only added to the beginning of the historical Japanese and not to the contemporary Japanese.

As the parallel corpus did not have the title information, we searched the sentences of the source language, i.e., historical Japanese, in The Complete Collection of Japanese Classical Literature, the original dataset, and automatically identified it. In this process, a few errors happened especially for

[4] Chikamatsu-Monzaemon-shū has three volumes, Genji monogatari has six volumes, Heike monogatari has two volumes, and Kokin Wakashū has two volumes.

Titles	Year	Style
Chikamatsu-Monzaemon-shū (Collection of Chikamatsu-Monzaemon's Stories)	1703-1720	Playbook
Genji monogatari (The Tale of Genji)	Before 1001	Tale
Gikeiki (The Chronicle of Yoshitsune)	Before 1400	War chronicle
Heichū Monogatari (Tales of Heichū)	Around 1000	Poem tales
Heiji monogatari (The Tale of Heiji)	Mid-13th century	War chronicle
Heike monogatari (The Tale of the Heike)	Before 1309	Epic account
Hōgen monogatari (The Tale of Hōgen)	Around 1220	War chronicle
Hōjōki (Square-jō record)	1212	Essay
Imose Yamaonna Teikin	1771	Playbook
Ise monogatari (The Tales of Ise)	Around 900?	Poem tales
Izumi Shikibu Nikki (The Diary of Lady Izumi)	1008	Poetic diary
Jikkinsyo	1252?	Folktales
Kagerō Nikki (The Mayfly Diary)	974	Poetic diary
Kaikou	1764	Playbook
Kanameishi	1663	Folktales
Kindaishōka	1209	Essay on poetry
Kokin Wakashū (Collection of Japanese Poems Of Ancient and Modern Times)	905	Anthology of the poetry
Kokin Wakashū Appendix	905	Anthology of the poetry
Kokka-Hachiron	1742	Essay on poetry
Konjaku Monogatarishū (Anthology of Tales from the Past)	Around 1120	Folktales
Koraifūteishō	1197	Essay on poetry
Maigetsu-shō	1219	Essay on poetry
Makura no Sōshi (The Pillow Book)	1001	Essay
Murasaki Shikibu Nikki (The Diary of Lady Murasaki)	1010	Diary
Matsuranomiya Monogatari (The Tale of the Matsura Palace)	1201	Tale
Mutsuwaki (Chronicle of Mutsu)	Late 11th century	War chronicle
Nihon Ryōiki (Miraculous Stories from The Japanese Buddhist Tradition)	822	Folktales
Nii-Manabi-Iken	1811	Essay on poetry
Ochikubo Monogatari (The Tale of Ochikubo)	Around 1000	Tale
Ōkagami (The Great Mirror)	1119	History book
Otogi Monogatari	1678	Folktales
Sanukinosuke Nikki (The Diary of Sanukinosuke)	1109	Diary
Sarashina Nikki (The Sarashina Diary)	1020	Nonfiction narrative
Shasekishū (Sand and Pebbles)	1283	Buddhist text
Shōbōgenzō Zuimonki (The Treasury of the True Dharma Eye: Record of Things Heard)	1235-1238	Dharma talks
Shōmonki, Masakadoki (Chronicle of Masakado)	Before 1099	War chronicle
Soga Monogatari (The Revenge of the Soga brothers)	Before 1285	War chronicle
Taketori Monogatari (The Tale of the Bamboo Cutter)	Around 900	Fictional prose narrative
Takitsuke Moekui Keshizumi	1677	Folktales

Table 3: List of the Literature Books 1

Titles	Year	Style
Tannishō (Lamentations of Divergences)	Around 1300	Buddhist text
Tosa Nikki (Tosa Diary)	934	Poetic diary
Toshiyori Zuinō	1113	Essay on poetry
Tsurezuregusa (Essays in Idleness)	1332	Essay
Tsutsumi Chūnagon Monogatari (Tales of the Riverside Middle Counselor)	Before 1271	Short stories
Tsutsumi Chūnagon Monogatari Appendix	Before 1271	Short stories
Uji Shūi Monogatari	1221	Short stories
Ukiyo Monogatari	1661	Folktales
Utsubo Monogatari (The Tale of the Hollow Tree)	Around 1000	Tale
Yamato Monogatari (Tales of Yamato)	9th - 10th century	Short stories
Yōkyoku-shū	Before 1573	Playbook

Table 4: List of the Literature Books 2

short sentences, because sometimes the same short sentences appeared in multiple pieces of literature. However, we used them without any corrections.

5 Experiments

We conducted two kinds of experiments: (1) the experiment using the parallel data and (2) the experiment using the title-added parallel data.

We used sonoisa/t5-base-japanese and T5 tokenizer in the hugging face library for Japanese T5 and its tokenizer. We conducted a grid search for the learning rate, epoch number, and repetition penalty in both experiments as shown in Table 5. We set the hyperparameters of T5 as shown in Table 6. The other parameters are set to the default values. Because sonoisa/t5-base-japanese model does not need lots of memory, we trained the models by using only an RTX3090. It took 2.5 hours per epoch.

We measured BLEU scores and the similarities between translations using Sentence-BERT (Reimers and Gurevych, 2019) for evaluation. For the calculation of BLEU score, we used Sacreblue (Post, 2018) [5]. We calculated the similarities between translations for reference because BLEU scores do not directly reflect the meanings of the sentences. We used sonoisa/sentence-bert-base-ja-mean-tokens-v2 to calculate the similarities[6]. We used default settings for the hyperparameters of the Sentence-BERT.

[5]https://huggingface.co/spaces/evaluate-metric/sacrebleu

[6]https://huggingface.co/sonoisa/sentence-bert-base-ja-mean-tokens-v2

6 Results

6.1 BLEU scores

Table 7 shows the best BLEU scores of the experiments using the parallel and title-added parallel data and their hyperparameters. It shows that the BLEU score of NMT with simple T5 using only the parallel data (27.50) is better than that of LSTM (19.95) (Takaku et al., 2020) and slightly lower than that of SMT (28.02) (Hoshino et al., 2014). It also shows that the BLEU score is better when the title-added parallel data was used than when the parallel data without the title was used. The BLEU score of NMT with T5 using the title-added parallel data (28.67) outperformed those of LSTM and SMT.

6.2 Similarities Using Sentence-BERT

Table 8 shows the average similarities between the translation of the systems and the reference translation calculated using Sentence-BERT. They were averaged over all the test examples. The BLEU scores are also shown in the table as a reference.

As shown in Table 8, the average similarity between the translation by the system trained with the parallel data and the reference translation was 0.787 and that between the translation by the system trained with the title-added parallel data and the reference translation was 0.784. The difference was 0.004 points, which is not significant. However, the system trained with the title-added parallel data outperformed the system trained with parallel data again.

Learning rate	0.0001, 0.0002, 0.0003, 0.0004, 0.0005, 0.0007, 0.001
Epoch number	1, 5, 10
Repetition penalty	1, 1.5, 2.0

Table 5: The hyperparameters of the experiments

max_seq_length	512
weight_decay	0.0
adam_epsilon	1e-8
warmup_steps	100
batch_size	128
gradient_accumulation_steps	4
n_gpu	1
early_stop_callback	False
fp_16	False
opt_level	01
max_grad_norm	1.0

Table 6: Hyperparameters of T5

7 Translation Analysis

Table 9 shows the translation examples of the experiments using the parallel and title-added parallel data. The first example is from Heiji Monogatari (The Tale of Heiji) and the second example is from Genji Monogatari (The Tale of Genji) [7]. In these two examples, we can see the reference translations are idiomatic, rather than literal translation.

For instance, in the first example, the original sentence does not contain the subject, who brought Tokiha (a woman) to someone who ordered someone to do so. Because omission of the subject often occurs in Japanese, this is a natural Japanese sentence. Nevertheless, the reference translation includes the subject; it says that Ito Musha, a specific warrior, brought her to someone who ordered him to do so. The translations of our systems do not include the subject, like the original sentence. In addition, the verb in the original sentence, "宣ふ", the basis form of "宣へ," literally means "say" rather than "order." However, the reference translation uses the verb "order" instead of "say" in this example, whereas the translations of the systems use the word "say." We think that this is probably because "order" is easier to understand than "say" in this context. However, it is an idiomatic translation again.

In addition, in the second example, the original sentence says "ここなる人々" literally means "people staying here." The translations of our systems are correct if the sentence was translated literally. However, the reference translation rephrased the word to "my sons." This is an additional explanation by a translator. Therefore, we believe that the BLEU scores of our systems tend to decrease because the reference translations are idiomatic. To evaluate these systems, literal translations are necessary.

Moreover, some Japanese verbs totally change the word when an honorific form is used. It is like the past tense of irregular verbs in English. In addition, some Japanese verbs have some variant honorific forms.

In the first example, two verbs, which mean bring someone and order, are written in normal form in the reference translation. However, when the T5 translation with the parallel data was used for the experiments, two verbs, which means bring someone and say, were written in honorific form. When the title-added parallel data was used, a verb that means "say" was written in another honorific form and the other two verbs, which mean bring someone and follow, were written in normal form.

In the second example, "おはせましかば," which means "were still alive" with old honorific form in the original sentence was translated into "生きていらっしゃったら," in the reference translation, which means the same but with contemporary honorific form. However, it was translated into "ご存命でいらっしゃったら" using our systems instead of "生きていらっしゃったら." "ご存命でいらっしゃったら" means "were still alive" again, and this is another expression that means the same with a different contemporary honorific form. In other words, these two expressions mean the same, however, the ways of honorification are different. These kinds of variations of expressions also tend to decrease the BLEU scores. Because we do not have the translation examples of the other systems, we cannot compare the results directly [8]. However, we believe that NMT tends to

[7]The title we used was Genji05, because we had 6 volumes for Genji Monogatari and it was the fifth one.

[8]We could not find any translation examples as same as (Takaku et al., 2020) because the dataset was the same but the

	Learning rate	Epoch number	Repetition penalty	BLEU
The parallel data	0.0002	10	1.0	27.50
The title-added parallel data	0.0005	5	1.5	28.67
LSTM (Takaku et al., 2020)				19.95
SMT (Hoshino et al., 2014)				28.02

Table 7: The best results of experiments using the parallel and the title-added parallel data

Data	Sentence-BERT	BLEU
Parallel data	0.784	27.49
Title-added data	0.787	28.57

Table 8: The average similarities of the translation of systems and the reference translation calculated using sentence-BERT

translate the texts with more variations than SMT does because SMT uses a dictionary. Therefore, these kinds of variations of expression probably happened more when NMT was used.

8 Conclusion and Future Work

In this paper, we translated historical Japanese into contemporary Japanese by fine-tuning Japanese T5, which is a large-scale pre-trained model. We expected that T5 would be useful for the translation because it compensates for the lack of parallel data. In addition, we proposed to add a book title from which the source sentence was extracted at the beginning of the input sentences. Because CHJ, the historical text corpus we used, ranges in periods when the texts were written and styles, we expected that the title of the books could give those kinds of information to the translation systems. The experiments revealed that T5 and giving the title of the books are effective for the translation; BLEU scores and the Sentence-BERT similarities of our system outperformed those of previous studies, which used SMT and LSTM, respectively. According to the analysis of the translation examples, we observed that the translations of our systems are more literal than the reference translations. However, the BLEU scores tend to decrease because the reference translations are idiomatic, rather than literal translation. The evaluation using the literal translations by human annotators is our future work.

data split was different. (Hoshino et al., 2014) did not contain the translation examples.

In addition, further experiments and analyses should be done to investigate the effectiveness of our method. For example, comparison to the setting where random titles were used and the setting where only periods and genres are given should be explored. Moreover, we are planning to compare our method to Large Language Models, such as chatGPT.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 22K12145, and collaborative research projects of National Institute for Japanese Language and Linguistics. We would like to thank Hikaru Yokono, who gave us helpful advice and the data set used in previous research.

References

Oshin Agarwal, Mihir Kale, Heming Ge, Siamak Shakeri, and Rami Al-Rfou. 2020. Machine translation aided bilingual data-to-text generation and semantic parsing. In *Proceedings of the 3rd International Workshop on Natural Language Generation from the Semantic Web (WebNLG+)*, pages 125–130, Dublin, Ireland (Virtual). Association for Computational Linguistics.

Isaac Caswell, Ciprian Chelba, and David Grangier. 2019. Tagged back-translation. In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 53–63, Florence, Italy. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Chris Chinenye Emezue and Bonaventure F. P. Dossou. 2021. MMTAfrica: Multilingual machine translation for African languages. In *Proceedings of the Sixth Conference on Machine Translation*, pages 398–411, Online. Association for Computational Linguistics.

Kshitij Gupta. 2022. MALM: Mixing augmented language modeling for zero-shot machine translation. In *Proceedings of the 2nd International Workshop*

	Sentence	BLEU	SBERT
Original English Translation	具して参れ」と宣へば、常葉を具して参りけり。 (Someone,the subject in omitted here) brought Tokiha (This is a woman's name.) to him because (he, the subject is omitted here.) said "Bring her to me."		
Reference English Translation	連れて来い」と命ずるので、伊藤武者は常葉を連れてきた。 Ito Musha (This is a warrior's name.) brought Tokiha to him because (he, the subject is omitted here.) ordered as "Bring her to me."		
The parallel data English Translation	連れて参れ」と言われるので、常葉を連れて参った。 (He, the subject is omitted here.) brought Tokiha to him because (he, the subject is omitted here again.) said "Bring her to me."	21.98	0.81
The title-added parallel data English Translation	ついて来い」とおっしゃるので、常葉を連れて行った。 (He, the subject is omitted here.) brought Tokiha to him because (he, the subject is omitted here again.) said "Follow me."	32.04	0.81

	Sentence	BLEU	SBERT
Original English Translation	故殿おはせましかば、ここなる人々も、 かかるすさびごとにぞ、心は乱らまし」とうち泣きたまふ。 (The subject is omitted here.) cried saying " If the late Lord were still alive, I'm sure even the people staying here would have had to worry their heads about these whimsical play things."		
Reference English Translation	故殿が生きていらっしゃったら、こちらの息子たち だってこうした気まぐれな遊び事で頭を 悩ませていたことでしょうに」とお泣きになる。 (The subject is omitted here.) cried saying " If the late Lord were still alive, I'm sure even my sons here would have had to worry their heads about these whimsical play things."		
The parallel data English Translation	亡き殿がご存命でいらっしゃったら、ここにいる 人々も、こうした遊び事に心を 奪われていたでしょうに」とお泣きになる。 (The subject is omitted here.) cried saying " If the late Lord were still alive, I'm sure even the people staying here would have been fascinated by these play things."	35.16	0.83
The title-added parallel data English Translation	亡き殿がご存命でいらっしゃったら、ここにいる 人たちも、こうした慰みごとのために心が乱れ たことでしょう」とお泣きになる。 (The subject is omitted here.) cried saying " If the late Lord were still alive, I'm sure even the people staying here would have been disturbed by these comforts."	28.26	0.74

Table 9: Translation examples using the parallel and title-added parallel data

on Natural Language Processing for Digital Humanities, pages 53–58, Taipei, Taiwan. Association for Computational Linguistics.

Sho Hoshino, Yusuke Miyao, Shunsuke Ohashi, Akiko Aizawa, and Hikaru Yokono. 2014. Machine translation from historical Japanese to contemporary Japanese using parallel corpus. In *Proceedings of the NLP2014, (In Japanese)*, pages 816–819.

Kanako Komiya, Nagi Oki, and Masayuki Asahara. 2022. Word sense disambiguation of corpus of historical Japanese using Japanese BERT trained with contemporary texts. In *Proceedings of the 36th Pacific Asia Conference on Language, Information and Computation*, pages 438–446, Manila, Philippines. De La Salle University.

Jiwei Li, Michel Galley, Chris Brockett, Georgios Spithourakis, Jianfeng Gao, and Bill Dolan. 2016. A persona-based neural conversation model. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 994–1003, Berlin, Germany. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 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. Association for Computational Linguistics.

Andrew Piper and Matt Erlin. 2022. The predictability of literary translation. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 155–160, Taipei, Taiwan. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving Language Understanding by Generative Pre-Training.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

SHOGAKUKAN. 2010. *The Complete Collection of Japanese Classical Literature (New Edition)*. SHOGAKUKAN.

Masashi Takaku, Tosho Hirasawa, Mamoru Komachi, and Kanako Komiya. 2020. Neural machine translation from historical Japanese to contemporary Japanese using diachronically domain-adapted word embeddings. In *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, pages 534–541, Hanoi, Vietnam. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need.

Francis Zheng, Edison Marrese-Taylor, and Yutaka Matsuo. 2022. A parallel corpus and dictionary for Amis-Mandarin translation. In *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities*, pages 79–84, Taipei, Taiwan. Association for Computational Linguistics.