Controlling Japanese Honorifics in English-to-Japanese Neural Machine Translation

Weston Feely* Eva Hasler† Adrià de Gispert†
SDL Research
*Los Angeles, CA, †Cambridge, UK
{wfeely,ehasler,agispert}@sdl.com

Abstract

In the Japanese language different levels of honorific speech are used to convey respect, deference, humility, formality and social distance. In this paper, we present a method for controlling the level of formality of Japanese output in English-to-Japanese neural machine translation (NMT). By using heuristics to identify honorific verb forms, we classify Japanese sentences as being one of three levels of informal, polite, or formal speech in parallel text. The English source side is marked with a feature that identifies the level of honorific speech present in the Japanese target side. We use this parallel text to train an English-Japanese NMT model capable of producing Japanese translations in different honorific speech styles for the same English input sentence.

1 Introduction

Languages differ in the way they express the same ideas depending on social context. In English different words or phrases are used in a more casual or familiar context compared to a more formal context. In languages such as Japanese or Korean formality distinctions are grammatically encoded using a system of honorifics. These honorifics are part of Japanese verbal morphology, which allows the same concept to be expressed in multiple levels of formality by altering the inflection of the main verb of the sentence. The examples in Table 1 show one sentence in three different levels of formality. In all three examples the meaning is the same, but the inflection of the main verb is different.

It is important to note that these formality distinctions in Japanese are not optional. All sentences must use one verb inflection or another, so speakers are always making a choice of what level of formality to use depending on social context.

Formality	Japanese sentence and transcription
Informal	駅の 近くに たくさんの お店が ある。 eki-no chikaku-ni takusan-no omise-ga aru
Polite	駅の 近くに たくさんの お店が あります。 eki-no chikaku-ni takusan-no omise-ga arimasu
Formal	駅の 近くに たくさんの お店が ございます。 eki-no chikaku-ni takusan-no omise-ga gozaimasu

Table 1: Three sentences meaning "There are many shops near the train station", in different levels of formality

For example, when speaking with family, close friends, or others of equal social status, the informal ある (aru "there are") is used. When speaking to superiors, strangers, or older individuals the polite expression あります (arimasu "there are") is used. When expressing deference or humility, the formal expression ございます (gozaimasu "there are") is used. In this paper we use the terms informal, polite, and formal to refer to these three levels of formality as shown in Table 1. Traditional Japanese grammars may make finer-grained, more nuanced distinctions than this.

While there is this nuance to Japanese grammar, in English there is no such distinction, so when translating from English into Japanese, a translator must choose one level of formality or another. This poses a challenge for English-Japanese NMT, since for a translation to be adequate it needs to both capture the meaning of the source sentence and use the appropriate level of formality.

We propose a method to allow English-Japanese NMT to produce translations in a particular level of formality, using an additional feature on the source side marking the desired level of formality to be used in the translation. With this feature provided at both training and test time, a single NMT system can learn to distinguish these levels of formality and produce multiple translations for the same input sentence. We evaluate our ap-

English Source	The number at the bottom of the list drops off.
Modified English Source	<polite> The number at the bottom of the list drops off.
Japanese Target	リストの一番下にある番号がリストから削除されます。 risuto-no ichiban shita-ni aru bangō-ga risuto kara sakujo saremasu

Table 2: Attaching a single token to the beginning of an English training data source sentence, based on the predicted formality of the Japanese target side

proach on multiple data sets and show that it successfully produces sentences in the requested level of formality. Apart from yielding more consistent outputs, it improves general translation quality as measured by BLEU on all data sets. We see particularly strong gains on the polite and formal portions of the test sets. We also release the following resources that were developed as part of our work towards formality-aware NMT

- A set of manual formality labels for a portion of the Tanaka corpus

- Code for a rule-based formality converter which can be applied as a translation post-processing step

We hope that these resources will spur further research on translation into Japanese.

2 Formality-Aware NMT

This section describes our approach for creating a formality-aware English-Japanese NMT system.

2.1 Choosing Formality in Translation

Our proposed method starts with identifying the formality of every Japanese target sentence in our parallel training corpus. We can determine that the Japanese sentence is informal, polite, or formal based on the verb inflection of the main verb of the sentence, which is often the last word in the sentence.

For example, in Table 2 the suffix ます (masu) at the end of the Japanese target sentence is a common politeness marker that identifies this as a polite sentence. This is particular to Japanese grammar, and from the English source sentence alone you cannot determine what level of formality the Japanese translation should have. So to inform our English-Japanese NMT system what formality level we are translating into, we attach the token *<polite>* to the beginning of the English source sentence. For every sentence pair in our training

corpus, we must attach such a token to the beginning of the English source side, depending on the formality of the Japanese target side.

At test time the resulting English-Japanese NMT model will need to be provided the same kind of informal, polite, or formal tokens at the beginning of every English input sentence to be translated. This allows the user of the NMT system to choose which level of formality they would like their Japanese translation to use. There are applications where these labels could be determined automatically from the context; we leave this for future work as our current data sets do not have context beyond the sentence level.

2.2 Automatic Identification of Honorifics

In order to label our training and test data with these formality tokens, we need to be able to identify the formality of a Japanese sentence automatically. To do this we look for the presence or absence of certain Japanese honorific verb forms as a heuristic. We created a set of common verbs and verbal inflections that correspond to each formality level, such as the informal expression じゃなかった (janakatta "was not"), the suffixes です (desu) and ます (masu), which attach to verb stems to express politeness, as well as several honorific and humble verbs such as なさいます (nasaimasu "to do" *honorific*) and 致します (itashimasu "to do" *humble*), which are used in formal social contexts to either show respect to the listener or show humility from the speaker, respectively. The full set of verb forms can be found in Table 3.

We apply our heuristics to a 21 million sentence Japanese monolingual corpus, composed of web-crawled text from multiple domains. We categorize sentences into three classes which we label informal, polite, or formal by looking for the verb forms in Table 3. We start with the formal verb forms. If any of these verbs are present we consider the sentence to be formal, if not then we proceed to looking for the polite verb forms, then the

Formality	Verb forms
Informal	だ, だった, じゃない, じゃなかった, だろう, da, datta, janai, janakatta, darou, だから, だけど, だって, だっけ, そうだ, ようだ dakara, dakedo, datte, dakke, souda, youda
Polite	です, でした, ない, なかった, ます, ました, ません, desu, deshita, nai, nakatta, masu, mashita, masen, ましょう, でしょう, ください, なさい, である, ではない mashou, deshou, kudasai, nasai, dearu, dewanai
Formal	ございます, いらっしゃいます, おります, なさいます, 致します, gozaimasu, irrashaimasu, orimasu, nasaimasu, itashimasu, ご覧になります, 拝見します, お目に掛かります, goranninarimasu, haikenshimasu, omenikakarimasu, おいでになります, 伺います, 参ります, 存知します, 存じ上げます, oideninarimasu, ukagaimasu, mairimasu, zonjishimasu, zonjiagemasu, 召し上がります, 頂く, 頂きます, 頂いて, 差しあげます, meshiagemasu, itadaku, itadakimasu, itadaite, sashiagemasu, 下さいます, おっしゃいます, 申し上げます kudasaimasu, osshaimasu, moushiagemasu

Table 3: Common verbs and suffixes for each level of formality, used as identifying heuristics.

informal verb forms. If none of the verb forms in Table 3 are present in the sentence it is ignored. From the original 21 million sentences, 1 million were unable to be categorized by our heuristics.

We hypothesize that a text classifier trained on the resulting 20 million sentences selected by our heuristics will learn more nuanced distinctions in word choice and style than using the heuristics alone, which only identify a small set of verb forms. We tokenize this data set with the KyTea morphological analyzer (Neubig, 2011b) and train a model on the tokenized monolingual data and labels with the text classification tools provided by the FastText (Joulin et al., 2017) toolkit, using word trigram features.

To evaluate our classifier's performance, we enlisted the help of a Japanese linguist to make formality judgments on a small test set of 150 Japanese sentences drawn from the publicly-available Tanaka corpus (Tanaka, 2001). Out of these 150 total sentences, 68 were labeled *informal*, 45 were labeled *polite* and 37 were labeled *formal* by the annotator. These sentences and annotations will be made publicly available alongside the publication of this paper.

Table 4 contains a precision-recall evaluation of our formality classifier, showing strong F1 scores for all three classes. The formality classifier output matches the exact classifications made by our

	Informal	Polite	Formal
Precision	1.00	0.82	0.72
Recall	0.74	0.91	0.97
F1	0.85	0.86	0.83

Table 4: Evaluation scores of labels produced by the formality classifier compared to gold test set labels for each formality category (n=150).

heuristic rules on this test set, but we hypothesize that it generalizes better to unseen text and therefore use it in our translation experiments. The results show that our classifier has higher precision on the informal category, but lower recall, and higher recall on the polite and formal categories, but lower precision.

2.3 Rule-Based Formality Conversion

We also compare our method of formality-aware NMT with a simple rule-based tool which converts a Japanese sentence from one level of formality to another. This is done by identifying the main verb in a Japanese sentence and either replacing the verb itself or just the verbal inflection with the inflection for the desired level of formality. The code will be made available open-source alongside this publication.

Rule-based formality conversion is non-trivial since there are many conjugations to consider for

a single verb, which differ based on the class the verb belongs to. For example, to convert the verb 歩きました (arukimashita "walked" *polite*)" to an informal inflection, the polite suffix ました (mashita) is removed from the stem of the verb and a new suffix is appended to create 歩いた (aruita). The き (ki) at the end of the verb stem marks this as a verb in a particular verb class. All verbs with *ki* at the end of their stem belong to the same class and have the same conjugation pattern.

In order to use this rule-based method to compare to our English-Japanese formality-aware NMT, we can simply take our baseline NMT system, trained without the formality tokens described above in section 2.1, and apply the rules to convert the NMT output into the desired level of formality. However, this rule-based method is imperfect and relies on tokenization and part-of-speech information from the KyTea morphological analyzer. Incorrect part of speech tags or tokenization that doesn't match our rule-based tool's dictionary will lead to errors in changing verbal inflection. In our evaluation, we show how using this rule-based method compares to our formality-aware NMT.

3 Evaluation

In this section, we evaluate the translation quality of our formality-aware NMT models as well as their ability to produce the desired formality level in the output.

3.1 Datasets

We use three publicly-available parallel data sets for our NMT experiments. The Asian Scientific Paper Excerpt Corpus (ASPEC) (Nakazawa et al., 2016), a corpus of scientific paper abstracts, the Japanese-English Subtitle Corpus (JESC) (Pryzant et al., 2018), a corpus of sentence-aligned movie and television subtitles, and the Kyoto Free Translation Task (KFTT) (Neubig et al., 2011a), a corpus of Wikipedia data about the city of Kyoto. In our experiments we use the standard training and test sets for each parallel corpus. We also use a proprietary parallel training data set which contains web-crawled data from a mix of domains and a corresponding test set. Training and test set sizes are reported in Table 5.

Dataset	Train	Test
Proprietary	23,781,990	300
ASPEC	1,000,000	1,812
JESC	3,237,376	2,001
KFTT	329,882	1,160

Table 5: Parallel training and test data set sizes in number of sentences

3.2 Experimental Setup

The English source of each bitext was tokenized with the Moses (Koehn et al., 2007) tokenizer.perl script and the Japanese target was tokenized with KyTea. We limit sentence length to 80 tokens on either side of the bitext and train a Moses truecaser for the English source side, except for the JESC data set because the English side of the JESC corpus is already entirely lowercased. We use 32k subword unit vocabularies (Sennrich et al., 2016b) separate for source and target. We use the resulting tokenized, truecased, subworded training data to train our NMT models.

Our experiments use the Transformer (Vaswani et al., 2018) to train NMT models. We use 512 hidden units, 6 hidden layers, 8 heads, and a batch size of 4096. We train for 200k training steps using the Adam optimizer (Kingma and Ba, 2015).

For each parallel corpus, we train a formality-aware NMT model by classifying the formality of the Japanese target side and attaching a corresponding feature to the beginning of each English source segment, identifying the target as being informal, polite, or formal. For comparison, we also train a baseline NMT model without these formality annotations.

3.3 Experimental Results

To evaluate our formality-aware NMT models, we first need to choose the right level of formality for each sentence in the test sets. We do this by applying the formality classifier to the test reference and prepending the predicted labels to the source side of each test sentence. We then provide this input to our formality-aware NMT models and compare the output to test set translations from our baseline NMT using BLEU (Papineni et al., 2002). We tokenize the Japanese MT output and reference using KyTea before computing BLEU. We evaluate on the overall test set, as well as each separate portion of the test set where the test reference was classified as informal, polite or formal. Table 6 shows

our results on the test set using BLEU.

3.3.1 Performance of Rule-Based Conversion

We first evaluate the performance of the rule-based conversion method described in Section 2.3. The rule-based tool currently has the capability to convert to informal or polite verbal inflections, lacking rules for formal verb inflections. Thus, we only report results on the informal and polite sections of our test sets.

As shown in Table 6, the rule-based method yields improvements on the informal test portion for all models except ASPEC where performance remains the same. On the polite portion, we only see gains for the JESC model but a notable decrease in performance on ASPEC and no changes for the proprietary and KFTT test sets. This shows that while it is possible to adjust the formality level through post-processing, it is a non-trivial task and will require more work to improve the coverage of the tool. However, the rule-based tool could also be used for other tasks such as creating additional synthetic training data.

3.3.2 Performance of formality-aware NMT

The BLEU scores in Table 6 show that on the overall test set, our formality-aware NMT models show an improvement over the baseline NMT models. This holds true for both the model trained on the proprietary training data set, and the models trained on the publicly-available training data sets. Out of the models trained on publicly-available data, the ASPEC model shows the smallest improvement (+0.3 BLEU), the KFTT model improves more (+0.9 BLEU), and the JESC model shows the highest improvement (+1.5 BLEU).

When looking at the individual portions of the test set, as identified by our classifier, we see a larger quality improvement for the model trained on proprietary data on the informal and formal sections of its test set, and a smaller improvement on the polite section. The ASPEC formality-aware NMT is not better on the informal section of its test set, but there are larger gains in quality on the polite and formal test sections. The JESC and KFTT models improve on all three sections, with the largest gains seen in the formal section. Finally, formality-aware NMT improves over the rule-based method for all models and test sections, indicating that the NMT model is more effective at producing the desired formality level in context.

3.3.3 Evaluating formality levels

Since choosing the appropriate formality level in Japanese is very important to conform with social norms, we want to show that our formality-aware NMT models can provide translations in the desired level of formality. As our test sets do not have gold labels from a human annotator for each reference, we use our formality classifier to predict the level of formality for both the MT output and the test reference and compute F1 scores using the predicted reference labels.

Our F1 comparison in Table 7 shows to what extent the formality-aware NMT output matches the predicted formality level of the reference translation when the system is provided with the correct input label. We can see that the F1 scores for the formality-aware NMT are high for all three levels of formality, above 0.9 in all cases. We also see a big improvement over the baseline NMT models for each test set, especially in the polite and formal categories. From this we conclude that our formality-aware NMT models can produce a translation in the desired level of formality.

An imbalance of the training data may partly explain the difference in quality improvement across the three formality sections of the test sets. Table 8 shows how much of the training data for each data set was classified as being informal, polite, or formal. The proprietary data set contains mostly polite and informal data. In contrast, the majority of the three publicly available data sets is informal data, with a much smaller portion of polite data. For all data sets there is very little formal data, leading to the weak baseline performance on that category. By modelling formality levels more explicitly, our models are better able to compensate the inherent bias towards informal style.

4 Analysis and Examples

To show some concrete examples of our formality-aware translations, Table 9 contains an example of the MT output from the JESC formality-aware NMT model and the corresponding JESC NMT baseline trained without formality annotations. For this single English source sentence, there are multiple different MT outputs depending on which formality label is attached to the source before passing it to the NMT model for translation. The informal expression ない (nai "there is not") is used in the MT output by both the baseline model and the formality-aware NMT model when

Dataset	Model	test BLEU			
		Overall	Informal	Polite	Formal
Proprietary	Baseline NMT	24.6	17.8	28.0	17.4
	Rule-Based Conversion	-	18.3	28.0	-
	Formality-Aware NMT	25.5	18.7	28.4	22.3
ASPEC	Baseline NMT	43.0	42.8	45.7	33.0
	Rule-Based Conversion	-	42.8	44.5	-
	Formality-Aware NMT	43.3	42.8	47.1	43.9
JESC	Baseline NMT	18.8	18.0	19.4	22.1
	Rule-Based Conversion	-	18.7	20.6	-
	Formality-Aware NMT	20.3	19.4	21.9	29.3
KFTT	Baseline NMT	24.6	26.0	18.1	11.4
	Rule-Based Conversion	-	26.4	18.1	-
	Formality-Aware NMT	25.5	26.5	20.7	19.3

Table 6: KyTea-tokenized BLEU, comparing baseline NMT, rule-based conversion and formality-aware NMT models on heldout test

Dataset	Model	F1		
		Informal	Polite	Formal
Internal	Baseline NMT	0.59	0.82	0.29
	Formality-Aware NMT	0.97	0.99	0.91
ASPEC	Baseline NMT	0.95	0.67	0.19
	Formality-Aware NMT	1.00	1.00	0.96
JESC	Baseline NMT	0.85	0.50	0.28
	Formality-Aware NMT	1.00	0.99	0.96
KFTT	Baseline NMT	0.93	0.41	0.00
	Formality-Aware NMT	1.00	0.96	0.74

Table 7: F1 scores for each formality category when comparing predicted labels for MT output and reference translation.

Train	Informal	Polite	Formal
Internal	41.5%	56.9%	1.6%
ASPEC	87.6%	11.2%	1.2%
JESC	80.7%	18.3%	0.9%
KFTT	90.9%	8.4%	0.7%

Table 8: Percentage of each training data set classified as informal, polite, or formal

'informal' is attached to the source segment. The polite expression ありません (arimasen "there is not") is used by the formality-aware NMT model when 'polite' is attached, and the formal expression ございません (gozaimasen "there is not") is used by the formality-aware NMT model when 'formal' is attached. All of these expressions have the same meaning, but correspond correctly to the desired level of formality.

Table 10 shows another example of the MT output from the internal formality-aware NMT model and the corresponding internal NMT baseline trained without formality annotations. Again there are multiple different MT outputs depending on which formality label is attached to the source. The informal expression 戦う (tatakau "to fight") is used in the MT output by the formality-aware NMT model when 'informal' is attached to the source segment. The polite expression 戦います (tatakaimasu "to fight") is used by the baseline NMT model and the formality-aware NMT model when 'polite' is attached, and the formal expression 戦いをいたします (tatakai-wo itashimasu "to do battle" *humble*) is used by the formality-aware NMT model when 'formal' is attached. All of these expressions have the same meaning, but correspond correctly to the desired level of formality. Here the formality-aware NMT model is closest to the reference when 'informal' is attached, but all three formality-aware NMT translations are

Source	There's nothing to apologize for.
Reference	謝ることなんかなにもないさ ayamaru koto nanka nanimo nai sa
NMT Model	MT Output and Transcription
Baseline	謝ることなんて何もない ayamaru koto nante nanimo nai
Formality-Aware - Informal	別に謝ることはない betsu ni ayamaru koto wa nai
Formality-Aware - Polite	謝ることはありません ayamaru koto wa arimasen
Formality-Aware - Formal	お詫びのことは何もございません owabi no koto wa nanimo gozaimasen

Table 9: Example output from JESC NMT baseline model and formality-aware NMT model, when each formality level is attached to the source segment.

equally adequate.

5 Related Work

Sennrich et al. (2016a) showed that side constraints can be added to the source side of a parallel text to provide control over the politeness of translation output in an English-German translation task. Following this paper's suggestion, we take a similar approach towards Japanese honorifics.

Niu et al. (2017) also use a similar approach, termed "Formality-Sensitive Machine Translation", in a French-English translation task. In (Niu et al., 2018) French-English parallel text with formality features is combined with English-English parallel text, where the source and target are of similar meaning but different formality, to create a multi-task model that performs both formality-sensitive MT and monolingual formality transfer.

In related work on Japanese-English NMT, Yamagishi et al. (2016) use a side-constraint approach to control the voice (active or passive) of an English translation. Takeno (2017) apply side constraints more broadly to control translation length, bidirectional decoding, domain adaptation, and unaligned target word generation.

Our paper follows the modeling approach introduced by Johnson et al. (2017), who showed that by adding a token to the source side of parallel text allows for training a single NMT model on data for multiple language pairs. Their token specifies the desired target language, allowing the user control over the language of machine translation output, even for source-target language pairs that were not seen during training, which they call "zero-shot"

translation. The same approach has been successfully used in other applications, such as in distinguishing standard versus back-translated translation parallel corpora (Caswell et al., 2019).

6 Conclusion

We have shown how the distinctions between levels of formality in the Japanese language can be learned by an NMT model, by identifying Japanese honorifics in parallel training data and labeling the source side with an additional feature. We find that this technique provides control over the honorifics present in the MT output and provides an improvement in translation quality, particularly in polite and formal sentences in each test set. This improvement holds for models trained on proprietary data as well as models trained on three widely-used publicly available Japanese data sets. In future work, we would like to explore augmenting the training data for each of the comparisons we showed. We would like to explore creating artificial English-Japanese data by doing a rule-based transformation of the Japanese side of the bitext into different formality levels. We would also like to do further human evaluation of our Japanese formality classifier and the NMT models we trained, and we may explore applying this technique to English-Korean NMT because Korean also has a similar system of honorifics.

References

Isaac Caswell, Ciprian Chelba, and David Grangier. 2019. Tagged back-translation. In *Proceedings of the Fourth Conference on Machine Translation*,

Source	King Arthur's knights do battle with a killer rabbit.
Reference	円卓の騎士たちはキラーラビットと戦う。 entaku-no kishitachi-wa kirā rabitto-to tatakau
NMT Model	MT Output and Transcription
Baseline	アーサー王の騎士はキラーウサギと戦います。 āsā ō-no kishitachi-wa kirā usagi-to tatakaimasu
Formality-Aware - Informal	キングアーサーの騎士たちはキラーウサギと戦う。 kingu āsā-no kishitachi-wa kirā usagi-to tatakau
Formality-Aware - Polite	キングアーサーの騎士たちはキラーウサギと戦います。 kingu āsā-no kishitachi-wa kirā usagi-to tatakaimasu
Formality-Aware - Formal	キングアーサーの騎士たちはキラーウサギと戦いをいたします。 kingu āsā-no kishitachi-wa kirā usagi-to tatakai-wo itashimasu

Table 10: Example output from internal NMT baseline model and formality-aware NMT model, when each formality level is attached to the source segment.

pages 53–63, Florence, Italy. Association for Computational Linguistics.

Melvin Johnson, Mike Schuster, Quoc V Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomáš Mikolov. 2017. Bag of tricks for efficient text classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431. Association for Computational Linguistics.

Diederik P Kingma and Jimmy Lei Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*. arXiv.org.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180. Association for Computational Linguistics.

Toshiaki Nakazawa, Manabu Yaguchi, Kiyotaka Uchimoto, Masao Utiyama, Eiichiro Sumita, Sadao Kurohashi, and Hitoshi Isahara. 2016. ASPEC: Asian scientific paper excerpt corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2204–2208. European Language Resources Association (ELRA).

Graham Neubig. 2011b. The Kyoto free translation task.

Graham Neubig, Yosuke Nakata, and Shinsuke Mori. 2011a. 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. Association for Computational Linguistics.

Xing Niu, Marianna Martindale, and Marine Carpuat. 2017. A study of style in machine translation: Controlling the formality of machine translation output. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2814–2819.

Xing Niu, Sudha Rao, and Marine Carpuat. 2018. Multi-task neural models for translating between styles within and across languages. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1008—-1021.

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. Association for Computational Linguistics.

Reid Pryzant, Youngjoo Chung, Dan Jurafsky, and Denny Britz. 2018. JESC: Japanese-English subtitle corpus. In *Proceedings of the 11th Language Resources and Evaluation Conference*. European Language Resource Association.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. Controlling politeness in neural machine translation via side constraints. In *Proceedings of NAACL-HLT 2016*, pages 35–40. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 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. Association for Computational Linguistics.

Nagata Masaaki Yamamoto Kazuhide Takeno, Shunsuke. 2017. Controlling target features in neural machine translation via prefix constraints. In *Proceedings of the 4th Workshop on Asian Translation*, pages 55–63.

Yasuhito Tanaka. 2001. Compilation of a multilingual parallel corpus. In *Proceedings of PACLING 2001*, pages 265–268.

Ashish Vaswani, Samy Bengio, Eugene Brevdo, François Chollet, Aidan Gomez, Stephan Gouws, Llion Jones, Łukasz Kaiser, Nal Kalchbrenner, Niki Parmar, Ryan Sepassi, Noam Shazeer, and Jakob Uszkoreit. 2018. Tensor2Tensor for neural machine translation. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, pages 193–199. Association for Machine Translation in the Americas.

Hayahide Yamagishi, Shin Kanouchi, Takayuki Sato, and Mamoru Komachi. 2016. Controlling the voice of a sentence in Japanese-to-English neural machine translation. In *Proceedings of the 3rd Workshop on Asian Translation*, pages 203–210.