Neural Machine Translation Strategies for Generating Honorific-style Korean

Lijie Wang*, Mei Tu*, Mengxia Zhai*, Huadong Wang*, Song Liu*, Sang Ha Kim†
* *Samsung Research China - Beijing (SRC-B), China*
† *Samsung Research (SR), Seoul, Korean*
{lijie.wang, mei.tu}@samsung.com, mengx.zhai@partner.samsung.com
{huadong.wang, s0101.liu, sangha01.kim}@samsung.com

***Abstract*—Expression with honorifics is an important way of dressing up the language and showing politeness in Korean. For machine translation, generating honorifics is indispensable on the formal occasion when the target language is Korean. However, current Neural Machine Translation (NMT) models ignore generation of honorifics, which causes the limitation of the MT application on business occasion. In order to address the problem, this paper presents two strategies to improve Korean honorific generation ratio: 1) we introduce honorific fusion training (HFT) loss under the minimum risk training framework to guide the model to generate honorifics; 2) we introduce a data labeling (DL) method which tags the training corpus with distinctive labels without any modification to the model structure. Our experimental results show that the proposed two strategies can significantly improve the honorific generation ratio by 34.35% and 45.59%.**

***Keywords*-honorific fusion training; data labeling; korean honorific**

I. Introduction

Neural Machine Translation has made great progress in recent years, the translation results have reached or even exceeded the level of human beings in some fields, especially after combining the LSTM /GRU [1, 2] or self-attention mechanism [3]. In NMT, the encoder encodes the source language into a high level of semantic representation and decoder generates the target translation, the systems comply with this kind of structure and achieve good performance in most cases. However, the generated translation fails to express target language phenomenon sometimes, e.g. honorific-style expressions.

Expressions containing honorifics are used to show respect towards a speaker's audience. It also indicates the level of formality and politeness of a situation. Therefore, honorifics are incredibly important in Korean culture and widely used for conversation and relationship-building, ignorance of using honorifics can be seen as very impolite, it is necessary to use honorifics in order to show respect to people like elders, those in superior positions. The following examples show honorific-style and non-honorific-style expressions when translating from English to Korean:

- Source: "What did you do last weekend?"
- Honorific target:

ji-nan ju-mar-e mu-eos-eul ha-syeot-seum-ni-kka?
지난 주말에 무엇을 하**셨습니까**?

- Non-honorific target:

ji-nan ju-mar-e mwol haen-ni?
지난 주말에 뭘 했니?

The two targets above are correct and the Korean words with bold are honorifics. The first target is a polite honorific-style expression, while the second target is a non-honorific-style expression that will be not appropriate for a formal occasion. Since the current encoder-decoder structures in NMT cannot capture the honorific information in the source sentence, it is hard to generate the honorific-style target as the final translation according to the context representation of the encoder. From the perspective of training, the reason that the current NMT techniques are difficult to generate honorifics mainly comes from two folds:

- Source sentence and target sentence contain asymmetric information. The current NMT can't distinguish the honorifics from the training data. The honorific and non-honorific sentences are mixed together. One source sentence corresponds to both honorific and non-honorific sentences sometimes. The data consistency is not good, thus making it difficult to learn.
- The translation model itself does not have a reward and punishment mechanism for the honorifics and non-honorifics, which leads to the translation paying no attention to the choice of honorifics.

There're two obvious existing methods to solve the above problems. The first method is to pick out the parallel corpus with honorific-style targets; the second method is to modify the translation results to honorific-style expressions by hard rule. However, the honorific-style data is only about 30% of the whole corpus, which is not enough to achieve good translation performance. While the hard-rule modification will always lead to disfluent translation. On the other hand, a complement rule-database is hard to build, making the second method less practical. Considering the data utilization rate and feasibility, we propose two strategies to generate the honorific-style translation:

- The first strategy is honorific fusion training which introduces honorific fusion training loss into the training loss function. The honorific fusion training loss is integrated under the minimum risk training (MRT) [4] framework and can help the translation model paying more attention to honorifics during training
- The second strategy is to annotate distinctive labels on the training corpus. From the perspective of information symmetry, the source sentence is marked with honorific or non-honorific tag according to target sentence. The encoder can capture the distinctive feature from the source sentence, which contributes to generating honorific or non-

978-1-7281-5014-7/19/$31.00

honorific translation based on the addictive tag in source side.

II. Related Work

The encoder-decoder NMT architectures with attention mechanism like GNMT [5], ConvS2S [6], Transformer [3] have achieved state-of-the-art in recent years. However there're still some challenges [7] at the model and data level after excluding decoding speed, previous work has focused on how to optimize the data and model space. Dual learning [8] and unsupervised NMT [9, 10] are proposed to solve the problem of insufficient training data; Deliberation net [11] is proposed to solve decoding problem that the beam search is not globally optimal [12]; MRT [4] and Adversarial-NMT [13] are proposed to update the parameter gradient based on sampling obtained by search. Moreover, the current NMT models are optimized by maximum likelihood estimation to make the translation result more fidelity. However it's uncertain whether the style of translation result is appropriate in some cases, this paper tries to address this issue at both model and data level according to the previous work.

From the perspective of the model, we learn from the text style transfer which is a sequence generation method in order to apply constraints to the NMT loss to guide the generation of results containing expected styles. Text style transfer [14, 15, 16] is a rephrasing task without changing the internal meaning of the context, and generative adversarial networks [17], back-translation [18] and Auto-Encoder [19] have been proposed to yielded state-of-the-art results. Despite the success of NMT and text style transfer, previous work is rarely involved how to accurately translate the source language to another target language that contains the desired style. Our solution is not only to translate the source language correctly, but also make the generated translation results meet the specific style rather than simple rephrasing.

From the perspective of data, the source language doesn't contain the characteristics of the target sentence style. Therefore, we borrow the method of labeling the language direction in one-to-many translation [20], and tag source language with the style label of the corresponding target language before training. After training, the model can learn the style corresponding to different tags.

III. Our Method

In this section, we proposed two methods to improve the honorific generation ratio. The first method introduces honorific fusion training which depends on an honorific classifier to help the NMT model to generate more honorific results. The second method adopts data labeling to add corresponding label at the end of the source Chinese sentence depending on whether the target is an honorific-style result or not before training, then the translation model can learn the honorific mode during training.

A. Honorific Fusion Training

Given parallel corpus X, Y, where $X = \{x_1, x_2, ..., x_s\}$, $Y = \{y_1, y_2, ..., y_s\}$ and (x_i, y_i) is the i-th aligned sentence pair, $M(\theta)$ represents the parameters of NMT model. The training goal is to minimize the total loss function in formula (1) with model parameter $M(\theta)$, θ is the parameter of NMT model:

$$Loss_{Total}(\theta) = \gamma Loss_{MT}(\theta) + (1-\gamma) Loss_{HE}(\theta) \tag{1}$$

Where $Loss_{MT}$ refers maximum likelihood estimation (MLE) loss, while $Loss_{HE}$ indicates our proposed honorific fusion training loss. γ is a hyper-parameter that controls the trade-off between honorific fusion training loss and MLE loss.

Honorific fusion training loss is defined as expected loss with respect to the posterior distribution in formula (2):

$$\begin{aligned} Loss_{HE}(\theta) &= \sum_{i=1}^{S} E_{y_i'|x_i;\theta}[Loss_D(y_i')] \\ &= \sum_{i=1}^{S} \sum_{y_i' \in y(x_i;\theta)} P(y_i'|x_i;\theta) Loss_D(y_i') \end{aligned} \tag{2}$$

Where the $Loss_D(y_i')$ is the honorific loss of the model prediction y_i', $y(x_i;\theta)$ is the all possible candidate translations for x_i with $M(\theta)$. $P(y_i'|x_i;\theta)$ is the probability of generating y_i given x_i and θ.

The proposed honorific loss is to measure how the model prediction y_i' is not likely to be an honorific-style sentence. In this paper, we use a pre-trained honorific classifier D to compute it with formula (3). D is based on convolutional neural network (CNN). $D(y_i')$ is the softmax output of classifier. Sentences used for training are labeled by rules in Section III.B. We follow [21] to build the honorific classifier, where the label 1 stands for honorific-style sample and 0 stands for non-honorific-style sample.

$$Loss_D(y_i') = -\log D(y_i') \tag{3}$$

In formula (2), the expectations are intractable to calculate due to the exponential search space of $y(x_i;\theta)$. Using a subset of the full search space to approximate the posterior distribution to alleviate this problem in formula (4):

$$Loss_{HE}(\theta) = -\sum_{i=1}^{S} \sum_{j=1}^{K} Q(y_{ij}'|x_i;\theta) \log(D(y_{ij}')) \tag{4}$$

Where the j-th sampled sentence is y_{ij}', K candidate sentences is sampled as a subset of the full search space, $Q(y_{ij}'|x_i;\theta)$ is a distribution defined on the subset in formula (5):

$$Q(y_{ij'}|x_i;\theta) = \frac{P(y_{ij}'|x_i;\theta)}{\sum_{j=1}^{K} P(y_{ij}'|x_i;\theta)} \tag{5}$$

The overall framework of our honorific fusion training is shown in Figure 1. The honorific fusion training contains MLE and honorific fusion training loss as follows:

1) Input sentence (x_i, y_i) into NMT to get probability matrix of prediction and evaluate the gap between reference and predicts, then get $Loss_{MT}(\theta)$ by cross entropy.

2) Generate K candidate sentences $\{y'_{i1}, y'_{i2}, \ldots, y'_{iK}\}$ by polynomial sampling with playback K times at each decoding step according to the output distribution [22].
3) Compute the approximating posterior probability of the sampled K sentences $Q(y'_{ij}|x_i;\theta)(1 \le j \le K)$ according to formula (5).
4) Compute honorific loss $Loss_D(y'_i)$ according to formula (3).
5) The total loss of sentence pair(x_i, y_i) is as follows:

$$Loss_{Total}(\theta) = \gamma Loss_{MT}(\theta) - (1-\gamma)\sum_{j=1}^{K} Q(y'_{ij}|x_i;\theta)\log D(y'_{ij}) \quad (6)$$

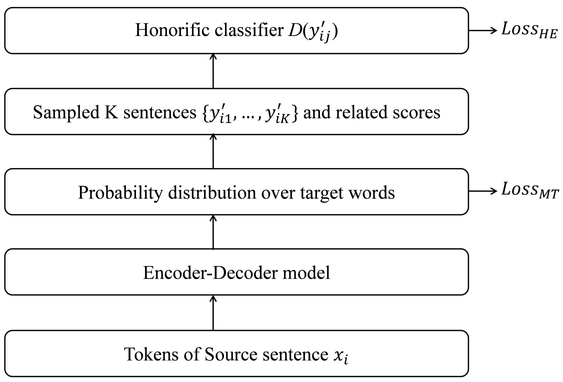

Figure 1. The honorific fusion training framework

The Algorithm 1 shows the full details of the proposed honorific fusion training. In this paper, the NMT model is pre-trained with only MLE loss firstly. The model is initialized with the pre-trained NMT parameters instead of training from scratch in honorific fusion training. The classifier D is also pre-trained upon the honorific and non-honorific corpus. The classifier is built following [21]. It is based on CNN with multiple convolution kernels. The classification performance on our test set will be reported in the experimental part. The honorific and non-honorific corpus are obtained from the target side of the parallel training corpus. Hard rules are applied to determine whether a sentence is an honorific-style sentence or not, which will be described in Session III.B. The hyper-parameters γ and K will be discussed in the experimental part. Based on pre-trained NMT model and classifier D, the honorific fusion training won't take too much training time in the fine-tune stage, and decoding time remains the same.

B. Data Labeling

Data labeling is used to automatically distinguish the target language in a one-to-many translation task [20], the target direction is labeled at the end/start of the source sentence to indicate the translation direction. The second strategy is to add distinctive labels on the training data

Algorithm 1 Honorific fusion training.

Input: Chinese-Korean parallel sentence: x, y and the target sentence length is T, K is maximum sampling times, γ is the hyper-parameter controlling trade-off for loss.

Output: $M(\theta)$

```
Load pre-trained NMT model M(θ), honorific classifier D
i ← 1
while i ≤ T do
    j ← 1
    while j ≤ K do
        Sampling y'_ij ~ P(y'_i|x, y_<i; θ), P_ij = P(y'_ij|x, y_<i; θ)
    end while
end while
P_j = ∏_{i=1}^{T} P_ij
Q(y'_j|x; θ) = P_j / Σ_{j=1}^{K} P_j   //score of j-th sampled result
Loss_HE(θ) = − Σ_{j=1}^{K} Q(y'_j|x; θ) × log D(y'_j)
    //honorific fusion training loss
Loss_MT(θ) = CE(y, y'_i)   //the cross-entropy loss
Loss_Total(θ) = γ Loss_MT(θ) + (1 − γ) Loss_HE(θ)
Update NMT model by gradient ∇_θ Loss_Total(θ)
```

to generate honorific-style translation. The data labeling method can capture honorific feature from the target sentence by tagging the labels at the end of source sentence. Label <p> or <np> is added at the end of the source sentence according to the aligned target sentence. If it is an honorific-style sentence, the corresponding label <p> is added at the end of source sentence, or else <np> is added. We did nothing else to the model except for the data labeling. When testing, we add label <p> at the end of every input sentence to generate honorific-style expression. In addition, the training time and decoding time won't increase too much with just a simple label added at the end of every sentences.

Table I
SPEECH LEVELS IN KOREAN

Name	Politeness	Formality
Hasipsio-che (하십시오체)	High	High
Haeyo-che (해요체)	High	Low
Hao-che (하오체)	Neutral	High
Hage-che (하게체)	Neutral	Neutral
Haera-che (해라체)	Low	High
Hae-che (해체)	Low	Low

The criteria above determine what are the honorifics. According to the Korean grammar[1] published by the National Institute of Korean Language, the Korean language experts summarize six speech levels based on the level of politeness and formality in the Korean dialogue as shown in Table I. Each level has its own unique set of verb endings which are used to indicate the level of formality and politeness.

[1] http://www.ilovekorean.net/files/pdf/korgrammar.pdf

Korean language experts classify the Korean sentence under high level of politeness into an honorific-style expression. As shown in Table I, both Hasipsio-che (another name is hapsyo-che or '합쇼체') and Haeyo-che meet the requirement of honorific. Referring to related paper [23], the verb endings of the above two honorific types are summarized as shown in Table II.

Table II
VERB ENDING OF TWO HONORIFIC TYPES

Type	Verb Ending with Hasipsio-che	Verb Ending with Haeyo-che
Declarative	-ㅂ/습니다	-아요/어요
Interrogative	-ㅂ/습니까	-아요/어요
Imperative	-(으)십시오/-(으)시지요	-아요/어요
Propositive	-(으)십시다	-아요/어요

IV. EXPERIMENTS

A. Datasets and Metrics

The paper focuses on improving honorific generation ratio which is defined as the number of sentences containing honorifics divided by the total number of sentences. Our translation task is conducted on Chinese into Korean direction. Our baseline and honorific fusion training experiments make use of an MT training set of about 7.2M Chinese-Korean parallel sentences, including 3.9M purchased data and 3.3M script crawled data. Before training, Chinese have been segmented and processed into subword units using byte-pair encoding (BPE) [24], eventually the source Chinese vocabulary size is about 4.4k. For Korean, tokenizer[2] in Moses and BPE are used to obtain the target Korean vocabulary and the size is about 4.2K. The randomly extracting 2,133 Chinese sentences are translated into Korean 4 references by crowdsourcing as the Normal Test Set (NTS), then we extract the parallel sentences in which the references containing honorifics in any of them are about 1,474, we set it as the Honorific Test Set (HTS). BLEU [25] score and honorific generation ratio are adopted when evaluating.

When training honorific classifier, 1M source Chinese sentences are randomly selected, the pre-trained and best performance NMT model is used to translate the selected Chinese into Korean and we get best 5 translation results of every sentences. Based on the classification criteria of honorific in section III.B, the Korean are classified to honorific-style sentences and non-honorific-style sentences with label 1 and 0. Finally, 5M labeled Korean sentences with 30% of them contain honorifics are obtained to train the honorific classifier.

When training NMT model with data labeling, based on the classification criteria of honorific in section III.B, the 7.2M processed Chinese-Korean parallel sentences are labeled at the end of corresponding source Chinese sentences according to the raw Korean sentences. According to statistics, honorific sentences account for about 30%.

[2]https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl

B. Setup

Transformer base model [3] is used as the baseline model. We use 2 Tesla P40 GPUs to pre-train the NMT model for honorific fusion training. The NMT experiments are carried out based on the OpenNMT-py [26] and the honorific classifier[3] is based on [21]. We use Adam [27] with $\beta_1 = 0.9$, $\beta_2 = 0.998$ and $\epsilon = 10^{-9}$ when training NMT. We vary learning rate with *warmup_steps* 4000 over the course of training.

C. Experiment Results

As shown in Table III, the accuracy of honorific classifier is 99.99% on a test set of 7k, the BLUE of honorific fusion training model (γ=0.7, K=50), and data labeling model in NTS are a little bit lower than baseline model, although the honorific generation ratio is much higher than baseline model. The reason that the BLEU of proposed two methods are a little bit lower than baseline model, we think it's caused by the references. Since the references are not for honorific expressions, and it's unfair to compare honorific-style results with NTS, then we test in HTS, the results show both honorific fusion training and data labeling method can not only improve BLEU, but also improve the honorific generation ration by 34.35% and 45.59% respectively . In order to compare the effectiveness of the proposed methods a step further, we use manual evaluation to further analyze the results in Table IV.

Table III
COMPARISON OF HONORIFIC FUSION TRAINING AND DATA LABELING

Test Set	Model	BLEU	HG Ratio
NTS	Baseline	**34.21**	44.82%
	HFT	33.57	79.23%
	DL	34.17	**99.40%**
HTS	Baseline	35.53	54.21%
	HFT	36.04	88.74%
	DL	**37.20**	**99.80%**

NTS represents Normal Test Set; HTS represents Honorific Test Set; baseline represents Transformer baseline model; HFT represents the honorific fusion training model; DL represents the data labeling model; HG Ratio represents honorific generation ratio.

D. The Human Evaluation

In this paper, we random take 300 Chinese sentences and their corresponding Korean results. We adopt Mean opinion score (MOS) as subjective evaluation from "excellent" to "bad" with range 5 to 1. In order to consider the correctness of the translation and the honorific-style expression, when the result Korean sentence is not an honorific-style expression, the sentence score will be subtracted by 0.2 points; when the result honorific-style Korean sentence doesn't conform to correct grammars, the sentence score will be subtracted by 1. The final average MOS is shown in Table IV, the MOS of honorific fusion training and data labeling method exceed the baseline model, and honorific-style results also increase greatly.

[3]https://github.com/junwang4/CNN-sentence-classification-pytorch-2018

Table IV
HUMAN EVALUATION

Model	MOS	HG Ratio
Baseline	4.41	48.67%
HFT	4.43	81.00%
DL	**4.48**	**95.00%**

During manual evaluation, we find the two proposed methods have their own characteristics. Honorific fusion training model will miss honorific verb endings in few cases to ensure the accuracy of the results, while DL will ensure the generation of honorifics to the greatest extent. There're some positive and negative impact on final results when using method DL. In few cases, the DL model will add respectful title for the character although we don't deliberately make the model to learn such honorific features, but in few cases the DL model will add some suffix '요 (yo)' to represent the honorific feature for the translation of single word, for example when translation '1.5 欧元 (ou yuan)', the result will be '1.5유로요 (yuro-yo)' , however it's grammatically wrong. We think the phrase translation ability of DL method in some cases should be further improved.

It is concluded that the result of data labeling model surpasses honorific fusion training model. Data labeling method not only has good effect but also has a simpler process without modifying the NMT model; the honorific fusion training provides a way to generate a specific style of the translation. We notice the limitation that the current methods we proposed can only generate honorific-style sentence endings. Next, we will further try to generate honorific words in the middle of sentences including nouns and pronouns, and try to verify the effectiveness of our methods in honorific expression in other languages like Japanese.

E. The Effect of Hyperparameters to HFT

In honorific fusion training experiments, the accuracy of honorific classifier is 99.99% on a test set of 7k. We adjust the hyper-parameter γ which is the parameter that controls trade-off between MLE and honorific fusion training loss, we evaluate in NTS. In the process, maximum sampling times K remaining unchanged and results are shown in Table V ,and the test set is NTS. The BLEU score continues to increase from 32.07 to 34.17 when K = 10 and γ increases from 0.1 to 0.9 at interval 0.2 (When γ = 1, the model turns to the standard Transformer baseline model), but the honorific generation ratio decreases from 84.76% to 54.95%. The increase of honorific generation ratio accompanied by the decrease of BLEU, BLEU decreases slightly but honorific generation ratio increases by 24.75% when γ = 0.7; in order to raise the honorific generation ration as much as possible without lowering the BLEU too much, then γ is set to 0.7 and change maximum sampling times K to observer the influence of hyper-parameter K.

In Table VI, when K increase from 1 to 400 and γ = 0.7, BLEU decreases from 34.30 to 33.35, and honorific generation ratio increases to 79.23% when K is set to 50, the honorific generation ratio increases by 34.41%, then both BLEU and honorific generation ratio will be basically stable.

Table V
ADJUSTMENT OF HYPER-PARAMETER γ

Model	BLEU	HG Ratio
Baseline	**34.21**	44.82%
HFT (γ=0.1, K=10)	32.07	**84.76%**
HFT (γ=0.3, K=10)	32.93	81.58%
HFT (γ=0.5, K=10)	33.23	79.18%
HFT (γ=0.7, K=10)	33.62	69.57%
HFT (γ=0.9, K=10)	34.17	54.95%

Table VI
ADJUSTMENT OF HYPER-PARAMETER K

Model	BLEU	HG Ratio
Baseline	34.21	44.82%
HFT (γ=0.7, K=1)	**34.30**	44.49%
HFT (γ=0.7, K=5)	33.82	64.56%
HFT (γ=0.7, K=10)	33.62	69.57%
HFT (γ=0.7, K=50)	33.57	**79.23**%
HFT (γ=0.7, K=100)	33.45	77.59%
HFT (γ=0.7, K=200)	33.36	78.11%
HFT (γ=0.7, K=400)	33.35	78.06%

As shown in Table V and Table VI, when hyper-parameter K is fixed, if γ becomes bigger, the proportion of MLE loss becomes larger, the BLEU will be higher and the honorific generation ratio will be lower because the model is more likely to notice the word associated with NTS reference and pays less attention to generate honorifics; when γ is fixed as the sample size K become bigger, the honorific generation ratio will be higher, but the BLEU will be a little bit lower in NTS but higher in HTS as shown in table III.

V. CONCLUSION

In this paper, we have introduced two training strategies to improve the honorific generation ratio in Chinese-Korean translation from the perspective of model and data respectively. The first strategy adopts honorific fusion training loss under MRT framework to guide NMT. The second strategy adopts data labeling which will label the source sentences with corresponding honorific labels of the target sentences. Experiments show both strategies can increase honorific generation ratio significantly.

REFERENCES

[1] Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Computation. 9(8), 1735–1780 (1997)

[2] Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014)

[3] Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems (2017)

[4] Shen, S., Cheng, Y., He, Z., He, W., Hua, W., Sun, M., Liu, Y.: Minimum risk training for neural machine translation. In: Proceedings of ACL 2015, pp. 1683–1692 (2015)

[5] Wu, Y., Schuster, M., Chen, Z., et al.: Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:16008144 (2016)

[6] Gehring, J., Auli, M., Grangier, D., et al.: Convolutional sequence to sequence learning. arXiv preprint arXiv:17003122 (2017)

[7] Koehn, P., Knowles, R.: Six challenges for neural machine translation. arXiv preprint arXiv:17003872 (2017)

[8] He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T., Ma, W.: Dual learning for machine translation. In: Proceedings of NIPS 2016 (2016)

[9] Artetxe, M., Labaka, G., Agirre, E., et al.: Unsupervised neural machine translation. arXiv preprint arXiv:17111041 (2017)

[10] Lample, G., Ott, M., Conneau, A., Denoyer, L., Ranzato, M.: Phrase-based & neural unsupervised machine translation. arXiv preprint arXiv:1804.07755 (2018)

[11] Xia, Y., Tian, F., Wu, L., Lin, J., Qin, T., Yu, N., Liu, T.: Deliberation networks: Sequence generation beyond one-pass decoding. In Proc. of NIPS (2017)

[12] Freitag, M., Al-Onaizan, Y.: Beam search strategies for neural machine translation. CoRR, abs/1702.01806 (2017)

[13] Wu, L., Xia, Y., Zhao, L., Tian, F., Qin, T., Lai, J., Liu, T.: Adversarial neural machine translation. arXiv preprint arXiv:1704.06933 (2017)

[14] Prabhumoye, S., Tsvetkov, Y., Salakhutdinov, R., Black, W.: Style Transfer Through Back-Translation. In Proceedings of ACL (2018)

[15] Jhamtani, H., Gangal, V., Hovy, E., Nyberg, E.: Shakespearizing modern language using copyenriched sequence-to-sequence models. arXiv preprint arXiv:1707.01161 (2017)

[16] Shen, T., Lei, T., Barzilay, R., Jaakkola, T.: Style transfer from non-parallel text by cross-alignment. In: Advances in Neural Information Processing Systems. (2017)

[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al.: Generative adversarial nets. In NIPS (2014)

[18] Edunov, S., Ott, M., Auli, M., Grangier, D.: Understanding back-translation at scale. arXiv preprint arXiv:1808.09381 (2018)

[19] Hinton, G., and Salakhutdinov, R.: Reducing the dimensionality of data with neural net-works. Science, 313 (5786), 504–507 (2006)

[20] Wang, Y., Zhang, J., Zhai, F., Xu, J., Zong, C.: Three Strategies to Improve One-to-Many Multilingual Translation , Association for Computational Linguistics, (2018)

[21] Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)

[22] Chatterjee, S., Cancedda, N.: Minimum error rate training by sampling the translation lattice. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Pro-cessing, pages 606–615. Association for Computational Linguistics (2010)

[23] Yin, F.: A contrastive Study on the honorific forms in Korean and Chinese, Jilin University (2016)

[24] Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, pp. 1715–1725 (2016)

[25] Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–3 Association for Computational Linguistics (2002)

[26] Klein, G., Kim, Y., Deng, Y., Senellart, J., Rush, A.: OpenNMT: Open-Source Toolkit for Neural Machine Translation. In Proceedings of ACL (2017)

[27] Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:146980 (2014)