Article

Context-Aware Neural Machine Translation for Korean Honorific Expressions

Yongkeun Hwang [1], Yanghoon Kim [1] and Kyomin Jung [1,2,*]

1 Department of Electrical and Computer Engineering, Seoul National University, Seoul 08826, Korea; wangcho2@snu.ac.kr (Y.H.); ad26kr@snu.ac.kr (Y.K.)
2 Automation and Systems Research Institute, Seoul National University, Seoul 08826, Korea
* Correspondence: kjung@snu.ac.kr

Abstract: Neural machine translation (NMT) is one of the text generation tasks which has achieved significant improvement with the rise of deep neural networks. However, language-specific problems such as handling the translation of honorifics received little attention. In this paper, we propose a context-aware NMT to promote translation improvements of Korean honorifics. By exploiting the information such as the relationship between speakers from the surrounding sentences, our proposed model effectively manages the use of honorific expressions. Specifically, we utilize a novel encoder architecture that can represent the contextual information of the given input sentences. Furthermore, a context-aware post-editing (CAPE) technique is adopted to refine a set of inconsistent sentence-level honorific translations. To demonstrate the efficacy of the proposed method, honorific-labeled test data is required. Thus, we also design a heuristic that labels Korean sentences to distinguish between honorific and non-honorific styles. Experimental results show that our proposed method outperforms sentence-level NMT baselines both in overall translation quality and honorific translations.

Keywords: automatic post-editing; context-aware machine translation; Korean machine translation; Korean honorifics; neural machine translation

1. Introduction

Neural machine translation (NMT) has shown impressive results on translation quality, due to the availability of vast parallel corpus [1], and the introduction of novel deep neural network (DNN) architectures such as encoder-decoder model [2,3], and self-attention based networks [4]. The performance of NMT systems has reached on par with human translators in some domains, and hence many commercial MT services, such as Google Translation, have adopted NMT as their backbone of translation systems [5].

Despite the significant improvement over the previous machine translation (MT) systems, NMT still suffers from language-specific problems such as Russian pronoun resolution [6] and honorifics. Addressing such language-specific problems is crucial in both personal and business communications [7] not only because the preservation of meaning is necessary but also many of these language-specific problems are also closely related to their culture. *Honorifics* are good example of these language-specific problems that conveys respect to the audience. In some languages including Korean, Japanese, and Hindi that use honorifics frequently, speaking the right honorifics is considered imperative in those languages.

In Korean, one of the most frequent usages of honorifics occurs in the conversation with people who are in superior positions, or elders [8]. As is shown in Figure 1, the source English sentence "Wait a minute, please.", which is the second utterance by the son, is translated into the target sentence "잠시만 기다려요." (jam-si-man gi-da-lyeo-yo) that is represented as *haeyo-che* (해요체) as the sentence ends with -요 (-yo). Haeyo-che is a type of Korean honorific reflecting the relationship between the two speakers.

Addressing such honorifics in MT is challenging since the definition of honorifics differs across different languages. For example, Korean has 3 major types of honorifics [8] and corresponding honorific expressions. In contrast, it is known that English has fewer types of honorifics compared to many other languages [9]; only titles, such as Mr. and Mrs., are frequently used in modern English. It is known that managing honorifics in translation is comparatively more complicated in English-Korean translation; the source language has a simpler honorific system compared to the target language. The source language with fewer honorifics provides fewer honorific features that are used to generate correct honorifics in the target side, as shown in Figure 1. Since the English verb "wait" can be translated into both the honorific style (기다려요, gi-da-lyeo-yo) and the non-honorific style (기다려, gi-da-lyeo), the model cannot determine the adequate honorific solely depending on the source sentence, and additional information is necessary such as the relationship between speakers.

Sentence	English	Korean
context_1	Come on, **dad**. Don't you even take something?	**아빠**, 뭐라도 안 드세요?
context_0	Okay, give me some coffee.	좋아, 그럼 커피 좀 줘.
source/target	Wait a minute, please.	잠시만 기다려요.

Figure 1. An example of Korean dialogue that is extracted from subtitles. The blue words are verbs that translated into polite form whereas the red words are impolite form, using Korean honorifics.

In this paper, we propose a novel method to remedy limitations from solely depending on source sentence by using *context*, which is represented by the surrounding sentences of the source sentence. In Figure 1, we can infer that this is a dialogue between a son and his father from the content of context_1, and the source sentence. Therefore, the model can determine that the source sentence should be translated into a polite sentence using honorifics, such as *haeyo-che* (해요체), if such context is taken into account.

To this end, we introduce a *context-aware* NMT to incorporate the context for improving Korean honorific translation. It is known that the context-aware NMT can improve the translation of words or phrases that need contextual information, such as pronouns that are sensitive to the plural and/or gender [10]. Considering above example that how the adequate honorific style can be determined using the context, we suggest that the context-aware NMT can also be used to aid the honorific-aware translation. To the best of our knowledge, this work is the first attempt to utilize context-aware NMT for honorific-aware translation.

We consider two types of context-aware NMT framework in our proposed method. First, we use a contextual encoder that takes context in addition to the source sentence as input. The encoder captures contextual information from the source language that is needed to determine target honorifics. Second, a context-aware post-editing (CAPE) system is adopted to take the context of translated target sentences for refining the sentence-level translations accordingly.

To demonstrate the performance of our method, an honorific-labeled parallel corpus is needed so we also developed simple and fast rule-based honorific annotation for labeling the test data. In the experiments, we compared our context-aware systems with context-agnostic models and we show that our method outperformed the context-agnostic baselines significantly in both the overall translation quality and translation of honorifics.

We hope that our proposed method improves the overall quality of Korean NMT and thus expanding the real-world use of NMT in communicating with Korean. Adequate use of honorifics can greatly improve the overall quality of Korean translations, especially in spoken language translation (SLT) systems. We suggest that MT systems for applications like movie/TV captioning, chatting can be benefited from our method.

Our contributions can be summarized in threefolds:

- We show that the NMT model with a contextual encoder improves the quality of the honorific translation regardless of the model structure. In our experiments, even the simplest model that concatenates all the contextual sentences with the source sentence can improve honorific accuracy. We also show that the NMT model with contextual encoder also outperforms the sentence-level model even when the model is explicitly controlled to translate to a specific honorific style.
- In addition to the contextual encoder, we demonstrate that the CAPE can improve honorifics of both the sentence-level NMT and contextual NMT by exploiting contextual sentences of the target language. Our qualitative analysis also reveals the ability of CAPE to improve the inconsistent use of honorifics of the NMT model with a contextual encoder.
- We also develop an automatic data annotation heuristics for labeling Korean sentences as honorific and non-honorific style. Our heuristics utilize Korean morphology to precisely determine the honorific style of a given sentence. We labeled our test set by using the heuristics and used it to validate the improvements of our proposed method.

The remaining part of this paper consists as follows: We briefly review the related works in Section 2 and introduce Korean honorifics in Section 3. Context-aware NMT methods are presented in Section 4. We introduce our methods in Section 5 then show the experimental results in Section 6. Finally we present our conclusion in Section 7.

2. Related Works

2.1. Neural Machine Translation

The NMT represents the translation directly via a DNN. This is different from the traditional methods such as statistical MT (SMT), which consists of a number of sub-components such as the translation model and the language model. Generally, most of the NMT models consist of two parts, one that takes the source sentence and the other that generates the target sentence, with each of them represented as a sequence of vectors. This framework is the so-called the encoder-decoder or sequence-to-sequence model [2,3]. The model is then trained on the parallel corpus, which consists of many pairs of the source and the target sentence.

The early NMT methods were composed of recurrent neural networks (RNNs) such as long short-term memories (LSTMs) [3]. Recently, the attention mechanism [11] has made a breakthrough in the field of NMT. It summarizes the sequence of a vector by finding the most relevant part of the sequence to the input. In the early stage, the attention mechanism was widely used as a sub-component of the model by attending to the encoded source sentence [11]. Recently, the Transformer [4] has exploited the attention mechanism as a backbone of the model, which consists of an attentional network followed by a feedforward network. This greatly improved the translation quality compared to other RNN-based methods with attention, and the Transformer is now widely used as a base of NMT and many other natural language processing (NLP) methods. In addition to the architectural improvement, the more sophisticated training methods such as back-translation [12] and the language model (LM) pretraining (e.g., BERT [13], MASS [14], and BART [15]) have also been studied to further improve the translation quality.

There have been a number of MT studies involving Korean. Because parallel corpora containing Korean are not as widely available as English and many European languages, a number of the existing works focused on low-resource MT settings. For example, Xu et al. [16] exploited out-of-domain and multilingual parallel corpora, and Jeong et al. [17] applied LM pretraining and back-translation. In addition, some other works have attempted to develop additional techniques to overcome the limitations of common low-resource MT methods. For example, Nguyen et al. [18] incorporated morphological information and word-sense disambiguation (WSD) on Korean source sentences to improve the translation into Vietnamese. Park et al. [19] focused on beam search decoding and experimented with various decoding settings including beam size to improve

translation quality without re-training the target NMT model. Although low-resource MT methods are out of scope in this paper, some methods including back-translation are closely related with our methods in training CAPE.

2.2. Controlling the Styles in NMT

Although the style of a generated translation also affects the quality of the machine translation, it has received little attention in the field of NMT. Since the source sentence contains insufficient information of the output style, most of the existing works have introduced a set of special tokens [20]. For example, to control the formality of the target sentence, one can add <F> at the beginning of the source sentence to translate formally or add <I> to translate informally. The model can attend to this token and extract the relevant linguistic features on training. This approach has been adopted in many subsequent works such as [21,22]. Some other works have addressed this problem as domain adaptation that treats each style as a domain [23] or adopted multitask learning of the machine translation and the style transfer problem to address the lack of a style-annotated parallel corpus [24], but the output is still controlled by the special tokens. By contrast, our approach can improve the honorific translation without using such kinds of special tokens by exploiting the contextual information of the surrounding text. In addition, our method can be combined with the methods using special tokens to further improve the accuracy of honorifics.

On the other hand, a few kinds of grammatical styles have addressed the style-controlled MT. The English formality [25] or the T-V distinction in European languages such as Spanish [7] are two common examples. Viswanathan et al. [7] have addressed the control of T-V distinction such as the use of a formal/informal form of second-person pronouns (*usted* vs. *tú*), as domain adaptation. Niu et al. [25] has shown that employing syntactic supervision can improve the control of English formality. Furthermore, few studies have addressed the honorifics of Asian languages such as Japanese [26] and Korean [22]. Wang et al. [22] used data labeling and reinforcement learning (RL) to enhance translation of Korean honorifics. However, they ignored contextual sentences and only relied on special tokens to control the honorifics.

2.3. Context-Aware NMT

The context-aware MT models focus on contextual information in the surrounding text [27] and either the context of the source or the target sentence can be considered. Exploiting the source side of contexts usually implements an additional encoder to represent the multiple contextual sentences efficiently [6,28,29]. On the other hand, the target-side contexts can be exploited by first translating a part of documents or discourses at the sentence level and then refining those translations. This can be implemented either by the use of multi-pass decoding or automatic post-editing (PE). The multi-pass decoder generates the translation at the sentence level first and then translates again by regarding the translated sentences as contexts [30,31]. On the other hand, the context-aware PE corrects the common and frequent errors of sentence-level models by considering both the target sentence and its contexts [10]. We choose to use both sides of contexts; the source side of context helps to choose suitable honorifics in the target sentence, whereas the target-side context is helpful in correcting inconsistencies of honorific translations since we are focusing on the honorifics in the target language.

On the other hand, many of the context-aware MT studies have been focused on improving pronoun resolutions such as choosing the correct gender or plural for pronouns. For example, Voita et al. [6,10] have addressed the translation of Russian, and Müller et al. [32] are focused on German pronoun resolution. To the best of our knowledge, our work is the first attempt to use context-aware NMT to control grammatical styles such as honorifics.

3. Addressing Korean Honorifics in Context

In this section, we present an overview of the Korean honorifics system and how the contextual sentence can be used to infer appropriate honorifics for translation.

3.1. Overview of Korean Honorifics System

Asian languages such as Korean, Japanese, and Hindi are well-known as having rich honorific systems to express formality distinctions. Among those languages, the use of honorifics is extensive and also crucial in Korean culture. In practice, Korean speakers are forced to choose appropriate honorifics in every utterance, and failing to do that can induce serious social sanctions including school expulsion [8]. Moreover, it is known that Korean honorific systems are very sophisticated among the well-known languages thus teaching how to use Korean honorifics appropriately is also considered challenging in Korean as a Second Language (KSL) education [8,33]

There are three types of Korean honorifics; subject honorification, object honorification, and addressee honorification. First, in the subject honorification, the speaker honors the referent by using honorific suffixes such as '-시-'(-si-), case particles such as '-께서'(-kke-seo), and so on:

1. 철수가 방에 들어가다. (cheol-su-ga bang-e deul-eo-ga-da; Cheolsoo goes to the room.)
2. 어머니께서 방에 들어가신다. (eo-meo-ni-kke-seo bang-e deul-eo-ga-sin-da; My mother goes to the room.)

In contrast to (1), the speaker's 어머니 (eo-meo-ni, mother) in (2) is honored by the following case particle '께서' (-kke-seo) and the honorific suffix '-신-' (-sin-) at the verb 가다 (ga-da; *go*). Second, object honorification is used when the referent of the object is of higher status (e.g., elder) to both the speaker and referent of the subject:

(a) 철수는 잘 모르는 것이 있으면 항상 **아버지**께 여쭌다. (*cheol-su*-neun jal mo-leu-neun geos-i iss-eu-myeon hang-sang **a-beo-ji**-kke yeo-jjun-da; Cheolsoo always ask his father about something that he doesn't know well.)
(b) **아버지**는 휴대폰에 대해 잘 모르는 게 있으면 항상 **철수**에게 묻는다. (**a-beo-ji**-neun hyu-dae-pon-e dae-hae jal mo-leu-neun ge iss-eu-myeon hang-sang *cheol-su*-e-ge mud-neun-da; Cheolsoo's father always ask him about mobile phones that he doesn't know well.)

In the example (a), Cheolsoo's **아버지** (a-beo-ji, father) is in the superior position both to 철수 (Cheolsoo) and the speaker. Therefore, '여쭌다' (yeo-jjun-da), which is an honorific from of the verb 묻는다 (mud-neun-da, ask), is used.

Finally, addressee honorifics are expressions of varying speech levels that are used to show politeness or closeness and are usually expressed as sentence endings in Table 1.

Despite that all 6 examples are translated as the same English sentence, each example has its own levels of formality and politeness and different usages. For example, '반말체' (*banmal-che*) and '해라체' (*haela-che*) are used between people with close relationships or used by the elderly when speaking to younger people. Conversely, '해요체' (*haeyo-che*) and '합쇼체' (*hapsio-che*) are used to honor the addressees and express politeness [8].

Table 1. Speech levels and sentence endings in Korean. Names are translated with respect to [8]. Each of the example sentences are a translation of: The weather is cold. Each underllined sentence ending corresponds to their addressee honorific.

Style and Name	Politeness	Formality	Example
합쇼체 (Hapsio-che; Deferential)	High	High	날씨가 춥습니다. nal-ssi-ga chub-<u>seub-ni-da</u>
해요체 (Haeyo-che; Polite)	High	Low	날씨가 추워요. nal-ssi-ga chu-wo-<u>yo</u>
하오체 (Hao-che; Semiformal)	Neutral	High	날씨가 춥소. nal-ssi-ga chub-<u>so</u>
하게체 (Hagae-che; Familiar)	Neutral	Low	날씨가 춥네. nal-ssi-ga chub-<u>ne</u>
반말체 (Banmal-che; Intimate)	Low	High	날씨가 추워. nal-ssi-ga chu-<u>wo</u>
해라체 (Haela-che; Plain)	Low	Low	날씨가 춥다. nal-ssi-ga chub-<u>da</u>

3.2. The Role of Context on Choosing Honorifics

As stated earlier, the relationship between speaker and audiences affects the use of Korean honorifics. For example, the student should use *haeyo-che* and *hapsio-che* as addressee honorifics when asking a teacher some questions. Since such social context is often reflected in utterances, readers may infer the relationship from text without knowing who are speakers and/or audiences.

In the Figure 1, we can infer that the source and the contextual sentence is consist of a dialogue between a dad and a son and the context_1 and the source sentence is utterances of the son, so the source English sentence should be translated into a polite Korean sentence as shown.

Figure 2 shows two another examples in our dataset. In (a), a dialogue between a person (context_0) and his/her superior (context_1). So their Korean translations are in polite (*haeyo-che*) and impolite (*banmal-che*) respectively. In addition, we can infer that the source sentence is also an utterance by the same person who told (context_0) as we can find the same pronoun <u>we</u> to refer themselves. So the sentence endings of translation should be as "중독 됐어요" (jung-dog dwaess-eo-**yo**) which has the same honorifics as context_0, instead of using *banmal-che*, such as "중독 됐어" (jung-dog dwaess-**eo**).

On the other hand, (b) shows the usage of *hapsio-che* which is frequently used for formal expressions in context_0 and the source sentence, as both of the sentences are ending with '-ㅂ니다' (-b-nida). The word *suspect* (용의자, yong-ui-ja) in context_0 give us a hint that the context_0 is told by police officers, prosecutors etc since the word is frequently used by those occupations. We can also infer that this dialogue is not held between those officers from the pronoun <u>you</u>, rather the utterances are told to a witness, etc. So the context_0 and the source sentence would be translated into formal Korean utterances, rather than informal sentences like "우린 빨리 제시카를 찾아야 **해**" (u-lin ppal-li je-si-ka-leul chaj-a-ya **hae**).

As shown in the examples, contextual sentences often have important clues for choosing appropriate honorifics in Korean translation. However prior approaches for honorific-aware NMT including [26] for Japanese, and [22] for Korean have ignored those contexts. Instead, they explicitly controlled the model to translate the source sentence into a specific honorific style, using special tokens for indicating the target honorific as [20].

Sentence	English	Korean
context_1	You're back.	자네들 또 왔구만.
context_0	Yes, **sir**, we are.	예, **어르신**.
source/target	We're addicted to **your** citrus.	**어르신**의 감귤류에 중독 됐어요.

(a)

Sentence	English	Korean
context_1	You need to relax, okay?	진정해 주실래요?
context_0	You are not a **suspect**.	당신은 **용의자**가 아닙니다.
source/target	We should find Jessica right now.	저희는 빨리 제시카를 찾아야만 합니다.

(b)

Figure 2. Two examples of Korean dialogue from our dataset, which are extracted from subtitles. The blue words are verbs that translated with polite and/or formal honorifics whereas the red words are translated with impolite and/or informal honorifics. The bold keywords are used to determine what types of honorifics should be used. The underlined pronouns indicate that the two utterances is told by the same speaker in (**a**) and the utterances are formal speech in (**b**).

4. Context-Aware NMT Frameworks

To utilize the contextual sentences in NMT, we introduce the context-aware NMT systems. These are divided into two categories: contextual encoders on NMT models and a CAPE system. Here we briefly review those systems before explaining our proposed method.

4.1. NMT Model with Contextual Encoders

Generally, NMT models are operated at the *sentence level*; it takes an input sentence in a source language and returns an output sentence in a target language. On the other hand, a contextual encoder in NMT is designed to handle one or more contextual sentences as input and extract the contextual representation. In our settings, NMT models are based on the Transformer [6], which is based on a stack of attentional networks. Each hidden layer in the Transformer consists of a self-attention mechanism followed by feedforward networks. Because of its performance and efficiency, the Transformer has been widely used in NMT, and many improvements have also been made including contextual encoders. We list five Transformer-based models in our experiments:

- **Transformer without contexts (TwoC)**: As a baseline, we have experimented with the TwoC model which has the same structure as [4]. TwoC does not use any contextual sentences and only incorporates the input and the target sentences.
- **Transformer with contexts (TwC)**: This is the simplest approach to incorporate contextual sentences with the Transformer [27]. TwC concatenates all contextual sentences and an input sentence and considers the concatenated sentence as a single-input sentence. Then, the output of the TwoC encoder is the output of a stacked Transformer encoder with concatenated source and contextual sentences.
- **Discourse Aware Transformer (DAT) [6]**: DAT handles a single contextual sentence with an extra context encoder that is also a stacked Transformer encoder. To handle multiple contextual sentences, we slightly modified DAT such that the contextual encoder takes a concatenation of contextual sentences. The context encoder has the same structure as the source encoder and even shares its weights. Encoded contextual sentences are integrated with an encoded source sentence by using a source-to-context attention mechanism and a gated summation.

- **Hierarchical Attention Networks (HAN) [28]**: HAN has a hierarchical structure with two-stage at every hidden layer in their contextual encoder. At the first level of the hierarchy, HAN first encodes each of the contextual sentences to sentence-level tensors using the stacked Transformer encoder as in [4]. Then each encoded sentence is summarized by word-level context-source attention, resulting in sentence-level representations. These sentence-level vectors are concatenated and again encoded with sentence-level context-source attention. Finally, encoded contextual sentences are integrated using a gated summation.

- **Hierarchical Context Encoder (HCE) [34]** HCE also exploits a similar hierarchical structure as HAN but uses different method to summarize word-level and sentence-level information. In the lower part of the hierarchy, the encoded sentence-level tensor is compressed into a sentence-level vector by a self-attentive weighted sum module, which is similar to that of [35]. The collection of sentence-level vectors is fed into another Transformer encoder layer that is the upper part of the hierarchy to encode the entirety of contextual information into a single tensor. Finally, the contextual information tensor is combined with the source encoder in a similar fashion as DAT.

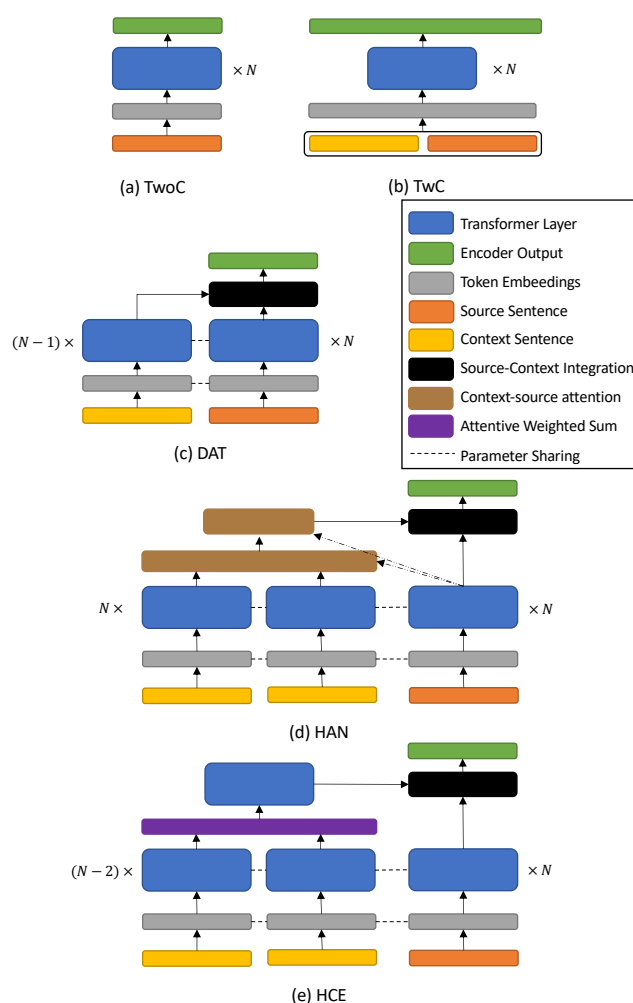All the model structures are described in Figure 3.

Figure 3. The structure of compared contextual encoders; (**a**) TwoC (**b**) TwC (**c**) DAT (**d**) HAN and (**e**) HCE.

4.2. Context-Aware Post Editing (CAPE)

CAPE is a variant of automatic post-editing (PE) systems (e.g., Vu et al. [36]). The PE fixes systematic errors that frequently occur in a specific machine translation system. Most of the PE operates at the sentence level; however, Voita et al. [10] suggested using PE to

correct inconsistencies between sentence-level translations of a context-agnostic MT system. Analogous to many existing PE systems, the CAPE itself is independent of a specific MT model and can therefore in principle be trained to correct translations from any black-box MT system including a context-aware NMT system.

The training and testing process of CAPE is illustrated in Figure 4. First, the translation inconsistency of the target NMT model is simulated by using a round-trip translation. For example, to refine an English to Korean NMT system, Korean sentences are translated into English using Korean to English NMT first; then, they are again back-translated into Korean with a target English to Korean NMT system. In this way, the errors of the NMT model can be represented as the difference and inconsistency between the original Korean sentences and its round-trip translations. Once these round-trip translations are prepared, the CAPE, which consists of a typical sequence-to-sequence model, is trained to minimize these gaps. At test time, the target NMT system translates each sentence first, and then the CAPE takes a group of such translations and produces fixed translations. Moreover, CAPE has been shown to improve the English to Russian translation of context-sensitive pronouns [10] such as deixis and ellipsis.

Figure 4. (**a**) Training a CAPE model requires a monolingual, discourse-/document-level corpus. Each consecutive text is segmented into a set of sentences first. Then, each sentence is translated and then back-translated. The resulting sentence group is concatenated again, and then the CAPE, which consists of a sequence-to-sequence model, is trained to minimize the errors of these round-trip translations. (**b**) At test time, a trained CAPE fixes sentence-level translations by taking them as a group.

5. Our Proposed Method—Context-Aware NMT for Korean Honorifics

In this section, we describe our proposed approach to generate appropriate Korean honorific expressions with context-aware NMT. We propose the use of context-aware NMT for translation of the honorific-styled sentence, which can improve the translation of honorifics without explicit control as done with special tokens. We also developed an

automatic honorific labeling method to label the parallel corpus so that evaluation of the honorific translations, and preparing training data when the system is allowed to control target honorifics as in [22]. The process of our proposed method is illustrated in Figure 5.

5.1. Using NMT with Contextual Encoder and CAPE for Honorific-Aware Translation

To capture contextual information that affects the use of Korean honorifics, our method exploits the context-aware models in two ways, as described in Section 4.

The first one is an NMT model with a contextual encoder (Section 4.1), which is trained to capture the dependency between the contents of contextual sentences of the source language and the usage of honorific expressions represented in the training data. For example, in Figure 1, the model can attend the noun *dad* in the context_1 to generate a translation in *haeyo-che*. In this way, the trained model can implicitly control the translation to generate appropriate honorific expressions according to the contextual sentences. In the experiments, we compare this approach against the NMT models that explicitly control the translation honorifics by introducing special tokens as in [22].

The second one is a CAPE (Section 4.2) for improving the inconsistent sentence-level translation of honorifics. As stated earlier, the CAPE is trained by recovering inconsistent round-trip translations that require a pretrained bidirectional sentence-level MT model. Therefore, we first train a TwoC model to translate both Korean-English and English-Korean using the same parallel corpus. Then, we sample round-trip translations from a separately constructed monolingual Korean corpus and train a CAPE to reconstruct the original Korean sentence from the sampled round-trip translations, as illustrated in Figure 4. Our CAPE model is implemented using the same Transformer model as the TwoC [4], so once the monolingual corpus and its round-trip translations are prepared, training CAPE is similar to training a TwoC. We also apply the CAPE to improve the NMT models with contextual encoders, such as HCE. Despite that the CAPE was originally intended to correct the errors of sentence-level MT similar to TwoC [10], it can complement the NMT with a contextual encoder. Importantly, the CAPE exploits the context information of the target language, and some types of inconsistency, such as inter-sentence disagreement of honorifics, can only be identified in the target language. In the experiments, we show that the CAPE can further improve the honorific translation of HCE as well by correcting the inconsistency of honorifics between sentences.
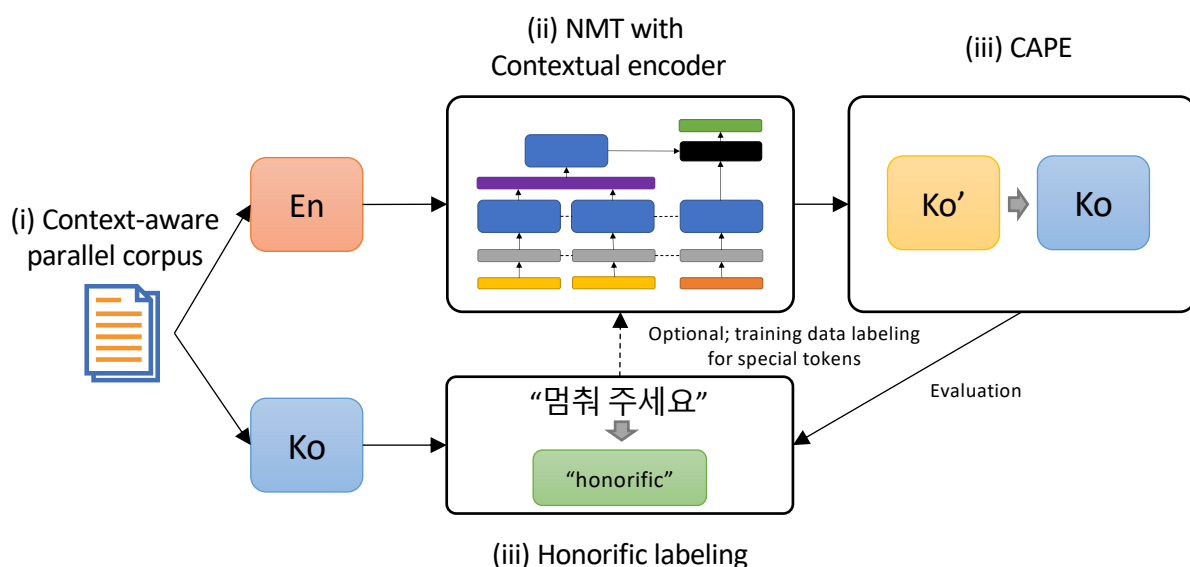

Figure 5. The process of our method, context-aware NMT for Korean honorifics. First we train NMT model with contextual encoder for English-Korean and Korean-English translation. Then we train CAPE to correct errors on those round-trip translations made by the NMT model. The automatic honorific labeling is primarily used for assessing honorific translation, but can also be used to label the training set if the NMT model uses special tokens to control target honorifics explicitly.

5.2. Scope of Honorific Expressions

Our work focuses on the translation of addressee honorifics, which is a key factor in determining whether the sentence is *honorific style*. From the 6 types of sentence endings in Table 1, the *haeyo-che* and *hapsio-che* are usually considered honorific styles that are used frequently by age–rank subordinates speaking to superiors [8,22]. Thus, we consider sentences having these two types of endings as *honorific sentences*, while others are *non-honorific* sentences. The target sentence in Figure 1 "잠시만 기다려**요**" (jam-si-man gi-da-lyeo-**yo**) whose ending is *haeyo-che*, is an example of an honorific sentence. In contrast, "잠시만 기다리**게**" (jam-si-man gi-da-li-**ge**) is a non-honorific sentence that is translated the same as in English according to our criteria since its ending is *hagae-che*.

5.3. Automatic Honorific Labeling

To assess the quality of honorific translation, we need to annotate the corpus into *honorific sentence* vs. *non-honorific* sentences. We developed heuristics using the above criteria to label the Korean sentences with honorific styles.

As illustrated in Figure 6, we first segment sentences into morphologies and obtain their part-of-speech (POS) tags. This ensures that our heuristic can correctly identify the proper sentence ending. In our implementation, the Kkma Korean tagger [37] is used to extract morphologies and POS tags. Once morphologies and POS tags are extracted, we then select *eomi* (어미) which is the sentence ending. We picked morphologies whose tag starts with 'EF' (http://kkma.snu.ac.kr/documents/index.jsp?doc=postag) in our implementation. We label sentences as honorific if their *eomi* is *hapsio-che* or *haeyo-che*. In some cases where the morphology tagger fails to extract word endings, we resort to sub-string matching with sentence-ending markers such as '?', or '.' to correctly extract the proper sentence ending.

This heuristic is used primarily to label the test set for evaluation of our method; however, it can also be used to label the training set for training NMT models with explicit control of honorifics. In this case, the honorific label is used to generate a special token if the translation honorific of the model is controlled by a special token.
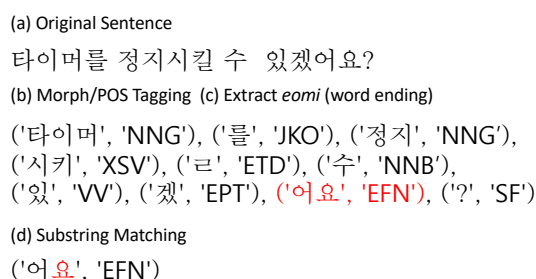
(a) Original Sentence

타이머를 정지시킬 수 있겠어요?

(b) Morph/POS Tagging (c) Extract *eomi* (word ending)

('타이머', 'NNG'), ('를', 'JKO'), ('정지', 'NNG'),
('시키', 'XSV'), ('ㄹ', 'ETD'), ('수', 'NNB'),
('있', 'VV'), ('겠', 'EPT'), ('어요', 'EFN'), ('?', 'SF')

(d) Substring Matching

('어요', 'EFN')

Figure 6. Tagging sentences into honorific or nonhonorific styles. The original sentence (a) "타이머를 정지시킬 수 있겠어요?" (ta-i-meo-leul jeong-ji-si-kil su iss-gess-eo-yo; *Can you shut off the timer?*) is segmented into morphologies with their part-of-speech (POS) tags. Then we use 'eomi's to classify the sentence.

6. Experiments

To verify how the context-aware models improve Korean honorifics in English-Korean translation, we conduct comprehensive experiments and analyses on how context-aware MT models translate Korean honorifics. First, we constructed an English-Korean parallel corpus with contextual sentences. Then, we train and compare the models described in Section 4. Finally, a qualitative analysis is conducted on some examples from our proposed method.

6.1. Dataset and Preprocessing

To the best of our knowledge, there are no English-Korean discourse-level or context-aware parallel corpora that are publicly available. Thus, we constructed an English-Korean

parallel corpus with contextual sentences. We took an approach similar to [34] by choosing to use bilingual English-Korean subtitles of movies and TV shows because these subtitles contain many scripts with honorific expressions.

We first crawled approximately 6100 subtitle files from websites such as GomLab.com. Then, we split these files into training, development, and test sets, which consist of 5.3*k*, 500, and 50 files, respectively. We applied a file-based split to make sure that contextual sentences are only extracted from the same movie/episode. Unlike other datasets such as OpenSubtitles2018 [38], our subtitle files contain both English and Korean sentences, so extracting bilingual sentence pairs is straightforward; we used timestamp-based heuristics to obtain those pairs. The resulting sentence pairs are 3.0*M*, 28.8*k*, and 31.1*k* pairs for training, development, and test sets, respectively. Some of the raw samples from our test sets are shown in Figure 7.

The contextual sentences are selected by using the timestamp of each subtitle, which contains the start time and end time in milliseconds. We assume that the sentences contain contextual information if they appear within a short period of time before the source sentence. Specifically, the start time of a contextual sentence is within *K* milliseconds from the start time of the source sentence. We set *K* as 3000 heuristically, and the maximum number of preceding contextual sentences is 2 for all experiments except those of Section 6.4.2. The final data contains 1.6*M*, 155.6*k*, and 18.1*k* examples of consecutive sentences in the training, development, and test sets, respectively.

For monolingual data to train the CAPE, we added 2.1*M* Korean sentences using an additional 4029 crawled monolingual subtitles. The resulting monolingual data consist of 5.1*M* sentences.

We finally tokenized the dataset using the wordpiece model [5], and the size of the vocabulary is approximately 16.5*k*. We also put a special token <BOC> at the beginning of contextual sentences to differentiate them from the source sentences.

Start (ms)	End (ms)	English	Korean
		...	
646819	649786	Gives us a view inside the house.	집 안 모습을 보여줍니다.
649786	652102	We have a clear heat signature of the occupants.	안에 있는 사람의 열이 분명히 감지되고 있습니다.
652102	655129	Size, stance, says male. We think he's armed.	몸집, 자세를 보면 남자입니다. 무기를 소지한 것 같군요.
655129	656482	Any sign of a woman?	여자는 없나요?
656482	659841	Second body, right here, prone, giving off heat.	두 번째 사람, 바로 저기요, 누워 있는데, 열이 감지됩니다.
		...	

Figure 7. Example parallel sentence pairs extracted from bilingual subtitles.

6.2. Model Training Details

For NMT models, we use model hyperparameters, such as the size of hidden dimensions and the number of hidden layers as the `transformer-base` [4], since all of the models in our experiment share the same Transformer structure. Specifically, we set 512 as the hidden dimension, the number of layers is 6, the number of attention heads is 8, and the dropout rate is set to 0.1. These hyperparameters are also applied to the CAPE model. For NMT models with additional encoders (DAT, HCE), we share the weights of encoders.

All models are trained with ADAM [39] with a learning rate of 1e-3, and we employ early stopping of the training when loss on the development set does not improve. We trained all of the models from scratch with random initialization, and we do not pretrain the model on a sentence-level task as in [22,28]. All the evaluated models are implemented by using the `tensor2tensor` framework [40].

6.3. Metrics

We measure the translation quality by BLEU scores [41]. For scoring BLEU, we use the `t2t-bleu` script (https://github.com/tensorflow/tensor2tensor/blob/master/tensor2

tensor/bin/t2t-bleu) which yields the same scores as Moses [42]. We first measure BLEU scores with original translations and we refer to these scores as *normal* BLEU scores. In addition, we also measure *tokenized* BLEU scores by tokenizing translations prior to scoring BLEU, as a common practice in the evaluation of Korean NMT [43].

For honorifics, we set the accuracy of honorifics as the ratio of translations with the same type of honorific style with respect to the reference translations. For example, if the reference translation of an English sentence "Yeonghee is cleaning." is "영희가 청소**해요**." (yeong-hui-ga cheong-so-**hae-yo**; *haeyo-che - honorific*) and the model translation is "영희가 청소**한다**." (yeong-hui-ga cheong-so-**han-da**; *banmal-che - non-honorific*), the translation is considered inaccurate.

6.4. Results

First, overall BLEU scores and honorific accuracy are compared among MT models with various types of contextual encoders. We also examine the varying performance of these models with respect to the number of contextual sentences and effects of CAPE for improving honorific translations.

6.4.1. Effect of Contextual Encoders

To evaluate the effect of contextual information on the translation of Korean honorifics, we first measure the performances of context-agnostic and context-aware models. The results are summarized on Table 2. As shown in the results, all the context-aware models (TwC, DAT, HAN, and HCE) outperform the context-agnostic model (TwoC) in terms of BLEU. The HCE shows a significant English-Korean BLEU improvement over TwoC of approximately 1.07/2.03 and the TwC, DAT, and HAN also show slight improvements. We later use Korean-English TwoC and HCE trained in this experiment for generating round-trip translations on CAPE experiment since the HCE performed best among the context-aware models in terms of BLEU. We also experimented with the models on Korean-English BLEU using the same dataset for comparison. All the context-aware models again outperformed the context-agnostic model in this experiment. Note that BLEU scores are lower in all English-Korean experiments compared to Korean-English BLEU in the same dataset. This is mainly due to the morphological-rich nature of Korean and the domain of the dataset, which consists of spoken languages.

Table 2. English<->Korean BLEU scores and accuracy (%) of honorifics for context-agnostic (TwoC) and context-aware (TwC, DAT, and HCE) NMT models. English-Korean BLEU scores are shown as (normal/tokenized) respectively. All the models are trained and tested without any honorific labels or explicit control of honorifics.

Models	BLEU		Accuracy	Accuracy
	En-Ko	**Ko-En**	**All Test Set**	**Polite Targets**
TwoC	9.16/12.45	23.81	64.34	39.27
TwC	9.6/13.2	24.35	66.85	**44.08**
DAT [6]	9.36/12.98	23.96	65.12	38.7
HAN [28]	9.50/13.08	24.54	66.3	42.26
HCE [34]	**10.23/14.75**	**26.63**	**67.94**	42.42

In addition to the BLEU scores, the context-aware models are also better in translation with correct Korean honorifics in English-Korean translation. In particular, the HCE has improved the honorific accuracy by 3.6%. Since showing politeness is considered important in Korean culture as discussed in Section 3.1, we also focus on the accuracy of the test sets which are polite target sentences. The TwC outperformed all other models in this set up to 4.81% compared to TwoC. The HAN and HCE also showed significant improvement over TwoC, while the DAT's accuracy is slightly lower than that of TwoC. We believe that such differences derive from how the model utilizes contextual information. Since we only use the sequence-level cross-entropy (CE) as a training objective, the more compact

representations of contextual encoders in DAT, HAN, and HCE can improve the main objective (translation quality), but considering the raw information of contextual sentences as in TwC could be more beneficial to honorific translation.

On the other hand, all of the results in Table 2 are from models that do not have any explicit control of honorifics and do not employ the honorific-annotated dataset. For comparison with prior works that forced the model to translate with specific honorifics as [22], we also include the results of NMT models with special tokens for controlling output honorifics in Table 3. In particular, the TwoC with special tokens is the same as the data labeling (DL) method in [22]. The training set was labeled the same as the test set, with the method described in Section 5.3. As shown in the results, both models are able to translate almost all the test examples with the same honorifics as their references, which is a similar result to that in [22]. Interestingly, both controlled models also improve the translation quality over their counterparts without control, and the HCE with special tokens again outperformed TwoC with special tokens on BLEU.

Table 3. English-Korean BLEU scores (normal/tokenized) and accuracy (%) of honorifics for models with explicit control of honorifics by special tokens on the input. All the models are forced to obtain the translation with the honorific style of the reference sentence.

Models	BLEU	Accuracy All Test Set	Accuracy Polite Targets
TwoC + Special Token	9.36/12.68	99.46	98.91
HCE + Special Token	**10.83/14.79**	**99.49**	**99.04**

In summary, the context-aware NMT models can improve not only the translation quality but also the accuracy of honorifics. While their improvements are less significant compared to the honorific-controlled models, they can nevertheless exploit the contextual information to aid in the correct translation of honorifics.

6.4.2. Effect of the Number of Contextual Sentences

The number of contextual sentences has a significant effect on the model performance since not all the contextual sentences are important in obtaining an adequate translation [44]. Such redundant information can hurt the performance. Since this number is dependent on the model and the data, we carry out experiments to examine the effect of the number of contextual sentences. As shown in Table 4, both the BLEU and accuracy of honorifics are the best on 2 contextual sentences, and then they decay as the number increases. Similar effects are also shown by the other context-aware NMT models, as displayed in Table 5.

Table 4. English-Korean BLEU scores (normal/tokenized) and accuracy (%) by the number of contextual sentences on HCE

# Contextual Sents.	BLEU	Accuracy All Test Set	Accuracy Polite Targets
1	9.23/12.88	65.42	40.31
2	**10.23/14.75**	**67.94**	**42.42**
3	9.83/13.49	66.56	41.93
4	9.31/12.92	64.8	39.27
5	8.98/12.09	63.3	36.48

Table 5. English-Korean BLEU scores (normal/tokenized) and accuracy (%) by the number of contextual sentences on all of the context-aware NMT models

Models	# Contextual Sents.	BLEU	Accuracy All Test Set	Accuracy Polite Targets
TwC	2	9.6/13.2	66.85	44.08
	5	8.23/11.41	61.21	38.05
DAT	2	9.36/12.98	65.12	38.7
	5	8.02/11.2	60.94	33.2
HAN	2	9.5/13.08	66.3	42.26
	5	8.55/11.74	63.1	36.6
HCE	2	10.23/14.75	67.94	42.42
	5	8.98/12.09	63.3	36.48

6.4.3. Effect of CAPE

Finally, scores with or without context-aware postediting (CAPE) are provided in Table 6. The CAPE improved TwoC by 0.87/1.93 on BLEU and outperformed TwC and DAT on honorific accuracies by approximately 3 to 4%. The improvement in honorific accuracy suggests that CAPE can also repair the inconsistency of honorifics. We additionally applied CAPE to HCE. The result shows that HCE with CAPE also outperformed the vanilla HCE, supporting our hypothesis.

Table 6. English-Korean BLEU scores (normal/tokenized) and accuracy (%) of honorifics for models with/without CAPE.

Models	BLEU	Accuracy All Test Set	Accuracy Polite Targets
TwoC	9.16/12.45	64.34	39.27
+CAPE	**10.03/14.38**	**67.5**	**43.81**
HCE	10.23/14.65	67.94	42.42
+CAPE	**10.55/15.03**	**69.16**	**46.51**

6.5. Translation Examples and Analysis

We show some translation examples in Figures 8 and 9. As discussed in Section 5, the honorific sentences are mostly used when a subordinate such as a child is talking to superiors such as his/her parents. Figure 8 shows two examples of these situations. In (a), context and source sentences are a conversation between a mother and her child. This can be speculated from the contextual sentences; the child is talking but the mom urges him/her to continue eating. The TwoC completely ignores the contextual sentences, so such a situation is not considered. Thus, TwoC translates the source sentence as a non-honorific style using the non-honorific sentence ending 때 (ttae), which is *banmal-che*. In contrast, the translation of HCE is an honorific sentence since its sentence ending is 요 (yo), which is *haeyo-che*, the same as the reference. This is an example that shows HCE's context-awareness that helps translation of honorific-styled sentences.

En (Context_1)	Life must go on as it always has. (언제나처럼 인생은 계속되어야죠.)
En (Context_0)	Come on, let's eat. (어서 먹자.)
En (Source)	How's mom?
Ko (TwoC)	엄마는 어때?
Ko (HCE)	엄마는 어떠세요?
Ko (Reference)	엄마는 어때요?

(a)

En (Context_1)	Daddy! (아빠!)
En (Context_0)	Hey, I'm here. (어, 나 여기 있어.)
En (Source)	I'm sorry. I should have listened to **you**.
Ko (TwoC)	미안해. **네** 말을 들었어야 했는데.
Ko (HCE)	미안해. **네** 말을 들었어야 했어.
Ko (TwC)	죄송해요. **아빠** 말을 들었어야 했는데요.
Ko (Reference)	미안해요. **아빠** 말을 들을 걸 그랬어요.

(b)

Figure 8. Example translations of different NMT models. The sentences are given in a sequence, from context_1 to source. The reference translation of each contextual sentence is given in (). In (**a**), a mother and her child are talking to each other. The context-aware model (HCE) can infer this situation using contextual sentences and translate the source sentence with an appropriate honorific style. Similarly, in (**b**) a dad and his child are talking, but only a translation from TwC has the correct honorific style. Note that translations of the verb sorry and the 2nd person pronoun **you** also differ among models despite that all the translations have the same meaning as the source sentence.

On the other hand, *Daddy!* in context_1 of (b) and the content of context_1 directly indicate that the source sentence is spoken by a dad's child. Despite such direct hints, HCE failed to correctly identify the proper honorific style, resulting in *banmal-che* (해 (hae) and 어 (eo)). However, the TwC correctly translated the source sentence as an honorific sentence using *haeyo-che* (해요 (haeyo) and 데요 (daeyo)). Note that there are two sentence segments in the source and translations, and the honorific style of the two segments agrees in all the model translations and the reference. One interesting observation is that TwC has translated verb sorry as 죄송-하다 (joesong-hada) instead of 미안-하다 (mian-hada) and the 2nd person pronoun **you** as 아빠 (appa; daddy) instead of 네 (ne; you) like HCE. As the former is resulting as a more polite translation and the latter is closer to the reference so this example can be viewed as a clue that TwC's context-awareness is better than that of HCE. We suggest that TwC's simple and direct use of contextual sentences can perform better than the abstract representation of contextual sentences in HCE when the contextual sentences are simple and short.

En	My <u>condolences</u>. / Skip the sympathy. This is a <u>business</u>. / My father met with you right before he died.
Ko (TwoC)	<u>조의</u>를 표합니다. / 조문은 필요없어. 이건 <u>사업</u>이야. / 아버지가 죽기 직전에 널 만났어.
Ko (HCE)	<u>명복</u>을 빕니다. / 조문은 필요없어. 이건 <u>사업</u>이야. / 아버지가 죽기 직전에 당신을 만났어요.
Ko (HCE+CAPE)	<u>명복</u>을 빕니다. / 조문은 필요없어요. 이건 <u>사업</u>이에요. / 아버지께서 돌아가시기 직전에 당신을 만났어요.
Ko (Reference)	고인의 <u>명복</u>을 빕니다. / 조문은 필요없어요. 이건 <u>비즈니스</u>입니다. / 아버지께서 돌아가시기 직전에 당신을 만났죠.

Figure 9. Example of a translation made by HCE and its correction by CAPE. The second and third sentence segments are the utterance of the same speaker. HCE's translations are inconsistent in honorifics since honorifics of the second and third segments do not agree. The CAPE successfully corrected that inconsistency. Note that CAPE also fixed the subject honorification, resulting in a more polite translation. Note that the underlined nouns are differ among models, despite that all the translations have the same meaning.

Finally, Figure 9 shows how the CAPE corrects the inconsistent use of honorifics. These 3 sentence segments are obtained from a scene held in a funeral home. Considering the content of the sentences, we can assume that the 2nd and 3rd segments are the utterances of the same speaker. However the honorific styles of HCE translations do not agree on *banmal-che* for the 2nd segment and *haeyo-che* for the 3rd. CAPE corrected this inconsistency by looking at the translated Korean sentences. In addition, CAPE also amended the 3rd sentence segment by modifying the subject honorification, replacing both the case particle for the subject (his father) from -가 (-ga) to 께서 (-kkeseo) and the verb 죽기 (jukgi) to 돌아가시기 (doragasigi); both are translated as *died*. Considering that a deceased person is generally highly honored in Korean culture, the CAPE's correction results in a more polite and thus adequate honorific-styled sentence. Although the subject honorification is out of scope in this paper, this shows the CAPE's ability to capture various honorific patterns observed in the training corpus and correct translations.

7. Conclusions

In this paper, we have introduced the use of context-aware NMT to improve the translation of Korean honorifics. By using contextual encoders, the context-aware NMT models can implicitly capture the speaker information and translate the source sentence with proper honorific style. In addition, context-aware postediting (CAPE) is adopted to improve the honorific translation by correcting the inconsistent use of honorifics between sentence segments. Experimental results show that our proposed method can improve the translation of Korean honorifics compared to context-agnostic methods both in BLEU and honorific accuracy. We also demonstrated that the use of context-aware NMT can further improve the prior methods which use special tokens to control honorifics translation. Qualitative analysis on sample translations supports the effectiveness of our method on exploiting contextual information for improving translations of honorific sentences. In the future, we will extend our method to other Asian languages such as Japanese, and Hindi which also has complex and widely-used honorifics system.

References

1. Esplà, M.; Forcada, M.; Ramírez-Sánchez, G.; Hoang, L. ParaCrawl: Web-scale parallel corpora for the languages of the EU. In Proceedings of the MT Summit XVII, Dublin, Ireland, 19–23 August 2019; pp. 118–119.

2. Cho, K.; Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734.

3. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS), Montréal, Canada, 8–13 December 2014; pp. 3104–3112.

4. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.

5. Wu, Y.; Schuster, M.; Chen, Z.; Le, Q.V.; Norouzi, M.; Macherey, W.; Krikun, M.; Cao, Y.; Gao, Q.; Macherey, K.; et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. [Online]. September 2016. Available online: https://arxiv.org/abs/1609.08144 (accessed on 30 June 2021).

6. Voita, E.; Serdyukov, P.; Sennrih, R.; Titov, I. Context-Aware Neural Machine Translation Learns Anaphora Resolution. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 15–20 July 2018; pp. 1264–1274.

7. Viswanathan, A.; Wang, V.; Kononova, A. Controlling Formality and Style of Machine Translation Output Using AutoML. In *Annual International Symposium on Information Management and Big Data*; Springer: Cham, Switzerland, 2019; pp. 306–313.

8. Brown L. Questions of appropriateness and authenticity in the representation of Korean honorifics in textbooks for second language learners. *Lang. Cult. Curric.* **2010**, *23*, 35–50.

9. Xiao, Z.; McEnery, A. Two approaches to genre analysis: Three genres in modern american english. *J. Engl. Linguist.* **2005**, *33*, 62–82.

10. Voita, E.; Sennrich, R.; Ivan, T. Context-Aware Monolingual Repair for Neural Machine Translation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 877–886.

11. Bahdanau, D.; Cho, K.; Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv* **2014**, arXiv:1409.0473.

12. Sennrich, R.; Haddow, B.; Birch, A. Improving Neural Machine Translation Models with Monolingual Data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL), Berlin, Germany, 7–12 August 2016; pp. 86–96.

13. Devlin, J.; Chang, M-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.

14. Song, K.; Tan, X.; Qin, T.; Lu, J.; Liu, T-Y. MASS: Masked Sequence to Sequence Pre-training for Language Generation. In Proceedings of the 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 10–15 June 2019; pp. 5926–5936.

15. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *arXiv* **2019**, arXiv:1910.13461.

16. Xu, G.; Ko, Y.; Seo, J. Low-resource Machine Translation by utilizing Multilingual, Out-domain Resources. *J. KIISE* **2019**, *40*, 649–651.

17. Jeong, Y.; Park, C.; Lee, C.; Kim, J.s. English-Korean Neural Machine Translation using MASS. In Proceedings of the 31st Annual Conference on Human & Cognitive Language Technology (HCLT), Busan, Korea, 11–12 October 2019; pp. 236–238.

18. Nguyen, Q-P.; Anh D.; Shin, J.C.; Tran, P.; Ock, C.Y. Korean-Vietnamese Neural Machine Translation System With Korean Morphological Analysis and Word Sense Disambiguation. *IEEE Access* **2019**, *7*, 32602–32616, doi: 10.1109/ACCESS.2018.2851281.

19. Park, C.; Yang, Y.; Park, K.; Lim, H. Decoding Strategies for Improving Low-Resource Machine Translation. *MDPI Electron.* **2020**, *9*, 1562–1577.

20. Sennrich, R.; Haddow, B.; Birch, A. Controlling Politeness in Neural Machine Translation via Side Constraints. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), San Diego, CA, USA, 12–17 June 2016; pp. 35–40.

21. Chu, C.; Dabre, R.; Kurohashi, S. An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL), Vancouver, Canada, 30 July–4 August 2017; pp. 385–391.

22. Wang, L.; Tu, M.; Zhai, M.; Wang, H.; Liu, S.; Kim, S.H. Neural Machine Translation Strategies for Generating Honorific-style Korean. In Proceedings of the 2019 International Conference on Asian Language Processing (IALP), Shanghai, China, 15–17 November 2019; pp. 450–455.

23. Michel, P.; Neubig, G. Extreme Adaptation for Personalized Neural Machine Translation. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia, 15–20 July 2018; pp. 312–318.

24. Niu, X.; Rao, S.; Carpuat, M. Multi-Task Neural Models for Translating Between Styles Within and Across Languages. In Proceedings of the 27th International Conference on Computational Linguistics (COLING), Santa Fe, NM, USA, 20–26 August 2018; pp. 1008–1021.

25. Niu, X.; Carpuat, M. Controlling Neural Machine Translation Formality with Synthetic Supervision. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020.

26. Feely, W.; Hasler, E.; Gispert, A. Controlling Japanese Honorifics in English-to-Japanese Neural Machine Translation. In Proceedings of the 6th Workshop on Asian Translation (WAT), Hong Kong, China, 4 November 2019; pp. 45–53.

27. Tiedemann, J.; Scherrer, Y. Neural Machine Translation with Extended Context. In Proceedings of the Third Workshop on Discourse in Machine Translation (DiscoMT), Copenhagen, Denmark, September 8 2017; pp 82–92.

28. Miculicich, L.; Ram, D.; Pappas, N.; Henderson, J. Document-Level Neural Machine Translation with Hierarchical Attention Networks. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium, 31 October–4 November 2018; pp. 2947–2954.

29. Maruf, S.; Martins, A.; Haffari, G. Selective Attention for Context-aware Neural Machine Translation. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, USA, 2–7 June 2019; pp. 3092–3102.

30. Xiong, H.; He, Z.; Wu, H.; Wang, H. Modeling Coherence for Discourse Neural Machine Translation. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HW, USA, 27 January–1 February 2019.

31. Voita, E.; Sennrich, R.; Titov, I. When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), Florence, Italy, 28 July–2 August 2019; pp. 1198–1212.

32. Müller, M.; Rios, A.; Voita, E.; Sennrich, R. A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation. In Proceedings of the Third Conference on Machine Translation (WMT), Brussels, Belgium, 31 October–1 November 2018; pp. 61–72.

33. Byon, A.S. Teaching Korean Honorifics. *Korean Lang. Am.* **2000**, *5*, 275–289.

34. Yun, H.; Hwang, Y.; Jung, K. Improving Context-Aware Neural Machine Translation Using Self-Attentive Sentence Embedding. In Proceedings of the AAAI Conference on Artificial Intelligence, New York City, NY, USA, 7–12 February 2020.

35. Lin, Z.; Feng, M.; Yu, M.; Xiang, B.; Zhou, B.; Bengio, Y. A structured self-attentive sentence embedding. In Proceedings of 5th International Conference on Learning Representations (ICLR), Toulon, France, 24-26 April 2017.

36. Vu, T.; Haffari, G. Automatic Post-Editing of Machine Translation: A Neural Programmer-Interpreter Approach. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 3048–3053.

37. Lee, D.; Yeon, J.; Hwang, I; Lee, S. KKMA : A Tool for Utilizing Sejong Corpus based on Relational Database. *J. Kiise Comput. Pract. Lett.* **2010**, *16*, 1046–1050.

38. Lison, P.; Tiedemann, J.; Kouylekov, M. OpenSubtitles2018: Statistical Rescoring of Sentence Alignments in Large, Noisy Parallel Corpora. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, 7–12 May 2018.

39. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.

40. Vaswani, A.; Bengio, S.; Brevdo, E.; Chollet, F.; Gomez, A.; Gouws, S.; Jones, L.; Kaiser, Ł.; Kalchbrenner, N. Parmar, N.; et al. Tensor2Tensor for Neural Machine Translation. March 2018. Available online: http://arxiv.org/abs/1803.07416 (accessed on 30 June 2021).

41. Papineni, P.; Roukos, S.; Ward, T.; Zhu, W. BLEU: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 6–12 July 2002; pp. 311–318.

42. Koehn, P.; Hoang, H.; Birch, A.; Callison-Burch, C.; Federico, M.; Bertoldi, N.; Cowan, B.; Shen, W.; Moran, C.; Zens, R.; et al. Moses: Open Source Toolkit for Statistical Machine Translation. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL) Demo and Poster Sessions, Prague, Czech Republic, 24–29 June 2007; pp. 177–180.

43. Jung, Y.; Park, C.; Lee, C.; Kim, J. English-Korean Neural Machine Translation using MASS with Relative Position Representation. *J. Kiise* **2020**, *47*, 1038–1043.

44. Kang, X.; Zhao, Y.; Zhang, J.; Zong, C. Dynamic Context Selection for Document-level Neural Machine Translation via Reinforcement Learning. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 2242–2254.