香港公共衛生資訊的神經網絡翻譯質量評估（英譯漢）

蕭世昌
恒生管理學院翻譯學院
蕭世恩
香港大學公共衛生學院

摘要

神經機器翻譯把神經網絡及深度學習應用在自動翻譯中，近年取得矚目成就，獲得廣泛應用，如谷歌、微軟、Systran 及搜狗。此技術也可用於健康資訊翻譯，促進有關資訊的跨語言傳遞，這對於雙語或多語言社會的公共衛生至為重要；香港奉行中英雙語為法定語言，即為一例。今天神經機器翻譯系統日增，本文旨在評測神經機器翻譯在翻譯香港公共衛生資訊的水平。本文集中比較谷歌翻譯、微軟翻譯及搜狗翻譯三大知名翻譯系統，首先概述公共衛生資訊、神經機器翻譯及有關翻譯系統，然後利用 BLEU 方法對三個系統進行自動評估，用於評測的原文取自由香港衛生署發出的新聞稿。本研究也會採用譯文排序的方式作人工評估。基於評估結果，本文針對公共衛生資訊的神經機器翻譯，在系統設計及應用兩方面提出改善建議。研究成果期望有助於加深了解神經機器翻譯在公共衛生領域之應用。

關鍵詞：公共衛生資訊　神經機器翻譯　翻譯質量評估　翻譯科技　深度學習

On the Quality of Neural Machine Translation of Public Health Information in Hong Kong (English-Chinese Translation)

SIU Sai Cheong
School of Translation, Hang Seng Management College

SIU Sai Yan
School of Public Health, The University of Hong Kong

Abstract

Neural machine translation, which involves the application of neural networks and deep learning to automatic translation, has achieved state-of-the-art results and has been widely adopted in recent years. Given the promising performance of the technology, we could apply neural machine translation systems to the translation of health information, which plays a crucial role in public health in bilingual/multilingual societies, such as Hong Kong, where both Chinese and English are the official languages. Considering an increasing number of neural machine translation systems on the market, we aim to assess the quality of neural machine translation of public health information in Hong Kong by evaluating the output of three neural machine translation systems: Google Translate, Microsoft Translator and Sogou Translator. This paper first gives an overview of public health information, neural machine translation, and the three translation systems. This is followed by an automatic evaluation of the systems on press releases published by the Department of Health in Hong Kong in the last two months, with the use of the BLEU automatic evaluation metric. In addition, there is a complementary human evaluation by adopting the translation ranking method. Based on the evaluation results, we will suggest possible ways to improve the development and use of neural machine translation systems for the translation of public health information. It is hoped that this present work will help enhance the understanding of neural machine translation in the public health domain.

Keywords: public health information, neural machine translation, machine translation evaluation, translation technology, deep learning

1. Background

Timely and accurate translation of public health materials plays an important role in promoting health in bilingual or multilingual communities, such as Hong Kong, where not only Chinese and English but also ethnic minorities' languages such as Hindi, Nepali and Urdu are spoken. The use of automatic translation tools may be useful in this regard, especially after the recent rapid development of neural machine translation systems, which have reportedly outperformed their predecessors, such as statistical machine translation systems, and have been adopted by major developers of online machine translation tools, including Google (Wu et al., 2016), Microsoft (Microsoft Translator, 2016), Systran (Crego et al., 2016), and Sogou (Wang et al., 2017).

Despite the growing popularity of neural machine translation, there have been few studies of its application to the public health domain, except a few papers on the feasibility of statistical machine translation and its incorporation into the translation procedure. This paper therefore aims to study neural machine translation of public health materials in Hong Kong by selecting three online machine translation engines, namely Google Translate (n.d.), Microsoft Translator (2018a) and Sogou Translator (2018), and focusing on the evaluation of their machine translation outputs of selected press releases published by the Department of Health.

The structure of the paper is as follows: Section 2 discusses the publication of public health information, with a focus on press releases published by the Department of Health, and the global research trend in machine translation for public health. Section 3 explains the latest developments in the field of neural machine translation. Section 4 conducts automatic and manual evaluation of the three neural machine translation engines. Section 5 studies common issues in the machine outputs and suggests ways to make better use of neural machine translation in the public health domain.

2. Public Health Information and Research on its Machine Translation

Public health departments play a pivotal role in disseminating accurate and comprehensible health promotion messages to the public (Regidor et al., 2007), which is essential for not only educating the public on important health issues, but also ensuring the health of the nation at large (Maylahn Fleming & Birkhead, 2013).

Public health information is available in different formats including pamphlets, posters, press releases, web pages, TV announcements, video clips and even mobile applications. In Hong Kong, for example, such information is primarily made available by the Department of Health and the Hospital Authority. The former is the "Government's health adviser and agency to execute health policies and statutory functions" (Chan, 2016) and its major service areas include health education, surveillance and epidemiology; many units have been set up within the Department, such as the Centre for Health Protection, Drug Office, Primary Care Office, and Office for Regulation of Private Healthcare Facilities. The latter is responsible for the management of public hospital and is accountable to the Hong Kong Government through the Secretary for Food and Health (Hospital Authority, 2018). In addition to government departments and statutory bodies, private medical organizations and hospitals also provide public health information (see, for example, Hong Kong Sanatorium & Hospital (n.d.)).

Given the wide variety of health promotion materials, ranging from environmental health, infection control and clinical service to primary care, the production of quality cross-lingual public health materials could be costly and time-consuming. In the light of the rapid development of machine translation technology, the global public health research community has an increasing interest in automatic translation of public health information. For example, Kirchhoff et al. (2011) investigated the feasibility of translating health promotion materials with the use of a freely available statistical machine translation website. Gathering documents from the websites of public health agencies in the United States and comparing the machine-assisted and human translations of the documents, the authors revealed the overall equivalency between machine-translated and manually translated materials. However, as the study only analyzed the translations of documents from English to Spanish, it did not conclude whether the proposed approach was applicable to other languages. To assess the feasibility of machine translation involving other language pairs, Turner et al. (2015) conducted follow-up research to examine the performance of Google Translate in translating English health promotion documents into Chinese. Experiments showed that more work was needed to improve the translation quality, given the many problems in the automatic translation results, such as errors of word sense and word order. The authors suggested that the less satisfactory results were due to the very divergent syntactic structures and linguistic differences between English and Chinese.

3. New Possibilities Offered by Neural Machine Translation

Technically speaking, there are different approaches to machine translation, such as example-based machine translation by using translation examples (Nagao, 1984) and statistical machine translation by developing language and translation models (Brown et al., 1988; Brown et al., 1990; Brown, S. Della Pietra, V. Della Pietra, & Mercer, 1993; Koehn, Och, & Marcu, 2003). In recent years, we have seen the growing popularity of neural machine translation, one of the latest approaches to machine translation that involves deep learning, the use of layers of artificial neurons (Cho, Merriënboer, Bahdanau, & Bengio, 2014; Firat, Cho, & Bengio, 2016).

Neural machine translation generally consists of three steps. The first step is word embedding: Each token in the source text is mapped on to a vector, which is a sequence of numbers attempting to capture the meaning of the token concerned. The second step is encoding: Based on the results of word embedding, a neural machine translation system encodes the source text and converts it to a sequence of vectors, an internal representation of the text, using multiple layers of artificial neurons. The third step is decoding: The internal representation is decoded, and the most probable tokens in the target language are identified for the generation of the target text, again with the involvement of multiple layers of artificial neurons.

More specifically, there are ways to implement a neural machine translation system, including the use of recurrent neural networks (such as a network consisting of long-short term memory cells or gated recurrent units) and the attention mechanism (Bahdanau, Cho, & Bengio, 2014; Luong, Pham, & Manning, 2015; Wu et al., 2016); the use of convolutional neural networks (Gehring et al., 2017); and the use of self-attentional transformers (Vaswani et al., 2017).

From the discussion above, it is apparent that neural machine translation takes into consideration not only word similarity (through word embedding), but also a wider context of

the source text (sentences instead of just individual words and phrases) (Diño, 2017). Probably because of these features, neural machine translation has outperformed its predecessors such as statistical phrase-based machine translation in many machine translation projects (see Bojar et al., 2016 for some examples).

As a result, recent years have witnessed rapid growth of neural machine translation systems. Google announced its Google Neural Machine Translation system in 2016 (Wu et al., 2016). Microsoft launched its neural machine translation engine in the same year (Microsoft Translator, 2016) and announced in March 2018 that they had "developed an AI machine translation system that can translate with the same accuracy as a human from Chinese to English" (see Microsoft Translator, 2018b and Hassan et al., 2018). Other online machine translation providers in Mainland China, including Tencent (2018) and Sogou (2018), have also incorporated neural machine translation technology into their systems. Sogou, for example, built neural machine translation systems for the WMT 2017 news translation tasks, and ranked first among 20 systems in the Chinese-English translation task and ranked third among 16 systems in the English-Chinese translation task (Wang et al., 2017).

4. Evaluation of Neural Machine Translation of Public Health Press Releases

To have a better idea of the quality of neural machine translation of public health information published in Hong Kong, we focus on bilingual press releases published by the Department of Health, which are available on the following websites: http://www.dh.gov.hk/english/press/press.html (in English) and http://www.dh.gov.hk/tc_chi/press/press.html (in Traditional Chinese). They are published on a need basis and mainly cover the following public health issues: updates on cases of infectious diseases, food poisoning clusters, preventive measures against diseases, follow-up investigations in response to detection of virus, identification of problematic cosmetic, medical or pharmaceutical products or devices, and health promotion activities.

We created a dataset of 20 randomly selected press releases published between January and March 2018. For the sake of the diversity of the selected texts, we studied the articles selected and replaced similar ones (with only slight differences such as dates and numbers) with additional articles randomly selected. The selected texts consist of 9,472 tokens in English and cover topics such as Middle East Respiratory Syndrome (MERS), dengue fever, oral health, food poisoning, slimming products, influenza B infection, precautions against cold weather, Legionnaires' disease, undeclared Western drug ingredients in powder prescribed by a Chinese medicine practitioner, avian influenza, and acute gastroenteritis.

We translated the selected press releases by using three neural machine translation engines: Google Translate, Microsoft Translator and Sogou Translator. Google Translate supports over 100 languages and offers text translation (neural machine translation for languages including Chinese), speech translation, image translation, and offline translation (Google, n.d.). Similarly, Microsoft Translator provides text translation for more than 60 languages supported (neural machine translation for 21 languages including Chinese), speech translation for 10 languages, and image translation (Microsoft Translator, 2017). Sogou Translator is the first commercial translation system that is fully based on neural machine translation in Mainland China and supports some 60 languages (Sogou, 2018b).

Prior to translation, the press releases in English were first divided into sentences, and they were aligned with the corresponding sentences extracted from the official press releases in Chinese for the automatic evaluation to be discussed below (note that some of the sentences in English/Chinese, after alignment, may comprise two or more sentences because of sentence combination; the use of "sentences" in the remaining sections of this paper refers to post-alignment sentences, which could be single sentences or consist of multiple sentences). For the evaluation of Google Translate and Microsoft Translator, the English sentences had been stored in an HTML file before they were translated by the two engines with their web page translation function. For Sogou Translator, since its web page translation feature did not directly support English-Traditional Chinese translation at the time of evaluation, the English press releases were translated into Simplified Chinese using the translation interface on the main page of the tool, and the translation results were converted to Traditional Chinese with Microsoft Translator.

We conducted automatic evaluation of the machine translation outputs by adopting the Bilingual Evaluation Understudy (BLEU) metric (Papineni, Roukos, Ward, & Zhu, 2002), which has been widely adopted in the machine translation research community. The BLEU score is a measure of n-gram precision with a value between 0 and 1, comparing a machine translation output with a reference translation (i.e. the aligned Chinese sentences from the official press releases in our case). Formally, the BLEU score is computed as follows:

$$BLEU = \min(1, e^{1-\frac{r}{c}}) \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

where c is the total length of the test set in words, r is the total length of the reference translation in words, p_n is the modified n-gram precision, w_n is the weight of the precision, and N refers to the maximum length of n-grams to be considered.

We computed the following BLEU scores: (1) sentence-based BLEU scores for individual sentences in the target language using the "add-1" smoothing technique (i.e. adding 1 to the number of matched n-grams and total number of n-grams for $n \geq 2$; see Lin & Och (2004), as cited in "Smoothing 2" in Chen & Cherry (2014)) and (2) overall BLEU scores of all the translated sentences with the smoothing method that combines length-scaled pseudo-counts and match interpolation (see "Smoothing 7" in Chen & Cherry (2014)). Chinese sentences, including the official version and the online translations, were tokenized with Jieba (2017), and the BLEU scores were computed with Natural Language Toolkit (NLTK, 2017).

Table 1 summarizes the sentence-based BLEU scores in (1), and Table 2 shows their distributions (in percentage of sentences of the target text translated by each engine). Table 3 shows the overall BLEU scores in (2).

	Google	**Microsoft**	**Sogou**
1st Qu.	0.0514	0.0426	0.0538
Median	0.0729	0.0613	0.0800
Mean	**0.0879**	**0.0762**	**0.1008**
3rd Qu.	0.1150	0.0931	0.1261
Max	0.3976	0.3593	0.4518

Table 1: Summary of sentence-based BLEU scores

Sentence BLEU	Google (%)	Microsoft (%)	Sogou (%)
<0.05	24.27	35.73	20.22
0.05≤x<0.10	44.27	42.7	44.27
0.10≤x<0.15	19.55	12.81	19.33
0.15≤x<0.20	7.87	4.04	6.97
0.20≤x<0.25	1.35	1.8	3.15
0.25≤x<0.30	1.35	2.02	2.92
0.30≤x<0.35	0.22	0	0.9
0.35≤x<0.40	**1.12**	**0.9**	**1.8**
0.40≤x<0.45	0	0	0
≥0.45	0	0	**0.45**

Table 2: Distribution of sentence-based BLEU scores

	Google	Microsoft	Sogou
Overall BLEU with smoothing	0.2430	0.2319	**0.2470**

Table 3: Overall BLEU scores

As shown in Tables 1 and 2, Sogou Translator attained the highest average sentence-based BLEU score (0.1008, compared with 0.0879 of Google Translate and 0.0762 of Microsoft Translator) and a higher percentage of sentences with a score greater than 0.30 (3.15%, compared with 1.34% of Google Translate and 0.9% of Microsoft Translator) in the translation of the press release dataset. Sogou also attained the highest overall BLEU score (higher than Google Translate and Microsoft Translator by 0.004 and 0.151 respectively). In terms of the overall BLEU score, the difference between Sogou Translator and Google Translate (difference = 0.004) was smaller than that between Google Translate and Microsoft Translator (difference = 0.011).

To complement the BLEU evaluation, we conducted human evaluation by using 100 sentences randomly drawn from the dataset. We used the method of translation ranking (see, for example, Toral & Way (2018)), which has been a common way of conducting human assessment of automatic translation system in machine translation competitions in recent years, and invited a translation teacher to participate in the evaluation. For each of the 100 English sentences, he was given the corresponding official version in Chinese, together with the three machine translation outputs that were arranged randomly. He ranked the translated sentences by considering their quality and followed the instructions below, which were based on the ones proposed by Toral & Way (2018):

1. Evaluation method: Please consider whether all the information is preserved in the ST. If all translations convey the meaning to a similar extent, consider if they are natural and compare the number of errors in each translation.
2. If translation A is better than translation B, rank the first translation higher than the second (i.e. Translation A: rank 1 and Translation B: rank 2).
3. If Translation A and Translation B have equal quality, rank the two translations equally (rank 1 for both Translations A and B).

4. Use the highest rank possible. For example, given three translations A, B, C, if Translations A and B are of equal quality and better than C, then rank 1 should be given to both Translations A and B, and rank 2 to Translation C. Do not use lower rankings (i.e. rank 2 for A and B and rank 3 for C).

Table 4 shows the proportions of different ranks of translations of each of the translation engines.

	Google (%)	**Microsoft (%)**	**Sogou (%)**
Rank 1	38	22	**55**
Rank 2	35	**46**	35
Rank 3	27	**32**	10
Total	100	100	100

Table 4: Results of human ranking

The results of human ranking in Table 4 were consistent with the results of automatic evaluation. Sogou Translator ranked first in terms of the proportion of translations given rank 1 (Google Translate and Microsoft Translator ranked second and third respectively). As for the aggregate proportion of translations given ranks 2 or 3, Microsoft Translator ranked first, followed by Google Translate and Sogou Translator.

5. Translation Issues and Suggestions

The three systems sometimes provided readable outputs with little post-processing, especially translations with a BLEU score equal to or higher than 0.25. These translations tend to be short sentences, and here are some examples.

1. **Source Text:** Poultry should be cooked thoroughly. (Department of Health, 2018a)
   Google Translate: 家禽應徹底煮熟。
2. **Source Text:** • Ensure proper personal hygiene; (Department of Health, 2018b)
   Google Translate: •確保個人衛生;
3. **Source Text:** Investigations are ongoing. (Department of Health, 2018c)
   Microsoft Translator: 目前正在進行調查。
4. **Source Text:** Maintain good indoor ventilation; (Department of Health, 2018c)
   Microsoft Translator: 保持良好的室內通風;
5. **Source Text:** Read carefully the label instructions first; (Department of Health, 2018d)
   Sogou Translator: 先仔細閱讀標籤說明；
6. **Source Text:** All patients are in stable condition. (Department of Health, 2018e)
   Sogou Translator: 所有病人情況穩定。

However, there are still some common issues in the machine translation results. The issues are categorized and presented in Table 5. Relevant translation errors are marked with the symbol “(X)” for easy reference.

Translation Issue	Examples and Discussion
Issue 1: Abbreviations	**Source Text:** Update on latest **MERS** situation in Oman (Department of Health, 2018f) **Reference:** 阿曼中東呼吸綜合症最新情況(Department of Health, 2018g) **Google Translate:** 更新阿曼最新的 **MERS** 情況 **Microsoft Translator:** 阿曼**最新匯率情況(X)**的最新情況 **Sogou Translator:** 阿曼**市場匯率(X)**最新情況 **Discussion:** "MERS" is the abbreviation of "Middle East respiratory syndrome", but the first three letters "MER" in the abbreviation, which could refer to the "market exchange rate", might have caused confusion and led to the use of the expressions "最新匯率情況" and "市場匯率" in some of the translations.
Issue 2: Proper nouns	**Source Text:** People who have purchased the products may submit them to the **Drug Office** of the DH at Room 1856, **Wu Chung House**, 213 Queen's Road East, Wan Chai, during office hours for disposal. (Department of Health, 2018h) **Reference:** 已購買上述產品的市民可於辦公時間內將有關產品交予灣仔皇后大道東二一三號**胡忠大廈**一八五六室衞生署**藥物辦公室**銷毀。(Department of Health, 2018i) **Google Translate:** 購買產品的人士可於辦公時間將其送交灣仔皇后大道東 213 號胡忠大廈 1856 室衛生署**藥物辦事處(X)**處理。 **Microsoft Translator:** 購買該等產品的人士, 可于辦公時間內, 在灣仔皇后大道東 213 號**五湧樓(X)**1856 室向**弥鍍 ' p 禁毒處(X)**提交。 **Sogou Translator:** 市民如購買該等藥物，可于辦公時間內，將該等藥物交回灣仔皇后大道東 213 號胡忠大廈 1856 室衛生署**毒品調查科(X)**處理。 **Discussion:** Proper nouns such as "Drug Office" and "Wu Chung House" are mistranslated into "藥物辦公室"/"弥鍍 ' p 禁毒處"/"毒品調查科" (instead of "藥物辦公室" in the reference) and "五湧樓" (instead of "胡忠大廈").
Issue 3: Word sense	**Source Text:** He **attended** the Accident and Emergency Department (AED) of Alice Ho Miu Ling Nethersole Hospital yesterday (March 7) and was **admitted** today. (Department of Health, 2018j) **Reference:** 他於昨日（三月七日）**前往**雅麗氏何妙齡那打素醫院急症室**求診**，並於今日**入院**。(Department of Health, 2018k) **Google Translate:** 他於昨日（三月七日）**出席(X)**雅麗氏何妙齡那打素醫院急症室，並於今日**受理(X)**。 **Microsoft Translator:** 他昨日 (3 月 7 日) **出席(X)愛麗絲(X)**何妙齡那打素醫院急症室 (AED), 並於今日入院。 **Sogou Translator:** 他昨日(三月七日)前往雅麗氏何妙齡那打素醫院急症室求診，並於今日入院。 **Discussion:** The expressions "出席" and "受理" do not fit the context and should be replaced with "前往……求診" and "入院" respectively. In addition, the proper noun "Alice" should be translated

	into “雅麗氏” instead of “愛麗絲”.
Issue 4: Numbers	**Source Text:** According to the latest information, 2 144 cases have been reported to the WHO (with at least 750 deaths), including **1 927 in 10 Middle East countries comprising 1 769 in the Kingdom of Saudi Arabia, 86 in the United Arab Emirates**, 28 in Jordan, 19 in Qatar, 11 in Oman, six in Iran, four in Kuwait, two in Lebanon, and one each in Yemen and Bahrain. (Department of Health, 2018f) **Reference:** 根據最新資料，世衞至今獲通報 2 144 宗個案，至少 750 人死亡。當中 10 個中東國家確診共 1927 宗，**其中沙特阿拉伯王國有 1 769 宗，阿拉伯聯合酋長國 86 宗，約旦 28 宗**，卡塔爾 19 宗，阿曼 11 宗，伊朗六宗，科威特四宗，黎巴嫩兩宗，以及也門和巴林各一宗。(Department of Health, 2018g) **Google Translate:** 根據最新資料，世衛組織報告了 2 144 例（至少 750 人死亡），其中包括 10 個中東國家的 1 927 例，其中沙特阿拉伯王國 1 769 例，阿拉伯聯合酋長國 86 例，**阿拉伯聯合酋長國 28 例，約旦(X)**，卡塔爾 19 人，阿曼 11 人，伊朗 6 人，科威特 4 人，黎巴嫩 2 人，也門和巴林各 1 人。 **Microsoft Translator:** 根據最新資料, 已向衛生組織報告了 2 144 起案件 (至少有 750 人死亡), **包括 927 個中東國家的 1 10 個(X)**, 沙烏地阿拉伯王國 1 769 個, 阿拉伯聯合大公國 86 個, 28 個在約旦, 19 個在卡塔爾。, **11 在阿曼, 六在伊朗(X)**, 四人在科威特, 兩個在黎巴嫩, 一個在葉門和巴林。 **Sogou Translator:** 根據最新資料，向世衛組織報告了 2 144 例病例(至少 750 人死亡)，其中 10 個中東國家 1 927 例，其中沙烏地阿拉伯王國 1 769 例，阿拉伯聯合大公國 86 例，約旦 28 例，卡塔爾 19 例，阿曼 11 例，伊朗 6 例，科威特 4 例，黎巴嫩 2 例 ，葉門和巴林各 1 例。 **Discussion:** The presence of a space between the one thousand’s place and the hundred’s place (such as the space between “1” and “9” in the number “1 927”) could lead to translation errors. Problems include wrong numbers (e.g., the wrong number of cases in Middle East countries and wrong number of Middle East countries in the expression “927 個中東國家的 1 10 個”), missing numbers (e.g., the missing number of cases in the expression “約旦”) or classifiers (e.g., “11 在阿曼, 六在伊朗”), and repetition (e.g., repetition of “阿拉伯聯合酋長國”).
Issue 5: Sentence structure and word order	**Source Text:** The second case is a 6-year-old girl **with good past health**. (Department of Health, 2018l) **Reference:** 第二宗個案涉及一名**過往健康良好**的六歲女童。(Department of Health, 2018m) **Google Translate:** 第二種情況是一名**身體健康的(X)**6 歲女孩。 **Microsoft Translator:**第二個案件是一個 6 歲的女孩**與良好的過去健康(X)**。 **Sogou Translator:** 第二個病例是一名 6 歲女孩，過去健康情況良好。 **Discussion:** The translation of the prepositional phrase “with good

	past health" was unsatisfactory in two of the outputs. The use of the expression "身體健康的" without the Chinese translation of "past" conveys meaning opposite to the source text. The expression "與良好的過去健康" is a word-for-word translation of the English original. Neither its location nor the expression itself is clear and grammatical.
Issue 6: Punctuation marks	**Source Text:** Do not try to use **salt, vinegar, wine and wasabi** to kill bacteria as they are not effective; and (Department of Health, 2018o) **Reference:** 不要嘗試使用鹽、醋、酒及日本芥末殺菌，因為均沒有殺菌效用；及(Department of Health, 2018p) **Google Translate:** 不要嘗試用鹽，醋，酒和芥末(X)來殺死細菌，因為它們沒有效果;和 **Microsoft Translator:** 不要試圖使用鹽, 醋, 葡萄酒和芥末(X)殺死細菌, 因為它們是不有效的;和 **Sogou Translator:** 不要嘗試使用鹽、醋、酒和芥末來殺滅細菌，因為它們沒有效果；和 **Discussion:** In the translation of the list "salt, vinegar, wine and wasabi", two of the engines fail to use enumeration commas (i.e. "、") in Chinese.
Issue 7: Omission or incomplete translation	**Source Text:** As of yesterday (March 8), 11 cases had been confirmed this year, all of which were imported cases. The cases were imported from **the Philippines (five), Cambodia (two), India (one), Indonesia (one), Singapore (one) and Thailand (one)**. (Department of Health, 2018d) **Reference:** 截至昨日（三月八日），今年共錄得 11 宗個案，均為外地傳入，分別從菲律賓（五宗）、柬埔寨（兩宗）、印度（一宗）、印尼（一宗）、新加坡（一宗）和泰國（一宗）傳入。(Department of Health, 2018q) **Google Translate:** 截至昨日（3 月 8 日），今年已確認 11 起，均為外來案件。 這些病例是從菲律賓（五），柬埔寨（二），印度（一），印度尼西亞（一），新加坡（一）和泰國（一）進口的(X)。 **Microsoft Translator:** 截至昨日 (3 月 8 日), 今年已有 11 宗個案獲證實, 全部均為進口個案。案件從菲律賓 (五)、柬埔寨 (二)、印度 (一)、印尼 (一)、新加坡 (一) 和泰國 (一) 進口(X)。 **Sogou Translator:** 截至昨日(三月八日)，本年共證實十一宗個案，全部為外地傳入個案。 這些病例來自菲律賓(5 例)、柬埔寨(2 例)、印度(1 例)、印尼(1 例)、新加坡(1 例)和泰國(1 例)。 **Discussion:** The underlined section in the source text provides figures of dengue fever cases, but only one of the engines adds classifiers to the numbers for clarification. The use of expressions "進口" and "案件", which correspond to "imported" and "cases", is also less desirable.
Issue 8: Problematic addition	**Source Text:** The CHP has also stepped up surveillance for possible human cases. **Letters will be sent** to doctors and hospitals today urging them to report any suspected cases. (Department of Health, 2018r)

	Reference: 中心已就可能出現的人類感染個案加強監測，今日**亦會發信**予醫生及醫院呼籲他們呈報可疑個案。(Department of Health, 2018s) **Google Translate:** 衛生防護中心也加強了對可能的人類病例的監測。信件將在今天發送給醫生和醫院，敦促他們報告任何疑似病例。 **Microsoft Translator:** 此外, 生防護中心亦加強監察可能發生的人類個案。**警方(X)**今日 **(星期五)(X)** 將信件送交醫生及醫院, 促請他們舉報懷疑個案。 **Sogou Translator:** 生防護中心亦加強監察可能出現的人類個案。**警方(X)**今日會發信給醫生及醫院，促請他們舉報懷疑個案。 **Discussion:** In the translation of the passive voice expression "letters will be sent", problematic expressions such as "警方", which was not in the source text, are added. Another problematic addition "(星期五)" is found in one of the translations.
Issue 9: Wrong characters	**Source Text: DH** holds World Oral Health Day publicity event (with photo) (Department of Health, 2018t) **Reference:** **衞生署**舉辦世界口腔健康日推廣活動（附圖）(Department of Health, 2018u) **Google Translate:** **衛生署**舉行世界口腔健康日宣傳活動（附圖） **Microsoft Translator:** **弥鍍 ' p (X)**舉行世界口腔健康日宣傳活動(带照片) **Sogou Translator:** **生署(X)**舉辦世界口腔健康日宣傳活動(附圖) **Discussion:** "DH" is the abbreviation of the Department of Health, the Chinese name of which is "衞生署". The Chinese character "衞" in the name is not properly shown in two of the translations, one of which even contains an unreadable expression "弥鍍 ' p" ,which is seemingly a result of wrong internal encoding.
Issue 10: Collocation	**Source Text:** They should **seek medical advice** and avoid school till 48 hours after **the fever has subsided**. (Department of Health, 2018e) **Reference:** 他們應**求診**及避免上學，直至**退燒**後四十八小時。(Department of Health, 2018v) **Google Translate:** 他們應該**尋求醫療建議**，並在**發燒平息(X)**後 48 小時內避免學校。 **Microsoft Translator:** 在**發燒消退** 48 小時後, 他們應該**尋求醫療建議**, 避免上學。 **Sogou Translator:** 他們應**尋求醫療諮詢**，避免上學，直到**發燒消退**後 48 小時。 **Discussion:** The use of "平息" in the expression "發燒平息" is unidiomatic in terms of Chinese collocation. The expressions "發燒消退" and "尋求醫療建議" (or "尋求醫療諮詢") are also less common in Hong Kong.

Table 5: Common issues in the machine translation outputs

Based on the above discussion, we would like to make the following recommendations on the use and development of neural machine translation for the public health domain. From the user's point of view, pre-editing or the use of controlled language could be useful. Special attention should be given to the clarification of abbreviations, division of longer sentences, and simplification of sentence structure. In the stage of post-editing, special care should be taken to such issues as word order, word selection, collocation patterns, inappropriate omission or addition, the use of punctuation marks, and the translation of abbreviations, proper nouns and numbers. In terms of system development, as suggested by Wolk, Marasek and Glinkowski (2015), generic web-based machine translation systems usually produce lower quality translations as the systems are not trained for specific domains. We therefore propose that we could explore the possibilities of building a domain-specific neural machine translation system for public health information by, for example, using a combination of in-domain and out-domain bilingual documents as training data.

6. Concluding Remarks

This paper examined the application of neural machine translation technology to English-Chinese translation of public health information published by the Department of Health. The three machine translation engines we considered were sometimes able to generate readable and publishable results, and Sogou Translator tended to get higher scores in our automatic and manual tests. Meanwhile, we identified some common issues in the outputs of the systems, and they deserve our attention in the pre-editing or post-editing stages of machine translation. In the future, we should conduct more in-depth analysis of the quality of neural machine translation for public health by expanding our testing datasets (e.g., increasing the size of datasets and the number of language pairs), deriving pre-editing and post-editing guidelines, considering other forms of public health information (e.g., pamphlets and video clips), and increasing the number of machine translation engines to be evaluated. It is hoped that this present work will help enhance the understanding of neural machine translation in the public health domain.

References

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

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., ... & Negri, M. (2016). Findings of the 2016 Conference on Machine Translation. In *ACL 2016 First Conference on Machine Translation (WMT16)* (pp. 131-198). The Association for Computational Linguistics.

Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., & Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics*, *16*(2), 79-85.

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, *19*(2), 263-311.

Brown, P., Cocke, J., Pietra, S. D., Pietra, V. D., Jelinek, F., Mercer, R., & Roossin, P. (1988). A statistical approach to language translation. In *Proceedings of the 12th conference on Computational linguistics-Volume 1*(pp. 71-76). Association for Computational Linguistics.

Chan, H. Y. (2016). Message from the Director. *Department of Health.* Retrieved from http://www.dh.gov.hk/english/aboutus/aboutus_md/aboutus_md.html

Chen, B., & Cherry, C. (2014). A systematic comparison of smoothing techniques for sentence-level BLEU. In *Proceedings of the Ninth Workshop on Statistical Machine Translation* (pp. 362-367).

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Crego, J., Kim, J., Klein, G., Rebollo, A., Yang, K., Senellart, J., ... & Enoue, S. (2016). SYSTRAN's Pure Neural Machine Translation Systems. *arXiv preprint arXiv:1610.05540*.

Diño, G. (2017). 3 reasons why neural machine translation is a breakthrough. *Slator*. Retrieved from https://slator.com/technology/3-reasons-why-neural-machine-translation-is-a-breakthrough/

Department of Health. (2018a). Public urged to stay alert against avian influenza (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from https://www.info.gov.hk/gia/general/201801/19/P2018011900682.htm

Department of Health. (2018b). CHP investigates outbreak of acute gastroenteritis at kindergarten-cum-child care centre in Tuen Mun (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201801/19/P2018011900531.htm

Department of Health. (2018c). CHP investigates case of severe paediatric influenza B infection (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201802/18/P2018021800564.htm

Department of Health. (2018d). Update on number of dengue fever cases (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/09/P2018030900537.htm

Department of Health. (2018f). Update on latest MERS situation in Oman (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/16/P2018031600399.htm

Department of Health. (2018g). Update on latest MERS situation in Oman (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/16/P2018031600401.htm?fontSize=1

Department of Health. (2018h). Public urged not to buy or consume slimming product from unknown sources or of doubtful composition (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201802/22/P2018022200618.htm

Department of Health. (2018i). Public urged not to buy or consume slimming product from unknown sources or of doubtful composition (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201802/22/P2018022200624.htm?fontSize=1

Department of Health. (2018j). CHP investigates suspected food poisoning case related to wild mushrooms (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from https://www.info.gov.hk/gia/general/201803/08/P2018030800752.htm

Department of Health. (2018k). CHP investigates suspected food poisoning case related to wild mushrooms (Chinese Version). *The Government of the Hong Kong Special*

Administrative Region Press Releases. Retrieved from https://www.info.gov.hk/gia/general/201803/08/P2018030800750.htm?fontSize=1

Department of Health. (2018l). CHP investigates three cases of severe paediatric influenza B infection (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201802/22/P2018022200618.htm

Department of Health. (2018m). CHP investigates three cases of severe paediatric influenza B infection (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201802/22/P2018022200624.htm?fontSize=1

Department of Health. (2018o). CHP investigates suspected food poisoning outbreak (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201801/08/P2018010800880.htm

Department of Health. (2018p). CHP investigates suspected food poisoning outbreak (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201801/08/P2018010800877.htm?fontSize=1

Department of Health. (2018q). Update on number of dengue fever cases (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/09/P2018030900542.htm?fontSize=1

Department of Health. (2018r). CHP urges public to be cautious in handling chilled poultry to guard against avian influenza (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201801/23/P2018012300717.htm

Department of Health. (2018s). CHP urges public to be cautious in handling chilled poultry to guard against avian influenza (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201801/23/P2018012300724.htm?fontSize=1

Department of Health. (2018t). DH holds World Oral Health Day publicity event (with photo) (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/11/P2018030900606.htm

Department of Health. (2018u). DH holds World Oral Health Day publicity event (with photo) (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from http://www.info.gov.hk/gia/general/201803/11/P2018030900604.htm?fontSize=1

Department of Health. (2018e). CHP investigates two influenza outbreaks at primary schools (English Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from https://www.info.gov.hk/gia/general/201801/29/P2018012900801.htm

Department of Health. (2018v). CHP investigates two influenza outbreaks at primary schools (Chinese Version). *The Government of the Hong Kong Special Administrative Region Press Releases.* Retrieved from https://www.info.gov.hk/gia/general/201801/29/P2018012900829.htm?fontSize=1

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. *arXiv preprint arXiv:1705.03122*.

Google. (n.d.). Languages. *Google Translate*. Retrieved from https://translate.google.com/intl/en/about/languages/

Google Translate [Computer software] (n.d.). Retrieved from https://translate.google.com.
Firat, O., Cho, K., & Bengio, Y. (2016). Multi-way, multilingual neural machine translation with a shared attention mechanism. *arXiv preprint arXiv:1601.01073*.
Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., ... & Liu, S. (2018). Achieving Human Parity on Automatic Chinese to English News Translation. *arXiv preprint arXiv:1803.05567*.
Hospital Authority. (2018). Introduction. *Hospital Authority*. Retrieved from http://www.ha.org.hk/visitor/ha_visitor_index.asp?Content_ID=10008&Lang=ENG&Dimension=100&Parent_ID=10004#
Hong Kong Sanatorium & Hospital. (n.d.). Health Info. *Hong Kong Sanatorium & Hospital*. Retrieved fro http://www.hksh-hospital.com/en/health-info
Jieba [Computer software] (2017). Retrieved from https://github.com/fxsjy/jieba
Kirchhoff, K., Turner, A. M., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: a feasibility study. *Journal of the American Medical Informatics Association*, 18(4), 473-478.
Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 conference of the North American chapter of the Association for Computational Linguistics on human language technology-Volume 1* (pp. 48-54). Association for Computational Linguistics.
Lin, C. Y., & Och, F. J. (2004). Automatic evaluation of machine translation quality using longest common subsequence and skip-bigram statistics. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics* (p. 605). Association for Computational Linguistics.
Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. *arXiv preprint arXiv:1508.04025*.
Maylahn, C., Fleming, D., & Birkhead, G. (2013). Health departments in a brave new world. *Preventing Chronic Disease*, 10, E41. http://doi.org/10.5888/pcd10.130003
Microsoft Translator. (2016, November 16). Microsoft Translator launching neural network based translations for all its speech languages. [Web log post]. Retrieved from https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/
Microsoft Translator. (2017). Languages. *Microsoft Translator*. Retrieved from https://translator.microsoft.com/help/articles/languages/.
Microsoft Translator [Computer software] (2018a). Retrieved from https://www.bing.com/translator
Microsoft Translator. (2018b, March 14). Neural machine translation reaches historic milestone: human parity for Chinese to English translations. [Web log post]. Retrieved from https://blogs.msdn.microsoft.com/translation/2018/03/14/human-parity-for-chinese-to-english-translations/.
Nagao, M. (1984). A framework of a mechanical translation between Japanese and English by analogy principle. *Artificial and Human Intelligence*, 351-354.
NLTK [Computer software] (2017). Retrieved from https://www.nltk.org/
Regidor, E., de la Fuente, L., Gutiérrez-Fisac, J. L., de Mateo, S., Pascual, C., Sánchez-Payá, J., & Ronda, E. (2007). The role of the public health official in communicating public health information. *American Journal of Public Health*, 97(Suppl 1), S93–S97. http://doi.org/10.2105/AJPH.2006.094623
Sogou. (2018). Machine Translation (Chinese version). *Sogou DeepI*. Retrieved from http://deepi.sogou.com/fanyi?from=translatepc
Sogou Translator [Computer software] (2018). Retrieved from https://fanyi.sogou.com/

Tencent. (2018). Machine Translation (Chinese version). *Tencent Cloud*. Retrieved from https://cloud.tencent.com/product/tmt

Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text?. *arXiv preprint arXiv:1801.04962*.

Turner, A. M., Dew, K. N., Desai, L., Martin, N., & Kirchhoff, K. (2015). Machine translation of public health materials from English to Chinese: A feasibility study. *JMIR Public Health and Surveillance*, 1(2).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention Is all you need. *arXiv preprint arXiv:1706.03762*.

Wang, Y., Cheng, S., Jiang, L., Yang, J., Chen, W., Li, M., ... & Yang, H. (2017). Sogou neural machine translation systems for WMT17. In *Proceedings of the Second Conference on Machine Translation* (pp. 410-415).

Wołk, K., Marasek, K., & Glinkowski, W. (2015). Telemedicine as a special case of machine translation. *Computerized Medical Imaging and Graphics*, 46, 249-256.