

Detecting Free Translation in Parallel Corpora from Attention Scores

Qi Chen^{1,2}

Oi Yee Kwong^{1*}

Jingbo Zhu²

¹Department of Translation, The Chinese University of Hong Kong

²Natural Language Processing Lab, Northeastern University, China

chenqi.neu@foxmail.com, oykwong@arts.cuhk.edu.hk, zhujingbo@mail.neu.edu.cn

Abstract

In this study, we propose a method for extracting free translation examples from bilingual parallel corpora based on an innovative use of attention scores. Preliminary results show that the approach is promising and paraphrases at both sentential and sub-sentential levels covering diverse surface forms could be identified. The extracted data, upon further filtering, have great potential to supplement the example sentences available in existing bilingual dictionaries in an effective and systematic way.

1 Introduction

In any language, the same meaning can often be expressed in alternative ways, or paraphrased. The recognition and generation of such meaning-equivalent forms are deemed important for various natural language processing (NLP) applications (Madnani and Dorr, 2010). In the inter-lingual context, typically in translation, there is inevitably a trace of paraphrasing, whether or not it is employed consciously as a strategy in the process. Notwithstanding the different interpretation of terminology and research objectives across disciplines (e.g. translation vs paraphrase, literal translation vs free translation, etc. in translation studies, or paraphrase generation vs query expansion in the NLP community), access to such context-sensitive equivalents is essential especially when fluency, in addition to fidelity, is concerned, for machine and human translation alike. This study is thus motivated

by the need to mine useful examples of free translation for human translators' reference on the one hand, and for inspiring machine translation's further improvement on fluency on the other. Paraphrase, in this paper, is therefore used in a slightly restricted sense to refer to alternative expressions in a target language which are not only semantically equivalent, but also fulfill other contextual criteria to be qualified as translation of the text in a source language. Along the continuum of equivalence, free translation (as opposed to literal translation) will be of more interest to us.

Three types of paraphrases are often included in paraphrase databases (e.g. Ganitkevitch et al., 2013; Ganitkevitch and Callison-Burch, 2014): lexical, phrasal, and syntactic paraphrases. Although the inclusion of these paraphrase types has recognized that paraphrases could go beyond the replacement by synonyms or synonymous phrases, there is nevertheless a restriction of syntactic category, as it at most allows the substitution of expressions that match a given syntactic type, even when non-constituent phrases are accommodated (e.g. Callison-Burch, 2008), as long as labels in syntactic trees are used as the point of reference.

Paraphrases in different phrasal categories and syntactic constructions are particularly important in translation, as individual possible renditions would have their strengths and weaknesses in a given context with a certain literary style for a specific communicative purpose. In addition to fidelity, this is even more salient when fluency is concerned across language pairs with very distinct linguistic properties, where literal translation is not always the

* Corresponding author

best or even natural option. Consider an example for English-Chinese translation:

I still have vivid memories of my childhood.

我	对	童年	还有	清晰	的	记忆。
wǒ	duì	tóngnián	háiyǒu	qīngxī	de	jìyì
I	for	childhood	still-have	clear	DE	memory

童年	的	事	我	还	记得	清清楚楚。
tóngnián	de	shì	wǒ	háiyǒu	jìde	qīngqīngchǔchǔ
childhood	DE	event	I	still	remember	clear

小时候	的	点点滴滴	至今	记忆犹新。
xiǎoshíhòu	de	diǎndiǎndīdī	zhìjīn	jìyìyóuxīn
young-days	DE	drips-and-drabs	so-far	memory-still-fresh

As shown in the above example, the first rendition can be considered most literal among the three. The noun “childhood” and the noun phrase “vivid memories” in the source text are expressed as the same syntactic categories in the target text, as “童年” and “清晰的记忆” respectively. In the second and third renditions, however, “childhood” is expanded into a noun phrase in the target text (“童年的事” and “小时候的点点滴滴”), while “vivid memories” is translated with a verb phrase and a four-character idiomatic expression respectively (“记得清清楚楚” and “记忆犹新”). The current study is interested in extracting such examples of free translation from parallel corpora.

Our proposed method makes innovative use of attention scores, going by the following main assumption. In the neural machine translation framework, the encoder works out a sequence of context vectors for a source sentence; and given a source sentence, the decoder computes a probability distribution over the translation by an attention mechanism over the context vectors of the source sentence (Bahdanau et al., 2014; Luong et al., 2015). At each state, the probability of the next target symbol is updated by a softmax function. For most work on machine translation, the focus would primarily be placed on the more strongly correlated parts in the resulting alignment from the attention mechanism, which often indicate relatively faithful or standard, if not necessarily literal, translation. In the case of other possible and perhaps more fluent renditions, the correlation between the source and the target would be

much weaker. For the above example, it is thus expected that “vivid memories” would match very well with “清晰的记忆” but would be found to correlate more weakly with “记得清清楚楚”. But provided that the bilingual parallel corpus is of good quality, such weakly correlated parts are likely to correspond to free yet more fluent translation. So by ruling out the known parts for literal translation (or at least the most common renditions), the remaining less well-aligned parts may contribute to paraphrastic expressions, although they might be less frequent and are probably restricted in terms of literary style and communicative purposes.

In Section 2, we will review related NLP work on paraphrase extraction and generation. In Section 3, we will introduce our proposed method for finding free translation from bilingual parallel corpora. Section 4 describes the experimental setup. Section 5 discusses preliminary results and future plans, followed by a conclusion in Section 6.

2 Related work

For two decades by now, methods on paraphrase extraction and generation have mostly been data-driven (Madnani and Dorr, 2010). Monolingual or bilingual corpora may be used, sometimes also with the help of existing lexical resources (e.g. Wu and Zhou, 2003).

Earlier methods primarily rely on distributional similarity for finding paraphrases from identical surrounding context (e.g. Barzilay and McKeown, 2001; Ibrahim et al., 2003). For example, Barzilay and McKeown (2001) utilized a monolingual corpus consisting of multiple English translations of the same novels by foreign authors. The approach takes advantage of the many words shared by the parallel translations, assuming that phrases in aligned sentences appearing in similar contexts are paraphrases. They used a co-training algorithm based on contextual and lexico-syntactic features of paraphrases, which produced more than 9,000 pairs of lexical paraphrases including relations other than synonymy. Human judgment, with or without context, was solicited.

Lexical paraphrases, which often involve a replacement with synonyms or hypernyms, do not give a complete account of paraphrase itself or serve ap-

plications adequately. Phrasal and sentential paraphrases are indispensable (e.g. Barzilay and Lee, 2003; Bannard and Callison-Burch, 2005; Callison-Burch, 2008; Ganitkevitch et al., 2011). For phrasal paraphrases, syntactic categories often become a basic point of reference. Bannard and Callison-Burch (2005), for instance, used a bilingual parallel corpus and obtained English paraphrases by pivoting through foreign language phrases. With reference to phrase-based statistical machine translation, they aligned phrases in the corpus. Those mapping to the same phrase in another language are considered candidates and ranked by a paraphrase probability defined in terms of two translation model probabilities. The extracted paraphrases were judged by two native English speakers in terms of both meaning and grammaticality.

Callison-Burch (2008) even constrained paraphrases to have the same syntactic type as the original phrase, and reported higher performance. Sentential paraphrase has also been extracted in a syntax-driven way exploiting a set of meaning-preserving transformations (Ganitkevitch et al., 2011). However, such constraints may not be appropriate for our current purpose, as the kind of paraphrase, or free translation, that we find useful often appears as different syntactic constructions on the one hand, and may not be accompanied by regular and predictable transformation patterns on the other.

The pivoting approach has stayed in the mainstream of paraphrase extraction, with large paraphrase databases like the PPDB and multilingual PPDB created in the meantime (Ganitkevitch et al., 2013; Ganitkevitch and Callison-Burch, 2014). Mallinson et al. (2017) revisited the approach from the perspective of neural machine translation (NMT), without reference to any underlying grammar or creating any phrase table. In particular, pivoting is done with the NMT model, in the form of one-to-one back-translation or multi-pivoting through the K -best translations. Their system, PARANET, makes use of the attention mechanism for identifying semantically equivalent parts between the paraphrase sentence and the source sentence, with each word of the paraphrase sentence attending to words in the pivot sentence and each word in the pivot sentence attending to words in the source sentence.

Before we move on, it should be noted that while

this study shares the objective of paraphrase extraction with previous studies, we have a subtly different motivation. While previous studies intend to extract or generate paraphrases for a sentence in the same language, our current study sets out to identify or mine different free translations in a given language (in the form of paraphrase, as opposed to metaphrase and imitation in translation terminology) of the same source text in another language. Hence, although we are also inspired by the NMT model here, we do not need explicit pivoting either from the corpus data itself or generated by machine translation. The pivoting can be considered inherent in the task itself. Moreover, the way we use the attention weighting is for detecting the different possibilities for translating the same source text. The pairwise comparison is between the source sentence and the target sentence, but not between any two target sentences (or sub-sentential units). This is also why it makes sense to pay attention to the more weakly correlated parts instead of those with higher attention scores, as the former would more likely signal the variations we look for, otherwise the word-for-word and more literally translated parts would be found more correlated.

3 Method

3.1 Overall Design

Our method is composed of the following modules as illustrated by Figure 1.

First, we split a selected phrase into all possible subphrases, and then search those subphrases by hash and get all source sentences that contain those subphrases. In this way, we will not be limited to a complete phrase but can give different attention to each subphrase, so that we can focus on some translations of rare words and ignore some literal translations. Moreover, the output is allowed to have some degree of variation relative to the input.

Second, for each pair of source sentence and target sentence, we get the attention weight matrix among the source words and target words by NMT, as shown in Figure 2. We also match all the words in the sentence pair against a dictionary, as shown in the attention score module in Figure 1. If a match is found, then the corresponding mask is set to -1 , otherwise it is set to 1. The attention weight and

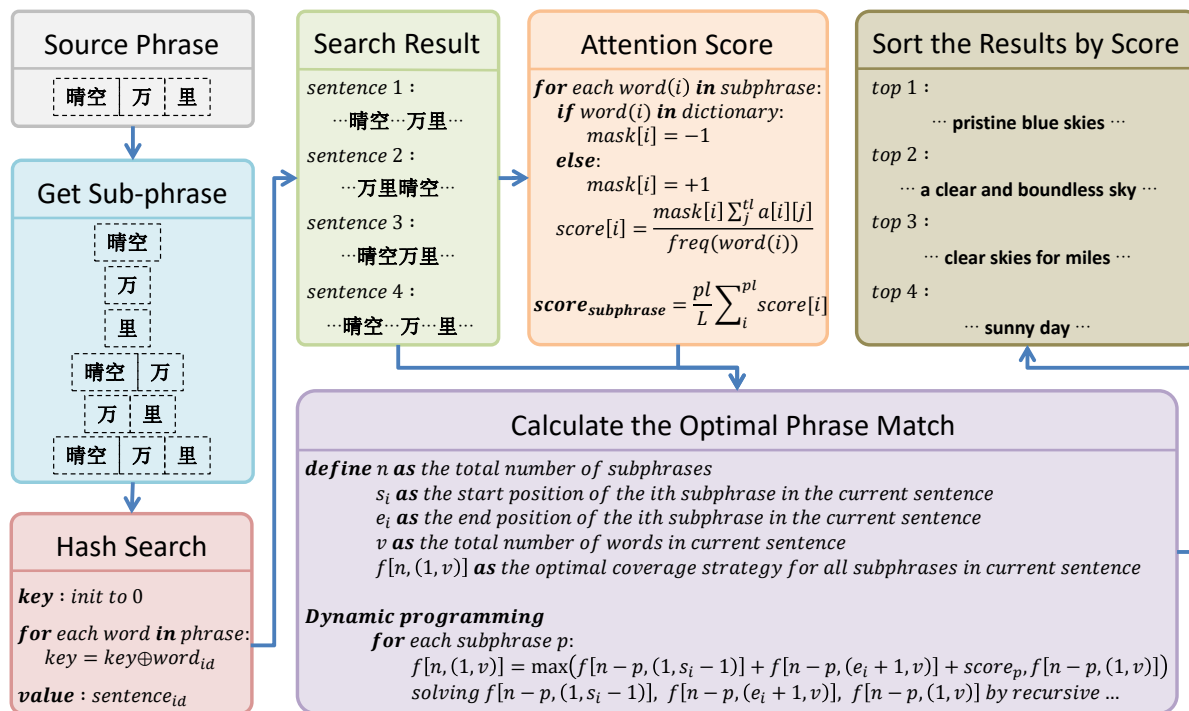


Figure 1: Overall architecture diagram of the system. In the hash search module, $word_{id}$ and $sentence_{id}$ are keys that uniquely identify a word or a sentence. In the attention score module, tl means the length of the target sentence and pl means the length of the current subphrase, L means the maximum length between the first word and the last word of the subphrase in the source sentence, a means the attention matrix obtained from NMT, $freq(word)$ is the frequency of a word in the corpus.

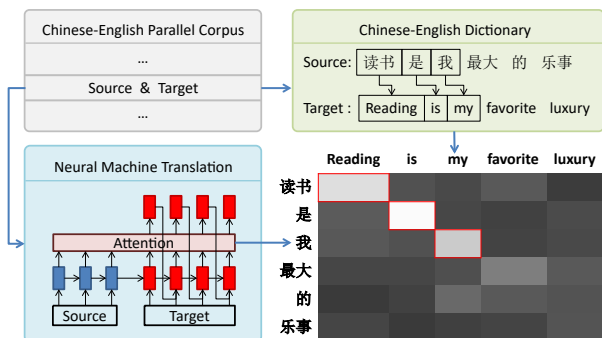


Figure 2: For each source and target sentence pair, on the one hand we get the attention weight matrix of them by NMT, what we focus on are the words which do not have significant alignment probability in the matrix, since they are more likely to be paraphrased. On the other hand, we find out all the word matchings in the sentence pair in an existing dictionary, which can help us more accurately locate the literal translation of the words.

the mask are then multiplied. The reason for doing this is that we assume the translations found in the dictionary are more literal translations, so we penalize them on the score when they appear in the candidate sentences. We also make assumptions about the possibility of each word being paraphrased, that is, the words with higher frequency are more likely to be translated literally, and those with lower frequency are more likely to be paraphrased. Hence each score is divided by the frequency of the corresponding word.

In addition, for each subphrase, we do not only sum up the scores of the words in the subphrase but also give each subphrase a penalty weight pl/L , where pl is the length of the current subphrase and L is the maximum length between the first and the last word of the subphrase in source sentence. By this method, the score mechanism includes the evaluation of the degree of discretization between each

subphrase. If the words in a subphrase are far away from each other in a source sentence, it will be severely penalized.

Third, we find out the optimal phrase match with each matching result by the score of each subphrase. We use dynamic programming to do this and the state transfer equation is shown in the optimal phrase match module in Figure 1. This module mainly solves two problems in the matching process as shown in Figure 3:

- First, if there are more than one matching result of subphrase A and B, as we used length penalty measure in the previous scoring process, the matching results that are far apart can be filtered out.
- Second, if subphrase AB is a match, but the word B is confirmed as a literal translation, then we can split subphrase AB to subphrase A and subphrase B and filter B by the state transfer equation. In this way, the final score is not affected by B, and our attention will be focused on the degree of paraphrase of A.

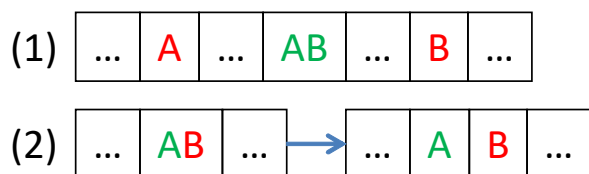


Figure 3: The ambiguity matches that may occur in the matching process.

Finally, we sort all the sentence pairs by their optimal matching scores, as shown in Figure 1. For each sentence pair, if $score \in [-\infty, \theta]$, we treat it as noise or possible omission, where θ is an empirical threshold. For other sentence pairs, it is assumed that the smaller the score, the more interesting example of free translation the selected phrase will offer.

3.2 Cross Entropy Loss

The cross entropy loss function is a commonly used loss function for NMT. Its formula is as follows:

$$E(t, y) = -\frac{1}{n} \sum_j^n t_j \log y_j \quad (1)$$

where t represents the target sentence and y_j represents the output probability of each word t_j .

The loss value can reflect the accuracy of sentence translation to a certain extent. We will use the score for a preliminary ordering of the sentence pairs to filter obvious noise first.

4 Experimental Setup

We tested our approach on both English-Chinese and Chinese-English sentence pairs. For the translation task, the NIST12 OpenMT was used to train the NMT system and we used NIST 2006 (MT06) as the validation set.

Our system was implemented based on the Nema-tus open-source toolkit (Sennrich et al., 2017). Both the encoder and the decoder consisted of two LSTM layers. By default, the hidden layer dimension was set to 1024, and the embedding dimension was set to 512. We limited our word vocabularies to the most frequent $30K$ words for both source and target.

5 Preliminary Results and Discussion

Given that the task concerned in this study is not exactly the same as previous attempts on paraphrase extraction, and the nature of the task itself is more open-ended, for the time being we will present the results from a more qualitative perspective, as a proof of concept to start with. From this we will identify potential development and applications, to arrive at a feasible and meaningful evaluation model in the near future.

5.1 Initial Outcomes

Large corpora are bound to contain noise in one way or another, which may arise from a complete mismatch between the source and the target sentence, an incompatible alignment at the sentence level, or at other times simply very poor translation. Such noisy sentence pairs will not be of much use in subsequent paraphrase identification anyway, so they need some filtering to start with. To distinguish the available sentence pairs by their usefulness, we ordered them by the *loss* score, which is essentially a measure for the translation quality of the whole sentence in general. Some examples are shown in Table 1, in ascending order of their *loss* scores.

#	Chinese Sentence	English Sentence	Keep?
1	一个新闻工作者你不能一时冲动啊。	A journalist cannot act rashly .	Y
2	不过,我必须指出,我绝非因一时的冲动而仓卒提出对策。	But I must stress that my actions are not based on impulse .	Y
3	因此,我可证明接受政府的提案,并非一时冲动的决定,而是已经我们商讨已久。	Therefore, I can prove that I have not rashly accepted the government’s proposal because we have discussed the issue for a long time.	Y
4	他来回奔跑近6000里地,只为了一时冲动想要见她。	He ran nearly 6,000 li to and fro, just for the spur-of-the-moment desire to meet her.	Y
5	每一件事能够一直乐在其中,实在需要一分耐心,而不是一时冲动。	To keep on finding pleasure in anything takes a bit of patience, not just a momentary infatuation .	Y
6	自杀不是解决问题的方法,自杀很多时都是一时的冲动,如果事前可以聆听和辅导,情况是可以避免的。	Mr President, I am not laying the blame on the media because this is absolutely unfair.	N

Table 1: Filtering the examples with *loss* scores

What makes a sentence pair useful, with respect to its potential for providing an example of free translation, will depend on at least two factors. One is the overall literalness of the translation of the whole sentence. Hence if the whole sentence is more or less literally or very faithfully translated, the NMT model should find a relatively high probability for the target sentence given the source sentence in general, hence a low loss. Sentences on the low end of this continuum may thus be too faithful to make a good example for free translation, and those on the other extreme may just be too messy for the purpose. As shown in Table 1, among the examples with the phrase “一时冲动”, Sentence 1 (“A journalist cannot act rashly”) is relatively well translated according to the NMT model, in comparison with Sentence 5 (“To keep on finding pleasure in anything takes a bit of patience, not just a momentary infatuation”). On the one hand, the phrase is embedded in different contexts in the two sentences, which may really demand different ways of translation in practice. On the other hand, Sentence 1 may have an advantage of being shorter in length and using a slightly more common English phrase in the translation (“act rashly” vs “a momentary infatuation”). But after all, Sentences 1 and 5 would both be more useful for our purpose than, say, Sentence 6 (“Mr President, I am not laying the blame on the media

because this is absolutely unfair”), which happens to be a complete mismatch with the source sentence, that is, noise.

The other factor regarding the usefulness of a sentence pair for providing a free translation example has to do with the subphrase in focus, whether it has been rendered in a relatively straightforward way in the target sentence and finds a highly correlated chunk in it. This would be preliminarily revealed by the attention score. As illustrated by the English-Chinese examples in Table 2, for each source phrase, “extravagant luxury” and “extravagant habit”, the first translation resulted in a lower attention score than the second translation. Judging by the actual wording and the context, comparatively the first translation can be considered more “free” and the second translation more “literal”. For instance, “extravagant luxury” as “奢侈的享受” and “extravagant habit” as “浪费的现象” are quite literally translated in view of the word order and the syntactic structure, especially as “奢侈” and “浪费” are often found in dictionaries as the context-free equivalents for “extravagant”. On the other hand, the first translation for each of the example source phrases obviously demonstrates a more contextually dependent rendition. Our method thus works as expected for such cases, where relatively literal and relatively free translations are found for a source text chunk.

Source Phrase	English Sentence	Chinese Sentence
extravagant luxury	I had always considered them to be an extravagant luxury .	我总觉得,这是玩物丧志。
	Frequenting coffee shops was considered an extravagant luxury .	坐咖啡馆已经是一种奢侈的享受了。
extravagant habit	Social habits have become extravagant , and people try to get rich overnight, and take risks, which causes social disorder.	而由於社会风气奢靡,负担不起的人想一夜致富,只好冒险,治安问题就来了。
	The President said that society is full of luxurious and extravagant habits , which are detrimental to the people's wellbeing.	但是总统以为,今天我们所感不足的,就是在这许多进步之中,社会风气没有达到同等程度的进步,以致形成了奢侈、浪费的现象,對於社会人心产生了不良的影响。

Table 2: Revealing the degree of literalness by attention scores

At other times, however, the difference on the degree of literalness may be less distinct when a truly literal translation is not found or may not even be possible. This is observed more often in Chinese-English translation, particularly for four-character phrases or idioms which make a condensed syntactic structure and can be used in different ways in a sentence. Referring to the examples in Table 1 again, upon filtering the poorly aligned sentences, the remaining ones are ordered by their attention scores for the phrase in focus. We thus get the following possible renditions for “一时冲动” (including disconnected cases as “一时的冲动”), in ascending order of their attention scores:

- (actions) based on impulse
- (act / do something) rashly
- a momentary infatuation
- spur-of-the-moment (desire)

Strictly speaking none of these is particularly literal, while it is also difficult to say which is more free than the others. The variation in syntactic forms of the translations is obvious, among which renditions in the form of nouns, noun phrases, adverbs, and even adjectives are found. On the one hand, Chinese words are notorious for being categorially fluid. It is not uncommon for a word to take up different grammatical functions in a sentence at different times. On the other hand, as mentioned in the introduction, when fluency is concerned, one often has to move away from the syntactic form of the source sentence

and express the meaning in another syntactic form which would sound more natural and idiomatic in the target language, especially when prompted by particular contexts, literary styles, as well as communicative needs. As seen in the examples, the noun phrase “一时冲动的决定” in Sentence 3 is not rendered as “a decision based on impulse”, but simply as an adverb used together with a verb phrase “(have not) rashly (accepted ...)”. It is such examples, indeed, that we aim at collecting, for they are most valuable to help translators realize the subtle relation between fidelity and fluency.

The preliminary results and observations we have thus far serve as proof of concept, showing that the method we propose for detecting free translation examples based on attention scores is practically feasible and potentially effective, although we have yet to work toward a more convincing evaluation model. In fact, the deployment of the attention scores may need to be more refined for pinpointing the precise chunk in the target sentence which corresponds to the non-literal translation. For instance, the attention score of the subphrase in focus may need to be considered with the rest of the sentence taken into account. Hence if the subphrase in focus is not well aligned with any part of the target sentence, and if the rest is more or less poorly aligned, it would indicate a less favourable situation. On the other hand, if the subphrase in focus is not well aligned but the rest is quite well aligned, it will suggest a better chance

for obtaining a free translation example for the subphrase part. More testing is needed, and it is an art to strike a balance between various considerations for different threshold settings.

5.2 Potential Applications and Evaluation

Existing bilingual dictionaries and other resources (e.g. Cambridge Dictionary¹, Collins Dictionary², iCIBA³, Baidu Translate⁴) tend to offer context-free and context-sensitive translation equivalents as conventionally distinguished by lexicographers (Atkins and Rundell, 2008). Context-free translations are thus given as general equivalents at the lexical or phrasal level, while context-sensitive translations are provided in the form of example sentences, which are to be explored by the more proficient users on their own. It is the latter that would be more important and useful for human translators in practice. Nevertheless, where example sentences are available in most existing resources, if not limited in number and variety, they are presented simply as a list of sentences gathered from various sources (often web sources) in no specific order. Similar examples may be far apart, and repetitive cases are not streamlined.

The approach for detecting free translation examples proposed in this study could contribute in this regard. As shown in the results above, the extracted examples ordered by their attention scores provide a rough indication of their interestingness, with respect to illuminating variations in translation. The extraction process itself may be able to mine additional examples from corpora to supplement existing dictionaries. Meanwhile, the method could also help in systematically organizing the example sentences.

An *in vitro* evaluation framework could thus be outlined as follows: The baseline will be the set of bilingual example sentences from a given dictionary or lexical resource. Run the extraction method on a given bilingual parallel corpus together with the sentences from the dictionary. Consider the ordered list of examples extracted. Ask human judges to evaluate the usefulness of the additionally found examples and see whether the organized presentation makes more sense from the point of a user.

¹<https://dictionary.cambridge.org/>

²<https://www.collinsdictionary.com/>

³<http://www.iciba.com/>

⁴<https://fanyi.baidu.com/>

Furthermore, since the method allows the extraction of examples at the sentential and sub-sentential levels, the lexical or phrasal example pairs could be deployed for enhancing the navigational means in dictionary access.

5.3 Ongoing and Future Work

The preliminary outcomes based on the attention scores are being more closely scrutinized, to further observe the relation between the scores and the translation quality of the sentence pairs as well as the translation strategy employed. Further utilization of the scores for fine-grained extraction of subphrase units from the source and target sentences is underway. The following directions have also been planned for future work:

- To extend the method so as to group together examples demonstrating a similar strategy of translating the word or phrase in focus.
- To study thoroughly the effectiveness of the method, aiming at more insightful observations with respect to the nature of English-Chinese and Chinese-English translation.
- To test the method on other English-Chinese parallel corpora, especially with texts of different literary styles.
- To work toward a stringent evaluation framework, supplementing qualitative observations with quantitative measures at a larger scale.

6 Conclusion

We have thus proposed a method for extracting free translation examples from bilingual parallel corpora based on an innovative use of attention scores. For proof of concept, the preliminary results and observations show that the approach is promising. Free translations of lexical and phrasal units covering diverse surface forms could be identified and ordered in terms of their literalness. Further testing and investigation is underway. Although a more formal evaluation framework has yet to be derived, free translation examples extracted by the method are expected to supplement existing bilingual dictionaries in an effective and systematic way.

Acknowledgments

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 14616317).

References

- B.T. Sue Atkins and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Dzmitry Bahdanau, KyungHyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv e-prints, abs/1409.0473*.
- Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 597–604, Ann Arbor.
- Regina Barzilay and Lillian Lee. 2003. Learning to paraphrase: An unsupervised approach using multiple-sequence alignment. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 16–23, Edmonton.
- Regina Barzilay and Kathleen R. McKeown. 2001. Extracting Paraphrases from a Parallel Corpus. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, pages 50–57, Toulouse.
- Chris Callison-Burch. 2008. Syntactic Constraints on Paraphrases Extracted from Parallel Corpora. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 196–205, Honolulu.
- Juri Ganitkevitch and Chris Callison-Burch. 2014. The Multilingual Paraphrase Database. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pages 4276–4283, Reykjavik, Iceland.
- Juri Ganitkevitch, Chris Callison-Burch, Courtney Napoles, and Benjamin Van Durme. 2011. Learning Sentential Paraphrases from Bilingual Parallel Corpora for Text-to-Text Generation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1168–1179, Edinburgh, Scotland.
- Juri Ganitkevitch, Benjamin Van Durme, and Chris Callison-Burch. 2013. PPDB: The Paraphrase Database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 758–764, Atlanta, Georgia.
- Ali Ibrahim, Boris Katz, and Jimmy Lin. 2003. Extracting structural paraphrases from aligned monolingual corpora. In *Proceedings of the International Workshop on Paraphrasing*, pages 57–64, Sapporo.
- Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective Approaches to Attention-based Neural Machine Translation. *arXiv e-prints, abs/1508.04025*.
- Nitin Madnani and Bonnie J. Dorr. 2010. Generating Phrasal and Sentential Paraphrases: A Survey of Data-Driven Methods. *Computational Linguistics*, 36(3):341–387.
- Jonathan Mallinson, Rico Sennrich, and Mirella Lapata. 2017. Paraphrasing Revisited with Neural Machine Translation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 881–893, Valencia, Spain.
- Rico Sennrich, Orhan Firat, Kyunghyun Cho, Alexandra Birch, Barry Haddow, Julian Hirschler, Marcin Junczys-Dowmunt, Samuel Läubli, Antonio Valerio Miceli Barone, Jozef Mokry, and Maria Nădejde. 2017. Nematus: a Toolkit for Neural Machine Translation. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 65–68, Valencia, Spain.
- Hua Wu and Ming Zhou. 2003. Synonymous Collocation Extraction Using Translation Information. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo.