Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 142 (2018) 308–314

Procedia Computer Science

www.elsevier.com/locate/procedia

The 4th International Conference on Arabic Computational Linguistics (ACLing 2018), November 17-19 2018, Dubai, United Arab Emirates

Paraphrasing Arabic Metaphor with Neural Machine Translation

Manar Alkhatib[a]*, Khaled Shaalan[a]

[a] Faculty of Engineering and IT, British University in Dubai, Dubai , UAE.

{ manaralkhatib09@gmail.com , khaled.shaalan@buid.ac.ae }

Abstract

The task of recognizing and generating paraphrases is an essential component in many Arabic natural language processing (NLP) applications. A well-established machine translation approach for automatically extracting paraphrases, leverages bilingual corpora to find the equivalent meaning of phrases in a single language, is performed by "pivoting" over a shared translation in another language. Neural machine translation has recently become a viable alternative approach to the more widely-used statistical machine translation. In this paper, we revisit bilingual pivoting in the context of neural machine translation and present a paraphrasing model based mainly on neural networks. Our model describes paraphrases in a continuous space and generates candidate paraphrases for an Arabic source input. Experimental results across datasets confirm that neural paraphrases significantly outperform those obtained with

Keywords: Neural Machine Translation; Paraphrasing; Metaphor; Arabic language

1. Introduction

Automatic machine translation is one of the major problems in natural language processing. It has proved both the most enticing task and the least approachable. Since its introduction in the field, many approaches have been applied, from traditional rule-based methods to the more recent statistical methods [1]. Still, as anyone who has spent a few minutes on Google Translate®, an online free translator that uses statistical machine translation, will testify, there is still a long way to go before this problem can be considered solved in any useful fashion.

However, the efficiency of a machine translation system is deeply dependent on the language pair under consideration. While there are still certain grammatical structures to be considered, for example metaphors that are often not translated appropriately, statistical machine translation between language pairs such as French and English is considered to have obtained enough accuracy to be somewhat useful in practice [2].

* Corresponding author. Tel.: +971-52-929 0281.
E-mail address: manaralkhatib09@gmail.com

1877-0509
10.1016/j.procs.2018.10.493

Neural machine translation in English [3-5] has become a significant alternative to the widely used statistical machine translation system [6], evidenced by the successful entries in the WMT'15 and WMT'16 conferences. For Arabic language translation, however, neural machine translation (NMT) is a new machine translation approach that has not led to remarkable improvements, particularly concerning human evaluation, compared to rule-based and statistical machine translation (SMT) systems [7].

While everyone may be familiar with the concept of paraphrase in its most fundamental sense, there is still a room for elaboration on how paraphrases may be automatically generated or extracted for use in language processing applications. The rest of this section formalizes the notion of a paraphrase and scopes out the collusion of the Arabic language.

Our goal in this paper is, therefore, to introduce the first result on automatic Arabic translation using paraphrases with neural machine translation employing a bilingual corpus and WordNet [8]. We also aim to solve the problem of word sense disambiguation and metaphors in both directions (i.e. Ar→En and En→Ar). The experiments reveal that our paraphrase neural machine translation system shows superior performance compared to the standard phrase-based system. We used METEOR [9], which measures translation quality *averaged* over all the sentences in a corpus.

1.1 What is a Paraphrase?

The principle of semantic equivalence most generally defines the concept of paraphrasing: A paraphrase is an alternative surface form in the same language describing the same semantic content as the original form. The idea of paraphrasing has been examined in conjunction with and employed in many natural language processing applications. Given the difficulty inherent in analyzing such a complex task, an unfortunate but necessary improvement is to impose specific limits on the scope of our discussion [10].
Individual lexical items sharing the same meaning are usually referred to as lexical paraphrases, or, more commonly, synonyms, for example, (حار, har, hot) versus (دافئ, dafe', warm), and (تناول, tnAwl, eat) versus (استهلك, Asthlk, consume). However, lexical paraphrasing cannot be limited strictly to the notion of synonymy. There are several other methods such as hyperymy, where one of the words in the paraphrastic relationship is either more general or more particular than the other, for example, (رد, rd, reply) and (قول, qwl, say) [10].

1.2 Linguistic Description

In Arabic, machine translation (MT) is bound to face many problems in producing exact coherent translations between Arabic and English. When evaluating the output of MT, the transferred meaning is the most significant focus point. Semantics is a critical aspect of translation both as a theory and in its application; it therefore, requires our utmost attention. Very few systems have addressed the problem of Arabic syntactic generation within MT and in Interlingua-based multilingual translation in particular, due both to the language complexity and to a lack of resources [11].

The use of metaphor is ubiquitous in natural language text, and it is a severe bottleneck in automatic text understanding. Improving methods to identify and deal with metaphors is an open problem in Arabic natural language processing, especially in its Machine Translation. The complexities involved in any metaphor can semantically modify the meaning of the machine-translated text. This makes metaphors a vital research area for computational and cognitive linguistics; their automatic identification and interpretation is indispensable for any semantics-oriented Arabic natural language processing [12]. Table 1 shows different examples of metaphors

Table1: Examples of different metaphors

Arabic metaphor	Transliteration	Translation
تحديات تنال من هوية الأمة واستقرار مجتمعاتها على نحو يستهدف الروابط بين	tHdyAt tnal mn hwyt Alomt wa estqrar mjtmeatha	The challenges that compromise the identity of the nation disrupting the stability of their

دولها وشعوبها و تفكيك نسيج مجتمعاتها	mjtm'eatha 'ela nhw ysthdf alrwabt byn dwlha wsh'ewbha wa tfkyk nsyj mjtm'eatha	societies and specifically target bonds between their states and peoples, causing disintegration of societies relations.
نريد حاليا صناعة التاريخ	Nryd halya snaet AltArykh	We want now to make the history

2. Related Work

In the literature, previous work on using neural networks for Arabic translation has mainly focused on using neural networks to induce an additional feature for phrase-based statistical machine translation systems (see, e.g., [13]; [14]). The paraphrase database project (PPDB) has paraphrase resources for multiple languages, including Arabic. The paraphrases are achieved using parallel bilingual corpora by implementing the pivot method, where one language is used as a bridge or for common meaning representation [15].

Turker-assisted paraphrasing is used to improve English-Arabic MT [16]. A comparison between several paraphrase acquisition techniques on sentential paraphrasing is given in [17], but it does not carry experiments on Arabic sentential paraphrasing. To the best of our knowledge, there is no study that has solved for or adequately covered the metaphor and word sense ambiguity of Arabic language.

3. Neural Paraphrasing

In this section, we present our Arabic paraphrasing approach, which is based on NMT. It uses neural machine translation to first paraphrase the Arabic metaphor to a pivot the language (Modern Standard Arabic) with the same meaning, and then translates it to English. In the following, we shortly overview the basic encoder-decoder NMT framework and then explain how it can be extended to paraphrasing.

3.1 NMT Background

NMT has shown promising results lately [18-20]. Most NMT methods follow the encoder-decoder framework proposed by [21], which, as its name indicates, typically consists of two RNNs. The encoder is a recurrent neural network (RNN) that reads the source sentence and compresses the meaning into a sequence of vector representations. Next, the decoder RNN takes the vector representation and generates the target sentence word by word. The decoder stops once a particular symbol denoting the end of the sentence is generated.

For a language pair, an encoder takes in a source sentence $X = \{x_1, \ldots, x_{T_x}\}$ as a sequence of linguistic symbols and generates a sequence of context vectors $V = \{h_1, \ldots, h_{T_x}\}$. Our Arabic Neural Paraphrasing method uses a bidirectional RNN, where each context vector h_t is the sequence of the forward and the backward RNN's hidden states at time *t*. The decoder is a conditional RNN language model that, given a source sentence, generates a probability distribution over the translation. The decoder's hidden state is updated at each time t':

$$D_{t'} = RNN(D_{t'-1}, y_{t'}, V_{t'}) \ldots..(1)$$

The update uses the previously hidden state zt0−1, the previous target symbol $y_{t'-1}$ and the time-dependent context $V_{t'}$, calculated by an attention mechanism $\alpha_{t,t'}$ over the source sentences' context vectors:

$$V_{t'} = \sum_{t=1}^{TX} \alpha_{t't} h_t \ldots..(2)$$

The probability of the target sentence $Y = \{y_1, \ldots, y\}$ is the product of the probabilities of the symbols within this sentence:

$$P(Y|X) = \prod_{t'=1}^{T_Y} p(y_{t'}|y < t', X) \ldots..(3)$$

Please refer to [19];[22];[23] for more details.

4. Arabic Neural Paraphrasing

Our approach to Arabic paraphrasing is the pivot method which is inspired by Bannard and Callison-Burch [15]. Pivoting is often used in MT to overcome the deficiency of parallel data, i.e., when there is no direct translation path found from the source language to the target. Instead, pivoting takes advantage of indirect paths by an intermediate language.
The concept dates back at least to 1997 when Kay observed that ambiguities in the translation from one language into another may be resolved if a translation through a third language is possible [24]. This approach has met with success in traditional phrase-based SMT [25];[26], and more recently in NMT systems [27]. In our case of paraphrasing, pivoting offers a path from the ambiguities of Arabic to English, through a translation to simple Arabic forms. In other words, we translate a source sentence into a pivot language and then translate the pivoted phrase into the target language. Pivoting using NMT ensures that the entire sentence is considered when choosing a pivot. Contextual information is thus considered when translating, which allows for a more accurate pivoted sentence. This approach also places greater emphasis on capturing the full meaning of the sentence, a crucial aspect of paraphrasing.

To extract paraphrases, we first obtain a parallel corpus through Arabic metaphor and English. We prune the corpus to only those containing sentences with less than 60 words each. We tokenize words of those sentences using the Stanford NLP Arabic Tokenizer [28]. Then, we perform sentence alignment in order to calculate the conditional probabilities for our paraphrase equation. Consequently, we run the corpora on GIZA++ [29], the alignment tool most widely-used with MT involving Arabic [30]. Once we have a database of paraphrase mappings, we can then replace phrases with their corresponding paraphrases by selecting the phrases with the highest probability.

A crude approach to pivoting is one-to-one translation. The ambiguous source sentence *A1* is translated into English phrase E, which has a similar or exact metaphorical meaning. The English E is then translated back into Arabic, producing the intermediate pivot phrase, which gives a probability distribution over English sentences, *E*. This substitution approach was used by Bannard and Callison-Burch [15]. Our approach to obtain a paraphrase is summarized in the following mathematical equation:

$$P(E|A1, A2): P(E|A1, A2) = P(E|A2) \quad (4)$$

In encoder-decoder models, care is taken during each decoding step to indicate which words are the relevant source words. In our case, each word of the paraphrase relates to words in the pivot sentence, and each word in the pivot sentence relates to words in the source sentence.

Table2: An example for Arabic Neural Paraphrasing Approach

Original	احمر وجه الوزير حين ظهر علي الشاشة وطعن في كلامه
Translation	The minister's face blushed when he appeared on the screen and when doubted his words
Paraphrasing	احمر وجه الوزير خجلاً شديداً حين ظهر علي الشاشة وشكك في كلامه

An example of this approach is given in Table 2, where close attention has successfully identified the semantically equivalent parts of two sentences. Beyond providing interpretable paraphrasing, attention scores can be used in both generation and classification tasks. Furthermore, our approach can readily be used to perform text generation via an NMT which takes the advantage of semantic processing offered by the WordNet database [8]. Fig1 shows and example of alignment process.

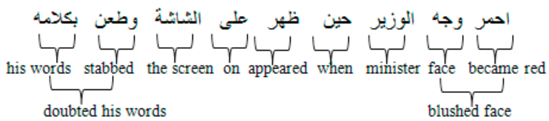

Fig.1: Alignment process example

5. Evaluation

Designing the appropriate automated metrics for evaluating machine translations is challenging due to the variety of acceptable translations for each different sentence. Favorite metrics produce scores mainly based on matching the sequences of words in the system translation to those in one or more reference translations. The metrics primarily differ in how they account for reordering and synonyms.

5.1. Measuring the Results

METEOR [31] is an evaluation measure that calculate a one-to-one alignment between mapping words in a candidate with a reference translation. If a word matches multiple other words, choice is given to the alignment that reorders the words the least, with the amount of reordering estimated by the number of crossing alignments. Alignments are first created for exact matches between words. Additional alignments are then created by repeatedly running the alignment procedure over unaligned words, first allowing for matches between word stems [32], and then allowing matches between words listed as synonyms in WordNet [8]. After realizing the final alignment, METEOR calculates the candidate translation's unigram Precision (P) and Recall (R), $P = \frac{matches}{length\ trans}$, and $R = \frac{matches}{length\ ref}$, respectively. These two values are then gathered into a weighted harmonic mean (5). To penalize reordering's, this value is computed by a fragmentation penalty based on the number of parts the two sentences would need to be broken into to allow them to be reordered with no crossing alignments, $P_{\beta,\gamma} = 1 - \gamma(\frac{chunks}{matches})^{\beta}$

$$F_{\alpha} = \frac{PR}{\alpha P + (1-\alpha)R} \quad (5)$$

$$METEOR_{\alpha,\beta,\gamma} = F_{\alpha}.P_{\beta,\gamma} \quad (6)$$

The free parameters $\alpha, \beta, and\ \gamma$ can be used to tune the metric to display human judgments on a specific language and to adjust to any variation of the evaluation task (e.g., ranking candidate translations vs. reproducing judgments of translations' adequacy and fluency). We used cosine similarity to compute the matches between our proposed system and human translation.

METEOR was implemented to explicitly address the weaknesses in BLEU [33]. It evaluates a translation by computing a score based on explicit word-to-word matches between the translation output and a reference translation. If more than one reference translation is available, the given translation is scored against each reference independently, and the best score is reported [33].

5.2. Setup

The training dataset consists of a corpus of bilingual metaphors comprising 90k sentences extracted manually from Arabic rhetoric books. The corpus includes almost all of the Arabic metaphors with their translation into English by a bilingual group. The group consisted of native speakers of Arabic who had lived in the United States for the past several years and who had worked as annotators.

The test dataset corpus consists of 386 Arabic metaphor sentences, including the headlines. The corpus covers political, sport and art topics extracted using "HTML Text Extractor"[1] [34] from the homepage of the Egypt State Information Service (SIS)[2] over 5 weeks[3] in 2018.

The best evaluation metric to use here is the one that ultimately points to the best translations according to human judges. We performed a human evaluation of selected models using the METEOR score to measures the translation quality averaged over all the sentences in the corpus. These evaluations used two

---

[1] HTML Text Extractor is a program that extracts only raw text (i.e., without HTML or script format)
[2] The web link for the Egypt State Information Service (SIS) is http://www.sis.gov.eg
[3] from April 1st, 2018 to May 5th ,2018

encoder-decoder NMT models (one-to-one pairs): Arabic → English and English → Arabic, as illustrated in Table 3, which displays the METEOR scores with our model and with the Google translator.

Table 3: METEOR score evaluations %

Translation	Our Model	Google translator
Arabic – English	86.9%	53.5%
English – Arabic	94.1%	57.8%

We also calculated the cosine similarity between our model and human judgment, as shown in Table 3. We considered the datasets from SIS for the human judgment result, as they were translated manually, i.e. gold standard dataset. Table 4 shows the high correlation between human evaluation and our model, in both directions. We examined the dataset to determine if the results were biased by sampling or data peculiarities. For sentence pairs, including the headline and the following sentence, both human and evaluation metric scores were high, at 89.9% and 92.1%, respectively.

Table4: Similarity match % for our model and human judgment

Translation	Cosine Similarity
Arabic – English	89.9%
English – Arabic	92.1%

6. Conclusion

We propose a framework for paraphrasing NMT which solves the ambiguity in Arabic metaphors and introduces an auxiliary score to measure the sufficiency of translation candidates. The advantage of the proposed approach is two-fold. First, it improves metaphor translation, thereby producing better translation candidates. Second, it consistently improves the translation performance of NMT by using paraphrasing combined with the use of the pivoting method. We applied the pivoting method to construct a large coverage paraphrase database for Arabic metaphors that includes over 90K phrase pairs. Experimental results show that the two advantages indeed help our approach to improve translation performance consistently, particularly when compared to the Google translator. Our work offers encouraging results in terms of its correlation with human judgment.

7. References

[1] Greenstein, Eric, and Daniel Penner. (2015) "Japanese-to-English Machine Translation Using Recurrent Neural Networks." (2015): 1-7.

[2] Mallinson, Jonathan, 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, vol. 1, pp. 881-893.

[3] Kalchbrenner, Nal, and Phil Blunsom. (2013) "Recurrent continuous translation models." *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing,* pp. 1700-1709.

[5] Cho, Kyunghyun, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. (2014) "On the properties of neural machine translation: Encoder-decoder approaches." *arXiv preprint arXiv*:1409.1259.

[6] Koehn, Philipp, Franz Josef Och, and Daniel Marcu.(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.

[7] Bentivogli, Luisa, Arianna Bisazza, Mauro Cettolo, and Marcello Federico. (2016) "Neural versus phrase-based machine translation quality: a case study." *arXiv preprint arXiv*:1608.04631 (2016).

[8] Alkhatib, Manar, Azza Abdel Monem, and Khaled Shaalan. (2017) "A Rich Arabic WordNet Resource for Al-Hadith Al-Shareef." *Procedia Computer Science* 117 (2017): 101-110.

[9] Isabelle, Pierre, Colin Cherry, and George Foster. (2017) "A challenge set approach to evaluating machine translation." *arXiv preprint arXiv*:1704.07431

[10] Shaalan, Khaled, Ashraf Hendam, and Ahmed Rafea. "Rapid development and deployment of bi-directional expert systems using machine translation technology. (2012) " *Expert Systems with Applications* 39, no. 1 (2012): 1375-1380.

[11] Monem, Azza Abdel, Khaled Shaalan, Ahmed Rafea, and Hoda Baraka. (2008) "Generating Arabic text in multilingual speech-to-speech machine translation framework." *Machine Translation* 22, no. 4 (2008): 205.

[12] Alkhatib, Manar, and Khaled Shaalan.(2018) "The Key Challenges for Arabic Machine Translation." *In Intelligent Natural Language Processing: Trends and Applications*, pp. 139-156. Springer, Cham,.

[13] Devlin, Jacob, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul. "Fast and robust neural network joint models for statistical machine translation.(2014) " *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), vol. 1, pp. 1370-1380.

[14] Setiawan, Hendra, Zhongqiang Huang, Jacob Devlin, Thomas Lamar, Rabih Zbib, Richard Schwartz, and John Makhoul. (2015) "Statistical machine translation features with multitask tensor networks." *arXiv preprint arXiv*:1506.00698.

[15] Bannard, Colin, and Chris Callison-Burch.(2005) "Paraphrasing with bilingual parallel corpora." *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 597-604. Association for Computational Linguistics.

[16] Denkowski, Michael, Hassan Al-Haj, and Alon Lavie. "Turker-assisted paraphrasing for English-Arabic machine translation. (2010). " In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 66-70. Association for Computational Linguistics.

[17] Bouamor, Houda, Aurélien Max, and Anne Vilnat.(2010) "Comparison of paraphrase acquisition techniques on sentential paraphrases." In *International Conference on Natural Language Processing*, pp. 67-78. Springer, Berlin, Heidelberg.

[18] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. (2014.) "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pp. 3104-3112.

[19] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. (2014) "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473*.

[20] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. (2015) " Effective approaches to attention-based neural machine translation". *arXiv preprint arXiv*:1508.04025

[21] Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. (2014) "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078*.

[22] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. (2014) "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pp. 3104-3112.

[23] Cho, Kyunghyun, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. (2014) "Learning phrase representations using RNN encoder-decoder for statistical machine translation." *arXiv preprint arXiv:1406.1078* .

[24] Kay, Martin. (1997) "The proper place of men and machines in language translation." *machine translation* 12, no. 1-2 (1997): 3-23.

[25] Wu, Hua, and Haifeng Wang. (2007) "Pivot language approach for phrase-based statistical machine translation." *Machine Translation* 21, no. 3: 165-181.

[26] Utiyama, Masao, and Hitoshi Isahara.(2007) "A comparison of pivot methods for phrase-based statistical machine translation." In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pp. 484-491.

[27] Firat, Orhan, Baskaran Sankaran, Yaser Al-Onaizan, Fatos T. Yarman Vural, and Kyunghyun Cho. (2016) "Zero-resource translation with multi-lingual neural machine translation." *arXiv preprint arXiv:1606.04164*.

[28] Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. (2014) "The Stanford CoreNLP natural language processing toolkit." In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pp. 55-60.

[29] Och, Franz Josef, and Hermann Ney. (2003) "A systematic comparison of various statistical alignment models." *Computational linguistics* 29, no. 1: 19-51.

[30] Al-Raisi, Fatima, Abdelwahab Bourai, and Weijian Lin. (2018) "NEURAL SYMBOLIC ARABIC PARAPHRASING WITH AUTOMATIC EVALUATION." *Computer Science & Information Technology*: 1.

[31] Cer, Daniel, Christopher D. Manning, and Daniel Jurafsky. (2010) "The best lexical metric for phrase-based statistical MT system optimization." In Human Language Technologies: The 2010 *Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 555-563. Association for Computational Linguistics.

[32] Pasha, Arfath, Mohamed Al-Badrashiny, Mona T. Diab, Ahmed El Kholy, Ramy Eskander, Nizar Habash, Manoj Pooleery, Owen Rambow, and Ryan Roth. (2014) "MADAMIRA: A Fast, Comprehensive Tool for Morphological Analysis and Disambiguation of Arabic." In LREC, vol. 14, pp. 1094-1101.

[33] Banerjee, Satanjeev, and Alon Lavie. (2005) "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments." *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*.

[34] Sakre, Mohammed M., Mohammed M. Kouta, and Ali MN Allam. (2009) "Automated Construction of Arabic-English Parallel Corpus." *Journal of the Advances in Computer Science 3*.