

# Grammatical Error Correction Considering Multi-word Expressions

Tomoya Mizumoto Masato Mita Yuji Matsumoto

Nara Institute of Science and Technology  
8916-5 Takayama, Ikoma, Nara 630-0192, Japan

{tomoya-m, mita.masato.mz2, matsu}@is.naist.jp

## Abstract

Multi-word expressions (MWEs) have been recognized as important linguistic information and much research has been conducted especially on their extraction and interpretation. On the other hand, they have hardly been used in real application areas.

While those who are learning English as a second language (ESL) use MWEs in their writings just like native speakers, MWEs haven't been taken into consideration in grammatical error correction tasks. In this paper, we investigate the grammatical error correction method using MWEs. Our method proposes a straightforward application of MWEs to grammatical error correction, but experimental results show that MWEs have a beneficial effect on grammatical error correction.

## 1 Introduction

Publicly usable services on the Web for assisting second language learning are growing recently. For example, there are language learning social networking services such as Lang-8<sup>1</sup> and English grammar checkers such as Ginger<sup>2</sup>. Research on assistance of second language learning also has received much attention, especially on grammatical error correction of essays written by learners of English as a second language (ESL). In the past, three competitions for grammatical error correction have been held: Helping Our Own (Dale and Kilgarriff, 2011; Dale et al., 2012) and CoNLL Shared Task (Ng et al., 2013; Ng et al., 2014).

<sup>1</sup><http://lang-8.com>

<sup>2</sup><http://www.gingersoftware.com>

Most previous research on ESL learners' grammatical error correction is targeted on one or few restricted types of learners' errors. ESL learners make various kinds of grammatical errors (Mizumoto et al., 2012). For dealing with any types of errors, grammatical error correction methods using phrase-based statistical machine translation (SMT) are proposed (Brockett et al., 2006; Mizumoto et al., 2012). Phrase-based SMT carries out translation with phrases which are a sequence of words as translation units. However, since phrases are extracted in an unsupervised manner, an MWE like "a lot of" may not be treated as one phrase. In machine translation fields, phrase-based SMT considering MWEs achieved higher performance (Carpuat and Diab, 2010; Ren et al., 2009).

In this paper, we propose a grammatical error correction method considering MWEs. To be precise, we apply machine translation methods considering MWEs (Carpuat and Diab, 2010) to grammatical error correction. They turn MWEs into single units in the source side sentences (English). Unlike typical machine translation that translates between two languages, in the grammatical error correction task, source side sentences contain errors. Thus, we propose two methods; one is that MWEs are treated as one word in both source and target side sentences, the other is that MWEs are treated as one word in only the target side sentences.

## 2 Related work

Research on grammatical error correction has recently become very popular. Grammatical error correction methods are roughly divided into two types; (1) targeting few restricted types of errors (Rozovskaya and Roth, 2011; Rozovskaya and Roth, 2013; Tajiri et al., 2012) and (2) targeting

any types of errors (Mizumoto et al., 2012). In the first type of error correction, classifiers like Support Vector Machines have mainly been used. In the second type, statistical machine translation methods have been used. The only features for grammatical error correction that have been considered in many of previous works are token, POS and syntactic information of single words, and features considering two (or more) words as a whole such as MWEs have never been used.

There is the work dealing with collocations, a kind of MWEs, as target of error detection (Futagi et al., 2008). Our method is different in that we are aiming at correcting not MWEs but other expressions like articles, prepositions and noun numbers as targets considering MWEs.

A lot of research for identifying MWEs and constructing MWE resources have been conducted (Schneider et al., 2014; Shigeto et al., 2013). In addition, there is some research in natural language processing applications using MWEs; i.e., statistical machine translation (Carpuat and Diab, 2010; Ren et al., 2009), information retrieval (Newman et al., 2012) and opinion mining (Berend, 2011).

Our task is very similar to the research of SMT using MWEs (Carpuat and Diab, 2010; Ren et al., 2009). However we are in different situation where incorrect words may be included in source sentence side, thus identifying MWEs in source side may make mistakes.

### 3 Multi-word expressions

MWEs are defined as expressions having “idiosyncratic interpretations that cross word boundaries (or spaces)” (Sag et al., 2002). In this paper, we mainly deal with fixed expressions that function either as adverbs, conjunctions, determiners, prepositions, prepositional phrases or pronouns.

#### 3.1 Multi-word expressions in native corpora and learner corpora

ESL learners also use a lot of MWEs in their writings just like native speakers. For comparing MWEs usages of ESL learners and native speakers, we prepare a native corpus and a learner corpus. We use the MWE data set from (Shigeto et al., 2013), MWE-annotated Penn Treebank sections of OntoNotes Release 4.0<sup>3</sup> as the native corpus. We

<sup>3</sup><https://catalog.ldc.upenn.edu/LDC2011T03>

Table 1: The rate of overlap of multi-word expressions from Penn Treebank section of OntoNotes and Lang-8 Learner Corpora

top number	rate of overlap
10	30.0%
20	45.0%
30	46.7%
40	57.5%
50	54.0%
70	57.1%
120	66.7%
170	66.5%

use Lang-8 Learner Corpora<sup>4</sup> as the learner corpus<sup>5</sup>.

Table 1 shows the rate of overlap of multi-word expressions from Penn Treebank section of OntoNotes and Lang-8 Learner Corpora in taking top  $N$ . Although they are in different domains, MWEs used by learners overlap about 60% with those used by native speakers.

The occurrence frequency of MWEs obeys the Zipf’s law. In the learner corpus, top 70 MWEs cover about 50%, top 120 MWEs cover about 80% and top 170 MWEs cover 90% of all the MWEs in the corpus by token count.

#### 3.2 Advantage of using Multi-word Expressions for Grammatical Error Correction

There are two advantages to use MWEs in grammatical error correction. The first advantage is that it prevents translation of correct parts of MWEs to other words. To illustrate this, let us consider the following example:

He ate sweets, for example ice and cake.

This sentence does not have grammatical errors, thus error correction systems does not need to correct it. However, the system might correct the word “example”, into the following:

He ate sweets, for examples ice and cake.

This is because the system has no knowledge of MWEs.

<sup>4</sup><http://cl.naist.jp/nldata/lang-8/>

<sup>5</sup>MWEs are automatically tagged by tools which explained in 5.1.

The second advantage is that the system becomes capable of considering longer contexts when using MWEs. To illustrate this, let us consider the following example:

I have a lot of red apple.

Without considering MWEs, the system takes “I have a”, “have a lot”, “a lot of”, “lot of red”, “of red apple” as word 3-grams, unable to consider the relationship between “a lot of” and “apple”.

#### 4 Grammatical error correction methods using multi-word expressions

In this section, we describe our error correction method with MWEs. We use statistical machine translation approaches for grammatical error correction. We apply MWEs to the phrase-based SMT.

##### 4.1 Error correction with phrase-based SMT

The error correction method with phrase-based SMT was proposed for the first time by (Brockett et al., 2006). Although they used phrase-based SMT for grammatical error correction, they only handled one error type, *noun number*. Mizumoto et al. (2012) also used phrase-based SMT, however they targeted all error types. In this paper, we use phrase-based SMT which many previous research used for grammatical error correction.

##### 4.2 Error correction methods considering multi-word expressions

We propose two methods for grammatical error correction considering MWEs. Previous research of machine translation using MWEs (Carpuat and Diab, 2010) handled MWEs in source side sentences by simply turning MWEs into single units (by conjoining the constituent words with underscores). We essentially apply their method to grammatical error correction; however, in our case identifying MWEs might fail because source side sentences contain grammatical errors. Therefore, we propose and compare the following two methods.

**Using MWEs in both source side and target side** In this method, MWEs are considered in both source side and target side. We show an example in the following:

Source: I have a.lot.of pen.  
Target: I have a.lot.of pens.

Table 2: Results of grammatical error correction

		P	R	F
Baseline (without MWEs)		30.1	32.9	31.4
Source: w/ MWE, Target: w/ MWE	70 (50%)	27.3	37.8	31.7
	120 (80%)	30.0	34.9	32.2
	170 (90%)	27.9	38.2	32.3
	All	29.2	32.8	30.9
Source: w/o MWE, Target: w/ MWE	70 (50%)	30.1	35.1	32.4
	120 (80%)	29.3	36.9	32.9
	170 (90%)	29.8	36.7	32.9
	All	31.3	29.4	30.4

**Using MWEs in target side** In this method, MWEs are considered only in target side. We show an example in the following:

Source: I have a lot of pen.  
Target: I have a.lot.of pens.

We train both language model and translation model using texts of considering MWEs.

## 5 Experiments of grammatical error correction using multi-word expressions

### 5.1 Experimental settings

We used cicada 0.3.0<sup>6</sup> for the machine translation tool. This includes a decoder and a word aligner. As the language modeling tool we used expgram 0.2.0<sup>7</sup>. We used ZMERT<sup>8</sup> as the parameter tuning tool.

For automatic identifying MWEs, we use AMALGr 1.0<sup>9</sup> (Schneider et al., 2014). The MWE identification tool is re-trained using the MWE data set tagged by (Shigeto et al., 2013) on the Penn Treebank sections of OntoNotes Release 4.0. This is because their annotation was more convenient for our purpose.

The translation model was trained on the Lang-8 Learner Corpora v2.0. We extracted English essays which were written by ESL learners whose native language is Japanese from the corpora and cleaned the noise with the method proposed in (Mizumoto et al., 2011). As the results, we got 629,787 sentence pairs. We used a 5-gram

<sup>6</sup>[http://www2.nict.go.jp/univ-com/multi\\_trans/cicada/](http://www2.nict.go.jp/univ-com/multi_trans/cicada/)

<sup>7</sup>[http://www2.nict.go.jp/univ-com/multi\\_trans/expgram/](http://www2.nict.go.jp/univ-com/multi_trans/expgram/)

<sup>8</sup><http://cs.jhu.edu/~ozaidan/zmert/>

<sup>9</sup><https://github.com/nschneid/pysupersensetagger>

Table 3: Examples of system outputs

Learner	Last month, she gave me a lot of rice and <u>onion</u> .
Baseline	Last month, she gave me a lot of rice and <u>onion</u> .
with MWE	Last month, she gave me a lot of rice and <u>onions</u> .

language model built on corrected sentences of the learner corpora. Konan-JIEM Learner Corpus<sup>10</sup> (Nagata et al., 2011) are used for evaluation and development data. We use 2,411 sentences for evaluation, and 300 sentences for development.

## 5.2 Experimental Result

As evaluation metrics, we use precision, recall and F-score. We compare phrase-based SMT without using MWEs (baseline) with the two methods explained in 4.2. In addition, we varied the number of MWEs used for training the translation model and the language model. This is because MWEs that appear few times may introduce noises. We use top 70 (50%), 120 (80%) and 170 (90%) MWEs described in 3.1.

Table 2 shows the experimental results. The methods considering MWEs achieved higher F-score than baseline except for the case that uses All MWEs. In addition, using more MWEs increases the F-score.

## 5.3 Discussion

Using all MWEs shows worse results because infrequent MWEs become noise in training and testing.

We got better results when we use MWEs only in the target side. This is likely because learners tend to fail to write MWEs correctly, only writing them in partial forms. One cause of deterioration of precision is that a single word like “many” is wrongly corrected into an MWE like “a lot of”, although it is actually not incorrect.

There are two reasons why the performance improved considering MWEs. The first reason is that the system becomes capable of considering the relationship between MWEs which are made up of a sequence of two or more lexemes and words lie adjacent to MWEs. We show an example of system results in Table 3. Although the baseline system did not correct the example, the system considering MWEs was able to correct this error. This is

<sup>10</sup><http://www.gsk.or.jp/en/catalog/gsk2015-a/>

because the system was able to consider the MWE “a lot of”.

The second reason is that the probabilities of translation model and language model are improved by handling MWEs as single units. Let us consider the two sentences, “There are a lot of pens” and “There is a pen.” as examples of language model. Without considering MWEs, the word 3-grams, “There are a” and “There is a”, have high probability. With considering MWEs, however, the former trigram becomes to “There are a\_lot\_of pens” and then the probabilities of trigrams that should not be given high probability like “There are a” come to low. The correction performance of articles and prepositions that are likely to become a component word of MWEs is considered to improve by this revision. The number of true positive for article as compared with baseline and MWE (170) of only target side are 190 and 227, respectively. Likewise, the number of true positive for preposition as compared with them are 108 and 121, respectively.

## 6 Conclusion

We proposed a grammatical error correction method using multi-word expressions. Our method proposes a straightforward application of MWEs to grammatical error correction, but experimental results show that MWEs have quite good effects on grammatical error correction. Experimental results show that the methods considering MWEs achieved higher F-score than baseline except for the case that uses all MWEs. We plan to use more multi-word expressions which we did not handle in this paper, such as phrasal verbs. Moreover, we plan to conduct grammatical error correction considering MWEs which contain gaps that are dealt with (Schneider et al., 2014).

## References

- Gábor Berend. 2011. Opinion Expression Mining by Exploiting Keyphrase Extraction. In *Proceedings of IJCNLP*, pages 1162–1170.
- Chris Brockett, William B. Dolan, and Michael Ga-

- mon. 2006. Correcting ESL Errors Using Phrasal SMT Techniques. In *Proceedings of COLING-ACL*, pages 249–256.
- Marine Carpuat and Mona Diab. 2010. Task-based Evaluation of Multiword Expressions: a Pilot Study in Statistical Machine Translation. In *Proceedings of HLT-NAACL*, pages 242–245.
- Robert Dale and Adam Kilgarriff. 2011. Helping Our Own: The HOO 2011 Pilot Shared Task. In *Proceedings of ENLG*, pages 242–249.
- Robert Dale, Ilya Anisimoff, and George Narroway. 2012. HOO 2012: A Report on the Preposition and Determiner Error Correction Shared Task. In *Proceedings of BEA*, pages 54–62.
- Yoko Futagi, Paul Deane, Martin Chodorow, and Joel Tetreault. 2008. A Computational Approach to Detecting Collocation Errors in the Writing of Non-Native Speakers of English. *Computer Assisted Language Learning*, 21(4):353–367.
- Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. 2011. Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. In *Proceedings of IJCNLP*, pages 147–155.
- Tomoya Mizumoto, Yuta Hayashibe, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. 2012. The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings. In *Proceedings of COLING*, pages 863–872.
- Ryo Nagata, Edward Whittaker, and Vera Sheinman. 2011. Creating a Manually Error-tagged and Shallow-parsed Learner corpus. In *Proceedings of ACL-HLT*, pages 1210–1219.
- David Newman, Nagendra Koilada, Jey Han Lau, and Timothy Baldwin. 2012. Bayesian Text Segmentation for Index Term Identification and Keyphrase Extraction. In *Proceedings of COLING*, pages 2077–2092.
- Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadiwinoto, and Joel Tetreault. 2013. The CoNLL-2013 Shared Task on Grammatical Error Correction. In *Proceedings of CoNLL Shared Task*, pages 1–12.
- Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, Raymond Hendy Susanto, and Christopher Bryant. 2014. The CoNLL-2014 Shared Task on Grammatical Error Correction. In *Proceedings of CoNLL Shared Task*, pages 1–14.
- Zhixiang Ren, Yajuan Lü, Jie Cao, Qun Liu, and Yun Huang. 2009. Improving Statistical Machine Translation Using Domain Bilingual Multiword Expressions. In *Proceedings of Workshop on MWE*, pages 47–54.
- Alla Rozovskaya and Dan Roth. 2011. Algorithm Selection and Model Adaptation for ESL Correction Tasks. In *Proceedings of ACL-HLT*, pages 924–933.
- Alla Rozovskaya and Dan Roth. 2013. Joint Learning and Inference for Grammatical Error Correction. In *Proceedings of EMNLP*, pages 791–802.
- Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann A. Copestake, and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. In *Proceedings of CICLing*, pages 1–15.
- Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith. 2014. Discriminative Lexical Semantic Segmentation with Gaps: Running the MWE Gamut. *Transactions of the Association for Computational Linguistics*, 2:193–206.
- Yutaro Shigeto, Ai Azuma, Sorami Hisamoto, Shuhei Kondo, Tomoya Kouse, Keisuke Sakaguchi, Akifumi Yoshimoto, Frances Yung, and Yuji Matsumoto. 2013. Construction of English MWE Dictionary and its Application to POS Tagging. In *Proceedings of Workshop on MWE*, pages 139–144.
- Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. 2012. Tense and Aspect Error Correction for ESL Learners Using Global Context. In *Proceedings of ACL*, pages 198–202.