**Paris is rain.* or *It is raining in Paris?*: Detecting Overgeneralization of *Be*-verb in Learner English

Ryo Nagata[1,2][0000−0002−5893−3335], Koki Washio[3][0000−0003−1385−4067], and Hokuto Ototake[4][0000−0002−6502−5570]

[1] Konan University, 8-9-1 Okamoto, Higashinada, Kobe, Hyogo 658-8501, Japan
nagata-cicling@ml.hyogo-u.ac.jp
[2] Japan Science and Technology Agency, PRESTO, 4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan
[3] The University of Tokyo, 3-8-1 Komaba, Meguroku, Tokyo 153-8902, Japan
kokiwashio@g.ecc.u-tokyo.ac.jp
[4] Fukuoka University, 19-1 Nanakuma 8-Chome, Fukuoka 814-0180, Japan
ototake@fukuoka-u.ac.jp

Abstract. This paper addresses the detection of overgeneralization of *be*-verb found in learner English. It is an error where the subject and complement are not semantically equivalent in a *be*-verb sentence as in **Paris is rain.* This type of error often appears in the writing of learners whose native language has a *be*-verb equivalent that has usages other than those which English *be*-verb does. This paper presents a method for detecting overgeneralization of *be*-verb by predicting through word embeddings whether a given subject and complement pair is semantically equivalent or not. It also presents a method for determining the hyperparameters in the method efficiently and effectively. Experiments show that the present method outperforms four baseline methods based on corpus statistics and WordNet ontology despite the fact that it is a rather simple method. Looking into the detection results reveals the performance limitations of the method.

Keywords: Grammatical error detection/correction · Overgeneralization of *be*-verb · Learner English.

1 Introduction

Although grammatical error detection/correction has made tremendous progress with the advent of neural network-based methods, there still exist errors to which researchers have paid far less attention. **Overgeneralization of *be*-verb**, as in **Paris is rain.*, is one of the typical examples; as far as we know, there has been no work on its detection. It is an error where *be*-verb links the subject and complement which are not semantically equivalent[5]; in the example sentence

[5] The precise definition is introduced in Sect. 2.

above, the subject *Paris* is NOT *rain*, which violates the basic English rule of *be*-verb (or strictly copula *be*), that subject-complement [11] should be semantically equivalent to its subject.

Overgeneralization of *be*-verb often appears in learner English. This is especially true for learners whose native language has a *be*-verb equivalent that has usages other than English *be*-verb does. A typical example is Japanese. For instance, the following expressions are valid in the corresponding Japanese expressions: **Airplanes are danger.* (correctly *Airplanes are dangerous.*), **The meeting is five.* (correctly, *The meeting is scheduled at five.* or *The meeting is at five.*), and **Paris is rain* (correctly, *It is raining in Paris.*)[6]. In other words, overgeneralization of *be*-verb is (at least partly) ascribed to language transfer or mother tongue interference. This suggests that it will likely be beneficial to such groups of learners to give them feedback explaining how *be*-verb functions in English and why such expressions as **Paris is rain.* and **Airplanes are danger.* are not correct. To achieve it, one has to recognize overgeneralization of *be*-verb in learner English in the first place, distinguishing it from the other error types.

Unfortunately, however, previous error detection/correction methods would not suit this application. Previous error-specific methods, which typically rely on an unannotated native corpus, solve error detection/correction as a classification problem as in article and preposition error detection/correction methods [4, 6] (i.e., selecting the correct article or preposition). This way of detection/correction does not apply well to this type of error; it is not a problem to select the correct *be*-verb, but to determine whether a given subject and complement pair is semantically equivalent. It is not trivial at all how to detect this type of error as a classification problem, relying solely on an unannotated native corpus. Another typical way is to predict directly whether a given word or phrase is correct or not as found in the method [6]. However, it would not be practical at all considering the fact that it requires a learner corpus annotated with overgeneralization of *be*-verb, which are rare at present; as far as we know, such publicly available data do not exist. Machine translation-based methods [7] and neural network-based methods [2, 14] would probably be capable of detecting/correcting part of overgeneralization of *be*-verb, but not of distinguishing it from the other error types because they detect/correct multiple error types simultaneously. It is crucial to detect it as overgeneralization of *be*-verb in order to realize such feedback as mentioned above. Besides, it is often the case that overgeneralization of *be*-verb requires the rewrite of the whole structure as in **Paris is rain.* $\rightarrow$ *It is raining in Paris.*, which would be difficult for all the methods above.

In view of this background, this paper presents a method for detecting overgeneralization of *be*-verb. It uses word embedding vectors (simply, word embeddings) to predict whether a given subject and complement pair is semantically equivalent or not, which plays the central role in the error detection procedure.

[6] Admittedly, such expressions as *I am coffee.* can also be correct in English. However, they are used in limited contexts and situations, and the usage rarely appears in the writings of learners of English (e.g., essay writing).

They have been shown to be effective in detecting historical meaning changes [3] and differences in meaning between loan words and their originals [12]. These results imply that they will also likely be effective in the present task. The present method detects the triple of *be*-verb, its subject, and its complement as overgeneralization of *be*-verb if a given pair is predicted not to be semantically equivalent.

The contributions of this paper are four-fold: (i) it presents the first-ever method that detects overgeneralization of *be*-verb; (ii) it also presents a method for determining the hyperparameters of the method efficiently and effectively; (iii) the resulting method significantly outperforms four baselines based on corpus statistics and WordNet [9] ontology; (iv) it investigates detection results to show the performance limitations empirically and theoretically.

2 *Be*-verb Sentence and Overgeneralization of *be*-verb

In this paper, the *be*-verb sentence, which is the target of error detection, is defined as follows:

***be*-verb sentence**: sentence consisting of S, V, C where S is a noun phrase (NP) that is the subject of the sentence, V is the *be*-verb, and C is also an NP that is its subject-complement [11].

Hereafter, subject-complement will be referred to just as complement.

Under this definition, the *be*-verb sentence has the following two basic usages [11]:

(a) Identification:
e.g., Kevin is my brother.

(b) Characterization:
e.g., Dwight is an honest man.

The usages (a) and (b) are roughly summarized as the rule that the subject and complement in a *be*-verb sentence should be semantically equivalent. For example, *Kevin* is a *brother* and *Dwight* is a *man*. Hereafter, the rule will be referred to as **subject complement equivalence**.

Learners of English often violate this rule as in the examples we have already seen in Sect. 1. This is especially true for the writer whose native language has a *be*-verb equivalent that has usages other than the above two. A typical example is Japanese; the three erroneous examples are all valid in Japanese. This type of error is defined as **overgeneralization of *be*-verb** in this paper. Part of the reasons why it occurs is that the other usages in the native language are negatively transferred into English. Considering this, it would be useful to explain to this group of learners why the usage is erroneous and how English *be*-verb functions.

Note that only characterization attributes normally allow reversal of subject and complement without affecting the semantic relation. Learners might also violate this. However, this paper excludes this type of error from the detection target. We will discuss this problem again in Sect. 5.

3 Proposed Method

3.1 Detection Procedure

The procedure of the proposed method is as follows:

Step (1) Input
Step (2) Subject complement pair extraction
Step (3) Subject complement equivalence check
Step (4) Error detection
Step (5) Postprocessing
Step (6) Output

In Step (1), each sentence is read from the detection target text. In Step (2), its dependency structure is obtained by using a parser. Then, its subject complement pairs are extracted from the parse (if any); only head nouns are extracted. In Step (3), the extracted pairs are examined as to whether they are semantically equivalent or not based on word embedding. The details are described in Subsect. 3.2. In Step (4), it is determined whether they are correct or not; if they are predicted to be semantically equivalent in Step (3), then they are judged to be correct; otherwise, erroneous (overgeneralization of *be*-verb). In Step (5), as a postprocessing step, those that contain one of the following words are filtered out to achieve a better error detection performance: *it*, *they*, *this*, *these*, *that*, *those*, *thing*, *things*. These are the words that can refer to a wide variety of things, and thus it would be hard to predict whether a given pair is semantically equivalent or not. Filtered-out pairs are always judged to be correct. Finally, in Step (6), the detection result is output, either 0 (correct) or 1 (erroneous). Alternatively, the triple (subject, *be*-verb, and complement) are marked in the target sentence when the result is erroneous.

3.2 Subject complement Equivalence Check

Figure 1 shows the big picture of subject complement equivalence check. As already mentioned, subject complement equivalence check is done based on word embeddings. They are learned in advance from a large native corpus. Before learning, all words are put into lowercase to decrease the vocabulary size.

To formalize the subject complement equivalence check procedure, let s and c be a subject and its corresponding complement in a *be*-verb sentence, respectively. Also, let $\boldsymbol{v}_s$ and $\boldsymbol{v}_c$ be their corresponding word embeddings (i.e., vectors), respectively, and $\cos(\boldsymbol{v}_s, \boldsymbol{v}_c)$ be the cosine similarity between the two.

Then, the check is done by the following function with a threshold θ:

$$f(\boldsymbol{v}_s, \boldsymbol{v}_c) = \begin{cases} 1 \; (\cos(\boldsymbol{v}_s, \boldsymbol{v}_c) > \theta \\ 0 \; (\text{otherwise}) \end{cases} \tag{1}$$

where 1 and 0 denote that the pair is semantically equivalent and not, respectively (Subsect. 3.3 will shortly describe how to determine the threshold θ).

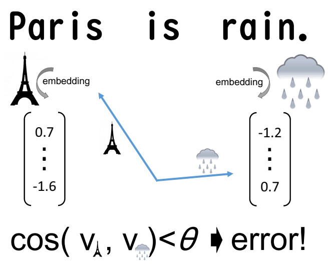

Fig. 1. Illustration of Subject Complement Equivalence Check.

Eq (1) can be interpreted as follows. The similarity between subject and complement is measured by the cosine of the corresponding two vectors. Then, the check whether or not the subject complement pair is semantically equivalent is approximated to be the problem of determining whether or not the pair is similar enough in terms of the angle between the two vectors as illustrated in Fig. 1. This may seem a too crude approximation, but it works well in practice as shown in Sect. 4. For the moment, take as an example the erroneous sentences **Paris is rain.*, **Airplanes are danger*, and **The meeting is five*. Their corresponding cosine similarities are 0.05, 0.08, and −0.04, respectively, which shows that their vectors are almost orthogonal[7]. This observation agrees well with our intuition that each subject has little or nothing to do with its corresponding complement. Unlike these erroneous pairs, the values become much larger for correct expressions such as *Paris is a city.* (0.24), *Airplanes are a machine.* (0.21), *The meeting is a gathering (of someone).* (0.57).

3.3 How to Determine Hyperparameters

Performance of the proposed method greatly depends on the values of its hyperparameters including the threshold[8] θ in Eq. (1) and those for word embeddings such as the dimension of the vector and the window size. Ideally, it would be best to determine them with a development set, that is, a learner corpus with which overgeneralization of *be*-verb errors are manually annotated. Unfortunately, however, it would be not possible considering that there exists no such learner corpus at present.

To overcome this problem, the proposed method automatically generates pseudo-training data from a native corpus as shown in Fig. 2. To achieve this, it first extracts subject complement pairs from a native corpus just as in Step (2) described in Subsect. 3.1 ((1) *Extraction* in Fig. 2). It discards the pairs whose

[7] The cosine similarities were calculated by the word embeddings with the window size of 10 and the dimension of 200 whose details are described in Sect. 4.

[8] Strictly, the threshold θ is rather a parameter of the method than a hyperparameter. However, it will be referred to as a hyperparameter for the simplicity of explanation in this paper.

subject or complement does not appear[9] in a learner corpus in order to create a learner corpus-like training data set ((2) *Filtering* in Fig. 2). It would be safe to say that (almost) all these pairs are free from overgeneralization of *be*-verb because they are from a native corpus. In other words, they can be regarded as correct instances consisting of subjects and complements that likely appear in learner English. It then generates pseudo-erroneous pairs from them by sampling out their subjects and complements independently ((3) *Random Sampling* in Fig. 2). Namely, it randomly chooses a subject from one of them and a complement from another to make a pseudo-erroneous pair. An exception is that it excludes those already found in the correct instance set to avoid including possibly correct pairs. Finally, it merges correct and pseudo-erroneous instances into a pseudo-training data set ((4) *Merging* in Fig. 2).

One thing we should take care of is that we have to determine the ratio of pseudo-erroneous instances to correct ones, which corresponds to the error ratio of overgeneralization of *be*-verb in learner English. This paper assumes that it is empirically given considering the fact that a learner corpus annotated with overgeneralization of *be*-verb is not publicly available.

The proposed method uses the resulting pseudo-training data just as a standard development set. In other words, it applies Step (3)–(6) in Subsect. 3.1 to them to estimate its performance with an arbitrary set of values for the hyperparameter; it selects the setting that maximizes performance (F-measure, for example).

Here, it should be emphasized that the present development set is a pseudo-one (i.e., automatically generated). For this reason, the estimated best setting may not perform well on a learner corpus. Even if it were a real one, the resulting setting may not, suffering from other problems such as overfitting.

To reduce the problem, the proposed method takes a vote of the detection results obtained through word embeddings with different settings (different window sizes and different dimensions, for example). For this, it learns a number of them from a native corpus. It determines the threshold θ for each of them by the same method as above. Finally, it takes a vote of their detection results (whether correct or not) to determine whether a given pair is really correct or not. This way of detection will likely reduce the influence from the problems above.

4 Evaluation

We chose the Konan-JIEM Learner Corpus fifth edition [10] (KJ) as our detection target. We manually annotated it with overgeneralization of *be*-verb. We first extracted *be*-verb sentences, and in turn subject complement pairs from it by using the LexicalizedParser of Stanford Parser Ver.3.5.0[10]. As a result, we

[9] Note that occurrences other than as a subject or a complement are considered when the subject and complement in question are checked whether they appear in learner English.

[10] `https://nlp.stanford.edu/software/lex-parser.shtml`. Sentences longer than 50 tokens are excluded from parsing.

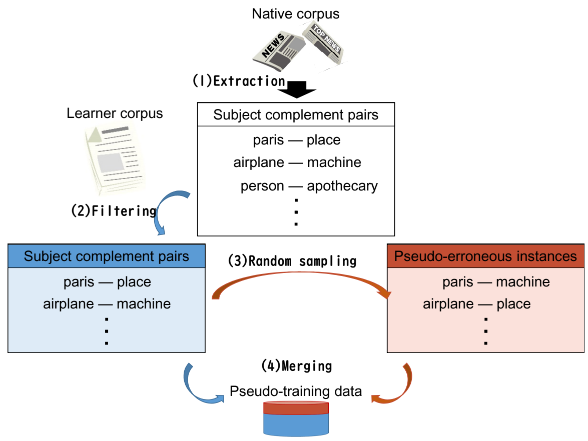

Fig. 2. Procedure for Generating Pseudo-training Data Set.

obtained 294 subject complement pairs. The first and second authors independently annotated them with *correct* or *erroneous*. After that, they discussed and solved disagreements, which identified 83 errors.

We obtained word embeddings[11] from news.en-00001-of-00100 to news.en-00099-of-00100 of one Billion Word Language Model Benchmark[12]. We used 15 sets of them with different settings of their hyperparameters (the combinations of the dimensions ranging over 200, 400, 800 and the window sizes ranging over 5, 10, 15, 20, 25). We did not include words appearing less than five times in the word embeddings; we did not apply the proposed method to subject complement pairs containing one of these words (i.e., they were always regarded as correct when they appeared in a subject complement pair in KJ). The other hyperparameters were fixed as described in the footnote.

We used the same native corpus to determine the values of the threshold θ in Eq (1). We applied the method described in Subsect. 3.3 to it to obtain a pseudo-training set. We set the ratio of pseudo-erroneous instances to the whole training data to 28% (=83/294), which equals the error rate in KJ. This resulted in a pseudo-training set consisting of 155,402 correct instances and 60,434 erroneous instances. We selected the value of the threshold θ that maximizes F-measure on it, ranging over $0 \leq \theta \leq 1$ with an interval of 0.01. Note that because we used the prior knowledge about the error rate in the target corpus, the evaluation was not strictly done by a blind test.

For comparison, we implemented four baseline methods. The first one was simply based on co-occurrence of the subject complement pair in question. Namely, it detected as errors those that did not appear in the native corpus above. The second one was based on Pointwise Mutual Information (PMI) be-

[11] We used the word2vec software (`https://github.com/tmikolov/word2vec`) with the options: -negative 25 -sample 1e-4 -iter 15 -cbow 1 -min-count 5

[12] `http://www.statmt.org/lm-benchmark/`

Table 1. Detection Performance.

Method	Accuracy	Recall	Precision	$F_{1.0}$
Proposed (given best hyperparameter)	0.786	0.614	0.622	0.618
Proposed (estimated best hyperparameter, voting)	**0.762**	0.578	**0.578**	**0.578**
Proposed (estimated best hyperparameter)	0.643	0.639	0.414	0.502
Majority class	0.713	—	—	—
Co-occurrence	0.673	0.530	0.436	0.478
Lin's similarity (given best hyperparameter)	0.599	0.735	0.389	0.509
PMI (given best hyperparameter)	0.480	0.759	0.321	0.451
Is-a recognition (given best hyperparameter)	0.411	**0.816**	0.282	0.419

tween a given subject complement pair. It detected as errors those whose PMI was smaller than a threshold. We set it to the one that maximized F-measure on KJ to show the upper bound of its performance. We defined the probability of co-occurrence in PMI as that of subject complement pairs in the *be*-verb sentences in the native corpus. We also defined the probability of single word occurrence as the unigram probability of each word in the native corpus. We estimated all probabilities by Laplace Estimator. The third one was based on the Lin's similarity [8] calculated from WordNet. Similar to the second one, it detected as errors those whose similarity was less than a threshold, which we determined the same way as in the PMI-based method. The fourth one was an adaptation of the neural network-based method [13] for recognizing semantic relations between given word pairs. We trained it[13] so that it can predict whether a given word pair has the *is-a* relation or not; we detected it as an error if a given pair was predicted not to have the *is-a* relation. We used the same corpora and the same parser as in the proposed method to implement the four baselines. Also, we excluded from error detection subject complement pairs that appeared less than five times in KJ (they were always predicted to be correct as in the proposed method). Note that the evaluation for the PMI-based and Lin's similarity-based methods were also not strictly done by a blind test because their thresholds were optimized on KJ.

To evaluate performance, we used accuracy, which was defined as the number of instances whose subject complement equivalence was correctly predicted divided by the number of subject complement pairs. We also used recall, precision, and F-measure.

Table 1 shows the results. For comparison, Table 1 includes the performances of the proposed method with the hyperparameter setting optimized on KJ (threshold $\theta = 0.10$; window size: 10 words; and dimension: 200), that with the best setting estimated from the pseudo-training data ($\theta = 0.10$; window size: 5 words; and dimension: 800) without voting, and another baseline that always predicts as correct; they are denoted as *Proposed (given best hyperparameter)*, *Proposed (estimated best hyperparameter)*, and *Majority class* in Table 1, respectively.

[13] BLESS[1], which contains the information about semantic relations, was used as the training data for recognizing semantic relations.

Table 1 reveals that the co-occurrence-based method is a strong baseline, which outperforms the other two baselines that exploit other sources of information (corpus statistics and WordNet). Besides, the thresholds in the latter two are optimized on KJ. Nevertheless, the former achieves a better accuracy. This implies that subject complement co-occurrences obtained from a large native corpus is a good source of evidence to tell that a given pair is correct. At the same time, its accuracy is still low even compared to the majority class baseline. This also implies that it suffers from the data sparseness problem even when it uses such a large native corpus as the one Billion Word Language Model Benchmark. This probably applies to the PMI-based method, too. In addition, PMI is suitable for measuring how correlated a pair is but not to predict whether it is semantically equivalent or not. Contrary to our expectation, the Lin's similarity-based method does not perform well either even though it is based on a kind of semantic knowledge (WordNet). Note that the similarity is not defined for pairs containing a proper noun or a pronoun (except I) because they are not included in the synsets of WordNet. This partly explains why it does not perform well. Similarly, the method based on *is-a* recognition does not work well either. Part of the reasons is ascribed to the miss-match between the target language (learner English) and the semantic relation corpus *BLESS* (native English).

In contrast, the proposed method with voting outperforms even the best-performing baseline (co-occurrence) both in accuracy and F-measure; the difference in accuracy is statistically significant at a significance level of 0.05 (McNemar's test, $p = 0.015$). Importantly, its accuracy and precision are especially high compared to the three baselines. This property of the proposed method is particularly preferable in applications to language learning assistance that put more emphasis on precision over recall.

Given the best setting of the hyperparameters, the proposed method improves further, achieving an accuracy of 0.786 and an F-measure of 0.618. In contrast, its performance degrades when it relies solely on the best estimated setting from the pseudo-training data set, suggesting that the estimation can be unreliable in some cases. The proposed method with voting successfully overcomes the problem, achieving much better performance.

The evaluation results are summarized as follows. The word embedding-based method is effective in predicting whether or not a given subject complement pair is semantically equivalent and in turn in detecting overgeneralization of *be*-verb despite the fact that it is a rather simple method requiring no manually-annotated learner corpus. All it does is use the information about the error rate of overgeneralization of *be*-verb in the target text. At the same time, it is crucial to set the hyperparameters properly. The voting method aptly avoids selecting just one setting.

5 Discussion

We investigated the detection results of the proposed method with the given best hyperparameter setting. It gave a cosine similarity of below zero to 20 subject

Table 2. Error Typology and its Recall.

Error Type	Recall	
Whole structure rewrite e.g., **Japan was a winter. → It was winter in Japan.*	0.47	(9/19)
Change of subject/complement to another noun e.g., **My job is an acceptance. → *My job is a receptionist.*	0.50	(9/18)
Change of *be*-verb to another verb e.g., **I can be fun. → I can have fun.*	0.50	(5/10)
Change of complement to participle e.g., **I was warry about it. → I was worried about it.*	0.88	(7/8)
Change of complement to adjective e.g., **Airplanes are danger. → Airplanes are dangerous.*	1.0	(7/7)
Addition of preposition after *be*-verb e.g., **The story is basketball. → The story was about basketball.*	0.50	(3/6)
Reversal of subject and complement e.g., **Fruits are bananas. → Bananas are fruits.*	0.00	(0/1)
Other	0.57	(8/14)
TOTAL	0.58	(48/83)

complement pairs out of 294, which were accordingly detected as overgeneralization of *be*-verb. Only 60% (12 instances) of them were actually erroneous. At first sight, this seemed that it was not a good measure for subject complement equivalence check. Looking into the pairs, however, revealed that five out of eight false positives were in proper nouns and that the accuracy on common nouns was much higher (83%=10/12) than on them (0.25%=2/8). We observed high accuracy at the other end of the cosine similarity, too; 48 pairs received a cosine similarity of more than 0.3, which were determined as correct. Most of them (41) were actually correct use, achieving an accuracy of 85%. These results show that the proposed method is effective in detecting overgeneralization of *be*-verb at least in common nouns.

Proper nouns were found to be problematic in the whole data set. One possible reason for this is that proper nouns from the writer's native language (such as Japanese names and places in the present case) were less frequent than other English common nouns in the native corpus used to learn word embeddings. Consequently, their word embeddings were likely to be less reliable. Also, some had coincidently the exact same spelling as an English proper noun. Examples were *Kobe* (a Japanese city) and *Himeji Oden* (a kind of Japanese local food). In the native corpus used as training data for word embeddings, they both often appeared as a person name (*Kobe Bryant* and *Greg Oden*, both basketball players); we actually sampled 100 instances of *Kobe* out of it and recognized 87% of them as a person (mostly, *Kobe Bryant*). This explains well why false positives occurred in these proper nouns in KJ as in *Kobe is a nice place.* They appeared seven times in total, four of which had a detection failure, suggesting that proper nouns require special care in order to achieve better performance.

For better understanding of its detection tendency, we classified the 83 errors into subcategories according to the treatment they require to become a valid

English sentence. As a result, we recognized seven types; Table 2 shows them with their corresponding recall. It shows that the proposed method is capable of detecting all types of error, except **Reversal of subject and complement**, to some extent.

Let us finally discuss the performance limitations of the proposed method. One of the major limitations is that it cannot distinguish between senses in homographs and even polysemes. In other words, a noun, a homograph/polyseme or not, is represented by one word embedding vector. This property of word embeddings leads to false positives and negatives as found in the detection failures in the proper nouns described above. It may mitigate this problem to encode the information on the entire NPs of the head nouns (subject and complement) in question. It can be achieved, for example, by taking the averages of their word embeddings or encoding them into LSTM [5]. It would be interesting to see how well such encoded vectors work on this problem.

Another performance limitation is that it does not at all target the reversal of subject and complement errors as in **Fruits are bananas*. Obviously, it gives the same value of the cosine similarity for a subject complement pair and its reversal. Accordingly, it always predicts both to be correct or incorrect, failing to detect this type of error. This is one of the performance limitations of the proposed method in theory. At the same time, this type of error is relatively infrequent even in learner English as shown in Table 2. This is probably because it is not a problem of language use but a problem of logic. Therefore, after a certain age, language learners are expected to have much less trouble with it; what poses the difficulty for them is the interference from their native language when the *be*-verb equivalent has usages other than those which English *be*-verb does. Considering this, the limitation will likely be less problematic in practice.

6 Conclusions

This paper addressed the problem of the detection of overgeneralization of *be*-verb. The presented method solved it as a problem of predicting the subject complement equivalence through word embeddings. Evaluation showed that it outperformed the three baselines exploiting corpus statistics and the WordNet ontology. Detailed investigations of the results revealed its performance limitations.

Acknowledgments

This work was supported by Japan Science and Technology Agency (JST), PRESTO Grant Number JPMJPR1758, Japan

References

1. Baroni, M., Lenci, A.: How we BLESSed distributional semantic evaluation. In: Proc. of the GEMS 2011 Workshop on GEometrical Models of Natural Language

Semantics. pp. 1–10. Association for Computational Linguistics (2011), http://aclweb.org/anthology/W11-2501
2. Chollampatt, S., Taghipour, K., Ng, H.T.: Neural network translation models for grammatical error correction. In: Proc. of 25th International Joint Conference on Artificial Intelligence. pp. 2768–2774 (2016), http://www.ijcai.org/Proceedings/16/Papers/393.pdf
3. Hamilton, W.L., Leskovec, J., Jurafsky, D.: Diachronic word embeddings reveal statistical laws of semantic change. In: Proc. of 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1489–1501 (2016), http://www.aclweb.org/anthology/P16-1141
4. Han, N.R., Chodorow, M., Leacock, C.: Detecting errors in English article usage by non-native speakers. Natural Language Engineering **12**(2), 115–129 (2006)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997). https://doi.org/10.1162/neco.1997.9.8.1735
6. Izumi, E., Uchimoto, K., Saiga, T., Supnithi, T., Isahara, H.: Automatic error detection in the Japanese learners' English spoken data. In: Proc. of 41th Annual Meeting of the Association for Computational Linguistics. pp. 145–148 (2003)
7. Junczys-Dowmunt, M., Grundkiewicz, R.: Phrase-based machine translation is state-of-the-art for automatic grammatical error correction. In: Proc. of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1546–1556 (2016), https://aclweb.org/anthology/D16-1161
8. Lin, D.: An information-theoretic definition of similarity. In: Proc. of 15th International Conference on Machine Learning. pp. 296–304 (1998), http://dl.acm.org/citation.cfm?id=645527.657297
9. Miller, G.A.: WordNet: A lexical database for English. Communications of the ACM **38**(11), 39–41 (1995)
10. Nagata, R., Sakaguchi, K.: Phrase structure annotation and parsing for learner English. In: Proc. of 54th Annual Meeting of the Association for Computational Linguistics. pp. 1837–1847 (2016). https://doi.org/10.18653/v1/P16-1173
11. Quirk, R., Greenbaum, S., Leech, G., Svartvik, J.: A Comprehensive Grammar of the English Language. Longman, New York (1985)
12. Takamura, H., Nagata, R., Kawasaki, Y.: Analyzing semantic change in Japanese loanwords. In: Proc. of 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. pp. 1195–1204 (2017), http://www.aclweb.org/anthology/E17-1112
13. Washio, K., Kato, T.: Filling missing paths: Modeling co-occurrences of word pairs and dependency paths for recognizing lexical semantic relations. In: Proc. of 2018 Conference of the North American Chapter of the Association for Computational Linguistics. pp. 1123–1133 (2018). https://doi.org/10.18653/v1/N18-1102, http://aclweb.org/anthology/N18-1102
14. Yannakoudakis, H., Rei, M., Andersen, Ø.E., Yuan, Z.: Neural sequence-labelling models for grammatical error correction. In: Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 2795–2806 (2017), https://www.aclweb.org/anthology/D17-1297