The Effect of Learner Corpus Size in Grammatical Error Correction of ESL Writings

Tomoya Mizumoto[1] *Yuta Hayashibe*[1]
Mamoru Komachi[1] *Masaaki Nagata*[2] *Yuji Matsumoto*[1]
(1) Nara Institute of Science and Technology, Nara, Japan
(2) NTT Communication Science Laboratories, Kyoto, Japan
tomoya-m@is.naist.jp, yuta-h@is.naist.jp, komachi@is.naist.jp,
nagata.masaaki@lab.ntt.co.jp, matsu@is.naist.jp

Abstract

English as a Second Language (ESL) learners' writings contain various grammatical errors. Previous research on automatic error correction for ESL learners' grammatical errors deals with restricted types of learners' errors. Some types of errors can be corrected by rules using heuristics, while others are difficult to correct without statistical models using native corpora and/or learner corpora. Since adding error annotation to learners' text is time-consuming, it was not until recently that large scale learner corpora became publicly available. However, little is known about the effect of learner corpus size in ESL grammatical error correction. Thus, in this paper, we investigate the effect of learner corpus size on various types of grammatical errors, using an error correction system based on phrase-based statistical machine translation (SMT) trained on a large scale error-tagged learner corpus. We show that the phrase-based SMT approach is effective in correcting frequent errors that can be identified by local context, and that it is difficult for phrase-based SMT to correct errors that need long range contextual information.

Keywords: ESL, grammatical error correction, statistical machine translation.

1 Introduction

English as a Second Language (ESL) learners' writings contain various kinds of grammatical errors. Recent growth in corpus annotation of learner English allows detailed analysis of grammatical errors in learners' writings. Konan-JIEM Learner Corpus (hereafter referred to as KJ Corpus)[1] is one such corpus composed of English essays written by Japanese college students. Table 1 shows the distribution of errors found in KJ Corpus[2]. The most frequent error type is *article* errors, followed by *noun number* and *preposition* errors. It is not surprising that frequent types of errors account for the most errors, but it should be noted that there are many different types of errors in learner corpus.

Thus far, a lot of studies have been made on automated error correction in regard to errors ESL learners make. However, most previous studies of second language learning deal with one or a few restricted types of learners' errors. For example, there are studies on *preposition* errors (Rozovskaya and Roth, 2011), *verb selection* errors (Liu et al., 2011), *tense* errors (Tajiri et al., 2012), *verb form* errors (*agreement* and *tense*) (Lee and Seneff, 2008), *preposition* and *article* errors (Dahlmeier and Ng, 2011) and *spelling*, *article*, *preposition* and *word form* (*agreement* and *tense*) errors (Park and Levy, 2011). Recently, Swanson and Yamangil (2012) presented a detailed analysis on correcting all types of errors in the Cambridge Learner Corpus, but their task is different from the others in that their goal is to detect errors and select error types given both the original and corrected text, which is not often available in practice.

Some types of errors like agreement errors can be corrected by simple rules using heuristics, while others like preposition errors are difficult to correct without statistical model trained on native corpora and/or learner corpora. It was not until recently that large scale learner corpora became widely available for grammatical error correction. However, little is known about the effect of learner corpus size in ESL grammatical error correction.

In this paper, we conduct experiments in error correction targeting all types of errors using a large scale error-annotated learner corpus to see the effect of corpus size in grammatical error correction. We build an error correction system with phrase-based statistical machine translation (SMT) technique. Also, we create a large scale error-tagged corpus of learner English from the web. We then analyze the results of error correction by breaking down the error types and discuss the strength and weakness of the example based approach using a large scale but noisy learner corpus.

The main contribution of this work is two-fold:

- To our knowledge, it is the first attempt to use a large scale learner corpus to correct all types of errors.
- We show the effect of learner corpus size on the phrase-based SMT approach and show its advantages and disadvantages.

In the following, we briefly overview related work of grammatical error correction in Section 2. Then we describe our grammatical error correction system and large scale error-annotated learner corpus in Section 3. Section 4 shows our experimental results and discusses the effect of corpus size on different error types.

2 Related work

Even though there are many works on error correction in learners' English, only a few target multiple various kinds of grammatical errors.

[1]`http://www.gsk.or.jp/catalog/GSK2012-A/catalog_e.html`

[2]Spelling errors are excluded from target of annotation in KJ Corpus.

Types	Proportion (%)	Types	Proportion (%)
article	19.23	verb other	4.09
noun number	13.88	adverb	3.59
preposition	13.56	conjunction	2.04
tense	8.77	word order	1.34
lexical choice of noun	7.04	noun other	1.30
lexical choice of verb	6.90	auxiliary verb	0.88
pronoun	6.62	other lexical choice	0.74
agreement	5.25	relative	0.42
adjective	4.30	interrogative	0.04

Table 1: The distribution of errors on KJ Corpus.

First, Brockett et al. (2006) proposed an error correction model with phrase-based SMT. Even though their model can deal with all types of errors, they evaluated their method only on noun number errors using an artificial data, partly because there was no large scale learner corpus available at the time. We would like to emphasize that our work is the first attempt to use a real world large learner corpus with phrase-based SMT technique. We will show that phrase-based SMT especially suffers from data sparseness.

Second, Park and Levy (2011) attempted to correct various kinds of errors with a noisy channel model using a large scale unannotated corpus of learner English. Ours differs from their work in that we use a large scale error-tagged corpus annotated by the wisdom of crowds. In addition, they targeted only spelling, article, preposition and word form errors, while we do not restrict error types.

Third, Han et al. (2010) developed a preposition correction system using a large scale error-tagged corpus of learner English. They built a maximum entropy-based model for preposition errors trained on learner and native corpora. We also take advantage of a large scale error-tagged corpus of learner English, but use phrase-based SMT to deal with various kinds of errors and to fully exploit the learner corpus.

Recently, Dahlmeier and Ng (2012) presented a beam-search decoder for correcting spelling, article, preposition, punctuation and noun number errors. They reported that their discriminative model achieves considerably better results than an SMT baseline trained on a few hundreds of sentences. As we will see later, we observed a similar tendency in preposition error correction when we trained a phrase-based SMT system on a small learner corpus. However, in this work, we exploit a large scale error-annotated corpus extracted from the web to overcome the data sparseness problem.

3 Using a large scale learner corpus with phrase-based SMT for grammatical error correction

3.1 Error correction with phrase-based SMT

We use phrase-based statistical machine translation (Koehn et al., 2003) to conduct unrestricted error correction. There are several studies about grammatical error correction using phrase-based statistical machine translation (Brockett et al., 2006; Mizumoto et al., 2011; Ehsan and Faili, 2012). Although Brockett et al. (2006) corrected English learners' error using phrase-based statistical machine translation, they only targeted mass noun errors. Mizumoto et al. (2011) dealt with unrestricted types of learners' errors, but their target is not English but Japanese. Ehsan and Faili (2012) applied an SMT framework to English and Persian grammatical error correction, but used artificially created learner corpora.

The well-known statistical machine translation formulation using a log-linear model (Och and Ney, 2002) is defined by:

$$\hat{e} = \arg\max_{e} P(e|f) = \arg\max_{e} \sum_{m=1}^{M} \lambda_m h_m(e,f) \tag{1}$$

where e represents target sentences (corrected sentences) and f represents source sentences (sentences written by learners). $h_m(e,f)$ is a feature function and λ_m is a model parameter for each feature function. This formulation finds a target sentence e that maximizes a weighted linear combination of feature functions for source sentence f. A translation model and a language model can be used as feature functions. The translation model is commonly represented as conditional probability $P(f|e)$ factored into the translation probability between phrases. The language model is represented as probability $P(e)$. The translation model is learned from sentence-aligned parallel corpus while the language model is learned from target raw corpus.

3.2 Crowdsourcing annotation of a large scale corpus of learner English

We use data from a language learning social networking service Lang-8 [3] to train the error correction system using statistical machine translation. In Lang-8, language learners post their writing on the Lang-8 site to be corrected by native speakers. We can obtain pairs of learner's sentence and corrected sentence in large scale from Lang-8. Mizumoto et al. (2011) first presented an approach to extract a learner corpus from the web, but we differ from them in that we create a learner corpus of English rather than Japanese. Also, unlike (Tajiri et al., 2012), we propose to use metadata of users to determine the L1 of English learners. Because our test corpus (KJ Corpus) is written by Japanese college students, we would like to use the same kind of data; it is out side of the scope of this paper to see the effect of learners' L1.

We crawled blog entries found in Lang-8 as of December 2010. We used writings in Lang-8 written by Japanese ESL learners for translation model and language model of error correction system with SMT. There are 509,116 sentence pairs in English writings written by Japanese L1 English learners. However, we need to filter noisy sentences because it may be hard to align them if the sentences are drastically changed from the original learner's sentences, resulting in degraded performance on phrase-based SMT approach. Therefore, we calculate the edit distance between a learner sentence and the corrected sentence using a dynamic programming algorithm, and retain sentences whose numbers of both insertions and deletions is equal to or less than 5 words [4]. As a result, we obtain 391,699 sentence pairs.

4 Experiment: Effect of learner corpus size in grammatical error correction

We carried out an experiment on grammatical error correction with SMT-based system using a large scale learner corpus. To see the effect of corpus size, we compare a system using Lang-8 Corpus (large scale learner corpus) with different sizes and a system using KJ Corpus (small scale corpus). In order to get a closer look at the effect of error correction methods, we also experimented on the preposition error correction task using a maximum entropy model as a discriminative baseline and SMT-based models as our proposal for all error correction.

[3] `http://lang-8.com/`

[4] We use 6 as a distortion-limit for Moses, therefore we chose the edit distance to be smaller than the distortion-limit.

4.1 Tools and experimental data

We used Moses 2010-08-13 [5] with default parameters as a decoder and GIZA++ 1.0.5 [6] as an alignment tool to implement an error correction system with phrase-based SMT. We applied grow-diag-final-and (Och and Ney, 2003) heuristics for phrase extraction. The number of extracted phrases are 1,050,070 (245 MB) using all data of Lang-8 Corpus. We used 3-gram as a language model trained on the corrected text of Lang-8 Corpus.

Next, we built the maximum entropy model (Berger et al., 1996) as a multi-class classifier baseline for preposition error correction (Sakaguchi et al., 2012). We used the implementation of Maximum Entropy Modeling Toolkit [7] with its default parameters. We incorporated surface, POS, WordNet, parse and language model features described in (Tetreault et al., 2010) and (De Felice and Pulman, 2008). POS and parse features were extracted using the Stanford Parser 2.0.2. This system achieves recall of 18.44, precision of 34.88 and F-measure of 24.12 trained and tested on the CLC FCE dataset (Yannakoudakis et al., 2011), which ranked the 4th out of 13 systems at the HOO 2012 Shared Task (Dale et al., 2012).

We use KJ Corpus as a test data. KJ Corpus consist of 170 essays, containing 2,411 sentences. When we experiment on a system using KJ Corpus, we perform 5-fold cross validation.

4.2 Evaluation metrics

For the evaluation metrics, we use automatic evaluation criteria. To be precise, we use recall, precision and F-measure.

Recall and precision for each type of errors are calculated from true positive, false positive and false negative based on error tags in KJ Corpus. The word which does not have any tag in KJ Corpus does not affect precision for each type of errors [8]. For example, let us consider the following:

learner: He talked <u>to</u> me _ his life <u>of</u> Kyoto, and he took me _ Kyoto university.
correct: He talked <u>to</u> me <u>about</u> his life <u>in</u> Kyoto and he took me <u>to</u> Kyoto university.
system: He talked _ me _ his life <u>on</u> Kyoto, and he took me <u>to</u> Kyoto university.

In this example, the system deletes preposition "to", which does not have any tag. Thus, precision = 1/2, recall = 1/2 for *preposition* errors and precision = 1/3, recall = 1/2 for Total scores.

4.3 Experimental results

Table 2 shows error correction results for each type of errors on different corpora. We compared SMT systems trained on KJ Corpus, Lang-8 Corpus with the same amount of data with KJ Corpus, and full Lang-8 Corpus. With very few exceptions, the larger the size of learner corpus, the higher the accuracy. In addition, using the larger corpus, precision tends to increase more than recall.

Table 3 presents F-measures for each type of error varying the corpus sizes (2K, 10K, 20K, 100K, 200K, 300K, All (390K)). As we will see later in the next section, there are two types of errors in which learner corpus size matters.

Table 4 shows the performance of preposition error correction. Perhaps not surprising, but it still deserves attention that SMT model trained on all Lang-8 Corpus clearly outperformed other two

[5] `http://http://www.statmt.org/moses/`
[6] `http://code.google.com/p/giza-pp/`
[7] `https://github.com/lzhang10/maxent`
[8] The total score is calculated using all the correction output with and without any tag.

Training Corpus	KJ Corpus			Lang-8 Corpus (2K)			Lang-8 Corpus (390K)		
	Recall	Prec	F	Recall	Prec	F	Recall	Prec	F
article	0.187	0.531	0.277	0.187	0.571	0.282	**0.359**	**0.761**	**0.488**
noun number	0.207	0.603	0.308	0.136	0.671	0.226	0.199	**0.710**	0.311
preposition	0.137	0.375	0.201	0.092	0.319	0.143	**0.262**	**0.585**	**0.361**
tense	0.102	0.170	0.128	0.043	0.088	0.058	0.080	0.149	0.104
lexical choice of noun	0.035	0.114	0.054	0.033	0.152	0.054	**0.182**	**0.443**	**0.258**
lexical choice of verb	0.070	0.161	0.098	0.065	0.200	0.098	**0.192**	**0.324**	**0.241**
pronoun	0.075	0.220	0.112	0.040	0.143	0.063	0.150	**0.367**	**0.213**
agreement	0.236	**0.604**	0.340	0.125	0.483	0.199	0.228	0.469	0.307
adjective	0.151	0.326	0.206	0.056	0.286	0.094	**0.389**	**0.522**	**0.446**
verb other	0.089	0.139	0.109	0.147	0.333	0.204	**0.286**	**0.419**	**0.340**
adverb	0.265	0.450	0.333	0.214	0.429	0.286	0.292	0.432	0.349
conjunction	0.100	0.417	0.161	0.091	0.714	0.161	0.115	**0.546**	0.190
word order	0.500	0.025	0.048	0.667	0.050	0.093	**0.750**	0.075	0.136
noun other	0.182	0.222	0.200	0.143	0.167	0.154	**0.571**	**0.429**	**0.490**
auxiliary verb	0.056	0.167	0.083	0.100	0.400	0.160	0.100	**0.400**	0.160
other lexical choice	0.167	0.200	0.182	0.000	0.000	0.000	**0.357**	**0.455**	**0.400**
relative	0.111	0.250	0.154	0.182	0.667	0.286	0.091	**0.500**	0.154
interrogative	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total	0.149	0.147	0.148	0.113	0.205	0.146	0.247	**0.275**	**0.260**

Table 2: Result for each type of errors by statistical machine translation. Bold face indicates that one system's result is equal or greater by more than 0.1 points than the other systems' result.

systems. MaxEnt does slightly better than SMT when they are trained on the same small corpus. Unfortunately, we were not able to use Lang-8 Corpus since it took too long to train.

4.4 Discussion

We can classify errors into two types: (1) errors which get better correction by increasing corpus size and (2) errors which have little relationship with corpus size. The first type of errors includes *article*, *preposition*, *lexical choice of noun*, *lexical choice of verb*, *adjective*, and *noun other*. On the other hand, the second type of errors comprises *noun number*, *tense*, *agreement*, *adverb*, *conjunction*, *word order*, *auxiliary verb*, *relative* and *interrogative*. We can expect to improve performance (both recall and precision) for errors that require wide coverage lexical knowledge, such as lexical choice errors, by using a much larger corpus with phrase-based SMT. In contrast, we may say that errors which involve larger context such as tense errors are difficult to correct with phrase-based SMT. We discuss the result while looking at examples of two of the former type of errors (*article* and *lexical choice of noun*) whose F-measures improve with increasing corpus size, and three of the latter type of errors (*noun number*, *tense* and *agreement*), whose F-measures do not change or even degrade.

Table 5 shows examples of article and lexical choice of noun. These are the examples that phrase-based SMT failed to correct using KJ Corpus. Because we can acquire a lot of pairs of an error phrase and its correction by increasing the size of the learner corpus, the phrase-based SMT was able to correct them using Lang-8 Corpus.

Table 6 shows examples of noun number, tense and agreement errors. The first example of noun

Training Corpus	KJ	Lang-8						
		2K	10K	20K	100K	200K	300K	390K
article	0.277	0.282	*0.390	*0.420	*0.443	*0.459	*0.475	*0.488
noun number	0.308	0.226	0.214	0.238	0.270	0.300	0.319	0.311
preposition	0.201	0.143	0.192	0.226	*0.333	*0.336	*0.344	*0.362
tense	0.128	0.058	0.066	0.058	0.081	0.096	0.089	0.104
lexical choice of noun	0.054	0.054	0.124	0.133	*0.189	*0.216	*0.250	*0.258
lexical choice of verb	0.098	0.098	0.087	0.138	*0.196	*0.232	*0.232	*0.241
pronoun	0.112	0.063	0.131	0.150	0.177	0.195	0.213	0.213
agreement	0.340	0.197	0.224	0.248	0.260	0.284	0.307	0.307
adjective	0.206	0.094	0.165	0.219	*0.413	*0.426	*0.426	*0.446
verb other	0.109	0.204	0.240	0.311	0.291	*0.340	0.308	0.340
adverb	0.333	0.286	0.286	0.302	0.333	0.349	0.349	0.349
conjunction	0.161	0.161	0.161	0.191	0.161	0.191	0.191	0.191
word order	0.048	0.093	0.093	0.091	0.091	0.091	0.091	0.136
noun other	0.200	0.154	0.286	0.286	*0.531	*0.490	*0.490	*0.490
auxiliary verb	0.083	0.160	0.160	0.083	0.083	0.160	0.160	0.160
other lexical choice	0.182	0.000	0.095	0.095	0.400	0.400	0.400	0.400
relative	0.154	0.285	0.154	0.154	0.154	0.154	0.154	0.154
interrogative	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total	0.148	0.146	0.180	0.200	0.239	0.247	0.254	0.260

Table 3: Results (F-measure) for error correction by SMT varying the learner corpus sizes. Asterisks indicate that the difference of result using Lang-8 Corpus and result using KJ Corpus is statistically significant ($p < 0.01$).

System	Training corpus	Recall	Precision	F-measure
Maximum entropy-based model	KJ Corpus	0.165	0.407	0.235
Phrase-based SMT	KJ Corpus	0.137	0.375	0.201
Phrase-based SMT	Lang-8 Corpus (390K)	**0.262**	**0.585**	**0.362**

Table 4: Result for preposition error correction on KJ Corpus.

number was corrected using Lang-8 Corpus with phrase-based SMT since the error is one of the common learners' expressions. The second was not corrected using Lang-8 Corpus with phrase-based SMT because "dools" [9] is slightly displaced from "a big", and a proper noun "snoopy" is inserted between "dools" and "a big". It is hard to correct this kind of error with Phrase-based SMT, even using artificial data such as in Brockett et al. (2006). To solve this problem, we need to conduct generalization using POS or consider dependency relations.

The first example of a tense error was corrected using both KJ Corpus and Lang-8 Corpus with phrase-based SMT. One of the reasons why the baseline system was able to correct the error is that it requires only local context to correct and is very frequent even in a small leaner corpus. In the second example, the system fails to find tense agreement in the complex sentence. Tense error is difficult to correct for phrase-based SMT since it involves global context (Tajiri et al., 2012).

The first example of agreement error was corrected using Lang-8 Corpus with phrase-based SMT. This is because the phrase pair correcting "Flowers is" to "Flowers are" is frequent and the language

[9] The word "dools" written by a learner is also a spelling error.

	learner	correct
article	I like a chocolate very much.	I like _ chocolate very much.
lexical choice of noun	my cycle was injured, but i wasn't.	my bicycle was damaged, but i wasn't.

Table 5: Examples of system output for article and lexical choice of noun error

	learner	correct
noun number 1	I read various type books.	I read various types of books.
*noun number 2	There is a big snoopy dools in my room.	There is a big snoopy doll in my room.
tense 1	If I 'll live in saitama, I must have ...	If I _ live in saitama, I must have ...
*tense 2	The weather is very sunny, so we were ...	The weather was very sunny, so we were ...
agreement 1	Flowers is very beautiful.	Flowers are very beautiful.
*agreement 2	I think, reading comics are not "reading"	I think, reading comics is not "reading"

Table 6: Examples of system results for noun number, tense and agreement errors. Asterisks indicate that the SMT system using full Lang-8 Corpus failed to correct the errors.

model probability of "Flowers are" is also higher than "Flowers is". The second example is one that the system failed to correct since the pattern is unseen in the learner corpus and thus the system has no way to capture the relation between the subject "reading" and "are". To solve this problem, it needs to get the subject-verb relation considering a dependency structure.

As for prepostion error correction, we suspect that there are two reasons why the SMT-based model using full Lang-8 Corpus outperformed the MaxEnt model. First, due to the small amount of training data in KJ Corpus (2,000 sentences), the MaxEnt model failed to build a high performance system. Second, the high performance of the SMT system may be attributed to the fact that both KJ Corpus and Lang-8 Corpus were written by Japanese native speakers. Also, the reason why the MaxEnt model achieved better result than SMT when trained on the same small corpus is possibly because KJ Corpus is too small to learn variations in learner English by phrase-based SMT approach, while a discriminative model can exploit a small dataset using rich features.

Conclusion

We tackled the task of ESL grammatical error correction of all types of errors using a large scale corpus of learner English with phrase-based SMT technique. Previous research focused on restricted types of errors due to the small amount of learner corpora. We overcome this problem by training an error correction system on a large scale error tagged corpus extracted from the web.

We found that the size of corpus is critical to improve phrase-based SMT approach. However, the degree of improvement varies across error types. Phrase-based SMT is effective in correcting frequent errors which require only local context. For example, there is a clear improvement in increasing the size of learner corpus for correcting *article*, *preposition*, *lexical choice* and *adjective* errors, while there is little improvement for correcting *agreement* and *tense* errors.

Acknowledgement

We would like to thank Yangyang Xi for granting permission to use text from Lang-8.

References

Berger, A. L., Pietra, V. J. D., and Pietra, S. A. D. (1996). A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, 22(1):39–71.

Brockett, C., Dolan, W. B., and Gamon, M. (2006). Correcting ESL Errors Using Phrasal SMT Techniques. In *Proceedings of COLING-ACL*, pages 249–256.

Dahlmeier, D. and Ng, H. T. (2011). Grammatical Error Correction with Alternating Structure Optimization. In *Proceedings of ACL-HLT*, pages 915–923.

Dahlmeier, D. and Ng, H. T. (2012). A Beam-Search Decoder for Grammatical Error Correction. In *Proceedings of EMNLP*, pages 568–578.

Dale, R., Anisimoff, I., and Narroway, G. (2012). HOO 2012: A Report on the Preposition and Determiner Error Correction Shared Task. In *Proceedings of BEA*, pages 54–62.

De Felice, R. and Pulman, S. G. (2008). A Classifier-Based Approach to Preposition and Determiner Error Correction in L2 English. In *Proceedings of COLING*, pages 169–176.

Ehsan, N. and Faili, H. (2012). Grammatical and Context-Sensitive Error Correction Using a Statistical Machine Translation Framework. *Software: Practice and Experience*.

Han, N.-R., Tetreault, J., Lee, S.-H., and Ha, J.-Y. (2010). Using an Error-Annotated Learner Corpus to Develop an ESL/EFL Error Correction System. In *Proceedings of LREC*, pages 763–770.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical Phrase-Based Translation. In *Proceedings of HLT-NAACL*, pages 48–54.

Lee, J. and Seneff, S. (2008). Correcting Misuse of Verb Forms. In *Proceedings of ACL-HLT*, pages 174–182.

Liu, X., Han, B., and Zhou, M. (2011). Correcting Verb Selection Errors for ESL with the Perceptron. In *Proceedings of CICLing*, pages 411–423.

Mizumoto, T., Komachi, M., Nagata, M., and Matsumoto, Y. (2011). Mining Revision Log of Language Learning SNS for Automated Japanese Error Correction of Second Language Learners. In *Proceedings of IJCNLP*, pages 147–155.

Och, F. J. and Ney, H. (2002). Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In *Proceedings of ACL*, pages 295–302.

Och, F. J. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.

Park, Y. A. and Levy, R. (2011). Automated Whole Sentence Grammar Correction Using a Noisy Channel Model. In *Proceedings of ACL*, pages 934–944.

Rozovskaya, A. and Roth, D. (2011). Algorithm Selection and Model Adaptation for ESL Correction Tasks. In *Proceedings of ACL*, pages 924–933.

Sakaguchi, K., Hayashibe, Y., Kondo, S., Kanashiro, L., Mizumoto, T., Komachi, M., and Matsumoto, Y. (2012). Naist at the hoo 2012 shared task. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 281–288.

Swanson, B. and Yamangil, E. (2012). Correction Detection and Error Type Selection as an ESL Educational Aid. In *Proceedings of NAACL: HLT*, pages 357–361.

Tajiri, T., Komachi, M., and Matsumoto, Y. (2012). Tense and Aspect Error Correction for ESL Learners Using Global Context. In *Proceedings of ACL*, pages 198–202.

Tetreault, J., Foster, J., and Chodorow, M. (2010). Using Parse Features for Preposition Selection and Error Detection. In *Proceedings of ACL*, pages 353–358.

Yannakoudakis, H., Briscoe, T., and Medlock, B. (2011). A New Dataset and Method for Automatically Grading ESOL Texts. In *Proceedings of ACL*, pages 180–189.