Building a TOCFL Learner Corpus for Chinese Grammatical Error Diagnosis

Lung-Hao Lee[1,2], Yuen-Hsien Tseng[2], Li-Ping Chang[3]
[1]MOST Joint Research Center for AI Technology and All Vista Healthcare, National Taiwan University
[2]Graduate Institute of Library and Information Studies, National Taiwan Normal University
[3]Mandarin Training Center, National Taiwan Normal University
[1]No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan
[2,3]No. 162, Sec. 1, Heping East Rd., Taipei 10610, Taiwan
{lhlee, samtseng, lchang}@ntnu.edu.tw

Abstract

This study describes the construction of a TOCFL learner corpus and its usage for Chinese grammatical error diagnosis. We collected essays from the Test Of Chinese as a Foreign Language (TOCFL) and annotated grammatical errors using hierarchical tagging sets. Two kinds of error classifications were used simultaneously to tag grammatical errors. The first capital letter of each error tags denotes the coarse-grained surface differences, while the subsequent lowercase letters denote the fine-grained linguistic categories. A total of 33,835 grammatical errors in 2,837 essays and their corresponding corrections were manually annotated. We then used the Standard Generalized Markup Language to format learner texts and annotations along with learners' accompanying metadata. Parts of the TOCFL learner corpus have been provided for shared tasks on Chinese grammatical error diagnosis. We also investigated systems participating in the shared tasks to better understand current achievements and challenges. The datasets are publicly available to facilitate further research. To our best knowledge, this is the first annotated learner corpus of traditional Chinese, and the entire learner corpus will be publicly released.

Keywords: computer-assisted language learning, second language acquisition, grammatical error diagnosis, interlanguage analysis

1. Introduction

Annotating learners' inappropriate usage of written language is an important task for learner corpus research (Díaz-Negrillo and Fernández-Domínguez, 2006; Tono, 2003). From a linguistic perspective, annotated learner corpora are valuable resources for research in second language acquisition (Swanson and Charniak, 2013), foreign language teaching (Wang and Seneff, 2007), and contrastive interlanguage analysis (Granger, 2015). In engineering, such language resources can be used to develop natural language processing techniques for educational applications, such as automatic essay scoring (Yannakoudakis et al., 2011), assessment report generation (Sakaguchi et al., 2013), and native language identification (Malmasi and Dras, 2015).

Automated grammatical error detection and correction are important research directions and a number of competitions have been organized to encourage innovation (Leacock et al., 2014). For example, Helping Our Own (HOO) is a series of shared tasks for correcting errors in non-native English texts (Dale and Kilgarriff, 2011, Dale et al., 2012). The CoNLL 2013/2014 shared tasks aimed to correct grammatical errors for learners of English as a foreign language (Ng et al., 2013; 2014). The NLPTEA workshops had hosted a series of shared tasks for Chinese grammatical error diagnosis (Yu et al., 2014; Lee et al., 2015b; Lee et al., 2016b). Recently, the IJCNLP 2017 shared task 1 has focused on Chinese grammatical error diagnosis (Rao et al., 2017). All these competitions require annotated learner corpora for system development and evaluation.

To our best knowledge, only two previous studies have manually annotated learner corpora for Chinese as a foreign language (CFL). One is the HSK Dynamic Composition Corpus constructed by the Beijing Language and Culture University (Cui and Zhang, 2011; Zhang and Cui, 2013); the other is the Jinan Chinese Learner Corpus (Wang et al., 2015). The target language of these two studies is simplified Chinese, and no traditional Chinese learner corpus is available for public research. This pilot study thus aims to build such a learner corpus of traditional Chinese to expand research resources, especially for the study of linguistic differences or similarities among Chinese learners around the world.

This study annotated grammatical errors in texts collected from learner essays written as part of the Test Of Chinese as Foreign Language (TOCFL). The TOCFL learner corpus contained 2,837 essays written by learners originating from a total of 46 different mother-tongue languages. Chinese native speakers were trained to annotate these essays using hierarchical error tags, and 33,835 inappropriate grammatical usages were identified. We then used the Standard Generalized Markup Language to format annotated errors and their correct usages. Some of the annotated sentences were used for shared tasks hosted by the NLPTEA workshops. We also investigated systems participating in the shared tasks using various approaches for automated Chinese error diagnosis.

The rest of this paper is organized as follows. Section 2 briefly reviews existing learner corpora from around the world. Section 3 describes the process of TOCFL learner corpus annotation. Section 4 presents the annotation results. Section 5 investigates the shared tasks for Chinese grammatical error diagnosis based on the TOCFL learner corpus. Conclusions are finally drawn in Section 6.

2. Related Work

The Longman Learner Corpus is the first to collect essays and exam scripts written by learners of English (Gillard and Gadsby, 1998). The International Corpus of Learner English (ICLE) consists of argumentative essays written by advanced English learners from different native language backgrounds (Granger, 2003). The Cambridge Learner Corpus (CLC) is established to assist English Language Teaching/Training (ELT) publishers to create various learning aids including dictionaries and ELT course books (Nicholls, 2003). The NUS Corpus of Learner English (NUCLE) is annotated for the development and evaluation

of grammatical error correction systems (Dahlmeier et al., 2013).

Additional learner corpora exist for European languages. The Lund CEFLE is a leaner corpus of texts in French produced by adolescent Swedish learners of French (Granfeldt et al., 2006). The Error-Annotated German Learner Corpus (EAGLE) is a corpus of beginning learners with grammatical error annotation (Boyd, 2010). The ASK corpus is a learner corpus of Norwegian as a second language that contains essays collected from language tests (Tenfjord et al., 2006). The CzeSL corpus is a learner corpus of Czech that has been annotated using multi-layer error types (Hana et al., 2010). The Hungarian Learner Corpus is composed of student journals annotated for learner errors using tagging sets from different linguistic categories, including phonology, morphology and syntax (Dickinson and Ledbetter, 2012).

Recently, learner corpora have been established for Asian languages. The HSK Dynamic Composition Corpus contains simplified Chinese essays written by learners of Chinese, annotated with different error types (Cui and Zhang, 2011; Zhang and Cui, 2013). The Jinan Chinese Learner Corpus is a collection of texts produced for educational applications (Wang et al., 2015). Lang-8 contains a Japanese learner corpus extracted from a language learning and exchange social network service (Mizumoto et al., 2011). Linguistic properties of Korean particle errors have been outlined and annotated in collected learner writings (Lee et al., 2012).

The present study follows the research trend of worldwide learner corpora construction to build the TOCFL learner corpus. To the best of our knowledge, this is the first traditional Chinese learner corpus to be publicly available for research.

3. Annotation

The Steering Committee for the Test Of Proficiency-Huaya (SC-TOP) aims to develop and promote the Test Of Chinese as a Foreign Language (TOCFL) to assess the proficiency of CFL learners. The TOCFL writing test references the proficiency levels of the Common European Framework of Reference (CEFR) (Little, 2006). The testing principle is task orientation, which evaluates learners' ability to express their thoughts in the context of real-world situations. There are 4 available levels of the TOCFL writing test, which are described as follows.

(1) *Waystage level* (A2): test takers have to write a note and describe a story after looking at four pictures.
(2) *Threshold level* (B1): test takers are asked to write relatively detailed personal letters that describe their experiences and feelings about events they have encountered.
(3) *Vantage level* (B2): test takers are asked to write a functional letter highlighting definite purposes or to develop an argument to express personal opinions of specific events.
(4) *Proficiency level* (C1): test takers are asked to write an essay or report that gives reasons in support or against a particular point of view or explains provided figures and tables.

Test takers choose an exam level based on their current perceived level of Chinese proficiency. The proficiency-level evaluation is based on the appropriateness of the test takers' responses to situational tasks, compositional structure and completeness, syntax correctness, and the use of a suitably wide range of appropriate vocabulary. Each evaluation is conducted by at least two Chinese teachers and is then scored on a 0-5 point scale, where a score of 3 is a passing grade.

We collected learners' essays that had been given a score of 3 or above. Lower-scoring essays were ignored because of the difficulty in interpreting the learners' intended meaning and to annotate possible errors. In addition to learners' written texts for the TOCFL test, we collected all accompanying metadata including the corresponding CEFR level, evaluated score, and learner's native language. The TOCFL writing test is computer-based, and Chinese-typing ability is a requirement for all test takers. Spelling errors of Chinese characters frequently arise from confusion among multiple-character words that are phonologically and visually similar but semantically distinct. Spelling errors were corrected in our collected learners' written texts.

We then annotated grammatical errors to analyze inappropriate linguistic usages. Hierarchical tag sets were used in annotating grammatical errors (Lee et al., 2016a). Table 1 shows two kinds of error classifications used to simultaneously to tag grammatical errors. The first capital letter denotes the coarse-grained surface differences, while the subsequent lowercase letters denote the find-grained linguistic category. The coarse-grained error types originate from comparing erroneous sentences with the correct usages. There are 4 error types: missing, redundant, incorrect selection, and word ordering errors of linguistic components (also called PADS error types, denoting errors of Permutation, Addition, Deletion, and Selection). The fine-grained error types focus on representing linguistic concepts. A total of 36 error types were distributed into word-level errors (16 cases), grammatical function-level errors (11 cases), sentence pattern-level errors (7 cases), and mixture errors (2 cases).

Native Chinese-speaking annotators were trained to follow our annotation guidelines for the error-tagging task. There were cases in which our annotators knew something was wrong, but were unable to select the appropriate annotation. In such cases, the annotators discussed the errors amongst themselves to seek agreement on the appropriate tag. A tagging editor was used to help annotators to insert error tags and rewrite the contextually correct usage for the learner corpus (Lee et al., 2014). This editor provides an error analysis function to further assist annotators in easily finding incorrect or inconsistent tagging instances during the annotation process. Some example annotations are given in Table 2. In sentence 1(a), there is a missing object "他" (he), so it was annotated using a tag [Mobj]. The correction with English translation is shown in 1(b). In sentence 2(a), the tag [Rdet] represents that "第" (the) is a redundant word that should be deleted. The word "兩" (two) in sentence 3(a) is an incorrect numeral (denoting as the tag [Snum]), where the contextually correct one should be "二" (two/second). Sentence 4(a) has a word ordering error. The phrase "七點鐘" (seven o'clock) had been tagged as [Wtime], that means this phrase should precede the word "起床" (get up).

After annotating grammatical errors and providing their correct usages, we used the Standard Generalized Markup Language (SGML; ISO 8879:1986) to format the learner texts into the following four parts.

(1) *Essay*: unique identification number, writing style, article title, the obtained score, and writing date.
(2) *Learner*: learner's mother-tongue language and his/her corresponding language proficiency of CEFR.
(3) *Text*: learners' original written texts.
(4) *Mistake*: paragraph in which a grammatical error occurs along with the positions of the starting and ending character, in which each character or punctuation mark has a counting position value of 1.

Target Modification Taxonomy	
Missing (*M*), Redundancy (*R*), Incorrect Selection (*S*), Word Ordering Error (*W*)	
Linguistic Category Classification	
Word-level	action verb (*v*), auxiliary (*aux*), stative verb (*vs*), noun (*n*), pronoun (*pron*), conjunction (*conj*), preposition (*p*), numeral (*num*), demonstrative (*det*), measure word (*cl*), sentential particle (*sp*), aspectual particle (*asp*), adverb (*adv*), structural particle (*de*), question word (*que*), plural suffix (*plural*)
Grammatical Function-level	subject (*sub*), object (*obj*), noun phrase (*np*), verb phrase (vp), preposition phrase (*pp*), modifier (*mod*), time expression (*time*), place expression (*loc*), transitivity (*tran*), separable structure (*vo*), [numeral /determiner+measure] phrase (*dm*),
Sentence Pattern-level	complex noun clause (*rel*), 把 sentence (*ba*), 被 sentence (*bei*), 讓 sentence (*rang*), 是 sentence (*shi*), 有 sentence (*you*), other patterns (*pattern*)
Mixture	formation (*form*), ambiguity of syntactic or meaning (*sentence*)

Table 1: Error Tags for grammatical error annotations.

1(a) * 他請我教 [*Mobj*] 日文
1(b) 他請我教他日文
(He asked me to teach him Japanese.)

2(a) * 我會在第 [*Rdet*] 一樓等你
2(b) 我會在一樓等你
(I will wait for you on the first floor.)

3(a) * 傑克是一個兩 [*Snum*] 年級的高中生
3(b) 傑克是一個二年級的高中生
(Jack is a second-year senior high school student.)

4(a) * 他平常起床七點鐘 [*Wtime*]
4(b) 他平常七點鐘起床
(He usually gets up at seven o'clock.)

Table 2: Example annotations and their corrections.

Figure 1 shows an example of our built TOCFL learner corpus. This essay was given an identification number "0612", and was written in "2009年5月" ('May 2009') by a "B1"-level learner with "韓語" ('Korean') as his/her L1. The topic of this "記敘文" ('narrative-style') essay is "最難忘的購物經驗" ('My memorable shopping experience'), and was given a score of "4". A number of errors were annotated. For example, the first error occurred in the first paragraph starting and ending at position 105, meaning that a word "是" ('is') is missing between "有時候" ('sometimes') and "比" ('than'). It was annotated using an error tag "Mshi", in which the first capital letter denotes the coarse-grained surface difference, while the subsequent lowercase letters denote the fine-grained linguistic category. In the fourth paragraph, "那時候" ('that time') was put in a wrong position resulting in a "Wtime" error tag denoting a time phrase in the wrong order and the word "了" ('le') was annotated as "Rasp" denoting a redundant aspectual particle that should be removed. The annotation can yield the correct sentence "記得那時候讀書讀得不太好" ('Recalling that I didn't study well at that time').

```
<ESSAY id="0612" style="記敘文" title="最難忘的購物經驗" score="4" date="2009年5月">
<LEARNER>
<L1>韓語</L1>
<CEFL>B1</CEFL>
</LEARNER>
<TEXT>
<P>
…… 在星巴克讀書，雖然需要付錢，可是我覺得有時候比圖書館好多的地方。
</P>
<P>
我喝咖啡的時候，常喝美式咖啡。……
</P>
<P>
有一天，我去星巴克，……
</P>
<P>
…… 那時候記得讀書讀得不太好了。以後也不常去那裏的星巴克了。
</P>
</TEXT>
<MISTAKE paragraph="1" start_off="105" end_off="105">
<TYPE>Mshi</TYPE>
<CORRECTION>是</CORRECTION>
</MISTAKE>
……
<MISTAKE paragraph="4" start_off="33" end_off="37">
<TYPE>Wtime</TYPE>
<CORRECTION>記得那時候</CORRECTION>
</MISTAKE>
<MISTAKE paragraph="4" start_off="45" end_off="45">
<TYPE>Rasp</TYPE>
<CORRECTION>null</CORRECTION>
</MISTAKE>
</ESSAY>
```

Figure 1: An essay in our TOCFL learner corpus.

4. Results

Table 3 provides descriptive statistics for our TOCFL learner corpus. The B1 level occupies near a half of the corpus. In total, about one million characters were collected and annotated covering 62 different essay titles.

CEFR	#Title	#Essay	#Char	Ratio%
A2	21	850	131,684	29.96%
B1	24	1,388	540,286	48.93%
B2	14	503	280,239	17.73%
C1	3	96	50,079	3.38%
Total	62	2,837	1,002,288	100%

Table 3: Statistics of TOCFL learner corpus.

Figure 2 shows the distribution of the learners' mother-tongue languages, with the top 10 languages accounting for 88%, and another 36 languages accounting for the remaining 12%. Slightly more than one fourth of the sample spoke Japanese as their first language, followed by English, Vietnamese, Korean, and Indonesian.

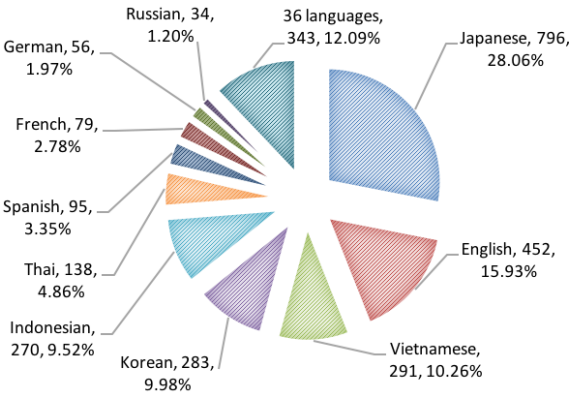

Figure 2: Distribution of learners' first languages.

Figure 3 shows the distribution of the 33,835 total errors. The most common error type was incorrect selection of linguistic components (13,278 cases or 39% of the total), followed by missing errors (12,155 cases), redundancy (6,066 cases) and word ordering errors (2,336 cases). Figure 4 further shows the histogram of the top 10 error tags (accounting for about 47% of all errors) among the total 124 distinct error tags. The most common errors were related to the incorrect selection of verbs (Sv) and nouns (Sn). Half of these errors are categorized as missing word-level linguistic components.

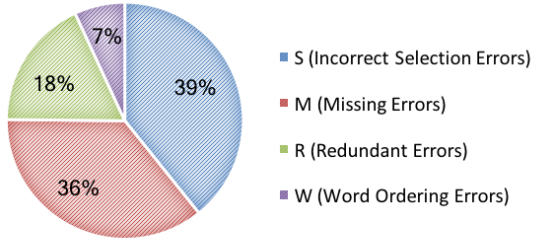

Figure 3: Distribution of all error tags.

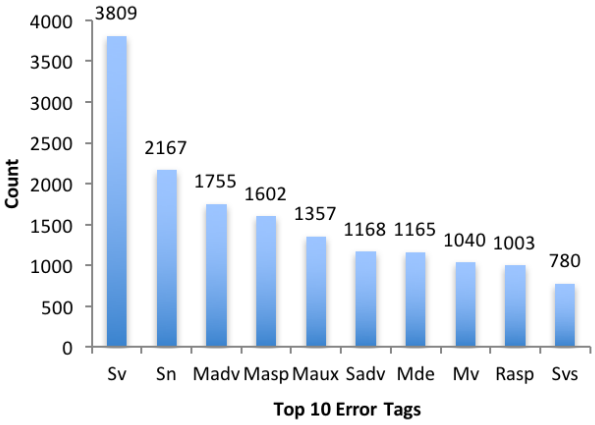

Figure 4: Distribution of the top 10 error tags.

We also developed and implemented a special-purpose retrieval system for the TOCFL learner corpus, which is available online at http://tocfl.itc.ntnu.edu.tw, to facilitate interlanguage analysis for second language acquisition (Lee et al., 2015a). Take the '讓' (*rang4* 'to make') sentence for example, we can choose the main error type S and the sub-type *rang*. Figure 5 shows the search results. We found that learners usually confuse '讓' (*rang4* 'to make') with '把' (*ba3* 'disposal marker'), '對' (*dui4* 'to someone'), and '給' (*gei3* 'to give'). It is difficult to check each sentence individually to find all erroneous and confused usages and this retrieval system effectively reduces required analysis time. Moreover, we can limit the search results for specific words, such as '把' (*ba3* 'disposal marker'), which will benefit observation and analysis. In addition to the filtering function, we can also select specific learners' attributes such as the learners' mother tongue or their L2 proficiency, thus increasing the ease and efficiency of interlanguage analysis.

Figure 5: Search results of rang4 sentences.

5. Shared Tasks

A part of our annotated TOCFL learner corpus was used to organize a series of shared tasks for Chinese grammatical error diagnosis hosted by NLPTEA workshops. These tasks seek to develop NLP techniques to automatically identify grammatical errors in Chinese sentences. In the NLPTEA 2014 workshop (Yu et al., 2014), the developed system is used to check a sentence for grammatical errors and, if found, identify the error type. In the NLPTEA 2015 workshop (Lee et al., 2015b), the system was required to additionally indicate the range of occurring errors. In the

NLPTEA 2016 workshop (Lee et al., 2016b), the task is basically the same, except that the target sentence may contain more than one error, and the HSK Dynamic Composition Corpus was also included for the task.

In addition to the provided data sets, participating systems were allowed to use the other language resources which should be found in the system description reports. We briefly describe them as follows.

For the NLPTEA 2014 shared task, the KUAS&NTNU system used manually constructed and automatically generated rules to identify grammatical errors (Chang et al., 2014). The UDS system designed an n-gram frequency-based approach to detect grammatical errors (Zampieri and Tan, 2014). The NTOU system defined several features to train SVM classifiers for error detection (Lin and Chan, 2014). The NCYU system adopted word segmentation and part-of-speech tagging techniques to identify missing and redundant error types (Yeh et al., 2014). To compensate for data insufficiency for supervised machine learning, the TMU system extracted a Chinese learner corpus from the Lang-8 website, and used it as a parallel corpus for phrase-based statistical machine translation for grammatical error identification (Zhao et al. 2014).

For the NLPTEA 2015 shared task, The HITSZ system presented an ensemble learning based method to detect and identify grammatical errors (Xiang et al. 2015). The SCAU system adopted a hybrid model by integrating rule-based and n-gram statistical methods for grammatical error diagnosis (Wu et al., 2015b). The CYUT team built an error diagnosis system based on the Conditional Random Fields (CRF) (Wu et al., 2015a). The NTOU system proposed two sentence likelihood functions based on frequencies of Google n-grams to diagnose grammatical errors (Lin, & Chen, 2015). The NCYU system also used statistical word and part-of-speech patterns based CRFs to detect grammatical errors (Yeh et al., 2015). The TMU examined corpus augmentation and explored syntax-based and hierarchical phrase-based translation models for use in this task (Zhao et al. 2015).

For the NLPTEA 2016 shared task, the ANO system and CYUT-III system diagnosed grammatical errors based on the CRF (Chen et al., 2016a; Liu et al., 2016) along with word order sensitive embedding approaches (Chou et al., 2016). The NTOU system generated and scored correction candidates for grammatical error diagnosis (Chen et al., 2016b). The HIT system adopted long short-term memory (LSTM) neural networks to identify grammatical errors (Zheng et al., 2016). The PKU system presented a model based on bidirectional LSTM (Huang, & Wang, 2016). The NCYU system proposed the structure of the recurrent neural network using LSTM to detect grammatical errors (Yeh et al., 2016). The YUN-HPCC system built single word embeddings based convolutional neural networks and LSTM neural networks for this task (Yang et al., 2016).

In terms of performance, a good system should have a high F1 score and a low false positive rate. Overall, none of the participating systems provided satisfactory results when measuring different metrics (i.e. False Positive Rate, Accuracy, Precision, Recall and F1), indicating the difficulty of developing systems for effective grammatical error diagnosis, especially in the context of CFL.

Recently, neural network-based deep learning techniques have shown promising results in identifying Chinese grammatical errors. However, a large amount of training data is needed to train and fine-tune the parameters of these complex networks. Another challenge is raised by the biased distribution of error types in the training instances reflecting real-world errors caused by CFL learners, emphasizing the importance of annotated learner corpora in tackling this research problem.

Shared tasks can be used to meaningfully compare the performance of various techniques using the same data sets and evaluation metrics. All evaluations encourage the proposal of unorthodox and innovative approaches which could lead to breakthroughs. Following each shared task, all gold standard data sets and evaluation tools are made publicly available for research purposes as follows:

- NLPTEA 2014 shared task: http://ir.itc.ntnu.edu.tw/lre/nlptea14cfl.htm
- NLPTEA 2015 shared task: http://ir.itc.ntnu.edu.tw/lre/nlptea15cged.htm
- NLPTEA 2016 shared task: http://ir.itc.ntnu.edu.tw/lre/nlptea16cged.htm

6. Conclusions

This study presents the construction of a TOCFL learner corpus based on 2837 learner essays annotated using hierarchical tagging sets. Our error tags are used to simultaneously represent two kinds of grammatical error classifications. The first capital letter of each error tags denotes the coarse-grained surface differences, while the subsequent lowercase letters denote the fine-grained linguistic categories. Native Chinese-speaking annotators were trained to follow our annotation guidelines for the error-tagging task, and identified 33,835 grammatical errors. We then used SGML to format the annotations and their corresponding corrections along with learners' accompanying metadata. Parts of the annotated TOCFL learner corpus have been used to organize shared tasks for Chinese grammatical error identification. We also investigate participating systems to better understand the current capabilities and challenges. The shared-task datasets are publicly available to facilitate further research. We plan to release the entire learner corpus in fully annotated SGML format in the hopes that this resource can facilitate future development in related research areas.

7. Acknowledgements

This study was partially supported by the Ministry of Science and Technology (MOST), Taiwan under the grant: MOST 103-2221-E-003-013-MY3, MOST 105-2221-E-003-020-MY2, MOST 106-2221-E-003-030-MY2, and MOST 107-2634-F-002-019-. We also thank all the annotators for their efforts in the error-tagging task.

8. Bibliographical References

Boyd, A. (2010). EGALE: an error-annotated corpus of beginning learner German. Proceedings of 7th International Conference on Language Resources and Evaluation (LREC'10), pages 1-6.

Chang, T.-H., Sung, Y.-T., Hong, J.-F. and Chang, J.-I (2014). KNGED: a tool for grammatical error diagnosis of Chinese sentences. Proceedings of the 1st Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 48-55.

Chen, P.-L., Wu, S.-H., Chen, L.-P. and Yang, P.-C. (2016a). CYUT-III system at Chinese grammatical error diagnosis task. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 63-72.

Chen, S.-H., Tsai, Y.-L. and Lin C.-J. (2016b). Generating and scoring correction candidates in Chinese grammatical error diagnosis. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 131-139.

Chou, W.-C., Lin, C.-K., Liao, Y.-F. and Wang, Y.-R. (2016). Word order sensitive embedding features/conditional random field-based Chinese grammatical error detection. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 73-81.

Cui, X. and Zhang, B. (2011). The principles for building the "International Corpus of Learner Corpus". *Applied Linguistics*, 2011(2):100-108.

Dahlmeier, D., Ng. H. T. and Wu, S. M. (2013). Building a large annotated corpus of learner English: the NUS corpus of learner English. Proceedings of the 8th Workshop on the Innovative Use of NLP for Building Educational Applications (BEA'13), pages 22-31.

Dale, R. and Kilgarriff, A. (2011). Helping our own: The HOO 2011 pilot shared task. Proceedings of the 13th European Workshop on Natural Language Generation (NLG'11), pages 1-8.

Dale, R., Anisimoff, I. and Narroway, G. (2012). HOO 2012: A report on the preposition and determiner error correction shared task. Proceedings of the 7th Workshop on the Innovative Use of NLP for Building Educational Applications (BEA'12), pages 54-62.

Díaz-Negrillo, A. and Fernández-Domínguez, J. (2006). Error tagging systems for learner corpora. *Revista española de lingüística aplicada (RESLA)*, 19:83-102.

Dickinson, M. and Ledbetter. S. (2012). Annotating errors in a Hungarian learner corpus. Proceedings of 8th International Conference on Language Resources and Evaluation (LREC'12), pages 1659-1644.

Gillard, P. and Gadsby, A. (1998). Using a learners' corpus in compiling ELT dictionaries. In S. Granger, & G. Leech (Eds.), *Learner English on Computer*. Addison Wesley Longman, London, pages 159-171.

Granfeldt, J., Nugues, P., Ågren, M., Thulin, J., Persson, E. and Schlyter, S. (2006). CEFLE and direct profil: a new computer learner corpus in French L2 and a system for grammatical profiling. Proceedings of 5th International Conference on Language Resources and Evaluation (LREC'06), pages 565-570.

Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL*, 37(3):538-546.

Granger, S. (2015). Contrastive interlanguage analysis: a reappraisal. *International Journal of Learner Corpus Research*, 1(1):25-74.

Hana, J., Rosen, A., Škodová, S. and Štindlová, B. (2010). Error-tagged learner corpus of Czech. Proceedings of the 4th Linguistic Annotation Workshop (LAW'10), pages 11-19.

Huang, S. and Wang, H. (2016). Bi-LSTM neural networks for Chinese grammatical error diagnosis. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 148-154.

Leacock, C., Chodorow, M., Gamon, M. and Tetreault, J. (2014). Automated grammatical error detection for language learners, 2nd edition, Morgan & Claypool Publishers.

Lee, L.-H., Chang, L.-P., Liao, B.-S., Cheng, W.-L. and Tseng, Y.-H. (2015a). A retrieval system for interlanguage analysis. Proceedings of 23rd International Conference on Computers in Education (ICCE'15), pages 599-601.

Lee, L.-H., Chang, L.-P. and Tseng, Y.-H. (2016a). Developing learner corpus annotation for Chinese grammatical errors. Proceedings of the 20th International Conference on Asian Language Processing (IALP'16), pages 254-257.

Lee, L.-H., Lee, K.-C., Chang, L.-P., Tseng, Y.-H., Yu, L.-C. and Chen, H.-H. (2014). A tagging editor for learner corpus annotation and error analysis. Proceedings of 22nd International Conference on Computers in Education (ICCE'14), pages 806-808.

Lee, L.-H., Rao, G., Yu, L.-C., Xun, E., Zhang, B. and Chang, L.-P. (2016b). Overview of the NLP-TEA 2016 shared task for Chinese grammatical error diagnosis. Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 40-48.

Lee, L.-H., Yu, L.-C. and Chang, L.-P. (2015b). Overview of the NLP-TEA 2015 shared task for Chinese grammatical error diagnosis. Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 1-6.

Lee, S.-H., Dickinson M. and Israel, R. (2012). Developing learner corpus annotation for Korean particle errors. Proceedings of the 6th Linguistic Annotation Workshop (LAW'12), pages 129-133.

Lin, C.-J. and Chan, S.-H. (2014). Description of NTOU Chinese grammar checker in CFL 2014. Proceedings of the 1st Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 75-78.

Lin, C.-J. and Chen, S.-H. (2015). NTOU Chinese grammar checker for CGED shared task. In Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 15-19.

Little, D. (2006). The Common European Framework of Reference for languages: content, purpose, origin, reception, and impact. *Language Teaching*, 39(3):167-190.

Liu, Y., Han, Y., Zhuo, L. and Zan, H. (2016). Automatic grammatical error detection for Chinese based on conditional random field. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 57-62.

Malmasi, S. and Dras, M. (2015). Large-scale native language identification with cross-corpus evaluation. Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'15), pages 1403-1409.

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. Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP'11), pages 147-155.

Ng, H. T., Wu, S. M., Briscoe, T., Hadiwinoto, C., Susanto, R. H. and Bryant, C. (2014). The CoNLL-2014 shared task on grammatical error correction. Proceedings of the

18th Conference on Computational Natural Language Learning (CoNLL'14), Shared Task, page 1-14.

Ng, H. T., Wu, S. M., Wu, Y., Hadiwinoto, C. and Tetreault, J. (2013). The CoNLL-2013 shared task on grammatical error correction. Proceedings of the 17th Conference on Computational Natural Language Learning (CoNLL'13), Shared Task, page 1-12.

Nicholls, D. (2003). The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT. Proceedings of the Corpus Linguistics 2003 Conference (CL'03), pages 572-581.

Rao, G., Zhang, B. and Xun, E. (2017). IJCNLP-2017 Task 1: Chinese Grammatical Error Diagnosis. Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP'17), Shared Tasks, pages 1-8.

Sakaguchi, K., Arase, Y. and Komachi, M. (2013). Discriminative approach to fill-in-the-blank quiz generation for language learners. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL'13), pages 238-242.

Swanson, B. and Charniak, E. (2013). Extracting the native language signal for second language acquisition. Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'13), pages 85-94.

Tenfjord, K., Meurer, P. and Hofland, K. (2006). The ASK Corpus – a language learner corpus of Norwegian as a second language. Proceedings of 5th International Conference on Language Resources and Evaluation (LREC'06), pages 1821-1824.

Tono, T. (2003). Learner corpora: design, development and applications. Proceedings of the Corpus Linguistics 2003 Conference (CL'03), pages 800-809.

Wang, C. and Seneff, S. (2007). Automatic assessment of student translations for foreign language tutoring. Proceedings of the 2007 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT'07), pages 468-475.

Wang, M., Malmasi, S. and Huang, M. (2015). The Jinan Chinese learner corpus. Proceedings of the 10th Workshop on the Innovative Use of NLP for Building Educational Applications (BEA'15), pages 118-123.

Wu, S.-H., Chen, P.-L., Chen, L.-P., Yang, P.-C. and Yang, R.-D. (2015a). Chinese grammatical error diagnosis by conditional random fields. Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 7-14.

Wu, X., Huang, P., Wang, J., Guo, Q., Xu, Y. and Chen, C. (2015b). Chinese grammatical error diagnosis system based on hybrid model. Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 117-125.

Xiang, Y., Wang, X., Han, W. and Hong, Q. (2015). Chinese grammatical error diagnosis using ensemble learning. Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15). pages 99-104.

Yannakoudakis, H., Briscoe, T. and Medlock, B. (2011). A new dataset and method for automatically grading ESOL texts. Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL'11), pages 180-189.

Yang, J., Peng, B., Wang, J., Zhang, J. and Zhang, X. (2016). Chinese grammatical error diagnosis using single word embedding. In Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 155-161.

Yeh, J.-F., Hsu, T.-W. and Yeh, C.-K. (2016). Grammatical error detection based on machine learning for Mandarin as second language learning. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 140-147.

Yeh, J.-F., Lu, Y.-Y., Lee, C.-H., Yu, Y.-H. and Chen, Y.-T. (2014). Detecting grammatical error in Chinese sentence for foreign. Proceedings of the 1st Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 62-68.

Yeh, J.-F., Yeh, C.-K., Yu, K.-H., Li, Y.-T. and Tsai, W.-L. (2015). Conditional random field-based grammatical error detection for Chinese as second language. In Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 105-110.

Yu, L.-C., Lee, L.-H. and Chang, L.-P. (2014). Overview of grammatical error diagnosis for learning Chinese as a foreign language. Proceedings of the 1st Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 42-47.

Zampieri, M. and Tan, L. (2014). Grammatical error detection with limited training data: the case of Chinese. Proceedings of the 1st Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 69-74.

Zhang, B. and Cui, X. (2013). Design concepts of the construction and research of the inter-language corpus of Chinese from global learners. *Language Teaching and Linguistic Study*, 2013(5):27-34.

Zhao, Y., Komachi, M. and Ishikawa, H. (2014). Extracting a Chinese learner corpus from the Web: grammatical error correction for learning Chinese as a foreign language with statistical machine translation. Proceedings of the 1st Natural Language Processing Techniques for Educational Applications (NLPTEA'14), pages 56-61.

Zhao, Y., Komachi, M. and Ishikawa, H. (2015). Improving Chinese grammatical error correction with corpus augmentation and hierarchical phrase-based statistical machine translation. Proceedings of the 2nd Natural Language Processing Techniques for Educational Applications (NLPTEA'15), pages 111-116.

Zheng, B., Che, W., Guo, J. and Liu, T. (2016). Chinese grammatical error diagnosis with long short-term memory networks. Proceedings of the 3rd Natural Language Processing Techniques for Educational Applications (NLPTEA'16), pages 49-56.