Generating and Scoring Correction Candidates in Chinese Grammatical Error Diagnosis

Shao-Heng Chen, Yu-Lin Tsai, and Chuan-Jie Lin
Department of Computer Science and Engineering
National Taiwan Ocean University
No 2, Pei-Ning Road, Keelung, Taiwan ROC
{shchen.cse, yltsai.cse, cjlin}@ntou.edu.tw

Abstract

Grammatical error diagnosis is an essential part in a language-learning tutoring system. Based on the data sets of Chinese grammar error detection tasks, we proposed a system which measures the likelihood of correction candidates generated by deleting or inserting characters or words, moving substrings to different positions, substituting prepositions with other prepositions, or substituting words with their synonyms or similar strings. Sentence likelihood is measured based on the frequencies of substrings from the space-removed version of Google n-grams. The evaluation on the training set shows that Missing-related and Selection-related candidate generation methods have promising performance. Our final system achieved a precision of 30.28% and a recall of 62.85% in the identification level evaluated on the test set.

1 Introduction

Although that Chinese grammars are not defined as clearly as English, Chinese native speakers can easily identify grammatical errors in sentences. This is one of the most difficult parts for foreigners to learn Chinese. They are often uncertain about the proper grammars to make sentences. It is an interesting research topic to develop a Chinese grammar checker to give helps in Chinese learning. There have been several researches focusing on Chinese (Wu *et al*., 2010; Chang *et al*., 2012; Yu and Chen, 2012; Lee *et al*., 2014).

There are 3 evaluation tasks focusing on Chinese grammatical error diagnosis. CGED 2014 (Yu *et al*., 2014) defined four kinds of grammatical errors: redundant, missing, selection, and disorder. At most one error occurred in one sentence. The evaluation was based on error detection and error classification in sentence level. CGED 2015 (Lee *et al*., 2015) further required the positions of the errors. CGED 2016 tested on the ability of finding multiple errors in one sentence.

This paper is organized as follows. Section 2 describes Chinese grammatical error diagnosis task. Section 3 defines the sentence likelihood scoring function. Section 4 explains how correction candidates are generated for different error types. Section 5 clarifies the details of decision making. Section 6 shows the evaluation results and Section 7 concludes this paper.

2 Task Definition

The task of Chinese grammatical error diagnosis (CGED) is defined as follows. Given a sentence, a CGED system should first decide if there is any error occurring in the sentence. If so, report its error type, starting and ending positions.

Errors are divided into four types: redundant, missing, selection, and disorder, which are shortly explained here. All examples are selected from CGED-2015 training set.

Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, pages 131–139, Osaka, Japan, December 12 2016.

- Redundant: some unnecessary character appears in a sentence

  [A2-0598, Redundant, 3, 3]
  (X) 他是真很好的人 (*He is a really very good man.)
  (O) 他是很好的人 (He is a very good man.)

- Missing: some necessary character is missing in a sentence

  [B1-0046, Missing, 4, 4]
  (X) 母親節一個禮拜就要到了 (*Mother's Day is coming in one week.)
  (O) 母親節再一個禮拜就要到了 (Mother's Day is coming in one more week.)

- Selection: a word is misused and should be replaced by another word

  [B1-1544, Selection, 1, 2]
  (X) 還給原來的地方只花幾秒鐘而已
  (*It only takes a few seconds to return it to its original place.)
  (O) 放回原來的地方只花幾秒鐘而已
  (It only takes a few seconds to put it back to its original place.)

  Note that sometimes a SELECTION error looks like a missing or redundant case rather than a misused word. It is because there are many one-character words in Chinese. An example is given as follows.

  [B1-1546, Selection, 5, 5]
  (X) 關於跟你見的事 (*About the seeing with you…)
  (O) 關於跟你見面的事 (About the meeting with you…)

- Disorder: some words' locations should be changed

  [B1-2099, Disorder, 4, 6]
  (X) 當然我會<u>一定</u>開心 (*Of course I will be <u>certainly</u> happy.)
  (O) 當然我<u>一定</u>會開心 (Of course I will <u>certainly</u> be happy.)

CGED systems were evaluated in 3 levels: detection level for whether each sentence has errors; identification level for what type the error is; and position level for where the error appears (in terms of Chinese characters). Evaluation metrics are well-known accuracy, precision, recall, and F-measure.

3 Sentence Likelihood Scoring

The systems proposed in this paper were based on our previous work of (Lin and Chen, 2015). Our contributions include proposing candidate generation for Selection-type errors (described in Section 4), and observing the effects of factors in the candidate generation methods and sentence scoring functions. We also examined how to propose multiple errors in a given sentence so that our system can be evaluated on the CGED 2016 test set.

In our previous work (Lin and Chen, 2015), we have defined a sentence likelihood scoring function to measure the likelihood of a sentence to be common and correct. This function uses frequencies provided in the Chinese Web 5-gram dataset in a way described as follows.

Chinese Web 5-gram[1] consists of real data released by Google Inc. which were collected from a large amount of webpages in the World Wide Web. Entries in the dataset are unigrams to 5-grams. Frequencies of these n-grams are also provided. Some examples from the Chinese Web 5-gram dataset are given in the left part of Table 1.

In order to avoid interference of word segmentation errors, we decided to use substrings instead of word n-grams as the scoring units of likelihood. When scoring a sentence, frequencies of all substrings in all lengths are used to measure the likelihood.

[1] https://catalog.ldc.upenn.edu/LDC2010T06

Gram	Text	Freq	Length	Text	Freq
Unigram	稀釋劑	17260	9	稀釋劑	17260
Bigram	蒸發量 超過	69	15	蒸發量超過	69
Trigram	能量 遠 低於	113	15	能量遠低於	113
4-gram	張貼 色情 圖片 或	73	18	張貼色情圖片或	73
5-gram	幸好 我們 發現 得 早	155	24	幸好我們發現得早	155

Table 1. Examples of Google N-grams (before and after Space Removal)

Frequencies of substrings are derived by removing space between n-grams in the Chinese Web 5-gram dataset. For instances, n-grams in the left part of Table 1 will become the strings in the right part, where length of a substring is measured in bytes and a Chinese character often occupies 3 bytes in UTF-8 encoding. Note that if two or more different n-grams are transformed into the same substring after removing the space, they become one entry and its new frequency is the summation of their original frequencies. Simplified Chinese words were translated into Traditional Chinese in advanced.

Some notations are explained as follows. Given a sentence S, let $SubStr(S, n)$ be the set of all substrings in S whose lengths are n bytes, and ***Google String Frequency*** $gsf(u)$ be the frequency of a string u in the modified Chinese Web 5-gram dataset. If a string does not appear in that dataset, its gsf value is defined to be 1 (so that its logarithm becomes 0).

Equation 1 gives the equation of ***length-weighted string log-frequency score*** $SL(S)$. Each substring u in S contributes a score of the logarithm of its Google string frequency weighted by u's length n. The value of n starts from 6, because most content words are not shorter than 6 bytes (i.e. two Chinese characters).

$$SL(S) = \sum_{n=6}^{len(S)} \left(n \times \sum_{u \in SubStr(S,n)} \log(gsf(u)) \right) \qquad \text{Eq 1.}$$

This function was also explained in the work of Lin and Chu (2015). Please refer to that paper for examples of how to compute the sentence generation likelihood scores.

4 Correction Candidate Generation

4.1 Character or Word Deletion (Case of Redundant)

Generating correction candidates in the case of Redundant type is quite straightforward: simply removing any substring in an arbitrary length. However, in order not to generate too many unnecessary candidates, we only do the removal under three special cases: removing one character, removing two-adjacent characters, and removing one word whose length is no longer than two Chinese characters. The examples are as follows, where org is the original sentence and new is a correction candidate.

[B1-0764] org: 我 很 想到 跟 你 見面
<u>(*By removing characters*)</u>
new: 很想到跟你見面(*by removing* 我)
new: 我想到跟你見面(*by removing* 很)
……
new: 我很想到跟你見(*by removing* 面)
new: 想到跟你見面 (*by removing* 我很)
new: 我到跟你見面 (*by removing* 很想)
……
new: 我很想到跟你 (*by removing* 見面)
<u>(*By removing one word*)</u>
new: 很想到跟你見面(*by removing* 我)
new: 我想到跟你見面(*by removing* 很)
new: 我很跟你見面 (*by removing* 想到)
……
new: 我很想到跟你 (*by removing* 見面)

Given a sentence with n characters constituting m words (whose lengths are not longer than 2 Chinese characters), removing one character will generate n candidates, removing two adjacent characters will generate $n-1$ candidates, and removing one word will generate m candidates.

Obviously all candidates generated by removing one word were also generated by other two methods. We would like to see the efficiency of each method in terms of accuracy and the size of candidate set.

4.2 Character Insertion (Case of Missing)

The idea of generating correction candidates in the case of Missing type is to insert a character or a word into the given sentence. But it is impractical to enumerate candidates by inserting every known Chinese characters or words. We observed the CGED 2015 training set (Lin and Chen, 2015) and collected 34 characters which were commonly missing in the essays written by Chinese-learning foreign students. Table 2 shows some of these frequent missing characters in the training data. These 34 characters occurred at least three times and covered 73.7% of the missing errors in the CGED 2015 training set.

Char	Freq	Char	Freq	Char	Freq
的	74	有	24	要	13
了	65	會	18	在	12
是	44	就	17	過	12
都	34	很	16	讓	11

Table 2. Examples of Frequent Missing Characters

Insertion happens between characters or words as usual. Examples of insertion between characters are as follows.

[B1-0764] org: 我 很 想到 跟 你 見面
(*By inserting between characters*)
new: $\boxed{\text{的}}$我很想到跟你見面 (*by inserting* 的 *before* 我)
new: 我$\boxed{\text{的}}$很想到跟你見面 (*by inserting* 的 *between* 我 *and* 很)
......
new: 我很想到跟你見面$\boxed{\text{的}}$ (*by inserting* 的 *after* 面)
new: $\boxed{\text{了}}$我很想到跟你見面 (*by inserting* 了 *before* 我)
......
new: 我很想到跟你見面$\boxed{\text{買}}$ (*by inserting* 買 *after* 面)

Given a sentence with n characters constituting m words in total, insertion between characters will generate $34\times(n+1)$ candidates and insertion between words will generate $34\times(m+1)$ candidates.

4.3 Substring Moving (Case of Disorder)

Generating correction candidates in the case of Disorder type is also straightforward: simply moving any substring in any length to another position to its right (not to its left so that no duplication will be produced). Examples of substring moving are as follows.

[B1-0764] org: 我 很 想到 跟 你 見面
(*By moving a substring to a new position between characters*)
new: 很$\boxed{\text{我}}$想到跟你見面 (*by moving* 我 *to the position between* 很 *and* 想)
new: 很想$\boxed{\text{我}}$到跟你見面 (*by moving* 我 *to the position between* 想 *and* 到)
......
new: 很想到跟你見面$\boxed{\text{我}}$ (*by moving* 我 *to the position after* 面)
new: 想$\boxed{\text{我很}}$到跟你見面 (*by moving* 我很 *to the position between* 想 *and* 到)
......
new: 面$\boxed{\text{我很想到跟你見}}$ (*by moving* 我很想到跟你見 *to the position after* 面)

Given a sentence with n characters, there are $(n-k)$ substrings whose lengths are k ($1 \le k \le n-1$). A substring with lengths k at the position t ($1 \le t \le n-k+1$) can be moved to $(n-k-t+1)$ new positions. By summing on all k and t, there will be $(n^3-n)/6$ candidates by moving substrings to positions between characters. Similarly, there will be $(m^3-m)/6$ candidates by moving substrings to positions between words for a sentence with m words. The number will grow fast if the given sentence is long.

4.4 Preposition Substitution (Case 1 of Selection)

In our observation, some selection errors are misuse of prepositions. But unlike the case in English, it is not the most major errors in selection type.

To generate the correction candidates for preposition substitutions, we first extracted all prepositions in the Academia Sinica Balanced Corpus (ASBC for short hereafter, cf. Chen *et al*., 1996). An input sentence is word-segmented and POS-tagged automatically before hand. Correction candidates are generated by replacing each preposition (whose POS is "P") in the given sentence by other prepositions. Examples of preposition substitution are as follows, where only the word "跟" (with) is a preposition.

> [B1-0764] org: 我 很 想到 跟(P) 你 見面
> (*By moving a substring to a new position between characters*)
> new: 我很想到$\boxed{在}$你見面 (*by replacing* 跟 *by* 在)
> new: 我很想到$\boxed{為}$你見面 (*by replacing* 跟 *by* 為)
> ......

There are 243 prepositions in ASBC. Given a sentence containing *k* prepositions, $243\times k$ correction candidates will be generated.

4.5 Synonym Substitution (Case 2 of Selection)

In our observation, another type of selection errors is misuse of words which are synonyms. As we known, even synonyms cannot freely replace each other without considering context.

To generate the correction candidates for synonym substitutions, we consulted a Chinese thesaurus, Tongyici Cilin[2] (the extended version; Cilin for short hereafter). A given sentence is word-segmented before hand. Correction candidates are generated by replacing each word in the given sentence by its synonyms in Cilin if any. Examples of synonym substitution are as follows.

> [B1-0764] org: 我 很 想到 跟 你 見面
> (*By moving a substring to a new position between characters*)
> new: 我很$\boxed{悟出}$跟你見面 (*by replacing* 想到 *by its synonym* 悟出)
> new: 我很$\boxed{想開}$跟你見面 (*by replacing* 想到 *by its synonym* 想開)
> ......
> new: 我很想到跟你$\boxed{相會}$ (*by replacing* 見面 *by its synonym* 相會)

The number of candidates depends on the number of Cilin terms and their synonyms in a given sentence. Generally the number is not too large.

4.6 Similar String Substitution (Case 3 of Selection)

In our observation, we found a special type of selection errors that the misusing words were lexically similar to the correct ones. It should be the case when the writer tried to use a word but misused another word with similar looking, such as "仔細" (carefully) and "細節" (details).

To generate but not over-generate the correction candidates for similar string substitutions, we first collected all 2-character words in ASBC and Cilin. Correction candidates are generated by replacing each 2-character word in the given sentence by another 2-character word having at least one character in common, such as "仔細" and "細節" where "細" appears in both words, or "合適" (suitable, *adjective*) and "適合" (suiting, *verb*) where both characters appear in both words. Examples of similar string substitution are given in the next page.

The number of candidates depends on the number of 2-character words and their similar words in a given sentence. Generally the number is not small.

---

[2] http://ir.hit.edu.cn/
http://www.ltp-cloud.com/

[B1-0764] org: 我 很 想到 跟 你 見面
(*By moving a substring to a new position between characters*)
new: 我很$\boxed{\text{想出}}$跟你見面 (*by replacing* 想到 *by a similar string* 想出)
new: 我很$\boxed{\text{思想}}$跟你見面 (*by replacing* 想到 *by a similar string* 思想)
......
new: 我很想到跟你$\boxed{\text{面向}}$ (*by replacing* 見面 *by a similar string* 面向)

5 Error Detection and Classification

5.1 Error Decision

All correction candidates, as well as the original sentence, are scored by the sentence likelihood function in Eq 1. They are ranked according to their likelihood scores. If the top-1 sentence is the original sentence, report it as a "Correct" case. Otherwise, output the first top 2 candidates as errors by reporting their corresponding error types and occurring positions. If the top-2 candidate conflicts with the top-1 candidate in position (i.e. they are overlapped), discard it and take the next candidate until 2 errors are reported or the rank of the original sentence is reached. Moreover, if more than 2 candidates have the same scores, report them all (if no position confliction).

The choice of 2 is based on the average errors appearing in a sentence in the CGED 2016 training set, which are 1.314 errors in one sentence. To increase recall, we decide to propose 2 errors for each sentence. We have also done an observation by propose only 1 error for one sentence. We found that the precision was not improved but the recall was greatly harmed.

5.2 Selection Error Classification Fixing

For a correction candidate of the Redundant type, if the deleted character appears in a multi-character word in the original sentence, it should be a Selection-type error. An example is given as follows.

[B1-0764] Redundant => Selection
(X) 我 很 想$\boxed{\text{到}}$ 跟 你 見面 (*I really want to $\boxed{\text{to}}$ meet you.)
(O) 我 很 想 跟 你 見面 (I really want to meet you.)

In this example, the deleted character "到" appears in a 2-character word "想到" in the original sentence. This error will be classified into Selection type because the word "想到" (think-of) should be corrected into "想" (want). Our system will check the word boundary in the original sentence to fix this error type classification.

Similarly for a correction candidate of the Missing type, if the inserted character appears in a multi-character word in the new sentence, it should be a Selection-type error. An example is given as follows.

[B1-1047] Missing => Selection
(X) 我 真 很 怕 (*I am real scared.)
(O) 我 真$\boxed{\text{的}}$ 很 怕 (I am $\boxed{\text{really}}$ scared.)

In this example, the inserted character "的" appears in a 2-character word "真的" in the new sentence. This error will be classified into Selection type because the word "真" (real) should be corrected into "真的" (really). Our system will check the word boundary in the new sentence to fix this error type classification.

6 Experiments

6.1 Evaluating of Correction Candidates in Individual Methods

Table 3 shows the evaluation results when each candidate generation method is used individually. These methods were evaluated on the whole CGED 2016 training set. The names in the "Method" column represent the following candidate generation methods:

- RDN_char: deleting one character
- RDN_2char: deleting two adjacent characters
- RDN_word: deleting one word
- MIS_char: inserting frequent missing characters between characters
- MIS_word: inserting frequent missing characters between words
- WDO_char: moving substrings based on characters
- WDO_word: moving substrings based on words
- SEL_P: substituting prepositions
- SEL_CLN: substituting with synonyms in Cilin
- SEL_SIM: substituting with similar 2-character words
- SEL_R1C: fixed Selection type from RDN_char cases
- SEL_R2C: fixed Selection type from RDN_2char cases
- SEL_M1C: fixed Selection type from MIS_char cases
- SEL_M1W: fixed Selection type from MIS_word cases

Method	#Cands	P	R	F1	Method	#Cands	P	R	F1
RDN_char	433613	13.72	3.03	4.97	SEL_P	6433812	10.03	9.88	9.95
RDN_2char	390096	6.92	0.40	0.75	SEL_CLN	10862488	8.22	25.68	12.46
RDN_word	281069	13.62	3.21	5.20	SEL_SIM	49465905	5.00	9.67	6.60
MIS_char	16215314	6.48	15.56	9.15	SEL_R1C	(206825)	8.03	0.13	0.26
MIS_word	11337266	6.54	15.41	9.18	SEL_R2C	(140564)	0.60	0.01	0.02
WDO_char	14247107	1.47	2.16	1.75	SEL_M1C	(152175)	8.28	14.49	10.53
WDO_word	4278827	1.36	1.91	1.59	SEL_M1W	(151986)	8.33	14.47	10.58

Table 3. Evaluation Results of Individual Candidate Generation Methods

In Table 3, the "#Cands" columns show the number of correction candidates generated by different methods. Note that the numbers of candidates of the fixed Selection types are included in the Redundant and Missing sets. Moreover, only those candidates whose scores are higher than the original sentences can have the chance to fix their error types. P, R, and F1 stand for precision, recall, and F1 measure, respectively. All evaluations were done in position level (cf. Section 2).

As we can see in Table 3, Selection-related methods achieved better performance. Maybe it is because Selection is the major error type in the dataset. The missing-related methods achieved good recalls while the Redundant-related methods achieved good precisions, but recall dominates the experimental results in this observation. The missing-related methods also provided many correct candidates for Selection type.

The Disorder-related methods were surprisingly poor. Although a correct answer in this type could be generated, too many incorrect candidates were also generated and lowered the rank of the correct candidate. The performance of RDN_2char was also poor. We will discard these two methods in the final system.

6.2 Evaluating of Correction Candidates by All Methods on the Training Data

We have tried every combination of generation methods and evaluated their performances when proposing top-2 candidates. Table 4 shows the performance of the best system and its comparisons evaluated in position level. P, R, and F1 stand for precision, recall, and F1 measure, respectively.

The best system used only Missing-related and Selection-related candidate generation methods. By comparing S2 with S1, adding candidates from Redundant-related method did not affect the performance at all. But by removing Missing-related candidates (S3 and S4), the performance would drop in a certain degree.

To see the effect of 3 different Selection-related methods, each method was discarded from the best system. As we can see, synonym substitution was the best because its absence (S6) decreased the performance the most.

However, if we chose the best system as our final system, we would never be able to capture the Redundant-type and Disorder-type errors. We still used them in the final systems.

#	Systems	P	R	F
S1	MIS_char, SEL_M1C, SEL_M1W, SEL_P, SEL_CLN, SEL_SIM	9.11	30.95	14.08
S2	MIS_char, SEL_M1C, SEL_M1W, SEL_P, SEL_CLN, SEL_SIM, RDN_char	9.11	30.95	14.08
S3	MIS_char, SEL_P, SEL_CLN, SEL_SIM	8.64	29.03	13.32
S4	SEL_P, SEL_CLN, SEL_SIM	8.40	27.53	12.87
S5	MIS_char, SEL_M1C, SEL_M1W, SEL_CLN, SEL_SIM	8.91	30.26	13.76
S6	MIS_char, SEL_M1C, SEL_M1W, SEL_P, SEL_SIM	7.66	23.05	11.50
S7	MIS_char, SEL_M1C, SEL_M1W, SEL_P, SEL_CLN	9.07	30.81	14.02

Table 4. Evaluation Results of Overall Systems on the Training Data

	Detection Level			Identification Level			Position Level		
	P	R	F	P	R	F	P	R	F
S8	51.73	100.00	68.19	30.28	62.17	40.73	2.03	13.55	3.53
S9	51.73	100.00	68.19	29.78	62.85	40.41	2.02	13.50	3.52

Table 5. Evaluation Results of Final Systems on the Test Data

6.3 Evaluating on the Test Data

Two final systems S8 and S9 were evaluated on the CGED 2016 test set. Table 5 shows the evaluation results. The definitions of S8 and S9 are as follows:

- S8: RDN_char, RDN_word, MIS_char, SEL_*
- S9: RDN_char, RDN_word, MIS_char, SEL_*, WDO_word

As we can see in Table 5, the two systems have similar performance. It means that the candidates from Missing-related and Selection-related methods dominate the systems. But the system with the Disorder-related method WDO_word is a little worse than the system without using it, which is consistent to our previous observation.

7 Conclusion

This paper describes the design of our Chinese grammatical error diagnosis system. Correction candidates corresponding to 4 error types are first generated. The sentence likelihood scores of these candidates are measured based on web frequencies provided in the space-removed version of Google n-grams. Top-2 candidates are reported as errors.

Redundant correction candidates are generated by deleting characters or words; Missing candidates are generated by inserting frequently missed characters into position between characters or words; Disorder candidates are generated by moving sequences of characters or words to different positions; Selection candidates are generated by substituting prepositions with other prepositions and substituting words with their synonyms in Tongyici Cilin.

The best system uses the candidates generated for Missing and Selection types. Adding candidates for Redundant type does not affect the performance, but adding candidates for Disorder type harms the performance.

When evaluating on CGED 2016 test set, our final system achieved a precision of 30.28% and a recall of 62.85% in the identification level, which is better than our system in 2015. The final system proposed in this paper used candidates for Redundant, Missing, and Selection types.

The performance looks not good enough, which means that the task is very hard. We need to find out the reason why Redundant and Disorder candidates cannot improve the performance. More rules or features should be discovered in the future.

Acknowledgement

This research was funded by the Taiwan Ministry of Science and Technology (grant MOST 105-2221-E-019-071-).

Reference

Ru-Yng Chang, Chung-Hsien Wu, and Philips Kokoh Prasetyo (2012) "Error Diagnosis of Chinese Sentences Using Inductive Learning Algorithm and Decomposition-Based Testing Mechanism," *ACM Transactions on Asian Language Information Processing*, Vol. 11, No. 1, article 3.

Keh-Jiann Chen, Chu-Ren Huang, Li-ping Chang, and Hui-Li Hsu (1996) "Sinica Corpus: Design Methodology for Balanced Corpora," *Proceeding of the 11th Pacific Asia Conference on Language, Information and Computation*, pp. 167-176.

Lung-Hao Lee, Liang-Chih Yu, Kuei-Ching Lee, Yuen-Hsien Tseng, Li-Ping Chang, and Hsin-Hsi Chen (2014) "A Sentence Judgment System for Grammatical Error Detection," *Proceedings of the 25th International Conference on Computational Linguistics* (*COLING '14*), pp. 67-70.

Lung-Hao Lee, Liang-Chih Yu, and Li-Ping Chang (2015) "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* (*NLP-TEA 2*), *the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing* (*ACL-IJCNLP 2015*), pp. 1-6.

Chuan-Jie Lin and Shao-Heng Chen (2015) "NTOU Chinese Grammar Checker for CGED Shared Task," *Proceedings of The 2nd Workshop on Natural Language Processing Techniques for Educational Applications* (*NLP-TEA 2*), *the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing* (*ACL-IJCNLP 2015*), pp. 15-19.

Chuan-Jie Lin and Wei-Cheng Chu (2015) "A Study on Chinese Spelling Check Using Confusion Sets and N-gram Statistics," *International Journal of Computational Linguistics and Chinese Language Processing* (*IJCLCLP*), Vol. 20, No. 1, pp. 23-48.

Chung-Hsien Wu, Chao-Hong Liu, Matthew Harris, and Liang-Chih Yu (2010) "Sentence Correction Incorporating Relative Position and Parse Template Language Models," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 18, No. 6, pp. 1170-1181.

Chi-Hsin Yu and Hsin-Hsi Chen (2012) "Detecting Word Ordering Errors in Chinese Sentences for Learning Chinese as a Foreign Language," *Proceedings of the 24th International Conference on Computational Linguistics* (*COLING '12*), pp. 3003-3017.

Liang-Chih Yu, Lung-Hao Lee, and Li-Ping Chang (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*), pp. 42-47.