

# CYUT-III System at Chinese Grammatical Error Diagnosis Task

**Po-Lin Chen, Shih-Hung Wu\***  
Chaoyang University of Technology,  
Taichung, Taiwan, R.O.C  
streetcatsky@gmail.com

**Liang-Pu Chen, Ping-Che Yang**  
IDEAS, Institute for Information Industry,  
Taipei, Taiwan, ROC.  
{eit, maciac Clark}@iii.org.tw

\*Contact author: shwu@cyut.edu.tw

## Abstract

This paper describe the CYUT-III system on grammar error detection in the 2016 NLP-TEA Chinese Grammar Error Detection shared task CGED. In this task a system has to detect four types of errors, including redundant word error, missing word error, word selection error and word ordering error. Based on the conditional random fields (CRF) model, our system is a linear tagger that can detect the errors in learners' essays. Since the system performance depends on the features heavily, in this paper, we are going to report how to integrate the collocation feature into the CRF model. Our system presents the best detection accuracy and Identification accuracy on the TOCFL dataset, which is in traditional Chinese. The same system also works well on the simplified Chinese HSK dataset.

## 1 Introduction

Chinese essay writing is hard for foreign learners, not only on the aspect of learning pictograph Chinese characters but also on that of learning Chinese grammar that has no strong syntax rules. An automatic grammar error detection system might help the learners to get instant feedback when they are writing an essay in a computer aided language learning environment (Shiue and Chen, 2016).

In order to develop a grammar error detection system with the statistical natural language processing technology, developers need a large learner corpus (Chang et al., 2012). However, currently there is no publicly available large learner corpus in Chinese essay writing. That puts off the research in this field. The NLP-TEA workshop has been holding a Chinese Grammar Error Detection (CGED) shared task in the workshop for two years since 2014 (Yu et al., 2014) (Lee et al. 2015). They provides a set of learner corpus and a clear definition on 4 major Grammar error types in the foreign learner corpus. The shared tasks stimulated the research and drew many participants.

The goal of the shared task is to develop a system that can detect the four types of grammar errors in learner corpus. Comparing to the task definition of CGED in 2014 and 2015, the major difference in this year is the sentences might contain multiple errors. And the organizers provide two data sets: one is in traditional Chinese, the TOCFL dataset; the other is in simplified Chinese, the HSK dataset. Figure 1 and 2 are examples of the four error types, where redundant word is abbreviated 'R', missing word 'M', word selection error 'S', and word ordering error 'W'.

Based on the conditional random fields (CRF) model, we build a linear tagger that can detect the errors in learners' essays. The major improvement of our system is integrating the collocation feature into the CRF model. Since there is no publicly available Chinese collocation dataset, we will also report how we collect collocation.

The paper is organized as follows: Section 2 describes our methodology, section 3 shows our system architecture, section 4 is the discussion, and the final part is the conclusions.

### TOCFL (Traditional Chinese)

#### Example 1:

Input: (sid=A2-0007-2) 聽說妳打算開一個慶祝會。可惜我不能參加。因為那個時候我有別的事。當然我也要參加給你慶祝慶祝。

Output: A2-0007-2, 38, 39, R

(Note: "參加" is a redundant word)

#### Example 2:

Input: (sid=A2-0007-3) 我要送給你一個慶祝禮物。要是兩、三天晚了，請別生氣。

Output: A2-0007-3, 15, 20, W

(Note: "兩、三天晚了" should be "晚了兩、三天")

#### Example 3:

Input: (sid=A2-0011-1) 我聽到你找到工作。恭喜恭喜！

Output: A2-0011-1, 2, 3, S

A2-0011-1, 9, 9, M

(Notes: "聽到" should be "聽說". Besides, a word "了" is missing. The correct sentence should be "我聽說你找到工作了")

#### Example 4:

Input: (sid=A2-0011-3) 我覺得對你很抱歉。我也很想去，可是沒有辦法。

Output: A2-0011-3, correct

Figure 1. Examples of TOCFL (Traditional Chinese) from 2016 NLP-TEA CGED shared task [<http://nlpte2016.weebly.com/shared-task.html>]

### HSK (Simplified Chinese)

#### Example 1:

Input: (sid=00038800481) 我根本不能了解这妇女辞职回家的现象。在这个时代，为什么放弃自己的工作，就回家当家庭主妇？

Output: 00038800481, 6, 7, S

00038800481, 8, 8, R

(Notes: "了解" should be "理解". In addition, "这" is a redundant word.)

#### Example 2:

Input: (sid=00038800464) 我真不明白。她们可能是追求一些前代的浪漫。

Output: 00038800464, correct

#### Example 3:

Input: (sid=00038801261) 人战胜了饥饿，才努力为了下一代作更好的、更健康的東西。

Output: 00038801261, 9, 9, M

00038801261, 16, 16, S

(Notes: "能" is missing. The word "作" should be "做". The correct sentence is "才能努力为了下一代做更好的")

#### Example 4:

Input: (sid=00038801320) 饥饿的问题也是应该解决的。世界上每天由于饥饿很多人死亡。

Output: 00038801320, 19, 25, W

(Notes: "由于饥饿很多人" should be "很多人由于饥饿")

Figure 2. Examples of HSK (Simplified Chinese) from 2016 NLP-TEA CGED shared task [<http://nlpte2016.weebly.com/shared-task.html>]

## 2. Methodology

Our system is based on the conditional random field (CRF) (Lafferty et al., 2001). CRF model can cooperate with various kind of linguistic features. We believe that the word itself, its POS, and the appearance of collocation words or not are the major components. In our system, we use the template technology to generate 49 combinatorial features. The technology is briefly described in the following sub-sections.

### 2.1. Conditional Random Fields

CRF has been used in many natural language processing applications, such as named entity recognition, word segmentation, information extraction, and parsing. To perform different tasks, it requires different feature sets and labelled training data. The CRF can be regarded as a sequential labelling tagger. Given a sequence data  $X$ , the CRF can generate the corresponding label sequence  $Y$  based on the trained model. Each label  $Y$  is taken from a specific tag set, which needs to be defined in different tasks.  $X$  is a data sequence to be labelled, and output  $Y$  is a corresponding label sequence. While each label  $Y$  is taken from a tag set, how to define and interpret the label is a task-dependent work for the developers.

Mathematically, the model can be defined as:

$$P(Y|X) = \frac{1}{Z(X)} \exp(\sum_k \lambda_k f_k) \quad (1)$$

where  $Z(X)$  is the normalization factor,  $f_k$  is a set of features,  $\lambda_k$  is the corresponding weight. In this task,  $X$  is the input sentence, and  $Y$  is the corresponding error type label. As in the previously work, we define the tag set as: {O, R, M, S, D}, corresponding to no error, redundant, missing, selection, and word ordering respectively (Chen et al., 2015). Figure 3 shows a snapshot of our working file. The first column is the input sentence  $X$ , and the fourth column is the labelled tag sequence  $Y$ . The second column is the Part-of-speech (POS) of the word in the first column. The combination of words and the POSs will be the features in our system. The POS set used in our system is listed in Table 1, which is a simplified POS set provided by CKIP<sup>1</sup>.

Our system is built on the base of CRF++ (Kudo, 2007), a linear-chain CRF model software developed by Kudo<sup>2</sup>. In the training phase, a training sentence is first segmented into terms. Each term is labelled with the corresponding POS tag and error type tag. Then our system uses the CRF++ leaning algorithm to train a model. The features used in CRF++ can be expressed by templates. The format of each template is %X[row, col], where row is the number of rows in a sentence and column is the number of column as we shown in Figure 3. The feature templates used in our system are the combination of terms and POS of the input sentences. All the templates are listed in Table 2. An example on how a sentence is represented is given in Table 3. For example, the first feature template is “Term+POS”: if an input sentence contains the same term with the same POS, the feature value will be 1; otherwise the feature value will be 0. The second feature template is “Term+Previous Term”: if an input sentence contains the same term bi-gram, the feature value will be 1; otherwise the feature value will be 0.

Term	POS	collocation	Tag
一	DET	N	O
個	M	N	O
小時	N	N	O
以前	POST	Y	O
我	N	Y	O
決定	Vt	Y	O
休息	Vi	N	O

Figure 3 A Snapshot of a training sentence example in our system

<sup>1</sup> <http://ckipsvr.iis.sinica.edu.tw/>

<sup>2</sup> <http://crfpp.sourceforge.net/index.html>

Table 1. Simplified CKIP POS tags<sup>3</sup>

POS	
A	Adjective
C	Conjunction
POST	Postposition
ADV	Adverb
ASP	Tense marker
N	Noun and pronoun
DET	Article and Numeral
M	Chinese classifier
Nv	Nominalization
T	Chinese particles
P	Preposition
Vi	Intransitive verbs
Vt	Transitive verbs

## 2.2. Collocation

Collocation is useful lexicon knowledge for error correction in language learning (Ferraro et al., 2014). In his computational linguistic research papers, Smadja defined that collocations has four characteristics (Smadja, 1993). Firstly, collocations are arbitrary combinations of any lexicon, not syntactic or grammatical combinations. Secondly, collocations are domain depended, which means collocations are like terminology in one domain and it is hard to understand for outsider. Thirdly, collocations are recurrent, that means collocations are not exceptions, but rather often are repetitions in a given context. Lastly, collocations are cohesive lexical clusters, the presence of one word of a collocation often implies the rest of the collocation will appear in the context.

(Manning and Schütze, 1999) defined that a COLLOCATION is an expression consisting of two or more words that correspond to some conventional way of idea delivering. And there are three characteristics. The first is the non-compositionality, i.e. the meaning of the expression cannot be predicted from the meaning of the parts. The second is the non-substitutability, i.e. substitute near-synonyms for the components of a collocation will not be a collocation. The last is the Non-modifiability that is collocations cannot be freely modified with additional lexical material or through grammatical transformations.

## 2.3. The collection of Chinese collocation pairs

In the experiments, two methods are used to collect collocation pairs. The first is to select manually some collocation pairs from publicly available printed collocation dictionaries. We collect 80,040 collocation pairs (Chen et al., 2016). The second method is to use T-score to determine if the pair in a corpus is collocation or not.

We extract collocation from 874 correct sentences provided by NLP-TEA2. After word segmentation and POS tagging, our system focuses on content words, i.e. nouns, verbs, adverbs and adjectives only. Using the T-test technic, our system extracts 7,746 collocation pairs from all possible 10,581 pairs. The null hypothesis is: two terms appears independently, not a collocation pair.

The T-test formula is:

$$t = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}} \quad (2)$$

where  $\bar{x}$  is the sample mean,  $s^2$  is the sample variance,  $N$  is the sample size, and  $\mu$  is the mean of the distribution. If the t statistic is above a threshold, we can reject the null hypothesis. The null hypothesis here is that the two words are independent (Manning and H. Schütze, 1999).

<sup>3</sup> National Digital Archives Program, “CKIP POS,” <http://ckipsvr.iis.sinica.edu.tw/>

Table 2. Sample statistics of word frequency in the training set

Sample pairs	Term 1	# of term 1	Term 2	# of term 2	# of term 1 and 2 in one sentence
Pair 1	繼續(continue)	7	工作(work)	14	4
Pair 2	媽媽(Mother)	9	台灣(Taiwan)	26	1

For example, in our corpus with total  $N=4869$  terms, the frequency of the term “continue” is 7, the frequency for term ” work” is 14, and the frequency of ” continue work” is 4, then we can calculate the t-score as follows:

$$H_0: P(\text{continue work}) = P(\text{continue}) * P(\text{work}) = \frac{7}{4869} * \frac{14}{4869} = 4.1337 * 10^{-6} = \mu$$

Since  $4.1337 * 10^{-6}$  is near 0, thus  $s^2 = P(1 - P) \approx P$ . There are 4 times these two terms appear together in one sentence, therefore:

$$\bar{x} = \frac{4}{4869} \approx 8.21523 * 10^{-4}. \text{ Then we can get the T-score:}$$

$$\text{T-Score}(\text{continue work}) = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}} \approx \frac{8.21523 * 10^{-4} - 4.1337 * 10^{-6}}{\sqrt{\frac{8.21523 * 10^{-4}}{4869}}} \approx 1.98994.$$

This t value of 1.98994 is larger than 0.96817, the threshold we chose. So we can reject the null hypothesis that “continue work” occurs independently and it is a collocation.

For the second example, frequency of the term “Mother” is 9, the frequency for term ”Taiwan” is 26, and the frequency of ”Mother Taiwan” is 1, then we can calculate the t-score as follows:

$$\mu = P(\text{Mather Taiwan}) = P(\text{Mather}) * P(\text{Taiwan}) = \frac{9}{4869} * \frac{26}{4869} \approx 9.870435 * 10^{-6}$$

$$\text{And } \bar{x} = \frac{1}{4869} \approx 2.05380 * 10^{-4}$$

Again, according to Bernoulli trial, since  $\bar{x}$  is very small,  $s^2 \approx \bar{x}$ .

$$\text{T-Score}(\text{Mather Taiwan}) = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{N}}} \approx \frac{2.05380 * 10^{-4} - 9.870435 * 10^{-6}}{\sqrt{\frac{2.05380 * 10^{-4}}{4869}}} \approx 0.95194.$$

This t value of 0.95194 is not larger than 0.96817, the threshold we chose. So we cannot reject the null hypothesis that “Mother Taiwan” occurs independently and it is not a collocation.

Table 3. Templates and the corresponding value

Template	Corresponding Features
U01:%x[0,0]/%x[0,1]	Term+POS
U02:%x[0,0]/%x[-1,0]	Term+previous Term
U03:%x[0,0]/%x[-1,1]	Term+previous POS
U04:%x[0,1]/%x[-1,0]	POS+previous Term
U05:%x[0,1]/%x[-1,1]	POS+previous POS
U06:%x[0,0]/%x[-1,0]/%x[-1,1]	Term+previous Term+previous POS
U07:%x[0,1]/%x[-1,0]/%x[-1,1]	POS+previous Term+previous POS
U08:%x[0,0]/%x[-2,0]	Term+previous previous Term
U09:%x[0,0]/%x[-2,1]	Term+previous previous POS
U010:%x[0,1]/%x[-2,0]	POS+previous previous Term
U011:%x[0,1]/%x[-2,1]	POS+previous previous POS
U012:%x[0,0]/%x[-2,0]/%x[-2,1]	Term+previous previous Term+previous previous POS
U013:%x[0,1]/%x[-2,0]/%x[-2,1]	POS+previous previous Term+previous previous POS
U014:%x[0,0]/%x[1,0]	Term+next Term
U015:%x[0,0]/%x[1,1]	Term+next POS
U016:%x[0,1]/%x[1,0]	POS+next Term
U017:%x[0,1]/%x[1,1]	POS+next POS

U018:%x[0,0]/%x[1,0]/%x[1,1]	Term+next Term+next POS
U019:%x[0,1]/%x[1,0]/%x[1,1]	POS+next Term+next POS
U020:%x[0,0]/%x[2,0]	Term+next next Term
U021:%x[0,0]/%x[2,1]	Term+next next POS
U022:%x[0,1]/%x[2,0]	POS+next next Term
U023:%x[0,1]/%x[2,1]	POS+next next POS
U024:%x[0,0]/%x[2,0]/%x[2,1]	Term+next next Term+next next POS
U025:%x[0,1]/%x[2,0]/%x[2,1]	POS+next next Term+next next POS
U026:%x[0,0]/%x[0,2]	Term+C
U027:%x[0,0]/%x[-1,2]	Term+previous C
U028:%x[0,2]/%x[-1,2]	C+previous C
U029:%x[0,0]/%x[-1,0]/%x[-1,1]/%x[-1,2]	Term+previous Term+previous POS+previous C
U030:%x[0,1]/%x[-1,0]/%x[-1,1]/%x[-1,2]	POS+previous Term+previous POS+previous C
U031:%x[0,0]/%x[-2,2]	Term+previous previous C
U032:%x[0,1]/%x[-2,2]	POS+previous previous C
U033:%x[0,0]/%x[1,2]	Term+next C
U034:%x[0,1]/%x[1,2]	POS+next C
U036:%x[0,1]/%x[2,2]	POS+next next C
U037:%x[0,0]/%x[-2,0]/%x[-2,1]/%x[-2,2]	Term+previous previous Term+previous previous POS+previous previous C
U038:%x[0,1]/%x[-2,0]/%x[-2,1]/%x[-2,2]	POS+previous previous Term+previous previous POS+previous previous C
U039:%x[0,0]/%x[1,0]/%x[1,1]/%x[1,2]	Term+next Term+next POS+next C
U040:%x[0,1]/%x[1,0]/%x[1,1]/%x[1,2]	POS+next Term+next POS+next C
U041:%x[0,0]/%x[2,0]/%x[2,1]/%x[2,2]	Term+next next Term+next next POS+next next C
U042:%x[0,1]/%x[2,0]/%x[2,1]/%x[2,2]	POS+next next Term+next next POS+next next C
U043:%x[-1,1]/%x[0,1]/%x[1,1]	previous POS+POS+next POS
U044:%x[-1,0]/%x[0,0]/%x[1,0]	previous Term+Term+next Term
U045:%x[-1,0]/%x[0,0]/%x[1,1]	previous Term+Term+next POS
U046:%x[-1,0]/%x[0,1]/%x[1,2]	previous Term+POS+next POS
U047:%x[-1,1]/%x[0,0]/%x[1,0]	previous POS+Term+next Term
U048:%x[-1,1]/%x[0,0]/%x[1,1]	previous POS+Term+next POS
U049:%x[-1,1]/%x[0,1]/%x[1,0]	previous POS+POS+next Term
U050:%x[-1,0]/%x[0,1]/%x[1,1]	previous Term+POS+next Term

### 3. System architecture

Our system flowchart is shown in Figure 4. The training phrase consists of two steps: 1. Collecting collocation. 2. Training the CRF with the help of collocation detection, word segmentation and POS tagging results. In the first training phrase, a large Chinese corpus is used as the training set. After the word segmentation and POS tagging, the corpus is used to collect collocations as we described in section 2.2. In the second training phrase, the collocations appear in the same sentence or not is used as one separate feature for CRF tagger training.

The test phrase is straightforward. The test sentence is first segmented into words with POS tag, after detecting the appearance of collocation terms or not, the sentence is prepared as the input of CRF model. The CRF model will give one output tag to each term. The tag indicate error detection, error type, and also error position at the same time.

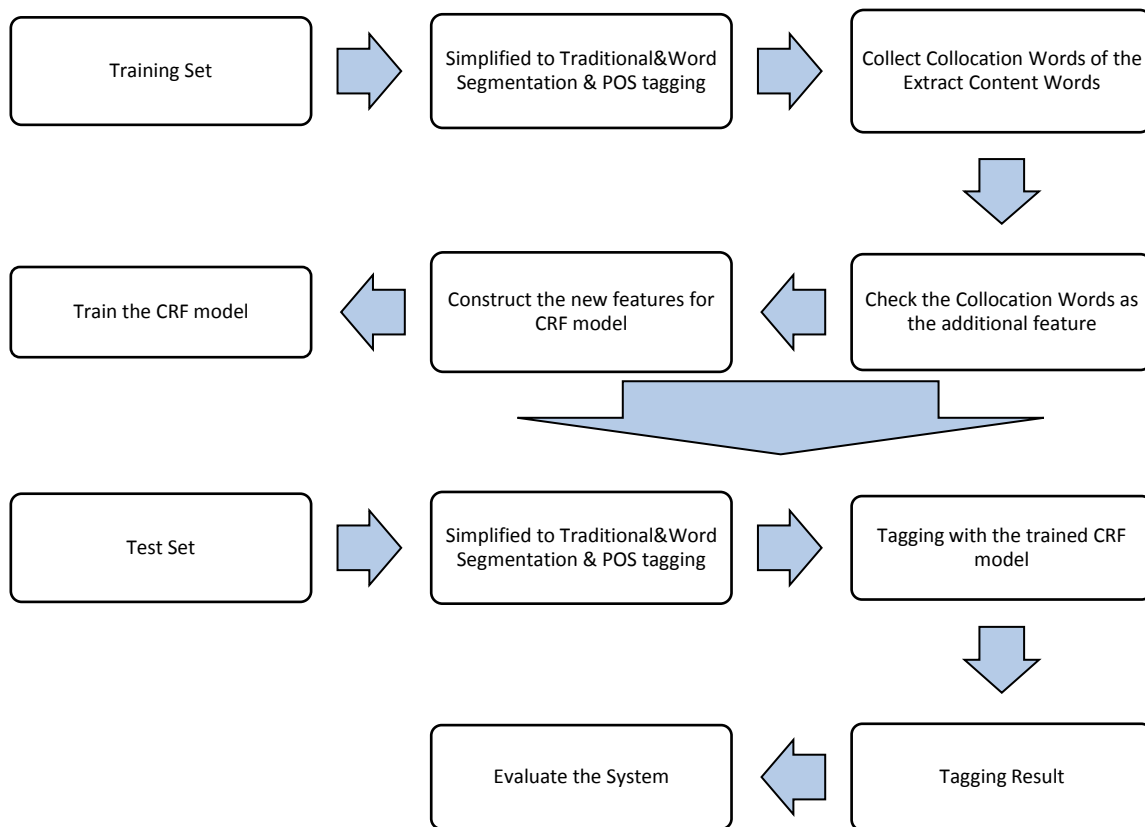


Figure 4. System Flowchart

## 4. Experiments

The system evaluation metrics of CGED shared task includes three levels. We focus on the identification level: this level is a multi-class categorization problem. All error types should be identified, i.e., Redundant, Missing, Word ordering, and Selection. The metrics used are accuracy, precision, recall, and F1-score.

### 4.1. Experiment Settings

We send respectively three runs for both data set this year, and the major difference for each experiment settings is the size of training set. Our system is based on traditional Chinese processing, the simplified Chinese is translated into traditional Chinese by Microsoft Word in advance. Our training data consists of data from NLP-TEA1 (Chang et al., 2012) Training Data, Test Data, and the Training Data from NLP-TEA2 and NLP-TEA3. Table 4 shows the number of sentences in our training set.

Run1 settings: Use all the available data as the training set. For TOCFL, the training set is the union of the training sets in the NLP-TEA1, NLP-TEA2, and TOCFL in NLP-TEA3. For HSK, the training set is the union of the one used for TOCFL and HSK in NLP-TEA3.

Run2 settings: Almost the same as those in Run1, the only difference is the correct sentences are excluded from the training set. We believe that they provide no help for finding errors.

Run3 settings: Almost the same as those in Run1. The difference is how our system treats the continuous errors. If two errors of the same type occurred continuously, our system will combine them as one longer error. For example, two errors of the same type:

A2-0019-1, 10, 12, S

A2-0019-1, 13, 13, S

will be reported as:

A2-0019-1, 10, 13, S.

Table 4. Training set size

size	NLP-TEA1	NLP-TEA2	NLP-TEA3
Redundant	1830	434	10010
Correct	874	0	0
Selection	827	849	20846
Disorder (word ordering)	724	306	3071
Missing	225	622	15701

## 4.2. Experimental results

In the formal run of NLP-TEA-3 CGED shared task, there are 5 participants, and each team submits 3 runs in TOCFL, totally 15 runs. There are 8 participants in HSK, totally 21 runs. Table 5 shows the false positive rate. Our system has 0.082 false positive rate. The average of all runs is calculated from 15 runs for TOCFL and 21 runs for HSK.

Table 6, Table 7, and Table 8 show the formal run result of our system compared with the average in Detection level, Identification level, and Position level respectively. Our system achieves the highest Accuracy in Detection Level(TOCFL) and Identification-Level (TOCFL). The numbers in boldface are the best performance among all formal runs.

Table 5: The false positive rate in Detection Level (the lower the better)

Submission	False Positive Rate (TOCFL)	False Positive Rate (HSK)
CYUT&III-Run1	0.3470	0.4016
CYUT&III-Run2	0.3558	0.4191
CYUT&III-Run3	0.3635	0.4016
Average of all runs	0.4812	0.4956

Table 6: Performance evaluation in Detection Level

	Detection Level(TOCFL)				Detection Level(HSK)			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Run1	<b>0.5955</b>	0.6259	0.5419	0.5809	0.6141	0.6003	0.6304	0.615
Run2	<b>0.5955</b>	0.6236	0.5501	0.5846	0.6118	0.5951	0.644	0.6186
Run3	0.5941	0.6205	0.5545	0.5856	0.6141	0.6003	0.6304	0.615
Average of all formal runs	0.5442	0.5700	0.5679	0.5455	0.5627	0.5807	0.6237	0.5688

Table 7: Performance evaluation in Identification Level

	Identification-Level (TOCFL)				Identification-Level (HSK)			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
CYUT-Run1	<b>0.5154</b>	0.46	0.3021	0.3647	0.5714	0.5306	0.4376	0.4797
CYUT-Run2	0.5133	0.4567	0.3061	<b>0.3666</b>	0.5662	0.5238	0.4509	0.4846
CYUT-Run3	0.5078	0.4472	0.3001	0.3592	0.5715	0.5306	0.4352	0.4782
Average of all formal runs	0.39118	0.32647	0.27321	0.2716	0.4555	0.4310	0.3705	0.3720



Table 8: Performance evaluation in Position Level.

	Position-Level (TOCFL)				Position-Level (HSK)			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
CYUT-Run1	0.3113	0.1461	0.1089	<b>0.1248</b>	0.3202	0.2037	0.2138	0.2086
CYUT-Run2	0.3061	0.1432	<b>0.1092</b>	0.1239	0.3143	0.2034	0.2225	0.2125
CYUT-Run3	0.3088	0.1196	0.0768	0.0935	0.3304	0.1814	0.144	0.1605
Average of all formal runs	0.2402	0.0846	0.0459	0.0597	0.2892	0.2059	0.1366	0.1529

## 5. Discussion and Conclusions

This paper reports our approach to the NLP-TEA-3 CGED Shared Task evaluation. By integrating the collocation as an additional feature into CRF model, we build a system that can achieve the task. The approach uniformly deals with the four error types: Redundant, Missing, Selection, and Word ordering.

Our system presents the best accuracy in detection level, best accuracy and F1 in identification level, and best recall and F1 in position-level at the TOCFL official run.

Due to the limitation of time and resource, our system is not tested under different experimental settings. In the future, we will use a larger corpus to extract more collocations to improve the performance on error diagnosis.

## Acknowledgments

This study is conducted under the "Online and Offline integrated Smart Commerce Platform (3/4)" of the Institute for Information Industry which is subsidized by the Ministry of Economic Affairs of the Republic of China.

## References

- 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*, 11(1), article 3, March.
- Tao-Hsing Chang, Yao-Ting Sung, Jia-Fei Hong, Jen-I chang. 2014. KNGED: a Tool for Grammatical Error Diagnosis of Chinese Sentences.
- Po-Lin Chen, Wu Shih-Hung, Liang-Pu Chen, Ping-Che Yang, Ren-Dar Yang. 2015. Chinese Grammatical Error Diagnosis by Conditional Random Fields, in *Proceedings of The 2nd Workshop on Natural Language Processing Techniques for Educational Applications*, pages 7–14, Beijing, China, July.
- Po-Lin Chen, Wu Shih-Hung, Liang-Pu Chen, Ping-Che Yang. 2016. Improving the Selection Error Recognition in a Chinese Grammar Error Detection System, *IEEE IRI 2016*, Pittsburgh, Pennsylvania, USA, July.
- Gabriela Ferraro, Rogelio Nazar, Margarita Alonso Ramos, and Leo Wanner. 2014. Towards advanced collocation error correction in Spanish learner corpora. *Lang. Resour. Eval.* 48, 1, pp. 45-64.
- Michael Gamon. 2011. High-Order Sequence Modeling for Language Learner Error Detection, in *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 180–189, Portland, Oregon, June.
- Lafferty, A. McCallum, and F. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Intl. Conf. on Machine Learning*.
- Lee, Lung-Hao, Liang-Chih Yu, and Li-Ping Chang. 2015. Overview of the NLP-TEA 2015 shared task for Chinese grammatical error diagnosis. In *Proceedings of the 2nd Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA 2015)*. 1-6.
- Manning, C. D. and H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*, The MIT Press.
- National Digital Archives Program, "CKIP POS," <http://ckipsvr.iis.sinica.edu.tw/>, 2015.
- F. Smadja. 1993. "Retrieving Collocation from Text: Xtract," *Computational Linguistics*, Vol. 19, No. 1, pp. 143-177.
- Yow-Ting Shiue and Hsin-Hsi Chen. 2016. "Detecting Word Usage Errors in Chinese Sentences for Learning Chinese as a Foreign Language." *Proceedings of 10th Language Resources and Evaluation Conference*, 23-28 May 2016, Portorož, Slovenia.
- Taku Kudo. 2007. "CRF++: Yet Another CRF toolkit", <https://taku910.github.io/crfpp/>.

- Jui-Feng Yeh, Yun-Yun Lu, Chen-Hsien Lee, Yu-Hsiang Yu, Yong-Ting Chen. 2014. Detecting Grammatical Error in Chinese Sentence for Foreign.
- Yu, Liang-Chih, Lung-Hao Lee, and Li-Ping Chang. 2014. Overview of grammatical error diagnosis for learning Chinese as a foreign language. In Proceedings of the 1st Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA 2014). 42-47.
- 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, 18(6), 1170-1181.
- Shih-Hung Wu, Hsien-You Hsieh. 2012. Sentence Parsing with Double Sequential Labeling in Traditional Chinese Parsing Task. Second CIPS-SIGHAN Joint Conference on Chinese Language Processing, pages 222–230.