

# NLP for Chinese L2 Writing: Evaluation of Chinese Grammatical Error Diagnosis

Gaoqi RAO (饶高琦)

Beijing Language and Culture University

8.May.2018 @ LREC



# Introduction

- ▶ With increasing influence of China, L2 Chinese teaching has approached in width and depth, attracting increasing NLPers attention.
- ▶ Chinese grammatical error diagnosis (CGED) is one of the most important evaluation to in this field to provide a common setting so that researchers who approach the tasks using different linguistic factors and computational techniques can compare their results.



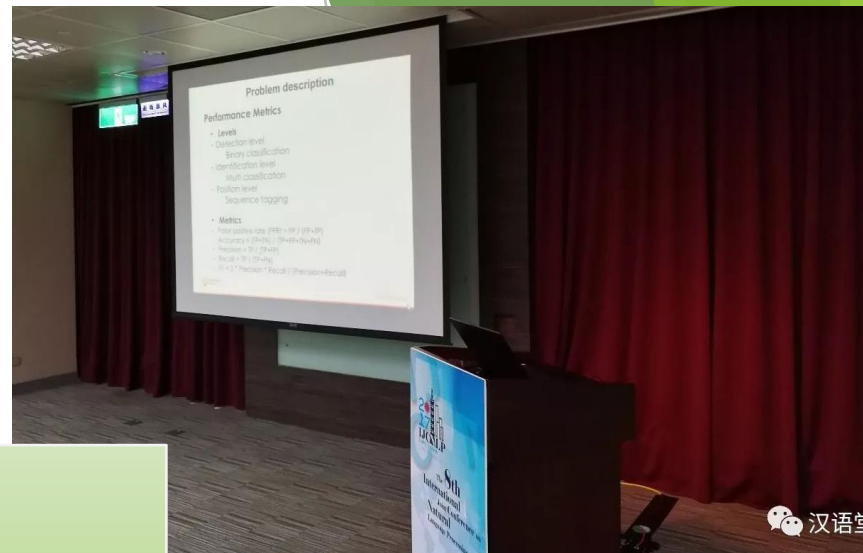
# Introduction

- ▶ Chinese is an “isolated-like” language in Sino-Tibetan language family. Grammar and lexicon made huge trouble for L2 learners from mother language of other topology or language families
- ▶ Therefore we aim grammar a center point to push the R&D for computer assistant learning



学中文  
[xuézhōngwén]  
Learn Chinese





CGED-1@ICCE2014  
CGED-2@ACL2015  
CGED-3@COLING2016  
CGED-4@IJCNLP2017



学中文  
[xuézhōng  
Learn Chinese



# Task Description

- ▶ The developed tool is expected to identify the error types and its position at which it occurs in the sentence
- ▶ Four **PADS** error types are included in the target modification taxonomy, to decrease the data sparsity caused by multi error types

redundant string, R (**A**ddition) “我爱逛饶河夜市过”

missing string, M (**D**eletion) “我爱逛饶河夜┆”

selection error, S(**S**ubstitution) “我爱走饶河夜市”

word ordering error, W(**P**ermutation) “我爱饶河夜市逛”

The input sentences may contain at least one of defined error types

Errors are described in offset of chars, instead of words

One track in CGED2014、CGED2015；TOCFL (Test Of Chinese as a Foreign Language)

Two tracks in CGED2016；TOCFL and HSK

One track in CGED2017、CGED2018；HSK

# Task Description

TOCFL (Traditional Chinese) ◦	HSK (Standard Chinese) ◦
<ul style="list-style-type: none"><li>● 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 ◦ (Notes: “參加” is a redundant word) ◦</li><li>● Example 2 ◦ 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”. ◦</li><li>● Example 3 ◦ 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 ◦</li></ul>	<ul style="list-style-type: none"><li>● 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.) ◦</li><li>● 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 ◦</li><li>● 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”) ◦</li></ul>

# Task Description

<DOC> ↵

<TEXT id="A2-0005-1"> ↵

我聽說你打算開一個慶祝會。對不起，我要參加，可是沒有空。你開一個慶祝會的時候我不能會參加，是因為我在外國  
做工作。 ↵

</TEXT> ↵

<CORRECTION> ↵

我聽說你打算開一個慶祝會。對不起，我要參加，可是沒有空。你開慶祝會的時候我不能參加，是因為我在外國工作。 ↵

</CORRECTION> ↵

<ERROR start\_off=" 31" end\_off=" 32" type="R"></ERROR> ↵

<ERROR start\_off=" 42" end\_off=" 42" type="R"></ERROR> ↵

<ERROR start\_off=" 53" end\_off=" 53" type="R"></ERROR> ↵

</DOC> ↵

# 任务描述

<DOC> 。

<TEXT id="A2-0005-1"> 。

我聽說你打算開一個慶祝會。對不起，我要參加，可是沒有空。你開一個慶祝會的時候我不能會參加，是因為我在外國做工作。 。

</TEXT> 。

<CORRECTION> 。

我聽說你打算開一個慶祝會。對不起，我要參加，可是沒有空。你開慶祝會的時候我不能參加，是因為我在外國工作。 。

</CORRECTION> 。

<ERROR start\_off="31" end\_off="32" type="R"></ERROR> 。

<ERROR start\_off="42" end\_off="42" type="R"></ERROR> 。

<ERROR start\_off="53" end\_off="53" type="R"></ERROR> 。

</DOC> 。

<DOC> 。

<TEXT id="200210543634250003\_2\_1x3"> 。

对于“安乐死”的看法，向来都是一个极具争议性的题目，因为毕竟每个人对于死亡的观念都不一样，怎样的情况下去判断，也自然产生出很多主观和客观的理论。每个人都有着生存的权利，也代表着每个人都能去决定如何结束自己的生命的权利。在我的个人观点中，如果一个长期受着病魔折磨的人，会是十分痛苦的事，不仅是病人本身，以致病者的家人和朋友，都是一件难受的事。 。

</TEXT> 。

<CORRECTION> 。

对于“安乐死”的看法，向来都是一个极具争议性的题目，因为毕竟每个人对于死亡的观念都不一样，无论在怎样的情况下去判断，都自然产生出很多主观和客观的理论。每个人都有着生存的权利，也代表着每个人都能去决定如何结束自己的生命。在我的个人观点中，如果一个长期受着病魔折磨的人活着，会是十分痛苦的事，不仅是病人本身，对于病者的家人和朋友，都是一件难受的事。 。

</CORRECTION> 。

<ERROR start\_off="46" end\_off="46" type="M"></ERROR> 。

<ERROR start\_off="56" end\_off="56" type="S"></ERROR> 。

<ERROR start\_off="106" end\_off="108" type="R"></ERROR> 。

<ERROR start\_off="133" end\_off="133" type="M"></ERROR> 。

<ERROR start\_off="151" end\_off="152" type="S"></ERROR> 。

</DOC> 。



# Data Sets

## Training Sets

Evaluation	Track	Unit	Errors ↴	R	M	S	W
CGED2016 ↴	TOCFL ↴	10,693 ↴	24,492 ↴ (100%) ↴	4,472 ↴ (18.3%) ↴	8,739 ↴ (35.7%) ↴	9,897 ↴ (40.4%) ↴	1,384 ↴ (5.7%) ↴
	HSK ↴	10,071 ↴	24,797 ↴ (100%) ↴	5,538 ↴ (22.3%) ↴	6,623 ↴ (26.7%) ↴	10949 ↴ (44.2%) ↴	1,687 ↴ (6.8%) ↴
CGED2017 ↴	HSK ↴	10,449 ↴	26,448 ↴ (100%) ↴	5,852 ↴ (22.1%) ↴	7,010 ↴ (26.5%) ↴	11,591 ↴ (43.8%) ↴	1,995 ↴ (7.5%) ↴

# Data Sets

## Training Sets

Evaluation	Track	Unit	Errors ↴	R	M	S	W
CGED2016 ↴	TOCFL ↴	10,693 ↴	24,492 ↴ (100%) ↴	4,472 ↴ (18.3%) ↴	8,739 ↴ (35.7%) ↴	9,897 ↴ (40.4%) ↴	1,384 ↴ (5.7%) ↴
	HSK ↴	10,071 ↴	24,797 ↴ (100%) ↴	5,538 ↴ (22.3%) ↴	6,623 ↴ (26.7%) ↴	10949 ↴ (44.2%) ↴	1,687 ↴ (6.8%) ↴
CGED2017 ↴	HSK ↴	10,449 ↴	26,448 ↴ (100%) ↴	5,852 ↴ (22.1%) ↴	7,010 ↴ (26.5%) ↴	11,591 ↴ (43.8%) ↴	1,995 ↴ (7.5%) ↴

# Data Sets

## Testing Sets

Evaluation	Track	Unit	Correct Unit	Erroneous Unit	Errors	R	M	S	W
CGED2016	TOCFL	3,528	1,703 (48.3%)	1,825 (51.7%)	4,103 (100%)	782 (19.06%)	1,482 (36.12%)	1,613 (39.31%)	226 (5.51%)
	HSK	3,011	1,539 (51.1%)	1,472 (48.9%)	3,695 (100%)	802 (21.71%)	991 (26.82%)	1620 (43.84%)	282 (7.63%)
CGED2017	HSK	3,154	1,173 (48.4%)	1,628 (51.6%)	4,876 (100%)	1,062 (21.78%)	1,274 (26.13%)	2,155 (44.20%)	385 (7.90%)

# Data Sets

## Testing Sets

Evaluation	Track	Unit	Correct Unit	Erroneous Unit	Errors	R	M	S	W
CGED2016	TOCFL	3,528	1,703 (48.3%)	1,825 (51.7%)	4,103 (100%)	782 (19.06%)	1,482 (36.12%)	1,613 (39.31%)	226 (5.51%)
	HSK	3,011	1,539 (51.1%)	1,472 (48.9%)	3,695 (100%)	802 (21.71%)	991 (26.82%)	1620 (43.84%)	282 (7.63%)
CGED2017	HSK	3,154	1,173 (48.4%)	1,628 (51.6%)	4,876 (100%)	1,062 (21.78%)	1,274 (26.13%)	2,155 (44.20%)	385 (7.90%)

# Performance Metrics

confusion matrix:

$$\text{False Positive Rate} = \text{FP} / (\text{FP} + \text{TN})$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1} = 2 * \text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})$$

Confusion Matrix		System Results	
		Positive (Erroneous)	Negative (Correct)
Gold Standard	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)



# Performance Metrics

- ▶ The criteria for judging correctness are determined at 4 levels as follows:

**Detective-level:** Binary classification of a given unit, that is, correct or incorrect

**Identification-level:** multi-class categorization of error types

**Position-level:** In addition to identifying the error types, this level also judges the occurrence range of the grammatical error

**Correction-level:** In the CGED2018, the participant systems are required to offer 0 to 3 recommended corrections to error types of missing and selection. The amount of the correction to recommend depends on the trust computation at each error. More recommendation would increase the recall, but somehow reduce precision, since the gold standard only offers one correction to each error.

- ▶ **Difficulty:** Detective < Identification < Position

# Submission & Results ( CGED2016 )

<b>Participant (Ordered by abbreviations of names)</b>	<b>#TRuns</b>	<b>#HRuns</b>
NLP Lab, Zhengzhou University ( <b>ANO</b> )	0	2
Beijing Foreign Studies University ( <b>BFSU-TZT</b> )	0	0
Beijing Information Science and Technology University ( <b>BISTU</b> )	0	0
Central China Normal University ( <b>CCNU</b> )	0	1
Chaoyang University of Technology ( <b>CYUT</b> )	3	3
Harbin Institute of Technology ( <b>HIT</b> )	0	3
Institute of Computational Linguistics, Peking University ( <b>PKU</b> )	3	3
Saarland University & Harvard Medical School ( <b>MAZA</b> )	0	0
National Chiao Tung University & National Taipei University of Technology ( <b>NCTU+NTUT</b> )	3	0
National Chiayi University ( <b>NCYU</b> )	3	3
National Taiwan Ocean University ( <b>NTOU</b> )	0	0
NLP Lab, Zhengzhou University ( <b>SKY</b> )	0	3
Beijing Sogou Inc. ( <b>Sogou</b> )	0	0
Dublin City University & National Taiwan University ( <b>TWIRL</b> )	0	0
School of Information Science and Engineering, Yunnan University ( <b>YUN-HPCC</b> )	3	3

15 Participants, 9 made submissions

# CGED2016·TOCFL

TOCFL Submission	FPR	Detection-level				Identification-level				Position-level			
		Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1
CYUT-Run1	0.3470	<b>0.5955</b>	0.6259	0.5419	0.5809	<b>0.5154</b>	0.4600	0.3021	0.3647	0.3113	0.1461	0.1089	<b>0.1248</b>
CYUT-Run2	0.3558	<b>0.5955</b>	0.6236	0.5501	0.5846	0.5133	0.4567	0.3061	<b>0.3666</b>	0.3061	0.1432	<b>0.1092</b>	0.1239
CYUT-Run3	0.3635	0.5941	0.6205	0.5545	0.5856	0.5078	0.4472	0.3001	0.3592	0.3088	0.1196	0.0768	0.0935
NCTU+NTUT-Run1	<b>0.1362</b>	0.5442	<b>0.6593</b>	0.2460	0.3583	0.5110	<b>0.4892</b>	0.1224	0.1958	<b>0.4603</b>	<b>0.2542</b>	0.0483	0.0811
NCTU+NTUT-Run2	0.2913	0.5530	0.6000	0.4077	0.4855	0.4793	0.4036	0.1982	0.2659	0.3784	0.1644	0.0639	0.0920
NCTU+NTUT-Run3	0.3200	0.5612	0.6013	0.4504	0.5150	0.4773	0.3993	0.2185	0.2824	0.3613	0.1521	0.0668	0.0928
NCYU-Run1	0.5602	0.5507	0.5559	0.6542	0.6011	0.3577	0.2749	0.2862	0.2805	0.1728	0.0074	0.0056	0.0064
NCYU-Run2	0.9612	0.5218	0.5202	<b>0.9726</b>	<b>0.6779</b>	0.2328	0.2265	<b>0.4744</b>	0.3066	0.0231	0.0129	0.0195	0.0155
NCYU-Run3	0.8491	0.5363	0.5307	0.8959	0.6665	0.2653	0.2384	0.4134	0.3024	0.0580	0.0130	0.0163	0.0145
PKU-Run1	0.2284	0.5210	0.5739	0.2871	0.3828	0.4575	0.3418	0.1173	0.1747	0.3844	0.0996	0.0263	0.0416
PKU-Run2	0.7205	0.5258	0.5292	0.7556	0.6224	0.3242	0.2792	0.3712	0.3187	0.1381	0.0680	0.0824	0.0745
PKU-Run3	0.5250	0.5349	0.5467	0.5907	0.5678	0.3705	0.2729	0.2192	0.2431	0.2331	0.0872	0.0651	0.0745
YUN-HPCC-Run1	0.6289	0.5420	0.5444	0.7014	0.6130	0.2211	0.1588	0.3196	0.2122	0.0886	0.0002	0.0002	0.0002
YUN-HPCC-Run2	0.5931	0.5026	0.5167	0.5918	0.5517	0.2322	0.1675	0.3136	0.2184	0.0991	0	0	null
YUN-HPCC-Run3	0.3382	0.4847	0.5030	0.3195	0.3908	0.4023	0.2810	0.1359	0.1832	0.2797	0.0012	0.0005	0.0007

# CGED2016·HSK

HSK Submission	FPR	Detection-level				Identification-level				Position-level			
		Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1	Acc.	Pre.	Rec.	F1
ANO-Run1	0.5601	0.5473	0.5297	0.6596	0.5876	0.4723	0.4244	0.4292	0.4268	0.3687	0.2910	0.2460	0.2666
ANO-Run2	0.6517	0.4779	0.4738	0.6135	0.5346	0.2977	0.2243	0.2535	0.2380	0.1157	0.0046	0.0046	0.0046
*CCNU-Run1	0.3294	0.4988	0.4811	0.3193	0.3838	0.4012	0.2425	0.1324	0.1713	0.2806	0.0187	0.0089	0.0121
CYUT-Run1	0.4016	0.6141	0.6003	0.6304	0.6150	0.5714	0.5306	0.4376	0.4797	0.3202	0.2037	0.2138	0.2086
CYUT-Run2	0.4191	0.6118	0.5951	0.6440	0.6186	0.5662	0.5238	0.4509	0.4846	0.3143	0.2034	0.2225	0.2125
CYUT-Run3	0.4016	0.6141	0.6003	0.6304	0.6150	0.5715	0.5306	0.4352	0.4782	0.3304	0.1814	0.1440	0.1605
HIT-Run1	0.4334	0.6377	0.6111	0.7120	0.6577	0.5683	0.5146	0.5219	0.5182	0.4781	0.4034	0.3691	<b>0.3855</b>
HIT-Run2	0.4327	0.6370	0.6108	0.7099	0.6566	0.5744	0.5224	0.5094	0.5158	0.4756	0.3970	0.3483	0.3711
HIT-Run3	0.4516	0.6370	0.6071	0.7296	<b>0.6628</b>	0.5565	0.5002	<b>0.5447</b>	<b>0.5215</b>	0.4475	0.3695	<b>0.3697</b>	0.3696
NCYU-Run1	0.2820	0.5526	0.5629	0.3798	0.4535	0.4554	0.3259	0.1877	0.2382	0.3301	0.0244	0.0095	0.0136
NCYU-Run2	0.9467	0.5042	0.4964	<b>0.9755</b>	0.6580	0.2687	0.2588	0.5263	0.3470	0.0312	0.0158	0.0217	0.0183
NCYU-Run3	0.9818	0.4846	0.4864	0.9721	0.6484	0.2227	0.2195	0.3578	0.2721	0.0148	0.0081	0.0089	0.0085
PKU-Run1	0.7706	0.4972	0.4910	0.7772	0.6018	0.3104	0.2717	0.3991	0.3233	0.1106	0.0523	0.0674	0.0589
PKU-Run2	0.8070	0.5022	0.4945	0.8254	0.6185	0.3144	0.2765	0.3594	0.3125	0.1016	0.0595	0.0923	0.0724
PKU-Run3	0.8213	0.5058	0.4968	0.8478	0.6265	0.3062	0.2694	0.3586	0.3076	0.0896	0.0520	0.0863	0.0649
SKY-Run1	0.0695	0.6523	0.8326	0.3614	0.5040	0.6605	0.8235	0.2732	0.4102	0.6073	0.6153	0.1783	0.2765
SKY-Run2	<b>0.0481</b>	0.6579	<b>0.8746</b>	0.3505	0.5005	0.6765	<b>0.8821</b>	0.2972	0.4446	0.6376	0.7054	0.2217	0.3373
SKY-Run3	0.0559	<b>0.6659</b>	0.8652	0.3750	0.5232	<b>0.6849</b>	0.8744	0.3185	0.4669	<b>0.6477</b>	<b>0.7144</b>	0.2430	0.3627
YUN-HPCC-Run1	0.5608	0.5191	0.5069	0.6026	0.5506	0.3485	0.2800	0.3879	0.3252	0.0654	0.0024	0.0062	0.0035
YUN-HPCC-Run2	0.7122	0.4949	0.4886	0.7113	0.5793	0.3092	0.2681	0.4565	0.3378	0.0373	0.0022	0.0070	0.0034
YUN-HPCC-Run3	0.2710	0.5058	0.4902	0.2724	0.3502	0.4306	0.2886	0.1448	0.1928	0.2701	0.0010	0.0005	0.0007

# CGED2017·HSK

TEAM	RUNs	False Positive Rate	Detection Level				Identification Level				Position Level			
			Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
	run1	0.6172 (724/1173)	0.6439	0.686	0.7986	0.738	0.488	0.4791	0.5657	<b>0.5188</b>	0.2547	0.2169	0.2752	0.2426
ALI_NLP	run2	0.6607 (775/1173)	<b>0.6465</b>	0.6792	<b>0.8284</b>	0.7464	0.4654	0.453	0.6006	0.5164	0.2264	0.1949	<b>0.2941</b>	0.2344
	run3	0.3052 (358/1173)	0.6173	0.7597	0.5714	0.6523	<b>0.5513</b>	0.6007	0.3756	0.4622	<b>0.4121</b>	<b>0.3663</b>	0.213	<b>0.2693</b>
	run1	<b>0.098 (115/1173)</b>	0.4721	<b>0.7894</b>	0.2176	0.3411	0.4337	0.5474	0.106	0.1776	0.3775	0.2773	0.0418	0.0727
BNU	run2	0.1355 (159/1173)	0.4794	0.758	0.2514	0.3776	0.4412	0.5527	0.131	0.2118	0.3735	0.2818	0.0515	0.0871
	run3	0.1893 (222/1173)	0.5181	0.7547	0.3448	0.4733	0.4696	0.5707	0.1786	0.2721	0.3798	0.2968	0.0715	0.1152
CVTE	run1	0.1441 (169/1173)	0.4756	0.7459	0.2504	0.3749	0.4461	<b>0.606</b>	0.1214	0.2023	0.3314	0.118	0.0204	0.0348
	run2	0.3154 (370/1173)	0.539	0.708	0.4528	0.5523	0.4711	0.5391	0.2057	0.2978	0.2602	0.1093	0.0465	0.0653
NTOUA	run1	1 (1173/1173)	0.6281	0.6281	1	<b>0.7716</b>	0.3211	0.3211	<b>0.6099</b>	0.4207	0.0212	0.0212	0.0958	0.0348
	run2	1 (1173/1173)	0.6281	0.6281	1	<b>0.7716</b>	0.3889	0.3889	0.506	0.4398	0.018	0.018	0.082	0.0295
YNU-HPCC	run1	0.6513 (764/1173)	0.5796	0.65	0.7163	0.6816	0.4218	0.4219	0.4217	0.4218	0.1778	0.1262	0.1191	0.1225
	run2	0.7383 (866/1173)	0.5891	0.6417	0.7829	0.7053	0.3879	0.3825	0.4575	0.4167	0.1426	0.1056	0.1191	0.112
	run3	0.6104 (716/1173)	0.5311	0.6298	0.6148	0.6222	0.3979	0.4086	0.3298	0.365	0.1702	0.0981	0.0698	0.0816

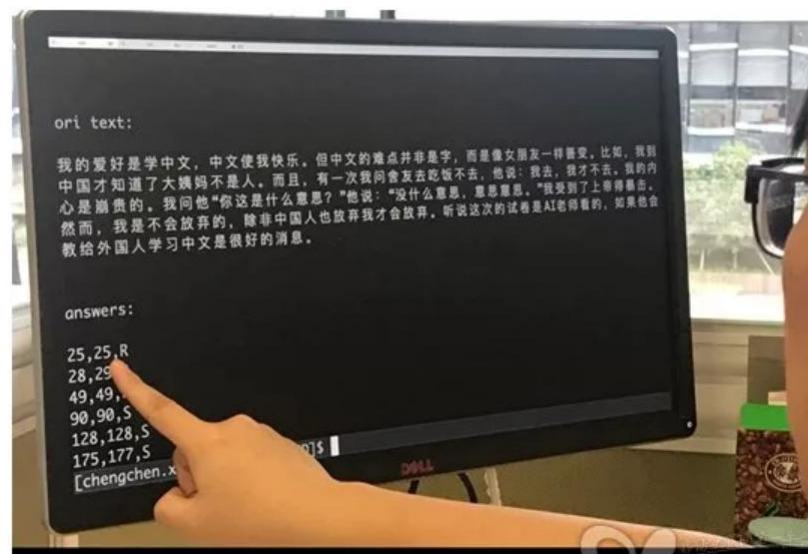


# CGED2016

<b>Participant (Ordered by abbreviations of names)</b>	<b>#TRuns</b>	<b>F1</b>	<b>#HRuns</b>	<b>F1</b>
NLP Lab, Zhengzhou University ( <b>ANO</b> )	0	-	2	0.2666
Central China Normal University ( <b>CCNU</b> )	0	-	1	0.0121
Chaoyang University of Technology ( <b>CYUT</b> )	3	<b>0.1248</b>	3	0.2125
Harbin Institute of Technology ( <b>HIT</b> )	0	-	3	<b>0.3855</b>
Institute of Computational Linguistics, Peking University ( <b>PKU</b> )	3		3	0.0724
National Chiao Tung University & National Taipei University of Technology ( <b>NCTU+NTUT</b> )	3	0.0745	0	-
National Chiayi University ( <b>NCYU</b> )	3	0.0155	3	0.0183
NLP Lab, Zhengzhou University ( <b>SKY</b> )	0	-	3	0.3627
School of Information Science and Engineering, Yunnan University ( <b>YUN-HPCC</b> )	3	0.0007	3	0.0035

# CGED2017

Participant (Ordered by abbreviations of names)	#Runs	F1
ALI_NLP	3	<b>0.2693</b>
BNU_ICIP	3	0.1152
CVTER	2	0.0653
NTOUA	2	0.0348
YNU-HPCC	3	0.1255



# CGED2018

TEAM	DIP	Correction
DM_NLP	3	3
AutoNlp	3	3
BUPT	3	3
CYUT-III	3	1
ECNU	3	0
HFL	3	3
IIT (BHU)	1	0
<u>ling@Cass</u>	3	3
NCYU	3	3
NTOUA	1	0
PkU_ICL	3	3
UIUC	2	0
walker	1	0

20 Participants, 13 made submissions, 8 recommend correction

# What's New from CGED 2015

- ▶ HSK joined !
- ▶ Training Set Doubled, Testing Set Tripled (Each Track!)
- ▶ Testing Setence --> Testing Unit
  - ▶ Multi-Sentence Diagnosis
- ▶ One Error each Sentence --> Multi-Errors
- ▶ Deep Learning Comes !

Team	Approach	Word/Character Embedding	Additional Resources
CYUT	CRF	---	NLP-TEA-1&NLP-TEA-2
HIT	CRF+LSTM networks	Character Embedding	---
NCTU+NTUT	W2V+CRF	Word Embedding	Sinica Balanced Corpus v4.0 LDC Chinese Gigaword v2 CIRB0303 Taiwan Panorama Magazine TCC300 Wikipedia(ZH_TW) NLP-TEA-1&NLP-TEA-2
NCYU	RNN+LSTM networks	Word Embedding	NLP-TEA-1&NLP-TEA-2
PKU	Bi-LSTM networks	Word Embedding	NLP-TEA-1&NLP-TEA-2
SKY	Ngram+CRF	---	---
YUN-HPCC	CNN/LSTM networks	Word Embedding	Wikipedia(ZH)

# What's New from CGED 2016

- ▶ Only HSK !
- ▶ Testing set was mixed in correct sentences, by sampling real data.
- ▶ ~~Deep Learning comes!~~
- ▶ CRF disappeared!



# 反思

- ▶ DL as data consuming methods, didn't bring revolution
- ▶ How to collect L2 writings effeciently
- ▶ Agreement in teachers' correction
- ▶ **PADS** should not be a good description of errors
- ▶ How to collaborate with cognitive linguistics and language aquisition studies
- ▶ Grammatical to Semantical
- ▶ Are there words in L2 learners' mind?
- ▶ How did they think when wrote: “|是我们运动员非常重要的”
- ▶ Chinese L2 study becoming lower in age: data and processing

# 反思

```
<TEXT id="200405109523201836 2 1x1">  
最好不要增长吸烟者是对的，可是尊重吸烟者也是应该的。  
</TEXT>  
<CORRECTION>  
吸烟者最好不要再增加是对的，可是尊重吸烟者也是应该的。  
</CORRECTION>  
<ERROR start_off="6" end_off="7" type="S"></ERROR>  
<ERROR start_off="1" end_off="10" type="W"></ERROR>  
  
<TEXT id="200505109525202573 2 3x1">  
这种的话，大多数的孩子会更发挥自己的自己的能力。  
</TEXT>  
<CORRECTION>  
这样的话，大多数的孩子会更好地发挥自己的自己的能力。  
</CORRECTION>  
<ERROR start_off="1" end_off="2" type="S"></ERROR>  
<ERROR start_off="13" end_off="13" type="S"></ERROR>  
<ERROR start_off="14" end_off="14" type="M"></ERROR>
```

# Future Work

- ▶ CGED 2018 in conjunction with ACL2018 in NLPTEA
  - ▶ Australia • Melbourne (19<sup>th</sup> July)
  - ▶ CGED always come with mild winters



# Acknowledgments

Consultation:

Lung-hao LEE@ NTNU

Proofreading:

Qi GONG, Peilan TANG, Ping LUO, Jie CHANG @ BLCU

Funds:



**语言资源高精尖创新中心**

Beijing Advanced Innovation Center for Language Resources



**北京语言大学**

BEIJING LANGUAGE AND CULTURE UNIVERSITY

# THANK YOU!

Find Data!

[www.cged.science](http://www.cged.science)

