# Overview of the NLP-TEA 2015 Shared Task for Chinese Grammatical Error Diagnosis

**Lung-Hao Lee**, National Taiwan Normal University

**Liang-Chih Yu**, Yuan Ze University

**Li-Ping Chang**, National Taiwan Normal University

# Introduction

- The NLP-TEA 2015 shared task features a Chinese Grammatical Error Diagnosis (CGED) task, providing an evaluation platform for the development and implementation of NLP tools for computer-assisted Chinese learning

# Shared 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
  - Mis-Ordering (**P**ermutation)
  - Redundancy (**A**ddition)
  - Omission (**D**eletion)
  - Mis-Selection (**S**ubstitution)

- For the sake of simplicity, the input sentence is selected to contain one defined error types

# Testing Examples

- Example 1
  Input: (sid=B2-0080)　他是我的以前的室友
  Output: B2-0080, 4, 4, Redundant

- Example 2
  Input: (sid=A2-0017)　那電影是機器人的故事
  Output: A2-0017, 2, 2, Missing

- Example 3
  Input: (sid=A2-0017)　那部電影是機器人的故事
  Output: A2-0017, Correct

- Example 4
  Input: (sid=B1-1193)　吳先生是修理腳踏車的拿手
  Output: B1-1193, 11, 12, Selection

- Example 5
  Input: (sid=B2-2292)　所以我不會讓失望她
  Output: B2-2292, 7, 9, Disorder

# Data Preparation

- The essay section of the computer-based <span style="color:red">Test of Chinese as a Foreign Language (TOCFL)</span>

- Native Chinese speakers were trained to manually annotate grammatical errors and provide corrections corresponding to each error

# Training Set

- This set included 2,205 selected sentences

- Error types were categorized as redundant (430 instances), missing (620), selection (849), and disorder (306)

- Each sentence is represented in SGML format

```
<DOC>
<SENTENCE id="B1-1120">
我的中文進步了非常快
</SENTENCE>
<MISTAKE start_off="7" end_off="7">
<TYPE>
Selection
</TYPE>
<CORRECTION>
我的中文進步得非常快
</CORRECTION>
</MISTAKE>
</DOC>
```

# Dryrun Set

- A total of 55 sentences were given to participants to familiarize themselves with the final testing process.

- The purpose is <span style="color:red">output format validation only</span>

- No matter which performance can be achieved that will not be included in our official evaluation.

# Test Set

- This set consists of 1,000 testing sentences

- Half of these sentences contained no grammatical errors, while the other half included a single defined grammatical error: redundant (132 instances), missing (126), selection (110), and disorder (132)

# Performance Metrics

- Correctness is determined at three levels
  - Detection-level
  - Identification-level
  - Position-level

| Confusion Matrix | | System Result | |
|---|---|---|---|
| | | Positive (Erroneous) | Negative (Correct) |
| **Gold Standard** | Positive | **TP** | **FN** |
| | Negative | **FP** | **TN** |

- Metrics
  - False positive rate (FPR) = FP / (FP+TP)
  - Accuracy = (TP+TN) / (TP+FP+TN+FN)
  - Precision = TP / (TP+FP)
  - Recall = TP / (TP+FN)
  - F1 = 2 * Precision * Recall / (Precision+Recall)

# Evaluation Examples

- FPR = 0.5
- Detection-level  Acc. = 0.75, Pre.=0.67, Rec.=1,  F1=0.8
- Correction-level Acc. = 0.625, Pre.=0.5, Rec.=0.75,  F1=0.6
- Position-level Acc. = 0.5, Pre.=0.33, Rec.=0.5,  F1=0.4


- **System Results**
  "B1-1138, 7, 8, Disorder", "A2-0087, 12, 13, Missing", "A2-0904, 5, 6, Missing", "B1-0990, correct", "A2-0789, 2, 5, Disorder", "B1-0295, correct", "B2-0591, 3, 3, Redundant" and "A2-0920, 4, 5, Selection"

- **Gold Standard**
  "B1-1138, 7, 10, Disorder", "A2-0087, 12, 13, Missing", "A2-0904, correct", "B1-0990, correct", "A2-0789, 2, 3, Selection", "B1-0295, correct", "B2-0591, 3, 3, Redundant" and "A2-0920, correct"

# 13 Participants and 18 Submitted Runs

| Participant (Ordered by abbreviations of names) | #Runs |
|---|---|
| Adam Mickiewicz University on Poznan (**AMU**) | 0 |
| University of Cambridge (**CAM**) | 0 |
| Chinese Academy of Sciences (**CAS**) | 0 |
| Confucius Institute of Rutgers University (**CIRU**) | 0 |
| Chaoyang University of Technology (**CYUT**) | 3 |
| Harbin Institute of Technology Shenzhen Graduate School (**HITSZ**) | 3 |
| Lingage Inc. (**Lingage**) | 0 |
| National Chiayi University (**NCYU**) | 3 |
| National Taiwan Ocean University (**NTOU**) | 3 |
| National Taiwan University (**NTU**) | 0 |
| South China Agriculture University (**SCAU**) | 3 |
| Tokyo Metropolitan University (**TMU**) | 3 |
| University of Leeds (**UL**) | 0 |
| **Total** | **18** |

# Testing Results

| Submission | False Positive Rate | Detection Level | | | | Identification Level | | | | Position Level | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Acc. | Pre. | Rec. | F1 | Acc. | Pre. | Rec. | F1 | Acc. | Pre. | Rec. | F1 |
| CYUT-Run1 | 0.096 | 0.584 | 0.7333 | 0.264 | 0.3882 | 0.522 | 0.5932 | 0.14 | 0.2265 | 0.504 | 0.52 | 0.104 | 0.1733 |
| CYUT-Run2 | **0.082** | 0.579 | **0.7453** | 0.24 | 0.3631 | **0.525** | **0.6168** | 0.132 | 0.2175 | **0.505** | **0.5287** | 0.092 | 0.1567 |
| CYUT-Run3 | 0.132 | 0.579 | 0.6872 | 0.29 | 0.4079 | 0.505 | 0.5182 | 0.142 | 0.2229 | 0.488 | 0.45 | 0.108 | **0.1742** |
| HITSZ-Run1 | 0.956 | 0.509 | 0.5047 | 0.974 | 0.6648 | 0.173 | 0.2401 | 0.302 | 0.2675 | 0.031 | 0.0185 | 0.018 | 0.0182 |
| HITSZ-Run2 | 0.938 | 0.505 | 0.5027 | 0.948 | 0.657 | 0.149 | 0.201 | 0.236 | 0.2171 | 0.036 | 0.0105 | 0.01 | 0.0103 |
| HITSZ-Run3 | 0.884 | 0.51 | 0.5056 | 0.904 | 0.6485 | 0.188 | 0.2273 | 0.26 | 0.2425 | 0.068 | 0.0221 | 0.02 | 0.021 |
| NCYU-Run1 | 0.48 | 0.53 | 0.5294 | 0.54 | 0.5347 | 0.354 | 0.2814 | 0.188 | 0.2254 | 0.274 | 0.0551 | 0.028 | 0.0371 |
| NCYU-Run2 | 0.396 | 0.567 | 0.5724 | 0.53 | 0.5504 | 0.423 | 0.3793 | 0.242 | 0.2955 | 0.343 | 0.1715 | 0.082 | 0.111 |
| NCYU-Run3 | 0.374 | **0.607** | 0.6112 | 0.588 | 0.5994 | 0.463 | 0.4451 | 0.3 | **0.3584** | 0.374 | 0.246 | 0.122 | 0.1631 |
| NTOU-Run1 | 1 | 0.5 | 0.5 | 1 | 0.6667 | 0.117 | 0.1896 | 0.234 | 0.2095 | 0.005 | 0.0099 | 0.01 | 0.01 |
| NTOU-Run2 | 0.914 | 0.531 | 0.5164 | 0.976 | **0.6754** | 0.225 | 0.2848 | **0.364** | 0.3196 | 0.123 | 0.149 | **0.16** | 0.1543 |
| NTOU-Run3 | 0.948 | 0.519 | 0.5098 | 0.986 | 0.6721 | 0.193 | 0.2605 | 0.334 | 0.2927 | 0.093 | 0.1238 | 0.134 | 0.1287 |
| SCAU-Run1 | 0.62 | 0.505 | 0.504 | 0.63 | 0.56 | 0.287 | 0.2383 | 0.194 | 0.2139 | 0.217 | 0.0801 | 0.054 | 0.0645 |
| SCAU-Run2 | 0.636 | 0.503 | 0.5023 | 0.642 | 0.5637 | 0.279 | 0.2337 | 0.194 | 0.212 | 0.209 | 0.0783 | 0.054 | 0.0639 |
| SCAU-Run3 | 0.266 | 0.503 | 0.5056 | 0.272 | 0.3537 | 0.416 | 0.2692 | 0.098 | 0.1437 | 0.385 | 0.1192 | 0.036 | 0.0553 |
| TMU-Run1 | 0.478 | 0.516 | 0.5162 | 0.51 | 0.5131 | 0.313 | 0.1787 | 0.104 | 0.1315 | 0.27 | 0.0363 | 0.018 | 0.0241 |
| TMU-Run2 | 0.134 | 0.524 | 0.5759 | 0.182 | 0.2766 | 0.479 | 0.4071 | 0.092 | 0.1501 | 0.449 | 0.1928 | 0.032 | 0.0549 |
| TMU-Run3 | 0.35 | 0.546 | 0.5581 | 0.442 | 0.4933 | 0.42 | 0.3519 | 0.19 | 0.2468 | 0.362 | 0.1745 | 0.074 | 0.1039 |

# Summary

- It is a really difficult task to develop the computer-assisted Chinese learning tool, since there are only target sentences without the help of their context

- None of the submitted systems provided superior performance. In general, this research problem still has long way to go.

# Conclusions

- All submissions contribute to the common effort to produce an effective Chinese grammatical diagnosis tool

- The individual reports in the shared task proceedings provide useful insight into Chinese language processing

# Future Work

- **NLP-TEA-3 Workshop in COLING 2016**
  - To be bided
  - Osaka, Japan

- **The Shared Task**
  - Chinese Grammatical Error Diagnosis

# Acknowledgments

- Ministry of Education, Taiwan
  - Aim for the Top University Project
  - Center of Learning Technology for Chinese, NTNU
  - Innovative Center for Big Data and Digital Convergence, YZU
- Ministry of Science and Technology, Taiwan
  - International Research-Intensive Center of Excellence Program
  - Grant no.: MOST 104-2911-I-003-301

# THANK YOU

- All data sets with gold standards and evaluation tool are publicly available for research purposes at

[http://ir.itc.ntnu.edu.tw/lre/nlptea15cged.htm](http://ir.itc.ntnu.edu.tw/lre/nlptea15cged.htm)