AUTOMATED ESSAY MARKING TOOL FOR ESL WRITING
BASED ON HEURISTICS

Nazlia Omar
Universiti Kebangsaan Malaysia
Malaysia
no@ftsm.ukm.my

Nur Asma Mohd Razali
Universiti Kebangsaan Malaysia
Malaysia
nurasma84@yahoo.com

Saadiyah Darus
Universiti Kebangsaan Malaysia
Malaysia
adi@ukm.my

Abstract

Although computers and artificial intelligence have been proposed as tools to facilitate the evaluation of student essays, they are not specifically developed for Malaysian ESL (English as a second language) learners. A marking tool which is specifically developed to analyze errors in ESL writing is very much needed. Though there are numerous techniques adopted in automated essay marking, research on the formation and use of heuristics to aid the construction of computer-based essay marking system has been scarce. Thus, this paper aims to introduce new heuristics that can be used to mark essays automatically and detect grammatical errors in tenses. This approach, which uses natural language processing technique, can be applied as part of the software requirement for a CBEM (Computer Based Essay Marking) system for ESL learners. The preliminary result based on the training set shows that the heuristics are useful and can improve the effectiveness of automated essay marking tool for writing in ESL.

Keywords
Computer based essay marking system, heuristics, errors and ESL writing, natural language processing

1. INTRODUCTION

English has become a language for communication worldwide and it has been enrolled as part of courses in university. While writing is a good way to stimulate learning and critical thinking, writing skill is considered a very important skill, since it is a way to get the message across. Writing essays enables students to create avenues for their voices to be heard through communication particularly in writing. However, marking essays manually is an overwhelming task and time consuming for educators [1], especially when the size of the class increases. Larger classes reduce opportunities for learners' feedback on written work and learners in turn may have little contact with their lecturers. A study carried out by Darus et al. [3, 4, 5] shows that students prefer to receive feedback on errors in essay and their type [2], coherence of the text and organization of ideas. In Malaysian educational context, English is used as a second language (ESL). ESL learners are those learners whose English is not their primary or first language (L1) but they are fairly able to write, read, speak or understand English.

This paper aims to facilitate the task of essay marking especially to Malaysian ESL learners by automating the marking process. The task focuses on the identification of the grammatical errors in tenses made by the learners. Though there are numerous techniques adopted in automated essay marking, research on the formation and use of heuristics to aid the construction of computer-based
essay marking system has been scarce. Thus, this paper aims to introduce new heuristics that can be used to mark essays automatically and detect grammatical errors in tenses.

2. RELATED RESEARCH

The use of computers and media for language learning and teaching has become two disciplines, known as computer-assisted language learning (CALL) and intelligent CALL (ICALL). Past research in the area of CALL and ICALL has shown that using technology can enhance the ability of student to communicate and learn. Artificial Intelligence and natural language processing techniques offer many possibilities to improve CALL systems. However, the application of these technologies is not adequate and still requires more research. In this paper, research prototypes that explore an automatic analysis of grammatical ESL writing and supplying meaningful feedback are reviewed. The research prototypes discussed here touch some of the new possibilities and directions for future investigations.

Two Distance Applications Support English Distance Learning [6] is a system that was developed through multimedia database and Internet Technologies called English multimedia corpus. English articles, dialogs, and videos are included in the system. To construct English multimedia corpus, semantic query and Link grammar are applied during implementation of the system. The system concentrated to teach English grammar and allowed teachers to understand the most frequent mistakes. The main function of this system is to query the English sentence pattern through keywords from the English multimedia corpus. The second function is to detect grammatical errors in written English.

A Rule-Based Style and Grammar Checker [7] is a project aimed to provide an Open Source style and grammar checker for the English language. This tool uses Rule-based checking technique. In this approach, the system generates a text and returns a list of possible error. For its purposes, each of the word in the text is assigned its part-of-speech tag and each sentence is split into chunks, e.g. noun phrase. The parsing results are processed further by matching all the checker's pre-defined error rules. If a rule is matches, the text is supposed to contain error at the position of the match. Each rule includes an explanation of error and is shown to user.

A Sentence Analyzer and Viewer for Detecting Grammatical Errors [8] is a system that aims to identify grammatical errors in a sentence and has been implemented using Prolog language. In English, agreement is required for every sentence at the clause and noun-phrase levels. This system is able to detect agreement errors in sentence and display helpful messages to learners. The tool uses a few of stages in natural language processing which accesses tokenization, lexical analysis and syntactic analysis to analyze the learners' input.

English Grammar Checking [9] is a system that applied Artificial Immune System (AIS) based technique as an approach for grammar checking. The motivation of the study comes from human immune system, which is able to identify external harmful entity from the self cells in human body. This grammar checker has been implemented using JAVA. It uses Reuters-21578, Distribution 1.0 text as a corpus and is tagged for part of speech by using MontyLingua parser. It can detect any grammatical construct outside the corpus and identify it as an error.

GRADES [10] is Grammar Diagnostic Expert System that can detect and explain grammatical errors of non-native speakers. It is designed to diagnose errors found by native Japanese adult who are learning ESL. It contains small lexicon of words and can be expanded easily. GRADES diagnosis through a classification process where an error category is considered and pattern matching rules are used. An explanation is generated to help the user learn why the sentence is ungrammatical.

3. LINGUISTIC ANALYSIS OF ESL WRITING

Ferris [11] has pointed out that written errors made by adult second language learners are often quite different from those made by NES (Native English Speakers). Common ESL writing errors are morphological errors, lexical errors, syntactic errors and mechanical errors. Truscott [12] noted that different types of errors may need varying treatment in terms of error correction. In addition, lecturers need to understand that different students may make distinct types of errors and they should understand to prioritize error feedback for individual students. Lim Ho Peng [13] produced a classification scheme of errors that were displayed by Malaysian high school students and the result of the error analysis is shown in Table 1.
Table 1. Types of errors in ESL essays

<table>
<thead>
<tr>
<th>Errors</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense</td>
<td>One of them went in and <strong>steal</strong> a chain.</td>
</tr>
<tr>
<td>Articles</td>
<td>He ran to the Ahmad’s house right away.</td>
</tr>
<tr>
<td>Agreement</td>
<td>Azman <strong>do</strong> not know anything about it.</td>
</tr>
<tr>
<td>Infinitive</td>
<td>When it <strong>began</strong> moved I heard a sound.</td>
</tr>
<tr>
<td>Pronouns</td>
<td>He told <strong>she</strong> about his village.</td>
</tr>
<tr>
<td>Word Order</td>
<td>What work <strong>I must</strong> do?</td>
</tr>
<tr>
<td>Negative Constructions</td>
<td>She had <strong>not</strong> brother to take care of her.</td>
</tr>
<tr>
<td>Lexical</td>
<td>The noise was <strong>on</strong> our back.</td>
</tr>
<tr>
<td>Mechanics</td>
<td>It was quite far <strong>form</strong> my village.</td>
</tr>
<tr>
<td>Use of Typical Malaysian Words</td>
<td>I <strong>dapati</strong> that the cat was outside the house.</td>
</tr>
</tbody>
</table>

### 3.1 Analysis of Errors

400 written essays by undergraduate students who enrolled in Written Communication course at the School of Language Studies and Linguistic, Faculty of Social Sciences and Humanities, Universiti Kebangsaan Malaysia were gathered. Markin 3.1 [14] was used to speed up the process of analyzing the essays of errors. The annotation buttons in the software are first customized accordingly based on the error classification scheme. Markin 3.1 is a correction tool that allows instructors to mark written work submitted by students in the form of electronic documents. Markin 3.1 provides five marking facilities; annotation buttons, add feedback, add comment, add a grade, and compile error statistics. Figure 1 shows the marked essays using Markin 3.1.

Fig. 1. Original marked essays using Markin 3.1
3.2 Results on Error Analysis

The results on the error analysis using Markin 3.1 software is as shown in Table 2. The average is calculated by dividing the number of errors with the total numbers of essays. Six most common errors that are made by learners are as follows: tenses, prepositions, articles, word choice, mechanics, and verb to be. Based on the result of the analysis, it seems that the most errors that need to be analyzed by the error analysis marking tool are tenses.

<table>
<thead>
<tr>
<th>Errors</th>
<th>No. of errors</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Tense</td>
<td>1595</td>
<td>3.99</td>
</tr>
<tr>
<td>2. Articles</td>
<td>1204</td>
<td>3.01</td>
</tr>
<tr>
<td>3. Subject verb agreement</td>
<td>631</td>
<td>1.58</td>
</tr>
<tr>
<td>4. Other agreement errors</td>
<td>520</td>
<td>1.30</td>
</tr>
<tr>
<td>5. Infinitive</td>
<td>145</td>
<td>0.36</td>
</tr>
<tr>
<td>6. Gerunds</td>
<td>292</td>
<td>0.73</td>
</tr>
<tr>
<td>7. Pronouns</td>
<td>696</td>
<td>1.74</td>
</tr>
<tr>
<td>8. Possessive and attributive</td>
<td>232</td>
<td>0.58</td>
</tr>
<tr>
<td>9. Word order</td>
<td>194</td>
<td>0.49</td>
</tr>
<tr>
<td>10. Incomplete structures</td>
<td>253</td>
<td>0.63</td>
</tr>
<tr>
<td>11. Negative construction</td>
<td>56</td>
<td>0.14</td>
</tr>
<tr>
<td>12. Prepositions</td>
<td>1468</td>
<td>3.67</td>
</tr>
<tr>
<td>13. Mechanics</td>
<td>1105</td>
<td>2.76</td>
</tr>
<tr>
<td>14. Miscellaneous uncategorizable</td>
<td>1117</td>
<td>2.79</td>
</tr>
<tr>
<td>15. Word choice</td>
<td>1123</td>
<td>2.81</td>
</tr>
<tr>
<td>16. Word form</td>
<td>629</td>
<td>1.57</td>
</tr>
<tr>
<td>17. Verb to be</td>
<td>820</td>
<td>2.05</td>
</tr>
</tbody>
</table>

Table 2. Result of error analysis

4. AUTOMATED ESSAY MARKING TOOL BASED ON HEURISTICS

The next step in this study is the development of techniques and algorithm for detecting and analyzing tense error in learners’ essays. Based on these algorithms, the error-analysis marking tool will be developed. The approach used in the present study is similar to the study carried out by [15]. The heuristics are newly formed and developed based on the observation of essays written by ESL learners at Universiti Kebangsaan Malaysia. The rules of grammar are developed from [16] and details of the heuristics can be found in [17].

4.1 Overview of Heuristics

Heuristics enable someone to learn and know something through their own experience [18]. Heuristics represent an indefinite assumption [19] often guided by common sense, to provide good but not necessarily optimal solutions to difficult problems, easily and quickly [20]. Research on the formation and use of heuristics to aid the construction of CBEM system has been scarce.

The only existing work that proposes a large number of heuristics to be used in developing CBEM system is Argument Partitioning and Annotation Program (APA) [21]. APA selects the dictionary-based clue words, terms and non-lexical structures. APA uses partition, and annotate arguments specify syntactic structure and syntactic contexts. APA uses parsed essays to identify syntactic structure in essays.
**Heuristic to identify error in simple present tense:**

If noun or pronoun is in the set of heuristic Simple Present 1, check all verbs after each noun. If verb is tagged as VB, the sentence has no grammatical error.

You neglect your diet and take many unhealthy foods to release your tension.

You/PRP neglect/VB your/PRP$ diet/NN and/CC take/VB many/JJ unhealthy/JJ foods/NNS to/TO release/VB your/PRP$ tension/NN ./.

Note that “You” is a pronoun and “neglect” is tagged as VB, so the sentence has no grammatical error. The example above illustrates this heuristic.

**Heuristic to identify error in present progressive:**

If noun or pronoun is in the set of heuristic Present Progressive 3 followed by ‘are’, check the verb. If verb is tagged as VBG, the sentence has no grammatical error.

People are collecting the rubbish at Taman Negara.

People/NNS are/VBP collecting/VBG the/DT rubbish/NN at/IN Taman/NNP ./.

In this example, the noun ‘people’ is followed by ‘are’ and the verb ‘collecting’ is tagged VBG, so the sentence has no grammatical error.

**Heuristic to identify error in simple future:**

If noun or pronoun is in the set of heuristic Simple Future 1 followed by ‘will’, or “shall”, check the verb. If verb is tagged as VB, the sentence has no grammatical error.

The staff will write the letters later.

The/DT staff/NN will/MD write/VB the/DT letters/NNS later/RB ./.

“Staff” is tagged as noun and followed by “will”, the immediate verb is tagged as VB, so this sentence has no grammatical error.

### 4.1. Results on training data set

A manual test was carried out to test the new heuristics. Thirty writing samples were selected as the training dataset. Table 3 presents the results on the training dataset. Among the 30 writing samples, only Essay no. 10 and no. 13 have low percentage correctness. Most of these errors resulted from passive sentence and the use of modal like ‘can’, ‘can not’, ‘can not be’ and ‘will not be’ in the essays. The overall percentage of correctness of the training data set is 84%. This indicates that heuristics-rule based approach is viable and can be selected for the implementation.
<table>
<thead>
<tr>
<th>Essays</th>
<th>Detect error in tenses</th>
<th>Percentage correctness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>24</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>26</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>29</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

|        | Average | 84% |

Table 3. Result on training dataset

5. SYSTEM OVERVIEW

The automated marking tool architecture mainly consists of two stages: the natural language stage and the logical stage. The natural language stage comprises of parsing the natural language input. Natural language processing is involved in the process of reading natural language input (in the form of essays) into the system. The process begins by reading a plain input text file containing sentences written in English. For this purpose, a parser is used to parse the English sentences to obtain their part-of-speech (POS) tags before processing further. POS assigns each word in an input sentence into its proper part of speech such as noun, verb and determiner to reflect the word’s syntactic category [22]. The parser used is Memory-based Shallow Parser (MBSP) [23, 24]. The parsed text is then being
fed into the system to detect error in tense. The final result consists of type of errors in tenses for the natural language input. The final result that is displayed with a friendly user interface can be used as feedback for students to enhance the quality of their writing. Figure 1 depicts the architecture of automated marking tool for ESL Writing. The task requires several steps to be carried out, each of which is listed as follows:

- Step 1: Part of speech tagging using Memory-Based Shallow Parser
- Step 2: Read natural language input text into system
- Step 3: Apply heuristics approach
- Step 4: Produce final result

5.1 Algorithm

Figure 2 is a part of the algorithm to apply a selection of heuristics to detect errors in simple present tense. In this algorithm, ‘check’ refers firstly to the identification of a match, which meets the heuristic’s condition. Heuristics are applied to any relevant words in the sentences that meet the heuristics’ criteria.
6. CONCLUSION AND FUTURE WORK

We have described an approach of developing an automated marking tool for tenses for ESL writing. Based on the observation of English sentence structure and grammar, the formation of heuristics can be further developed to improve the accuracy of the result. For future work, weights may need to be assigned to each of the heuristic in order to support the level of confidence of each of the heuristics.

The next step is the development of algorithm based on the heuristics and rules-based approach for detecting errors in learners’ essays. Based on these algorithms, the automated marking tool will be developed.

Acknowledgement:

This study is funded by MOSTI (Ministry of Science, Technology and Innovation, Malaysia) with the following research code: e-ScienceFund 01-01-02-SF0092.
References