Recognizing Syntactic Errors in Written Philippine English

Allen Michael Gurrea  
chuchachicho@yahoo.com

Anderson Liu  
liu.anderson@gmail.com

Deryk Ngo Vincent  
omnibonta@yahoo.com

Jemie Que  
francisveto@yahoo.com

Ethel C. Ong  
onge@dlsu.edu.ph

College of Computer Studies, De La Salle University  
2401 Taft Ave. Manila Philippines 1004

Abstract

This paper reports on the error identification component of POPSICLE, a computer-aided language learning software that has been designed to help Filipino language speakers learn standard English. The software assesses the English grammar proficiency of the learner based on an input document, identifies the grammatical errors committed in the document, and generates grammar lessons that are tailor fit to the individual needs of the learner. The learner is given opportunities to correct and learn from his mistakes. The software maintains a user model that tracks an individual learner’s English grammar proficiency, his position and path toward acquiring English, the dialogue history containing the text generated by the system during the current tutorial session, the evaluation scores for each of the teaching strategies employed, and a concise log of explanations attempted by the system over the learning period of the current user.

Key Words: Mal-Rules, English Grammar Errors, English Grammar Tutor, Computer Assisted Language Learning, Intelligent Tutoring System

1. INTRODUCTION

The Interactive Computer Identification and Correction of Language Errors (ICICLE) is an intelligent tutoring system (ITS) being developed at University of Delaware [4, 6, 7, 8]. It aims to tutor deaf students by employing natural language generation techniques. It has a goal of generating tutorial feedback based from the student’s performance. Each feedback is individualized depending on the user model the system uses that include student’s language knowledge and learning style, the context of previous sessions or explanations [5].

ICICLE interacts with a student through the user inputs and system response. It goes through a series of processing in the form of a cycle that reviews the user’s input by performing syntactical analysis to determine grammatical errors and responds accordingly through tutorial feedback. The feedback will inform the user about the nature of errors found and instruct him or her on how to correct them. The user then re-submits his or her corrected writing and the cycle will begin again [2].

The system recognizes errors by using mal-rules, also called error production rules. These rules extend the set of sentences accepted by the grammar to include those containing the specified errors. The mal-rules themselves are derived from an error taxonomy which was the result of an analysis of writing samples [8].

The outline of the paper will be as follows. First the POPSICLE system will be presented, followed by the methodology in creating the mal-rules. Then, examples of common errors of Filipinos learning the English language and their corresponding mal-rules will be provided.

2. POPSICLE

POPSICLE is an ITS patterned after ICICLE. Its primary function is to tutor second-language learners of English. Essential to this goal is the correct analysis of the source and nature of user-generated language errors and the production of tutorial feedback regarding student performance which is both correct and individualized, taking into
consideration the language knowledge, proficiency, and learning style of the student, as well as the context of previous explanations and related concepts in the domain.

As patterned from the ICICLE system, POPSICLE also applies the Second Language Acquisition Theory of Stephen Krashen to represent the user’s position and path toward acquiring written English as a second language. The theory states that when a person is learning a second language, he produces his own internalized grammar that represents his hypothesis of that language. According to his “Principles and Practice in Second Language Acquisition” [10], that internalized grammar is produced from the person’s primary language as a guide for learning a new language.

However, in the local setting, Tagalog cannot be used as a pattern to learn the English language since the two languages differ on their structural design. For example, the English grammar requires that the subject comes before the verb. On the other hand, the Tagalog language can place the subject after the verb. It is important to consider the differences of the languages in order to address the common problems that second language learners face.

Therefore, following Krashen’s theory and seeing that Computer-Assisted Language Learning (CALL) software that consider Filipino as the first language are not available locally, there is a need for a tutoring system that can support the learning progress of the Filipino students acquiring English as a second language.

This presents the dissimilarity of POPSICLE against the ICICLE system. They vary from the target population and the domain. As discussed earlier, ICICLE aims to tutor deaf students whose primary language is the American Sign Language (ASL) whereas the POPSICLE system is dedicated specifically for students learning the English grammar and whose primary language is Tagalog.

With this at hand, the POPSICLE system will make use of a parser with coverage of English that has been augmented by error-production rules or mal-rules much like ICICLE. However, because of the difference in target population, a different set of mal-rules is required.

3. MAL-RULES
A mal-rule is an error grammar production rule that can be used to parse a given input sentence to determine its grammar errors. It makes use of specific form of the words to properly tag each words correctly, for example, Allen = PROPERNOUN, wrote = PASTVERB)

The problematic grammatical features of the Philippine English can be found in Gonzalez’s entry in The Oxford Companion to the English Language as cited by [1]. In the English grammar, the tense/aspect system of verbs is problematic, as depicted in the following instances:

1. The use of the past perfect for the simple past, e.g., He had been sick last week.
2. The use of continuous tenses for habitual aspect, e.g., He is going to school every day.

Also problematic is the use of the article. There is also the confusion between he and she, particularly in combination with kinship terms (his husband, her wife) especially in spoken English. Local adaptations of the rules appear in verb-preposition combinations (based from, result to) as well as co-occurrence rules between verbs and nouns (we enjoyed, I can’t afford).

A subvariety of Philippine English called Colegiala English (commonly known in the Philippines as conyo) has produced more hindrances to the correct usage of the English grammar. Such features of this variety include the following:

1. The frequent usage of common Tagalog function words (particles, demonstrative pronouns, conjunctions, and enclitics, like kasi, parang, bale) in English sentences
2. The use of Tagalog content words in English sentences (pikon, bukol, bastos, make + Tagalog content words, as in make pakalbo, make kento)
3. The use of Tagalog exclamatives in English sentences (baduy, grabe, daya, kadiri)

It is common to try and explain a concept in English to a Filipino learner by simply translating it to Tagalog. However, this will not work in teaching grammar to Filipino learners since the grammatical
structure of Tagalog is much different from that of English. In his work, *English Sentence Patterns for Filipinos* [9], Wilson tries to solve the problem by teaching Filipino learners not only by presenting patterns of phrases in a sentence but rather patterns of the sentences themselves. These patterns are the basis for POPSICLES’ mal-rules.

Furthermore, [9] classifies students into “beginning”, “intermediate”, and “advance” not according to age but according to the number and value of the students’ habitual mistake. He defines habitual mistakes as those that are made in situations like conversations where their attention is focused on the content rather than on the rules of grammar. These deviations from the patterns are commonly made by Filipinos learners and are thus considered as mal-rules.

The addition of mal-rules into the parser grammar of the system will enable it to capture errors common to the target population. In the following section, we will discuss how we were able to create the mal-rules for the common errors of our target population.

3.1. Methodology in Creating the Mal-Rules
In order to create the mal-rules for POPSICLE, there were several procedures that were followed. Since the POPSICLE system is designed to cater to the needs of Filipinos learning the English language, mal-rules had to be taken from a general medium. However, it was also taken into consideration that POPSICLE will first and foremost be used by DLSU, more specifically by Lasallian students. The first task was to collect information on common errors that students of La Salle makes on written English. With the cooperation of the Department of English and Applied Linguistics (DEAL), documents from their English Learning Laboratory (ELL) were the primary material for survey of the common errors. Past works by students were examined and the frequency of such errors was noted. The file was then passed on to the faculty of ELL for classification.

Mal-rules for the system do not only rely on the data found in ELL. Because of the variety of users the system intends to address, the complete set of mal-rules is an intersection of the data acquired from the ELL and from a book on English language specifically for Filipino learners. In the book of R. Wilson [9], numerous errors commonly made by Filipino learners are listed. From here, several sample sentences were taken and along with the ELL data were translated into the mal-rules of POPSICLE.

To be able to parse a sentence, POPSICLE will use the same context-free grammar (CFG) of the Automated Essay Evaluator (AEE) [3]. AEE is a system that checks and grades an essay of a certain corpus. POPSICLE uses an extended version of the grammar of AEE since the latter does not explicitly check the grammar of the user (e.g. subject tenses). A sample CFG used by AEE is shown in Figure 1. The CFG is also extended to include the mal-rules.

```
SENTENCE : INDEPENDENT_CLAUSE
           | COORDINATE_CLAUSE
           | DEPENDENT_CONJUNCTION
           | QUESTION_CLAUSE
           | COMPLEX

INDEPENDENT_CLAUSE : ADVERBIAL SUBJECT
                     | VERB_PHRASE
                     | ADVERBIAL1 SUBJECT
                     | ADVERBIAL1 COORDINATE_ADVERBIAL
                     | ADVERBIAL1 <comma> SUBJECT
                     | ADVERBIAL1 <comma> VERB_PHRASE
                     | ADVERBIAL1 <comma> <comma> VERB_PHRASE

QUESTION_CLAUSE : AUXILIARY NOUN_PHRASE
                 | VERB_PHRASE
                 | DIRECT_OBJECT
                 | <question> ADVERBIAL VERB_PHRASE
                 | <question> NOUN_PHRASE <auxiliary>
                 | NOUN_PHRASE VERB_PHRASE
                 | <question> ADVERBIAL NOUN_PHRASE
                 | <question> ADJECTIVE VERB_PHRASE

VERB : ADVERB <verb> ADVERB
       | ADVERB <verb> ADVERB
       | COORDINATE_VERB

OBJECT : DIRECT_OBJECT
         | ADJECTIVE
         | INDIRECT_OBJECT DIRECT_OBJECT
```
The process of deriving a mal-rule for an error category is as follows:

(1) A sample erroneous sentence from a given error category as well as its corresponding corrected version are acquired, e.g.,

Sentence with grammar error: Why you are late again?
Corresponding corrected sentence: Why are you late again?

(2) Using the corrected version of the sample sentence, the erroneous sentence is traced through the grammar shown in Figure 1 until its corresponding parse is determined. For example, the grammar contains the following production rule:

\[
\text{QUESTION_CLAUSE:} <\text{question}> \\
\quad <\text{auxiliary}> \text{NOUN_PHRASE} \\
\quad \text{VERB_PHRASE}
\]

The sample sentence with grammar error failed after <question>.

(3) We then use the sample erroneous sentence and form its CFG following the format of the parse we made earlier with the corrected version to derive a mal-rule. Going back to our example, the mal-rule generated is as follows:

\[
<\text{question}> \text{NOUN_PHRASE} \\
<\text{auxiliary}> \text{VERB_PHRASE}? 
\]

(4) The production rule label is changed to properly brand the incorrect element of the sentence structure. The standard is the appending of “ERR_” before the certain element is used to avoid irregularities.

\[
\text{ERR_QUESTION_CLAUSE}
\]

All possible tracks and combinations of the grammar must be exhausted. Each expandable element that contains the original point of error must be appended with the standard “ERR_”.

The following section contains sample common errors and their corresponding mal-rule.

### 3.2. Generated Mal-Rules

**Error Category 1**

<table>
<thead>
<tr>
<th>Correct Sentence 1-1</th>
<th>Incorrect Sentence 1-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why are you late again?</td>
<td>Why you are late again?</td>
</tr>
<tr>
<td>ERR_QUESTION_CLAUSE</td>
<td>NOUN_PHRASE</td>
</tr>
<tr>
<td>&lt;auxiliary&gt;</td>
<td>VERB_PHRASE</td>
</tr>
</tbody>
</table>

**Correct Sentence 1-2**

I opened the door for him.

**Incorrect Sentence 1-2**

I opened him the door.

**Error Category 2**

<table>
<thead>
<tr>
<th>Correct Sentence 2-1</th>
<th>Incorrect Sentence 2-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is a cat on the bench.</td>
<td>There are a cat on the bench.</td>
</tr>
<tr>
<td>ERR_INDEPENDENT_CLAUSE</td>
<td>SUBJECT</td>
</tr>
<tr>
<td>ERR_VERB_PHRASE</td>
<td>VERB</td>
</tr>
<tr>
<td>ERR_OBJECT</td>
<td>ERR_OBJECT</td>
</tr>
<tr>
<td>PLURAL_NOUN_PHRASE</td>
<td>VERB_PHRASE</td>
</tr>
</tbody>
</table>

**Correct Sentence 2-2**

The girls have won the game.

**Incorrect Sentence 2-2**

The girls has won the game.

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-1</th>
<th>Incorrect Sentence 3-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>They have given their grandmother a present.</td>
<td>They have given their grandmother a present.</td>
</tr>
<tr>
<td>ERR_INDEPENDENT_CLAUSE</td>
<td>SUBJECT</td>
</tr>
<tr>
<td>ERR_VERB_PHRASE</td>
<td>PRESENTVERB</td>
</tr>
<tr>
<td>NOUN_PHRASE</td>
<td>NOUN_PHRASE</td>
</tr>
</tbody>
</table>

**Correct Sentence 3-2**

I should have found you.

**Incorrect Sentence 3-2**

I should have find you.

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>

**Error Category 3**

<table>
<thead>
<tr>
<th>Correct Sentence 3-3</th>
<th>Incorrect Sentence 3-3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you get it?</td>
<td>Did you get it?</td>
</tr>
</tbody>
</table>
4. CONCLUSION AND FURTHER WORK

In this paper, we presented some examples of the mal-rules that are augmented into the grammar being used by POPSICLE. These mal-rules will enable the system to recognize some of the English grammar errors we expect the users to commit. Further additions of mal-rules to the grammar are possible by following the specified format.

The detection of English grammar errors is one of the modules of the POPSICLE system. In order to generate learner-sensitive tutorial response, the system must be able to track the progress of each user. As such, POPSICLE maintains a comprehensive user model that tracks an individual learner’s English grammar proficiency, his position and path toward acquiring English as a second language, the dialogue history containing the text generated by the system during the current tutorial session, the evaluation scores for each of the teaching strategies employed, and a concise log of explanations attempted by the system over the learning period of the current user which can encompass more than one session. This is the next direction of POPSICLE.

5. REFERENCES


