

**DISCOURSAL ASPECTS OF INTERLANGUAGE IN THE  
ARGUMENTATIVE ESSAYS OF UITM SARAWAK DIPLOMA  
STUDENTS**



**RESEARCH MANAGEMENT INSTITUTE (RMI)  
UNIVERSITI TEKNOLOGI MARA  
40450 SHAH ALAM, SELANGOR  
MALAYSIA**

**BY:**

**AFFIANA KADRI  
PM DR. SIMON PHILLIP POTLEY@FAIZAL HAKIM**

**DECEMBER 2015**

## Contents

1.	Letter of Report Submission.....	1
2.	Letter of Offer (Research Grant).....	2
3	Acknowledgements .....	3
4.	Enhanced Research Title and Objectives.....	4
5.	Introduction .....	4
	5.1. Proposed Executive Summary .....	4
	5.2. Enhanced Executive Summary .....	4
	5.3. Introduction .....	4
	5.4. Brief Literature Review .....	7
	5.5. Methodology .....	8
	5.6. Results and Discussion .....	10
	5.7. Conclusion and Recommendation.....	11
	5.8. References/Bibliography.....	12
6.	Research Outcomes.....	14
7.	Appendices.....	15

## 4. Research Objectives

Objectives:

1. To investigate the English interlanguage of Malaysian undergraduate students using a mini-corpus of argumentative essays collected from BEL 260 students at the UiTM Sarawak Samarahan Campus
2. To identify and examine the discourse-level and grammatical errors in the writing produced by BEL 260 students
3. To provide constructive proposals for improving the teaching of writing on the BEL 260 course and beyond, based on the error data derived from the corpus

## 5. Report

### 5.1 Proposed Executive Summary

This project aims to investigate the English Language errors produced by undergraduate students preparing for the Malaysian University English Test (MUET) in a large Malaysian public university. Errors are seen as one of the ways in which the *Interlanguage* of language learners can be understood. The Interlanguage refers to the incomplete knowledge of grammar structures and the strategies used to learn them which are characteristic of language learners at particular stages of their learning process.

This project will collect a mini-corpus of argumentative essays written by BEL 260 students at the UiTM Sarawak Samarahan Campus. BEL 260 is a course designed to prepare students at Diploma level for their Malaysian University English Test. A corpus is a text database containing language examples collected from real-life writers and speakers.

Once the data has been collected, it will be digitised in computer-readable form. This data can then be analysed using error analysis and lexical analysis software in order to identify, classify and examine the discourse-level and grammatical errors in the writing produced by BEL 260 students.

By examining the errors made by students on the BEL 260 course, it should be possible to gain a detailed picture of those aspects of written English which are causing students the most difficulty. Once such information is known, it should then be possible to provide constructive proposals for improving the teaching of writing on the BEL 260 course and beyond.

#### **Problem Statement:**

Despite having undergone a semester of English language instruction in their BEL 120 course, UiTM Sarawak BEL 260 students still seem to be having difficulties with some aspects of English language writing skills. Students' written performances show a high frequency of basic grammatical and lexical errors, evidencing a lack of grammatical and lexical proficiency.

As well as this, many students do not seem to be able to use complete and structurally complex sentences, based on lecturers' observations of essays submitted in class, nor do

### 5.3 Introduction

This project aims to investigate the English Language errors produced by undergraduate students preparing for the Malaysian University English Test (MUET) in a large public university. Errors are seen as one of the ways in which the **interlanguage** of language learners can be understood. The interlanguage refers to the incomplete knowledge of grammar structures and the strategies used to learn them which are characteristics of language learners of particular stage of their learning process.

This project will collect a mini-corpus of argumentative essays written by BEL260 students at the UiTM Sarawak Samarahan Campus. BEL260 is a course designed to prepare students at Diploma level for their Malaysian University English Test that they will take before the complete their Diploma programme. A corpus is a text database containing language examples collected from real-life writers and speakers.

Once the data has been collected, it will be digitized in computer readable form. This data can then be analysed using error analysis and lexical analysis software in order to identify, classify and examine the discourse-level and grammatical errors in the writing produced by BEL260 students.

By examining the errors made by students on the BEL260 course, it should be possible to gain a detailed picture of those aspects of written English which are causing students the most difficulty. Once such information is known, it should then be possible to provide constructive proposals for improving the teaching of writing in the BEL260 course and beyond.

## 5.4 Brief Literature Review

A computer corpus is a computer database containing authentic linguistic examples to be used for research and pedagogical purposes such as teaching and materials creation.

Up to very recently, most research on computer corpora<sup>1</sup> has been carried out using 'native speaker' or 'expert' data, such as texts produced by native English speakers, or native German speakers, etc.<sup>2</sup> For instance, one of the most widely used English native-speaker corpora is the 100-million word British National Corpus (BNC) (Burnard and Aston, 1998).

Recently, however, there has been a growing interest in collecting together the language used by non-native speakers – learners – with a view to answering research questions about second language acquisition, and perhaps improving the way in which languages are taught.

At the forefront of this research effort are such scholars as Prof. Sylviane Granger in Louvain, Belgium, and Prof. Emeritus Geoffrey Leech in Lancaster, UK. Leech (in Granger 1998:xiv) explains the reasoning behind research on learner corpora in the following way:

“...let us suppose that higher education teacher X, in a non-English-speaking country, teaches English to her students every week, and every so often sets them essays to write, or other written tasks in English. Now, instead of returning those essays to students with comments and a sigh of relief, she stores the essays (of course with the students' permission) in her computer, and is gradually building up, week by week, a larger and more representative collection of her students' work. Helped by computer tools ... she can extract data and frequency information from this 'corpus', and can analyse her students' progress as a group in more depth...”

Granger's approach has spawned a large number of recent studies which have set out to build and utilize learner corpora in a wide range of settings. The result of this work is the International Corpus of Learner English (ICLE) which is based in Louvain, Belgium. ICLE is a world-wide learner corpus project that has allowed workers in different countries where English is spoken or taught to construct learner corpora for different purposes (Granger 1998; Granger et. al 2002).

There is a growing interest in learner corpora in Malaysia as evidenced by the MEC (Malaysian English Corpus) begun in the University of Malaya (Knowles et al, 2003, Knowles, 2004), and the CALES (Corpus-based Archive of Learner English in Sabah/Sarawak) which is almost complete at UiTM Sarawak (Botley, Dealwis, Metom and Izza, 2005, Botley, Metom and Dillah, 2007).

Learner corpus data can be used for many purposes, such as:

- Designing and analyzing corpora and developing computer tools for exploiting them, such as Meunier, 1998.
- Studying in detail various aspects of English grammar, lexis and discourse (Ringbom, 1998; Granger and Rayson, 1998, Petch-Tyson, 1998, Botley et. al, 2005, 2007)
- Studying in detail the 'interlanguage' observed among learners of English as a second language (Altenberg, 2002; Housen, 2002 and Aarts and Granger, 1998)

---

<sup>1</sup> The Latin plural of 'corpus' is 'corpora'. This is widely used by corpus linguists rather than 'corpuses'.

<sup>2</sup> See McEnery and Wilson (2001) for an excellent introduction to the general field of corpus linguistics.