A CORPUS-BASED ARCHIVE OF LEARNER ENGLISH IN SABAH/SARAWAK (CALES PHASE 2)

INSTITUTE OF RESEARCH, DEVELOPMENT
AND COMMERCIALISATION,
UNIVERSITI TEKNOLOGI MARA,
40450 SHAH ALAM,
SELANGOR, MALAYSIA

PREPARED BY:

PROF. MADYA DR. SIMON BOTLEY
@ FAIZAL HAKIM (KETUA)
PUAN LILLY METOM
PUAN DOREEN DILLAH

8th MAY 2007

8th May, 2007

Prof. Dr. Azni Ahmad
Assistant Vice-Chancellor (Research)
Institute of Research, Development and Commercialisation
Universiti Teknologi MARA
40450 Shah Alam
Selangor
Malaysia

Dear Prof. Dr. Azni,

REF: FINAL REPORT FOR PROJECT: "A CORPUS-BASED ARCHIVE OF LEARNER ENGLISH IN SABAH/SARAWAK (CALES PHASE 2)"

With reference to the above, we hereby submit three (3) copies of the final report of this project, as required by URDC regulations. We sincerely hope that it meets with your approval, and we are sincerely sorry for the late submission of this report.

Thank you.

Yours faithfully

Prof. Madya Dr. Simon Botley @ Faizal Hakim
Project Head

TABLE OF CONTENTS

PAGE

ABSTRACT

This report describes the second phase of the learner corpus project called CALES (Corpus-based Archive of Learner English in Sarawak). The original project collected 89,000 words of learner writing in the form of argumentative essays written by students taking English proficiency courses in UiTM's Sarawak Branch Campus (Botley et al, 2005). This new project has increased this total to over 356,000 words of digital text, and has collected essays from three different institutions in order to further expand and enrich the corpus.

The project follows the methodological principles laid down by the International Corpus of Learner English (ICLE) project in Belgium (Granger et. al., 2002). The data was digitised and analysed in order to investigate different types of language error. A number of observations were made concerning some prominent error categories in the data, and their pedagogical implications were explored.

It is hoped that these findings will further contribute to our understanding of the way in which Malaysian learners of English actually perform in their writing. Also, it is hoped that the outcomes of this project will form a foundation for a larger-scale ongoing corpus-building enterprise in the future.

CHAPTER 1: INTRODUCTION

1.2 Research Background and Problem Statement

Educators in EFL (English as a Foreign Language) are all too aware of the errors, or performance features[1] frequently found in writing produced by students of English. However, EFL educators are often unable to make full use of the information revealed by such features in order to help students to improve their language performance. One reason for this is a lack of reliable and permanent examples taken from real student texts. Such examples could then be used as a source of reference to help teachers predict the features students may display in their writing and speech, and then do something about them in a systematic and principled manner.

At the moment, most EFL educators rely upon their professional experience to predict what kinds of features will be displayed in the writing of a particular non-native-speaker group. For instance, it is widely known that Malaysian learners of English regularly under-use the definite article, and turn non-countable nouns onto countable ones (*a staff*, rather than *a member of staff*).

Errors such as these may be traced back to the L1 which in most cases in Malaysia is Bahasa Melayu, a language which does not have a system of definite and indefinite articles, and in which the notion of countability is somewhat different to that in English (see Botley, Haykal and Monaliza, 2005 for a recent discussion of this issue).

[1] Here, we prefer the term 'performance features', because common terms such as 'errors' or 'mistakes' can be considered judgemental and prescriptive. Furthermore, see the section on definition of terms below.