

CORPUS LINGUISTICS BASED ERROR ANALYSIS OF FIRST YEAR UNIVERSITI TEKNOLOGI MALAYSIA STUDENTS' WRITING

Sarimah Shamsudin & Nurul Ros Adira Mahady
Fakulti Pendidikan,
Universiti Teknologi Malaysia

ABSTRACT : The ability to write in English among Malaysian university students is generally not at the most satisfactory level although English is considered as a second language. There has been a growing research interest in the analysis of errors students make in their English writing. The purpose of this study is to identify the errors made by first year UTM students in their writing. Besides that, this study also seeks to find out how much students know about the errors that they produce in writing besides investigating how they react towards these errors. For this study, 66 questionnaires were distributed to first year UTM students from the Faculty of Mechanical Engineering and the Faculty of Civil Engineering. Besides that, students' samples of paragraph were also used to collect the intended data. Findings from the study show that from the 66 paragraph samples analyzed, a total of 1202 errors were found and then tagged according to the types of error. Besides that, findings from the questionnaire show that many of the students are not sure about their English proficiency level and most of them agreed that they would like to improve their English writing by addressing the errors that they make. The paper concludes with the overall summary of the study, limitations of the study as well as the pedagogical implications of the study.

ABSTRAK : Kemahiran penulisan dalam bahasa Inggeris di kalangan pelajar universiti di Malaysia secara amnya tidak berada di tahap yang paling memuaskan walaupun bahasa Inggeris dianggap sebagai bahasa kedua. Terdapat minat yang tinggi di kalangan penyelidik bahasa untuk menjalankan kajian yang melibatkan analisis kesalahan bahasa yang dihasilkan pelajar dalam penulisan bahasa Inggeris mereka. Tujuan kajian ini adalah untuk mengenalpasti kesalahan-kesalahan yang dihasilkan oleh para pelajar UTM tahun pertama di dalam penulisan mereka. Selain itu, kajian ini juga bertujuan untuk mengenalpasti sejauh mana para pelajar tahu tentang kesalahan yang dilakukan dalam penulisan mereka, di samping menyelidik bagaimana reaksi mereka terhadap kesalahan tersebut. Untuk kajian ini, 66 kertas soal selidik telah diedarkan kepada pelajar UTM tahun pertama dari Fakulti Kejuruteraan Mekanikal dan Fakulti Kejuruteraan Awam. Selain itu, sampel perenggan pelajar juga digunakan untuk mengumpul data yang dikehendaki. Dapatan kajian menunjukkan bahawa daripada 66 sampel perenggan yang dianalisa, terdapat 1202 kesalahan yang dikenalpasti dan dilabel mengikut jenis kesalahan. Selain itu, dapatan kajian daripada soalselidik menunjukkan ramai pelajar yang tidak pasti tentang tahap kemahiran bahasa Inggeris mereka dan kebanyakan daripada pelajar tersebut bersetuju bahawa mereka ingin memantapkan penulisan bahasa Inggeris mereka dengan menangani kesalahan yang dilakukan. Laporan ini diakhiri dengan kesimpulan keseluruhan kajian, kekangan kajian dan implikasi pedagogi kajian.

Katakunci : *ability to write in English, Faculty of Mechanical Engineering, Faculty of Civil Engineering*

PENGENALAN

The English language has long been incorporated into the Malaysian school curriculum as a compulsory subject. Besides that, in 2002 the government announced that the teaching of Science and Mathematics would be done in English starting from 2003. Therefore, it is a must for every citizen to familiarize himself to the English language starting from at least 7 years of age. This is backed by a growing number of English exposure including media coverage and English in written forms. This further increases the importance of English language proficiency among students.

However, there are still a large number of university students who do not get a good grasp of the English language as they enter university. The problem is revealed when broken English is used by university students in their academic writing. There have been reports and complaints by many on the poor command of English, for example in a newspaper article, a veteran English teacher puts forward that even intelligent young graduates too have trouble getting ideas across in English. Besides that, there was a study done by Marzuki and Zainal (2004) on errors produced by students in their report writing in an examination situation, in which their study found that students produce grammatical errors, which resulted in the production of flawed sentences in their paragraph development. The findings also showed that students produced grammatical errors and errors related to vocabulary. Another study on error analysis is by Ishak (2002) in which she analyzed errors in the descriptive and narrative writing of form two secondary school students. Her study has also shown that errors are unavoidable in the writing of essay. Therefore, in relation to these studies, the lack of English language proficiency among UTM students is a serious matter and must be researched into as it somehow reflects the English proficiency level of university students in Malaysia.

STATEMENT OF PROBLEM

Students are prone to make errors while writing. This is either because they do not realize the errors that they make, or do not know how to correct them. Therefore, the errors made would be repeated as they do not deem their mistake as a digression from the right way of doing it or unable to correct the errors.

OBJECTIVES OF STUDY

The objectives of the study are as follows:

- 1 To examine the frequency of errors made by first year UTM students in their writing.
- 2 To detect the awareness of first year UTM students on errors that they make in writing.
- 3 To recognize first year UTM students' response and perception towards the errors that they make in writing.

SIGNIFICANCE OF STUDY

There have been a number of studies done on error analysis in Malaysia. However, there is hardly any study done using the corpus linguistics based error analysis approach. This study will benefit teachers of ESL in Malaysia, where they are able to detect and understand students'

mistakes, and later have the initiative to produce the teaching materials that is remedial to the spotted weaknesses. Error analysis is useful in second language learning because this will reveal the problematic areas that students face in their writing, and the errors could be focused on. As for the students, this study will benefit them in the sense that they will gain insight into their own errors that they usually make while writing. The realization of self-made errors will make them aware of these errors in the next writings.

SCOPE OF STUDY

This study is intended to cover the scope within the context of UTM, mainly because it was conducted on the subject of first year UTM students only. As for the analysis of errors, a learner corpus of the subjects' writing of paragraph will be analyzed for grammatical and lexical errors only, and not other types of errors relating to structure or semantics. Other than that, learner corpus is collected from the writing of paragraph, and therefore does not cover long texts such as essays or reports. As for the corpus software utilized to analyze data, not all of the features it provides are utilized for the purpose of this study.

RESEARCH METHODOLOGY

Research Design : This research utilizes the samples of paragraphs written by students with a topic that is predetermined. Besides that, the survey research method is also applied in which data will be collected from students using questionnaires that consist of structured and open-ended items. The rationale of adopting this method is because it is able to provide efficient data collection over a broad population.

Research Population and Sample : The group of subject chosen for the study are 66 first year UTM students from the Faculty of Mechanical Engineering and the Faculty of Civil Engineering, UTM. These students serve as a sample representing the population of first year university students from the faculties of Mechanical Engineering and Civil Engineering. The sampling type is a clustered and probability type of sampling, that is random sampling unit based on faculty grouping. Other attributes such as gender, age, race and socio-economic status are not given emphasis. The reason for choosing first year students is most of them have just recently entered the university from school or pre-university level, therefore their English writing would change (e.g: the academic research writing which has not been learnt yet). Besides that, the outcome of the research is possibly beneficial for the design of English teaching material for first year students.

Research Instruments

Questionnaires

The instrument used to collect data is questionnaire. A set of questionnaire containing a total of 29 questions, which is divided into four sections, was developed based on the objectives and purposes of the research. The questionnaire was adapted from the study on error analysis by Ishak (2002). Structured questions were utilized including the Likert scale to determine the degree of respond. Using the Likert scale, options would be from one extreme (strongly disagree) to another extreme (strongly agree) with a not sure option in the middle. Open-ended questions were also used in the questionnaire. The different sections or parts of the questionnaire are:

- Part A : Demographic information
- Part B : Awareness of English writing proficiency
- Part C : Awareness on errors made in writing
- Part D : Recommendation

Students' paragraphs

To collect data, the main material used is the first hand material, which is students' writing. The writing is in the form of paragraphs, and are produced by students for the writing section of the English for Academic Communication (UHB1412) course, in particular the paragraph writing assignment. These texts will be transferred to digital form, and the data will be derived from the machine-readable texts to search for the corpus of errors made by first year students. The samples of paragraph for the pilot study was argumentative writing, while for the real study, the samples of paragraphs were that of narrative writing, entitled "The problems that I encountered during my first month in UTM".

FINDINGS

Based on the students' responses, we could see that most of the respondents can be classified under intermediate level of English language proficiency. Although most of them scored 4B and above for their SPM English paper, the grading given by the British examination board tells us that not all of these high scorers in the SPM English paper are as proficient as the grades suggest. This is further supported by their MUET scores, where most of the students obtained a Band 3. Based on the MUET grading scale, Band 3 scorers are termed as "modest user", where their command in the language is moderate. Therefore, we are looking at a dispersion of students who are mostly able to use appropriate language but with noticeable inaccuracies. A small group of students belong to the low level proficiency, and the remaining few could be categorized under high level proficiency.

Most of the respondents answered "not sure" for all of the items. This shows that most of the students do not give too much thought to their English proficiency level. However, some items were more inclined to be agreed with. Almost half the students (48.48%) agreed that people can understand the English that they write. More than two thirds of them (36.36%) answered that they are aware of their English writing proficiency. As for the other items, there is an inclination for students to disagree and strongly disagree with the statements, such as writing English for personal reasons (40.91%), which somehow reflects whether students are good enough in the language to be using it for other purposes than that for the academics. There is also a major percentage of disagreement and strong disagreement for the statement professing their English writing is good with the total of 33.33% indicator. Although, 22.73% of students think that they are able to write well. Next, a total more than half or 42.43% of students say that they have not been told by other people that their English is good. This would further confirm their perception towards their English proficiency. This is further proved by the 43.94% of students who disagree and strongly disagree that they always score high marks for their English writings. An overall review of Part B in the questionnaire shows that many students do have some idea whether they are good in writing English or not, although most of them answered that they were not sure for most items.

An outstandingly high percentage of errors made on tenses that is 257 errors or 21.38% out of the total errors. This is in congruent with students' perception of the type of error that they make in writing. This shows that while students do know the type of error that they usually make, they would still probably repeat the same type of error in their next writing. This trend in their writing raises the question of why students would make the same errors again even when they know they are actually making mistakes in their writing. This is most probably because students do not know how to repair these errors even if they want to, as what they have answered in Part B of the questionnaire. The high number of errors in tenses is also because of the type of writing which is the narrative writing. As narrative writing is a writing that tells a story, therefore there would be more need for the use of tenses compared to other types of writing.

The next highest type of error is verb where a total of 168 errors were registered, making up 13.98% of the occurrence or error. Surprisingly, this also matches the second highest type of error as students think they make in writing. As narrative writing would also involve the uses of verbs, there is a lot of this type of errors found in their writing as they are used comparatively more than other types of grammatical item. The third highest tagged error is word choice or lexical, where from the discussion above, we can see that respondents generally mix up their native language with the second language and therefore this becomes the third most occurring error in their writing.

To compare these findings with the previous studies, we will first consider the study done by Marzuki and Zainal (2004). In their study where they analyzed the errors of writing reports in an examination situation, they found that the most occurring error is the subject-verb agreement with 95.5% occurrence, followed by verb related errors and errors in using the articles with 89.8% occurrence each, and the third most frequent type of error is singular-plural in nouns with 74.6% occurrence. As for the study done by Ishak (2002) on narrative writing of secondary school students, the most error found is on tenses with 229 errors on tense, 92 errors on article and 88 errors on prepositions. Therefore, findings from this study do not produce similar result as to the study done by Marzuki and Zainal but for Ishak's study, there is a similar result where the writings analyzed in her study also contain the most errors on tenses. As mentioned by Ishak, the narrative writings that she analyzed contained more errors due to the fact that students had to use the past tense of verbs in relating their past experience, same as what respondents of this study uses while writing their paragraphs on narrative writing.

SUMMARY OF STUDY

This study aims to look into the type of errors made by first year UTM students in their writing. In this study, students' paragraphs were tagged for errors which are accumulated later using the corpus linguistics software Monoconc. How much students know about the errors that they produce in writing is also investigated as well as their perception towards the errors that they make using feedback from questionnaires.

From this study, it is found that the first year students in this university were still producing grammatical and lexical errors in their writing. Students' writing in a second language are faced with social and cognitive challenges related to second language acquisition, where CAH and EAH can be used to explain the causes of these errors. From the 66 paragraphs analyzed, a total

of 1202 errors were tagged for errors according to the type of error. The findings show that the types of errors made in the order of most occurring to the least occurring are tenses with 257 errors (21.38%), verbs with 168 errors (13.98%), word choice or lexical with 128 errors (10.65%), spelling with 111 errors (9.23%), prepositions with 106 errors (8.82%), singular/plural with 104 errors (8.65%), article with 86 errors (7.15%), subject-verb agreement with 85 errors (7.07%), pronouns with 79 errors (6.57%), adjective with 33 errors (2.75%), infinitive 'to' with 26 errors (2.16%), adverb with 10 errors (0.83%), and lastly possessives with only 9 errors detected (0.75%).

Findings from the questionnaire show that most students do know the types of errors they make, yet still repeat the same mistakes in their future writing. The results from the questionnaire also reveal that many of the students think they know what type of errors they make in writing and are not sure about their English proficiency level. Most of them agreed that they would like to improve their English writing by addressing the errors that they make. Students most probably have some idea whether they are good or not because they were constantly tested and graded in school before either from examinations, tests, English assignments, etcetera. Although they are aware of their errors, they feel that they are not good enough in English and this is probably why they do not write well in the language. The students also do not know other people's point of view as for most of them English is not used outside the classroom, and therefore they can only know about their English writing from their teachers or grading.

REFERENCE

- Adolphs, S. (2006). *Introducing Electronic Text Analysis*. New York : Routledge.
- Al Hommos, M. D. A. and Dweik, B. S. (2007). *ERIC Patent No. ED458795* Retrieved on August 2008, from <http://www.eric.ed.gov/>
- Anglis, P. J. (1975). *Sentence-combing, error analysis and the teaching of writing*. In Burt, M and Dulay, H. (Eds.) *New Directions in Second Language Learning Teaching and Bilingual Education*. (229-299). Washington, DC: TESOL.
- Bennett, W. A. (1973). An applied linguistic view of errors of syntax. *Audiovisual Language Journal*. 2, 123-125.
- Biber, D., Conrad, S. and Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge : Cambridge University Press.
- Corder, S. P. (1974). *Error Analysis*. In Allen, J. P. B. and Corder, S. P. (Eds.) *Techniques in Applied Linguistics*. (60-72). Oxford: Oxford University Press.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Dulay, H., Burt, M., and Krashen, S. (1982). *Language Two*. Oxford: Oxford University Press.
- Elliott, A. B. (1983). *Errors in English*. Singapore: Singapore University Press.
- Fried, V. (1968). Comparative linguistic analysis in language. *Modern Language Teaching*. 38-46.
- Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.
- Lado, R. (1957). *Linguistics across Cultures*. Ann Arbor: University of Michigan Press.
- Langan, J. (2000). *College Writing Skills. 5th Edition*. Atlantic Community College, UK: McGraw Hill.

- Mackey, W. F. (1965). Bilingual Interference: its Analysis and Measurement. *Journal of Communication*, 15 (4), 239-249.
- Marzuki, Sahirah and Zainal, Zaidah (2004) Common Errors Produced By UTM Students In Report Writing. Universiti Teknologi Malaysia. (Unpublished Project Report.) University Teknologi Malaysia Repository
- Svartvik, J. (Ed.) (1991). *Directions in Corpus Linguistics Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. Berlin : Walter De Gruyter.
- Odlin, T. (1989). *Language Transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.