The Structural Linguistics Patterns of the Written Component of the Malaysian University English Test (MUET)

Manvender Kaur Sarjit Singh[1]*, Sarimah Shamsudin[2], Hishamuddin Isam[3], NaginderKaur[4], Gurmit Singh Pertap Singh[5], Anita Kanestion[6]

[1]School of Education and Modern Languages, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, [2]Language Academy, UTM Kuala Lumpur Campus, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia, [3]School of Education and Modern Languages, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, [4]Universiti Teknologi Mara, Arau, Perlis, Malaysia, [5]Awang Had Salleh Graduate School, Universiti Utara Malaysia, Sintok, Kedah, Malaysia, [6]Awang Had Salleh Graduate School, Universiti Utara Malaysia, Sintok, Kedah, Malaysia. *Email: manvender@uum.edu.my

ABSTRACT

The core of this paper is reflecting actual language use as the "performance" of the language learners. The areas of investigation involves examining the structural linguistics patterns used by the language learners while preparing the essays, in terms of using the part-of-speech (POS) and the sentence level syntactical analysis of the most frequently used POS, which reflects the distributional patterns of the linguistic components used. The methodology applied is fundamental as it tends to investigate the linguistic components in the compiled genre-specific corpus. Computer-based syntactical studies are limited as it requires hard work and long hours in order to key-in the data and then there is the complex analytic method of describing the findings. This paper presents steps taken for corpus compilation and moves identification of the written Malaysian University English Test, in short MUET. It comprises a move analysis and a multidimensional analysis conducted using a compiled representative corpus of MUET essays. As a descriptive and a corpus-based study, it explored the written essays produced by ESL learners in three matriculation colleges in Malaysia. Language and language use are commonly analyzed for competence and performance. Competence is best described as the internalized linguistic knowledge as acquired by the learners while the notion of "performance" is best defined as the external evidence of language competence. This paper proposes structural linguistics investigations such as frequency analyses, sentence level syntactical analyses, distributional patterns of sentence level linguistic structural patterns and subject-verb agreement analyses reflecting the writers' knowledge of applying their grammatical linguistics knowledge into their written output in various different contexts.

Keywords: Education Management, Malaysian University English Test, Performance
JEL Classifications: I21, I23

1. INTRODUCTION

According to the Malaysian Education Blueprint 2013-2025, "the Malaysian education system has come under increased public scrutiny and debate, as parents' expectations rise and employers voice their concern regarding the system's ability to adequately prepare young Malaysians for the challenges of the 21st century" (p. E-1). Among the three specific objectives of the blueprint is; "Understanding the current performances and challenges" which outlines the need to close the achievement gaps (equity). The need to focus future education according to

the actual needs of the learners can be realized by conducting well-placed research within the prospective requirements of its context. This particular research focused on learners' needs to master the written component of the Malaysian University English Test (MUET). The written component of the MUET is a crucial part of the larger examination which test learners' English language proficiency level before entering the tertiary level of education in Malaysia. The corpus-based structural linguistics investigation will be conducted using the computer-assisted corpus analysis (CACA) approach (Manvender, 2014). This approach will be adopted due to its fundamental nature

of intensive exploration of the written texts according to the established strategies used.

Evaluating and accessing the structural linguistics patterns in a written genre is a difficult task. The compilation of a representative corpus accommodates such an analysis. The term "representative corpus" refers to a corpus that is specifically compiled in order to supplement a particular study and is not meant to be used for any other purposes. However, the complicated work of assigning each sentence according to its linguistic constitution is an additional toil leading to limited studies in the area of sentence level linguistics investigations. It can be a time consuming and an expensive task. The present study aims to present an applicable approach that could be used to assist linguistics investigations such as move analysis, part-of-speech (POS) frequency analyses, sentence level syntax analyses, distributional patterns of sentence level linguistic structural patterns and subject-verb agreement analyses. It tends to describe the procedures involved in an uncomplicated method of analysis that highlights the linguistic patterns in the texts as well as provide an opportunity for the researcher to examine the overall structure of the sentences produced. Likewise, this study will demonstrate an investigation into the move structures used in a genre-specific representative corpus compiled using 48 written essays prepared by students in 3 matriculation colleges in Malaysia. However, for the purpose of this article, from the compiled corpus, only essays with Band 5 and 6 were examined.

A structural linguistics analysis is conducted to describe the linguistic features used in the texts and to show how these features are combined and used to accommodate the ultimate communicative purpose of the entire genre. According to Halliday (1994), functional grammar accounts for how language is used in every text. Everything which is written or said "… unfolds in some context of use" (Halliday; 1994. p. xiii). Within a structural linguistics analysis is the move-based analysis which allows speakers of English to comprehend the macro level organization of the linguistic structures in the genre and also have a control over the micro level of linguistic features naturally used in the texts of their chosen disciplines and professions (Swales, 1990; Bhatia, 1993, 2008, 2012; Bhatia et al. (2011), Cheng, 2012; Cheng, 2014). In order to understand the meaning composed in a sentence it is necessary to first organize and process the sentence into meaningful communicative moves and then to analyze the grammatical units as composed in the moves.

1.1. MUET
Malaysian University English Test or MUET for short, which was introduced in 1999, is a pre-requisite assessment for enrolment into various different courses offered in Malaysian public and private universities and colleges. The universities and colleges set different target band scores for different courses offered. In order to graduate from the universities or the colleges, students are required to satisfactorily obtain the required MUET score and are often advised to take the MUET as soon as possible to avoid delay in their graduation. MUET is a test that assess learners' English language proficiency level and is set by the Malaysian Examination Council. There are four components of the MUET assessment: Listening, speaking, reading and writing. Each

component is allocated 45-120 marks, with an aggregate score of 300. The scores are then graded according to six different bands, ranging from band one, which is the lowest score, to band six which is the highest score for the MUET assessment. Each band has an aggregate from 100 for the lowest band to 300 for the highest band. The Writing component is allocated 90 marks and makes up to 30% of the overall marks for the MUET score. Generally tested as Paper Four, the writing component comprises of one summary writing and one composition writing to be completed within one and half hour. The writing component has two compulsory components consisting of composition and information transfer from non-linear texts. Students' have been facing problems while completing the writing component. However, exploratory studies into the writing component of MUET have been scarce. Yusup (2012) conducted an item evaluation of the reading component of the MUET. A study conducted by Hamzah and Abdullah (2009) identified lack of metacognitive learning strategies as the main cause for ESL learners shying away from using English language. Jalaluddin et al. (2009) found that differences in language structures to be one of the reason leading to the problems acquiring a second language such as English language. As far as the writing component of the MUET is concerned, there is yet a single study to emerge. Recently, there have been calls for the integration of genre analysis and corpus-based investigations in order to understand language use and to address the fundamental structures of genres including the written genres. The Computer-assisted Corpus Analysis or CACA for short was developed to assist text analysis (Manvender, 2014).

1.2. CACA
Creating and investigating a corpus has been acknowledged to be a useful technique in order to understand the underlying structural constructs of written texts (Bhatia, 2008, 2012; Bhatia et al. (2011), Cheng, 2012; Cheng, 2014). A corpus-based approach is used to study "real life" language use (McEnery and Wilson, 1996). Biber et al. (1998. p. 4) presented the fundamental characteristics of a corpus-based analysis as being an empirical analysis, analyzing the actual patterns of language use in natural texts; utilizing large and principal collection of natural texts, known as "corpus" as the basis for the analysis; making extensive use of computers for the analysis; and applying both the qualitative and quantitative analytical techniques. Further elaborating the advantages of a corpus-based approach, they identified computer-based corpus analysis as providing consistent and reliable analyses of learner corpus. In addition, according to Biber et al. (1998. p. 4), the goal of corpus-based approach is to report quantitative findings and most of all, to explore the importance of the findings in order to learn the patterns of language being used in real-life context. In order to allow comprehensive descriptions of a collection of texts, it is necessary to use a tool (a corpus) that accommodates such an analysis and also enables a critical discovery of elements that make up the body of the texts.

Compilation of a corpus has always been conducted within a specific purpose and can be a useful tool to provide information related to language use especially to identify and to analyze complex "association patterns;" the term used by Biber and

Finegan (1994. p. 5) to indicate the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features. According to Halliday, a learner corpus consists of structured collections of text specifically compiled for linguistic analysis and they are large and also are representative of a language as a whole (Halliday, 1978). A corpus can be a useful tool to provide information related to language use in a specific discourse community especially to identify and to analyze complex "association patterns;" the term used by Biber and Finegan (1994. p. 5) to indicate the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features.

Although limited, the development of corpus has also received attention in Malaysia. The EMAS corpus, developed by researchers from Universiti Putra Malaysia, consists of untagged and unedited written data by 800 students of primary and secondary schools. An ongoing project, the Malaysian Corpus of English being developed by researchers from Universiti Malaya, is compiling data in the form of written essays by undergraduates of Universiti Malaya.

The multiple potentialities of this approach will include a manual tagging of moves in the corpus and a computer-assisted POS tagging, utilizing a tagger that is available online. Depending on the size of the investigated corpus, the POS tagger can be used online or purchased for a minimum payment. The frequency of the related linguistic constitutions was computed using concordance software that is available online and can be downloaded via internet. This paper will first elaborate the methodology involved in the tagging of the moves and the POS followed by the proposed analyses. Finally, it will discuss the potentials and the limitations of the approach.

2. METHODOLOGY

2.1. Participants
For the purpose of this particular research, random and purposive sampling method is applied to the research design. Purposive sampling is often chosen in qualitative research due to the fact that it allows an extensive scope of issues to be explored (Lincoln and Guba, 1985). Purposive sampling can be very useful when there is a need to reach a targeted sample quickly and when proportional sampling is not a concern. Participants for a purposive sampling are selected based on specific characteristics such as location, gender, race and easy accessibility to data. In this particular research, the participants were selected due to their representativeness of the criteria to be researched upon; the MUET writing component, their score for the written component of the MUET essays and location of the participants. The corpus was compiled using 48 written essays, each with MUET scores between Band 4, 5 and 6. There were 20 essays with a score of Band 4 each, 20 essays with a score of Band 5 each, and 8 essays with a score of Band 6 each. The justification of selecting essays with scores of Band 4, 5 and 6 is to provide insights from good to best MUET essays, as the findings of this particular analysis will be used to support the main research in terms of developing a written framework for the teaching of MUET essays in Malaysia.

2.2. Data Collection
The data for this study was collected through written texts prepared by students who were enrolled in matriculation colleges in 3 states in Malaysia, namely in Kedah, Perlis and Pulau Pinang. The written texts were gathered and used to create a genre-specific representative corpus of the writing component of the MUET. An important point to be made before going further is that in this study, only written texts was used and compiled into a corpus due to the assumption that authentic writing represents language use more closely than speech. The data for the corpus compilation was collected in the following phases of the study.

2.2.1. Phase 1
Access to the data was gained with visits to the selected locations. Written consent letters were provided to the persons-in-charge. Data, in the form of the written texts, was collected from each location.

2.2.2. Phase 2
The written texts collected were used to create a corpus where the written texts were first collected and saved into a folder in the computer. Each document was then converted into the plain text format, using the AVS document converter which can be downloaded online for free. The document was then saved in the Notepad++ 5.9.3 format for easy removal of unnecessary or confidential data. The saved files became the main corpus for the analysis. A specific name was given to the compiled corpus, in order to reflect the written texts and the structural linguistics investigation. Names of the students were removed and each written essay was given a code in order to protect the identity of the author. Specific codes were allocated to the individual content of the corpus, according to the locations of the participants. Subsequently, the compiled corpus was edited in order to conceal the names and colleges of the selected participants. This step was crucial in order to address the assured level of confidentiality of the data gathered. Next, the corpus was saved as a RAW Corpus file in softcopy, to be used for the POS tagging process for the structural linguistics analysis.

2.2.3. Phase 3
In this stage, the RAW Corpus was used to tag each POS used in the texts. The POS tagging was conducted via online CLAWS C7 Tagger available online at http://www.ucrel.lancs.ac.uk/claws/trial.html. The POS tagging was done in the horizontal form, for easier manual texts recognition and to assist the examination of in-text POS used. This step is crucial in the proposed approach as it generates a tagged version of the corpus, to be used in the frequency analysis. For a researcher, this type of tagging may be useful in order to accommodate the analysis of the structural patterns of the linguistic forms. In order to have a clearer picture of the patterns, the sentences in the texts could be further fragmentized according to its POS. The term 'fragmentized' is used to show the breaking up of the linguistic structures of a sentence to its various forms and allocating a partition in the form of asterisks in between the linguistic forms, for example:

Sentence:

However_RR,_, the_AT main_JJ issue_NN1 of_IO BA_NN1 is_ VBZ that_CST it_PPH1 requires_VVZ long_RR computational_ JJ time_NNT1 as_II31 well_II32 as_II33 numerous_JJ computational_JJ processes_NN2 to_TO obtain_VVI a_AT1 good_JJ solution_NN1,_, especially_RR in_II more_RGR complicated_JJ issues_NN2._.

Syntactical fragmentation:

//RR//AT//JJ//VBZ//NN1//IO//NN1//VBZ//CST//PPH1//VVZ// RR//JJ//NNT1//II31//II32//II33//JJ//JJ//NN2//TO//VVI//AT1// JJ///NN1//RR//II//RGR//JJ//NN2//.//
(Text used: CMWC1C6)

The syntactical layout reveals the ways in which words are combined and used in sentences. Thus, it is important to comprehend the codes used for the tagging as it requires an indication of the codes in order to evaluate the syntactic formation of the sentence.

2.2.4. Phase 4
During this phase, the POS tagged corpus was uploaded to the concordance software, namely; AntConc 3.4.4w WINDOWS (2014). The frequency of each POS was computed and tabulated. This particular paper presents the findings from the tabulated data analysis of the POS found in the compiled corpus. However, the reporting is limited only to essays with Band 5 and 6 from the compiled corpus.

3. DATA ANALYSIS

3.1. The CMWC (Corpus of MUET Written Component)
The corpus compiled for this particular research was developed using the essays written by students from 3 matriculation colleges in the Northern states in Malaysia. The students were preparing for the written component of the MUET and the essays gathered for the corpus are part of the preparatory classroom exercises. The essays were assessed by the teachers involved in the MUET preparatory program in the selected matriculation colleges. The teachers who assessed the selected texts were those with more than 5 years of assessing experience. The compiled corpus is named CMWC representing the purpose of corpus compilation for this specific study where C stands for Corpus, M stands for MUET, W stands for Written and C stands for Component. In short, CMWC stands for Corpus of MUET Written Component. The numbers 1, 2, 3 and so on, represents the number of essays in the corpus while M is used to represent Malay, C for Chinese, I for Indians and O for others, followed by the Band score 5 or 6. The corpus is then horizontally tagged.

3.2. The 5 Most Frequently Used POS in CMWC
The frequency computed by AntConc concordance software was analyzed according to the different tags allocated to the different linguistics constituents as used in the CLAWS C7 tagger. For the purpose of this paper, the focused was on the most frequently used

POS, which was found to be the LEXICAL VERBS; the base form, past tense, -ing participle form, past participle and the –s form.

3.3. Frequency of LEXICAL VERBS
According to the CLAWS coding, the LEXICAL VERBS are tagged as VV0 for the base form, VVD for the past tense form, VVG for the -ing particle form, VVN for the past participle form and VVZ for the –s form of the lexical verbs in the corpus. The frequency of the lexical verbs used in the text corpus of CMWC is shown in Table 1, which shows that the most recurring form of lexical verbs used by the students is the past participle form of verbs coded as VVN. The corpus analysis of texts coded as CMWC1C6 shows that the past participle verb form (VVN) is used 37 times, while the –s form of the lexical verbs (VVZ) is used 31 times followed by the -ing participle form (VVG) with 25 occurrences while the base form of lexical form (VV0) appears 16 times in the texts. The least used form of lexical verb in the texts coded as CMWC1C6 is the past tense form (VVD) with only 8 occurrences.

The corpus analysis shows that most of the students who achieved Band 5 and Band 6 in the written component of MUET have high tendency to use the past participle form of the verbs. This findings contradicts with the teaching and learning practices being employed by the matriculation teachers in the selected matriculation colleges, as one of the teachers who was showed the results from the corpus analysis was rather shocked as according to her the students were usually taught to use the base form of the lexical verb and the -ing form of the lexical verb. Teaching of the past participle form of the verbs were not given a precise emphasis as most of the students were found to have difficulties using the past participle forms of verbs due to the grammatical rules that governs its' usage. However, the findings showed that the students who are good at comprehending and using the past participle form of verbs are students who scores Band 5 and 6 in MUET. It was concluded that perhaps these students were equipped with certain grammatical ability that proves to be beyond the grasp of other students, thus reflecting their ability of using the past participle form of verbs accurately.

The findings from the corpus analysis also showed that most of the students avoided using the past tense form of the lexical verb. When asked, the teacher claims that perhaps it is due to the nature of the essay question which may not have required the students to write using the past tense form of verbs.

4. CONCLUSION AND IMPLICATIONS

This study has indicated that various linguistics structural analyses may be accomplished by applying a computer-based

Table 1: Frequency of LEXICAL VERBS in CMWC1C6

Corpus files/POS	VV0	VVD	VVG	VVN	VVZ	TOTAL
CMWC1C6	16	8	25	37	31	117

UCREL CLAWS C7 Tagset coding: VV0: Base form of lexical verb (e.g., give, work), VVD: Past tense of lexical verb (e.g., gave, worked), VVG: -ing particle of lexical verb (e.g., giving, working), VVN: Past participle of lexical verb (e.g., given, worked), VVZ: -s form of lexical verb (e.g., gives, works)

corpus analysis (CACA) method and these investigations do not necessarily have to be time consuming and expensive. This paper has identified an applicable approach that is comprehensive and convenient. Significantly, the present paper presents an approach that is descriptive in nature. The fundamental base of the frequency analysis begins with a corpus compilation followed by the computerized POS tagging. The POS tagging is useful to obtain a general descriptive view of the linguistic constituents in the individual text. The suggested computer-based tagging approach has been demonstrated in an analysis of the frequency of LEXICAL VERBS used in the text corpus, using the computer-based computations and tagging. Applying the similar approach, this paper has also proposed a sentence level syntactic fragmentation suitable to accommodate the analysis of various linguistics constituents and the analysis of the distributional patterns of the linguistic forms in the sentences. However, this approach has some limitations as it is insufficient to highlight the linguistic errors without being manually detected by the researcher. The suggested frequency analysis using the AntConc concordance software does not recognize variations of the codes, in order to compute the total of an indicated linguistic constituent, computing only similar coding. The proposed method of micro level analysis is most useful in a quantitative analysis of a text corpus. However, it is not denied that similar method of analysis may also be applied to a qualitative analysis. In order to validate a quantified analysis, the micro level analysis should be supported with a statistical analysis of the POS. On the other hand, validation of micro level qualitative data analysis could be supported with a selecting and training a human coder. Variation in coding is determined using the Cohen's Kappa calculation in the SPSS software. The findings of the macro and micro level analyses are useful to shed some lights on creating a framework to teach specific courses under the English for specific purposes domain.

REFERENCES

Bhatia, V.K. (1993), Analyzing Genre: Language use in Professional Settings. London: Longman.

Bhatia, V.K. (2008), Genre analysis, ESP and professional practice. English for Specific Purposes. 27(2):161-174.

Bhatia, V. (2012), Critical reflections on genre analysis. Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE), (24), 17-28.

Bhatia, V.K., Laurence, A., Noguchi, J. (2011), ESP in the 21st Century: ESP Theory and Application Today. In the Proceedings of the JACET 50th Commemorative International Convention (JACET 50).

Biber, D., Conrad, S., Reppen, R (1998), Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press.

Biber, D., Finegan, E. (1994), Intra-textual variation within medical research articles. In: Osdtidijk, N., de Haan, P., editors. Corpus-Based Research into Language. Amsterdam, Netherlands: Rodopi.

Cheng, S.W. (2012), That's it for today: Academic lecture closings and the impact of class size. English for Specific Purposes, 31, 234-248.

Cheng, W. (2014), Corpus analyses of professional discourse. In: Bhatia, V., Bremner, S., editors. The Routledge Handbook of Language and Professional Communication. London and New York: Routledge. p13-25.

Halliday, M.A.K. (1994), An Introduction to Functional Grammar (Second edition). London: Edward Arnold.

Halliday, M.A.K. (1978), Language as Social Semiotic the Social Interpretation of Language and Meaning. London: Edward Arnold.

Hamzah, M.S.G., Abdullah, S.K. (2009), Analysis on metacognitive strategies in reading and writing among Malaysian ESL learners in four education institutions. European Journal of Social Sciences, 11(4), 676-683.

Lincoln, Y.S., Guba, E.G. (1985), Naturalistic Inquiry. Beverly Hills, CA: Sage Publications, Inc.

Manvender K. Sarjit. S. (2014), A Corpus-Based Genre Analysis of Quality, Health, Safety and Environment Work Procedures in Malaysian Petroleum Industry. Unpublished thesis. Universiti Teknologi Malaysia, Skudai, Johor, Malaysia.

McEnery, T., Wilson, A. (1996), Corpus Linguistics. Great Britain: Edinburgh University Press.

Swales, J. (1990), Genre Analysis: English in Academic and Research Settings. Cambridge: Cambridge University Press.

Yusup, R.B. (2012), Item evaluation of the reading test of the Malaysian University English Test (MUET). Masters by Coursework & Shorter Thesis, Melbourne Graduate School of Education, The University of Melbourne.