

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

Manvender Kaur Sarjit Singh<sup>a\*</sup>, Sarimah Shamsudin<sup>b</sup>, Hishamuddin Isam<sup>a</sup>,  
NaginderKaur<sup>c</sup>, Gurmit Singh Pertap Singh<sup>d</sup>, Anita Kanestion<sup>d</sup>

\* Corresponding author: Manvender Kaur Sarjit Singh, manvender@uum.edu.my

<sup>a</sup>School of Education and Modern Languages, Universiti Utara Malaysia, Sintok, manvender@uum.edu.my

<sup>b</sup>Language Academy, UTM Kuala Lumpur Campus, Universiti Teknologi Malaysia

<sup>c</sup>Universiti Teknologi Mara, Arau, Perlis

<sup>d</sup>Awang Had Salleh Graduate School, Universiti Utara Malaysia, Sintok

### Abstract

<http://dx.doi.org/10.15405/epsbs.2016.08.57>

The need to focus future education according to the actual needs of the learners can be realized by conducting a well-placed research within the prospective requirements of its context. However, evaluating and accessing the actual linguistics needs of learners is undeniably a challenging undertaking. This paper presents steps taken to conduct a descriptive and a corpus-based study. It explored the written essays produced by ESL learners. The areas of investigation involved examining the strategies used by the language learners while preparing the essays, the frequency of part-of-speech (POS) used in the essays and sentence level syntactical analysis of the most frequently used POS. The methodology applied is fundamental as it tends to investigate the linguistic constituents in the compiled corpus. Computer-based syntactical studies are limited as it requires hard work, long hours and complex analytic method of describing the findings. In contrast, this article demonstrated an uncomplicated method of analysis and also encourages the use of existing part-of speech (POS) tagging software available online. The findings show that the students are using mainly the past participle form of the lexical verb (VVN) when preparing the written essays. As recommendations for future research, this paper proposes structural linguistics investigations such as frequency analyses, sentence level syntactical analyses (SLSA), examining the 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 be conducted within various contexts where language is used.

© 2016 Published by Future Academy www.FutureAcademy.org.uk

**Keywords:** Malaysian University English Test (MUET); Computer-Assisted Corpus Analysis (CACA); structural linguistics analysis; corpus analysis; part-of-speech (POS)

## 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” (pp. 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 research focuses on learners’ needs to master the aspects of writings particularly in the *Malaysian University English Test* (hereafter, 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. However, students seem to face difficulty scoring in the written component of the MUET. Table 1 shows the results for the four components of the MUET examination held in November 2014. The table indicates that the score of Band 5 and 6 is the lowest for the written component which is represented by 0.06 (for Band 6) and 3.18 (for Band 5). This is found to be sufficient to justify the need of research investigating the reason behind it. The present study was conducted to highlight the linguistics constituents used by the authors who scored Band 4, 5 and 6. It is crucial to investigate how the essays are written to achieve good scores, in order to form a writing framework to be used to teach the written component of MUET.

**Table 1.** MUET Result for NOVEMBER 2014 (Adapted from Annual Report of Malaysian Examination Council)

Component/ MUET Band	Listening (%)	Speaking (%)	Reading (%)	Writing (%)	Total
6	0.09	0.13	0.37	0.06	0.01
5	2.67	3.78	5.27	3.18	2.01
4	14.11	24.48	16.79	26.86	17.80
3	19.49	48.68	34.43	56.98	49.00
2	37.49	8.85	33.39	11.48	27.88
1	26.15	4.08	9.85	1.43	3.30
Total	100	100	100	100	100

Evaluating and accessing the structural linguistics elements of writings is undeniably a challenging undertaking. It can be expensive and rather time-consuming. Assigning each word in the text according to its linguistics constitution is an additional toil, leading to limited studies in the area of sentence level linguistics investigations. However, using a compiled representative corpus in order to assist the structural analysis has been found to be accommodative, in terms of providing easy access to the linguistics constituents in the texts and allowing sentence level linguistics investigations. The present research presents an applicable approach to examine the linguistics structures and constituents of the written essays produced by Malaysian ESL learners for the MUET. The representative corpus was compiled using 48 written essays prepared by students in 3 matriculation colleges in Malaysia.

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: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;). 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 analyse the grammatical constituents as composed in the moves.

In the present study, the corpus-based structural linguistics investigation was conducted using the computer-assisted corpus analysis (CACA) approach (Manvender, 2014). This approach was adopted due to its fundamental nature of intensive exploration of the written texts according to the established strategies used by the students.

## **2. Malaysian University English Test (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 assesses 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 to 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 percent 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. Rusilah 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, Awal & Bakar (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).

### **3. Computer-Assisted Corpus Analysis (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; Manvender, 2012, 2014). According to McEnery and Wilson (1996), a corpus-based approach to text analysis is fundamentally used to study real life language use. This view was supported by Biber et al. (1998:4) stating that a corpus-based analysis is “an empirical analysis, analysing 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”.

In addition, according to Biber et al. (1998: 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, especially to identify and to analyse complex “association patterns”; the term used by Biber et al. (1994:5) to indicate “the systematic ways in which linguistic features are used in association with other linguistic and non-linguistic features”. According to Biber et al. (1994), among various advantages of a corpus-based approach, is providing consistent and reliable analysis of learner corpus.

The multi-potentiality of a raw corpus includes but not limited to computer-assisted part-of-speech (POS) tagging and hand-tagging of identified moves. As the size of the corpus compiled for the present research is considerably small, POS tagging was conducted using the online trail version of CLAWS tagger and the frequency of the linguistic constituents was computed using the AntConc concordance software that is available online and can be downloaded via internet. The following section elaborates the methodology employed in the research.

## **4. Methodology**

### *4.1. Sampling frame for the corpus compilation*

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 & 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.

#### *4.2. 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 corpus of the writing component of the MUET. The data for the corpus compilation was collected in the following phases of the study:

##### *4.2.1 Phase 1*

Access to the data was gained with visits to the selected locations. Written consent letters was provided. Data, in the form of the written texts was collected from each location.

##### *4.2.2 Phase 2*

The written texts collected were used to create a corpus. The written texts were first collected and saved into a folder in the computer. Next, each document was converted into plain text using the AVS document converter. The new document was saved in the Notepad++ 5.9.3 format for easy removal of unnecessary or confidential data. The saved files represent the raw 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. 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 POS tagging.

##### *4.2.3 Phase 3*

In this stage, the RAW Corpus file was used to tag each part-of-speech (POS) used in the texts. The POS tagging was conducted via online CLAWS C7 Tagger. The POS tagging was done in the horizontal form, for easier manual texts recognition.

##### *4.2.4 Phase 4*

During this phase, the POS tagged corpus was then uploaded to the concordance software; 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.

## 5. Data Analysis

### 5.1. The Corpus of MUET Written Component (CMWC)

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, namely Perlis, Kedah and Pulau Pinang. 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 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 4, 5 or 6. The codes used are representative of the students who wrote the essays.

Data from the tagged CMWC corpus was analysed using to mixed method of data analysis, consisting of both qualitative and the quantitative analyses of the tagged texts. Table 2 shows an example of horizontally tagged texts in the corpus. The corresponding sentences were further examined using the sentence level syntactical analysis (SLSA) in order to understand the textual layout of the POS in the text.

**Table 2.** Horizontal tagging

However <sub>RR</sub> , <sub>,</sub> the <sub>AT</sub> main <sub>JJ</sub> issue <sub>NN1</sub> of <sub>IO</sub> BA <sub>NN1</sub> is <sub>VBZ</sub> that <sub>CST</sub> it <sub>PPH1</sub> requires <sub>VVZ</sub> long <sub>RR</sub> computational <sub>JJ</sub> time <sub>NNT1</sub> as <sub>II31</sub> well <sub>II32</sub> as <sub>II33</sub> numerous <sub>JJ</sub> computational <sub>JJ</sub> processes <sub>NN2</sub> to <sub>TO</sub> obtain <sub>VVI</sub> a <sub>AT1</sub> good <sub>JJ</sub> solution <sub>NN1</sub> , <sub>,</sub> especially <sub>RR</sub> in <sub>II</sub> more <sub>RGR</sub> complicated <sub>JJ</sub> issues <sub>NN2</sub> . <sub>.</sub>
--

(Text used: CMWC1C6)

The SLSA provided evidence related to the word combinations used in the corpus. Understanding the word combination used is deemed necessary in order to understand the general structure of the sentence. For example, the sentence from corpus text coded as CMWC1C6 is further fragmented according to its individual POS, in order to further understand the use of various POS in the sentence;  
Sentence:

However<sub>RR</sub> ,<sub>,</sub> the<sub>AT</sub> main<sub>JJ</sub> issue<sub>NN1</sub> of<sub>IO</sub> BA<sub>NN1</sub> is<sub>VBZ</sub> that<sub>CST</sub> it<sub>PPH1</sub> requires<sub>VVZ</sub> long<sub>RR</sub> computational<sub>JJ</sub> time<sub>NNT1</sub> as<sub>II31</sub> well<sub>II32</sub> as<sub>II33</sub> numerous<sub>JJ</sub> computational<sub>JJ</sub> processes<sub>NN2</sub> to<sub>TO</sub> obtain<sub>VVI</sub> a<sub>AT1</sub> good<sub>JJ</sub> solution<sub>NN1</sub> ,<sub>,</sub> especially<sub>RR</sub> in<sub>II</sub> more<sub>RGR</sub> complicated<sub>JJ</sub> issues<sub>NN2</sub> .<sub>.</sub>

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)

### 5.2 The 5 Most Frequently Used PART-OF-SPEECH (POS) in CMWC

The AntConc concordance software was used to compute the frequency of each of the POS used in the corpus. For the purpose of this paper, the focus was on the most frequently used POS. The most frequently used POS was identified as the forms of the LEXICAL VERBS; the base form, past tense, -ing participle form, past participle and the –s form. Among these forms, the past participle form was found to be most used by the students, while preparing the essays.

### 5.3 Frequency of LEXICAL VERBS

The tags used for the LEXICAL VERBS are; 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. The frequency of the lexical verbs used in the corpus of CMWC is shown in Table 3 below.

As highlighted in Table 2, the most recurring form of lexical verbs used is the past participle form of the lexical verbs (VVN). In the text of the corpus coded as CMWC1C6, 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 and the base form of lexical form (VV0) appearing 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. As shown in Table 3, the frequency analysis indicates that students with higher band scores (Band 4, 5 and 6) usually use past participle and the –s form of the lexical verb compared to the other forms of the lexical verb.

Table 3. Frequency of LEXICAL VERBS in CMWC1C6

FILE/CODE	VV0	VVD	VVG	VVN	VVZ	TOTAL
CMWC1C6	16	8	25	37	31	117
<b>UCREL CLAWS C7 Tagset Coding:</b>						
<b>VV0 - base form of lexical verb (e.g. give, work)</b>						
<b>VVD - past tense of lexical verb (e.g. gave, worked)</b>						
<b>VVG - -ing particle of lexical verb (e.g. giving, working)</b>						
<b>VVN - past participle of lexical verb (e.g. given, worked)</b>						
<b>VVZ - -s form of lexical verb (e.g. gives, works)</b>						

## 6. Discussion and Conclusion

As a conclusion, this paper has presented a reliable and easy method of assessing and examining written texts prepared by students. Using a corpus-based approach to text analysis, various linguistics constituents can be evaluated in terms of use and misuse. Applying a computer-assisted corpus analysis or CACA, for short (Manvender, 2014) to the compiled corpus, was found to be a useful method in terms of understanding and highlighting the use of linguistics constituents by students in their writings. A task which would otherwise be hectic and tedious, with the help of CACA is found to be easily applicable.

From the analysis, it was found that students with good band scores are those who are capable of constructing sentences with a variety of linguistic forms, using suitable linguistics constituents according to the sentence structures. Knowing the linguistics constituents and their functions shows that

the students are aware of the accurate grammatical component and are able to use them appropriately in each sentence constructed. Students with such knowledge tend to overcome the grammatical difficulty faced while writing and elaborating ideas in essays.

The analysis has provided some evidence regarding the use of POS by students who are preparing themselves for the MUET. The findings from this study show that the students are using mainly the past participle form of the lexical verb when preparing the written essays. As teachers, it would be necessary and beneficial to know the students' ability to use POS in order to provide further assistance in learning the POS where required, especially when the students are not aware of the tenses and the forms to use in the writings. The fundamental aim of the present research is to highlight the use of POS in written essays. Relying on the findings from this research, the researchers aim to investigate further in terms of the strategies or moves used by the students when preparing the essays. This will assist the researchers to develop a writing framework necessary for the teaching of genre-specific essays to ESL students in Malaysia.

Similar approach of corpus-based analysis is recommended to be applied in various contexts, in order to understand the fundamental aspects of language use. The findings of such a research may be used to enhance the teaching and learning practices.

This research was conducted with the support of Ministry of Higher Education Malaysia's Research Acculturation Collaborative Effort (RACE) Grant Fasa 3 2015/2017

## References

- Bhatia, V.K. (1993). *Analyzing Genre: Language use in professional settings*. London: Longman.
- Bhatia, V.K. (2012). Critical reflections on genre analysis. *Iberica*, 24, 17-28.
- Bhatia, V. K. (2008). Genre analysis, ESP and professional practice. *English for Specific Purposes*, 27(2), 161-174.
- Bhatia, V. K. (2004). *Worlds of Written Discourse: A Genre-Based View*. London and New York: Continuum.
- Biber, D. et al. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. United Kingdom.
- Biber, D. and Finegan, E. (1994). Intra-textual variation within medical research articles. In N. Ostdtjik & P. de Haan (Eds.). *Corpus-based research into language* (pp. 201-221). Amsterdam, Netherlands: Rodopi.
- 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 Sciences*, 11(4), 676 - 683.
- Halliday, M A K. (1978) *Language as Social Semiotic the Social Interpretation of Language and Meaning* London: Edward Arnold.
- Jalaluddin, N. H., Awal, N. M., & Bakar, K. A. (2009). Linguistics and environment in English language learning: towards the development of quality human capital. *European Journal of Sciences*, 9(4), 627 - 642.
- Lincoln, YS. & Guba, EG. (1985). *Naturalistic Inquiry*. Newbury Park, CA: Sage Publications.
- Malaysian Education Blueprint, 2013-2025. Ministry of Education, Malaysia.
- Manvender Kaur, Yasmin H-Z & Shamsudin (2012). A Computer-Assisted Corpus Analysis (CACA) of Professional Discourse. In *Sino-US English Teaching Journal*, Vol. 9, June 2012.
- Manvender K. Sarjit S. (2014). A Corpus-Based Genre Analysis of the Quality, Health, Safety and Environment work procedures in Malaysian Petroleum Industries. Unpublished PhD Thesis: Universiti Teknologi Malaysia.
- McEnery, T. & Wilson, A. (1996). *Corpus Linguistics*. Edinburgh University Press, Great Britain.
- Rusilah Yusup (2012). Item Evaluation of the Reading Test of the Malaysian University English Test (MUET). Unpublished Master Thesis. University of Melbourne.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.