

RESEARCH ARTICLE

A Corpus-based Analysis of Ten Most Frequently Used Part-of-speech Constituents in Descriptive Writing

Anas Sa'idu Muhammad^{1, *}, Manvender Kaur Sarjit Singh²

¹Department of Nigerian Languages, Faculty of Arts and Islamic Studies, Bayero University, PMB 3011, Kano, Nigeria.

²School of Languages, Civilisation and Philosophy, Awang Had Salleh Graduate School, Universiti Utara Malaysia, 26000 Sintok, Kedah, Malaysia.

ABSTRACT - This study is set to examine variation in the use of Part-of-Speech (POS) constituents in the contents of descriptive writing by Nigerian undergraduates using a corpus-based approach. The objective is to identify the ten most-frequently-used POS constituents they used. It also aims to identify the level of frequency variation and distribution patterns of the POS constituents used based on gender and ethnic groups. This study adopted the EAP theory and the usage-based theory. The method applied is quantitative descriptive research design and the samples include thirty (30) Nigerian undergraduates that were identified via purposive sampling. The corpus is subjected to a computer-assisted corpus analysis (CACA) where 33, 274 words were tagged and compiled. The findings revealed NN, NN1, II, AT, JJ, TO, NN2, RR and CC as the student's ten most frequently-used POS constituents, as they achieve higher frequencies. In terms of gender variation, the findings revealed that males achieved a higher frequency of the words used, as equated with females, even though the level of variation at 0.90% is below the statistical level of significance. Also, in terms of their ethnic groups, no significant level of variation was identified, although the Igbo attained a high frequency of words-usage compared to the Yorùbá and the Hausa. Among the theoretical implications derived in this study are the deployments of the corpus-based, genre-based, EAP and ESP-oriented approaches into Nigerian ESL classroom activities.

ARTICLE HISTORY

Received : 15 August 2022

Revised : 25 September 2023

Accepted : 1 November 2023

Published : 20 November 2023

KEYWORDS

Corpus analysis

Descriptive writing

Distribution pattern

Nigerian students

Part-of-speech constituents

INTRODUCTION

A part-of-speech (POS) in English syntax refers to word classes or rather words used to construct sentences (Mustanoja, 2016). Historically, grammarians have drawn on a broad range of methods to shed light on English POS using different forms of syntactic and phraseological analysis, mostly within the postulation of generative grammar, transformational generative grammar, systemic functional grammar, descriptive grammar and even within binary distinction (content and function words), among others (Chomsky, 1957; Fontaine, 2013; Görlach, et al., 1987; Halliday, 1994; Halliday & Webster, 2009; Leech, 2015; Susan, 2002). However, Nigerian learners of English find difficult the mastery and usage of POS, such as verbs (mostly tenses), prepositions, adjectives, etc., principally in English writing for academic excellence, career advancement and for the specific aim of facilitating research or for global business-oriented pursuits (Egbulonu, 2015). In the Nigerian academic setting, English has existed in the country since the advent of British colonisation, which dated from the missionary era of 1899 to the early 1960s when Nigeria gained its independence (Kamal, 2004; Ojetunde, 2012). Since then, Nigeria has English in existence as a second language (ESL) and as an official one (EOL). On a regular basis, studies have shown that Nigerian ESL learners still face challenges in the use of English, particularly in classifying POS. Broadly, Muhammad (2017) asserted that they first move into English learning for the sake of academic purposes, as they predominantly originate from the native Nigerian ethnic groups with Hausa, Igbo and Yorùbá as the major native languages allowed to operate side-by-side with English.

As the Nigerian language education policy stand, her undergraduates often start learning English writing right from pre-university levels or, in fact, it can be traced to start from the kindergarten level, including the nursery, primary, secondary, colleges and polytechnic levels and extends up to the university as well as post-university levels (Akinoye, 2015; Ojetunde, 2012). Still, Nigerian ESL students often come across challenges in the use of English POS as a subject of classroom-based activity and research. Moreover, it is made clear that the challenges in the use of POS are not restricted to Nigerian ESL learners, but also their counterparts in India (Rogers, 2002), Indonesia (Kalajahi, 2014; Mutiara, 2014; Purwanti, 2013), Malaysia (Akbari, 2009; Singh, 2014), Saudi Arabia (Al-Dubib, 2013; Sawalmeh, 2013), Chinese (He, 2020; Jichun, 2015; Pramoolsook & Qian, 2013) and Koreans (Ryoo, 2013; Zhang, 2013; Zheng & Park, 2013), among others. Therefore, the objective of this study is to identify the ten (10) most frequently-used POS constituents manifesting in the designed corpus (Biber, 2006; Hyland, 2009). It also aims to identify the level of frequency and the distribution patterns of POS constituents based on gender and ethnic groups (Tagliamonte, 2013). Based on these objectives, three research questions were created in order to guide the justification of this study as shown below:

1. What are the ten most-frequently-used POS constituents in the designed and compiled corpus?
2. How does the corpus indicate variation in the length of the words used based on gender and ethnic group?
3. How does the corpus indicate variation in the ten most-frequently-used POS constituents based on gender and ethnic groups?

In the long run, looking from the perspectives of these three research questions, the turning point for this study is based within the mainstream of two theories. First, the EAP theory is situated on the front line of the inventive practice of teaching and learning in ESL situations which oversees issues relating to proficiency, tests, performances, competencies and target language use in writing utilising both or either of the cognitive, behaviourist and pragma-linguistic orientations to the concepts of language learning and acquisition (Flowerdew, 2005; Hyland, 2003; Ibbotson, 2013; Jordan, 2002). Its focal target is to “help students perform well in their academic course, particularly English language learning” (Benesch 2008, p. xvi). The second is the usage-based theory that links up structural, functional, lexical and grammatical levels of use (Boye & Harder, 2012; Bybee, 2009; Ghalebi & Sadighi, 2015). The usage-based theory seeks explanation to “frequency of occurrence, usage pattern, variation and change are taken to provide cognitive representation” (Bybee & Beckner, 2010, p. 827). Therefore, this study adopted the theoretical principles of EAP and Usage-based theories to identify the 10 most-frequently-used POS constituents and the frequency and distribution patterns of the POS manifesting in corpus.

CORPUS LINGUISTICS AND ESL WRITING

The Corpus Linguistics approach studies language with the aid of computers to manipulate and analyse large bodies of language data and greatly affects “the methodological frame of linguistic enquiry” (Tognini-Bonelli, 2001, p.210). The use of computer-based corpus material is currently seen as a very useful methodological tool in linguistics and with the possibility of storing large electronic corpora on computers via concordance techniques as well as other computer-assisted methods (such as frequency analysis). Linguists can now observe characteristics and verify results, which are not apparent only by the use of the researcher’s intuition (Ge, 2015; He, 2020; Ko, 2016; Leung, 2016; Nation, 2001; Reppen, et al., 2002; Sinclair, 2005; Szudarski, 2018; Tognini-Bonelli, 2010; Vaughan & Clancy, 2013). In simple terms, corpus linguistics can be defined as the study of “the compilation and analysis of corpora” (Cheng, 2012, p. 6), which are large collections of “naturally occurring language texts chosen to characterize a state or a variety of language” (Sinclair 1991, p.171). Biber et. al. (1998, p.4) characterized corpus-based analysis as:

- a) it is empirical, analyzing the actual patterns of use in natural texts;
- b) it utilizes a large and principled collection of natural texts, known as a corpus, as the basis for analysis;
- c) it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- d) it depends on both quantitative and qualitative analytical techniques.

ESL writing continues to evolve as a central topic in applied linguistics and continues to remain an area of lively intellectual research and debate. Its complex, multifaceted nature seems to constantly evade adequate description and explanation (Carter, 2015; Indah, 2011). Purwanti (2013) sought to explore the causes of language errors in the descriptive writing of Indonesian ESL learners. She used the purposive sampling technique to attain 30 students of Class VIII from the MTs Amal Sholeh Sumogawe, Getasan. She analysed their writing using the strategy of surface taxonomy and comparative theories. In the findings, she reported 93 per cent of intra-lingual errors and 7 per cent of inter-lingual errors identified in their descriptive writing. Essentially, 39.4 per cent were on omission, 39.4 per cent on sentence misformation, 14.1 per cent on word disordering and the remaining 7.2 per cent in the usage of articles. As such, she concluded errors occurred in their descriptive writing due to their ignorance of grammatical rules and their deficiency in the knowledge of intercultural competence. In another related study by Sawalmeh (2013), his primary concern was to identify the reason that hindered Saudi Arabian undergraduates of the Ha’il University of the awareness of the proper use of linguistic rules (POS classes) in their English writing. Using simple random sampling, he obtained 32 representatives aged 19 and 20 years and assigned them to write on one of these five different topics: (a) Ha’il University campus (b) My city (c) Car accidents (d) Shopping or (e) My favourite season of approximately 150 to 300 words in a period of one hour. He used Corder’s (1967) taxonomy of language error analysis. His findings revealed that the Saudi Arabian English learners faced challenges in the use of verb tense, article, spellings, sentence fragmentation, word order, subject/verb agreement and pronouns.

Crossley et. al. (2013) examined learners’ linguistic competence in the use of cohesive devices using a computer-aided prompt-based technique. They used 701 argumentative writings composed of their samples in response to seven different prompts used in the Scholastic Achievement Test (SAT). While exploring the learners’ linguistic competence, they evaluated argumentative writing using the computational tool, including Coh-Metrix, to calculate lexical scores for word concreteness, word frequency and lexical overlap. Their findings revealed variation in their use of lexical, syntactic and rhetorical features, as they did produce greater variation in the use of nominalisations and abstract items in the argumentation. Thus, the results revealed 282 of the essays signified either low or high lexical overlap prompts, χ^2 (df=1, [n=442] = 29.968, $p < .001$), for an accuracy of 63.8%.

In a more specific explanation, Isma'il et al. (2014) compared i-learn web-based pages (i-LDWR) with the prescribed textbook and learning module in order to analyse the frequency level of competency in descriptive writing. The researchers employed 20 intermediate ESL students (Diploma class) taking English proficiency courses at MARA University of Technology, Malaysia. Basically, Malaysia is a country where English is also a second language (ESL). Thus, for learning descriptive writing, they allowed the participants to use i-LDWR for a period of 8 weeks and it was fruitful. Their findings reveal that the students provided a dynamic description in their writing, as they were able to provide a vivid description since i-LDWR provided them with productive learning sources, which promote their vocabulary awareness, particularly in the use of adjectives. In turn and with its strong commitment to corpus-based analysis, this study developed a small or mini-corpus within the contents of descriptive writing of Nigerian undergraduates to identify the 10 most-frequently-used POS constituents and their variation in terms of gender and ethnic groups manifesting in the corpus.

METHODOLOGY

In this study, descriptive research was utilised to reveal the achieved percentages and the frequency of occurrences in the use and patterning the POS constituents embedded in the corpus (Dörnyei, 2011; Johnson 2013; Leech, 2015). This provided the ample situation of attaining their homogeneous frequencies (percentages) and concordance (tokens). By contrast, the concordance ratios per the number of the occurrences of the total words were processed in the corpus. The percentages with graphic representations of the POS distributional patterns as guided by research questions, were also provided (Flowerdew, 2003; Schneider & Grigonyte, 2016).

Research Location and Sampling Strategy

The location of the present study was in the province of Bayero University, Kano (BUK). BUK is a public university located in Kano state in central Northern Nigeria. Stratified purposive sampling strategy (Creswell, 2014; Tavakoli, 2012) was used to attain 30 representatives' during the second semester of 2014/2015 session. This was done in order to attain sample representatives that were willing to participate in the study and fit in with the research question of this study, i.e., based on their gender (male and female) and ethnic groups (Hausa, Igbo and Yorùbá).

Research Instrument

Five research instruments were used in this study. The first instrument employed in this study was Descriptive Essay Writing (DEW). DEW refers to a continuous and subjective examination and a form of genre, which is often employed on learners to test their level of competence, performance, maturity and awareness of English writing skills (Gultom, 2016; Isma'il et al., 2014). In this context, all the participants were given a duration of one hour, thirty minutes (1hr, 30 minutes) to write a descriptive essay about BUK in terms of its buildings and its road channels, etc. The students were monitored and not allowed to use any source of dictionary information and even smartphones were not allowed. The second instrument used is Microsoft Word Document (MWD) where all the handwriting samples of the essays written by the 30 students were typed-set and saved in a folder. The third instrument used is the AVS document converter where the saved MWD was transferred in plain text. At this stage, all the confidential details of the samples were taken out of the texts, which were again saved in a different folder.

The fourth instrument used is the Online POS Constituent Likelihood Automatic Word-tagging System (CLAWS) free software provided by the University of Lancaster, UK accessible for tagging English texts for grammatical corpus connotations. Here, the plain texts were processed and tagged online using the POS CLAWS7 via a horizontal tagging system to individually tag each of the POS classes manifesting in the contents of the processed descriptive ESL writing. The fifth instrument used is the AntConc 3.4.3.w. software developed by Prof. Lawrence Anthony of the University of Waseda, Japan (Anthony, 2005). It guided this study to identify the tokens and frequency counts of the POS constituents manifesting in the corpus.

Compilation of the Corpus and Data Analysis

The corpus developed and compiled in this study was subjected to Singh's (2014) procedure of Computer-Assisted Corpus Analysis (CACA), which provided a quick, time-saving and easy procedure for corpus compilation. It allowed coding raw data with the guidance of a variety of online software for structural and syntactical linguistic analysis based on their function, class, frequency level and category (Habibi, Salleh & Singh, 2015). By its virtue, the CACA procedure guided the criteria on the coder reliability and the independent coding practices of corpus development and achieved a higher correlation of $K=0.941$ value (Singh, 2014), which makes it easy for replication. The researchers first piloted 16, 637 words (Muhammad & Singh, 2020). Subsequently, the current mini-learner corpus was generated (Beaugrande, 2001; Flowerdew, 2001; Ghadessy, et al., 2001; Henry & Roseberry, 2001; Hyland, 2008; Ragan, 2001) named Corpus of Bayero University (C-BUK) presented in Figure 1:

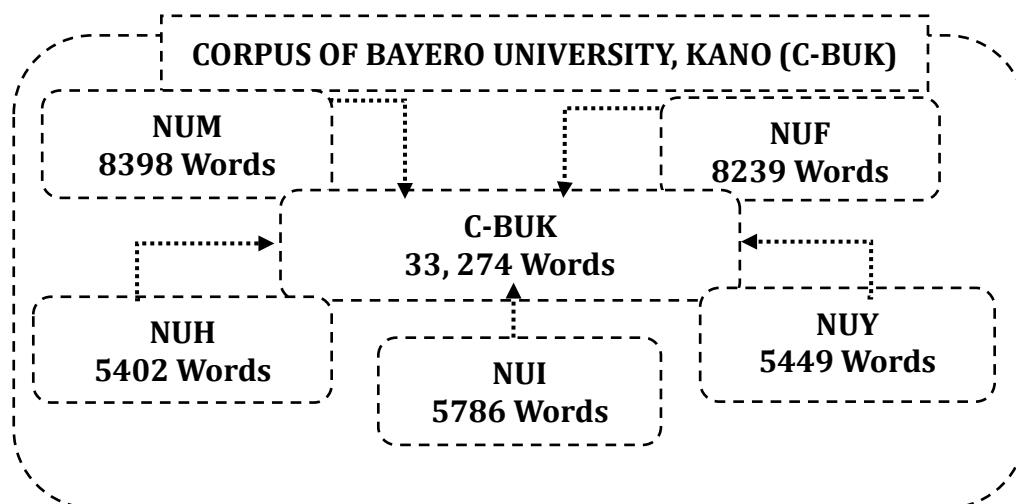


Figure 1. Compilation procedure of the C-BUK adapted from Muhammad and Singh (2020)

The corpus developed and compiled in this study was subjected to Figure 1 illustrates that C-BUK has 33,274 words and represented with five raw frames of sub-corpus demonstrating the working procedures of the corpus. Each sub-corpus is provided with its representative codes. Thus, the Male sub-corpus is tagged NUM, the Female NUF and the Hausa Group NUH, the Igbo Group NUI and then the Yorùbá Group NUY. For illustrative purposes, the C-BUK coded and tagged the grammatical and structural function of the POS constituents according to each sub-corpus, as shown in Figure 2:

The **AT** purpose **NN1** of **IO** the **AT** university **NN1** in **II** a **AT** green **JJ** and **CC** in **II** a **AT** huge **JJ** city **NN1** ._. The **AT** living **JJ** environment **NN1** is **VBZ** a **AT** little **JJ** discomfoting **NN1** because **CS** sometimes **RT** I **PPIS1** wake **VV0** up **II** to **II** creatures **NN2** like **II** snakes **NN2** in **II** my **APPGE** apartment **NN1** is **VBZ** highly **RR** tedious **JJ** and **CC** the **AT** most **RGT** interesting **JJ** or **CC** difficult **JJ** experiences **NN2** ._. I **PPIS1** manege **VV0** to **TO** leave **VVI** Nigeria **NP1** from **II** Aminu **NP1** Kano **NP1** ._. the **AT** university **NN1** made **VVD** me **PPIO1** feel **NN1** better **RRR** about **II** myself **PPX1** ._. It **PPH1** was **VBDZ** later **RRR** I **PPIS1** came **VVD** to **TO** realise **VVI** ._. Unfortunately **RR** ,_, both **DB2** the **AT** students **NN2** ,_, however **RR** ,_, my **APPGE** heart **NN1** starts **VVZ** pounding **NN1** madly **RR** ._. I **PPIS1** am **VBM** never **RR** sure **JJ** anything **PN1** greatest **JJT** nostalgic **JJ** about **II** feelings **NN2** whenever **RRQV** I **PPIS1** got **VVD** very **RG** nervous ._.

Figure 2. Samples of the annotated POS constituents and tag sets built from the corpus adapted from Muhammad and Singh (2020)

Figure 2 shows the nature of POS CLAWS and annotates the grammatical function of the POS constituents. Figure 2 further provides examples of the tagged POS constituents that were discussed in presenting the findings of the present study, which were shown in colours.

FINDINGS

The Most Frequently Used POS in C-BUK

POS as stated earlier, are the traditional word classes of nouns, pronouns, verbs, adjectives, adverbs, prepositions, conjunctions, and interjections. In this section, the overall frequency and concordance of the overall POS constituents manifesting in the contents of the students' descriptive ESL writing was presented using the CLAWS 7 set tag, as presented in Table 1.

Table 1. The 10 most frequently used POS constituents in the corpus per 33,274 words.

POS Codes	Interpretation of the 10 most-frequently-used POS Constituents (Codes)	Concordance (TOKENS)	Frequency (100%)	Count
NN	common noun, neutral for number (e.g. sheep, cod, headquarters)	3598	21.57	1
NN1	singular common noun (e.g. book, girl)	2579	15.46	2
II	general preposition	1417	8.48	3
AT	article (e.g. the, no)	1373	8.23	4
JJ	general adjective	1274	7.63	5
NN2	plural common noun (e.g. books, girls)	960	5.75	6
TO	infinitive marker (to)	959	5.74	7
RR	general adverb	652	3.90	8
CC	coordinating conjunction (e.g. and, or)	517	3.09	9
PPIS1	1st person sing. subjective personal pronoun (I)	473	2.83	10

Table 1 reveals the 10 most-frequently-used POS constituents in the corpus include the NN (Freq=3598= 21.57%), NN1 (Freq=2579= 15.46%), II (Freq=1417= 8.48%), AT (Freq=1373= 8.23%), JJ (Freq=1274= 7.63%), NN2 (Freq=960= 5.75%), TO (Freq=959= 5.74%), RR (Freq=652= 3.90%), and the CC (Freq=517= 3.09%). These results can be further observed in Figure 3.

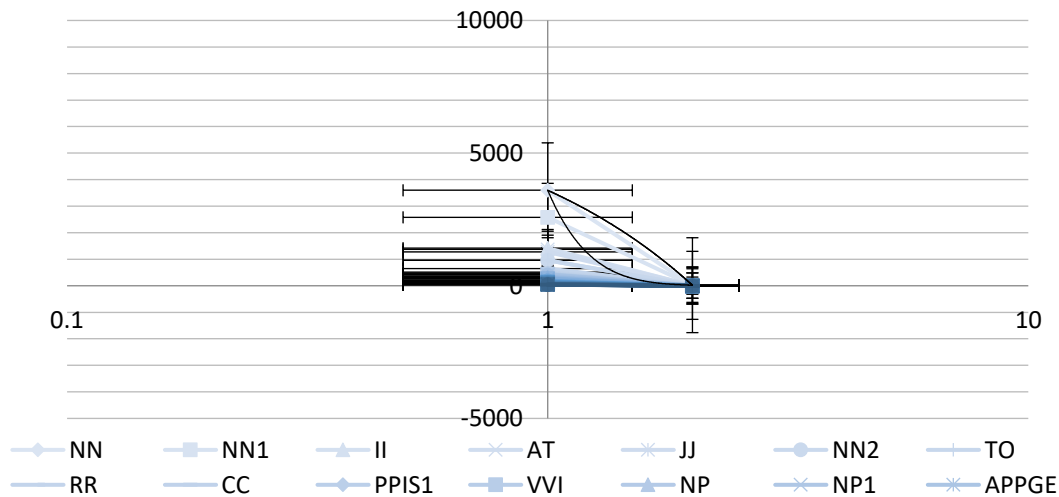


Figure 3. Most frequently used POS in the C-BUK.

The distribution pattern illustrated in Figure 3 shows that the students most frequently use six (06) of the eight (08) grammatical classes of POS, including nouns, pronouns, adverbs, adjectives, prepositions and conjunctions. They seldom use verbs (including tenses) and interjections in grammatical classes. Therefore, these findings are in line with Singh (2014), who argued that ESL learners feel much more at ease in the use of nouns, adverbs, adjectives, pronouns and conjunction POS classes in English writing. She supported her argument with the view that ESL learners do produce accurate and fluent writing, but they seem to be deficient in the use of other types of constituents due to their level of grammatical acquaintance, deficient structural knowledge in the formation of POS in sentences and also their low level of contextual awareness.

Variation in Word Count/Length based on Gender and Ethnic Groups

Word length refers to word count, but “not necessarily number of letters or phonemes or syllables at phonological or graphemic length” (Stefanowitsch, 2020, p.90). In this study, word count was operationalised to reveal the length of linguistic units at a word or constituent level based on their gender and ethnic group categories, as presented in Table 2.

Table 2. Gender and ethnic group variation in most word use per 33,274 words.

Nigerian Undergraduates		N	Tokens	Frequency	Count
Gender	Males	15	8398	50.48%	1
	Females	15	8239	49.52%	3
Ethnic Groups	Igbo	10	5786	34.78%	1
	Yorùbá	10	5449	32.75%	2
	Hausa	10	5402	32.47%	3

Table 2 clearly shows gender variation in the corpus. This stands to the fact that the male students portray a wide use of vocabulary, as they compose lengthy words in the corpus (Freq=8398 =50.48%) compared to the females (Freq=8239 =49.52%). The level of variation is at 0.96%, as shown in Figure 4.

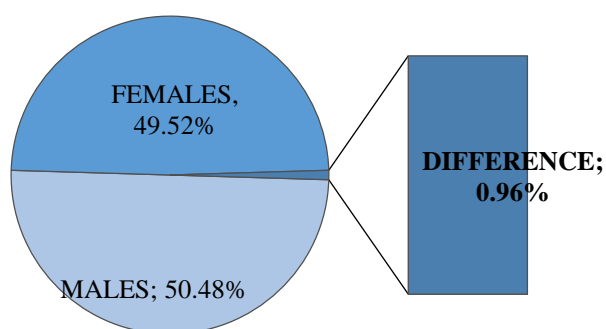


Figure 4. Distribution pattern of the word use based on gender per 33,274 words.

The distribution pattern illustrated in Figure 4 above on word count/length further shows that the males are more creative and show familiarisation with the use of POS in the corpus compared to the females. These results failed to agree with Muhammad and Nair (2015), who perceived that in terms of gender polarity, females generally perform better in ESL writing than males. The results identified have indicated that gender has a peculiar role in ESL writing.

In observing their variation for word count/length based on ethnic groups, the results reveal that the Igbo group achieves high-frequency of word count/length (Freq=5786 =34.78%), followed by the Yorùbá’ (Freq=5449 =32.75%) and then the Hausa group (Freq=5402 =32.47%). The level of variation is at 0.025%. The distribution pattern for word count/length based on ethnic group is further illustrated in Figure 5.

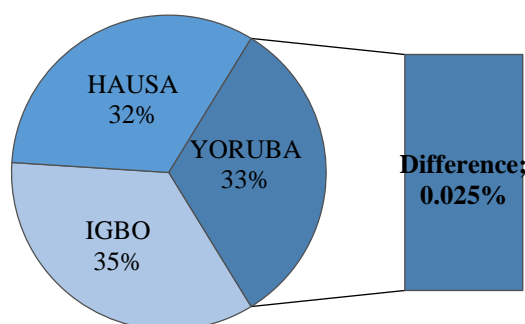


Figure 5. Distribution pattern of the word use based on ethnic groups per 33,274 words.

Figure 5 shows that ethnicity also plays a crucial role in expressing variations in ESL learners' writing, particularly when it comes to justifying their competence, performance, awareness and maturity in ESL writing. These results relate strongly to the assertions made by Kroskrity (1999) and Flowerdew (1993) that ethnolinguistic doctrines and Mother Tongue (MT) influence often limit the competence, performance, maturity and linguistic awareness of structures produced by ESL learners in an activity to guide their communicative events in ESL writing. Also, these results are parallel to Hyland's (2003) argument that ESL learners are influenced by their ethnolinguistic background, as they write and construct ideas in accordance to their nearest cognitive and psycholinguistic concepts. As such, MT influence cannot be easily avoided.

Variation in the Most Frequently Used POS based on Gender

In this section, the concordance (tokens) of the C-BUK was explored to identify the level of variation of the ten most-frequently-used POS constituents based on gender, as presented in Table 3.

Table 3. Variation in 10 most frequently used POS constituents based on gender per 33,274 words.

POS Codes	Female Tokens	Frequency	Male Tokens	Frequency	Count
NN	1846	21.88	1752	21.24	1
NN1	1352	16.02	1227	14.88	2
II	729	8.64	688	8.34	3
AT	686	8.13	687	8.33	4
JJ	647	7.67	627	7.60	5
TO	480	5.69	498	6.04	6
NN2	462	5.47	479	5.80	7
RR	324	3.84	328	3.97	8
PPIS1	259	3.07	259	3.14	9
CC	258	3.05	232	2.81	10

Table 3 shows gender variation in the 10 most frequently used POS. The results reveal that females achieve higher frequency in NN (Freq=1846 =21.88%), NN1 (Freq=1352 =16.02%), II (Freq=729 =8.64%), JJ (Freq=647 =7.67%) and CC (Freq=258 =3.05%). The males achieve high frequency over the females in the use of AT (Freq=687 =8.33%), TO (Freq=498 =6.04%), NN2 (Freq=479 =5.80%) and RR (Freq=328 =3.97%). However, in PPIS1 (Freq=259 =3.14%), they achieve an equal frequency. These results reflect the students' considerable level of competence in the use of POS constituents and reflect their maturity in the structure of clauses/T-Units, as well as their level of awareness in sentence construction based on gender. For this reason, Figure 6 below further illustrates the distribution patterns of the results.



Figure 6. Distribution pattern of 10 most frequently used POS constituents based on gender per 33,274 words.

The distribution pattern illustrated in Figure 6 shows that the females achieve high frequency in 5 (NN, NN1, II, JJ and CC) and the males in 4 (AT, TO, NN2 and RR) and both achieve equal frequency in 1 (PPIS1) of the 10 most-frequently POS used in the corpus. These results reveal that the females annotate mastery in the use of common nouns, singular nouns, general prepositions, general adjectives and coordination conjunctions in the descriptive ESL writing over the males. However, the males most frequently use articles, infinitive marker (to), plural common nouns and general adverbs over the females. These results agree with the affirmation of Labov, who said that “women use more variant devices than men do” (Labov, 1972, p. 243). Also, the results go in accordance with the affirmation made by Leech, et. al. (2001) that ESL learners require cognitive understanding to produce a text that is coherent in order to distribute the POS classes accurately. Likewise, the results are in line with the assertion of Ojetunde (2013) that Nigerian ESL students lack richness in vocabulary and are poor in the use of grammatical items (POS classes). In addition to this, the results reflect the view of Biria and Karimi (2015) that linguistic competence provides ESL students with the core requisite of acquiring recursive achievement. However, gender distinction often indicates their needs and complexity in terms of strategic writing skills.

Variation in 10 Most Frequently Used POS Based on Ethnic Groups

In this section, the concordance (tokens) of the C-BUK corpus was examined to identify the variation of the 10 most frequently used POS based on ethnic groups, as presented in Table 4.

Table 4. Variation in 10 most frequently used POS constituents based on ethnic groups per 33,274 words.

POS Codes	Igbo Tokens	Freq.	Yorùbá Tokens	Freq.	Hausa Tokens	Freq.	Count
NN	1262	21.77	1210	22.17	1126	20.74	1
NN1	902	15.56	856	15.68	821	15.12	2
II	505	8.71	480	8.79	459	8.45	3
AT	464	8.00	453	8.30	429	7.90	4
JJ	432	7.45	433	7.93	409	7.53	5
NN2	343	5.91	335	6.14	341	6.28	6
TO	313	5.40	305	5.59	282	5.19	7
RR	236	4.07	203	3.72	213	3.92	8
PPIS1	167	2.88	118	2.16	166	3.05	9
CC	185	3.19	182	3.33	150	2.76	10

Table 4 illustrates the level of ethnic group variation in the 10 most frequently used POS. Here, the ethnic group variation is interpreted in three different phases. In the first phase, the results reveal that the Igbo ethnic group consistently achieves high frequency in NN (Freq=1262=21.77%), NN1 (Freq=902= 15.56%), II (Freq=505= 8.71%), AT (Freq=464= 8.00%), RR (Freq=236= 4.07%), TO (Freq=313= 5. 40%) and CC (Freq=185= 3.19%) followed by the Yorùbá and then the Hausa. In the second, the results reveal that the Igbo group achieves high frequency in PPIS1 (Freq=167=2.88%), followed by the Hausa (Freq=166=3.05%) and then the Yorùbá (Freq=118= 2.16%). The same happens in NN2. The Igbo achieve a high frequency of 343 (5.91%), followed by the Hausa (Freq=341= 6.28%) and then the Yorùbá (Freq=343= 5.91%). In the third phase, the results reveal that the Yorùbá ethnic group achieves high frequency in JJ (Freq=433= 7.93%), followed by the Igbo (Freq=432= 7.45%) and then the Hausa group (Freq=409=7.53%). These results reflect the students’ considerable level of competence in the use of POS constituents and reflect their maturity in the structure of clauses/T-Units as well as their level of awareness in sentence construction based on ethnic groups. Thus, Figure 7 below further illustrates variation and the distribution patterns of the 10 most frequently used POS based on ethnic groups.

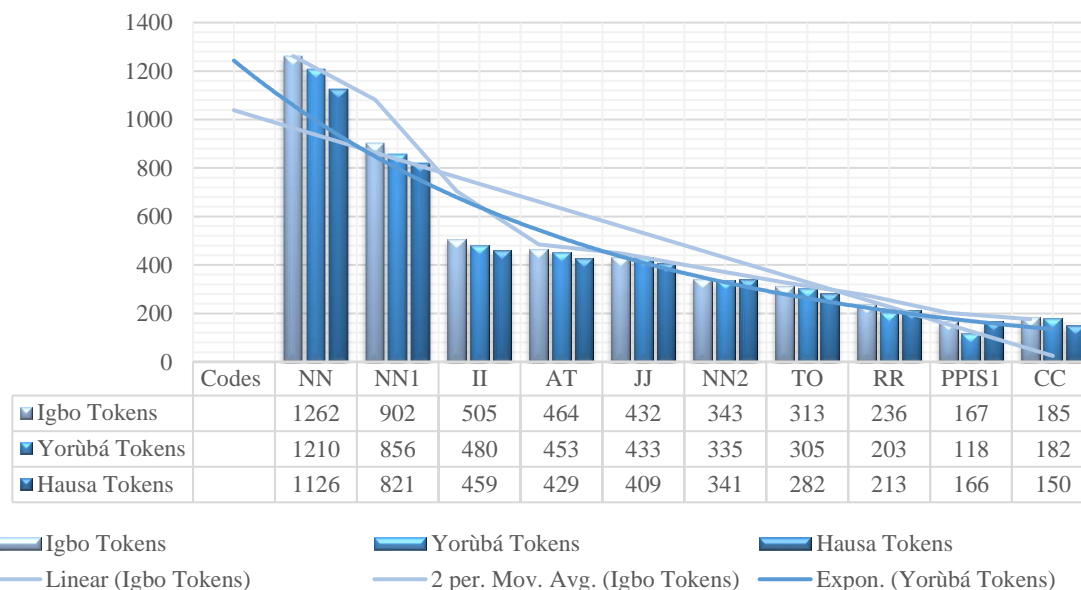


Figure 7. Distribution of 10 most-frequently used POS constituents based on ethnic groups per 33,274 words.

The distribution pattern illustrated in Figure 7 above shows that the Igbo ethnic group achieves high frequency in 9 of 10 POS constituents (including NN, NN1, II, AT, JJ, TO, RR, PPIS1 and CC). By twirl, the NN2 constituent, that is, the plural common noun is the only one that the Yorùbá group outsmarted the Igbo, so it is the one and only POS constituent that the Igbo ethnic group failed to achieve a high frequency in. In other words, these results reveal that the Igbo group annotates mastery in the use of common nouns, singular nouns, general prepositions, general adjectives, articles, infinitive marker (to), general adverb, first-person singular subjective personal pronouns and coordination conjunctions in the corpus over the Yorùbá and the Hausa ethnic groups. These findings could be tied to the assertion made by Egbulonu (2015) that the influx of English among Igbo ethnic groups has been accorded with high relativity. He affirms that Igbo communities use English more than their native language even in social gatherings, which is a rare case when compared to Hausa and Yorùbá ethnic groups. Thus, he fears the extinction of the Igbo language due to their inclination to use English. Hence, he recommended that Igbo should be taught and made a compulsory learning medium and not solely concentrate on improving their linguistic competence in English. Moreover, the findings are parallel to that of Muhammad and Singh (2020) that sociolinguistic factors, including ethnicity, heavily influence the competence of Nigerian ESL students and make them less linguistically competent in using the strategies and variant classes of POS in English writing. In this situation, the learners’ knowledge of English (the appropriate use of POS) often plays a key role in unfolding specific competence in ESL writing.

DISCUSSION

This study utilised a corpus-based approach, descriptive writing, as well as EAP and usage-based theoretical frames to evaluate variation in the Nigerian undergraduate students English writing. Thus, in view of the findings of this study, it is certain that the corpus-based approach is an important human strategic activity for understanding the level of competence, performance, maturity and awareness in Nigerian undergraduate students’ ESL writing. The research findings revealed a common phenomenon of students struggling when it comes to competence, maturity and awareness in the use of appropriate POS constituents in ESL writing. This often has an impact on their academic success. Majority of the students aspire for ESL writing fluency as a universal language that has raised its demand, particularly in various sectors’ and working spheres, such as ministries, media, agricultural, commercial, educational, medical and other

administrative sectors more challenging. In essence, this study calls for improving learners' strategic qualities with the aid of descriptive essays so as to influence their writing expressively and "encourage them to find their own voices to produce writing freshly and spontaneously" (Hyland 2003, p. 8). Basically, the act and process of descriptive writing improve students' cognitive knowledge and idea synthesis.

Muhammad (2017) revealed that descriptive writing is a genre which promotes aptitude and command in structuring sentences, in developing their grammatical functions and in proving the level of their vocabulary, as well as in testifying their inbuilt cognitive awareness in matters relating with the use of cohesive markers while organising their paragraphs, and so on. Learners with efficiency in the use of POS constituents will surely be able to write expressively and communicatively. In view of the findings of this study, it is certain that descriptive writing is an important human strategic activity for comprehending ESL learners' competence in writing. Therefore, the pedagogical contribution provided in this study is the need for EAP course design in order to provide learners with sufficient competence, maturity and awareness in ESL writing among Nigerian undergraduates. To further highlight the significance of EAP, Besnesch (2008, p. 11) outlined the seven steps that pay more attention to assessing ESL/EFL learners' achievement, competence, proficiency, maturity and situational needs in language pedagogy, including:

- a) specifying the learners'
- b) analysing their needs
- c) specifying their enabling objection
- d) identifying their selective or evolving learning materials
- e) identifying their needs for appropriate teaching strategies
- f) exploring their learning activities
- g) evaluating their needs through revision and provision of feedback on the learners' situational needs.

Along these same lines, previous studies have utilised the same EAP theoretical frame to evaluate English learners' writing skills. With this, special attention should be provided in the National Policy on Education (NPE), the Universal Education Board (UBE) and the State Universal Education Board (SUBEB) on intervention protocols concerning the strategic ways of enhancing students' competency in learning English writing. By the virtue of its pedagogical contributions, teachers, lecturers and researchers should pay more attention to exposing learners' awareness of the use of POS constituents in English writing in order for them to achieve a much higher level in terms of discursive competency and for wider knowledge coverage. In view of its theoretical contributions, the results identified in this study have raised our awareness of the level of expertise shown by the Nigerian undergraduates in descriptive writing, which signifies their word count/length and distribution patterns in ESL writing. By virtue of its practical contributions, the findings of this study have shown that through the use of a corpus-based approach, the classes of POS used by the students are easily identified because it provides them with an ample challenge of describing, reporting and even picturing vivid situations in ESL writing. Thus, Yoon and Jo (2014, p. 96) affirmed that studies in corpus linguistics have demonstrated the benefits of "corpora not only as linguistic a resource to improve students writing abilities but also as a cognitive tool to develop their learning skills and strategies".

CONCLUSION AND RECOMMENDATION

This study laid a basic ground for using descriptive writing in understanding the 10 most-frequently-used POS constituents in word count/length, distribution patterns, gender and ethnic group variation of the Nigerian undergraduates using a corpus-based approach. In view of the findings of this study, it is certain that the corpus-based approach is an important human strategic activity for understanding the level of competence, performance, maturity and awareness in students' ESL writing. The findings revealed that in the compiled C-BUK corpus the students most frequently used POS constituents in the form of the common noun neutral for number (NN), the singular common noun (NN1), the general proposition (II), the article (AT), the general adjective (JJ), the plural common noun (NN2), the (to) infinitive marker (TO), the general adverb (RR), the first person singular subjective personal pronoun (I), and the (e.g., and, or) as coordinating conjunction (CC). These results from the corpus belong to six (06) out of the eight (08) POS classes of nouns, pronouns, adverbs, adjectives, prepositions and conjunctions. Thus, the results indicate that they seldom use only the grammatical classes of verbs (including tenses) and interjections in the corpus. Moreover, the findings of this study revealed that the males achieve the highest frequency of word count/length equated to the females at a variation of a 0.90% level of difference. Then the ethnic groups' variation in word count/length revealed that the Igbo group attained the highest frequency followed by the Yorùbá and the Hausa groups and the variation is at a 0.025% level of difference. In the event of identifying variation in the 10 most-frequently-used POS based on gender, the females achieve high frequency in 5 (NN, NN1, II, JJ, and CC), the males in 4 (AT, TO, NN2, and RR) and both achieve equal frequency in 1 (PPIS1). Then in the ethnic group variation in the 10 most frequently used POS, the findings revealed that the Igbo ethnic group achieve the highest frequency in 9 of 10 POS constituents (NN, NN1, II, AT, JJ, TO, RR, PPIS1 and CC). By twirl, the NN2 constituent, that is, the plural common noun is the only POS constituent that the Yorùbá group outsmarted the Igbo ethnic group.

Part of the limitations of this study is that it could only adopt a descriptive type of research design to reveal the research findings using frequency counts and percentages. Future studies can use mixed-method, qualitative (such as syntactic parser), quantitative (such as Chi-Square, T-Test, ANOVA, MANOVA, correlational and other forms of inferential statistics) approaches to investigate Nigerian ESL student's usage of POS via corpus linguistics. In addition, this study could only use thirty (30) samples of students' descriptive writing using stratified purposive sampling. Future studies can use larger samples of written essays (such as the argumentative, narrative and the expository, among other forms of writing) and/or other forms of spoken and written discourse (such as newspapers, novels, fiction, legal documents, among other genres) to develop a larger corpus for linguistic analysis. Moreover, the corpus compilation procedure of this study is subjected to the CACA approach using POS CLAWS online software and AntConc software. Future studies can be conducted using different software, such as ConcGrams, W-Matrix, LANUS software, PAWS software, WordSmith Tools, and so on, for linguistic analysis among Nigerian ESL learners.

Likewise, this study could only limit its analysis to BUK undergraduates. Future studies can be made among other undergraduate students studying at various Nigerian institutions, such as Ahmadu Bello University, Zaria (ABU), University of Ibadan (UI), University of Calabar, Cross River (Unical), Usmanu Danfodiyo University, Sokoto (UDUS), Obafemi Awolowo University, Ile-Ife (OAU), University of Maiduguri (UNIMAID), Federal University Gusau (FUGUS), Lagos State University (LASU), University of Nigeria, Nsukka (UNN), to mention but a few. Among the theoretical and practical implications derived in this study are the deployments of the corpus-based and genre-based oriented approaches into the Nigerian ESL classroom activities. Fundamentally, pedagogical implications call for the organisation of writing seminars or workshops, extensive reading skills and the adoption of situated learning methods as some of the best ways of enhancing Nigerian undergraduates' ESL writing.

In view of the findings presented in this study, special attention should be provided to the National Policy on Education (NPE), the Universal Education Board (UBE) and the State Universal Education Board (SUBEB) for intervention protocols concerning strategic ways of enhancing students' competency in learning English writing. In view of this, we recommend that teachers, lecturers and researchers should pay more attention to exposing learners' awareness of the use of other types of linguistic constituents in their English writing in order for them to achieve a much higher level in terms of discursive competency and for wider knowledge coverage.

Among the theoretical implications derived in this study are the deployments of the corpus-based and genre-based oriented approaches into the Nigerian ESL classroom activities. As such, pedagogical implications call for the organisation of writing seminars or workshops, extensive reading skills and the adoption of situated learning methods as some of the best ways of enhancing Nigerian undergraduates' ESL writing.

REFERENCES

- Akbari, O. (2009). *A corpus based study on Malaysian ESL learners' use of phrasal verbs in narrative compositions* [Unpublished PhD Thesis]. University Putra, Selangor, Malaysia.
- Akinyeye, C.M. (2015). *Exploring the teaching and learning of English (L2) writing: A case of three junior secondary schools in Nigeria* [Unpublished PhD Thesis]. Faculty of Education, University of the Western Cape, South Africa.
- Al-Dubib, D.A. (2013). *Error analysis of subject-verb agreement in the writing of EFL Saudi female students': A corpus-based study* [Unpublished M.A. Dissertation]. Al-Imam Mohammad Ibn Saud Islamic University, Saudi-Arabia.
- Anthony, L. (2005). AntConc: Design and development of a freeware corpus analysis toolkit for the technical writing classroom. In IPCC 2005. Proceedings of International Professional Communication Conference (p.729-737). IEEE.
- Beaugrande, R.D. (2001). Large corpora, small corpora, and the learning of "language". In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 3-30). John Benjamins Publishing.
- Benesch, S. (2008). *Critical English for academic purposes: Theory, politics and practice*. Taylor and Francis e-Library Publishing.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins Publishing.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use*. Cambridge University Press.
- Biria, R. & Karimi, Z. (2015). Contributory role of pre-task planning in improving Iranian EFL learners' writing of argumentative essay: The case of complexity and accuracy. *Theory and Practice in Language Studies*, 5(5), 1026-1036.

- Boye, K. & Harder, P. (2012). A usage-based theory of grammatical status and grammaticalization. *Language*, 1-44.
- Bybee, J. (2009). Language universals and usage-based theory. In M.H. Christiansen, C. Collins & S. Edelman (Eds.), *Language Universals* (pp. 17-39). Oxford University Press.
- Bybee, J. L. & Beckner, C. (2010). Usage-based theory. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp.827-856). Oxford University Press.
- Carter, R. (2015). Teaching descriptive writing through visualization and the five senses. *English Teaching Forum*, 53(28), 37-40.
- Cheng, W. (2012). *Exploring Corpus Linguistics: Language in action*. Routledge Publishing.
- Chomsky, N. (1957). *Syntactic structures*. Moulton Publishing.
- Creswell, J.W. (2014). *Educational research: Planning, conducting and evaluating quantitative and qualitative research*, 4th edition. Pearson Publishing.
- Crossley, S.A., L.K. Varner, & D.S. McNamara, (2013). Cohesion-based prompt effects in argumentative writing. *FLAIRS: Proceedings of the 26th International Florida Artificial Intelligence Research Society Conference*, 24, 202-207.
- Dörnyei, Z. (2011). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford University Press.
- Egbulonu, F.O. (2015). The negative aspect of the English language of Igbo culture. *International Journal of Innovative Social Sciences and Humanities Research*, 3(3), 33-38.
- Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, 21(2), 231-244.
- Flowerdew, L. (2001). The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 363-380). John Benjamins Publishing.
- Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37(3), 489-511.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Fontaine, L. (2013). *Analysing English grammar: A systemic functional introduction*. Cambridge University Press.
- Ge, T. (2015). The use of modal verbs to express hedging in student academic writing. *Research in Corpus Linguistics*, 3, 37-46.
- Ghadessy, M., Henry, A., & Roseberry, R.L. (2001). *Small corpus studies and ELT: Theory and practice*. John Benjamins Publishing Company.
- Ghalebi, R., & Sadighi, F. (2015). The usage-based theory of language acquisition: A review of major issues. *Journal of Applied Linguistics and Language Research*, 2(6), 190-195.
- Görlach, M., Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., & Crystal, D. (1987). *A comprehensive grammar of the English language*. Longman Publishing.
- Gultom, H.L. (2016). *The effect of applying clustering techniques on the student's achievement in writing descriptive text* [Unpublished B.A. Dissertation]. State University of Medan, Sumatra, Indonesia.
- Habibi, H. Salleh, A.H. & Singh, M.K.S. (2015). The effect of reading on improving the writing of EFL students. *Pertanika: Journal of Social Sciences & Humanities*, 23(4), 1115 -1138.
- Halliday, M.A.K. (1994). *An introduction to functional grammar*. Edward Arnold Publishing.
- Halliday, M.A.K., & J.J. Webster, (2009). *Continuum companion to systemic functional linguistics*. Continuum Publishing.
- He, Q. (2020). A corpus-based study of transfers in English nominal groups. *Glotta Theory*, 10(1-2), 57-84.
- Henry, A. & Roseberry, R.L. (2001). Using small corpus to obtain data for teaching a genre. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 93-134). John Benjamins Publishing Company.
- Hyland, J. & L. Hamp-Lyons, (2002). EAP issues and directions. *Journal of English for Academic Purposes*, 1, 1-12.
- Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- Hyland, K. (2008). Small bits of textual material: A discourse analysis of Swales' writing. *English for Specific Purposes*, 27, 143-160.

- Hyland, K. (2009). *Teaching and researching writing*. Pearson Education Limited.
- Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in psychology*, 4, 255.
- Indah, P. (2011). Analysis on cohesive devices used in descriptive texts of senior high school books. Accessed <http://repo.iain-tulungagung.ac.id/1213/>
- Isma'il, N., S.I. Elias, Muthusamy, & P.D. Perumal, (2014). Learning to write descriptive essays in an ESL writing class with the aid of online i-Learn web-based resources. *Proceedings of the 8th International Conference on Technology, Education and Development* (pp. 3013-3017). Valencia, Spain.
- Jichun, P. (2015). A corpus-based study of errors in writing committed by Chinese students'. *Linguistics and Literature Studies*, 3(5), 254-258.
- Johnson, D.E. (2013). Descriptive statistics. In R.J. Podesva, & D. Sharma (Eds.), *Research Methods in Linguistics* (pp. 288-315). Cambridge University Press.
- Jordan, R. R. (2002). The growth of EAP in Britain. *Journal of English for Academic Purposes*, 1, 69-78.
- Kalajahi, S.A.R. (2014). *A corpus-based study of Malaysian ESL students use of discourse connectors in the upper and post-secondary argumentative writing* [Unpublished PhD Thesis]. University Putra, Selangor, Malaysia.
- Kamal, A. (2004). *Writing academic English in an ESL situation: A case study of students' of Bayero University, Kano* [Unpublished PhD Thesis]. Department of English and French, Bayero University, Kano-Nigeria.
- Ko, E. (2016). *A corpus-based study of variation and change in adverb placement across world Englishes* [Unpublished M.A. Dissertation]. Georgetown University, Washington, DC.
- Kroskrity P.V. (1999). Identity. *Journal of linguistic anthropology*, 9(1-2), 111-114.
- Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.
- Leech, G. (2015). Descriptive grammar. In D. Biber & R. Reppen, (Eds.), *Cambridge handbook of English corpus linguistics* (p.146-160). Cambridge University Press.
- Leung, R. C. (2016). A corpus-based analysis of textbooks used in the orientation course for immigrants in Germany: Ideological and pedagogic implications. *Journal of Language and Cultural Education*, 4(3), 154-177.
- Muhammad, A.S. & Nair, S.N. (2015). An analysis of descriptive essay writing among Nigerian undergraduates: An analysis of variance in English writing skills. *Australian Journal of Basic and Applied Sciences*, 9(37), 323-329.
- Muhammad, A.S. & Singh, M.K.S. (2020). A corpus-based variation in the processing of determiners in Nigerian undergraduates descriptive writing. *Applied Linguistics Research Journal*, 4(5), 22-38.
- Muhammad, A.S. (2017). *Variation in the descriptive writing skills of the Nigerian undergraduates* [Unpublished PhD Thesis]. School of Languages, Civilization and Philosophy, Awang Had Salleh College, University Utara, Malaysia.
- Mustanoja, T.F. (2016). *A middle English syntax: Part of speech*. John Benjamins Publishing Company.
- Mutiara, R. (2014). Teaching descriptive text writing through guided WH-questions: A pre-experiment study at the eighth-grade students of SMP Negeri 7 Pontianak in academic year 2013/2014. *West Kalimantan Scholars: Studies in English Language and Education*, 1 (1), 51-58.
- Nation, P. (2001). Using small corpora to investigate learner needs: Two vocabulary research tools. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 31-46). John Benjamins Publishing.
- Ojetunde, C. F. (2012). A study of the acquisition of 'their' and 'there' by selected Nigerian students'. *An International Journal of Language, Literature and Gender Studies*, 1(1), 65-75.
- Pramoolsook, I., & Qian, L. (2013). Comparative genre analysis of English argumentative essays written by Chinese English and non-English major students. *Arab World English*, 4(1), 213-223.
- Purwanti, I.S. (2013). *Descriptive analysis of grammatical errors in writing a descriptive essay among the 8th grade students in MTS Amal Shelley Sumogawe, Getasa, Semarang Regency in the academic year 2012/2013* [Unpublished B.A. Dissertation]. State Institute for Islamic Studies, Indonesia.
- Ragan, H. R. (2001). Classroom use of a systemic functional small learner corpus. In M. Ghadessy, A. Henry, & R.L. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and Practice* (pp. 207-236). John Benjamins Publishing.
- Reppen, R., Fitzmaurice, S.M., & D. Biber, (2002). *Using corpora to explore linguistic variation*. John Benjamins Publishing Company.

- Rogers, C.K. (2002). Syntactic features of Indian English: An examination of written Indian English. In R. Reppen, S.M. Fitzmaurice, & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 187-202). John Benjamins Publishing Company.
- Ryoo, M. (2013). A corpus-based study of the use of phrasal verbs in Korean EFL students' writing. *The Journal of Asia TEFL*, 10(2), 63-89.
- Sawalmeh, M.H.M. (2013). Error analysis of written English essays: The case of students of the preparatory year programme in Saudi-Arabia. *English for Specific Purposes World*, 40(14), 1-17.
- Schneider, G. & Grigonyte, G. (2016). Statistical sequence and parsing models for descriptive linguistics and psycholinguistics. In O. Timofeeva, A. Gardner, A. Honkajoki, & S. Chevalier (Eds.), *New Approaches to English Linguistics: Building bridges* (pp. 281-320). John Benjamins Publishing.
- Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language*. Oxford University Press.
- Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice* (pp. 1-16). Oxbow Books Publishing.
- Singh, M.K.S. (2014). *A corpus-based genre analysis of quality, health, safety and environment work procedures in Malaysian petroleum industry* [Unpublished PhD Thesis]. Universiti Teknologi Malaysia, Johor, Malaysia.
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Language Science Press.
- Susan, H. (2002). Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar. In R. Reppen, S.M. Fitzmaurice, & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 167-186). John Benjamins Publishing Company.
- Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge Publishing.
- Tagliamonte, S. A. (2016). Quantitative analysis in language variation and change. In S. Sesserego & Tejedo-Herero, F. (Eds.), *Spanish Language and Sociolinguistic Analysis* (pp. 3-33). John Benjamins Publishing.
- Tavakoli, H. (2012). *A dictionary of research methodology and statistics in applied linguistics*. Rahnama Publishers.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing.
- Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (eds.) *The Routledge Handbook of Corpus Linguistics* (pp. 14-27). Routledge Publishing.
- Vaughan, E. & Clancy, B. (2013). Small corpora and pragmatics. In J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics: New Domains and Methodologies* (pp. 53-76). Springer Dordrecht Heidelberg Publishing.
- Yoon, H. & Jo, J.W. (2014). Direct and indirect access to corpora: An exploratory case study comparing students' error correction and learning strategy use in L2 writing. *Language Learning and Technology*, 18(1), 96-117.
- Zhang, R. (2013). A corpus-based error analysis of students' writing in graded teaching classes. *Journal of Convergence Information Technology*, 8(10), 551-557.
- Zheng, C., & Park, T. (2013). An analysis of errors in English writing made by Chinese and Korean University students. *Theory and Practice in Language Studies*, 3(8), 1342-1351.