A Corpus-Based Variation in the Processing of Determiners in Nigerian Undergraduates Descriptive Writing

Anas Saidu Muhammad, Manvender Kaur Sarjit Singh

ABSTRACT

In this study, we built a learner corpus, specifically the C-BUK corpus, to investigate variation in the processing of determiners in the descriptive ESL writing of the Nigerian undergraduate. The theoretical assumption that guides this study is that of the usage-based theory. Basically, the method applied in this study is the quantitative descriptive research design within the multi-dimensional corpus-based orientation. The findings showed that the overall undergraduates attained the highest frequency in processing the determiner capable of pronominal function (DD). This was followed by the singular determiners (DD1), the wh-question determiner (DDQ), post-determiners capable of pronominal function (DA), and pre-determiners capable of pronominal function (DB). These prove variant to the DDQGE, DDQV, DB2, DAR, DAT, DA1, DA2, and the DD2 forms of determiners, which were least processed. In terms of their gender variation, the females achieved higher frequencies in processing the DD, DD1, and DB, which was in variance to the males that attain a higher frequency in DDQ and DA. In terms of their ethnic group variation, the Igbo group achieved higher rates in DD1, DDQ, and DB. The Hausa group and the Yorùbá group attained higher rates in DA and DD respectively. The pedagogical implications suggest for the organization of seminars and the adoption of situated learning methods in order to enhance the Nigerian undergraduates English writing systems.

Keywords: Corpus; Descriptive ESL writing; Determiners; Ethnic and gender variation; Nigerian undergraduates.

1. Introduction

In English syntax, a determiner is a functional element, otherwise called a pronominal modifier – that alludes to a word or a group of words that indicates, specifies, identifies and quantifies the noun or the noun phrase that comes after it. Practically, determiners are the most frequently used words in English and yet they are the most difficult grammatical structures for ESL learners (Gör latch, et al 1987; Klinge, 2008; Shin, 2012). Given the focal significance of determiner in English production, a lot of studies have identified them to fill in as a referent to a definite or an indefinite linguistic element expressing a specific person or thing, to a particular number or quantity (Dayal & Sağ, 2020; Dinković & Borucinsky, 2016; Fu, 2020; Gray, 2010; Reynolds,
In all actualities, many linguists have made insightful elucidation of the importance of determiners, as they simplify the explanation of the meaning of a referential element in a close or more distant stage and it sustains learner’s confusion in the process of demarcating the pronominal places of linguistic elements in lexical and grammatical processing at the sentence level (Freitheim & Amfo, 2008; Harder, 2008; He, 2020; Kleiber, 2008; Reynolds, 2013). However, in this study, it is argued that learners of English as a second language (ESL), particularly in the Nigerian settings, experience challenges in the processing of determiners and this stands obvious, as learners are often not clearly taught (Bennui, 2016; Ekundayo, 2017; Pine & Lieven, 1997; Tiffen, 1974) and are often discouraged from too much systemization of determiners in textualising coherent writing in English (Dipolog-Ubanan, 2016; Lenchuk & Ahmed, 2014; Yoo & Shin, 2020; Zhang, 2020). In this background, understanding the pattern of determiner processing in the descriptive ESL writing of the Nigerian undergraduate remains a challenging literacy experience and is the main focus of our study.

Basically, Nigeria exemplifies a type of geographically complex country which has within its bounded entities over 550 ethnic groups (Hausa, Igbo and Yorùbá as the major ones) with different native languages (L1), and Mother Tongues (MT) operating side-by-side with English language. This signifies its multilingual and multicultural requirement for discourse processes (Akinyeye, 2015). At this juncture, in its contextual settings of the academic discourse communities, of course, Nigeria for long has had English language as an official one and as an ESL. This originates right before the settlement of the British colonies in the country in 1886 (Emike & Iyiola, 2016). Considering the Nigerian scenario, the creativity to process determiners for a logical and cohesive ESL writing has critically proved to be challenging for ESL learners and is likewise identified to be very limited on account by previous studies (Ekundayo, 2017). To date, however, we are yet to identify a study that has focussed exclusively on variation in the processing determiners in descriptive ESL writing of the Nigerian undergraduates using the corpus-based approach as proposed in this study. Therefore, within the contextual point of this study, a corpus-based approach is employed to explore the processes of determiners embedded in the descriptive ESL writings of the Nigerian undergraduates. So, a learner corpus was developed to comparatively give a detailed idea of the Nigerian undergraduates processing of determiners and its homogeneity to variation in order to impart linguistic and sociolinguistic factors of gender and ethnic groups (Gries, 2015; Leech, 2015; Tagliamonte, 2013a; 2016). For such intention, this study uses a corpus-based procedure with the specific objective to examine the processes of determining patterns processed in the descriptive ESL writing of the Nigerian undergraduates. Therefore, the research questions governing this study are as follows:

1) What is the overall frequency of determiners processed in the Nigerian undergraduate descriptive ESL writing?

2) What is the level of variation of high and low processed determiners in the corpus based on gender?

3) What is the level of variation of high and low processed determiners in the corpus based on ethnic groups?

Going by these research questions 1-3, the significance of this study addresses the use of determiners among the Nigerian ESL learners, which on one hand will provide us with their level of intelligibility, awareness, maturity and knowledge on determinatives function and categories in English writing. On the other hand, it guides our identification of the synchronic and the homogenous account in their usage of determiners in ESL writing as a case point for language variation for an effective remedy (Tagliamonte, 2013a). In relation to this, this study objectively examines the variation in the logical ordering of gender and ethnicity expressing the use of determiners in the Nigerian undergraduate descriptive ESL writing. Attention is paid to the analytical nature of linguistic variation, thereby reporting frequencies and concordances that best reveal their level in the processing of determiners in terms of the sociolinguistic variables of ethnicity (Hausa, Igbo and Yorùbá) as well as their gender categories (male and female).
2. Literature Review

Theoretical framework

The usage-based theory guided this study. Usage-based, theory – an empirically-based theory links up structural and functional systems based on the actual knowledge of usage and processing of distinctive types of linguistic phenomena on lexical and grammatical levels of use (Boye & Harder, 2012; Tomasello, 2009). According to Bybee (2009), usage-based theory emphasizes on the token frequency of use, thereby strengthening or weakening linguistic units in a text-based research. It takes the construction of language and shares the idea that “languages conventionalize frequently used structures so that use directly shape structures” (p. 18). Ibbotson (2013) further revealed that the usage-based theory sees “productivity of language is as a result of knowledge generalised over usage events that usage has an effect on linguistic structure” (p. 6). Likewise, 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 principles of the usage-based to examine variation in the processing of determiners manifesting the Nigerian undergraduate descriptive ESL writing within the premise of corpus-based linguistics.

From the point of literature review, this study found instances where researchers exchange opinions on the overtly contextual confusion that says numbers of English learners do stumble on the appropriate use of determiners. Görlach, et al (1987) opines to the criticality and necessity for creating awareness to determiners processing in English writing. They express determiners as a set of discourse process of nominal groups in the shape of words, phrase, or affix that occur in performance with a noun or a noun phrase in the context of pressing out the reference of the used and patterned nouns and or noun phrases for meaning-making. In this realm, even though, Görlach, et al (1987) sort explain the determiners by the virtues of their functions when employed before the noun often acts as the head of the noun phrase or before its pre-modifiers that can expatiate as the definite reference and the indefinite reference. Yet, Gray (2010) expresses his dismay to the teachers, instructors, and textbook writers over their negligence towards explaining the roles of determiners in discursive practices. This proves that instructors have not fully made clear the situational linguistic context of determiners for learners in the event of classroom-based practices. Thus, he ascertains that ESL learners confusedly use demonstratives “this, these, that, and those” as determiners and then as pronouns, and this leads to a “vague referential”. In this way, as Fontaine (2013) further reveals:

Learners complicate identification of specific discourse processes of determiners in their written repertoires in terms of their variation and or deviation of the “deictic determiners (DD), e.g., a, the, this, that, those, these, my, his, their, John’s, my neighbour’s, my friend’s mother’s, the nice man’s, the partitive determiner (PD), e.g., the top (the top of the desk), the back (the back of the house), a part (a part of the book), the arm (the arm of the chair), a section (a section of the room), and the quantifying determiner (QD), e.g., a, some, many, one, two, six hundred, almost fifty, very few, about seven, two cups, a handful, a pinch”. (51-56)

Specifically, Reynolds (2013) expresses that learners find it challenging to distinguish the category of determiners and the functional roles of the specifiers. On his part, Reynolds (2014) perceives establishing acute knowledge on determiners develop learner’s inspiration for accuracy in the use of grammatical elements that guide their understanding to discourse characteristics in English writing. Exploring the processing of determiners of the Nigerian Educated English (ENE) and that of the Standard British English (SBE) within sociolinguistic lens, Ekundayo (2017) discovered that there are differences in the pattern of the use of articles, particularly definite article (the, a) as well as other types of determiners as a result of interference and intraference.

On another level of the situation, Lenchuk and Ahmed (2014) argue that developing the ESL/EFL learners’ awareness of the use and patterning of determiners in English writing systems facilitate their lexis and vocabulary awareness. A point in this case is a study by Bennui (2016) on the Thailand learners, who has English as a Foreign Language (EFL) experiences challenges in their
EFL writing repertoires due to their ethnic group inference. The Thai EFL learners commit high rate of errors in their patterning of the definite and the indefinite determiners in English writing. This occurs as they often translate the indefinite determiners of the "much-many and a-few-a little" in a duplicative or compounding way, which are present in their native Thai linguistic framework into English. This is justified as it was clear, the English specification and the rules of determiners in the context of use have to agree with the noun forms within a specific discourse process is absent in the Thai native language.

Another critical instance, in the use of determiners is found in a study conducted by Dinković and Borucinsky (2016). Their findings reveal that the Croatian language users inappropriately use and pattern determiners in their communicative events, which in turn, make their communicative repertoires chaotic. As well the Croatian learners could not differentiate English determiners in terms of their elements (possessive, demonstrative, and interrogatives), and functions (at the sentence level) in their writings. In another array of study, Dipolog-Ubanan (2016) reported high rate of errors in the Chinese EFL writing process of determiners. His findings hypothesized that the errors occur because Chinese language has determiners that function in the forms of singular nouns only. This is quite different from the English whose determiners function depending the number and/or if the first alphabet of the noun form is a vowel or a consonant. With these intuitions, this study’s exploration of the Nigerian undergraduates’ determiner processes provided basic abilities to their discourse “structural link between the determiner and the head of the nominal group” (Fontaine, 2013, p. 52).

3. Methodology

In reality, in order to present the findings of this study, a quantitative type of descriptive research design was utilised (Leech, 2015). We simply revealed the achieved concordance frequency hits, and the statistical percentage showcasing overall concordance and variation in determiners processing derived in the corpus in the form of tables and graphics (Johnson, 2013). This provides us with sufficient situation of attaining the variation of the concordance and frequencies.

3.1. Participants

In acquiring the participants, we utilised the technique of the stratified purposive sampling (Dörnyei, 2011; Tavakoli, 2012). During the second semester of the 2014/2015 academic session, 30 sample representatives were attained. The idea of using the stratified purposive sampling technique is to the objectives deliberated for this study in identifying the overall frequency usage and identify the variation in determiners processing in the descriptive ESL writing based on their gender and ethnic groups.

The location is within the province of the Bayero University, Kano (BUK). BUK is a Nigerian public university that is situated in Kano, a state that is located in the central Northern part of the country. The university strongly relates to English, as a specific discourse community, covering undergraduates originating from nearly all the ethnic groups in Nigeria. Therefore, we selected the research participants along with sociolinguistic variables of gender category to represent the males and females and then ethnicity to represent the Hausa, Igbo, and Yorùbá ethnic groups (Tagliamonte, 2016) to compose descriptive writing in English.

3.2. Instrument

Three research instruments were used in this study. The first instrument employed in this study is the descriptive essay writing. Basically, descriptive essay writing in the Nigerian academic discourse communities is regarded as a form of continuous assessment and a form of written discourse strategy (Muhammad, 2017). Specifically, Gultom (2016) says descriptive essay writing refers to a subjective creativity en route for ESL writing growth, practice and enhancement. In the opinions of Mutiara (2014) descriptive essay writing is a creative instrument that stands to serve as a testing tool, often employed on learners’ in order to experiment their level of awareness and competencies in a variety of pedagogical and classroom-based practices. In the event of this study, the sampled students were asked to compose a descriptive essay writing about Bayero University, Kano (BUK).
in English medium. They were given duration of one hour thirty minutes (1hr, 30 minutes) for the writing session. That is, they were required to vividly describe their main university in terms of its buildings and its road channels in English medium. In the event of the writing processes, the participants were monitored as they completed the descriptive writing. To justify the content validity and the face validity, they were not allowed to use any source of dictionary information and even their smartphones were not allowed where the action is taking place (Creswell, 2012). Upon completion, all the written essays were collected.

The second instrument used is the online version of the parts-of-speech (POS) Constituent Likelihood Automatic Word-Tagging System (CLAWS) uses rich text format. It was developed by the UCREL of the University of Lancaster, which was used to tag the British National Corpus. It is a free software accessible for tagging English texts for grammatical corpus connotations. The POS CLAWS tagger support three modes of text tagging – the horizontal, the vertical and that of the html accessed via the internet. And for the purposes of this study, we used the horizontal tagging system. It is vital to note that the POS CLAWS tagger has consistently achieved reliability of word tagging process having 96-97% accuracy and an error rate of 1.5% (Manvender, 2014, p. 155).

The third instrument used is the AntConc 3.4.3.w., software developed by Professor Lawrence Anthony, who is a Japanese Professor in the Faculty of Science and Engineering at Waseda University, Japan is utilised in this study. It is essentially designed to facilitate linguistic research on text corpora (small and large) and to be used in the classroom as an instrument of analysis (Anthony, 2005). It permits utilization of a powerful concordance, word cluster, word list and keyword frequency generators, tools for cluster and lexical bundle analysis, and a word distribution plot.

To further justify and check the validity as well as the reliability of the research instruments and the analysis done in this study. We used an inter-coder reliability assessment. We first selected two English lecturers that have taught English for over ten years in the university. They were given the research instrument to study. They affirm that the research instrument is understandable, safe and is valid to be used for the purposes of this study. Secondly, an inter-rater reliability test was carried out. The findings indicated that there is high inter-rater reliability between the two raters (lecturers) in their scoring. It showed inter-rater reliability with a significant Pearson correlation of \( r = 0.901, p = 0.01 \). The reliability test indicates the research instrument is highly reliable and can be positively utilised in the main study (Creswell, 2014; Dörnyei, 2011).

3.3. The Corpus Compilation Procedure

The corpus developed and compiled is subjected to Manvender’s (2014) Computer-Assisted Corpus Analysis (CACA). Crucially, CACA guides innovation of corpus-based framing and exposes students, as well as, researchers in the coding of raw data. By its virtue, CACA facilitates corpus-based analysis with the criteria of the coder reliability and the independent coding practices. CACA has achieved a higher correlation of validity of the value of \( K = 0.941 \), which makes it easy for replication. CACA is highlighted to provide a quick, time-saving, and an easy procedure for corpus compilation with the guidance of online software as shown in Figure 1:
Figure 1 showed that at the first stage, the samples of the descriptive writings were contextually hand written by 30 Nigerian undergraduates. They were typed-set in the Microsoft Word Document (MWD) and saved in a separate folder, as the “MWD Folder”. The saved MWD files were categorically transferred in plain text using the AVS text converter. All the private details existing in the descriptive texts were taken out, and were again saved in a different folder, as the “Plain Text Folder (PTF)”. This guided us to present the grammatical form of determiners embedded in the compiled corpus based on their function, class, frequency levels and categories. The learners-corpus is named the Corpus of Bayero University Kano (C-BUK). Due to the varying interest of this study, each sub-corpus of the C-BUK is provided with its representative codes. So, the C-BUK mini-corpus make available 16,637 words as shown in Figure 2 below in order to make evident variation in the processing of determiners in Nigerian undergraduates descriptive ESL writing. This, in turn, guided us to satisfy the basic aims and respond to the research objectives of this study in the following order:

Figure 2. The Compilation Procedure of the C-BUK

Figure 2 above showcases the Nigerian Undergraduates’ male’s sub-corpus is named “NUM” and that of the Females is “NUF”. Then sub-corpus of the Hausa Ethnic Group is “NUH”, the Igbo Ethnic Group is “NIH”. That of the Yorùbá Ethnic Group is “NYH”. The PTF files where each processed online with the help of the Parts of Speech Constituent Likelihood Automatic Word-tagging System (POS CLAWS). POS CLAWS7 is free software accessible for tagging English texts for grammatical corpus connotations. It is provided online by the University of Lancaster.

This study uses the horizontal tagging of the POS CLAWS7 system to individually tag each of the PTF. These coded data were saved in a different folder as the “POS Tagged CLAWS Folder” (PTCF). It is vital to be specific here; in each of the sample files, they represent five (5) PTCF files written by the Hausa females, five (5) PTCF files written by the Igbo females, and five (5) PTCF files written by the Yorùbá females to make up fifteen (15) PTCF files written by the females. After that, we selected five (5) PTCF files written by the Hausa males, five (5) PTCF files written by the Igbo males, and five (5) PTCF files written by the Yorùbá males to make up fifteen (15) PTCF files written by the males. So, the PTCF provides us with the initial development of a small-corpus that is enshrined with sub-corpus to develop a total of 16,637 words.

Basically, this study could not imitate a large corpus as it were in the British National Corpus
(NBC) that has over one hundred million words. Nor could it go by the Brigham Young University largest 450 million words Corpus of Contemporary American English (COCA), which entails wide texts from an array of genres. Needless to say, about going either trend of the Croatian National Corpus (CNC), or the Freiburg English Dialect Corpus (FRED), which have compiled 370 genres of texts recorded between 1970 and 1990 of approximately 2.4 million words to strengthen research on the morphosyntactic variation on oral and dialectal sources of speech. We tried to narrow down to the construction of a learner corpus across educational institutes for empirical investigation and pedagogic development. So, this study could only compile a mini-corpus to investigate the constituents of determiners processed in the Nigerian undergraduate descriptive ESL writing.

3.4. Data Analysis

In order to resolve on these fundamental issues on determiners, this study focuses solely on the POS tag-sets processed by the online CLAWS7. Table 1 depicts the codes of the POS CLAWS parsing and the interpretation of the grammatical function annotations of determiners embedded in the C-BUK. The codes and their interpretations read as:

<table>
<thead>
<tr>
<th>CODES</th>
<th>CLAWS S7 Interpretations</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA</td>
<td>after-determiner or post-determiner capable of pronominal function (for example, <em>such</em>, <em>former</em>, <em>same</em>)</td>
</tr>
<tr>
<td>DA1</td>
<td>singular after-determiner (for example, <em>little</em>, <em>much</em>)</td>
</tr>
<tr>
<td>DA2</td>
<td>plural after-determiner (for example, <em>few</em>, <em>several</em>, <em>many</em>)</td>
</tr>
<tr>
<td>DAR</td>
<td>comparative after-determiner (for example, <em>more</em>, <em>less</em>, <em>fewer</em>)</td>
</tr>
<tr>
<td>DAT</td>
<td>superlative after-determiner (for example, <em>most</em>, <em>least</em>, <em>fewest</em>)</td>
</tr>
<tr>
<td>DB</td>
<td>before determiner or pre-determiner capable of pronominal function (<em>all</em>, <em>half</em>)</td>
</tr>
<tr>
<td>DB2</td>
<td>plural before-determiner (<em>both</em>)</td>
</tr>
<tr>
<td>DD</td>
<td>determiner (capable of pronominal function) (for example, <em>any</em>, <em>some</em>)</td>
</tr>
<tr>
<td>DD1</td>
<td>singular determiner (for example, <em>this</em>, <em>that</em>, <em>another</em>)</td>
</tr>
<tr>
<td>DD2</td>
<td>plural determiner (for example, <em>these</em>, <em>those</em>)</td>
</tr>
<tr>
<td>DDQ</td>
<td>wh-determiner (for example, <em>which</em>, <em>what</em>)</td>
</tr>
<tr>
<td>DDQGE</td>
<td>wh-determiner, genitive (whose)</td>
</tr>
<tr>
<td>DDQV</td>
<td>wh-ever determiner, (for example, <em>whichever</em>, <em>whatever</em>)</td>
</tr>
</tbody>
</table>

The analyses of these thirteen (13) POS CLAWS constituents of determiners are given in Table 1. The AntConc 3.4.3.w., software guided us in enabling the concordance of the frequency to the coded determiners as presented in Figure 3 below:

Figure 3. AntConc POS CLAWS tagging of determiners of the C-BUK

Figure 3 above showed that AntConc 3.4.3.w. Software guided this study in identifying the
frequency concordance of the processed determiners in the descriptive ESL writing and the level of variation according to gender and ethnicity.

4. Result

Corpus-Based Taxonomy of the Determiners

In this section the taxonomies of determiners manifesting in the Nigerian undergraduate descriptive ESL writing using the corpus-based procedure. Globally, it is possibly rarely easier to come across competing approaches on gender and ethnic group variation in English writing and is distributed over deviant view of results, findings, and observations. This study makes the concept of gender and ethnicity to be regarded as a useful tool indicative of socially constructed variables in English writing skills. The global trend of the results on gender and ethnic group variation of the English native speakers and that of the non-native English users proved to be inconclusive. Hence, this study situated its analysis on identifying the Nigerian undergraduates’ discourse processing of determiners in descriptive ESL writing based on their gender and ethnic group’s variation as sociolinguistic variables.

The Overall Processed Determiners Across the Corpus

Research question 1 requires identifying the overall frequency of the concordances established for highly and lowliest categorical processed determiners embedded in the corpus. The highest forms of determiners processed by the Nigerian undergraduates in the C-BUK with over one hundred concordance are the DD and the DD1 constituents. Then, the DDQ and DA constituents are as well considered to have reached a higher concordance range as their concordance is very close to a hundred. However, the lowliest forms of determiners identified in this study are DB, DA2, DD2, DA1, DAT, DAR, and the DB2 constituents as they attained concordance that is much below one hundred percent. The concordance scores are categorically presented in Table 2 underneath:

<table>
<thead>
<tr>
<th>Codes</th>
<th>Overall Tokens</th>
<th>Overall % Frequency</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>270</td>
<td>1.62 %</td>
<td>1</td>
</tr>
<tr>
<td>DD1</td>
<td>165</td>
<td>0.99 %</td>
<td>2</td>
</tr>
<tr>
<td>DDQ</td>
<td>98</td>
<td>0.58%</td>
<td>3</td>
</tr>
<tr>
<td>DA</td>
<td>94</td>
<td>0.56%</td>
<td>4</td>
</tr>
<tr>
<td>DB</td>
<td>54</td>
<td>0.32 %</td>
<td>5</td>
</tr>
<tr>
<td>DA2</td>
<td>31</td>
<td>0.18 %</td>
<td>6</td>
</tr>
<tr>
<td>DD2</td>
<td>31</td>
<td>0.18 %</td>
<td>6</td>
</tr>
<tr>
<td>DA1</td>
<td>22</td>
<td>0.13%</td>
<td>7</td>
</tr>
<tr>
<td>DAT</td>
<td>14</td>
<td>0.08%</td>
<td>8</td>
</tr>
<tr>
<td>DAR</td>
<td>13</td>
<td>0.07 %</td>
<td>9</td>
</tr>
<tr>
<td>DB2</td>
<td>6</td>
<td>0.03 %</td>
<td>10</td>
</tr>
<tr>
<td>DDQV</td>
<td>2</td>
<td>0.01%</td>
<td>11</td>
</tr>
<tr>
<td>DDQGE</td>
<td>0</td>
<td>0.00%</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td>801</td>
<td>04.75%</td>
<td>801 (14.29%)</td>
</tr>
</tbody>
</table>

As presented in Table 2 above, the overall undergraduates processed thirteen constituents of determiners processed in the C-BUK with an overall frequency of 04. 75% out of the 801 tokens. The distribution pattern showed that the Nigerian undergraduates achieved highest percentage frequency in processing DD constituents, which are determiners capable of demonstrating the level of pronominal functions (Tokens=270, Freq=21.88%). The second highest form of processed determiner is the singular forms of functional categories after determiners that represent the DD1 constituent. The wh-determiner of the (DDQ) constituent is their third most processed (Tokens=1846, Freq=21.88%). The fourth is the DA constituent, which demonstrates their capability
in presenting pronominal function after a determiner (Tokens=94, Freq=0.56%). In justifying the other forms of determiners concordances, the scaling nodes of the variation in the Nigerian undergraduates overall processing of the lowliest categories of determiners in the corpus can be observed further as presented in Figure 4 below:

Figure 4. The overall frequency nodes of the determiners, per 16,637 words

Figure 4 above further showed that the overall Nigerian undergraduates prove to be deficient in processing determiners of the DDQGE constituent, which demonstrates their lowest ability in processing the genitive *wh*-determiners. In this situation, both the groups could not achieve a concordance hit and a percentage score in the corpus (Tokens=0, Freq=0.00%). Moreover, the DDQV constituent of the *wh-*ever, especially the *wh*-forms, including whichever, whatever, stand to be one of their lowest concordance hit achieved (Tokens=2, Freq=0.01%). In addition to this, the DB2 constituent was also found to be very low (Tokens=6, Freq=0.03%). In this, they could not utilise the plural forms that can be patterned, structure and situated before-determiner in the process of writing or speaking. Furthermore, the concordance of DAR constituent signifies their inability to process comparative after-determiners such as more, less, fewer. Likewise, the DAT constituent that signifies their use of superlative forms after-determiner, such as most, least, fewest, were found to be limited in the concordance hit (Tokens=14, Freq=0.08%). Another form of determiner that is low in occurrence is the DB constituent, which exhibit their knowledge in processing pronominal function before a determiner (Tokens=54, Freq=0.32%). That is, a number of the Nigerian undergraduates find it challenging to contextualise DB, DB2, DAR, DDQ, DDQGE, and DAT forms of determiners.

**Gender Variation of the Processed Determiners**

Research question 2 requires to identify the level of variation of high and low processed determiners in the corpus based on gender category. Basically, the findings indicate gender variation as the females processed more constituents of DD, DD1, DB, DA1, and DAT determiners. These ascertain a variation to the male counterparts that have processed more of the DDQ, DA, DD2, DA2, DAR, and DDQGE determiners. The interpretation of the research findings can be further observed as presented in Table 3 underneath:

**Table 3. Gender Variation of Determiners, per 16,637 Words**

<table>
<thead>
<tr>
<th>Codes</th>
<th>Male Tokens</th>
<th>Males %</th>
<th>Female Tokens</th>
<th>Females %</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>133</td>
<td>1.57 %</td>
<td>137</td>
<td>1.66 %</td>
<td>1</td>
</tr>
<tr>
<td>DD1</td>
<td>79</td>
<td>0.93 %</td>
<td>86</td>
<td>1.04 %</td>
<td>2</td>
</tr>
<tr>
<td>DDQ</td>
<td>53</td>
<td>0.62 %</td>
<td>45</td>
<td>0.54 %</td>
<td>3</td>
</tr>
<tr>
<td>DA</td>
<td>48</td>
<td>0.56 %</td>
<td>46</td>
<td>0.55 %</td>
<td>4</td>
</tr>
<tr>
<td>DB</td>
<td>25</td>
<td>0.29 %</td>
<td>29</td>
<td>0.35 %</td>
<td>5</td>
</tr>
<tr>
<td>DD2</td>
<td>20</td>
<td>0.23 %</td>
<td>11</td>
<td>0.13 %</td>
<td>6</td>
</tr>
<tr>
<td>DA2</td>
<td>16</td>
<td>0.18 %</td>
<td>15</td>
<td>0.18 %</td>
<td>7</td>
</tr>
</tbody>
</table>
As presented in Table 3, the concordance hit indicated gender variation of low and high frequencies in eleven (11) out of the thirteen determiner constituents processed in C-BUK with an overall frequency of 09.47% out of the 801 tokens. The distribution pattern of the overall concordance showed that the females achieved the highest ratio of concordance at a statistically significant level of 401 token with 04.81% frequency, compared to the male counterparts with a token of 400 and a frequency of 04.66%. In terms of the categorical determiner constituents, the females achieved a higher concordance of 137 (Freq=1.66%) in processing DD constituent and the males with 133 (Freq=1.57%). This showed that the females processed determiners capable of demonstrating the level of pronominal functions in the descriptive ESL writing than the male counterparts. In processing DD1 determiners, the females also achieved a higher score (Tokens=165, Freq=0.98%) compared to the males (Tokens=165, Freq=0.98%). This showcases the female’s high usage of the singular forms of functional categories after determiners in the descriptive ESL writing over the males. For illustrative purposes, the scaling nodes of the gender variation of the other forms of determiners can be further observed as presented in Figure 5 below:

![Figure 5. Gender variation of determiners, per 16, 637 words](image)

As presented in Table 3, the concordance hit indicated gender variation of low and high frequencies in eleven (11) out of the thirteen determiner constituents processed in C-BUK with an overall frequency of 09.47% out of the 801 tokens. The distribution pattern of the overall concordance showed that the females achieved the highest ratio of concordance at a statistically significant level of 401 token with 04.81% frequency, compared to the male counterparts with a token of 400 and a frequency of 04.66%. In terms of the categorical determiner constituents, the females achieved a higher concordance of 137 (Freq=1.66%) in processing DD constituent and the males with 133 (Freq=1.57%). This showed that the females processed determiners capable of demonstrating the level of pronominal functions in the descriptive ESL writing than the male counterparts. In processing DD1 determiners, the females also achieved a higher score (Tokens=165, Freq=0.98%) compared to the males (Tokens=165, Freq=0.98%). This showcases the female’s high usage of the singular forms of functional categories after determiners in the descriptive ESL writing over the males. For illustrative purposes, the scaling nodes of the gender variation of the other forms of determiners can be further observed as presented in Figure 5 below:

![Figure 5. Gender variation of determiners, per 16, 637 words](image)

As presented in Table 3, the concordance hit indicated gender variation of low and high frequencies in eleven (11) out of the thirteen determiner constituents processed in C-BUK with an overall frequency of 09.47% out of the 801 tokens. The distribution pattern of the overall concordance showed that the females achieved the highest ratio of concordance at a statistically significant level of 401 token with 04.81% frequency, compared to the male counterparts with a token of 400 and a frequency of 04.66%. In terms of the categorical determiner constituents, the females achieved a higher concordance of 137 (Freq=1.66%) in processing DD constituent and the males with 133 (Freq=1.57%). This showed that the females processed determiners capable of demonstrating the level of pronominal functions in the descriptive ESL writing than the male counterparts. In processing DD1 determiners, the females also achieved a higher score (Tokens=165, Freq=0.98%) compared to the males (Tokens=165, Freq=0.98%). This showcases the female’s high usage of the singular forms of functional categories after determiners in the descriptive ESL writing over the males. For illustrative purposes, the scaling nodes of the gender variation of the other forms of determiners can be further observed as presented in Figure 5 below:

![Figure 5. Gender variation of determiners, per 16, 637 words](image)
constituent based on gender. Together, the male and the female undergraduates achieved (Tokens=1, Freq=0.01%) concordance count. This showed that both the gender categories experience challenges in the ever form of determiners, especially the wh-forms, such as whichever, whatever, stand to be one of their lowest concordance hit achieved. In addition to this, there is no gender variation in processing the DB2 constituent and was also found to be very low (Tokens=6, Freq=0.03%). In this, they could not utilise the plural forms that can be patterned, structured and situated before-determiner in the process of writing.

**Ethnic Group Variation of the Processed Determiners**

Research question 3 requires to identify the level of variation of high and low processed determiners in the corpus based on ethnic grouping. In this study, the ethnic grouping refers to the three major Nigerian ethnic groups, namely the Hausa, Igbo and the Yorùbá. The findings showed the Igbo ethnic group achieves higher frequencies in DD1, DDQ, and DB, which prove to be varied to the Hausa’s and the Yorùbá’s that attain the higher frequency in DA and DD respectively. The interpretation can be further observed as presented in Table 4 underneath:

<table>
<thead>
<tr>
<th>Codes</th>
<th>Hausa Tokens</th>
<th>Hausa %</th>
<th>Igbo Tokens</th>
<th>Igbo %</th>
<th>Yorùbá Tokens</th>
<th>Yorùbá %</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>83</td>
<td>1.52%</td>
<td>91</td>
<td>1.57%</td>
<td>96</td>
<td>1.75%</td>
<td>1</td>
</tr>
<tr>
<td>DD1</td>
<td>55</td>
<td>1.01%</td>
<td>59</td>
<td>1.07%</td>
<td>51</td>
<td>0.93%</td>
<td>2</td>
</tr>
<tr>
<td>DDQ</td>
<td>26</td>
<td>0.47%</td>
<td>37</td>
<td>0.63%</td>
<td>35</td>
<td>0.64%</td>
<td>3</td>
</tr>
<tr>
<td>DA</td>
<td>34</td>
<td>0.62%</td>
<td>30</td>
<td>0.51%</td>
<td>30</td>
<td>0.54%</td>
<td>4</td>
</tr>
<tr>
<td>DB</td>
<td>12</td>
<td>0.22%</td>
<td>25</td>
<td>0.43%</td>
<td>17</td>
<td>0.31%</td>
<td>5</td>
</tr>
<tr>
<td>DD2</td>
<td>04</td>
<td>0.07%</td>
<td>15</td>
<td>0.25%</td>
<td>12</td>
<td>0.21%</td>
<td>6</td>
</tr>
<tr>
<td>DA2</td>
<td>09</td>
<td>0.16%</td>
<td>7</td>
<td>0.12%</td>
<td>15</td>
<td>0.27%</td>
<td>7</td>
</tr>
<tr>
<td>DA1</td>
<td>07</td>
<td>0.12%</td>
<td>10</td>
<td>0.17%</td>
<td>15</td>
<td>0.27%</td>
<td>8</td>
</tr>
<tr>
<td>DAT</td>
<td>08</td>
<td>0.14%</td>
<td>03</td>
<td>0.05%</td>
<td>03</td>
<td>0.05%</td>
<td>9</td>
</tr>
<tr>
<td>DAR</td>
<td>05</td>
<td>0.09%</td>
<td>04</td>
<td>0.06%</td>
<td>04</td>
<td>0.07%</td>
<td>10</td>
</tr>
<tr>
<td>DB2</td>
<td>01</td>
<td>0.01%</td>
<td>04</td>
<td>0.06%</td>
<td>14</td>
<td>0.01%</td>
<td>11</td>
</tr>
<tr>
<td>DDQV</td>
<td>00</td>
<td>0.00%</td>
<td>00</td>
<td>0.00%</td>
<td>02</td>
<td>0.03%</td>
<td>12</td>
</tr>
<tr>
<td>DDQGE</td>
<td>01</td>
<td>0.01%</td>
<td>00</td>
<td>0.00%</td>
<td>00</td>
<td>0.00%</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>245</td>
<td>04.44%</td>
<td>285</td>
<td>04.95%</td>
<td>271</td>
<td>4.90%</td>
<td>801   (14.29%)</td>
</tr>
</tbody>
</table>

As presented in Table 4 above, the concordance hit indicated ethnic group variation of low and high frequencies in the thirteen (13) processed determiners in the C-BUK. The overall frequency showed that there is an ethnic variation in the processing of determiners with a Tokens of 801 and a frequency of 14.29%. The overall distribution pattern showed that the Igbo group achieved higher concordance (Tokens=285, Freq=04.95%), followed by the Yorùbá group (Tokens=271, Freq=04.90%), then the Hausa group (Tokens=245, Freq=04.95%). The categorical distribution pattern of the constituents showed that the Yorùbá group achieved a higher concordance of 96 (Freq=1.66%) in processing the DD constituent. Followed by the Igbo group (Tokens=91, Freq=1.57%), then the Hausa group (Tokens=83, 1.52%). These findings showed that the Yorùbá group processed higher ranges of the determiners capable of demonstrating the level of pronominal functions in the descriptive ESL writing more than the Igbo group and the Hausa group and the difference is at a statistical significant level. For illustrative purposes, the scaling nodes of the ethnic group variation of determiners can be further observed as presented in Figure 6 below:
Figure 6 above further showed variation in the Nigerian undergraduates processing of determiners in the corpus based on ethnic groups. Here, the nodes showed that the Igbo group and the Yorùbá group were more deficient in processing the DDQGE constituent (Tokens=0, Freq=0.00%) compared to the Hausa group (Tokens=1, Freq=0.01%). This demonstrated that the Igbo group and the Yorùbá group encounter more of challenges to process the wh-determiners capable of genitive functions compared to the Hausa group. In processing DDQV constituent, the Hausa group and the Igbo group achieved (Tokens=00, Freq=0.00%) and the Yorùbá group attained (Tokens=02, Freq=0.03%). This showed that the Yorùbá group processed superlative forms of after-determiner, such as, most, least, fewest, than the counterparts of the Hausa group and the Igbo group.

5. Discussion and Conclusion

The findings of this study revealed that for each 16,637 words manifesting in the C-BUK, eight hundred and one words (801) with a frequency of 14.29% are determiners. As well as the findings revealed that the highest forms of determiners processed are the DD, DD1, DDQ and DA constituents. Also, the findings showed the lowliest determiners processed are DB, DA2, DD2, DA1, DAT, DAR, and DB2. These results support the findings of Tiffen (1974), Okunrinmeta (2014), Akinyeye (2015), and that of Muhammad (2017). They have argued on Nigerian English learner’s failure to portray their textual repertoires with comparative and contrastive features of a given descriptor in order to create an aesthetic, vivid and figurative imagery to the readers. These results also concur with Görlach, et al (1987), Gary (2010), Reynolds (2013) Lenchuk and Ahmed (2014), Dinković and Borucinsky (2016) as well as Bennui (2016) that asserts on a larger proportion of English learners are novice writers and they failed to show awareness of the function, category, and the analytical distinction of the determiners in English writing. Same is with Okunrinmeta (2014) who reports that from the year 1997 to 2012, records have shown that Nigerian ESL learners’ writing keeps falling down beyond the satisfactory level. He says in 1997, only 6.54 percent of the students who took the West African Senior School Certificate Examinations (WASSCE) got a C grade in English, while nearly 67 percent had a clear failure. In addition to this, only 29.59, 25.36, 34.48, and 29.94 percent passed English at credit level in 2004, 2005, 2006 and 2007 respectively. In 2012, only 38.1 percent of the candidates had a C grade in English Language (p. 27).

The findings of this study showed gender variation in the determiners processing with 09.47% frequency rate out of the 801 tokens. The females achieved higher concordance in DD, DD1, DB, DA1 and DAT constituents. The male students were variant as they attained higher concordance in DDQ, DA, DD2, DA2, DAR and DDQGE constituents. However, no gender variation was identified in BD2 and the DDQV constituents. These findings are parallel to that of Pichler (2010), Eckert (2012), Akinyeye (2015), Gries (2015), Chika (2016), Dipolog-Ubanan (2016), Eckert (2016), Oladimeji (2016), as well as Wayar and Saleh (2016) who asserts that gender is a linguistic variable that gives the multifunction and the multi-factorial diversity of the discourse features processed in a complex and sophisticated ways that gives the fundamental linguistic variety of the higher and lower
processes of discourse features.

Moreover, ethnic group variation was identified in the Nigerian undergraduates processing of determiners with 14.29% frequency rate out of the 801 tokens. The overall distribution pattern showed that the Igbo ethnic group achieved the highest ratio of concordance, followed by the Yorùbá group and then the Hausa group. In essence, the Igbo group processed more of the DD1, DDQ, DB, DD2, DA1, and the DB2 constituents. The Yorùbá group processed the DD, DA2, and the DDQV constituents than the Igbo and the Hausa ethnic groups. The Hausa group processed DA, DAT, DAR, and the DDQGE constituents than the Yorùbá group and the Igbo ethnic group. These findings are parallel to that of Obiamalu (2013) who asserts that the Igbo native language has no definite and indefinite article in their pattern of determiner usage. Igbo language has null-determiner head and the quantifiers, so they have less challenges when it comes to learning patterns of ESL determiners. In addition to this, the findings can be related to that of Lateef (2018), who also revealed that the Yorùbá native language has no article, and because of this they find it difficult to make use of articles and determiners where necessary. Furthermore, the findings can be related to that of Ekundayo (2017) who argued:

Educated Nigerians tend to “overgeneralize the use of article and determiners with noun phrases, applying them superfluously or omitting them where necessary and even sometimes using ‘the’ as a possessive determiner for ‘his’, ‘her’ and their’. They also often yoke similar and exclusive determiners together in nominal structures. (p. 139)

The implication provided in this study, is the need for EAP course design in order to provide learners with sufficient discourse competence in English writing as an ESL activity among Nigerian undergraduate. With this, special attention should be provided by the National Policy on Education (NPE), the Universal Education Board (UBE), and the State Universal Education Board (SUBEB) on intervention protocols concerning strategic ways of enhancing students’ competency in learning English. In view of this, we recommend; teachers, lecturers and researchers should pay more attention in exposing learners’ awareness in 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. It is almost found to be absent in the practised classroom-based discourse practices, even though they are needed for a successful and communicative writing.

There is also the need for further classroom-based activities to concentrate relatively more on the corpus-based and variationist sociolinguistic approaches. Over the years, corpus-based and the variationist sociolinguistic orientations have focused on the English discourse processes that eases classroom activities and facilitate knowledge sharing in multidimensional, multi-cultural and multi-complex pedagogic settings (Anthony, 2005; He, 2020; Labov, 1972, 1972b, 2010). Partly, moreover, most of the identified earlier studies, within the corpus-based linguistics and variationist sociolinguistics are multidimensional in nature, which are popularly framed within the genre, register, texts, gender, ethnicity, age and geography (Eckert, 2012; Oladimeji, 2016).

In the case of our research findings, however, this study is not without limitations. Firstly, it could only report findings from a mini-constructed corpus, it consists of thirty (30) samples of descriptive ESL writing. Future studies can consider a larger corpus with a descriptive writing sample of more than five hundred (500) descriptive essay writing. Secondly, it could only report the concordance frequencies. Future studies can use inferential statistics in the form of regression, quasi-experiments, to mention but a few ways to further explore the usage of determiners of the Nigerian ESL learners. Thirdly, it could only cover the quantitative type of research design to report the research findings. Future studies can adopt a mixed-method procedure to thoroughly examine the Nigerian ESL learner’s usage of determiners and in the same instance thereby reporting the factors that facilitate or challenges the Nigerian undergraduate usage of determiners in English. In addition to this, we could only employ patterns of descriptive essay writing of the Nigerian undergraduates to build the current C-BUK corpus. Future studies can develop a corpus using different forms of essay writings, including the argumentative essay writing, the narrative essay writing or even merge both types of descriptive, argumentative, narrative, expository writing to build a larger corpus. Moreover, in this study, we could only build a mini-learner corpus of sixteen
thousand, six hundred and thirty-seven (16,637) words. Future studies can work towards building a large corpus that can entail different types of essay writing in order to generate more than five hundred thousand, to one million or more words.

Despite these limitations, this study contributes to highlighting the academic issues relating to the determiner usage of the Nigerian undergraduates. It covers the theoretical and methodological areas of usage-based theory of language, second language acquisition and learning, English language in Nigeria, competence and performance, ethnolinguistics and identity, as well as the pedagogic issues that remain critical in the Nigerian academic setting. In like manner, this study has contributed strongly to methodology and its essential practices. This study is expected to be useful to the UNESCO, Federal Ministry of Education (FME), National Universities Commission (NUC), Nigerian Education Resource and Development Council (NERDC), National Teachers Institute (NTI), Universal Basic Education (UBE), National Association of Educational Administration and Planning (NAEAP), National Commission for Mass Literacy and Adult and Non-Formal Education (NCMLA & NE), among others, in the process of educational policy implementation in Nigeria.

In conclusion, the analysis presented in this study boulevards variation in the processing of determiners in the Nigerian undergraduate descriptive ESL writing framed within the corpus-based serene. In this study, gender and ethnicity, stand out as the most catching linguistic variables that demonstrate variation in the Nigerian undergraduates descriptive ESL writing processing determiners. In a specific manner, our study on gender and ethnic group is a kind of corpus-based and usage-based sociolinguistic, anthropological as well as ethnographic construction which caters for the flexibility and the variances of the linguistic variables to the assorted identity of people that shares similar ancestral languages, national experience, and sociocultural attributes.
References


https://doi.org/10.1016/j.jeap.2009.11.003

https://doi.org/10.1017/CBO9781139764377.004


https://doi.org/10.1075/sfcs.99.03har

He, Q. (2020). A corpus-based study of transfers in English nominal groups. Glottotheory, 10(1-2), 57-84. 
https://doi.org/10.1515/glot-2019-0003

https://doi.org/10.3389/fpsyg.2013.00255

https://doi.org/10.1017/CBO9781139013734.016

https://doi.org/10.1075/sfcs.99.16kle


https://doi.org/10.1017/S0047404500006576


https://doi.org/10.1017/CBO9781139764377.009

https://doi.org/10.18806/tesl.v31i2.1178

https://doi.org/10.3389/fpsyg.2020.01088

Johor, Malaysia.


https://doi.org/10.4236/ojml.2014.42026


https://doi.org/10.5430/wjel.v6n3p42


https://doi.org/10.1111/j.1467-9841.2010.00455.x


https://doi.org/10.1017/S0142716400009930


https://doi.org/10.18806/tesl.v30i2.1138


https://doi.org/10.18806/tesl.v31i2.1179


https://doi.org/10.1017/CBO9780511792519.025


https://doi.org/10.1002/9781118335598.ch6


https://doi.org/10.1075/ihll.8.01tag


https://doi.org/10.1017/CBO9780511576164.005


https://doi.org/10.1002/tesq.539