

EMERGENCE OF NEW PREDICTORS PROJECTING THE DEFINITE ARTICLE VARIABILITY: EVIDENCE FROM NIGERIAN ENGLISH

MAYOWA AKINLOTAN

Vrije Universiteit Brussels, Belgium
University of Texas at Austin, USA

Abstract. An important syntactic system in the noun phrase (NP) is the definite article system. The definite article system in emerging varieties of English, such as Nigerian and Indian, have been shown to manifest varying degrees of variability in their usages, given different contexts (Platt, Weber and Ho, 1984; Wahid, 2013; Akinlotan, 2016b; Akinlotan 2017). In addition to the fact that little has been done in quantifying this phenomenon, too little number of predictors explicating the scenarios where we might find certain definite article usages in different but specific contexts in emerging varieties has also been put forward. Following Wahid (2013), a revision of Hawkins's (1978) theoretical framework for the definite article usages, together with test statistic, the present study investigated 19276 tokens of *the*, spread across seven text types of academic, media, learner, interactional, popular, literary, and administrative. The study which Akinlotan expands (2017) shows that previously untested predictors of presence/absence of premodification and determiner structure, animacy and class of the head noun, and syntactic function of the NP, account for variability in the definite article usage in our corpus. In fact, these newly tested predictors show stronger influence than a well-known predictor of register (Biber et al., 1999; Wahid, 2013).

Key words: definite article usage, variability, register, Nigerian English, linguistic predictors, New Englishes, corpus linguistics

INTRODUCTION

Investigating variability in various linguistic choices has received some attention in the literature of non-native varieties of English. Results from different studies have shown evidence of idiosyncrasies in the various syntactic structures/systems studied. For instance, variability has been established in the use of the definite article in both the established and emerging varieties of English. Variability of the definite article system in Nigerian English has received no attention in literature. In the light of Wahid's (2013) and Akinlotan's (2017) important work on the variability of the definite article usages across eight different varieties of English, it is important to provide insights into Nigerian variety, at the same time testing more than two predictors of register and variety.

Until Wahid, many studies have either provided small datasets from which generalisations are made (see Lamidi, 2007, for the definite article in Nigerian English; Platt, Weber and Ho, 1984, for Singaporean and other Asian Englishes) or have focused on issues of omission, underuse, overuse, and misuse (Ionin, 2003; Lamidi 2007). Having provided some evidence from emerging varieties (Kortmann and Szmrecsanyi, 2005; Kortmann, 2006; Wahid, 2013; Akinlotan, 2016a; Akinlotan, 2017), identifying factors which influence the choices should suffice. In other words, apart from the established factors representing register and variety, which Wahid (2013) further attests to, it is important to find out whether the definite article usage relate to factors other than the often-tested ones.

On the basis that more specific explanations viz-a-viz predictors are required for clearer understanding of the syntactic, semantic and pragmatic idiosyncrasies characterising every variety, the present study shows the extent to which the definite article usages can be explained in terms of predictors which have never been tested before. In other words, in addition to register and syntactic functions, which have been found to be well-known factors motivating linguistic choices including the definite article usage, the following predictors representing (1) animacy of the head noun in the NP (2) noun class of the head noun in the NP (3) number of the head noun (4) presence or absence of a premodifier, and (5) presence or absence of any other determiner, are studied.

The choices of these factors are motivated by the reports of strong influences they have on linguistic choices involving alternation (see strong influence of animacy on genitive alternation in Nigerian English, the relationship between number disagreement and determiner structure in Akinlotan, 2016b, the relationship between presence or/and absence of a premodifier and other syntactic elements within the NP in Akinlotan and Housen, 2017). The hypothesis is that these factors, given their prior demonstration of a positive relationship with linguistic choices/alternation in the same variety, would shed light on the variability of the definite article usages in our variety, and can indeed be generalizable to similar varieties. In this way, it will be shown how previously known and unknown predictors compare in motivating choices, showing the extent to which our new candidate predictors explain the choices. Consequentially, it would then emerge whether Wahid's hypothesis about the strength of register as a good predictor is manifested in our corpus or whether syntactic function once again outweighs register, as Akinlotan and Housen (2017) found.

HAWKINS' DEFINITE ARTICLE FRAMEWORK

Hawkins (1978) provided a theoretical framework of the definite article usage, which is described in this section, following a further revision from Akinlotan (2017). First is the definite article use in the anaphoric sense, which points back to the referent which had been mentioned earlier in the text. For instance, 'the book' refers to 'an interesting book' in: *Fred was discussing an interesting book in his class. I went to discuss the book with him afterwards.*

The second type of use is the associative anaphoric usage. As it suggests, it is about inductive meaning processing, associating with the referent mentioned earlier. In the example, *Bill drove past our house in a car. The exhaust fumes were terrible*, ‘fumes’ is associated with ‘car’; therefore, associative anaphoric use has occurred.

The third use is what Hawkins (1978) termed the ‘Immediate situation use’, which has two sub-levels (1) the visible situation use and (2) the immediate situation use. In the visible situation use, what is being referred to is immediately present/visible, for instance, *Pass me the bucket*. On the other hand, what is being referred to in the immediate situation use is not immediately present/visible (e.g. *Beware of the dog*). These subcategories overlap, and can be collapsed into as contextual.

The next use is the ‘larger situation use’, which, in the present study, is still part of the contextual use as it deals with context, just as the visible, and the immediate situation use. The larger situation use, in the sense of Hawkins (ibid.: 115), requires the speaker’s and listener’s knowledge of larger context than it is in the case of visible and immediate situation uses which require knowledge of small context. Furthermore, this use also has two sub-levels; the specific knowledge use and the general knowledge use. In the specific knowledge use, some sort of specific knowledge about the larger situation is shared by the speaker and the hearer. For instance, the shared knowledge of a particular *Gibbet* is expressed in the following example, *The Gibbet no longer stands*. Whereas the general knowledge use, according to Hawkins (ibid.), refers to ‘a general knowledge of the existence of certain types of objects in certain types of situation’. General knowledge about the on-going event suggests that the hearer can correctly identify *the bridesmaids* in *Have you seen the bridesmaids?* Again, in the framework for this study, we collapse all of these uses into the contextual use, since they still require matching the referent to context in the world.

The fifth usage type is unfamiliar uses in a noun phrase with explanatory modifiers. As it suggests, it means this use refers to NPs containing modifiers that identify the referent in the applicable context. According to Hawkins (ibid.: 137), there are four structural types of modifiers that can explain the referent: (1) establishing relative clauses for example, *what’s wrong with Bill? Oh, the woman he went out with last night was nasty to him*. The argument (‘the woman’) in the clause, ‘The woman he (Bill) went out with last night was nasty to him’ makes the clause establishing ‘what’s wrong with Bill?’; (2) associative clause; which is a clause that contains the associates of the referent, in the example *I remember the beginning of the war very well*; (3) NP complements; in this case the NP is complemented by a clause that links the referent in the NP to a particular context, for example, ‘the fact’ in Bill’s amazement is linked to ‘there is so much life on Earth’, in *Bill is amazed by the fact that there is so much life on Earth* and (4) nominal modifiers; the presence of modifier aiding the identification of referent, for example, ‘red’ functions as a modifier, shedding light on the type of colour, in *I don’t like the colour red*. The sixth definite article usage by Hawkins is unfamiliar

uses with unexplanatory modifiers. As it suggests, the modifier does not explain the referent, for example, *My wife and I share the same secret*. The premodifier *same* does not explain 'the secret'. As the fifth and sixth usage shows, they can be collapsed into one category of structural use, which is proposed in the present study. Although Prince (1981, 1992), Wahid (2013), Akinlotan and Housen (2017) recategorized Hawkins' classification, their points of difference are not relevant to the objectives of the present study. In the methods section, a tabular presentation of the re-conceptualisation is presented.

Hawkins (1978) only provided us with a detailed yet overlapping classification of the definite article usage but with very little information as to which specific situations we expect to find certain uses. So far in the literature of New Englishes, the definite article usage has only been investigated in the contexts of register and variety as determinants (Wahid, 2013). Comparing different varieties of English, Wahid's expectation of variety emerging a better predictor than register was not clearly borne out. In other words, Wahid found register, rather than variety, a stronger predictor explicating uses in eight varieties of English. The fact that register outweighs variety validates Biber et al.'s (1999) claim that register is a very strong determinant constraining variation. On the basis of Wahid and Biber et al., we expect text type/register to emerge as a strong predictor constraining the choice of the definite article. It will be shown that certain text types/register prefers certain usage of the definite article to other. Akinlotan (2017) and Akinlotan and Housen (2017) have shown stiff competition between syntactic function and register. For instance, Akinlotan (2017) found variety, rather than register, asserting stronger predictive strength. In addition to these two known predictors representing (1) syntactic function, and (2) register, the present study proposes and engages now candidate predictors representing (3) animacy, (4) class, (5) number of the head noun, (6) presence or absence of a premodifier, and (7) presence or absence of another determiner. Only their independent effects will be shown in this study, leaving a multivariate analysis for future endeavor.

Determiners, including the definite article, have a strong semantic relationship with nouns or any other lexical items with which they appear (Akinlotan, 2016a). In other words, their syntactic and semantic processing is largely dependent on co-occurring elements. Given that the definite article could not be correctly identified or processed without recourse to the other occurring structural components in the entire NP which bears the morpheme, it is important to find out to which extent these structural and semantic elements assert the influence on choices of the definite article. For example, the NP, *the slow-moving car* consists of premodification and a noun, both of which are partly responsible for the classification of the definiteness manifested. Also, since Hawkins' classification can be distinguished on the basis of structure (see usage type 5 and 6) and of meaning (see usage types 1, 2, 3, and 4), then the syntactic position of the NP and the presence/absence of a premodifier and other determiner elements become important. Evidence for these expectations

has been provided in Schilk and Schaub (2016), Akinlotan and Housen (2017), and Akinlotan (2017). These works showed that the syntactic position of the NP can determine the presence or absence of a certain structural element, such as a premodifier, the definite or indefinite article. On the basis of these findings, we expect syntactic function that NP performs to influence preference for certain definite article usage, such that the definite article choices can be explained in terms of the grammatical position of the NP.

According to Hawkins' process of identifying different usages of the definite article, the presence and/or absence of a premodifier may impact on the easy identification of the referent in an NP. Relatedly, Wahid (2013: 33) also found that certain structural elements, such as a prepositional phrase, influence the usage choices of the definite article. Wahid also found that the syntactic usage of the definite article reflects some marked uses in which definiteness is marked on the postmodification. Therefore, on the basis of Akinlotan's (2017), Wahid's (2013) and Hawkins' (1978) argument that modifier is crucial to interpreting the definite article usage, we expect to find a positive relationship between the premodifier and the definite article usage in our corpus. More specifically, since certain uses (e.g. syntactic usage) are structurally more associated with the presence of a premodifier, we expect a premodified NP to significantly associate with syntactic usage.

Given the semantic import/feature of animacy, and on the basis of its asserted influence on structural choices in previous studies, it is also expected that some kind of relationship with usage type will suffice. The relationship between animacy and noun class implies that animate proper nouns are less likely to require the definite article for their identification. In other words, it is expected that certain usage types, such as contextual use and syntactic use, will have a negative relationship with NPs consisting of such animate nouns as the head noun. Also, on the basis of the fact that animacy has been shown to influence constructional choices in genitive alternation in our corpus, it is expected that animacy will relate with choices, alternating between one to the other. More specifically, we expect anaphoric use to associate with inanimate head noun and to disassociate with animate head noun.

The class of the head noun is also expected to assert some sort of relationship. Given that certain definite article usage, as shown in Hawkins' framework, is highly dependent on whether the referent is visible or not immediately visible, it is expected that the class of the head noun (concrete or abstract) will influence the definite article choices, such that a preferential pattern emerges. The expectation is also supported by Butler (2002) who showed that Japanese speakers considered the status of the head noun before using the definite article. What Butler has shown is the influence of mother tongues on the acquisition and use of the definite article. It is also expected that some transfer influences of the definite article system operating local Nigerian languages will suffice, such that the status of the head noun in an NP is considered before choosing the definite article usage. More specifically, it is thus expected that NPs

consisting of concrete head nouns will associate more with contextual use than with an abstract head noun.

Furthermore, the noun number, i.e. whether the head noun is singular or plural, has been shown to have influence on the determiner structure in Nigerian English NPs (Akinlotan, 2016b). Hence, it is expected that the number of the head noun will influence the definite article choices. Akinlotan (2016a) has also shown that combining determiners in the Nigerian English determiner system is rare, and when it occurs, the number of the head noun is likely to disagree. Lamidi (2007) has also shown that unconventional determiner combination (for instance, see ‘this’ and ‘your’ in ‘I like this your shoe’) in Nigerian English inhibits easy accessibility to and processing of the referent in the NP. Tagliamonte et al. (1997), who tested how different types of determiner influenced plural marking in Nigerian Pidgin English, also suggests that the number relates to determiner choice in Nigerian Pidgin. They found that a certain type of determiner associated with the nominal referent in plural marking. If this is the case in Nigerian Pidgin English, we can also expect a relationship between the number of the head noun and the definite article choices on the one hand, and that the determiner structure will show a positive relationship with the usage type on the other hand. Following Tagliamonte et al. (1997), Lamidi (2007) and Akinlotan (2016a), we hope to show the extent to which the definite article is combined with other determiner structures, and how this combination influences choices of the definite article usage.

Following the previous findings in text type, it is predicted that text type will assert the strongest influence of all the predictors. According to the previous findings in Akinlotan and Housen (2017) and Akinlotan (2017), it is expected that syntactic function will strongly compete with text type for influence. In the light of Fraurud’s (1990) claim that ‘the definite article is not most prominently used for its textual function’, it is expected that contextual use will emerge the most commonly used in our corpus.

DATA AND ANNOTATION PROCEDURES

Expanding significantly the datasets (and the hypotheses) from Akinlotan (2017), noun phrases appearing with the definite article were extracted from 15 text types (excluding academic, technical and social science) in the written component of the Nigerian section of the International Corpus of English. Antconc Tools (Anthony, 2014) were used for the extraction of a definite noun phrase (e.g. *the men that I saw yesterday*) from the text types. For the purpose of the present study, the 15 texts are reclassified. The reclassification is necessary for a clearer register classification, which Nigeria-ICE lacks (see Akinlotan and Housen, 2017, for Nigeria-ICE reclassification, and Akinlotan, 2017, for different conceptualisations of the definite article). Table 1 shows the reclassification of the text types, together with the frequencies of the token *the*.

*Table 1 Text type classification and the token extracted from them**

Text type in this study	Text type in the corpus	Tokens
Academic	Academic humanities (11) and social science (11)	2693
Administrative	Administrative (11) and business (18)	2632
Media	Reportage (26) and editorial (28)	2792
Learner	Student essay (12) and examinations (49)	2724
Literary	Novel	2929
Popular	Popular humanities (11), social science (11), natural science (11), and technology (14)	2747
Interactional	Social letter (48) and skill hobbies (25)	2759
7 categories	15 categories	19276

* In the brackets is given the number of texts used in each textual category. For instance, 11 texts (all of which were used in the present study) make up the academic textual categories. Examination category consists of 54 texts but only 49 were used in the extraction process.

For the purpose of a relatively balanced corpus, additional texts matching the ICE criteria for inclusion were accessed. The text types representing novel, skills and hobbies, and social letters did not return sufficient tokens. For the literary category, additional tokens were extracted from an online published literary work entitled ‘Cupid Risk Series’ (Whitman (ed.), 2000). This text publishes a number of Nigerian short stories from different backgrounds. For the interactional category, additional tokens were extracted from similar texts which have similar themes with those of ICE, and are addressed to similar audience on an online forum *Nairaland*. The same comment as above. The 19276 tokens were then coded for six usage types shown in Table 2. The six usage types used in this study emanate from Hawkins (1978), Prince (1981), Wahid (2013), and Akinlotan (2017).

Furthermore, these six choices of usage are analysed in the context of six variables discussed above. Following the ideas of Schilk and Schaub (2016), and Akinlotan and Housen (2017), two syntactic functions, subject and non-subject, are accounted for in the present study, though eight syntactic functions were completed in the actual coding. For example, the NP ‘the brilliant Hausa politician’ in the clause *the brilliant Hausa politician has promised to restructure the entire country* functions as the subject of the clause, whereas ‘the entire country’ functions as the direct object (non-subject) of the verb *restructure*. In other words, syntactic functions other than subject positions (e.g. direct object, prepositional phrase, apposition, subject complement, indirect object), are collapsed into a category called the non-subject syntactic function. For the animacy of the head noun, a binary scheme of animate or inanimate is applied.

Table 2 Comparison of the classification of the definite article usage in the study
(based on Hawkins, 1978; Prince, 1981; Wahid, 2013 and Akinlotan, 2017)

HAWKINS (1978)	WAHID (2013)	PRESENT STUDY
Anaphoric use	Textual	Anaphoric
Immediate situation i. Visible situation use ii. Immediate situation use	Situational	Contextual
i. Visible Situation Use		Contextual
Larger situation use i. specific knowledge in the larger situation ii. general knowledge in the larger situation		Contextual
Associative anaphoric use	Associative anaphoric use	Inductive
'Unfamiliar' uses in NPs with explanatory modifiers i. Establishing relative clauses ii. Associative clauses iii. NP complement iv. Iv. Nominal modifiers v. 'unexplanatory' modifiers	Structural	Structural
'Unfamiliar' uses in NPs with unexplanatory modifiers		Structural
	Non-referential e.g. He is the leader	Stylistic
	Generic e.g. The computer is a wonderful thing	Stylistic
	Idiomatic e.g. Grab the bull by its horns.	Stylistic
	Repair e.g. The, er, the project is useless.	Contextual
	Repeat e.g. The, the man came early today.	Contextual
	Unknown e.g. the guy (with insufficient context for identification)	Unfamiliar

Following Zaenen et al.'s (2004) animacy encoding, company, organisation and geographical names (e.g. NNPC, WHO, Lagos), and proper nouns that identify humans, animals, and supernatural beings are classified as animate, while all other entities are classified as inanimate.

Next is the number of the head noun, which is a two-way distinction of singular or plural. When a collective noun is used in the sense of a singular noun (e.g. *The President's team is poor* versus *The President's team are professionals*), such syntactic status is coded accordingly. The class of the head noun is coded as either abstract or concrete. In other words, the referent in the NP is taken to be physically identifiable (concrete) or not (abstract). Akinlotan and Housen (2017) have shown that the presence/absence of certain NP element (e.g. premodifier) influences the presence/absence of another element (e.g. postmodifier). Such a finding suggests that the presence/absence of certain NP element will indeed influence the choice of the definite article usage. More specifically, Akinlotan (2016b), Akinlotan and Housen (2017) show that structural components of the NP in Nigerian English are well predicted on the basis of their relations to one another. For modification as a variable, the presence or absence of a premodifier within the whole NP is coded for. Though not reported here, but in the actual coding the length of the premodifier is measured in 1, 2, 3, 4 or more words. In future research, it might be worthwhile to show whether the weight/length of a premodifier influences the choice of the definite article. In the present study, only the presence (premodified, i.e., a premodifier of any length) and absence (non-premodified) of premodification is reported. Another factor accounted for is the definite article combination with other determiner; thus, whether the definite article is used independently or cooperatively with another determiner of any sort, is annotated for.

RESULTS

In this section, the results of the variables tested are presented. While the definite article choices can be explained as a semantic/pragmatic category, the present study treats these choices as categorical; such that a token/usage analysed as anaphoric could not be further analysed as contextual. Such analysis is essential in the semantic/pragmatic approach, even though there are clear cases of the definite article usage that could not be analysed both ways. Using a cross tabulation, the preferential patterns are shown. These distributions are tested by means of chi squares test of independence showing whether there is a significant underlying relationship between the choices and the predictors. Given the focus of the study, which is to show the independent effect of different variables on the definite article choice, the chi square test of independence is able to provide insights into the underlying pattern. Each analysis is related to the hypothesis being tested, and followed by the test statistic result of the hypothesis, using the chi square test of independence. All analyses were carried out using SPSS.

1 ANIMACY

The distributions of usage types in relation to the animacy of the head noun in the NP are provided in Table 3 below.

Table 3 Article usage by animacy

	Ana-phoric		Con-textual		Inductive		Syntactic		Stylistic		Un-familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Inanimate Head	3269	20	6730	42	1607	10	3807	24	467	3	162	1	16042	100
Animate Head	1277	34	1120	35	327	10	350	11	120	4	40	1	3234	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

Since there is no prior evidence as to whether the animacy of the head noun in a noun phrase relates to the definite article usage type, we conducted a chi square test of independence to show what kind of relationship exists between the head animacy and usage type. The result ($\chi^2 (5) = 667.099$ $p < 0.0000$) shows that there is indeed a relationship between the animacy of the head noun in the definite NP and usage type. In other words, as the table shows, there is some kind of relationship between the head animacy and usage type: when the head in the NP is inanimate, the definite article is likely to function as contextual use (42%), which is not the case when the head is animate. When the head is animate, the article usage is likely to function as anaphoric (34%) or as contextual (35%). While contextual use (42%) is more likely to occur with inanimate head than with animate head (35%), anaphoric use (34%) is more likely to occur with animate head than with inanimate head (20%). A similar preferential pattern is found in the use of syntactic definite article which is more likely to occur in an NP with inanimate head (24%) than with an NP with animate head (11%). Furthermore, syntactic use is more likely to occur (24%) with inanimate head than with animate head (11%), whereas no pattern is found in inductive, stylistic, and unfamiliar uses, irrespective of the animacy status of the head in the NP. Our expectation of anaphoric use to associate with an inanimate head noun is not borne out. Rather, anaphoric use prefers animate head to inanimate head at 34 per cent versus 20 per cent respectively.

2 CLASS

Table 4 shows the relationship between the article usage and class of the noun – either concrete or abstract.

Table 4 Article usage by class

	Ana-phoric		Con-textual		Inductive		Syntactic		Stylistic		Un-familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Concrete Head	2970	36	3110	38	854	10	863	11	231	3	109	1	8137	100
Abstract Head	1576	14	4740	43	1080	10	3294	30	356	3	93	1	11139	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

As with the animacy of the head noun and article usage type, the relationship between the class of the head noun and the article usage type is also statistically significant $\{\chi^2 (5) = 1818.435 \text{ } p < 0.0000\}$. It reveals that some positive relationship exists between the class of the head noun and choice of the definite article usage. While contextual use is strongly preferred when the head is abstract (43%), the case is the opposite for a concrete head, which is closely distributed between anaphoric (36%) and contextual (38%) uses. While the competition for the most preferred article usage is between contextual and anaphoric (38% versus 36% respectively) when the head noun is concrete, the competition for the most preferred article usage when the head noun is abstract is found with contextual and syntactic (43% versus 30% respectively).

Similarly, while anaphoric use is more likely to occur with concrete head than with abstract head (36% versus 14% respectively), a clear preferential pattern is not found with the contextual use, though there is a very slight preferential pattern for concrete head. Furthermore, it is found that syntactic definite article usage associates more with abstract head (30%) than with concrete head (11%). In other words, abstract head syntactic use is more likely to be used than concrete head syntactic use (30% versus 11% respectively). As with animacy, there is no trend found in inductive, stylistic, and unfamiliar uses. The absence of a pattern suggests that whether the head noun is abstract or concrete is almost completely unrelated to the choice of the definite article usage. Again, our expectation that a concrete head noun rather than an abstract head noun is to attract contextual use is not borne out. Rather, abstract head noun contextual usage (43%) is more likely to be used than concrete head noun contextual usage (38%). One explanation of this result could stem from the fact that abstract head nouns may not be as easily identifiable in other usage types as it is the case with contextual use, where more information about the context/ referent is supplied.

3 NUMBER

Table 5 shows the number of a noun, singular or plural, relating to the choice of article usage.

Table 5 Article usage by number

	Ana- phoric		Con- textual		Inductive		Syntactic		Stylistic		Un- familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Singular Head	3708	24	6509	41	1412	9	3478	22	494	3	151	1	15752	100
Plural Head	838	24	1341	38	522	15	679	19	93	3	51	1	3524	100
Total	4546	24	7850	41	1934	10	4157	2	587	3	202	1	19276	100

As with animacy and class, there is no prior evidence suggesting that a positive relationship exists between the number of the head noun in the NP and the definite article usage type. In order to find out the kind of relationship that exists between the number and article usage, the chi square test was conducted on the above observed distribution. The result is statistically significant $\{\chi^2(5) = 125.264 \text{ } p < 0.0000\}$, showing that choices of the definite article usages are related to the number of the head noun in the NP. In other words, whether the head noun is singular or plural is a variable capable of explaining the definite article usage types, at least in our corpus. This result confirms the expectation (a similar finding by Tagliamonte et al. (1997) in Nigerian Pidgin English) that the number of the head noun would associate with certain choices of the definite article and some specific statements could be made. Specifically, contextual use is the most likely usage type to occur, irrespective of the number of the head noun of the NP. In other words, whether the head noun is in the plural or singular (41% and 38%), the contextual usage type is likely to be used. The inductive usage type is more likely to be used when the head noun is in the plural (15%) than when the head noun is in the singular (9%). At a very small degree, the contextual usage type is more likely to be used when the head noun is in the singular (41%) than when the head noun is in the plural (38%). This pattern is also the case of the syntactic usage type. As can be seen, a singular head noun associates with the syntactic usage type more than a plural head noun does (22% versus 19%). In addition, the anaphoric usage type is likely to be used, irrespective of the number of the head noun (24%).

4 PREMODIFICATION

The influence of occurrence or non-occurrence of a premodifier on the choice of the article usage is presented below.

Table 6 Article usage by premodification

	Ana-phoric		Con-textual		Induc-tive		Syntactic		Stylistic		Un-familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Un-Premodified	3113	25	4851	38	953	8	3248	26	364	3	118	1	12647	100
Premodified	1433	22	2999	45	981	15	909	14	223	3	84	1	6629	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

As we can see from the table, the choice of the definite article usage is related to the presence/absence of NP modification $\{\chi^2(5) = 592.806 \text{ } p < 0.0000\}$. In other words, the usage type can be explained in terms of premodification occurrence (i.e. whether a premodifier does or does not co-occur with the definite article). Since the presence of (pre)modification contributes to the real-world processing of the referent in an NP, the positive relationship shows that certain definite article usage types associate with a premodifier, while certain usage types dissociate with a premodifier.

Furthermore, it is very likely for the inductive usage to co-occur with a premodifier (15%) than not (8%), whereas the syntactic usage follows the opposite pattern. That is, it is more likely for the syntactic usage to disassociate with a premodifier (26%) than to associate with a premodifier (14%). On the other hand, this clear preferential pattern is not found with the anaphoric and contextual usage types, even though these usage types are the most used choices. For instance, it is found that the contextual usage is more likely to co-occur with a premodifier (45%) than not (38%). This pattern is different from what is found with the anaphoric usage. While the anaphoric usage is more likely to co-occur with a premodifier (22%), it is not likely to co-occur (25%). According to our expectation that the syntactic usage will significantly relate with a premodified NP, the table shows this not to be the case. Rather, it is more likely for a non-premodified NP to associate with the syntactic usage (26%) than with premodified NPs (14%). Akinlotan and Housen (2017) have shown that a premodifier within the NP in Nigerian English is more likely to be omitted because Nigerian NP is generally typical of simple structures transferred from the structures of local Nigerian languages.

5 DETERMINER STRUCTURE

The relationship between the determiner structure and article usage is given below.

Table 7 Article usage by determiner structure

	Anaphoric		Contextual		Inductive		Syntactic		Stylistic		Unfamiliar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Single	4421	24	7501	41	1839	10	3993	22	533	3	177	1	18464	100
Combined	125	15	349	43	95	12	164	20	54	7	25	3	812	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

On the basis of Tagliamonte et al.'s (1997) finding that there is a relationship between the determiner type and NP structure in Nigerian Pidgin English, it is expected that the determiner structure will have some influence on the choice of the definite article usage. A test statistic shows that the determiner structure has a relationship with the definite article usage $\{\chi^2 (5) = 97.792 \text{ } p < 0.0000\}$. In other words, the definite article usage can be predicted or explained on the basis of the definite article co-occurrence with another determiner. As can be seen, there are clear preferential patterns found in stylistic and unfamiliar usages, while this is not the case with the anaphoric and contextual usages. The table shows that the stylistic definite article usage is more likely to occur when the definite article co-occurs with another determiner (7%) than when the definite article does not co-occur with another determiner (3%). A similar scenario is found with the unfamiliar definite article usage, which is very likely to be used when the definite article does not co-occur with another determiner (3%) than when it does combine with another determiner (1%). Also similar is the anaphoric usage, which is more likely to be used when the definite article is not combined with another determiner (24%) than when the definite article is combined with another determiner (15%).

Small preferential difference is found in the contextual and inductive usages. Both usage types are more likely to be used when the definite article is combined with another determiner (43% and 12% respectively). Unlike other predictors tested so far, it is only this determiner-occurrence variable that really explains the stylistic and unfamiliar definite article usages. Our specific expectation that a combined determiner would strongly associate with the contextual usage is slightly borne out at a percentage difference of 2%. This sparse variation correlates with Akinlotan (2016b) and Lamidi (2007) that determiners are rarely combined in the Nigerian NP.

6 SYNTACTIC FUNCTION

The extent to which a syntactic function relates to the article usage is presented in Table 8.

Table 8 Article usage by syntactic function

	Ana-phoric		Con-textual		Inductive		Syntactic		Stylistic		Un-familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Subject	1289	27	1909	40	495	10	889	19	118	2	38	1	4738	100
Non-subj	3257	22	5941	41	1439	10	3268	22	469	3	164	1	14538	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

The syntactic function performed by the NP in which the definite article appears is associated with the choice of the definite article usage $\{\chi^2(5) = 69.158, p < 0.0000\}$; when the NP functions as the subject of a clause, the anaphoric usage is likely to be used (27% versus 22%). Meanwhile, the contextual usage is less likely to associate with an NP that functions as the subject of a clause. No preferential pattern is found with the inductive usage, while it is found that the syntactic usage is more likely to be used when the NP functions as the non-subject of a clause structure (22% versus 19%). As with the determiner predictor, the syntactic function of an NP shows that some explanation can be made about the usage choice of the stylistic definite article. That is, the stylistic definite article usage is more likely to be used when the NP functions as the non-subject of a clause structure (3%) than when the NP is positioned at the subject position (2%). In other words, the stylistic usage correlates with non-subject NPs.

Unlike a premodifier and the number predictors, the inductive definite article usage is not found showing preference or having a strong relationship with the syntactic function. As can be seen, the inductive article usage is likely to be used irrespective of the syntactic functions of the NP. In other words, whether the NP is positioned at the subject or non-subject positions has no influence on the choice of the inductive usage whereas influences of the premodifier and the number on the inductive usage are found. Also, it can be seen that the syntactic function has no relationship with the unfamiliar usage. On the basis of the previous findings (e.g. Akinlotan, 2016b; Akinlotan and Housen, 2017), which reported a strong influence of the syntactic function on structural choices, our expectation of a similar strong influence is not met. It is expected that the syntactic position of the overarching NP will influence the choice of the definite article usage, namely, the definite article choices can be explained in terms of the grammatical position of the NP that bear the morpheme. As it can be seen, this is not clearly borne out, though the anaphoric usage is more likely

to associate with the subject NP than with a non-subject NP. On the other hand, a non-subject NP is more likely to associate with both contextual and syntactic usages than any other usages.

7 TEXT TYPE

Text type which has been found to influence syntactic choices is presented in relation to the choice of the definite article.

Table 9 Article usage by text type

	Ana-phoric		Contextual		Inductive		Syntactic		Stylistic		Un-familiar		Total	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Academic	736	27	896	33	391	15	544	20	94	3	32	1	2693	100
Administrative	745	28	912	35	255	10	618	23	63	2	39	1	2632	100
Media	677	24	1013	36	333	12	632	23	107	4	30	1	2792	100
Learner	769	28	979	36	228	8	665	24	47	2	36	1	2724	100
Literary	477	16	1485	51	199	7	677	23	72	2	19	1	2929	100
Popular	660	24	1036	38	381	14	554	20	90	3	26	1	2747	100
Interactional	482	17	1529	55	147	5	467	17	114	4	20	1	2759	100
Total	4546	24	7850	41	1934	10	4157	22	587	3	202	1	19276	100

There is prior evidence suggesting that there is a strong relationship between the definite article usage and the text type that produces them (Biber et al., 1999; Wahid 2013). On the basis of these previous findings (Biber et al., 1999; Wahid, 2013), the chi square test of independence was carried out to show the extent to which this hypothesis is manifested in the definite article usage alternation. The result shows that there is a statistically significant relationship between the text type and the definite article usage $\{\chi^2 (30) = 801 \text{ } p < 0.0000\}$. In other words, there are some indicators that the text type determining the definite article usage can explain their distributions and preferential patterns. Noticeable and strong indicators emerge in the literary and interactional text types in which the contextual usage is very much the preferred choice (51% and 55% respectively). Unlike other predictors/context, the text type does not clearly explain the distribution of contextual usage to the extent that specific statements can be made.

This lack of a clear pattern is also reflected in the inductive usage, which is the least used choice at 7 per cent and 5 per cent respectively. The patterning of the inductive usage in a learner text type, which is the third least preferred at 8 per cent is related to anaphoric and contextual distributions. Apart from these two text types which clearly demonstrate the preferential pattern in the contextual and inductive usages, no other text type shows a strong relationship with

the definite article usages, so that the expected strong effect of text type could be ascertained. In other words, there is no significant preferential pattern which allows making specific statements. As it can be seen, the stylistic and unfamiliar usages show a similar trend irrespective of the text type which determines them. Similarly, the same trend, a sparse variation and obscure patterning, can also be seen with the anaphoric and syntactic usages, which are distributed irrespective of the text types that produce them. For instance, we can observe the significant explanatory power of the determiner predictor on the stylistic and unfamiliar usages, a scenario that is missing in the text type, a well-known variable of language use/constructional choices. In other words, the expectation that the text type will reveal much about variability in the definite article usage in our corpus is not met.

8 SYNTHESIS OF THE INDEPENDENT EFFECTS OF PREDICTORS TESTED

In this section, an analysis showing the aggregate effects of all predictors analysed is presented, allowing for a quick comparison among the effects.

Table 10 shows that some expectations are borne out, while some are not. There are clearly some divergences and convergences with the findings in the previous study by Akinlotan (2017). What the present study has added to the previous study is that there are more predictors than the common ones found in the literature that can shed light on the predictability of the definite article usage in different varieties of English. However, the emergence of these predictors, since they have not been tested and found in the article system of other varieties, could only be attested in Nigerian English variety, in the least. Therefore, this study has not only brought forward new predictors adding to the commonly tested ones but has also highlighted that the commonly tested ones such as the syntactic function and register can vary in their influences, given different scenarios that include datasets, idiosyncrasies of the variety in question, operationalization of the predictors (e.g. we have operationalized the register in a different way from the previous study). For instance, until now, the extent to which the aspect of the NP complexity influences the choice of the definite article usage has not been clear.

As the Table shows, there is a relationship between the definite article usage and (1) presence/absence of premodification, and (2) the syntactic functions of the NPs bearing the definite article. On the other hand, the expectation of a very strong relationship between the text type/ register and article usage is not met (Biber et al. 1999; Akinlotan and Housen, 2017). The re-classification of the text type may have impacted the desired results. Nevertheless, the register/text type still provides insights into where we are likely to find certain uses, and where we should least expect certain uses. At any rate, we have shown that the definite article usage may not only be predicted, but also that such prediction might be more insightful than others, depending on what variables are considered.

Table 10 Summary of effects of independent variables tested

	Anaphoric	Contextual	Inductive	Syntactic	Stylistic	Unfamiliar
Inanimate head	-	+(7%)		+(13%)	-	
Animate head	+(14%)	-		-	+(1%)	
Concrete head	+(22%)	-		-		
Abstract head	-	+(5%)		+(19%)		
Singular head		+(3%)	-	+(2%)		
Plural head		-	+(6%)	-		
Unpremodified	+(3%)	-	-	+(12%)		
Premodified	-	+(7%)	+(7%)	-		
Single	+(9%)	-	-	+(2%)	-	-
Combined	-	+(2%)	+(2%)	-	+(4%)	+(2%)
Subject	+(5%)	-		-	-	
Non-subject	-	+(1%)		+(3%)	+(1%)	
Literary	-	+(18%)	-	-	-	
Interactional	-	+(22%)	-	-	-	

The study further shows that certain usage choices are more likely to occur than others. The interplay of two systems (i.e. the definite system in English and that of Nigerian local languages) might partly influence the high preference of contextual uses, which demonstrate an intersection of meaning and interpretation in language use. Lamidi (2007) has argued that the use of the definite article in places where standard variety, such as British English, would omit it is traceable to the influence of the system in the mother tongue. As Table 10 shows, we can infer that the contextual use is the most diversified choices (preferential pattern in all variables are decided), followed by the syntactic, then anaphoric use.

Table 10 details conclusions that can be made from our univariate analysis by providing a summary of the predictors in relation to the choice of the definite

article. The sign + denotes a preference for a choice, while – shows a disfavor (or less preferred in relation to the corresponding variable). An empty cell shows that no preference is found (e.g. 10% versus 10% in the subject versus non-subject position in the inductive use). The extent of preference is calculated using higher percentages, irrespective of the gap. For instance, the difference between the contextual use in a syntactic function is one percent (1%) (41% favours the non-subject position, whereas 40% favours the subject position). The percentage difference (see also Akinlotan and Housen, 2017) is given in the brackets (e.g. 1%). Meanwhile, the difference between premodified and unmodified choices in inductive use (15% versus 8% respectively) is 7%, which is higher than 1% in the syntactic function. This difference shows the extent of influence, some sort of relative strength that can only be best captured with a further multifactorial regression analysis. However, we can still infer the scale of relative strength on the basis of percentage difference vis-à-vis the clearest predictive strength.

As can be seen, the highest difference is found in the class of the head noun (22% in the anaphoric use, and 19% in the syntactic use). Next is predictor animacy with 14 per cent and 13 per cent for the same choice of the definite article in the anaphoric and syntactic usage respectively. Third on the scale is a modification predictor with 12 per cent for the syntactic use, and 7 per cent each for the inductive and contextual uses. In the text type, we can see that two text types stand out: literary and interactional. The interactional text type has the highest percentage for the contextual use (55%), which, when deducted from the least percentage in the academic text type (33%), returns a very high preferential rate (22%). A similar situation is found in the literary text type (18%). Unlike other predictors, the variability in the text type is not spread across the choice of the definite article, as Akinlotan (2016a, 2016b) found in the study of the determiner structure in the same variety. Akinlotan (2016b) found that the text type provided in-depth insights into the variability of determiner choices. To sum up, it can be observed that the most predictable choice of the definite article is the anaphoric, syntactic, and contextual usage.

CONCLUSION

The present study has expanded the literature on the definite articles usages in Nigerian English, and in new varieties of English at large. Many previous studies have shown the tendency to omit, overuse, and underuse the definite article in Nigerian English (Lamidi, 2007; Akinlotan, 2016b), but none has ever shown the predictability of their usages. The study shows that the usage of the definite article in Nigerian English can indeed be predicted, and the hypothesis of omission, underuse and overuse is more likely traceable to contextual uses than other usage of the definite article. In other words, if contextual uses are highly interpretative, linking the meaning (semantics), context (pragmatics) and grammar (syntax), then influences from local Nigerian languages are very

much likely to be responsible for these patterns. This pattern appears to correlate with Lamidi (2007), who provided similar evidence in the interpretative sense of the definite article in Nigerian English variety. If Gut (2005) could find that speech rhythm in Nigerian English variety is distinctly different from the speech rhythm of Hausa, Igbo, and Yoruba (the three major languages in Nigeria), then the motivations for the patterns that emerged may not be completely ascribed to cross-linguistic influences.

No matter how small they might appear, according to Gries and Mukherjee (2010: 28), the speakers' other language will likely play a role in L2 construction choices. Although Wahid (2013) did not stress such influence, the significant preference for contextual uses in our corpus suggests that the different definite article systems of meaning and interpretation in Nigerian local languages are transferred, to some varying degree, into the Nigerian variety of English. Such interference is plausible given that Ionin (2003) has shown the existence of a strong relationship between the definite article acquisition and processing. In another perspective, the usage pattern might perhaps reflect the process of fossilization, including those temporary and permanent instances. Although it is difficult to measure, it can be argued that certain usages, such as unfamiliar or syntactic uses in 'older' varieties such as British English variety, may have been retained as contextual uses in the Nigerian English variety. Apart from the influences of transfer from local languages, other motivations, according to Mair (2002), may emanate from tendencies to hypercorrect and/or avoid certain usages/constructions.

Furthermore, the paper has shown the extent to which the choice of the definite article is projected by a number of previously tested and untested predictors. The effects of animacy, presence or absence of premodification, and the number of the choice of the definite article have been shown along the effect of a well-known predictor register (Biber et al., 1999). The hypothesis by Biber et al. that register is the most potent predictor of variation is not clearly found in our data. Perhaps this is due to the difference in datasets, which invariably shows that there are indeed structural differences between the Nigerian and British varieties of English. Biber et al.'s findings have resulted from the established varieties, especially American and British varieties of English. The stiff competition between the register and syntactic functions predicting choices corroborates Akinlotan and Housen's statement (2017) who, also, have shown the competition between the syntactic function and register predicting NP complexity. In other words, the hypothesis that register is a universal potent predictor of variation may well be rephrased as a regional potent predictor of construction choices. The limitations of the present study imply that these new predictors motivating the choice of the definite article may not be generalized onto other similar new varieties or established ones until they are tested, given that different varieties, old and new, are inherently peculiar (Akinlotan 2017). Also, a regression analysis of the predictors found significant in the present study is required in a further study, so that the extent of interaction among these predictors can be found.

REFERENCES

- Akinlotan, M. (2016a) The effects of structure and proficiency on determiner number (dis)agreement in Nigerian noun phrase. In A. M. Ortiz and C. Perez-Hernandez (eds.) *Proceedings of the CILC2016 8th International Conference on Corpus Linguistics* (pp. 1–8). Malaga: Spanish Association for Corpus Linguistics. Available from <http://easychair.org/publications/paper/> (epic series in language and linguistics) [Accessed on 4 August 2017]
- Akinlotan, M. (2016b) Genitive alternation in New Englishes: The case of Nigerian English. *Token: A Journal of English Linguistics*, 5 (1): 59–73.
- Akinlotan, M. and Housen, A. (2017) Noun phrase complexity in Nigerian English: Syntactic function and length outweigh genre in predicting noun phrase complexity. *English Today*: 1–8. DOI: 10.1017/S0266078416000626.
- Akinlotan, M. (2017) Predicting definite article usages in new varieties of English: syntactic function outweighs register. *Anglica: An International Journal of English Studies*, 26 (1): 101–122.
- Anthony, L. (2014) *Ant Conc (Version 3.4.3)* [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net> [Accessed on 12 March 2015].
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Butler, Y. G. (2002) Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*, 24: 451–480.
- Fraurud, K. (1990) Definiteness and the processing of NPs in natural discourse. *Journal of Semantics*, 7 (1): 395–433.
- Gries, S.T. and Mukherjee, J. (2010) Lexical gravity across varieties of English: An ICE-based study of n-grams in Asian Englishes. *International Journal of Corpus Linguistics* 15(4): 520–548.
- Gut, U. (2005) Nigerian English prosody. *English World-Wide*, 26 (2): 153–177.
- Hawkins, J. (1978) *Definiteness and Indefiniteness*. London: Croom Helm.
- Ionin, T. (2003). The Interpretation of 'the': A new look at articles in L2 English. *Proceedings of the 27th Annual Boston University Conference on Language Development*, 1 (pp. 346–357). Somerville MA: Cascadilla Press.
- Kortmann, B. (2006) Syntactic variation in English: A global perspective. In B. Aarts and A. McMahon (eds.) *Handbook of English Linguistics*. Malden, MA: Blackwell.
- Kortmann, B. and Szmrecsanyi, B. (2005) Global synopsis: Morphological and syntactic variation in English. In B. Kortmann, E. Schneider, K. Burridge, R. Meshtrier and C. Upton (eds.) *A Handbook of Varieties of English*, Vol. 2: *Morphology and Syntax*. Berlin: Mouton de Gruyter.
- Lamidi, M. T. (2007) The noun phrase structure in Nigerian English. *Studia Anglica Posnaniae*, 42 (2): 238–250.
- Mair, C. (2002) Creolisms in an emerging standard: written English in Jamaica. *English World-Wide*, 23 (1): 35–58.
- Platt, J., Weber, H. and Ho, M. L. (1984) *The New Englishes*. London: Routledge and Kegan Paul.
- Prince, E. (1981) Toward a taxonomy of given-new information. In P. Cole (ed.) *Radical Pragmatics*. New York: Academic Press.
- Prince, E. (1992) The ZP G letter: Subjects, definiteness, and information status. In W. C. Mann and S. A. Thompson (eds.) *Discourse description: Diverse Linguistic Analyses of a Fund-raising Text*. John Benjamins, Amsterdam/Philadelphia.

- Schilk, M. and Steffen S. (2016) Noun phrase complexity across varieties of English: Focus on syntactic function and text type. *English World Wide*, 37(1): 58–85.
- Tagliamonte, S. A., Poplack, S. and Eze, E. (1997) Pluralization patterns in Nigerian Pidgin English. *Journal of Pidgin and Creole Languages*, 12(1): 103–29.
- Wahid, R. (2013) Definite article usage across varieties of English. *World Englishes*, 32 (1): 23–41.
- Whitman, M. (ed.) (2000) *The Cupid Risk Book Series*. Available from <http://www.naijastories.com/CupidsRiskSeries.pdf> [Accessed on 12 September 2016].
- Zaenen, A., Carletta, J., Garretson, G., Bresnan, J., Koontz-Garboden, A., Nikitina, T., O'Connor, M. C., and Wasow, T. (2004) Animacy encoding in English: why and how. In *Proceedings of the Association for Computational Linguistics Workshop on Discourse Annotation* (pp. 118–125).

Mayowa Akinlotan's research interests include corpus linguistics, syntax, new varieties of English, variation linguistics, and quantitative sociolinguistics. His works have appeared in *English Today*, *Anglica*, *Token*, etc. Mayowa is currently with Vrije Universiteit Brussels (VUB), Belgium, and also currently a visiting scholar in the Linguistics Research Center at the University of Texas at Austin, USA. Email: mayowa.akinlotan@vub.be