

The relationship of Nigerian Pidgin English and Standard English in Nigeria: Evidence from copula constructions

Ogechi Agbo & Ingo Plag

Abstract

Deuber (2006) investigated variation in spoken Nigerian Pidgin data by educated speakers and found no evidence for a continuum of lects between Nigerian Pidgin and Nigerian English. Many speakers, however, speak both varieties, and both varieties are in close contact to each other, which keeps the question of the nature of their relationship on the agenda. This paper takes a new look at the relationship between Nigerian Pidgin and Nigerian English by choosing an approach opposite to Deuber's. We investigate 40 conversations in Standard Nigerian English by educated speakers as they occur in the International Corpus of English Nigeria (ICE-Nigeria), using the variability in copula usage as a test bed.

We first provide a variationist analysis in which we list and quantify the different copula constructions used by the speakers. We find a range of standard and non-standard constructions belonging to either of the two varieties. An implicational scaling analysis of the data reveals that there are strong implicational relationships between the different variants. Speakers vary, and they do so systematically along a cline from more typical to less typical Pidgin structures.

We argue, however, that the implicational patterning should not be interpreted as evidence for the existence of intermediate lects, but as the result of code mixing and style shifting. We analyzed our data in terms of Muysken's *Bilingual Speech* (2000) framework, which posits three kinds of code mixing: insertion, alternation and congruent lexicalization, with style shifting being an example of congruent lexicalization (p. 123). Based on the code mixing hypothesis, we can make two clear predictions. First, the spoken Nigerian English should exhibit the typical constellation of properties of congruent lexicalization. Second, all three types of code-mixing should be attested in our data, congruent lexicalization in particular.

Our analysis shows that both predictions are correct. A closer look at the individual conversations in our data set (personal relationships of the speakers, topic etc.) support the idea that with regard to English and Pidgin in Nigeria, we are dealing with language mixing for the purposes of style shifting. The speakers' competences in both languages help them to style-shift along an implicational scale, with topic, formality and social relationship as determinants of the code mix.

1 Introduction

In Nigeria, English co-exists with an English-based Pidgin, Nigerian Pidgin. Both languages serve as *lingua Franca* amidst over 400 other Nigerian languages spoken in Nigeria. The long existence of these two language varieties in Nigeria draws attention to the investigation of their mutual influence. These two languages have existed together for a long time in Nigeria and traditionally have been serving different purposes. English is the official language and serves prestigious functions as the language of the government, education, media etc in Nigeria. Nigerian Pidgin, on the other hand, is a contact Language that originated in Nigeria as a result of European contact with local languages, mostly through trade transactions. It is accorded a low prestige in Nigeria because it is not officially recognized and was mostly used by speakers who could not acquire formal education. Over the years, Nigerian Pidgin has gradually gained in importance, because the educated Nigerians also use it in communication. Research has shown that Nigerian Pidgin is now the language with the highest population of users, and also a first language in some minority groups in Nigeria (Igboanusi, 2008). The most striking change in status is its prominent use in tertiary institutions in Nigeria, which has also resulted in the modification of Nigerian Pidgin in different levels of language use (phonology, morphology, syntax etc). Given the status, use and long time co-existence of these languages, contact-induced mutual influence and change is to be expected (e.g. Thomason 2001, Sankoff 2001). This paper investigates the relationship between Nigerian English and Nigerian Pidgin in the speech of educated Nigerians.

A lot of research has been carried out on the co-existence of these two languages, focusing on the status of the two languages and speakers' attitudes towards them (e.g., Akande and Salami 2010; Balogun 2013 ; Osoba 2014; Ama kiri and Igani 2015; Oreoluwa 2015). Little empirical work has been done, however, on their mutual influence. Deuber (2006) approached this issue from the perspective of Nigerian Pidgin and found no evidence for the existence of a continuum of lects. The present study shifts the perspective and investigates Nigerian English as used by speakers who also speak Nigerian Pidgin. We address the following research questions:

1. Is there an influence of Nigerian Pidgin on the Nigerian English used by educated speakers?
2. If there is such an influence, how can this influence be analyzed? Is there a continuum? Or are we dealing with code-switching?
3. What are the possible explanations for the observed patterns of variation?

The data for this study is a set of conversations from International Corpus of English (ICE), Nigeria. ICE Nigeria is one of the world-wide varieties of English compiled for the analysis of linguistic structures. It represents both written and spoken genre by educated speakers, and we used the spoken casual conversations for our study. We concentrate on the use of copula constructions, as, across many non-standard varieties of English, these constructions have been shown in numerous studies to be highly variable (e.g. Ferguson 1971, Holm 1984, Winford 1990, McWhorter 1995, Rickford 1999, to mention only a few). Deuber (2006) also included the copula in her investigation, which will allow us to compare our results directly to hers.

This paper is organized as follows. The next section introduces the reader in more detail to the issues tackled in the present article. Section 3 describes the methodology and sections 4 and 5 present the empirical results. Section 6 concludes the paper.

2 Nigerian Pidgin and Nigerian English

2.1 The starting point: Deuber (2006)

The present work takes as its starting point the only available variationist study of the potential linguistic continuum in Nigeria, Deuber (2006). Two issues prompted Deuber's investigation of a linguistic continuum in Nigeria. The first is what she calls a "sweeping generalization" (2006:245) by Todd (1974), that Creole continuum can be found in every part of the world where an English based Creole co-exist with English. Another issue is Bickerton's (1975a) assertion that a linguistic continuum has emerged in Nigeria as far back as 1960, after Nigerian's independence. Bickerton attributed this to social mobility, which he believes is the vehicle for the development of a continuum. According to Bickerton, social mobility favours the development of a continuum as people with multilingual backgrounds move from one part of the country to another. Furthermore, due to emancipation, there was free movement of people across the country and there were opportunities for the uneducated to learn English. Because learning did not impart equally on every learner, intermediate varieties came into being, and speakers can be located along a continuum from the least standard variant (with influence from other local languages) to more standard variants.

Deuber (2006) pointed out, however, that there is no detailed empirical evidence to support these assumptions. Earlier, Agheyisi (1984) had contradicted Bickerton by saying that the relationship that exists between these language varieties in Nigeria cannot be described as a continuum in the sense of the Caribbean varieties.

To empirically investigate the possible existence of a continuum in Nigeria, Deuber investigated Nigerian Pidgin spoken by educated Nigerian speakers in Lagos, the southwestern part of Nigeria. She investigated variation in copula constructions, tense-aspect marking and verbal negation. With regard to copula constructions, Deuber interprets her data as evidence against the existence of a linguistic continuum. As a general result, she finds the two languages as two separate varieties with no evidence for intermediate varieties as found in the Anglophone Caribbean. Going back to Bickerton's assertion that social mobility enhances the development of a

continuum, Deuber wonders why after more than 40 years of Nigerian's independence, no such intermediate varieties have come into existence in Nigeria. One of her explanations is that the two situations differed with regard to the continuing presence of the substrates. The presence of the substrates in Nigeria may have been detrimental to the development of a continuum.

Deuber looked at the Pidgin-to-English continuum, focusing on the changes Nigerian Pidgin may have undergone as a result of the influence of English. The present work, however, looks at the relationship of Pidgin and English from a different angle. It investigates the (potential) English-to-Pidgin continuum, looking at changes Nigerian English has undergone as a result of the influence of Nigerian Pidgin. Like Deuber, we make use of corpus data for the analysis. In the analysis of the variation in the use of copula constructions we employ implicational scaling as a means to establish a linguistic continuum. Let us now turn to the linguistic phenomenon that we investigate, the copula.

2.2 The copula in English and Nigerian Pidgin

The English lexical verb BE is known as a linking verb, or 'copula'. There are other verbs in English like *seem*, *appear*, *look*, *sound*, *smell*, *taste*, *feel*, *become* etc which, apart from their lexical functions, also have linking function. We restricted the scope of this work to the English copula BE and its functional equivalents in Pidgin (see Faraclas 1996: 50 on other linking verbs in Nigerian Pidgin).

In English, copula constructions are overtly marked by a form of BE and they follow the same structural rules of English word order SVO except in cases of inversion (e.g. *Are you sure?*, *Here is an example*). The complement following the copula may be a noun phrase, an adjective phrase, or a prepositional phrase, as shown in (1).¹

- (1) a. She is a girl (*noun phrase*)
b. They are beautiful (*adjective phrase*)
c. The man is here (*prepositional phrase*)

¹ We follow the assumption that many adverbs are best analyzed as intransitive prepositions (Emonds 1972, Jackendoff 1973, Pullum & Huddleston 2002: 598-603, 612-617).

- d. The house is on a hill (prepositional phrase)

We also included constructions in which BE functions as an auxiliary (as in *He is cooking food*).

Nigerian Pidgin has a copula system different from that of English. Faraclas (1996:46) writes that the “space normally covered by copulas is divided roughly into two parts, each of which is coded by one of two basic copula verbs: the copula identity verbs *bi* and the copula locative/ existence verb *de*”. The third copula verb in Nigeria Pidgin is *na*, which also functions as a focal marker. Just like English, Nigerian Pidgin also has some other verbs that may have copulative functions, but we restrict ourselves to the three main Nigerian Pidgin copulas discussed in Faraclas’ grammar (1996): *bi*, *de*, *dè* (with a low tone marked by the grave accent) and *na*. The first three are also written as *be*, *dey* and *dèy*, respectively. The copula *bi* and the copula *na* are used as an equative copula as in (2), taken from Faraclas (1996:51).

- (2) a. Im bi man
he COP man
,He is a man‘
- b. Di wuman na sista
the woman COP Sister
'The woman is a Sister'.

One important distinction between *na* and the other copulas in Nigerian Pidgin is that *na* is always followed by a nominal element and can never take auxiliaries, negators or non-emphatic pronouns (Faraclas 1996:50). The other two forms can take noun phrase and adjective phrases as complements, and can also take auxiliaries and negators.

The copula *na* is also used as an emphatic marker in sentence-initial position, as in (3).

- (3) Na sista we di wuman bi
It is sister that the woman COP
'The woman is a Sister'.

Adjectives do not serve as complements to Pidgin copulas because there are no predicative adjectives in Pidgin. What is regarded as predicative adjectives in English can be analyzed as stative verbs in Nigerian Pidgin. Consider (4), taken from Faraclas (1996:221):

- (4) Di man fyar
the man fear
'The man is afraid'

According to Faraclas (p. 45), that there is no zero copula in Nigerian Pidgin. But we do find variation in the use of stative verbs, such that stative verbs may also be accompanied by the copula *de*. Observe the contrast between (5a) and (5b).

- (5) a. Ma pikin de smol (from Faraclas 1996:48)
my children COP small
'My children are small'
- b. Di sup swit (from Faraclas 1996:222)
the soup sweet
'The soup is sweet'

We will remain agnostic as to the kind of syntactic analysis one would want to assign to examples such as (4) and (5b), but we will refer descriptively to these constructions as 'zero copula' or 'zero' constructions.

The copula *dè* with a low tone occurs in preverbal position, where it marks imperfective aspect.² Faraclas (p. 186) labels this form as ‘auxiliary. (6) illustrates its usage.

- (6) A de kari nyam
I –comp. carry yam
'I am carrying yam '

To summarize, there are important differences, but also similarities, in the use of the copula between Pidgin and English, which opens up a space for variation across varieties. We will see in section 4 that this expectation is borne out by the facts: speakers vary between English and Pidgin constructions within the same conversation.

3. Methodology

3.1 Data

The data used for this study come from the *International Corpus of English, Nigeria* (ICE-Nigeria, <https://sourceforge.net/projects/ice-nigeria/>). The *International Corpus of English* is a collection of corpora of world-wide varieties of English that were compiled for the analysis of linguistic structures. The ICE varieties represent speech of educated speakers (Greenbaum 1996:6) and the corpora have been used in many studies (see, e.g., Deuber 2009, 2010; Bolton et al 2002; Oenbring 2010; Gut and Fuch 2013).

Like the corpora of other varieties, ICE-Nigeria represents both the written text category (academic writing, business letters, administrative writings etc) and the spoken category (conversations, broadcast news, parliamentary debate etc), with a total number of 1,010,382 words. The conversation part of ICE Nigeria represents spontaneous speech of speakers from different geopolitical zones in Nigeria. We made

² Faraclas (1996:202f) uses the terms ‘imperfect’ and ‘incompletive’ without further discussion. Given that the exact aspectual function of *dè* is not quite clear, we use ‘imperfective’ as the established cover term for all non-perfective aspects.

use of 40 conversations by 87 speakers. The nature of the conversations differs, depending on the interlocutors. There are group discussions among workers, friends, family, and university classmates featuring three to six speakers. Some conversations had two speakers, e.g. between husband and wife, or two friends. The topics of discussion are familiar ones that have to do with marriage, studies, food, work etc.

3.2 Sampling

We first extracted utterances with copula constructions or omissions. The extraction was done both systematically and manually as to include constructions with copula omission. We only sampled declarative sentences to avoid potential complications arising from *wh*-questions and inversions. The sampling of standard inflected forms of *BE*, (*is, am, are, was, were, be, being, been*) was restricted to 3rd person forms and infinitival *be*. Including other standard forms would only have increased the proportion of standard forms in the data set without any further insights into the variation between standard and non-standard forms. The resulting data set consisted of 1036 tokens of copula constructions with eight variant forms from both languages.

3.3 Coding

We entered the copula constructions into a spreadsheet according to their form and classified them according to the type of construction. We encountered one variant that does not seem to belong to either language. This variant involved the use of the copula form *is* without a subject. The values of the two variables *FORM* and *CONSTRUCTION* are as follows: English main verb *is*/auxiliary *is* ('inflected'), English infinitival *be* ('inflected'), and Pidgin invariant *be* ('invariantBE'), *'s* ('contracted'), *_is* ('no subject'), \emptyset ('zero'), equative *na* ('cop na'), focus-marker *na* ('foc na') and *de/dè* ('dey'). For both English and Pidgin we conflated the auxiliary usage and the corresponding main verb into one category. We also coded for other variables: *FILE* (the conversation number), *SPEAKER* (the identifying number of the speaker as it appears in the conversation), *GENDER* ('male' or 'female'), *AGE*, and *VARIETY* ('standard' or 'non-standard'). The subjectless form was coded as 'non-standard'.

4. Results

4.1 Variation

Let us first look at the distribution of standard and non-standard constructions in our data. This is given in figure 1 (the numbers above the bars give the number of observations for this category). We can see that about one third of the constructions are non-English constructions, which shows that Pidgin is part of the repertoire of the speakers in ICE-Nigeria.

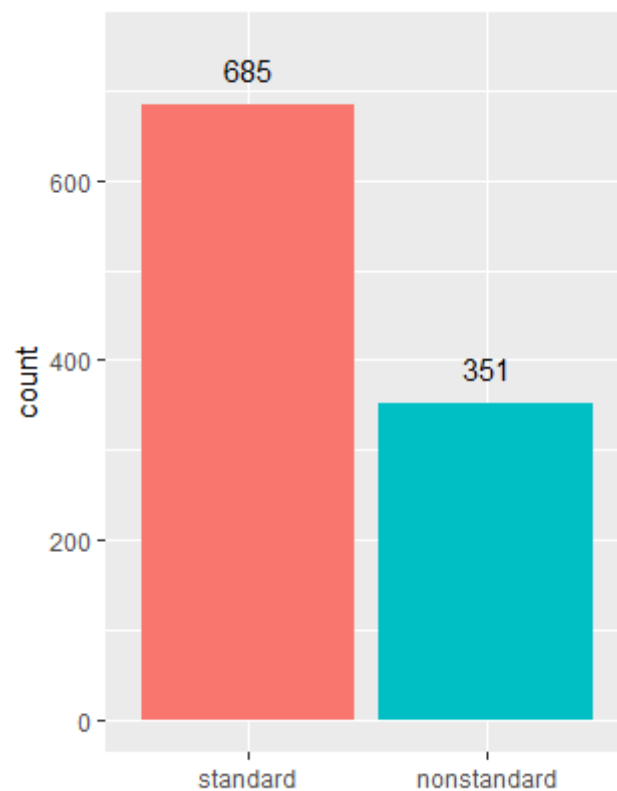


Figure 2: Distribution of the standard and non-standard forms

Figure 2 gives the distribution of the different constructions.

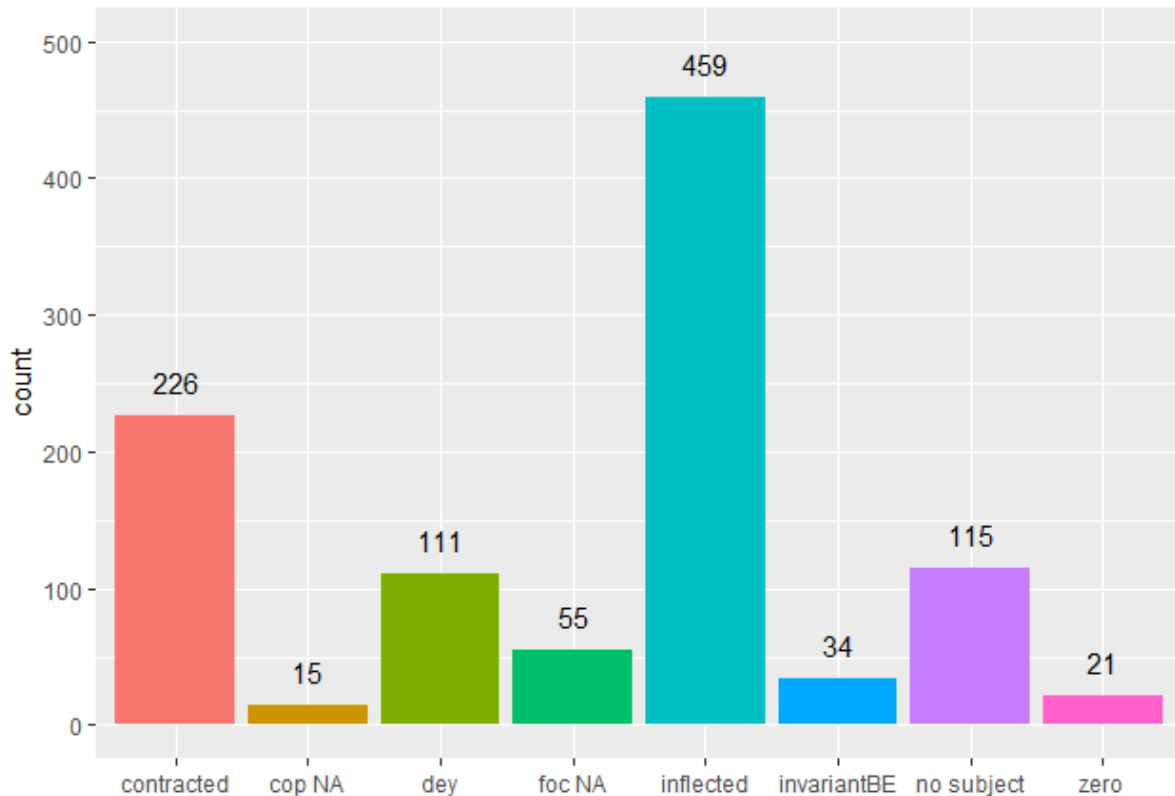


Figure 2: Distribution of copula constructions N = 1031

The inflected and contracted BE forms are predominant with 459 and 225 counts, respectively. Nigerian Pidgin copula forms are also used in non-negligible proportions, with *dey* and no-subject constructions being most frequent.

In order to get a better understanding of this variation we implemented an implicational scaling analysis. This kind of analysis has been standardly used (at least since Decamp (1971)) to test whether linguistic variation is structured in such a way that the presence of one feature implies the presence of other features. This model can be illustrated as a table where speakers are ranked using plus (+) or (-) sign according to the features they use. The “+” sign represents the presence of a standard form while the “-” sign represents the absence of a standard form. The features are then arranged in such a way that the most acrolectal speaker is placed on top while the most basilectal speaker is placed at the bottom. A continuum of lects is formed when all the plus (+) signs are in the same direction and all the minus (-) signs are in the same direction. This model has been used extensively in the Caribbean (see,

e.g., Bailey 1966; DeCamp 1971; Winford 1988; Patrick 1998a; Deuber 2009). Table 1 is an example of an ordered implicational scaling based on features from Jamaican speakers in Decamp’s work (1971, p355).

Table 1: Example of implicational scaling based on Decamp (1971:355). Linguistic features A-F are given in the top row, speakers 1-7 in the leftmost column

Speaker	B	E	F	A	C	D
5	+	+	+	+	+	+
1	+	+	+	+	+	-
6	+	+	+	+	-	-
2	+	+	+	-	-	-
7	+	+	-	-	-	-
3	+	-	-	-	-	-
4	-	-	-	-	-	-

Many speakers in our data set provided only a few copula sentences, which is detrimental for implicational scaling because it leads to empty cells in the table. We therefore decided to select only those speakers for which there are more than 20 utterances. This gave us eleven speakers for which the distribution of forms is given in table 2.

Table 2: Distribution of constructions by speaker (only speakers with more than 20 attestations)

Speaker	Inflected	Contracted	No Subject	Zero	Focal <i>na</i>	Invariant BE	Copula <i>na</i>	<i>dey</i>
02-1	31	14	2	0	0	0	0	0
02-2	15	13	4	0	0	0	0	0
03-3	11	6	4	0	0	0	0	0
04-2	4	0	1	2	32	21	8	72
04-2	6	3	0	1	21	5	7	31
05-1	23	9	9	1	0	0	0	0
06-1	27	15	4	1	0	0	0	0
07-2	12	9	4	0	0	0	0	0
09-1	17	1	9	2	0	0	0	0
12-1	9	7	11	0	0	0	0	0
15-2	29	0	1	0	0	0	0	0

Implicational scaling was possible, with only two of the 88 cells (2.2 percent) going against full scalability. The scaled table is given in table 3, in which we abstract away from the number of attestations by using a plus sign if a form is attested and a minus sign if a form is not attested.

Table 3: Implicational scaling

Speaker	Inflected	Contracted	No Subject	<	Zero	<	Focal <i>na</i>	Invariant BE	<	Copula <i>na</i>	<i>dey</i>
04-1	+	-	+		+		+	+		+	+
04-2	+	+	-		+		+	+		+	+
09-1	+	+	+		+		+	+		-	-
05-1	+	+	+		+		-	-		-	-
06-1	+	+	+		+		-	-		-	-
02-1	+	+	+		-		-	-		-	-
02-2	+	+	+		-		-	-		-	-
03-3	+	+	+		-		-	-		-	-
07-2	+	+	+		-		-	-		-	-
12-1	+	+	+		-		-	-		-	-
15-2	+	-	+		-		-	-		-	-

The implicational scale reveals four different patterns of usage, separated by grey columns. The further to the right we go in the table, the more non-standard forms are being used by the speakers. Note that in the leftmost pattern we could swap the columns labelled 'Contracted' and 'No Subject' since there is no clear evidence for an implication. This kind of situation reveals a general disadvantage in using implicational scaling. The non-presence of a particular construction in the speech of a given speaker could simply be a coincidence, especially if the number of attestations is rather low. However, the scalability measure of 97.8 percent shows that the variation is quite systematic and predictable.

How can we interpret the patterning of variation in table 3? Much in the spirit of analyses of creole continua one could say that the standard forms are on the left, and the more we go to the right, the more basilectal the forms become. Inflected forms would be the most acrolectal while Pidgin *dey* and copula *na* would be the most basilectal forms. We see that almost all the speakers used the English forms, which was expected as they represent educated speakers. Other speakers also move across the continuum and use English and Pidgin forms. Interestingly, the subjectless construction, which is neither a Pidgin construction, nor a standard construction is the only non-standard form used by the six speakers that do not use any Pidgin form. It seems to have an intermediate status between Pidgin and English.

If we interpret table 3 as a continuum, speaker 04-2 in conversation 04 at the top of the scale represents the acrolectal speaker while speaker 15-2 the basilectal speaker. Other speakers between them would be considered mesolectal speakers. What is also striking is the fact that the two speakers in conversation 04 used almost the same features and those in conversation 02 also used the same feature.

The obvious question now is whether we can interpret the scale shown in table 3 as a continuum of lects. Such an interpretation is perhaps not warranted since we have only one truly intermediate construction and otherwise constructions that are either clearly English or clearly Pidgin. This means that we are rather in a diglossic situation in which speakers seem to be able to mix the two varieties as need be. The next section will investigate this possibility in detail.

5. Structured variation: Code-mixing of English and Pidgin

If we interpret the scale in table 3 in the same way as comparable scales in the Caribbean have been interpreted, we should find different speakers identified as belonging to a particular lect. These lects will range from the near standard English variety at one end to Nigerian Pidgin at the other end. In between the two lects will be other intermediate lects. The speakers at the two ends, representing the acrolectal and basilectal speakers will not be able to understand each other because their lects are wide apart. They will rather understand only those intermediate speakers closer to their pole. In that case, no speaker has full command of both languages. Each individual has either full command of one language and a little knowledge of the other or no command of either language and a little knowledge of both. That is why they are placed on a continuum of different lects or codes.

This is not the case with our set of data. What we see in our data is that these speakers have full knowledge of the two languages in their repertoire, and are able to move from one pole to the other. We claim, however, that this continuum is not a continuum of lects, but a continuum of mixtures of two codes. This leads us to interpret our data as a case of code-mixing.³

We find different degrees of the mixing at different points of the continuum: some speakers with more mixing and some with less mixing. A closer investigation of each conversation shows that the code-mixing is as a result of style-shifting. The speakers' competence in both languages help them to style-shift along this continuum, with topic, setting, formality and social relationship as determinants of the code-mix. We have indirect evidence of this in the implicational scaling table with those conversations that have up to two speakers. The two speakers in conversations 04 used almost the same features and those in conversation 02 also used the same features. We interpret this as the speakers' attempt to style-shift to meet each others' footing depending on the topic, setting and social relationship. We therefore

³ In accordance with the literature (e.g. Bokamba 1988:24, Muysken 2000:1), we use 'code-switching' as the cover-term for the use of two languages in one conversation. 'Code-mixing' refers to switches between languages that occur intrasententially.

reinterpret our implicational scaling as reflecting a continuum of style, with style as a constellation of features that is determined by the extra-linguistic variables mentioned above.

Based on these considerations, we set up the hypothesis that we are dealing with code-mixing as a style-shifting device. To test this hypothesis empirically, we employ Muysken's bilingual speech framework (2000), which has been used fruitfully in many studies of code-mixing (e.g., Deuchar 2005, Deuchar, Muysken & Wang 2008, Lipski 2014). Muysken sets up a taxonomy of code-mixing to characterize different patterns of code-switching found in bilingual speech communities. In this framework style-shifting is considered a subtype of code-mixing called 'congruent lexicalization' (2000:123, see below). If we are dealing with code-mixing, we can derive two predictions from Muysken's work:

- We should find different patterns of switches that are common in uncontroversial cases of code-mixing.
- More importantly, we should find many switches that are instances of congruent lexicalization.

In what follows we will present the results of an analysis that tests these predictions

5.1 Code-Mixing

Muysken identifies three patterns of mixing that can be found in any bilingual setting:

Insertion: Material from one language is inserted into the structure of another language. This type of pattern follows the Matrix Language Framework theory proposed by Myer-Scotton (1993 et seq.) since the notion of 'insertion' presupposes a matrix language into which lexical material from the embedded language is inserted. This pattern has an a-b-a structure, where 'a' represents words from the matrix language A and 'b' represents words from the embedded language B. Example (7),

with Spanish-English code-switching as taken from Muysken (2000:5), illustrates the a-b-a structure. The words in italics represent the embedded language.

- (7) Yo anduve | *in a state of shock* | por dos dias
 a b a
I walked in a state of shock for two days
'I walked in a state of shock for two days' (Spanish/English)

The sentential structure, i.e. the order and categories of the major constituents, is that of the matrix language.

Alternation: The structure of a sentence is split between two languages. It has a structure that can be schematized as 'a-b'. This pattern of code-mixing is similar to the switching of codes between turns or utterances, in that each segment contains complete syntactic constituents, i.e. phrases, that fully reflect the structure of that language. See (8), from Muysken (2000:5).

- (8) Andale pues | *and do come again*
 a b
go-it well and do come again
'That's alright then, and do come again'

Congruent lexicalization: Material from two languages is inserted into a shared grammatical structure. Hence congruent lexicalization necessitates that the languages involved are structurally highly congruent. Similarities in the lexical material between the two languages further facilitates congruent lexicalization. The insertion of material may occur "more or less randomly" (Muysken 2000:8), as shown in (9), from Muysken (2000:5).

- (9) Bueno, | *in other words*, | el | *flight* | que sale de Chicago | *around three o'clock*
- a b a b a b
- Good in other words the flight that leaves from Chicago around three o'clock
- 'Good, in other words, the flight that leaves from Chicago around three o'clock'

To test for the occurrence of the different code-mixing pattern in our data set, we coded our data for the three types of code-mixing, using some of the diagnostic features proposed by Muysken. As shown in Deuchar, Muysken & Wang (2008), using this methodology, it is possible to establish a precise quantitative profile of the codeswitching properties in a given corpus.

5.2 Diagnostic Features

Muysken puts forward some diagnostic features that can be applied to bilingual speech data to determine which patterns of code-mixing (insertion, alternation or congruent lexicalization) occur, with every conversation involving code-switching potentially exhibiting a combination of the different types. We selected those diagnostic features that are informative for our task. They can be grouped under the headings 'constituency', 'elements switched' and 'properties'.

Constituency: Features under this heading code the kinds of constituents that participate in intrasentential switches. The following features are grouped under this heading:

- Single constituent: Any syntactic unit, for example a single lexical item, or a phrase (NP, VP, PP etc.)
- Several constituents: More than one phrase
- Non-constituent: Segments that are not full phrases in either language. This occurs when switches are located inside phrases.

- Nested a-b-a: Material from one language that forms one larger constituent with the surrounding material from the other language, as in $[a\ b\ a]_{XP}$
- Non-nested a-b-a: Material from one language (i.e., language b) in which the surrounding material from the other language (i.e., from language a) belongs to different constituents, as in, for example $[a_i]_{XP}\ b\ [a_j]_{YP}$, or $[a_i\ b]_{XP}\ [a_j]_{YP}$, or $[a_i]_{XP}\ [b\ a_j]_{YP}$. In these configurations, a_i and a_j belong to different constituents.

Elements switched: Under this heading we selected two features that are pertinent for the kinds of constructions we are investigating.

- Content word: Verb
- Function word: Auxiliary

Properties: This category refers to particular characteristics of the switched materials. Here we selected only linear equivalence because the other feature (so-called ‘triggering’) which would have been potentially useful, is rather ill-defined so that it cannot be reliably operationalized.⁴

- Linear equivalence: The word order in the two languages is the same.

These features are diagnostic of the three different code-mixing patterns in the way shown in table 4. A ‘+’ value for a feature indicates that the presence of the feature is indicative of a particular pattern. A ‘-’ value, in contrast, indicates that the presence of this feature is a clear indication that the switch is not of this type. A ‘0’ indicates that the feature is not diagnostic for the pattern in question, or, in our case, may show the absence of the feature from our data set. For illustration consider the feature ‘Non-constituent’. If it is present this is a clear indication of congruent lexicalization, if it is absent this indicates that we are not looking at a case of insertion or alternation. The presence of ‘Single constituent’ indicates that we are dealing with insertion, but it does not give evidence for or against the other types of mixing.

⁴ For example, based on Clyne 1967 and Broersma 2009, Lipski (2005:35) defines ‘triggering’ as “multi-word switches in which the choice of one of the words in the switch ... may lead to the switching of a longer string.” It is unclear how such a vague definition (“may lead to”) can be applied consistently.

Table 4. Diagnostic features based on Muysken (2000:230, table 8.1)

Selected Features	Insertion	Alternation	Congruent Lexicalization
Consistency			
Single constituent	+	0	0
Several constituents	-	+	0
Non-constituent	-	-	+
Nested a b a	+	-	0
Non nested a b a	-	+	+
Elements switched			
Content words	+	-	-
Function words	-	-	+
Properties			
Linear equivalence	0	+	+

5.3 Analysis: code mixing

We first selected all sentences featuring a Pidgin copula and a switch from English to Pidgin. Subjectless clauses and clauses with a zero copula were not included because covert elements cannot be assigned unambiguously. We then coded each segment involving a Pidgin copula according to the categories shown in table 4. Sentences such as those in (10) through (12), which are completely in Pidgin, were not considered.

(10) na wa o (con 9)

(11) Na wetin me I be dey think o .(con 04)

(12) Nebu hole abi wetin them dey call am (con 04)

If a sentence contained more than one Pidgin copula, each switch was analyzed separately. An example is given in (13), Pidgin segments are given in italics, the copulas are in bold:

- (13) a. **Na** some Africans **dey** control the place (con 4)
 b. **Na** one cave but the thing **dey** interesting anyway (con 4)

We also have some constructions in our data with two Nigerian Pidgin copulas in a single segment. In such cases, we analyse them as one switch. (14) illustrates this case (Pidgin segments are given in italics, the copulas are in bold):

- (14) a. The funny thing | **be** say *na* | for German so (con 04)
 b. Maybe | *na him* **be** say | I go move (con 04)

Due to the similarities in the lexicon of Nigerian English and Nigerian Pidgin, and the similarities in the pronunciation, it is often difficult to decide whether a given string is English or Pidgin. In cases of doubt, the word was taken to be English, unless it is used in a way that is alien to Nigerian English. Overall, the application of our criteria yielded 165 constructions which were subject to further coding.

Recall that under the code-mixing hypothesis we expected to find different types of code-mixing patterns (insertion, alternation and congruent lexicalization), and many instances of congruent lexicalization in particular.

Based on our coding of the switches, and following the procedure implemented, for example, by Deuchar (2005) and Lipski (2014), we assigned the value '1', if the switch corresponds to the expected value of the feature and '-1' if the switch negates the expected value of the feature. '0' was assigned if the expected value of the feature is neutral, or the feature does not occur in the switch. We then sum the scores for each pattern. For illustration, Table 5 gives a sample analysis based on the first three sentences in our data base. They are given in (15).

- (15) a. *Na them* **dey** | drive the biggest car for Nigeria
 b. *them* **dey** | send email
 c. I'm sure | **say na** | the guy spoil that room

The coding of the features is given in Table 5 in the second column from the left. Each ‘1’ that is listed in the remaining columns represents positive evidence for a particular pattern, each ‘-1’ evidence against a particular pattern. The sums of the scores are given in the bottom line of the table. As shown by Deuchar et al. (2008), the scores for each type of mixing can be added together in order to achieve an overall score for each type of mixing for the whole data set. The scores can be conceptualized as an index of the amount of evidence for a particular pattern in a given corpus.

Table 5 illustrates the approach with three sample constructions that were coded for the diagnostic features. In all three constructions, congruent lexicalization is the dominant pattern, i.e. the pattern for which we have the best evidence as given by our diagnostic features.

Table 5: Sample Analysis of code-mixing for three constructions. ‘I’ = Insertion, ‘A’ = alternation, ‘CL’ = congruent lexicalization

Diagnostic Features	Constructions			Score Construction 1			Score Construction 2			Score Construction 3		
	1	2	3	I	A	CL	I	A	CL	I	A	CL
Single Constituent	-	-	-	-1	0	0	-1	0	0	1	0	0
Several Constituents	-	-	-	1	-1	0	1	-1	0	1	-1	0
Non Constituent	+	+	+	-1	-1	1	-1	-1	1	-1	-1	1
Nested a-b-a	0	0	+	0	0	0	0	0	0	1	-1	0
Non Nested a-b-a	0	0	-	0	0	0	0	0	0	-1	1	1
Content Word	-	-	-	-1	1	1	-1	1	1	-1	1	1
Function Word	+	+	+	-1	-1	1	-1	-1	1	-1	-1	1
Linear Equivalence	-	+	+	0	-1	-1	0	1	1	0	1	1
Total				-3	-3	2	-3	-1	4	1	-1	5

Table 6 gives the overall results of the coding and counting of all 165 constructions.

Table 6: Counts of diagnostic features indicating different types of code-mixing in the data set

Diagnostic feature	Insertion			Alternation			Congruent lexicalization		
	1	-1	0	1	-1	0	1	-1	0
Single Constituent	73	92	-	-	-	165	-	-	165
Several Constituent	144	21	-	21	144	-	-	-	165
Non-constituent	87	76	2	88	75	2	62	80	23
Nested a-b-a	27	76	62	76	26	63	-	-	165
Non-nested a-b-a	25	76	64	72	29	64	63	25	77
Content Word	44	121	-	106	39	20	106	39	20
Function word	47	118	-	53	112	-	118	47	-
Linear Equivalence	-	11	154	111	54	-	112	43	-
Total (N = 1320)	447	-591	282	527	-479	314	471	-234	615

From the bottom row of Table 6 we can compute the total score that can then be used to find the dominant pattern in the data set. This total score is derived by adding the figures representing the evidence for and against a particular pattern ('0' values are ignored, as they do not speak for or against the given type of mixing). For instance, adding up the evidence for and against insertion across all codings gives us a total score of -144 (i.e. 447-591). If we do the same for the other two categories, as shown in table 7, we find that congruent lexicalization is clearly the dominant pattern.

Table 7: Total score for the three types of mixing

type of mixing	score index
Insertion	-144
Alternation	48
Congruent lexicalization	237

To summarize, the empirical analysis of the sentences with Pidgin copulas has shown that all types of switches occur, and that congruent lexicalization is the predominant type of mixing in this data set. This result is in line with the predictions that emerge from the hypothesis that our data present a case of code-mixing, and not a continuum

of lects. Furthermore, given the predominance of congruent lexicalization we have evidence that we are dealing with code-mixing as a style-shifting device.

The style-shifting hypothesis also predicts that social factors should play a role in the code mixing found in our data set. Such social factors may be the settings of the conversations, the topics discussed, or the relationships of the discussants. Due to the lack of pertinent information, the conversations from ICE-Nigeria do not allow for a systematic quantitative investigation of the possible influences of social factors, but there are some indications to that effect.

In our data, speakers use Pidgin forms mostly when talking to their fellow university friends and classmates, and they use them when talking about things that happen around them. Consider (17).

- (17) a. This guy him room na self con (Con 04)
b. You know him dey do him PhD. (Con 04)

Interestingly, when the students in ICE-Nigeria conversations talk to their lecturers, they consistently use only the standard forms, as illustrated in (18).

- (18) a. That's a professional course (Con 38)
b. Very soon I'll be through with my programme (Con 33)
c. I'd love to be a research fellow (Con 32)
d. He is the one going to supervise my project (Con 40)

We see here that the style-shift is a function of interlocutor, topic and relationship. The fact that the use of the non-standard features follows an implicational hierarchy does not speak against the idea of style-shifting. Rather, one could hypothesize that the use of particular forms is socially constrained, with the forms becoming less and less formal from one end of the scale to the other (see again Table 3).

6. Conclusion

In this paper we investigated the relationship between the standard Nigerian English and Nigerian Pidgin English. More specifically we looked at the potential mutual influence of the two languages in the speech of educated Nigerians. We used the English copula BE and its functional equivalents in Pidgin as our test case, analyzing conversations as recorded in ICE-Nigeria.

The data revealed a perhaps unexpected amount of variation among the different copula forms. Apart from the use of the standard copula variants, we also find different Pidgin copulas in the data and a subjectless construction that is neither found in English, nor in Pidgin. The variation lends itself to implicational scaling, showing a clear implicational pattern of usage.

We have argued that, unlike in some Caribbean varieties of English, the implicational pattern of variation should not be interpreted as a continuum of individual lects but as structured shifting. This view is supported by a number of considerations and findings:

- The speakers in our sample are fully competent in both Pidgin and English.
- The language situation in Nigeria is characterized by diglossia, with the two languages being traditionally used in different spheres.
- The predominant pattern of language mixing is congruent lexicalization, which in turn indicates style-shifting (Musken 2000)
- The qualitative evidence from the conversation suggest a strong relationship between speech situation (interlocutors, topic, personal relationship) and the variation in use of copula variants.

Comparing our results to those of Deuber (2006), we can say that our findings corroborate her conclusion that the type of continuum that is typical of the Caribbean does not exist in Nigeria. However, in contrast to Deuber, we found a significant amount of variation in the use of copula constructions in the speech of educated Nigerians, and this variation is structured implicationally. We argued that this

variation is best understood as language-mixing for the purposes of style-shifting. The pattern of mixing represents that of competent bilinguals with fluent knowledge of the structures of both languages. This helps them to style-shift along a stylistic continuum, with topic, settings, formality and social relationship as determinants.

References

- Agheyisi, R.N. (1977). *Language interlarding in the speech of Nigerians*. In P.E Kotey & Der-Houssikian (Eds), *Language and Linguistic Problems in Africa*, 97-110. Columbia: Hornbeam.
- Akande, A.T and Salami L. O. (2000). *Use and attitudes towards Nigerian Pidgin English among Nigerian Students*. In R.M. Millar(eds) *Marginal Dialects: Scotland, Ireland and beyond Aberdeen: Forum for Research on languages of Scotland and Ireland*. 70-89.
- Amakiri and Igani, (2015). *The effects of Pidgin on the teaching and learning of standard English in Nigeria*. *International Journal of English and Education*,4. 122-130
- Bailey; B. (1971) *Jamaican creole: Can dialect boundries be defined?* Hymes (ed) 341-48.
- Balogun, Temitope Abiodun (2013). *In defence of Nigerian pidgin*. *Journal of Languages and Culture* 4(5). 90-98.
- Bickerton, D. (1975a). *Can English and Pidgin be kept apart?* McElhanon, K.A (Ed.) *Tok Pisin I Go We?*. Port Moresby: Linguistic Society of Papua New Guinea 21-27.
- Bickerton, Derek, (1973). *The Nature of a Creole Continuum*. *Language* 49, 640-649.
- Bokamba, Eyamba G. 1988. Code-mixing, language variation, and linguistic theory. *Lingua* 76(1), 21–62.
- Broersma, M. (2009) Triggered codeswitching between cognate languages. *Bilingualism: Language and Cognition* 12(4): 447–462.
- Clyne, M. (1967) *Transference and triggering*. The Hague: Nijhoff.
- Crystal, David. (1998). *A Dictionary of Linguistic and Phonetics*. Oxford: Blackwell.
- Curnow ,Timothy J. (1999). *Towards a Cross-linguistic Typology of Copula Constructions*. Proceedings of the Conference of the Australian Linguistic Society 1. RCLT: Australian National University

- Decamp, David, 1971. *Towards a Generative Analysis of a Post-Creole Speech continuum*. Hymes (ed) 349-370.
- Deuber, Dagmar, (2006). *Aspect of Variation in Educated Nigerian Pidgin: Verbal Structures*. In Deumert Ana, Durrleman Stéphanie (eds.), *Structure and Variation in Language Contact*, 243-261. Amsterdam: Benjamins.
- Deuber, Dagmar, (2009). *The English We Speaking* In *Journal of Pidgin and Creole Languages* 24(1) 1-52 .Amsterdam: John Benjamins.
- Deuchar, Margaret (2005). *Minority Language survival: code-mixing in Welsh*. In Cohen et al (eds), *Proceedings of 4th International Symposium on Bilingualism*, 608-620. Somerville: Cascadilla Press.
- Deuchar, Margaret, Pieter Muysken & Sung-Lan Wang. 2007. *Structured Variation in Codeswitching: Towards an Empirically Based Typology of Bilingual Speech Patterns*. *International Journal of Bilingual Education and Bilingualism* 10(3). 298–340.
- Faraclas , Nicholas G. (1996) .*Nigerian Pidgin*. USA: Routledge.
- Greenbaum Sidney (1996). *Comparing English World Wide*. Oxford: Oxford University Press.
- Gut, Ulrike (2012). *Towards a codification of Nigerian English-The ICE Nigerian project*. *Journal of the Nigerian English Studies Association*. 15(1).1-13.
- Gut Ulrike and Fachs Robert (2013). *Progressive Aspect in Nigerian English*. *Journal of English Linguistics* 41(3). 243-267.
- Hasty, Daniel J. (2014). *We might should be thinking this way: Theoretical and methodological concerns in studies of syntactic variation*, in Zanuttini and Horn (eds.) *Micro-syntactic variation in North American English*.
- Honeybone, P. (2011). *Variation and linguistic theory*. In Maguire, Warren and McMahon (eds) *Analysing variation in English*, 151-177. Cambridge: Cambridge University press.
- Igboanusi, Herbert (2008). *Empowering Nigerian Pidgin: a challenge for status planning?* *World Englishes* 27(1). 68-82.
- Labov, Williams (1972): *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania press.

- Lipski, John. M. (2014). *Spanish – English code-switching among low-fluency bilinguals: towards an expanded typology*. Sociolinguistic Studies 8(1).
- Marcel den Dikken and Teresa O'Neill (2017). *Copula constructions in syntax*. Oxford Research Encyclopedia of Linguistics. Mark Aronoff (ed). Oxford University Press.
- Meyerhoff, Miriam (2013). *Syntactic variation and language change: the variationist framework and language change*. In Isabelle Leglise and Claudine Charmoreau (eds) *The interplay of variation and change in contact settings*, 23-51.
- Muysken, Pieter (2000) *Bilingual Speech: a typology of code-mixing*. Cambridge: Cambridge University press.
- Oenbring, R. A. (2010). *Corpus Linguistic Studies of Standard Bahamian English: A Comparative Study of Newspaper Usage*. International Journal of Bahamian 16. 51-62.
- Ojarikre, Anthony,(2015). *A Contrastive Study of the Segmental Phonology of Nigerian Pidgin English and Standard English*. Journal of Literature, Languages and linguistics. 12. 8-16.
- Oreoluwa Abraham Solomon (2015). *Problematic Co-existence of English and Nigerian Pidgin*. European Journal of English Language and Literature Studies, 3(4). 38-44.
- Osoba, Joseph B. (2014). *The use of Nigerian Pidgin in Media Advert*. International Journal of English Linguistics 4(2). 26-37.
- Patrick, Peter L. (1998). *Caribbean creoles and the speech community*. Presented in Society for Caribbean Linguistics XII. St. Lucia.
- R Development Core Team (2014). *R: A Language and environment for statistical computing*. Vienna: R foundation for statistical computing.<https://www.r-project.org/>(accessed 2017).
- Reppen et al (2002) (eds) *Using Corpora to Explore Linguistic Variation*. John Benjamins.
- Rickford, John R.(1998). *The Creole Origins of African American Vernacular English: Evidence from Copula absence*. In Salikoko S. Mufwene et al (eds) *African-American English: Structure, History and Use*, 154-200. New York: Routledge.

- Sankoff, Gullian (2001). *Linguistic outcomes of language contact*. In Chambers et al (eds) the handbook of language variation and change, 638-668. Oxford, UK: Blackwell
- Sebba , Mark (1977). *Contact Languages: Pidgins and Creoles*. London: Macmillian Press.
- Thomason, Sarah G. (2001) *Language Contact*. Edinburgh University Press: Edinburgh.
- Todd, L. (1974). *Pidgins and Creoles* . London: Routledge
- Wardhaugh, Roland and Fuller Janeth M.(2015).*An Introduction to Sociolinguistics* (5thed). USA: Wiley- Blackwel.
- Washabaugh, William (1977). *Constraining Variation in Decrolization* . *Language* 53(2) 329-352.
- Winford, Donald (1988). *The creole continuum and the notion of the community as locus of language*. *International Journal of the Sociology of Language* 71. 91-105.
- Winford, Donald (2002). *An Introduction to Contact Linguistics*. Blackwell .