


Introducing the Historical Corpus of English in Nigeria (HiCE–Nig)

FOLUKE OLAYINKA UNUABONAH ,
 ADEBOLA ADEBILEJE,
 ROTIMI OLANRELE OLADIPUPO, BERNARD FYANKA,
 MBA ODIM AND OLUWATENIOLA KUPOLATI

A database for investigating diachronic linguistic changes in Nigerian English

1. Introduction

Nigerian English (NigE) is a second language (L2) variety of English that has been domesticated, acculturated and indigenised within the Nigerian socio-cultural and linguistic contexts (Adegbija, 2004). Based on Schneider's (2007) Dynamic Model of the Development of New Englishes, scholars have shown that NigE is currently at the late stage of nativisation (stage 3) and is on the verge of entering the stage of endormormative stabilisation (stage 4) (see Gut, 2012; Collins, 2020). Nativisation, which typically begins with the declaration of independence, is a very active and important stage in which there are large-scale linguistic changes, especially during a time when English is usually the only official language (see Schneider, 2007). Although previous studies have investigated the historical development of English in Nigeria (see Taiwo, 2009), there are limited studies on the particular linguistic features that have changed over time, especially from the time Nigeria gained independence. It is very likely that the rapid increase in the number of universities and other educational institutions managed by Nigerians from independence, as opposed to previous management by Britons, would have affected the variety of English spoken in Nigeria and culminated in the development of NigE today.

Moreover, previous investigations of World Englishes indicate diachronic changes in the development of different varieties of English, which have been established through corpus-based methods. Such diachronic corpus-based studies have

focused on different varieties of English, such as British English (Baker, 2011), Singapore English (Hoffmann, Sand & Tan, 2012), New Zealand



FOLUKE O. UNUABONAH's areas of specialisation include discourse analysis and (corpus) pragmatics. She has published papers in journals such as World Englishes, Intercultural Pragmatics, International Journal of Corpus

Linguistics, and Journal of Pragmatics.

Email: unuabonahf@run.edu.ng



ADEBOLA ADEBILEJE's areas of specialisations are Applied English Linguistics (Second Language Acquisition), Sociolinguistics and English Syntax. She has published papers in international journals such as Corpus Pragmatics and

Cogent Arts & Humanities.



ROTIMI OLADIPUPO has research interests in the phonetics and phonology of English, sociophonetics, language contact and corpus linguistics. He has published articles in Sage Open, Journal of Pragmatics and World Englishes.



BERNARD B. FYANKA holds a Ph.D. in History and Strategic Studies from the University of Lagos, Akoka Lagos Nigeria. He is the author of International Patterns of Small Arms Control: An Evolutionary Assessment and the

Imperative for Strategic Control.



MBA O. ODIM has research interests in Machine and Deep Learning, Datamining, Database Systems and Information Security. He is a reviewer of a number of reputable journals and an awardee of the IBM Artificial Intelligence Analyst –

Mastery and Explorer Award 2019.



OLUWATENIOLA OLUWABUKOLA KUPOLATI is a sociolinguist whose research focus is on language use in diverse social contexts. She is also interested in contact linguistics and political debate discourse. She has

published articles in peer-reviewed journals, including Journal of Multilingual and Multicultural Development and Discourse & Communication.

English (Hundt, 2012), Philippine English (Collins, Borlongan & Yao, 2014; Collins, 2015), Australian English (Hundt, 2015), and Ghanaian English (Brato, 2018), without investigating diachronic changes that have affected the English language in Nigeria.

There is still limited understanding of the changes that may have affected the English language in Nigeria and the socio-historical factors that may have led to these changes. Although Fuchs and Gut (2015) examined the development of the progressive in NigE, this was done using the apparent-time approach, since real-time data from the past was not available. This underscores the need for a historical corpus of NigE, in order to examine linguistic changes that have taken place over time.

In the current paper, we present the Historical Corpus of English in Nigeria (HiCE–Nig), a 1,000,000-word corpus of written Nigerian English. The corpus comprises texts collected

from the period 1960–1999, which covers the early to middle nativisation phase of NigE. The text categories are similar to those in the International Corpus of English–Nigeria (ICE–Nig; Wunder, Voormann & Gut, 2010), which makes it a comparable corpus with ICE–Nig. Both corpora together can thus be used for tracing nativisation across time since ICE–Nig captures the late nativisation stage in the development of NigE. HiCE–Nig also benefits from the structural makeup of the Historical Corpus of English in Ghana (HiCE–Ghana; Brato, 2019a). This is also important considering that Nigeria and Ghana belong to the same region (West Africa) and share a similar colonial history. Thus, the onset of the nativisation period in both countries is close/similar. This synergy can facilitate comparative studies that examine linguistic changes across the region.

The remaining parts of this paper are as follows. Section 2 provides a brief outline of the sociolinguistic and historical development of NigE, Section 3 deals with methodological aspects in the compilation of the corpus, while Section 4 provides examples from the corpus. Section 5 concludes the study.

2. English in Nigeria

The earliest history of the English language in Nigeria dates back to the mid-16th century when the British arrived in the coastal areas of Nigeria for reasons of trade in both humans and non-human resources. From the middle of the 19th century, the Christian missionaries pioneered institutionalised Western education in Nigeria, and for about the next four decades they were in charge of language education in the country (Fafunwa, 1974; Taiwo, 1980). Thus, Christian education in Nigeria became a potent tool for spreading a type of Standard English (Ogu, 1992).

The involvement of the British colonial government in the educational sector of the country began to be felt in the 1880s (Awobuluyi, 1996). This was necessitated by the workforce need of the colonial administration. The colonial government promulgated various guidelines and ordinances in order to emphasise the learning of the English language. First, English was declared the language of instruction in schools. Second, a pass in English language became a prerequisite for certification which, invariably, presupposed that only those who passed and could speak English had access to job opportunities. These efforts encouraged the spread of English in Nigeria.

Other factors further helped in the spread of the English language, such as the multilingual nature

of Nigeria, admission criteria to enter schools and employment, societal attitudes towards speakers of the language, the place of English in the present language policy of Nigeria and the state of development of indigenous Nigerian languages. Today in Nigeria, English is the language of the media, government, education, religion, law, international communication and formal business transactions (Akindele & Adegbite, 1999; Jowitt, 2019). Although English is a second language for many Nigerians, English is now a first language for a growing number of young people in Nigeria (see Onabamiro & Oladipupo, 2019).

3. Compiling HiCE–Nigeria

This section focuses on the design, data collection, and challenges faced in compiling HiCE–Nig.

3.1 Design of HiCE–Nig

This section addresses the four critical areas of size, content, sampling, and representativeness, which are important in the design of a corpus.

3.1.1 Content and size

We decided at the beginning of the project that the content and size of the corpus should be similar to the written part of ICE–Nig for comparative purposes. However, this could not be completely achieved. In the first instance, similarity in content was not possible due to limited access to non-printed texts such as social letters and exam scripts. Very few social letters were obtained as these could only be collected from people personally, and in many cases, those contacted indicated that they had disposed of such letters when moving from one house to another. In the case of exam scripts, it was typical for universities to destroy such documents after a five-year period. To make up for this, other text categories such as those included in HiCE–Ghana were included, for example, letters to editors. In addition, we included written addresses or speeches produced by politicians, technocrats and military rulers. This is based on the belief that these texts were first written before being read to the public, and thus have similar features to written texts. Based on ICE–Nig, the texts were those produced by Nigerians who were over 18 years, had at least a secondary school education and mainly lived or resided in Nigeria, at the time of production. In some cases where such information could not be ascertained, these parameters were determined by the kind of materials collected, which could only have been produced by

Nigerians; for instance, news reports, academic writing, popular writings, novels and administrative writings. The place of publication of such materials also helped in authenticating that the materials collected were produced in Nigeria.

In addition, due to the period indicated for the collection of the materials, we decided to expand the size of the corpus from 400,000 words to 1,000,000 words (see Table 1 for all text categories). We also decided that the texts must have been written between 1960 and 1999 (a period of 40 years), which captures the beginning and the middle stages of the nativisation process. While 1960 was the year Nigeria gained independence, 1999 was chosen as the end period for the corpus since ICE–Nig contained texts produced from the early 2000s, which captures the late nativisation stage.

3.1.2 Sampling and representativeness

The data were sampled from a 40-year period, divided into two eras, i.e. 1960–1979 and 1980–1999, in order to achieve balance. These dates were also chosen based on socio-historical factors. The period of 1960–1979 covers independence to the post-civil war period, while 1980–1999 covers the period from the second republic to the beginning of the fourth republic. During this period, there was a changeover of government from the military to civilian rule in 1979, the first state-owned university was established in 1980 (Enugu State University of Technology), and another period of military rule ended in 1999. Thus, the corpus was divided into two sub-corpora, each containing 500,000 words, to form a total of one million words.

In order to ensure representativeness, we collected data from different ethnic groups in the country such as Ishan, Efik and Ishekiri. However, most of the materials were obtained from the three major ethnic groups: Hausa, Yoruba and Igbo. The ethnic groups and gender of some of the participants were largely obtained through the names of the text producers, as it was possible to obtain this information through their names. This, of course, was not possible for all text producers, as names were not listed in some documents while some names were unisex.

3.2 Data collection

Materials for the corpus were collected from different university departments, as well as national and university libraries across the country in order to ensure representativeness. In particular, administrative

Table 1: Target word count in thousands in HiCE–Nig, ICE–Nig and HiCE–Ghana

| Text categories | HiCE–Nig | ICE–Nig | HiCE–Ghana |
|-----------------------------------|--------------|------------------------|------------|
| Academic writing humanities | 60 | 20 | 30 |
| Academic writing natural sciences | 60 | 20 | 30 |
| Academic writing social sciences | 60 | 20 | 30 |
| Academic writing engineering | 60 | 20 | 30 |
| Administrative writing | 40 | 20 | 40 |
| Business letters | 40 | 30 | - |
| Editorials | 40 | 20 | 50 |
| Letters to editors | 40 | - | 50 |
| Novels | 80 | 40 | 80 |
| Popular writing humanities | 60 | 20 | 30 |
| Popular writing natural sciences | 60 | 20 | 30 |
| Popular writing social sciences | 60 | 20 | 30 |
| Popular writing technology | 60 | 20 | 30 |
| Press reportage | 100 | 40 | 100 |
| Skills and hobbies | 40 | 20 | 40 |
| Student essays | 80 | 20 | - |
| Written addresses | 60 | - | - |
| Total | 1,000 | 350¹ | 600 |

writings were obtained from newspapers, universities, and different government ministries and agencies; the National Archives in Kaduna was particularly helpful in this respect. Academic writings were largely taken from journals owned by academic associations in different fields, such as the *Journal of the Nigerian English Studies Association*, *Journal of English as*

a Second Language, *Journal of the Historical Society of Nigeria*, *Nigerian Journal of Science*, *Nigerian Journal of Nutritional Sciences*, *Nigerian Journal of Microbiology*, *Ife Psychologia*, *The Nigerian Journal of Economic and Social Studies*, *The Quarterly Journal of Administration*, *Journal of the Nigerian Society of Chemical Engineers*, *The Nigerian Engineer*, and *Nigerian Journal of Technology*. A number of these journals were obtained from departmental and university libraries in Obafemi Awolowo University and the University of Ibadan, while others were downloaded from open access journals in African Journals Online (AJOL) and open educational resources of the University of Nigeria, Nsukka. Newspapers and magazines, which contained editorials, letters to editors, popular writings, news reports, and records of skills and hobbies, were mainly collected from the Gandhi Library of the University of Lagos, and the National Archives in Ibadan and Kaduna. Notable among these newspapers were *West African Pilot*, *The Morning Post*, *Nigerian Tribune*, *Daily Express*, *Daily Times*, *New Nigerian*, *Punch*, and *the Guardian*. Student essays involved final-year long essays and were obtained from different departments such as English, Animal Science, Economics, History, Microbiology, Chemistry, and Mechanical Engineering. While the 1960 essays were collected from Obafemi Awolowo University, Ile-Ife, University of Ibadan, Ibadan, and Ahmadu Bello University, Zaria, most of the essays from the 1980s were collected from University of Port Harcourt, Port Harcourt, Imo State University, Owerri, and Ambrose Alli University, Ekpoma. Written addresses were obtained from the speeches that had been posted online as well as those published in newspapers, magazines and journals. In most cases, the materials were photocopied at the different sites before being brought to Redeemer's University for further processing.

There were some challenges during the collection of data. For example, there was restricted access to some texts in some university libraries, while many texts were published outside Nigeria in the 1960s. Most business letters found in the 1960s were written by non-Nigerians; thus, the business letter category for 1960–1979 is still pending. Newspapers from 1960 to 1979 had fewer pages and fewer sections, and in some cases, there were no letters to editors, editorials, or sections on science and technology, etc. Hence, some of these sections could not be expanded. Novels written around 1960 were mainly published in the United Kingdom; however, some were included considering that some of these

texts had been used in the identification of linguistic features of NigE (see Bamiro, 1991, 1994).

3.3 Instruments and materials

The instruments used in processing the data included laptops, phone cameras, and scanners with optical character recognition (OCR) software. The scanners were used to scan the photocopied documents, while the OCR software was used to extract the words from the scanned documents so that the materials could be stored in a format that could be analysed by different corpus software. In some cases, due to the fragility of some old texts, phone cameras were used to capture the texts, which were then printed and retyped. Some photocopied materials and some texts such as student essays that were produced using old typewriters were also retyped because the software could not read the materials accurately due to damage.

The data were then presented as .txt files, and were named using the same file naming patterns as in ICE–Nig with the addition of H1 and H2 representing the sub-corpus of 1960–1979 and 1980–1999, respectively, i.e. H1Ess_01.txt and H2Ess_01.txt for student essays. Where available, metadata on the transcribers' initials, place and time of text production, author's gender, age and ethnic group were also recorded in an Excel sheet. Although it was the intention of the compilers to carry out the annotation of the corpus, this was not possible at the time due to financial and other constraints. In the near future, annotation will be done.

4. Opportunities

The purpose of the ICE corpora is to supply resources for describing the distinguishing features of an English variety as well as to compare the varieties of World Englishes (Adegbite, 2012). Thus, different kinds of diachronic linguistic changes in NigE can be studied when HiCE–Nig is used in combination with ICE–Nig. Hence, using the AntConc corpus analysis toolkit (Anthony, 2015), orthographic, lexical, grammatical and discourse variation may be searched in the HiCE–Nig corpus in order to examine the development of the English language in Nigeria. For example, at the orthographic level, we can see changes in the spelling of items with local names, i.e. different spellings of *fufu* (fermented cassava, which is a staple food in some parts of West Africa) and *moin moin* (bean cake) in HiCE–Nig and ICE–Nig. In HiCE–Nig, *fufu* is spelt as 'foofoo', as shown in

(1) and (2), while in ICE–Nig, it is spelt as 'fufu', as exemplified in (3). In (2), *moin moin* is spelt as 'moyin-moyin' in HiCE–Nig and as 'moin moin' in ICE–Nig, as cited in (4).

- (1) To serve: Eat with yam flour (amala), eba or cassava **foofoo**. (HiCE–Nig_H2Skh_01.txt)
- (2) certain Nigerian lexemes have entered the English language – words like '**foofoo**', 'buba', '**moyin-moyin**', 'kiakia', etc. (HiCE–Nig_H2Ess_06.txt)
- (3) Fermented cassava roots are boiled and pounded into a thick paste called **fufu** which is consumed with soup. (ICE–Nig_ANsc_03.txt)
- (4) Generally, paps may be taken alone or with sugar or with bean cakes i.e. 'akara' or '**moin moin**' (made from *Vigna unguiculata* (L.)) (ICE–Nig_PNsc_01.txt)

At the lexical level, we see the overuse of fixed expressions such as *all and sundry*, as exemplified in (5), which is noted also in HiCE–Ghana (see Brato, 2019b). In HiCE–Nig, there were eight instances of the expression, while in written ICE–Nig, there were only two instances of the expression.

- (5) They protested, and their protests received the active support of **all and sundry**. (HiCE–Nig_H1Ed_20.txt)

At the discourse-pragmatic level, there is the entry of new discourse markers that may have developed from adverbs, such as the case of *moreso* from *more so* (see Unuabonah, 2019), as depicted in (6):

- (6) **Moreso**, the experts handling specialized areas in automobile repairs in the workshop have undergone refresher courses on injectors (HiCE–Nig_H2PTec_22.txt)

A preliminary investigation of some intensifiers shows a continuous reduction in the frequency of intensifiers such as *greatly* and *fairly* in (7) and (8), respectively. Both *greatly* and *fairly* had relative frequencies of 0.8, 0.4, and 0.2² in HiCE–Nig (1960–1979), HiCE–Nig (1980–1999), and ICE–Nig, respectively.

- (7) These Karamat should be distinguished from magic-making which was **greatly** condemned By the Fulani writers! (ICE–Nig_H1AHum_14.txt)
- (8) setting up a first class salon costs not less than N150, 000 but for a **fairly** equipped one, N50, 000 may be enough for a start. (HiCE–Nig_H2Skh_20.txt)

Other investigations show an increase in the frequency of stance markers such as *for sure* and

really, as cited in (9) and (10), respectively. *For sure* had relative frequencies of 0, 0.1, and 0.2 in HiCE–Nig (1960–1979), HiCE–Nig (1980–1999), and ICE–Nig, while *really* had relative frequencies of 1.2, 1.2, and 2.8, in the three NigE sub-corpora, respectively.

- (9) **For sure**, a headline that is bold and large will attract attention quicker than one that is small and not so bold. (HiCE–Nig_H2Ess_07.txt)
- (10) Well, I am not Chineke; I do not create children, ’**Really** I was not thinking of that.’ Madame reassured his wife. (HiCE–Nig_H1Nov_03.txt)

5. Conclusion

The creation of HiCE–Nig no doubt provides a rich database which can be explored to track the progress of nativisation of the English language in Nigeria or combined with ICE–Nig to identify and compare diachronic linguistic changes in NigE at different structural levels. HiCE–Nig also benefits from the structural makeup of the HiCE–Ghana (Brato, 2019a), which is also important considering that Nigeria and Ghana belong to the same region (West Africa), and share a similar colonial history and onset of the nativisation stage. This can also facilitate comparative studies that examine linguistic changes across the region. The inclusion of participants’ information, such as place and time of text production, author’s gender, and ethnic group in the corpus, where possible, is also useful for variation studies across linguistic levels. Finally, it is believed that the annotation of the data, when undertaken, will make the corpus much more informative and easier to search.

Acknowledgements

The compilation of the corpus was funded by Redeemer’s University, Ede, Nigeria. The authors appreciate the research assistants who worked on the project including Dr. Ayo Ojebode, Dr. Florence Daniel, Adesewa Odesanya, Terlanga Koko, Tinuade Onabamiro, Flora Ayodele and Aderonke Akinola. The authors are also grateful to colleagues who helped in providing access to many materials, among whom are Prof. Rotimi Taiwo, Dr. Omodesola Ogunraku (Obafemi Awolowo University, Ile–Ife), Dr. Uzoamaka Akwiwu (Imo State University, Owerri), Dr. Christian Eruteya (University of Port Harcourt, Port Harcourt), Faith Unuabonah (Ambrose Alli University, Ekpoma), Dr. Kemi Adesina, Dr.

Stephen Eyeh, Dr. Eniola Boluwaduro (Redeemer’s University) and many others who are not mentioned here. The authors sincerely thank Prof. Ulrike Gut for technical support on the project as well as for reading an earlier draft of this paper.

Notes

- 1 This excludes social letters (30,000) and exam scripts (20,000), which are not included in HiCE–Nig.
- 2 The categories searched in all three NigE sub-corpora include academic writings, administrative writings, editorials, novels, popular writings, press reports, skills/hobbies, and student essays.

References

- Adegbija, E. 2004. ‘The domestication of English in Nigeria.’ In V. Awonusi & E. A. Babalola (eds.), *Domestication of English Language in Nigeria: A Festschrift in Honour of Abiodun Adetugbo*. Lagos: University of Lagos Press, pp. 20–44.
- Adegbite, A. 2012. ‘The ICE Nigeria corpus as a data base for Nigerian English studies.’ *Ife Studies in English*, 10(1), 1–10.
- Akindele, F. & Adegbite, W. 1999. *The Sociology and Politics of English in Nigeria: An Introduction*. Ile–Ife: Obafemi Awolowo University Press.
- Anthony, L. 2015. *AntConc* (Version 3.4.4) [Computer Software]. Waseda University, Tokyo. Online at <<http://www.laurenceanthony.net>>.
- Awobuluyi, O. 1996. ‘Language education in Nigeria: Theory, policy and practice.’ Online at <<https://yeyeolade.wordpress.com/2009/05/15/>> (Accessed April 2009).
- Baker, P. 2011. ‘Times may change, but we will always have money: Diachronic variation in recent British English.’ *Journal of English Linguistics*, 39(1), 65–88.
- Bamiro, E. 1991. ‘Nigerian Englishes in Nigerian English literature.’ *World Englishes*, 10(1), 7–17.
- Bamiro, E. 1994. ‘Lexico-semantic variation in Nigerian English.’ *World Englishes*, 13(1), 47–60.
- Brato, T. 2018. ‘“Outdooring” the historical corpus of English in Ghana: Insights from the compilation of a historical corpus of New English.’ *English Today*, 34(2), 25–34.
- Brato, T. 2019a. ‘The historical corpus of English in Ghana (HiCE–Ghana): Motivation, compilation, opportunities.’ In A. Esimaje, U. Gut, & B. E. Antia (eds.), *Corpus Linguistics and African Englishes*. Amsterdam: John Benjamins, pp. 119–141.
- Brato, T. 2019b. ‘Lexical expansion in Ghanaian English from a diachronic perspective: A structural and semantic analysis.’ In A. Esimaje, U. Gut, & B. E. Antia (eds.), *Corpus Linguistics and African Englishes*. Amsterdam: John Benjamins, pp. 260–91.
- Collins, P., Borlongan, A. M. & Yao, X. 2014. ‘Modality in Philippine English: A diachronic study.’ *Journal of English Linguistics*, 42(1), 68–88.
- Collins, P. C. 2015. ‘Recent diachronic change in the progressive in Philippine English.’ In P. Collins (ed.), *Grammatical Change in English World-Wide*. Amsterdam: John Benjamins, pp. 271–96.

- Collins, P. C. 2020. 'Comment markers in world Englishes.' *World Englishes* (Online first). <<https://doi.org/10.1111/weng.12523>>.
- Fafunwa, A. 1974. *History of Education in Nigeria*. London: George Allen & Unwin Ltd.
- Fuchs, R. & Gut, U. 2015. 'An apparent time study of the progressive in Nigerian English.' In P. Collins (ed.), *Grammatical Change in English World-Wide*. Amsterdam: John Benjamins, pp. 373–87.
- Gut, U. 2012. 'Towards a codification of Nigerian English: The ICE Nigeria project.' *Journal of the Nigerian English Studies Association*, 15(1), 1–12.
- Hoffmann, S., Sand, A. & Tan, P. 2012. 'The corpus of historical Singapore English – A first pilot study on data from the 1950s and 1960s.' [Paper presentation] ICAME 31 Conference, Leuven, Belgium.
- Hundt, M. 2012. 'Towards a corpus of early written New Zealand English – News from Erewhon?' *Te Reo*, 55, 51–74.
- Hundt, M. 2015. 'Do-support in early New Zealand and Australian English.' In P. C. Collins (ed.), *Grammatical Change in English World-Wide*. Amsterdam: John Benjamins, pp. 65–86.
- Jowitt, D. 2019. *Nigerian English*. Berlin: Mouton de Gruyter.
- Ogu, J. N. 1992. *A Historical Survey of English and the Nigerian Situation*. Lagos: Kraft Books Ltd.
- Onabamiro, T. & Oladipupo, R. 2019. 'Forms and patterns of epenthesis of New English native speakers in Nigeria.' [Paper presentation] English Scholars' Association of Nigeria Conference, Abuja, Nigeria.
- Schneider, E. W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press
- Taiwo, C. O. 1980. *The Nigerian Education System: Past, Present and Future*. Lagos: Thomas Nelson.
- Taiwo, R. 2009. 'The functions of English in Nigeria from the earliest times to the present day.' *English Today*, 25(2), 3–10.
- Unuabonah, F. O. 2019. 'Frequency and stylistic variability of discourse markers in Nigerian English.' *Corpus Pragmatics*, 3, 249–271.
- Wunder, E.-M., Voormann, H. & Gut, U. 2010. 'The ICE Nigeria corpus project: Creating an open, rich and accurate corpus.' *ICAME Journal*, 34, 78–88.
-