

Corpus-based Study on African English Varieties*

Xiaohui Xu

Qingdao University of Science and Technology, Qingdao, China

Abstract—Corpus-based research is more and more used in linguistics. English varieties are used a lot in daily communications throughout the world. African English varieties are discussed in this paper, including West African English, East African English and South African English. Kenya and Tanzania corpus is the main target corpus while Jamaica corpus is used as a comparative one. The tool used is AntConc 3.2.4.

Index Terms—corpus, English varieties, African English, pidgins, creoles

I. INTRODUCTION

A. Standard English

There is an agreeable division among scholars that the whole world is divided into three circles: the Inner Circle, the Outer Circle, and the Expanding Circle. In the Inner Circle, English is spoken as mother tongue; in the Outer Circle, English is usually spoken as a second language; in the Expanding Circle, English is usually spoken as a foreign language.

Standard English is used in books, newspapers, magazines and nearly everything else that appears in print in the English-speaking world. This type of English is called “standard” because it has undergone standardization, which means that it has been subjected to a process through which it has been selected and stabilized, in a way that other varieties have not.

In the case of certain other languages, “selected” means that an official decision was made at a point for one particular dialect of a language to receive the standardization treatment, as opposed to any of the others. This is not what happened with English.

Standard English acquired its status more gradually and in an organic way. The ancestor of modern Standard English developed around the Royal Court in London, among the aristocracy and ruling elite. Because the elite were concentrated in London, this pre-Standard English was a dialect of a predominantly London-area type. Because it was associated with people who were of mixed geographical origins and who were unusually mobile and well travelled, this court dialect showed signs, from the very early records that we have, of being a mixed dialect. For example, the language of the Proclamation of Henry III shows a blending of Midland and southern features. The form of language which at last emerged over the centuries as the preferred way of writing among the governing classes had features which were not only south-eastern in origin but also in southern area and Midland.

B. Second Language Varieties of English

Kachru (1985) pointed out that the whole world is divided into three circles: The Inner Circle, the Outer Circle, and the Expanding Circle.

English of a distinctively East African type is spoken as a second language in Kenya, Tanzania and Uganda. These countries belong to the Outer Circle. The English spoken by Africans in Malawi, Zimbabwe, Zambia and South Africa is also of a very similar type.

English is a language which has more non-native speakers than native speakers. The non-native speakers can be divided into two types. First, there are speakers of English as a Foreign Language (EFL) who learn English as a tool of international communication.

People in Germany, Japan, or Morocco who have learnt English will normally expect to use it in interaction with people from other countries. Second, there are speakers of English as a Second Language (ESL). These are to be found in the nations where English is used as an official language, or as a language of education, by people who are not native speakers. There are many such countries in the world.

In Africa, there are large communities of native speakers of English in Liberia, South Africa, Zimbabwe and Kenya, but there are even larger communities in these countries of second language speakers. Elsewhere in Africa, English has official status, and is therefore widely used as a second language lingua franca in Ghana, Nigeria, Cameroon, Zambia and Uganda. It is also extremely widely used in education and for governmental purposes in Tanzania and Kenya.

In the Indian Ocean, Asian, and Pacific Ocean areas, English is an official language in Singapore, Hong Kong, the

* This paper is the result of the college research project of Education Department of Shandong Province (J16WC32) “Corpus-based Study of English Varieties in Africa and Sino-Africa Communications of Language and Culture” and Qingdao Social Science Program research project in 2016 (QSKL1601140) “the Study of English Varieties in “One Belt, One Road” Countries and Communications of Language and Culture.”

Philippines, and somewhere else in American-administered place. It is also very widely used as a second language in Malaysia, Bangladesh, Sri Lanka.

In many of these areas, English has become or is becoming indigenized. This means that the second language varieties of English, as a result of frequent use, have acquired relatively consistent, local norms of usage which are agreed by all speakers.

Gao Yihong, Xu Hongchen (2014) published a paper about the research of the attitude to English varieties during international games. Pan Zhangxian (2002), Hu Xiaoli (2008) and Zou Hang (2015) concluded Chinese English Varieties.

These varieties of English may differ from the English of native speakers elsewhere in the world, mainly as a result of influence from local languages. Thus native speakers of English may sometimes have some difficulty in understanding the non-native varieties. But it is more important for native speakers to understand it. After all, Americans may find it is difficult to understand, but no one would seriously suggest that this is a reason for changing it. There are some West Africans believe that they are speaking Standard English, or at least aim at speaking it.

Obviously, for Africa themselves, the margin for tolerance of differences can be much greater. Equally as important, we believe that native English speakers travelling to areas such as Africa should make the effort to improve their understanding of the non-native variety of English.

There are some well-established second language varieties of English in African English. The focus of the paper is on varieties of Standard English, the kind of English written and spoken by educated English speakers in Africa.

C. *English-based Pidgins*

Human beings appear to be biologically programmed to learn languages in early ages: children up to the age of 3 or 4 learn their native language rapidly and perfectly, and without any overt guide. Adults, on the other hand, tend to be rather poor language learners and only in rare instances manage to learn foreign languages so well that they can speak them like native speakers.

Whenever someone other than a small child tries to learn another language, the processes which originate from this imperfect learning ability will always occur. In particular, in the speech of such adult language learners, the language will be somewhat simplified and mixed.

Simplification is a complex concept, but it can best be understood as involving regularization and loss of redundancy. Regularization, obviously, means treating irregular forms as if they were regular, such as when an English learner said "I eated" rather than "I ate".

Loss of redundancy often involves the omission of grammatical material which is repeated elsewhere or is not absolutely necessary for conveying the message intended, as when a learner of English says she want rather than she wants. The grammatical category of third-person singular is conveyed only by the pronoun she rather than by the pronoun and the -s ending on the verb.

Mixing is a term which refers to the way in which language learners introduce elements from their own language into the target language. For example, a French-speaking learner of English may have a French accent in their English and may also use some French grammatical constructions and idioms.

Generally, compared to the language of native-speakers, adult learners' language will also be reduced. Because they do not know so much of the language, and because they use it for a restricted range of purposes, they will use fewer words, fewer grammatical constructions and fewer idiomatic devices.

When a language experiences such simplification, mixture and reduction, we can say that it has been subjected to the process of pidginization. When language learning takes place after an extended period pidginization will tend to be slight in a classroom.

Otherwise, if contact with the foreign language is minimal and short-lived, and the language is learned without formal guide, then pidginization may be extreme. In certain special social situations, it can happen that an extremely simplified, reduced and mixed form of language of this type comes to be useful as a means of communication among groups of people who have no native language in common. It may develop into a fixed form with norms that are shared by a large number of speakers which can subsequently be passed on to and learned by others. A language like this is referred to a pidgin. Todd (1984) discussed Pidgins and Creoles.

A development of this type occurred in West Africa, as a result of early colonization, Africanized, reduced form of English, acquired from limited contacts within traders and sailor, became useful as a lingua franca among different groups of the indigenous population. It then developed into the pidgin language that we call West African Pidgin English today. Until now, West African Pidgin English is widely used as a trading language along the coast of Africa from Gambia to Cameroon and Equatorial Guinea.

Compared with Standard English, this language is mixed-there are elements in its pronunciation, grammar and vocabulary which are due to indigenous African languages.

West African Pidgin English is a language with norms and grammatical rules that speakers have to learn in order to speak and understand the language correctly. Other well-known English-based pidgins are found in the South Pacific. Tok Pisin in Papua New Guinea, Bislama in Vanuatu and Solomon Islands Pidgin are closely related, and all are official languages in their respective countries.

D. *English-based Creoles*

Under certain social circumstances, a pidgin language uses as a trading language or lingua franca. Indeed, in some cases it can become the most important or even sole language of a community and be passed on to the next generation of children, later it will be their native language. When a pidgin language takes on a full range of social functions in this way and acquires native speakers, it is known as a creole. Like a pidgin, a creole is still relative to its source language, simplified and mixed. It is no longer reduced. Because the language now has to be used for all the purposes a native speaker needs to use a language for, the reduction that took place during pidginization has to be repaired by a process of expansion.

This expansion process is known technically as creolization. During creolization, vocabulary is developed and expanded, grammatical devices and categories are added to, and the language acquires a wide range of styles. Creole languages are thus perfectly normal languages although their histories are interesting and unusual, and are just as adequate means of communication and expression as any other language.

There are many English-based creoles in the world, and the number is growing: Tok Pisin, Bislama and Solomon Islands Pidgin are all currently going through the creolization process, as is West African Pidgin English, particularly in Nigeria and Cameroon. Presumably a majority of English-based creoles are spoken in the Atlantic Ocean area, which is a result of the slave-trade. As slaves from many different African ethnic groups were assembled on the coast of West Africa and transported to the Americas, they discovered, in their multilingual situation, that English-based pidgin varieties were an important means of communication, and in many parts of the Western Hemisphere, creole languages developed as a result of this need.

For example, an English-based creole is spoken in Surinam, in South America. Sranan, although historically related to English, is now a separate language.

II. AFRICAN ENGLISH VARIETIES

A. *South African English*

There are about 45 million people in South Africa. The African majority, about 70 percent of the population, speak Bantu languages such as Zulu, Xhosa, Sotho and Tswana. About 20 percent of the population speak Afrikaans, a language of European origin related to Dutch, and a small percentage speak languages of Indian origin. The surviving indigenous languages of the area, the Khoisan languages, are in a stronger position in neighboring Namibia and Botswana than they are in South Africa. Lanham (1967) concluded the pronunciation of South African English.

English has had significant numbers of speakers in South Africa since the 1820s thcolored mixed race and Indian-origin speakers.

English is also widely spoken as a second language. Forms of English which closely resemble South Africa English are also spoken natively in Zimbabwe, as well as by relatively small groups of whites in Namibia, Kenya and Zambia.

In what follows we pay our attention to the English of native speakers in South Africa.

1.1 South African English grammar and usage

There are some grammatical differences between South African English and Standard English.

1. A common use of the response question is *it?*, invariable for person, tense or auxiliary, which corresponds to the complex forms *do they, can't he, shouldn't we, will you*, etc. used in other varieties:

He's gone to cinema.-Oh, is it?

2. It is common to delete object noun phrases (NPs) after verbs which must have NPs in other varieties, e.g.:

Have you sent?

Have you got?

Did you put?

In the above sentences, the objects are all omitted.

3. Complement structures of adjective + infinitive occur where other varieties have adjective + *of*+ participle:

This plastic is capable to withstand heat. =This plastic is capable of withstanding heat.

4. Non-negative *no* occurs as an introductory particle:

How are you?

No, I'm fine, thanks.

The force of this is often to negate assumptions made in the preceding question or comment. Here "no" is not the usual meaning.

1.2 South African English lexis

There are some vocabularies originating from South African local languages. The better-known borrowings are:

from Zulu:

indaba means "conference"

from Afrikaans:

kraal means "African village"

dorp means "village".

Things are similar in different parts of Africa.

B. *West African English*

West African English is spoken by non-native speakers of English in Ghana, Nigeria, Liberia and Cameroon.

2.1 West African English grammar

West African English varies greatly from place to place: some of the forms occur in Nigeria, but not in Ghana or vice versa.

It also varies very much according to the education of the speaker and the form of the situation. Some forms given here are not used in writing. Where the grammar of West African English differs from that of other varieties of English, this is often because of the influence from local languages. This influence is most marked in less educated and informal styles.

Typical West African English grammatical forms include the following:

1. Pluralization of non-count nouns:

I like all my furnitures.

2. Omission of articles:

I am going to office.

3. The use of resumptive pronouns in some colloquial styles of English:

My sister, she's crazy.

or in relative clauses in a non-English manner:

The guests whom I invited them have arrived.

4. Formation of comparative clauses without using the comparative form of the adjective:

It is the youths who are skillful than the adults.

5. Absence of infinitival to after some verbs:

They ordered him do it.

6. No distinction between the reflexive pronoun themselves and the reciprocal pronoun each other:

"The like themselves" means "They like each other".

7. The use of a universal tag question-is it?:

We should go now, is it?

She has gone home, is it?

8. A non-English use of yes and no in answering questions:

Hasn't he come back yet?

Yes= He hasn't come back yet

No= He has come back

9. The use of progressive aspect with have while expressing a temporary state:

I am having a cold.

2.2 West African English lexis

Some differences in vocabulary between West African English and other varieties of English involve extensions or alterations to the semantic or grammatical function of English words. Others reflect usages of equivalent words from local languages, while still others are innovations. The list below gives a few examples by way of illustration. For example:

To bluff also means "to dress fashionably" or "to show off".

Guy means "an outgoing, self-assured young man".

Hot drink means "alcoholic spirits, liquor".

To take in also means "to become pregnant".

The difference may due to several things, including exposure to literary rather than colloquial English and the prestige of the written English.

C. *East African English*

East African English is spoken as a second language in Kenya, Tanzania and Uganda. The English spoken by Africans in Malawi, Zambia, Zimbabwe and South Africa is also of a very similar type.

Most of the local languages in most of the areas of eastern and southern Africa are members of the Bantu language family, while these have played an important role in influencing the nature of East African English.

There are many similarities between East African English and West African English. But there are still some differences.

A number of words from indigenous languages are used by East African English speakers even when speaking and writing English. These include the following:

askari means "policeman"

chai means "tea"

kibanda means "black market"

matatu means "taxi bus"

These words are all used in East African English because of the recognition of local people.

III. ANALYSIS OF KENYA AND TANZANIA CORPUS

Detailed research is carried out about Kenya and Tanzania corpus and a lot of discoveries are found out. Jamaica corpus is picked out as a comparative corpus for there are similarities while differences in their cultures. The corpus is ICE (International Corpus of English).

The target corpora are Kenya, Tanzania and Jamaica corpus. The top 20 vocabularies used in the corpus are listed. The vocabularies include nouns, verbs, adjectives and adverbials.

It can be seen that in the topic of Kenya, the problem of women and children is concerned. House, money, school and work are also concerned. In the process of speaking, things about like or do not like is discussed a lot. The common verbs say, tell, order, let are also used a lot. About adjectives, the topic of Africa, nation and society are common topics. The adjectives only, good, new and kind such positive words are used a lot. In the usage of adverbial, also, very, even, still and never are used a lot to describe a verb.

FIGURE1 KENYA

	noun	verb	adjective	adverbial
1	name	know	very	also
2	people	said	more	then
3	time	like	only	very
4	Kenya	think	most	just
5	women	see	well	more
6	government	make	same	only
7	children	told	much	even
8	development	order	another	most
9	country	made	good	well
10	language	let	right	over
11	day	tell	new	much
12	man	need	African	still
13	minister	give	national	right
14	education	find	last	really
15	house	look	kind	last
16	school	came	social	never
17	case	point	long	maybe
18	money	mean	different	down
19	work	done	down	without
20	world	put	public	within

It can be seen that in the topic of Tanzania, the problem of women and children are also concerned. Time, education and health are also concerned. Salaam is a special word here, showing their culture. In the process of speaking, things about like or do not like is also discussed a lot. The common verbs say, see, call, tell are also used a lot. Order, support and control are also used a lot. About adjectives, the topic of economics, nation and policy are common topics. The adjectives only, most, same and high such positive words are used a lot. In the usage of adverbial, very, much, without and within are used a lot to describe a verb. Especially, really and actually are often used.

FIGURE2 TANZANIA

	noun	verb	adjective	adverbial
1	people	said	very	very
2	Tanzania	like	only	only
3	government	know	well	well
4	name	say	most	most
5	development	use	national	much
6	country	take	economic	still
7	time	see	African	without
8	women	need	political	right
9	year	come	social	within
10	education	order	much	again
11	case	mean	international	back
12	language	found	same	especially
13	Africa	support	given	really
14	years	control	united	actually
15	areas	done	local	maybe
16	problems	called	high	already
17	children	find	general	down
18	health	told	public	off
19	number	want	last	too
20	salaam	give	foreign	together

It can be seen that in the topic of Jamaica, the problem of students and man are concerned. Time, work and university are also concerned. Laugh is used a lot. In the process of speaking, things about like or do not like is also discussed a lot. The common verbs say, see, go, tell are also used a lot. Think, mention and let are also used a lot. The adjectives unclear,

bold, only and both such words are used a lot. In the usage of adverbial, very, much, well and within are used a lot to describe a verb. Really, never and actually are often used.

FIGURE3 JAMAICA

	noun	verb	adjective	adverbial
1	people	quote	unclear	well
2	time	know	bold	right
3	Jamaica	like	well	more
4	words	Think	right	very
5	work	go	more	really
6	students	say	very	back
7	way	see	only	only
8	year	come	much	even
9	man	want	new	much
10	thing	work	most	down
11	university	mention	many	most
12	years	laughs	long	many
13	mean	make	little	still
14	education	take	last	too
15	school	tell	same	off
16	things	use	another	last
17	government	look	different	never
18	day	let	each	alright
19	fact	point	both	actually
20	water	find	alright	within

Because the numbers of each corpus is different, the numbers are listed in figure4.

FIGURE4

	Kenya	Tanzania	Jamaica
types	28350	25401	32849
tokens	813652	627767	1548596

After checking the numbers, bar chart is used to show the features of the usage. Then another tool log-likelihood ratio calculator is used to analyze the characteristics of the varieties.

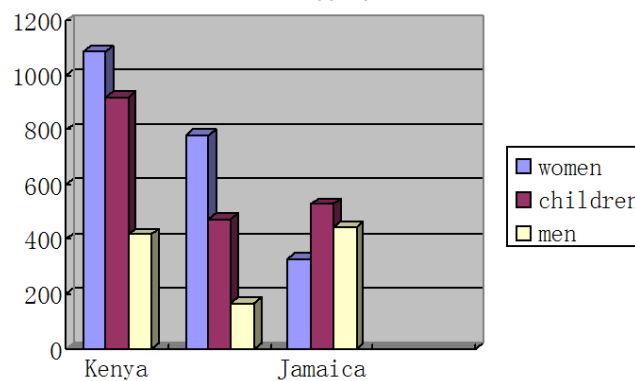
Of all the words listed, some special ones are picked out to do further analysis in bar chart.

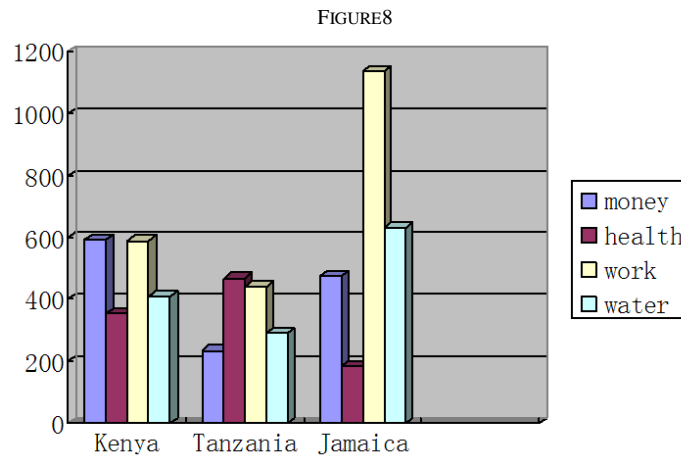
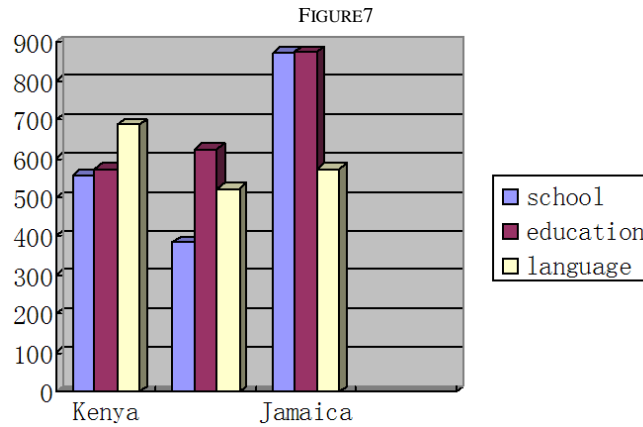
First, some nouns are picked out and the frequencies are listed in figure5. Because the number of Jamaica corpus is almost the total number of Kenya corpus and Tanzania corpus, the obvious difference can be easily observed in bar chart. And the result can be put into log-likelihood ratio calculator to test.

FIGURE5

	Kenya	Tanzania	Jamaica
women	1088	779	327
children	917	471	527
men	418	167	443
school	556	383	873
education	574	625	876
language	688	522	575
money	591	235	477
health	355	467	184
work	590	441	1136
water	409	290	632
salaam	4	457	0

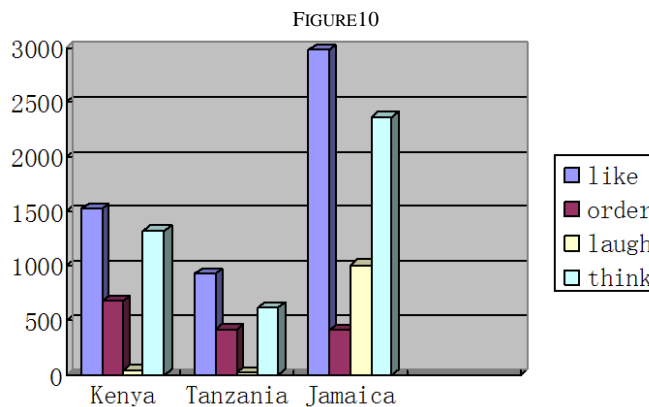
FIGURE6





Three bar charts are made to show the usage of nouns. It can be seen that the problem of women and children are cared a lot more in Kenya and Tanzania than in Jamaica. The topics of school and education are similar. Language is cared about in Kenya and Tanzania. The problem of work and water are similarly cared about, while money and health problems are cared about in Africa a lot more than in Jamaica. Especially health problem is cared about in Tanzania. The difference between each topic is substantial. Salaam is a word used in Tanzania to show politeness.

	Kenya	Tanzania	Jamaica
like	1518	924	2985
order	670	413	402
laugh	41	22	1009
think	1321	618	2361

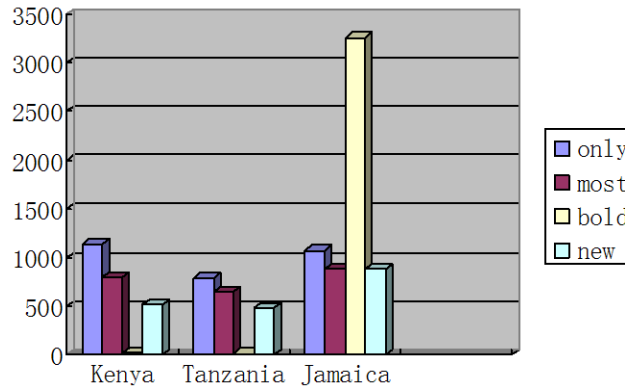


Some verbs are picked out to make comparisons. Every form of a certain verb is listed and the numbers are summed up in figure9. Obviously, the use of like is similar while the use of order in Kenya and Tanzania English are a lot more than in Jamaica English. The use of think is similar, while the use of laugh in Jamaica is a lot more than in Kenya and Tanzania English. It shows that comparatively people in South Africa are happier than people in Africa.

FIGURE11

	Kenya	Tanzania	Jamaica
only	1134	785	1070
most	800	641	876
bold	12	7	3260
new	514	485	881

FIGURE12

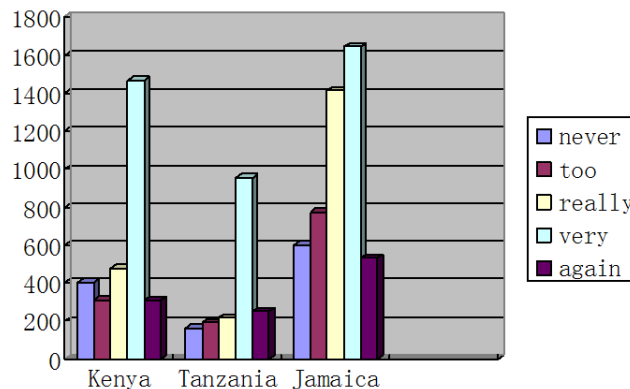


Some adjectives are picked out to make comparisons. It shows that only and most are used a lot more in Kenya and Tanzania English than in Jamaica English. Bold is quite special that used a lot in Jamaica English while not usually used in Africa. The difference in the usage of new is not substantial.

FIGURE13

	Kenya	Tanzania	Jamaica
never	399	162	597
too	310	190	770
really	475	216	1411
very	1462	956	1642
again	306	249	528

FIGURE14



Some adverbials are picked out to make comparisons. It shows that there are no substantial differences in the usage of “never, too and again”. Really is used a lot more in Jamaica English than in Kenya and Tanzania English, while very is used a lot more in Kenya and Tanzania English than in Jamaica English. That shows Africans like to use very to show stress while Jamaica people would like to use really to show it.

According to all the differences of the above, we can conclude that there are some similarities in the usage of words in Kenya and Tanzania English, while they are different from English of other places and sometimes the difference is substantial. So we can draw the conclusion that English varieties are somewhat influenced by their mother tongue and culture. That is consistent to the hypothesis.

IV. CONCLUSION

African English has its different characteristics and different country also has different features in the varieties. That may somewhat depend on the history, culture and daily life of local people.

It is obvious that corpus-based approach also has its limitations. In the future, studies should be carried out not only with corpus-based description but also with sociolinguistic and psycholinguistic data. There is still a wide scope for further research. Grammar, lexicology, phonetics and many other aspects should be paid attention to in further research.

ACKNOWLEDGEMENTS

I want to give my sincere thanks to Professor Nelson who has provided me with the free use of ICE and support.

REFERENCES

- [1] Gao Yihong & Xu Hongchen. (2014). University volunteers' attitudes towards global English before and after four great international activities. *Foreign Language Education, 1*, 43-48.
- [2] Hu Xiaoli. (2008). Current situation and prospect of Chinese English Varieties. *Foreign Languages in China, 2*, 27-30, 35.
- [3] Kachru, B. B. (1985). Institutionalized Second-Language Varieties. In S. Greenbaum & B. B. Kachru (eds), *The English Language Today*. Oxford: Pergamon, 211–226.
- [4] Lanham, L. (1967). *The Pronunciation of South African English*. Cape Town: Balkema.
- [5] Pan Zhangxian. (2002). Study of Chinese English varieties—review and expectation. *Foreign Languages Research, 6*, 24-27.
- [6] Todd, L. (1984). *Modern Englishes: Pidgins and Creoles*. Oxford: Blackwell.
- [7] Zou Hang. (2015). Study of Chinese English varieties under the perspective of biological evolutionism. *Foreign Language Education, 3*, 36-39.

Xiaohui Xu was born in Zibo, China in 1977. She received her M.A. degree in Foreign Linguistics and Applied Linguistics from Ocean University of China in 2010.

She is currently an associate professor in the School of Foreign Languages, Qingdao University of Science and Technology, Qingdao, China. Her research interests include Applied Linguistics and Language Policy.