Kenyan English idiomatic expressions:
They may sound frequent but that's not what corpus data show

Zipporah Otiso (Nairobi)

Abstract

Keen observers of Kenyan English usage will agree that idiomatic expressions such as *put into consideration,* rather than *take into consideration*, are certainly common in daily usage. The author of this paper set out to establish if that was indeed the case by having a sample of 122 respondents, all fourth-year university students, to choose between *put* and *take*. 77% of them chose *put*, which suggests that the expression *put into consideration* is indeed quite familiar to Kenyan English speakers. It was tested alongside another 19 idiomatic expressions. 17 out of the 20 were found to be familiar, though to varying degrees. But this familiarity was not reflected in corpus data from the *International Corpus of English* and the *Corpus of Global Web-based English*, where they were found to be rare. The same corpus data showed that this familiarity of the 20 expressions to Kenyan English speakers did not mean that they used them more frequently than their Standard International English counterparts. For example, the data showed that the "less familiar" *take into consideration* had more tokens in the Kenyan English components of the two corpora than the "more familiar" *put into consideration*. Nevertheless, the paper concludes that so-called Kenyan English idioms can still be claimed to be typical of Kenyan English, since they are practically absent from e. g. British English, its colonial ancestor.

1   Introduction

In this paper the phrase *idiomatic expressions* is used in a very wide sense to encompass both the traditional idioms and other fixed multi-word expressions. Regarding the former, there has been much discussion in the literature about the definition of the term *idiom* (cf. e. g. Nunberg/ Sag/Wasow 1994: 492–499; and Moon 1998: 2–5, for detailed definitions), going as far back in history as 1899 to Henry Sweet's definition, quoted by Skandera (2003: 42), which says that "the meaning of each idiom is an isolated fact which cannot be inferred from the meaning of which the idiom is made up". Skandera (2003) devoted his entire book on idioms in Kenyan English (hereafter KenE). The whole of the book's chapter 2 first reviews the different definitions of idiom and the criteria on which they were based, before offering the author's own definition, on which his analysis of idioms in KenE was based. It reads like this:

> An idiom is a conventionalized sequence of at least two words or free morphemes that is semantically restricted so that it functions as a single lexical unit, whose meaning—from a synchronic point of view—cannot or can only to a certain extent be deduced from the meanings of its constituents.
>
> (Skandera 2003: 60)

This definition can cater for semantically opaque idioms like *Don't count your chickens before they're hatched* and semi-transparent ones such as *bleed somebody dry*. But it cannot cater for the entirely transparent idioms like *spread like wildfire*, or even the multi-word expressions that are not figurative language at all, such as *take into consideration*. This paper discusses all three types of multi-word expressions, so long as they are structurally fixed and conventionalised, that is accepted as such by KenE speakers.

There are two broad categories of KenE idiomatic expressions (cf. Buregeya 2024): a) those that are structurally modelled on those used in Standard International English (hereafter StdIntE) but have altered the syntactic, morphological, or lexical composition of the latter; b) those that were coined through KenE usage, though using existing English words. An example of the former category is the verb phrase *add salt to injury* (for *add insult to injury* in StdIntE), while an example of the latter is the sentence *earth is hard* ('life can be difficult'). This paper will limit itself to the expressions in the first category because only they can enable a comparison between them and their StdIntE counterparts. The latter will be found in dictionaries traditionally associated with StdIntE, like the *Oxford English Dictionary* (and its smaller versions in the Oxford family of dictionaries), *Collins English Dictionary*, etc.

As a first stage, the comparison intended in this paper is aimed at establishing the extent to which the so-called KenE expressions are familiar to KenE speakers, compared with their StdIntE counterparts. It is a question of *extent* rather than that of *whether or* because it is not expected that the assumed KenE expressions have replaced the latter altogether. As Skandera (2003: 201f.) cautions us, we have to assume that both types of expressions coexist as variants in KenE usage. However, since the two authors who have written about idioms in KenE so far, namely Skandera (2000, 2003) and Buregeya (2007, 2019), argue that there are distinct KenE idioms, it can be equally assumed in this paper that the KenE expressions under study will be expected to be more familiar to KenE speakers; that is, these speakers should be able to recognise them as their preferred choice over the corresponding StdIntE expressions. At a second stage, the paper will relate this familiarity to the frequency of occurrence of select KenE expressions in corpus data with the aim of establishing whether there is a direct relationship between the two; that is, whether the assumed familiarity is reflected in a comparatively high presence of the same expressions in the two corpora that exist so far of KenE. The two are the Kenyan component of the *International Corpus of English* (ICE) –hereafter ICE-K – and the Kenyan component of the *Corpus of Global Web-based English* (GloWbE), hereafter GloWbE-KE.

2   Data and methodology

Buregeya (2024) discusses 72 "idioms and other fixed expressions" that have changed their structure in KenE. This paper is based on 20 of them chosen by the present author rather subjectively because no empirical indication exists so far of which KenE idioms are more frequent than which. The twenty under study are the following (with their StdIntE equivalents given in parentheses):

(i)   *Don't count your chicks before they hatch* (vs. *Don't count your chickens before they're hatched*)
(ii)   *master of ceremony* (vs. *master of ceremonies*)
(iii)   *to bear fruits* (vs. *to bear fruit*)
(iv)   *to put into consideration* (vs. *to take into consideration*)
(v)   *leave alone* (vs. *let alone*)
(vi)   *failure to which* (vs. *failure to do so/that*)
(vii)   *to sing to sb's tune* (vs. *to dance to sb's tune*)
(viii) *to pick sth/sb* – in the sense of 'to collect' or 'to lift up' (vs. *to pick up sb/sth*)
(ix)   *out of the blues* (vs. *out of the blue*)
(x)   *how comes* (vs. *how come*)
(xi)   *to add salt to injury* (vs. *to add insult to injury*)
(xii)   *to promise heaven on earth* (vs. *to promise the earth*)
(xiii) *human resource* (vs. *human resources*)
(xiv)   *to stay clear of* (e. g. *politics*) (vs. *to steer clear of*)
(xv)   *to milk sb dry* (vs. *to bleed sb dry*)
(xvi)   *to borrow a leaf from sb's book* (vs. *to take a leaf out of sb's book*)
(xvii) *to wreck havoc* (vs. *to wreak havoc*)
(xviii) *to play the devil's advocate* (vs. *to play devil's advocate*)
(xix)   *to make sb proud* (vs. *to do sb proud*)
(xx)   *women groups* (vs. *women's groups*).[1]

As can be seen from this list, the expressions concerned are of different complexity in length and grammatical structure. Regarding the latter, while some so-called KenE expressions differ from their StdIntE equivalents by a full word or even two words (e. g. *heaven on earth* vs. *the earth*), others are simply distinguished by a single bound morpheme, namely *-s*. But this *-s* is of much significance because the difference between *resource* and *resources* (in *human resource* vs. *human resources*) goes beyond the grammatical meaning of 'plural' denoted by *-s*: according to e. g. the *Collins English Dictionary Online*, the uncountable noun *resource* means 'capability, ingenuity, and initiative', as in *a person of resource and generosity*, which is totally different from the plural *resources*, whose meaning is the 'materials, money, and other things that they have and can use in order to function properly.' Similarly, the *-s* on *blues* (in *out of the blues*), which does not indicate plural, gives a semantically different meaning to *the blues*

---

[1] One may really wonder why the phrase *women groups* should be included in the list of idiomatic or fixed expressions. This is because the phrase was already pointed out in the very first publication called "Kenyan English", by Zuengler (1982: 116). Buregeya (2012: 470) comments that the phrase "is used in Kenyan English almost as if it were a set phrase", adding that "the phrase *women's groups* will be rarely found".

(whether it is 'a feeling of sadness' or even 'a type of music') from that of *the blue*, which means 'nowhere' and only in the idiomatic phrases *out of the blue* and *from the blue*.

To test whether the KenE expressions above were more familiar to KenE speakers, a gap-filling task was used which left to the respondent the option of choosing which word(s) to use to fill the gap, like this: *Congratulations. You have ____ our family proud.* Here either the verb form *made* was expected in the KenE version, or *done* was expected in the StdIntE one. This type of task was modelled on that used by Skandera (2003) as one of his sentence "completion tests" (ibid: 65) which he used to test his respondents' knowledge of specific idioms.

The familiarity of the 20 idiomatic expressions was tested on a sample of 122 respondents. These were all fourth-year university students at Kenyatta University and the University of Nairobi in Kenya. These are the two largest public universities in Kenya. As such, they admit students from all over the country, which means that these students represent a variety of ethnic backgrounds and, hence, of first languages (L1) of Kenya, as well as of socio-economic backgrounds. It was therefore assumed that factors such as L1, socio-economic status, and geography would not influence the choice of idiomatic expressions in the task given to the respondents. At the time of data collection, the respondents were all expected to be below 25 years of age. They were selected from two different classes: February 2020 (N: 72) and December 2022 (N: 50). Fourth-year university students were chosen to represent KenE speakers for two main reasons. Firstly, they had already been exposed to English long enough for their knowledge of English to qualify them as educated speakers of KenE.[2] (It is worth noting in passing that all the respondents were students training to become teachers of English and Literature.) Secondly, they can be claimed to accurately represent the stable, endogenous KenE that is generations-removed from the initial English spoken in Kenya which their forefathers were exposed to, which was still in significant contact with (mainly) British English as the language of the former colonial power. This sample of respondents was a primarily convenience one, to the extent that the researcher collected data from the two classes which were taught by lecturers who had offered to help collect the data. However, it had a random element to the extent that three different tasks testing three different aspects of English were distributed at the same time, and each student filled in only one of them. So, the task for the current study was completed by every third student in their seating arrangement.

In relation to the familiarity of the 20 expressions vis-à-vis their presence in corpora, *AntConc* (version 4.2.0) software was used to search for and analyse the relevant frequencies in ICE-K and ICE-GB (the British component of ICE). As for the frequencies of GloWbE-KE and GloWbE-UK (the British component of GloWbE), two important details need to be mentioned here: first, the "matching strings" were obtained by clicking on the "Chart" option on GloWbE's search interface, since this option gives both the raw figures and their normalised equivalents per million words. Second, where a verb was involved, to capture all its possible meanings the verb was put between square brackets, like this: [bear] fruits.

---

[2] According to the language policy in Kenya, English is the language of instruction from the fourth year of primary school, but it is already taught as a subject from the very first year, and in some places even at a pre-primary school level.

3   Results and discussion

Table 1 reports the results of how familiar the 20 idiomatic expressions were found to be by the respondents. Table 2 presents the frequencies of occurrence of the same expressions in both ICE-K and GloWbE-KE. For its part, Table 3 compares the normalised frequencies in GloWbE-KE and GloWbE-UK.

3.1   The assumed KenE idiomatic expressions are relatively quite familiar to KenE speakers.

This statement encapsulates the findings reported in Table 1. It should be pointed out from the outset that for all the 20 idiomatic expressions, more options were proposed by the respondents than just the two reported in the table. (This will explain why the percentages indicated in the table do not total 100.) However, only the KenE form and its StdIntE equivalent are reported in the table since it is the comparison between the two that is the focus of this paper. At items (1) and (18) the denominator is 50 because the two expressions concerned were not part of the initial task (for which N was 72), while at items (14) and (15) the denominator is 72 because the two expressions were not part of the second task (for which N was 50).

	Idiomatic expression	KenE word chosen for the gap	StdIntE word chosen for the gap
1.	*Congratulations! You have _____ our family proud.*	*made:* 49/50= 98%	*done:* 0/50= 0%
2.	*I'm sure your efforts will bear _____*	*fruits:* 106/122= 87%	*fruit:* 11/122= 9%
3.	*Be patient. Don't count your _____*	*chicks:* 92/122= 75%	*chickens:* 1/122= 1%
4.	*No wonder he is a master of _____ at weddings*	*ceremony:* 88/122= 72%	*ceremonies:* 28/122= 23%
5.	*The government needs to _____ that into consideration*	*put:* 94/122= 77%	*take:* 19/122= 16%
6.	*Kenya is already very far away from Europe, _____ alone the US.*	*leave:* 84/122= 69%	*let:* 30/122= 25%
7.	*I didn't expect it. It just came out of the _____.*	*blues:* 78/122= 64%	*blue:* 20/122= 16%
8.	*You will have to refund the loan, failure to ____, you will pay a higher interest on it.*	*which:* 72/122= 59%	*do that:* 7/122= 6%
9.	*He refused to _____ their tune.*	*sing:* 5/122: 4%	*dance to:* 81/122= 66%
10.	*The cake fell to the ground but the boy p____ and ate it.*	*picked it:* 34/122= 48%	*picked it up:* 0/72= 0%
11.	*How _____ you didn't notice it? It's so obvious.*	*comes:* 36/122= 30%	*come:* 66/122= 54%

	Idiomatic expression	KenE word chosen for the gap	StdIntE word chosen for the gap
12.	*They sacked him, and to add _____ to injury ...*	*salt:* 59/122= 48%	*insult:* 13/122= 11%
13.	*He had promised her ____earth.*	*heaven on:* 51/122= 42%	*the:* 33/122= 27%
14.	*Their company needs to invest in human _____.*	*resource* 29/72= 40%	*resources* 6/72= 8%
15.	*If men's groups are allowed, why not allow _____ as well?*	*women groups*: 29/72= 40%	*women's groups*: 29/72= 40%
16.	*They have been urged to ____ clear of politics.*	*stay:* 27/122= 22%	*steer:* 15/122= 12%
17.	*Some civil servants are intent on ____ the government dry.*	*milking:* 32/122= 26%	*bleeding:* 1/122= 1%
18.	*If you can't see how to do it, why don't you ____ a leaf___ my book?*	*borrow ... from:* 10/50= 20%	*take ... out of:* 0/50= 0%
19.	*If the rains continue like this, floods will ____ havoc in most of the country.*	*wreck:* 14/122= 11%	*wreak:* 1/122= 1%
20.	*Stop playing _____ advocate; we all know he attacked her.*	*the devil's:* 2/122= 2%	*devil's:* 1/122= 1%

Table 1: Respondents' choices between KenE words and their StdIntE counterparts

Table 1 shows that in 17 out of the 20 cases (that is 85%), the KenE idiomatic scored a higher percentage than its StdIntE counterpart. It is worth adding that in two cases the StdIntE version scored 0%: see No. 10: *picked it and ate it* vs. *picked it up and ate it*, and No. 18: *borrow a leaf from my book* vs. *take a leaf out of my book*. There are only 3 cases (15%) where, against the researcher's expectations, the KenE expression did not score a higher percentage than its StdIntE counterpart: No. 15: *women groups* vs. *women's groups*, where the score is a tie of 40% each, No. 8: *sing sb's tune* vs. *dance to sb's tune,* and No. 11: *how comes* vs. *how come*, where the KenE version scored a lower percentage.

Based on the figures above, it can be concluded that the vast majority of KenE idiomatic expressions are indeed more familiar to the KenE speakers than their StdIntE counterparts. This conclusion applies to both the expressions that seem to be quite familiar (that is those whose percentages are above 50% for the KenE forms) and those that do not seem to be familiar, e. g. Nos. 16 to 20, where the two options (KenE and StdIntE) total less than 40%, even when put together in each case. Clearly, an idiom like *play the devil's advocate* (vs. *play devil's advocate*), for example, seems to be extremely rare in KenE usage. (Most respondents simply said *stop playing the advocate* or *an advocate*, leaving out the word *devil*.) Concerning the expression *wreck* vs. *wreak havoc*, it is baffling how it does not seem to be familiar despite the fact that it will often be seen as a caption on TV news when there are floods in the country. (Most respondents used the verb *cause*, instead.) And although the KenE version *borrow a leaf* does not seem to be familiar, its StdIntE equivalent, *take out a leaf*, looks unheard-of in KenE. (Many respondents left the gap unfilled, while some filled it with simply *take a leaf from* or *pluck a leaf from*, making the sequence *out of* look *the* strange element in the entire phrase.)

3.2   The familiarity of assumed KenE idiomatic expressions is not reflected in corpus data

This statement sums up the findings in Table 2, which presents the frequencies of the 20 idiomatic expressions in ICE-K and GloWbE, both raw frequencies and their equivalents normalized per million words. It was deemed appropriate to use normalised frequencies because of the huge difference in size there is between the two corpora. Indeed, ICE-K, in both its spoken and written components combined, is only 791,695-words long (cf. Hudson-Ettle/Schmied 1999), while GloWbE-KE, which has only a written component, is 41,069,085 words long (cf. Davies 2013; and Davies/Fuchs 2015). However, although GloWbE lacks a component of spoken language, which is typically assumed to be more informal than written, it has a higher percentage of informal texts than ICE. Davies/Fuchs (2015: 4) state that "[a]bout 60 percent of the words for each country come from informal blogs, whereas the other 40 percent come from a wide variety of (often) more formal genres and text types." This makes GloWbE-KE a more likely source of the idiomatic expressions under study than ICE-K. This is because according to Nunberg/Sag/Wasow (1994: 493), one of the (six) properties of idioms is their "informality". They note that "[…] idioms are typically associated with relatively informal or colloquial registers and with popular speech and oral culture."[3]

	KenE expression (a) vs. StdIntE expression (b)	Their frequencies in ICE-K: raw and pmw in parentheses	Their frequencies in GloWbE-KE: raw and pmw in parentheses
1.	a) *pick sth/sb* b) *pick sth/sb up*	23 (29.05 pmw) 42 (53.05 pmw)	5,161 (125.69 pmw) 408 (9.94 pmw)
2.	a) *leave alone* b) *let alone*	7 (8.84 pmw) 9 (11.36 pmw)	121 (2.95 pmw) 519 (12.64 pmw)
3.	a) *women groups* b) *women's groups*	5 (6.31 pmw) 46 (58.10 pmw)	44 (1.07 pmw) 126 (3.07 pmw)
4.	a) *put into consideration* b) *take into consideration*	2 (2.52 pmw) 20 (25 pmw)	28 (0.68 pmw) 347 (8.45 pmw)
5.	a) *human resource* b) *human resources*	2 (2.52 pmw) 12 (15.15 pmw)	417 (10.16 pmw) 555 (13.52 pmw)
6.	a) *how comes* b) *how come*	2 (2.52 pmw) 14 (17.68 pmw)	59 (1.44 pmw) 221 (5.38 pmw)
7.	a) *borrow a leaf from* b) *take a leaf out of*	2 (2.52 pmw) 1 (1.26 pmw)	33 (0.80 pmw) 6 (0.15 pmw)
8.	a) *the devil's advocate* b) *devil's advocate*	2 (2.52 pmw) 1 (1.26 pmw)	6 (0.15 pmw) 15 (0.37 pmw)

---

[3] Furthermore, the spoken component of ICE-K does not seem to offer it an advantage over GloWbE-KE in relation to the presence or otherwise of idiomatic expressions because, as Hudson-Ettle/Schmied (1999: 6) acknowledge, there was difficulty in acquiring informal dialogue and certain other spoken categories in English" (ibid: 6) when the corpus was being designed.

	KenE expression (a) vs. StdIntE expression (b)	Their frequencies in ICE-K: raw and pmw in parentheses	Their frequencies in GloWbE-KE: raw and pmw in parentheses
9.	c) *wreck havoc* d) *wreak havoc*	1 (1.26 pmw) 2 (2.52 pmw)	20 (0.49 pmw) 61 (1.49 pmw)
10.	a) *of out the blues* b) *out of the blue*	1 (1.26 pmw) 1 (1.26 pmw)	15 (0.37 pmw) 71 (1.73 pmw)
11.	a) *add salt to injury* b) *add insult to injury*	0 (0.00 pmw) 3 (3.78 pmw)	5 (0.12 pmw) 31 (0.75 pmw)
12.	a) *bear fruits* b) *bear fruit*	0 (0.00 pmw) 3 (3.78 pmw)	34 (0.83 pmw) 128 (3.12 pmw)
13.	a) *don't count your chicks* b) *don't count your chickens*	0 (0.00 pmw) 0 (0.00 pmw)	4 (0.10 pmw) 0 (0.00 pmw)
14.	a) *failure to which* b) *failure to do that*	0 (0.00 pmw) 1 (1.26 pmw)	23 (0.56 pmw) 54 (1.32 pmw)
15.	a) *milk sb dry* b) *bleed sb dry*	0 (0.00 pmw) 0 (0.00 pmw)	5 (0.12 pmw) 2 (0.05 pmw)
16.	a) *make sb proud* b) *do sb proud*	0 (0.00 pmw) 0 (0.00 pmw)	93 (2.26 pmw) 29 (0.71 pmw)
17.	a) *master of ceremony* b) *master of ceremonies*	0 (0.00 pmw) 1 (1.26 pmw)	12 (0.29 pmw) 14 (0.34 pmw)
18.	a) *promise sb heaven on earth* b) *promise sb the earth*	0 (0.00 pmw) 0 (0.00 pmw)	1 (0.02 pmw) 2 (0.05 pmw)
19.	a) *sing sb's tune* b) *dance to sb's tune*	0 (0.00 pmw) 0 (0.00 pmw)	0 (0.00 pmw) 0 (0.00 pmw)
20.	a) *stay clear of* b) *steer clear of*	0 (0.00 pmw) 1 (1.26 pmw)	18 (0.44 pmw) 98 (2.39 pmw)

Table 2: Frequencies (raw and normalized) of the 20 idiomatic expressions in ICE-K and GloWbE-KE

Table 2 shows that 10 of the 20 expressions (i. e. half of them) do not appear even once in ICE-K: see Nos. 11 to 20. However, the same table also shows that some of their StdIntE equivalents, still in the bottom half of the table, do not appear even once either. This suggests some idioms are extremely rare, whether it is in non-native varieties (e. g. KenE) or native ones (e. g. BrE). In fact, speaking of idioms in native varieties of English, Simpson/Mendis (2003: 422) note that "a majority of them have frequencies in the range of 1 token or fewer per million words (Moon 1998)." And specifically in connection with "collocational analyses" that can be found in a small corpus, like ICE-K, Schmied (2004: 259) comments that the "standard 1-million word corpora can be too small […]." This should be even more relevant for ICE-K because it is less than 1 million words (just 791,695 words). Likewise, Davies' (2013) warns us about what not to expect in a corpus while searching for idioms: "Note […] how sensitive the frequency of idioms is to size. […] [I]n a tiny one million word corpus, there probably wouldn't be any

tokens [of specific idioms] at all". These quotations are quite reassuring, since they imply that it is not because a KenE idiom like *make sb proud* does not appear in ICE-Kenya at all that it is indeed equally absent from KenE usage. Anyone living in Kenya is likely to testify to hearing the expression used perhaps on a daily basis. The same applies to a phrase like *master of ceremony*, given that there are very frequent functions in Kenya where a master of ceremonies is needed, even though more often than not he/she will be referred to simply as MC. *Failure to which* is another frequent expression in KenE usage, and it is equally hard to imagine that it does not appear at all in ICE-K.

No less intriguing is the fact that, in an apparent contradiction of the findings reported in Table 1, KenE expressions which appear quite familiar to KenE speakers occur significantly less frequently than their StdIntE counterparts in corpus data. The three telling cases are *how comes* (2 occurrences) vs. *how come* (14 occurrences)*, human resource* (2 occurrences) vs. *human resources* (12 occurrences), and *women groups* (6 occurrences) vs. *women's groups* (46 occurrences). One explanation for these apparently counterintuitive findings is that the bulk of the corpus tokens may have been produced by a very small number of informants, which would mean that the StdIntE expressions in question are not as widely popular as the total number of their tokens would suggest. That seems to be the case of at least *women's groups*: out of its 44 occurrences in the sole written component of ICE-K, 30 (68%) appear in a single text called "Politics and Women's Groups" and produced by informant W2A018K; 6 more appear in another text (by W2A031K), and another 6 appear in a text by W2A036K, while informants W2C021K and W2C025K produced one each. This means that all in all, only 5 informants (out of the 200 who produced written texts, cf. Hudson-Ettle/Schmied 1999: 8) produced all the 44 tokens of *women's groups*.

Unfortunately, that explanation alone cannot be enough, because the figures in GloWbE-KE, which is by far a much bigger corpus (a whopping 41,069,085 words), though in some respects less representative (mainly because it lacks a spoken component, cf. Davies/Fuchs 2015: 26), also indicate that in the vast majority of cases the StdIntE expressions are still much higher than their KenE counterparts. It is only in two cases (namely those of *don't count your chicks …* and *borrow a leaf from …*) that the KenE versions are definitely more prevalent. This still points to the fact that while, on the one hand, there is a relatively high familiarity of the so-called KenE expressions to their users, on the other hand their frequency of occurrence in corpora is, in the majority of cases, lower than that of their StdIntE counterparts.

Nonetheless, as Buregeya (2024) remarks, the KenE expressions in question "can still be claimed to be typical of KenE if, by using e. g. frequencies normalized per million words in corpora, they can be shown to be more frequent in KenE than in other varieties of English." This is what the comparison in Table 3 is aimed at establishing. KenE will be compared with only BrE, its colonial ancestor. In this regard, Skandera (2003: 344) notes that "[a]s far as Kenyan English is concerned the use of idioms largely seems to follow British English usage patterns […]." The table compares data from the GloWbE corpus only, simply because there are no occurrences of any of the 20 structures in the British component of ICE (aka ICE-GB). A

key-word-in-context search using *AntConc* (version 4.2.0) produced 0 occurrences for each one of them.[4]

	Idiomatic expression	GloWbE-KE	GloWbE-UK
1.	*pick sth/sb* ('collect'; 'lift up')	125.69 pmw (5,161)	172.80 pmw (66,979)
2.	*leave alone*	2.95 pmw (121)	0.20 pmw (77)
3.	*women groups*	1.07 pmw (44)	0.02 pmw (6)
4.	*put into consideration*	0.68 pmw (28)	0.03 (13)
5.	*human resource*	10.16 pmw (417)	1.29 pmw (500)
6.	*how comes*	1.44 pmw (59)	0.15 pmw (60)
7.	*borrow a leaf from*	0.80 pmw (33)	0.01 pmw (4)
8.	*the devil's advocate*	0.15 pmw (6)	0.07 pmw (29)
9.	*wreck havoc*	0.49 pmw (20)	0.09 pmw (35)
10.	*of out the blues*	0.37 pmw (15)	0.03 pmw (13)
11.	*add salt to injury*	0.12 pmw (5)	0.00 pmw (1)
12.	*bear fruits*	0.83 pmw (34)	0.03 pmw (11)
13.	*don't count your chicks*	0.09 pmw (4)	0.00 pmw (0)
14.	*failure to which*	0.56 pmw (23)	0.00 pmw (1)
15.	*milk sb dry*	0.12 pmw (5)	0.03 pmw (10)
16.	*make sb proud*	2.26 pmw (93)	1.00 pmw (10)
17.	*master of ceremony*	0.29 pmw (12)	0.02 pmw (9)
18.	*promise sb heaven and earth*	0.02 pmw (1)	0.00 pmw (0)
19.	*sing sb's tune*	0.00 (0)	0.00 pmw (0)
20.	*stay clear of*	0.44 pmw (18)	0.37 pmw (144)

Table 3: Normalised frequencies for the 20 KenE idiomatic expressions in GloWbE-KE and GloWbE-GB with (raw figures in parentheses)

The normalised frequencies in Table 3 unequivocally show that the assumed KenE expressions are overwhelmingly more frequent in KenE than in BrE, and, thus, can be genuinely claimed to be typical of KenE.[5] The only exception seems to be the use of *pick sb/sth* instead of *pick sb/sth up*. This must be just an apparent exception arising from the fact that a large number of occurrences of the phrase must contain irrelevant cases, such as the uses of *pick* in its other many meanings of 'choose' (e. g. *pick a leader*), 'remove' (e. g. *pick one's nose*), etc. It became impractical (for the author) during the GloWbE search to separate the different uses. It is instructive to learn, though, that a key-word-in-context search of *pick sb/sth* in ICE-GB using *Ant-Conc* shows that the phrase was not used a single time in lieu of *pick sb/sth up*. So, all in

---

[4] Many of their StdIntE counterparts do not occur in the ICE-GB, either, including expressions like *women's groups* and *human resources*; but some do, like *let alone*, with 6 occurrences, and *take into consideration*, with just 1.

[5] Or at least in "African English", for some of them! This is because a GloWbE search shows that some of the so-called KenE expressions are almost as equally frequent (or if not more) in some other African varieties of English as in KenE. That is the case of *borrow a leaf from*, which appears at a per-million-word frequency of 0.94 in Nigerian English, against 0.80 in KenE.

all, the only genuine case where an assumed KenE expression is almost equally present in BrE is that of *stay clear of sth*, with a 0.44-pmw frequency against a 0.37-pmw one.

4   Conclusion

This paper set out first to establish the familiarity of twenty select idiomatic expressions assumed to be typical of KenE and then to relate this familiarity to the presence of the same expressions in corpus data. In summary, the paper has made the following observations: first, the assumed KenE idiomatic expressions are by and large familiar to the KenE speakers, that is familiar to at least the young generation used as respondents in the study. Second, on the whole the same expressions are, rather surprisingly, less frequent in corpus data than their StdIntE counterparts. Third, despite all that, they can still be argued to be typical of KenE, at least if reference is made to BrE, the colonial ancestor of KenE. Fourth, based on the corpora consulted, and beyond just KenE, in daily language we seem to notice idiomatic expressions rather easily because of their expressive nature, but they seem to be too infrequent in the language to be captured in any significant numbers (say of at least 1 token per million words) in a corpus, however large it is.

Studying the few tokens of those expressions that are available in existing corpora is still worthwhile, though: firstly, as hinted at in the preceding paragraph, those few tokens are still useful in distinguishing between dialects of English. For example, going by the figures in GloWbE, the expression *failure to which* appears significantly only in KenE: in 23 out of all the 43 tokens (i. e. 53%) from all the 20 varieties represented in the GloWbE corpus. (The second highest number of tokens, 9, that is 21%, is for Tanzanian English. Secondly, they can allow us to test Mair's (2013) notion of "epicentrality" in his "World System Model of World Englishes", which, if related to this specific example, seems to suggest that the use of *failure to which* is due to geographical proximity with a more powerful (if anything in terms of English usage) Kenya.

References

Buregeya, Alfred (2007): "Aspects of the vocabulary of Kenyan English: An overview". *Occasional Papers in Language and Linguistics* 3: 1–32.

Buregeya, Alfred (2012): "Kenyan English". In: Kortmann, Bernd/Lunkenheimer, Kerstin (eds.): *The Mouton world atlas of varieties of English*. Berlin, de Gruyter Mouton: 466–474.

Buregeya, Alfred (2019): *Kenyan English*. Berlin: de Gruyter Mouton.

Buregeya, Alfred (2024): "Adapting and creating English idioms and other fixed expressions in the Kenyan socio-cultural context". Submitted for publication, to be published in: Buregeya, Alfred et al (eds.): *Pragmatics of African Varieties of English*. Berlin: de Gruyter Mouton.

Collins English Dictionary. collinsdictionary.com/dictionary/english/dictionary [10.06.2024].

Corpus of Global Web-based English. english-corpora.org/glowbe/ [10.06.2024].

Davies, Mark (2013): "Introducing the 1.9 Billion Word Global Web-Based English Corpus (GloWbE)". *The 21st Century Text*. 21centurytext.wordpress.com/introducing-the-1-9-billion-word-global-web-based-english-corpus-glowbe/ [09.04.2024]

Davies, Mark/Fuchs, Robert (2015): "Expanding horizons in the study of World Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE)". *English World-Wide* 36/1: 1–28.

Hudson-Ettle, Diana M./Schmied, Josef (1999): *Manual to accompany the East African component of the International Corpus of English.* Chemnitz: Chemnitz University of Technology.

International Corpus of English (ICE). ice-corpora.uzh.ch/en.html [10.06.2024].

Mair, Christian (2013): "The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars". *English World-Wide* 34/3: 253–278.

Moon, Rosamund (1998): *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.

Nunberg, Geoffrey/Sag, Ivan A./Wasow, Thomas (1994): "Idioms". *Language* 70.3 (1994): 491–538.

Oxford English Dictionary. oed.com/?tl=true [10.06.2024].

Schmied, Josef (2004): "Cultural discourse in the Corpus of East African English and beyond: Possibilities and problems of lexical and collocational research in one million-word corpus". *World Englishes* 23/2: 251–260.

Simpson, Rita/Mendis, Dushyanthi (2003): "A corpus-based study of idioms in academic speech". TESOL *Quarterly* 37: 419–441.

Skandera, Paul (2000): "Research into idioms and the International Corpus of English". In: Mair, Christian/Hundt, Marianne (eds.): *Corpus Linguistics and Linguistic Theory*. Papers from the Twentieth International Conference on English Language Research on Computerized Corpora (ICAME 20). Amsterdam, Rodopi: 239–353.

Skandera, Paul (2003): *Drawing a Map of Africa: Idiom in Kenyan English*. Tübingen: Narr.

Zuengler, Jane E. (1982): "Kenyan English". In: Kachru, Braj B. (ed.): *The Other Tongue: English across Cultures.* Urbana, University of Illinois Press: 112–124.