Formality as a Feature of Postcolonial English in Kenya: a Contrastive Analysis Based on the *International Corpus of English*

Natalia Budohoska, ***PhD student***
Institute of English Studies, University of Warsaw, Poland
n.budohoska@uw.edu.pl

Abstract. In monolingual countries where English is the native language of the majority of population, it is used across all domains and registers, and by all members of society. In post-colonial, multilingual countries it comes to be used as a lingua franca in more restricted contexts. Because there is a significant difference between the role and use of English in a native and monolingual environment and a non-native, multilingual environment (Trudgill 1999) the following paper concentrates on differences in the formality level of English in postcolonial Kenya and Great Britain on the basis of parameters set up for this study. The data for this paper come from the *International Corpus of English* for East Africa and for Great Britain. The conclusions concerning the formality of English were drawn basing on a quantitative study employing the chi square test for evaluating significance of the features discussed and revealed a higher level of formality of English in the ICE-K. This study is an introduction to further qualitative research of characteristic morpho-syntactic features of English in Kenya.

Keywords: varieties of English, language contact, lingua franca, sociolinguistics.

1. Introduction

In a non-native language environment it is not sufficient to find large numbers of users of English in a wide range of socially significant domains on both the national and international levels to claim that a new variety of English is emerging. According to Platt, Weber & Mian (1984: 2-3), there are four essential elements which need fulfilling for a new variety to emerge and English in Kenya so far unquestionably fulfils three of them.

First of all, English in Kenya functions mostly in a non-native context, as the native English speakers are far outnumbered by second language speakers. Moreover, it is taught through the system of education, initially as an obligatory subject and later as the main medium of instruction. Thirdly, it serves the role of a *lingua franca* for people not having a language in common. The fourth condition allowing new varieties to be

established entails that a language should become localized and develop certain distinct linguistic features of its own. The extent to which this last element is true and how English in Kenya has been "(…) modeled, reshaped, acculturated and redesigned and – by doing so – enriched what was a Western medium (…)" (Kachru 2006: 338) remains to be extensively studied.

The work so far does provide strong assumptions that a new variety of English is developing and encourages further analysis to reliably answer the claim whether English in Kenya is in fact Kenyan English. So far the most significant contributors to the discussion about Kenyan English, though not dealing with the aspect of formality, have been Hancok & Angogo (1982), Zuengler (1982), Abdulaziz (1991), Schmied (1991) and Skandera (2003).

English in Kenya functions predominantly as a language of the media, education and professional life, i.e. domains which require a formal use of language. Family life and socializing is often done in Kiswahili or tribal languages, therefore English in Kenya is rarely used to express emotions or solidarity and this sociolinguistic situation is a prerequisite to possible differences in the formality level between British English and English in Kenya.

In many areas of the world, switching from informal to formal situations also involves switching from one language to another. In such cases, it is probable that neither of the two languages involved will have the full range of styles available to the speakers in monolingual situations. (Trudgill 1999: 119)

The aim of the present paper will be to provide a contrastive, quantitative data analysis of several features of formal language (phrases expressing personal opinion, contracted versus full forms, linking expressions and intensifiers) and to uncover statistically significant differences between the numbers of tokens extracted from the *International Corpus of English*.

2. The International Corpus of English (ICE)

The International Corpus of English (ICE) is an ongoing project aiming at collecting English language samples from various countries around the world where English functions as a first or second language. The goal of such a collection of subcorpora is to enable making comparisons between different varieties of English on an international level.

As this study concentrates on the analysis of English in Kenya, the Kenyan component (henceforth referred to as ICE-K) was extracted from the East African subcorpus of the ICE. As a reference for the English language in Kenya the British subcorpus of the ICE was consulted (henceforth referred to as ICE-GB). These two corpora were collected along similar principles and contain both written and spoken samples of parallel texts of similar sizes.

However, despite attempts to provide mirror reflections of samples of these two varieties of English, it turned out impossible to include identical numbers of the samples of identical text categories. Some spoken text categories were unobtainable at the time of the project or different word counts were obtained, sometimes additional written or spoken text categories, which do not have their counterparts in ICE-GB were added to the ICE-K or vice versa in order to fill this gap. After a careful investigation of the types of texts and the numbers of words in both of the corpora used in this study, slight modifications of the contents of the corpora diminished the imbalance in numbers to a reasonable level. Most importantly, due to the lack of corresponding categories in the ICE-K the whole section labelled as "written unscripted" texts in the spoken element of the ICE-GB, was omitted in further analysis.

		ICE-K	**ICE-GB**
Spoken	Dialogue - Private - Public	264,584	376,689
	Monologue	125,248	108,164
Subtotal		**389,832**	**484,853**

Written	Non-Printed - Correspondence - Non-professional writing - Legal presentations	101,003	104,105
	Printed - Academic writing - Creative writing - Instructional writing - Non-academic writing - Persuasive writing - Informational	320,952	319,476
Subtotal		**421,955**	**423,581**
Total		**811,787**	**908,434**

Table 1.

Contents of the ICE-K and the ICE-GB

Table 1 presents the contents and the number of words of the two corpora used for this analysis. In the majority of cases the labels used in this table are taken from the ICE-K since the corpora present a very similar content despite different labelling at times. Labels from ICE-GB were used only in the case of lack of an equivalent in the ICE-K.

3. Selection of data

According to Tognini-Bonelli (2001: 65) two types of corpus research need to be distinguished, i.e. corpus-based and corpus-driven. The former uses the corpus to verify previously made assumptions and as a source of examples. In the latter method the

researcher draws conclusions reflecting directly the data in the corpus. However, according to McEnery, Xiao & Tono "(…) the distinction between the corpus-based vs. corpus-driven is overstated (…) and idealized (…)" (2006: 8), as the first method might lead to disregarding inconvenient data and in the second method it is impossible to approach data in a completely naïve way. In order to provide reliable conclusions this study loosely applies the corpus-driven method as the point of departure and moves on to a corpus-based analysis.

As a starting point for the quantitative study of variation between English in Kenya and the British Standard the search tool Word List in the Ant.Conc search software was run on the data of the ICE-K. A list of 100 most common lexical items was devised together with the numbers of occurrences. Additionally the number of tokens was counted for the same items in the ICE-GB and the chi square test for statistical significance was run on this data.

After analysing the list of 100 most frequently appearing words in the ICE-K and comparing the frequency of the same items in the ICE-GB it was noted that several items reflect a statistically significant difference in the number of occurrences. Out of all these a list of related parameters was selected as the main point of further frequency overview.

	ICE-K	**ICE-GB**	**Significance**
Total words	811,787	908,434	
I	10,006	15,980	$p<0.05$
not	5,407	4,293	$p<0.05$
't	1,559	5,820	$p<0.05$
because	1,756	1,671	$p<0.05$
very	1,474	2,050	$p<0.05$

Table 2.

Statistically significant items selected for further analysis from the list of 100 most frequently appearing words in the ICE-K

Table 2 presents a list of items which were crucial in inspiring the direction of this study together with the number of occurrences in both corpora and also showing the level of statistical significance. Revealing a statistically significant difference in frequency, all these items were used as the basis for the four parameters referring to the level of

formality of the language samples in the ICE-K and ICE-GB.

The use of the personal pronoun *I* revealed significant differences in frequency and led to the assumption that there might be a difference in frequency of phrases used to express personal opinion between the two corpora, which is a parameter indicating the level of formality, since formal language is "impersonal" (Quirk et al. 1985: 26). A high level of subjective statements indicates a lower level of formality and a list of expressions of personal opinion was devised and tested for statistically significant differences in terms of formality of language between the ICE-K and ICE-GB.

A statistically significant difference was also observed in the frequency of *because* and the parameter of linking expressions was set up as a further indicator of the formality level of language. A higher frequency of linking expressions renders more organized utterances which reflect a higher level of formality (Quirk et al. 1985: 1535; Foley & Hall 2003: 338).

The use of contractions versus full forms is also a feature which is closely related to the formality of language and from the list of 100 most common words the contraction *'t* versus *not* inspired a parameter of verbs which form their negative with the adverb *not*. A high frequency of contractions lowers the level of formality (Quirk et al. 1985: 123; Swann 1995: 144, 216).

Finally the adverb *very* showed a statistically significant difference in frequency between the ICE-K and the ICE-GB, which has led to establishing the last parameter indicative of the level of formality of language for this study, i.e. the frequency of the use of intensifiers. An extended use of intensifiers is a feature of an informal character of the utterances (Quirk et al. 1985, p. 1510). Because an integral aspect of formality lies in the difference between spoken and written language, the context of use of the above parameters according to this distinction will also be taken into account.

4. Research methods

The data for this paper come from a frequency overview of the five following parameters: common phrases used to express personal opinion, contracted forms of verbs which form their negative counterparts by adding the adverb *not* versus full forms, common linking words, and popular intensifiers.

The pattern of analysis for the parameters under investigation begins with compiling lists of the most common lexical items representing each parameter which serve as the starting points for the frequency overview. The lists are based on the frequency of

appearance in the corpora and consulted with various sources including grammar books and websites describing English grammar and are arranged in tables in alphabetical order. Absolute occurrences for all these words were counted using the search tools assigned to ICE-K and ICE-GB. Unfortunately, due to differences in the construction of the corpora which caused technical limitations the same search software could not be used for both corpora. For the ICE-K, the Ant.Conc.3.2.m search software (available online) used was the search tool labelled *Concordance* and for the ICE-GB it was the ICE-Cup3 search software (concurrent with the ICE-GB) employing the search tool labelled *Text Fragment*. These both search tools worked according to the same principles and the results were used to make comparisons between the two corpora. All searches revealed the number of tokens together with the contexts allowing verification of the lexical items under investigation. Apart from providing quantitative data the concordance analyses permit exploring each token individually and in detail, rendering general claims about the language under scrutiny more reliable. The frequency numbers for lexical items in each parameter are later presented in four different ways, revealing further quantitative differences between the two corpora and allowing to draw tentative conclusions regarding the formality level of the Kenyan and British English based on the parameters analysed.

Firstly, the chi square test (Oakes 1998: 24-29) was run on the data, revealing statistically significant differences in the numbers of occurrence of the five parameters indicative of the level of formality set up for this study. Since the chi square test is unreliable when run on very small numbers, the minimal number of tokens to run such a test was 10 in at least one corpus. The value $p<0.05$ was established as the cut off point for statistically significant differences in results (McEnery & Wilson 2001, p. 85). After running the chi square test, results for the value $p<0.05$, i.e. statistically significant and of further interest in this study, were marked with an asterisk. One asterisk indicates a statistically significant higher frequency of a token for the ICE-GB, while two asterisks indicate a statistically significant higher frequency for the ICE-K. The total number of tokens is also given for each parameter and the chi square test is applied as well to reveal the statistical significance of the whole parameter.

Secondly, because the two corpora differ in size by approximately 90,000 words, direct comparisons between the numbers of occurrences should not be made. Frequency per one million words is therefore counted for all the statistically significant lexical items under investigation and presented on a bar chart.

Finally, numbers of tokens within each parameter in the written and spoken components of the corpora are presented.

5. Frequency overview

5.1. Phrases expressing personal opinion

Formal language and especially written language tends to be objective in expressing ideas and tends to avoid using the first person singular as the agent in a sentence presenting an action. It can therefore be assumed that the popularity of phrases expressing personal opinion may be indicative of the formality level in the two corpora. These phrases were selected on the basis of a list of words most frequently occurring in the ICE-K, devised using Ant.Conc search software.

	ICE-K	**ICE-GB**	**Significance**
as for me	5	3	—
I believe	36	100	p<0.05*
I feel	63	114	p<0.05*
I guess	20	44	p<0.05*
I know	191	381	p<0.05*
I mean	238	1,424	p<0.05*
I presume	1	14	p<0.05*
I promise	18	2	p<0.05**
I remember	48	84	p<0.05*
I say	133	260	p<0.05*
I suppose	15	196	p<0.05*
I suspect	4	16	p<0.05*
I think	729	1,753	p<0.05*
I wonder	22	56	p<0.05*
I'd say	4	14	p<0.05*
in my opinion	2	6	—
to my mind	2	2	—
Total	1,531	4,469	p<0,05*

Table 3.

Expressions of personal opinion: frequency overview in both corpora

Table 3 contains a list of common phrases used to express personal opinion, which is the first parameter under investigation indicating the level of formality of language. The numbers of occurrence for each phrase were given for the total number of tokens appearing in the present and past tenses, including the third person singular forms. Out of the 17 most common phrases found in the ICE-K and ICE-GB 3 did not reach the minimum of 10 tokens in at least one of the corpora and were disregarded in further analysis. Out of the remaining 14 phrases after running the chi square test, all reflected a statistically significant difference in the frequency levels. In all but one cases the phrases were more numerous in the ICE-GB and only the phrase *I promise*, was found to be more frequent in the ICE-K.

The total number of phrases expressing personal opinion was statistically significantly higher in the ICE-GB than the ICE-K according to the chi square test. Also the overall number of the personal pronoun *I* reflected a statistically significant higher frequency in the ICE-GB than the ICE-K.

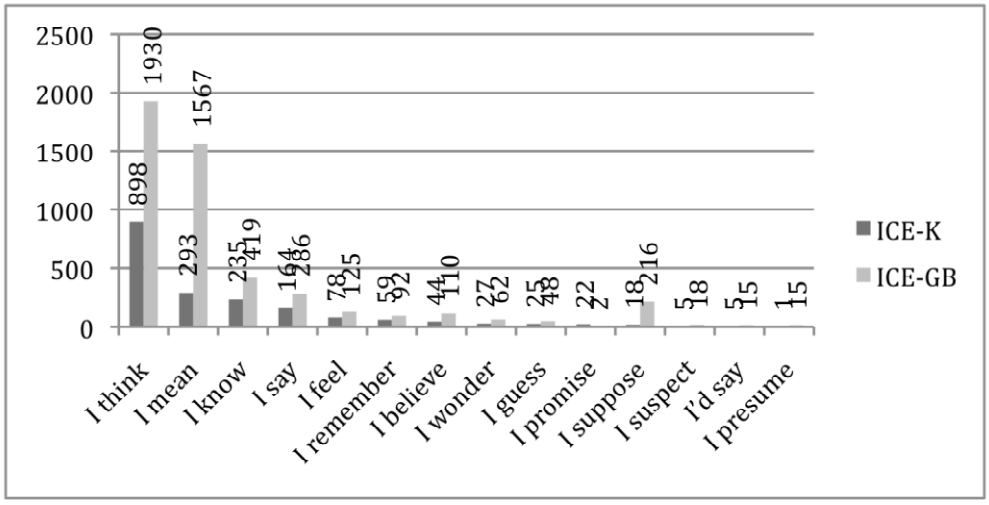

Figure 1.

Expressions of personal opinion: comparative frequency per one million words

Figure 1 presents the number of tokens per one million words on a bar chart illustrating each phrase in both corpora. As can be seen the pattern of frequency level is similar in both of the corpora with the highest difference for the two most frequent phrases, i.e. *I think* and *I mean*, although clearly the numbers are

consistently higher for the ICE-GB. In the case of ICE-K the most frequent phrase *I think* is 3 times as frequent than the following phrase *I mean*. In ICE-GB the difference between the first two phrases is much lower.

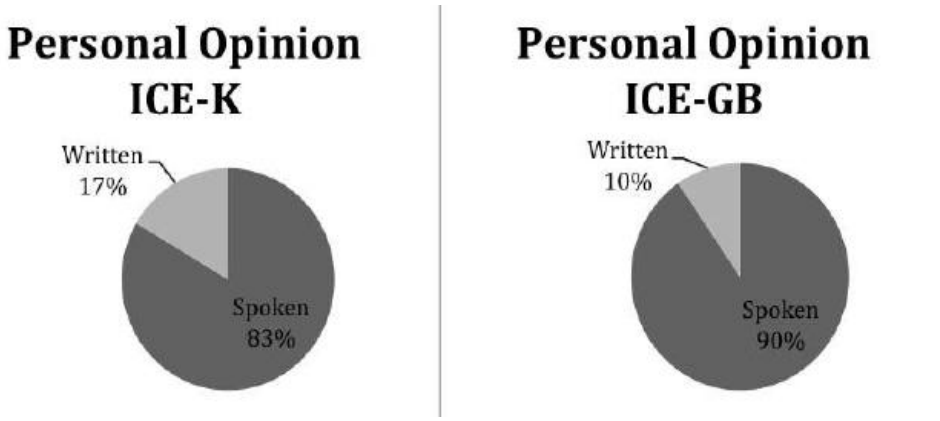

Figure 2.

Expressions of personal opinion: distribution

in spoken versus written ICE components

Despite a statistically significant difference of 3000 occurrences in the total number of tokens for this parameter the distribution between the written and spoken components of the ICE corpora is very similar. As can be deduced from Figure 2, expressions of personal opinion are overwhelmingly more popular in spoken language in Kenya (83%) and Great Britain (90%) confirming that they might be connected with a lower level of formality of language. An overall lower number of tokens representing the parameter of phrases expressing personal opinion in the ICE-K might be indicative of an overall higher formality level of English in Kenya.

5.2. Contracted forms

The use of contracted forms as opposed to full forms may be an indicator of the level of formality of language. Contractions are most commonly associated with the spoken, less formal use of language, while full forms with a written and more formal use of language. For the purpose of this study nouns which form their negatives by adding the adverb *not* were compared in terms of contracted and full forms.

	ICE-K	ICE-GB	Significance		ICE-K	ICE-GB	Significance
are not	422	179	p<0.05**	aren't	10	140	p<0.05*
cannot/ can not	482	176	p<0.05**	can't	173	548	p<0.05*
could not	161	69	p<0.05**	couldn't	38	182	p<0.05*
did not	541	133	p<0.05**	didn't	141	595	p<0.05*
do not	430	185	p<0.05**	don't	788	379	p<0.05**
does not	280	126	p<0.05**	doesn't	108	379	p<0.05*
had not	103	50	p<0.05**	hadn't	11	75	p<0.05*
has not	110	53	p<0.05**	hasn't	11	88	p<0.05*
have not	118	53	p<0.05**	haven't	20	308	p<0.05*
is not	655	385	p<0.05**	isn't	74	421	p<0.05*
shall not	21	12	—	shan't	0	2	—
should not	145	55	p<0.05**	shouldn't	14	52	p<0.05*
was not	323	175	p<0.05**	wasn't	34	346	p<0.05*
were not	153	53	p<0.05**	weren't	3	76	p<0.05*
will not	114	107	p>0.05	won't	43	168	p<0.05*
would not	133	81	p<0.05**	wouldn't	55	280	p<0.05*
Total	4,191	1,892	p<0.05**		1,523	4,039	p<0.05*

Table 4.

The *not* contractions vs. full forms: a frequency overview in both corpora

Table 4 contains a list of verbs which form their negative form by adding *not* in both full and contracted forms together with the number of occurrences in the ICE-K and the ICE-GB. Out of the 16 negative phrases only *shan't* did not fulfill the criterion of the minimum of 10 occurrences in at least one of the corpora and together with the full form

shall not was disregarded in further analysis. The chi square test was therefore run on the remaining 15 negative phrases in order to reveal statistical differences in the number of occurrences between contracted and full forms in both of the corpora under investigation. All of the contracted forms revealed a statistically significant difference and in the case of full forms only *will not* failed to do so. The negative phrase for the verb *will* was therefore also disregarded in further analysis.

All of the full forms were more commonly used in the ICE-K, and as for the contracted forms all except *don't* were more common in the ICE-GB. The overall statistically significant frequency proved to be almost entirely evenly distributed with full forms dominating in the ICE-K and contracted forms in the ICE-GB.

Finally, the chi square test was run on the total numbers of full and contracted forms in both of the corpora, thus further confirming the statistically significant difference. It should also be emphasised that the total numbers of occurrences of full and contracted forms of verbs forming their negatives with *not*, between the two corpora were very close although inversely proportional. The full forms in the ICE-K corresponded with the number of contracted forms in the ICE-GB and the other way round; the number of contracted forms in the ICE-K corresponded with the number of full forms in the ICE-GB.

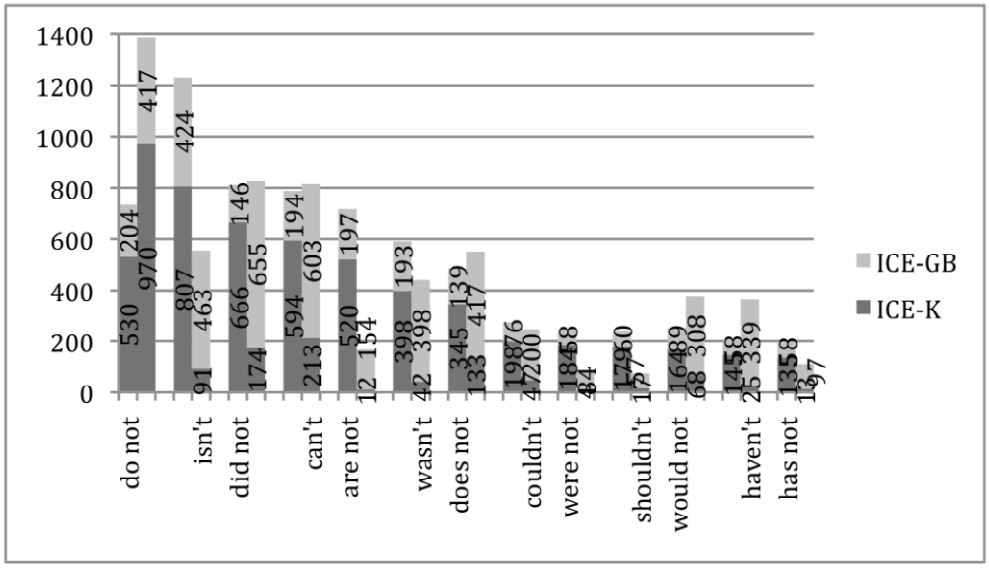

Figure 3.

Not contractions vs. full forms: a comparative frequency per one million words

The bar charts in Figure 3 represents the number of tokens for each negative phrase in full and contracted forms per one million words. The results are presented according to the highest frequency of a feature in the ICE-K. As regards the verbs *did*, *can*, *could* and *does* used in negative phrases the pattern of frequency is very similar in numbers illustrating the proportional difference between full and contracted forms in the two corpora. The remaining verbs exhibit no such clear reflections of numbers between the two corpora, although they continue to confirm the trend of full forms being more commonly used in the ICE-K and contracted forms in the ICE-GB.

The only exception to the pattern is in the ICE-K as regards the verb *do*, which in the ICE-K is not only more common in the contracted form than in the full form, but also has the highest overall frequency. In the ICE-GB, however, the pattern is preserved in the case of the verb *do* and the contracted form in more common than the full form as regards all the other verbs forming their negative inserting the adverb *not*.

In order to further illustrate the relationship between the full and contracted forms of the negative phrases under scrutiny, the number of tokens reflecting the contracted form was counted per one occurrence of the full form in each of the corpora.

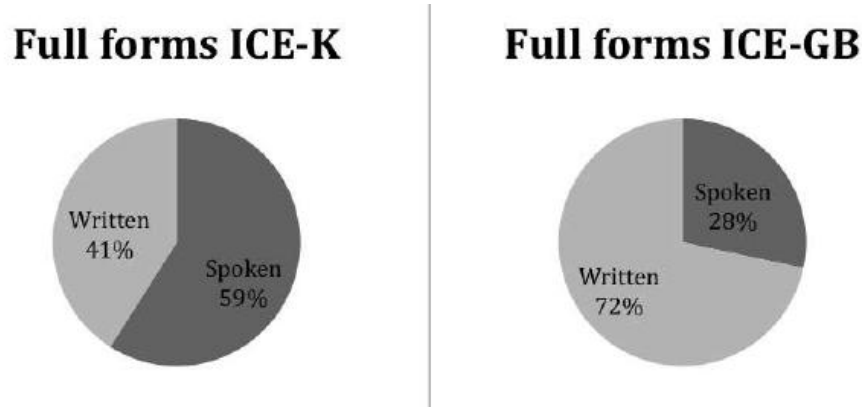

Figure 4.

Not contractions: distribution in spoken versus written ICE components

The difference in the number of tokens representing full forms of verbs which form their negatives with *not* was statistically significant and reaching around 2300 occurrences more in the ICE-K. What is interesting, however, is the distribution of

these items within the written and spoken components of the corpora as shown in Figure 4. In the ICE-K full forms are more popular within the spoken component reaching a result of 59%, while in the ICE-GB over 70% of full forms appear in the written component. In the case of the ICE-GB, therefore the results for full and contracted forms reflect the distinction of formality levels. In the ICE-K full forms are overall more popular and might be indicative of an overall higher level of formality of language.

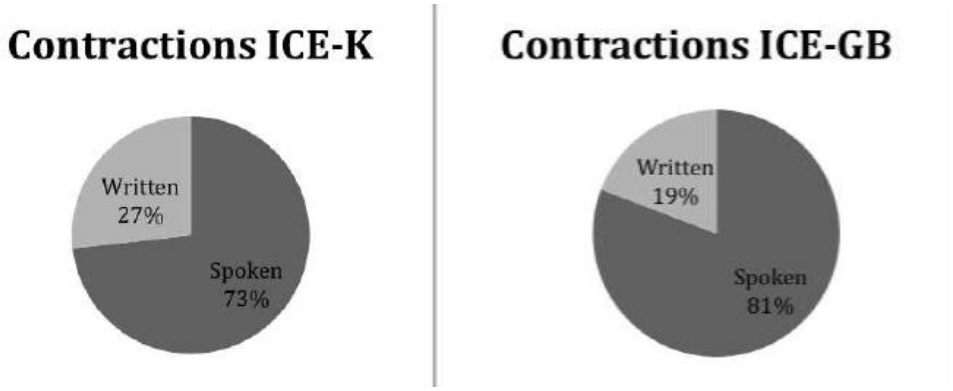

Figure 5.

Not full forms: distribution in spoken versus written ICE components

The difference between the total numbers of tokens for contracted forms of verbs which form their negatives with *not* was statistically significant and reaching around 3500 occurrences more in the ICE-GB. Despite this fact the distribution between the written and spoken components of the ICE corpora is similar with a majority of contracted forms appearing in spoken language a reflected in Figure 5.

5.3. Linking expressions

Because linking words are connected with organizing an utterance in a more formal way an assumption is made here that the overall number of linking words may indicate the level of formality of language used in each of the two corpora. Linking words most commonly appearing in the Word List of the ICE-K were selected as representative of this parameter.

	ICE-K	**ICE-GB**	**Significance**
although	229	342	p<0.05 *
despite	99	82	p<0.05 **
finally	85	95	p>0.05
firstly/first of all	53	73	p>0.05
for example	469	243	p<0.05 **
for instance	103	89	p>0.05
furthermore	24	25	p>0.05
however	436	545	p>0.05
in conclusion	3	3	—
in spite of	11	14	p>0.05
moreover	22	23	p>0.05
nevertheless	16	39	p<0.05 *
on the other hand	59	44	p<0.05 **
therefore	470	268	p<0.05 **
to sum up	2	3	—
though	145	439	p<0.05 *
whereas	35	98	p<0.05 *
Total	2,261	2,425	p<0.05 *

Table 5.

Linking expressions: a frequency overview in both corpora

Table 5 contains a list of basic linking expressions related to exemplifying, adding, contrasting and summarizing information, with the number of occurrences in each of the two corpora. On the basis of that the statistical significance in level of frequency is counted using the chi square test.

Using the corresponding search tools for the corpora it was found that out of the 18 linking expressions two, i.e. *in conclusion* and *to sum up,* did not fulfill the requirement

of a minimum of 10 tokens in at least one of the corpora, and were disregarded in further analysis.

After running the chi square test, out of the remaining 16 linking expressions eight showed a statistically significant difference on the frequency level, 4 linking expressions, i.e. *although*, *nevertheless*, *though* and *whereas* were more common in the ICE-GB and 4, i.e. *despite, for example, on the other hand* and *therefore*, were more common in the ICE-K.

The results of the chi square test in the total number of tokens in the parameter of linking expressions revealed a statistically significant difference on the level of frequency, with linking expressions being more frequently used in the ICE-GB.

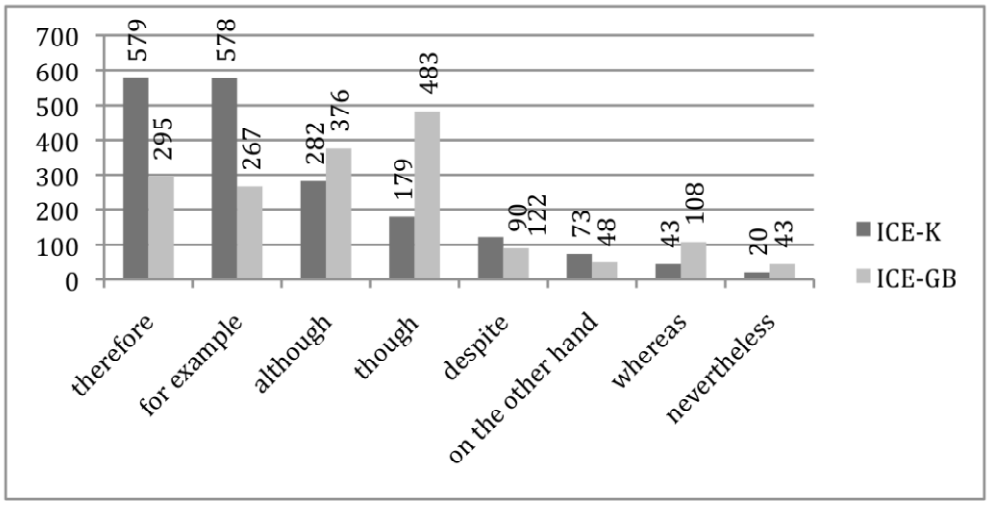

Figure 6.

Linking expressions: a comparative frequency per one million words

Figure 6 illustrates the number of occurrences of statistically significant linking expressions per one million words in both of the corpora presented from the highest to the lowest as reflected in the ICE-K. As can immediately be seen the difference in the number of occurrences between the two corpora is most visible in the case of *therefore*, *for example* and *though*. Also the order of the four most frequent linking expressions is different in the ICE-GB.

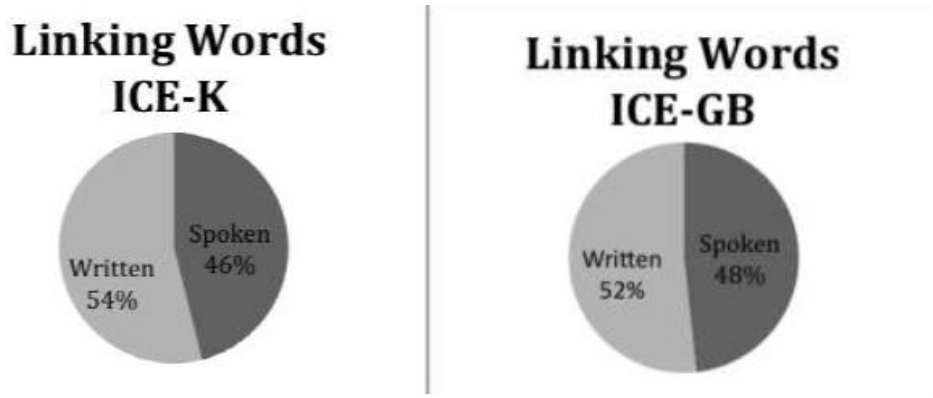

Figure 7.

Linking expressions: distribution in spoken versus written ICE components

The distribution of the total number of linking expressions in both spoken and written components of the corpora is almost identical in the ICE-K and ICE-GB as presented in Figure 7. Linking words are overall equally frequent in the written as in the spoken components of the corpora contradicting their relationship with formality of language.

5.4. Intensifiers

According to the *Cambridge Dictionary Online* an intensifier "is a word, especially an adverb or adjective, which has little meaning itself but is used to add force to another adjective, verb or adverb." An assumption is therefore made here that the extensive use of intensifiers is a sign of informal language through enhancing the emotional content of an utterance.

	ICE-K	**ICE-GB**	**Significance**
absolutely	22	142	$p<0.05$ *
awfully	0	9	—
completely	57	110	$p<0.05$ *
deeply	26	25	$p>0.05$
entirely	25	68	$p<0.05$ *

especially	261	142	p<0.05 **
extremely	51	81	p<0.05 *
fairly	36	99	p<0.05 *
hugely	1	3	—
immensely	5	12	p>0.05
incredibly	3	15	p<0.05 *
particularly	140	232	p<0.05 *
quite	238	774	p<0.05 *
rather	189	415	p<0.05 *
really	480	1301	p<0.05 *
slightly	25	112	p<0.05 *
terribly	2	31	p<0.05 *
totally	46	86	p<0.05 *
utterly	2	23	p<0.05 *
very	1,474	2,050	p<0.05 *
Total	3,083	5,730	p<0.05 *

Table 6.

Intensifiers: a frequency overview in both corpora

Table 6 presents a list of words selected as representative of the parameter of intensifiers, according to the Word List in the ICE-K together with the numbers of occurrences and the significance level of the difference between these numbers.

Out of the 20 intensifiers which were initially chosen, two, i.e. *awfully* and *hugely* did not fulfil the criterion of a minimum of 10 tokens in at least one of the corpora and were omitted when applying the chi square test, and in further analysis.

The chi square test run on 18 intensifiers revealed a statistically significant difference in frequency in 16 cases. The intensifiers *deeply* and *immensely* did not reveal any statistically significant differences in the number of tokens and were not dealt with in further analysis. In all but one of the statistically significant intensifiers

ICE-GB revealed a higher frequency. Only the intensifier *especially* revealed a higher frequency in the ICE-K.

The chi square test run on the total number of intensifiers in both of the corpora also revealed a statistically significant difference in frequency, with a larger number of tokens in the ICE-GB. This may lead to assumptions about the lower level of formality of language in this corpus.

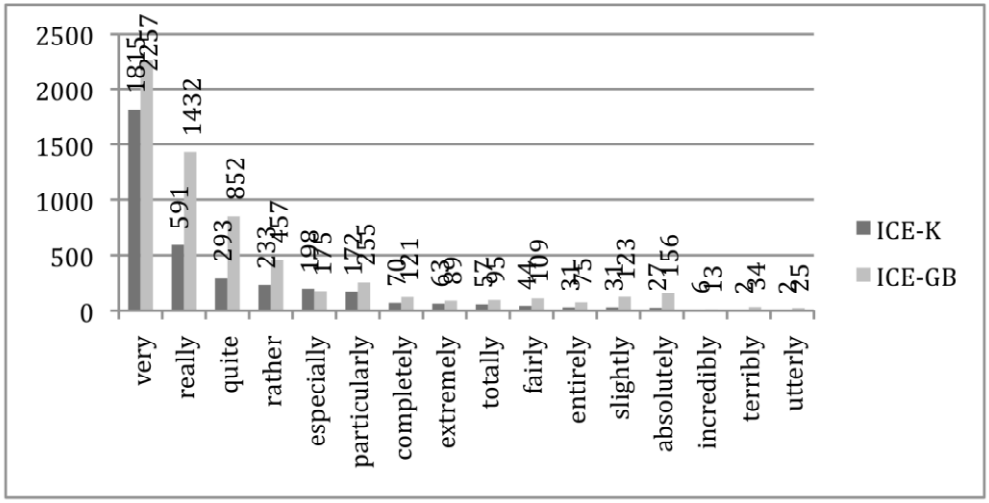

Figure 8.

Intensifiers: the comparative frequency per one million words

Figure 8 contains a bar chart presenting the number of tokens for each intensifier counted per one million words and presented from the highest to the lowest frequencies in the ICE-K. As can be seen the pattern of frequency level is similar in both of the corpora. The biggest difference between the two corpora can be observed between the numbers of tokens for the two most frequently used intensifiers, i.e. *very* and *really*. In the case of ICE-K the intensifier *very* has a 3 times higher frequency than the statistically second intensifier *really*, in ICE-GB the difference in the level of frequency between the first two intensifiers is only 1,5. The overall frequency of the use of intensifiers tends to be lower in the ICE-K with a much higher frequency of the most popular intensifier *very*.

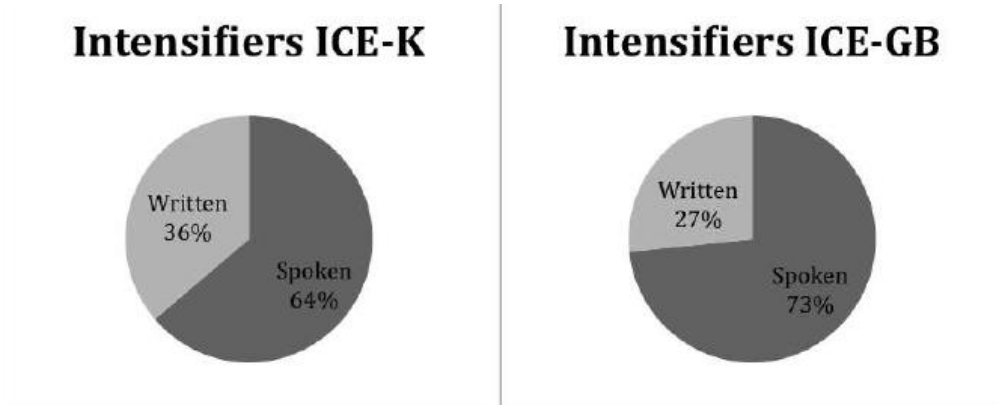

Figure 9.

Intensifiers: distribution in spoken versus written ICE components

As can be seen in Figure 9, according to expectations intensifiers are overall more frequently used within the spoken components of the corpora. Despite a 2700 difference in the number of tokens the distribution is quite similar between the two corpora: ICE-K 64% and ICE-GB 73%. The overall higher number of intensifiers in the ICE-GB confirms its lower level of formality as compared with the ICE-K.

6. Conclusions

This quantitative analysis serves as an introduction to a further qualitative analysis of the presented features in the ICE-GB and the ICE-K. On the basis of this analysis several tentative conclusions are drawn.

First of all, after careful analysis and slight modifications of the contents of the two corpora it can be stated that the ICE-GB and ICE-K are parallel corpora and the frequency of lexical items per one million words was counted in order to make direct comparisons between them.

On the basis of the list of 100 most common words in the ICE-K several parameters revealed a statistically significant difference in the number of tokens between the ICE-K and ICE-GB. Here belong prepositions, the pronoun *I*, contracted versus full forms of the adverb *not*, the linking word *because* and the adjective *very*. An assumption has been therefore made that these lexical items might be indicative of the following parameters connected to the level of formality of language: frequency

of expressions of personal opinion, contractions, linking words and intensifiers.

A quantitative analysis revealed numerous statistically significant differences in the frequency of these parameters between the ICE-GB and the ICE-K. The overall formality level on the basis of the analysed parameters revealed to be higher in the ICE-K.

This fact can be partially explained by the sociolinguistic situation of Kenya which has two official languages Kiswahili and English. Informal settings are generally dominated by Kiswahili and tribal languages, leaving the context of use for English limited mostly to professional life, education and the media. These domains are naturally associated with a formal register influencing the overall level of formality of the English language in Kenya.

It has been shown that there is a significant difference between the level of formality of English in a native and monolingual environment, e.g. Great Britain and a non-native, multilingual environment, e.g. Kenya as reflected in the ICE.

References

Abdulaziz, Mohamed (1991) "East Africa (Tanzania and Kenya)." [In:] Jenny Cheshire (ed.) *English around the World: Sociolinguistic Perspectives.* Cambridge: Cambridge University Press; 391-401.

Bailey, Richard, Manfred Görlach (eds.) (1982) *English as a World Language.* Ann Arbor: University of Michigan Press.

Bex, Tony, Richard James Watts (eds.) (1999) *Standard English. The Widening Debate.* London: Routledge.

Bolton, Kingsley, Braj Behari Kachru (eds.) (2006) *World Englishes. Critical Concepts in Linguistics.* Vols. 1-6. Oxon: Routledge Taylor & Francis Group.

Cambridge Dictionaries Online 2011: Cambridge University Press. http://dictionary.cambridge.org/ Accessed 10.06.2011

Cheshire, Jenny (ed.) (1991) *English around the World: Sociolinguistic Perspectives.* Cambridge: Press Syndicate of the University of Cambridge.

Foley, Hank, Diane Hall (2003) *Longman Advanced Learner's Grammar. A Self-study Reference and Practice Book with Answers.* Harlow: Pearson Education Limited.

Hancock, Ian, Rachel Angogo (1982) "English in East Africa." [In:] Richard Bailey, Manfred Görlach, (eds.) *English as a World Language.* Ann Arbor: University of Michigan Press. 306-323.

Kachru, Braj Behari (2006) "English as an Asian Language." [In:] Kingsley Bolton, Braj Behari Kachru (eds.) *World Englishes. Critical Concepts in Linguistics.* Vols. 1-6. Oxon: Routledge. vol. 2: 324-342.

Kachru, Braj Behari (ed.) (1982) *The Other Tongue: English across Cultures.* Urbana and Chicago: University of Illinois Press.

McEnery, Tony, Andrew Wilson (2001) *Corpus Linguistics. An Introduction.* 2nd ed. (1st ed. 1996) Edinburgh: Edinburgh University Press.

McEnery, Tony, Richard Xiao, Yukio Tono (2006) *Corpus-Based Language Studies. An Advanced Resource Book.* Oxon: Routledge.

Oakes, Michael Paul (1998) *Statistics for Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Platt, John Talbot, Heidi Weber, Ho Mian (1984) *The New Englishes.* London: Routledge and Kegan Paul.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, Jan Svartvik (1985) *A Comprehensive Grammar of the English Language.* Harlow: Pearson Longman.

Schmeid, Josef (1991) *English in Africa: An Introduction.* Harlow: Longman Group UK Limited.

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

Swan, Michael (1995) *Practical English Usage* 2nd ed. (1st ed. 1980) Oxford: Oxford University Press.

Tognini-Bonelli, Elena (2001) *Corpus Linguistics at Work.* Amsterdam: John Benjamins Publishing Co.

Trudgill, Peter (1999) "Standard English: What It Isn't." [In:] Tony Bex, Richard J. Watts (eds.) *Standard English. The Widening Debate.* London: Routledge. 117-128.

Zuengler, Jane (1982) "Kenyan English." [In:] Braj, Behari Kachru (ed.) *The Other Tongue: English across Cultures.* Urbana and Chicago: University of Illinois Press. 112-124.