

*Sergio Baldi, Maddalena Toscano
Università delgi Studi di Napoli*

Arabic Grammatical Loans in Contemporary Swahili Prose Texts¹

Abstract: The presence of Arabic loans in Swahili has not become subject of reliable corpus-based analyses so far. The influence of Arabic language on Swahili can be investigated in literary sources, but reference to whether writers are Muslims or not is essential for their differentiation. This article intends to investigate the presence of Arabic in contemporary prose texts written by Tanzanian authors from Zanzibar and from mainland. The electronic corpus has two sets, Tanzanian corpus and Zanzibarian corpus respectively which are almost equal in size. The reference list of Arabic loans has been extracted from two published sources. Using *Concordance*, a concordance software for text analysis, the frequency of words representing grammatical classes is tested. Differences in the two corpora have been indicated, as well as some shared occurrences of items of Arabic origin (mostly adverbs and conjunctions).

Keywords: Swahili, Arabic, loans

Introduction. The representation of Arabic loans in Swahili is either overestimated or underestimated, usually without any assessment based on reliable counting. On this topic there is just a couple of articles by Elena Bertoncini-Zúbková (1973; 1985), not to mention the one which is published in Italian and therefore not so accessible to a vast number of scholars. Statements on this subject

¹ Special thanks to Flavia Aiello who revised the translations of the Swahili examples.

are generally based on intuitive assumptions rather than on attested data.

Apart from the above mentioned articles, the works on Swahili literature do not provide any checked information about the percentage of Arabic loans in Swahili texts. It is obvious that the influence of Arabic must be different in what it is or was produced by writers who were Muslims with a good knowledge of Arabic, maybe from Zanzibar, and by others who were not Muslims and living in the mainland. In this paper we will try to investigate the representation of Arabic loans in Swahili grammatical words as manifested in an *ad hoc* tailored corpus.

The data on Arabic loans. The reference list of Arabic loans has been extracted from two sources: 1) Baldi's list of Arabic loan words in Swahili (Baldi 2008, from now on BL): it is a list of 1.880 items, each one of them has an indication of grammatical categories, source dictionary and the meaning in French. 2) Bertoncini-Zúbková's work on Swahili frequency words (Bertoncini-Zúbková 1973, from now on BZL): it is a Swahili frequency list containing 1.143 entries, coming from 3.700 words extracted from a count of 40.000 occurrences found in 100 sources (ancient texts, modern literary texts, journalistic texts, miscellaneous texts). For each item, information about the grammatical category and the meaning in English is provided, along with the rank with respect to the general frequency and the frequency relevant for each set of text. Loan words, with indication of their origin, are marked in italics.

The corpus. This is a raw corpus made of Swahili creative contemporary prose texts published mainly during the second half of XX century (1960-2002).

This is quite homogeneous corpus consisting of 226.486 tokens which represent mainland Tanzanian writers' texts and 186.355 tokens of Zanzibar writers' texts. The two groups of texts contain 29.071 and 29.775 different words respectively. The corpus has been divided into two sets of texts, according to the authors' origin (see References):

- Tanzanian Corpus (TC): it includes five works by four Tanzanian authors from mainland Tanzania,
- Zanzibarian Corpus (ZC): it includes six works by three Tanzanian authors from Zanzibar.

Attention was given to secure a comparable set of texts. In fact, the number of tokens differs in the two sets of texts, but the number of single words is quite similar; therefore, it was decided to maintain the selected texts.

Corpus	Initials	Words	Tokens
Tanzanian Corpus	(TC)	29.071	226.486
Zanzibarian Corpus	(ZC)	29.775	186.355
Total		59.239	412.841

The tool. The software used is *Concordance*, a powerful and flexible tool for text analysis that provides useful insight into electronic texts². It takes an electronic text as input and produces the list of appropriate words contained in the texts as output, along with giving the number of words, tokens, frequency and rank. It allows searching the context of each single token. It does not allow grammatical analysis, neither automatic morphological lemmatization nor syntax parsing. However, refined searches through the use of REGEX (regular expressions) are possible.

The analysis. Considering that the corpus does not have labels that allow refined search and that *Concordance* does not have functions for parsing (neither morphological nor syntactical), it was decided to start the analysis with invariable items, divided according to the grammatical categories (parts of speech that they represent).

We have decided to restrict the analysis to grammatical words for two main reasons:

- the Swahili corpus is a raw corpus, i.e. the forms attested in the corpus have no labels and it makes looking for inflected forms impossible,

² *Concordance* is available at <http://www.concordancesoftware.co.uk/>

–the software used for text analysis is adapted to searching for words and concordances of invariable forms but has no parsing functions to search for possible variants of basic lemma (except for the case of REGEX which can produce acceptable results).

Getting a list of grammatical items. The first step was to extract the items labelled as grammatical items from the two lists. Some items were assigned more than one label. The set of labels used by the two source lists were quite homogenous. Nevertheless, some adaptations were needed (a. and adj. = adjective; int. and interj.³ = interjection; part., pr. and prep. = preposition). The label v.⁴ (verb) was not relevant for our data. The final list of labels that we use is as follows:

1. adj.	adjective	5. int.	interrogative
2. adv.	adverb	6. interj.	interjection
3. conj.	conjunction	7. num.	numeral
4. idiom.	idiomatic	8. prep.	preposition

The next step was to extract the grammatical items from the two lists and to assemble them together. The first rough result was the list BL composed of around 160 items and the list BZL of around 60 items.

Some adaptations were needed, and as a result, some items had to be excluded. Here are few examples: *ina* is commonly listed with adjectival forms, but in the corpus it was found only as a constituent of the predicate ‘to have’ (cl. 3 and 9 + *na*); *wa* was used only in expressions, not as an isolated item, e.g. *kadha wa kadha*; the same with *bi*, as it was found only as *Bi(bi)*, and *amin*, found only as *Idd Amin*.

Graphic variants. The items representing one grammatical category were put together and listed according to the alphabetical order. In the case of alternative forms, the main representation was selected. For example, among the two forms *makusudi* and *kusudi*

³ interj. (Sw.: Ki Kiingizi).

⁴ *Wakati* which is indicated as loan from Arabic is marked with label v.

the first one (*makusudi*) was put as basic item, but the frequency of the two was checked through the whole text. For some items (*marhaba*) also the variants not included in the list were checked. Some missing items have been added, as *salala*, *masalala*, *msalale*; *masalkheri*; *minajili*; *minghairi*. Items with labels referring to verbs and nouns have also been maintained.

Phrases *badala ya*, *kadha wa kadha*, *semaa wa taa*, *lila na fila* were searched as expressions and not as isolated items. *Baada ya*, *baina ya* and similiar have been searched and counted only as basic items (*baada*, *baina* respectively). Ambiguous items such as *la*, *haya* have been searched separately and marked with “!”.

Expressions were counted according to the basic item, for example *kadha wa kadha* representing the item *kadha* has 30 occurrences (ZC). As 7 expressions of *kadha wa kadha* make 14 occurrences, there are 16 isolated occurrences of this item.

In items accompanied by multiple labels, as *-kabidhi* which was marked as v./adj., the v. was not maintained. The sequence of the labels is the one adopted by the source (see Appendix 1).

The further step was to transform the previous list into a list suitable to be used in *Concordance*. This implies canceling labels and numbers, dividing the items according to the categories and listing graphic variants as separate items (see the list in Appendix 2).

By using the lists indicated in Appendix 2, the searches for each grammatical category were conducted separately in the two sections of the corpus. The results are listed in Tab. 1. to Tab. 6. The number in parenthesis indicates the amount of items included in each separate list. The number of words indicates single occurrences, the number of tokens indicates how many occurrences for each single word have been attested.

Arabic adjectives

Tab. 1 – Arabic adjectives (83 items)							
TC adjectives: words 41, tokens 1224, t/w 30							
ZC adjectives: words 46, tokens 1096, t/w 24							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
1.adimu	2	1. adimu	4	1. kila	504	1. kila	521
2. ajabu	80	2. ajabu	83	2. ajabu	80	2. ajabu	83
3. ajibu	3	3. ajibu	1	3. muhimu	72	3. bora	75
4. amini	2	4. akali	4	4. makini	66	4. sawa	51
5. aushi	1	5. amini	2	5. bora	56	5. muhimu	31
6. awali	40	6 aula	1	6. kamili	52	6. rahisi	24
7. bora	56	7. awali	5	7. sawa	51	7.marehemu	23
8. dhaifu	14	8. badiri	1	8. rahisi	46	8. makini	23
9. dhalimu	2	9. bahili	1	9. awali	40	9. kamili	23
10. duni	12	10. bora	75	10. marehemu	37	10. haba	23
fani	6	dhaifu	14	halisi	27	hodari	22
fasihi	4	dhalimu	2	hafifu	24	sahihi	20
ghali	1	dhila	3	huru	20	haramu	20
haba	8	duni	4	sahihi	14	hafifu	15
hafifu	24	fasihi	2	dhaifu	14	dhaifu	14
halisi	27	ghali	3	hodari	12	imara	11
haramu	6	haba	23	duni	12	huru	11
hodari	12	hafifu	15	yakini	8	halisi	10
huria	7	halisi	10	haba	8	nadhifu	9
huru	20	haramu	20	huria	7	madhubuti	9
imara	6	hodari	22	mahututi	6	huria	7
kamili	52	huria	7	imara	6	thabiti	5
kila	504	huru	11	haramu	6	nadra	5

laghai	2	imara	11	fani	6	awali	5
madhubuti	3	jinai	1	nadhifu	5	stahili	4
mahiri	1	kamili	23	mahsusi	4	sharifu	4
mahsusi	4	karimu	1	fasihi	4	duni	4
mahututi	6	kila	521	madhubuti	3	akali	4
makini	66	laghai	2	ajibu	3	adimu	4
marehemu	37	madhubuti	9	maridhawa	2	wastani	3
maridhawa	2	makini	23	laghai	2	sunu	3
muhimu	72	marehemu	23	dhalimu	2	ghali	3
nadhifu	5	muhimu	31	amini	2	dhila	3
nadra	1	nadhifu	9	adimu	2	yakini	2
rahisi	46	nadra	5	thabiti	1	laghai	2
raufu	1	najisi	1	sunu	1	fasihi	2
sahihi	14	rahisi	24	raufu	1	dhalimu	2
sawa	51	safii	1	nadra	1	amini	2
sunu	1	sahihi	20	mahiri	1	safii	1
thabiti	1	sawa	51	ghali	1	najisi	1
yakini	8	sharifu	4	aushi	1	karimu	1
		stahili	4			jinai	1
		sunu	3			bahili	1
		thabiti	5			badiri	1
		wastani	3			aula	1
		yakini	2			ajibu	1

Looking at the first 10 most frequent items, we observe that 9 are the same in TC and ZC. The only marked difference is in the use of *awali* – ‘originally, initially, early’ (40 in TC and 5 in ZC), e.g.: *awali* in ZC as *-a awali* (3):

1. *Ilikuwa vigumu vilevile, ingawa si kama mara ya awali.*
‘It was also difficult, although it was not like the first time.’
2. *Kwa mara ya awali tangu alipotoroka kwao Mbiju, alitanabahi juu ya yale yaliyomfutu.*
‘For the first time since he fled Biju, he realized what he had forgotten.’
3. *Korja ya mishelisheli ilitamalaki kipande cha awali.*
‘About twenty breadfruit trees overhung from one side.’

(*awali* in ZC as *pale awali*)

4. *Mazungumzo baina ya huyu msichana na mzee, yalikuwa arefu kuliko pale awali.*
‘The conversation between the girl and the old man was longer than the previous one.’

awali in TC as *-a awali* (14)

5. *Wasiwasi ule wa awali ulikuwa umepotea kabisa.*
‘The previous concerns were gone.’
6. *[...] na hatimaye ukilipata mtu hukuona tofauti na yale ya awali!*
‘[...] and when you found it at last nobody noticed the difference with the first ones!’

awali in TC as *hapo awali* (11)

7. *[...] Ilionyesha wazi kwamba hapo awali ilikuwa na maji.*
[...] ‘Clearly showed that previously it had water.’

Arabic adverbs

Tab. 2 – Arabic adverbs (53 items)							
TC adverbs: words 35, tokens: 6375, t/w 182							
ZC adverbs: words 35, token 4083, t/w 117							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
1. afadhali	16	1. afadhali	25	1. sasa	943	1. sasa	791
2. asubuhi	188	2. aghalabu 1	15	2. sana	782	2. mara	455
3. baada	694	3. aghlabu 2		3. baada	694	3. baada	327
4. baadaye	104	4. asubuhi	76	4. wakati	625	4. wakati	322
5. bado	339	5. baada	327	5. mara	611	5. zaidi	236
6. bure	51	6. baadaye	117	6. zaidi	476	6. karibu	194
7. daima	42	7. bado	187	7. bado	339	7. bado	187
8. ghafla	135	8. bure	84	8. kabisa	331	8. hasa	177
9. hadhara	2	9. daima	46	9. karibu	261	9. labda	171
10. halafu	104	10. ghafla	103	10. subuhi	188	10. sana	162

Out of the first 10 most frequent items, 8 are the same, with a marked difference in the use of *sana* (TC 782 – ZC 162). Apart from following mainly verbal forms, *sana* in TC preferably follows adjectives (*-zuri* (46), *-kubwa* (40) and *-ingi* (39) as first cooccurrences), while in ZC it follows nouns (*nguvu* (8), *zamani* (4) and *-zuri* (4) as most frequent co-occurrences).

Samples of *sana* in TC

8. *Mwaka huu ulikuwa mwaka mzuri sana kwa wanakijiji hawa wa Bucho.*
‘This year was a great year for the villagers of Bucho.’
9. *Alikuwa mdogo sana. Kisha mzuri sana.*
‘She was very young. And very good.’
10. *Nyumba ilikuwa kubwa sana kwa familia ya watatu kama ya Dennis.*
‘The house was too big for a family of three like the one of Dennis.’

11. *Ilikuwa ni kazi kubwa sana kulinda vifaa na vyombo vya watafiti na wataalamu wake.*
‘It was a very difficult task to protect the equipment and utensils of the researchers and professionals.’
12. *Chumba cha maongezi kilikuwa kikubwa. Kilikuwa na picha nyingi sana ukutani.*
‘Living room was big. It had a lot of pictures on the wall.’
13. *Nimekwisha sikia jina lako mara nyingi sana humu mjini.*
‘I have heard your name very often in the city.’

Samples of sana in ZC

14. *Na yeye mwenyewe akajitupa chini kwa nguvu sana,*
‘And he fell on the ground forcefully.’
15. *Alipokwisha funikwa, shetan akaanza kutika na kuzungusha kichwa kwa nguvu sana.*
‘When he was covered, the demon began to shake and turned his head forcefully.’
16. *[...] jumba bovu lililobomoka zamani sana.*
‘[...] a dilapidated palace collapsed long ago’
17. *Sababu za kugoma sote tunazielewa, tumeshaelezana zamani; sana ni dharau, maonevu.*
‘The main reasons for the strike – we all knew it and we have already talked about – were the contempt and disrespect that we were showed.’
18. *Unaonaje tukipata nyumba nzuri sana... ?*
‘Don't you feel we got a very nice house..?’

Arabic interjections

Tab. 3 – Arabic interjections (36)							
TC Interjections: words 11, tokens 457, t/w 42							
ZC Interjections: words 15, tokens 402, t/w 27							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
1.ahsante 3	37	1.ahsante 10	19	1. basi	311	1. basi	240
asante 34		asante 9		2. hebu 38	63	2. hebu 61	62
2. ala	2	2. ala	16	ebu 25		ebu 1	
3. ashakum	2	3. basi	240	3. asante 34	37	3. naam	32
4. basi	311	4. ebu 1	62	ahsante 3		4. asante 10	19
5. ebu 25	63	hebu 61		4. laiti	20	asante 9	
hebu 38		5. hamadi	3	5. naam	7	5. ala	16
6. hasha	2	6. hasha	4	6. marahaba	6	6. laiti	9
7. laiti	20	7. hashakum	2	7. ole	5	7. wallahi	7
8. marahaba	6	8. labeka	1	8. yaa	2	8. hasha	4
9. naam	7	9. lahaulaa	2	9. hasha	2	9. hamadi	3
10.ole	5	10.laiti	9	10.ashakum	2	10.ole	2
yaa	2	marahaba	2	ala	2	marahaba	2
	457	naam	32			lahaulaa	2
		ole	2			hashakum	2
		salala	1			salala	1
		wallahi	7			labeka	1

Within the first 10 items 7 are the same. A marked difference within these 7 items is in the use of variants:

TC	ZC
hebu 38, ebu 25	63 hebu 61, ebu 1
asante 34, ahsante 3	37 ahsante 10, asante 9
	19

A marked difference between the two sources is also the quantitative use of *Ala!* ‘expression of annoyance’ (occurrences of *ala* ‘instrument, scabbard’ have not been considered).

Ala! in TC

19. *Ala! Kumbe huyu ni Mheshimiwa Luhala!*

‘Hey! So this is Mr. Luhala!’

20. *Ala! Simama hapo ulipo! Sasa ninataka jambo moja tu!*

‘Hey! Stop where you are! Now I want only one thing!’

Ala! in ZC

21. *Ala, bado unakokota miguu?*

‘Hey, are you still dragging feet?’

22. *Ala, jamani huyu mtoto yuko wapi?*

‘Hey, this child’s friend, where is he?’

Arabic numerals

Tab. 4 – Arabic numerals (30 items)							
TC Numerals: words 15, tokens 391, t/w 27							
ZC Numerals: words 12, tokens 261, t/w 22							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
arobaini	7	arobaini	2	mia	75	mia	59
elfu	27	edashara	1	sita	65	sita	47
hamsa	1	elfu	28	ishirini	56	saba	35
hamsini	21	hamsini	18	saba	41	ishirini	30
ishirini	56	ishirini	30	tisa	30	elfu	28
mia	75	mia	59	elfu	27	tisa	22
saba	41	saba	35	sabini	21	hamsini	18
sabini	21	sita	47	hamsini	21	thelathini	9
sita	65	sitini	9	themanini	18	sitini	9
sitini	8	thelathini	9	thelathini	11	arobaini	2
thelathini	11	tisa	22	tisini	8	tisini	1
theluthi	2	tisini	1	sitini	8	edashara	1
themanini	18			arobaini	7		
tisa	30			theluthi	2		
tisini	8			hamsa	1		

Here also 8 out of the first 10 items are the same.

A check of the collocates of 4 first most frequent items gave the following results:

Mia, sita, ishirini, saba in TC: shilingi mia; saa sita; miaka / milioni / mara / watu / magunia ishirini; saa / darasa / miaka saba

Mia, sita, saba, ishirini in ZC: shilingi, ngazi-mia, mia mia; miezi / miaka, saa sita; shilingi (elfu) ishirini

23. *Tokea siku ile waliofika kileleni juu ya ngazi-mia tamaa ya Sewa ilianza kuchipua na kuchanua.*

‘Since the day they reached the peak over the hundred feet, the desire of Sewa started to sprout and bloom.’

With *mia mia* the meaning seems to be rather ‘exactly hundred’ than ‘about/around hundred’, e.g.:

24. *Njiani aliufungua ule mkoba wake kutazama kilichosalia – noti tatu za shilingi mia mia na moja ishirini.*

‘On the way, he opened his wallet to look for what was left – three banknotes of one hundred shillings each and one of twenty.’

25. *Akathibitisha udhati wake kwa majani mawili ya shilingi mia mia.*

‘He proved his trustworthiness with two banknotes of a hundred shillings each.’

26. *Inspekta Fadhili ana hakika kwamba ukumbi huo aukuwa katika hali hiyo miezi sita nyuma.*

‘Inspector Fadhili is confident that the hall was not in the same state six months ago.’

27. “*Wazimu ncha saba*”, *alinong'ona*.
 “Countles spirits”, he whispered.’

28. *Asili ya mizimu ule, kama asili ya mizimu yote inavyokuwa, hadithi yenye ncha saba*.
 ‘The origin of that ghost, as the origin of all ghosts happens to be, is an endless story’

See also the expression *korja ya* ‘around twenty’ in sample 3.

Arabic conjunctions

Tab. 5 – Arabic conjunctions (28 items)							
TC Conjunctions: words 17, tokens 6929, t/w 408							
ZC Conjunctions: words 18, tokens 5638, t/w 313							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
aidha	1	aidha	3	1. kama	2109	1. kama	1394
ama	48	ama	55	2.lakini 1288	1289	2. lakini 1207	1207
au	394	au	647	walakini 1		walakini 6	
bali	87	bali	137	3. hata	899	3. tena	711
hadi	358	hadi	77	4. tena	606	4. au	647
hata	899	hata	561	5. au	394	5. hata	561
ila	15	ila	111	6. hadi	358	6. wala	284
ili	356	ili	175	7. ili	356	7. ili	175
jinsi	194	jinsi	120	8. kabla	243	8. bali	137
kabla	243	kabla	95	9. jinsi	194	9. jinsi	120
kama	2109	kama	1394	10.wala	147	10. ila	111
lakini 1288	1289	lakini 1201	1207	mradi	138	kabla	95
walakini 1		walakini 6		bali	87	hadi	77
minajili	13	madhali	3	ama	48	ama	55
mradi	138	mintarafu	1	yaani	32	yaani	38
tena	606	mradi	19	ila	15	mradi	19

wala	147	tena	711	minajili	13	madhali	3
yaani	32	wala	284	aidha	1	aidha	3
	6929	yaani	38			mintarafu	1

The table shows that 8 of the first most frequent 10 items are the same in the two sources. Looking at frequencies, we note much more occurrences of *kama* in TC than in ZC, whereas looking at colligates of *kama* we notice a higher co-occurrence of *kama si* in ZC (46) than in TC (18).

TC: kama 2109				ZC: kama 1394			
left		right		left		right	
na	44	vile	78	na	35	vile	76
hata	35	mtu	65	ilikuwa	28	si	46
ilikuwa	30	ni	44	si	18	kawaida	30
ni	29	kawaida	28	hata	14	kwamba	30
watu	25	watu	26	alikuwa	11	mtu	23

Arabic prepositions

Tab. 6 – Arabic prepositions (8 items) TC Prepositions: words 4, tokens 816, w/t 204,5 ZC Prepositions: words 4, tokens 457, w/t 92							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian copus		Tanzanian corpus		Zanzibarian corpus	
badala	136	badala	26	bila	435	bila	336
baina	9	baina	82	kuhusu	238	baina	82
bila	435	bila	336	badala	136	badala	26
kuhusu	238	kuhusu	13	baina	9	kuhusu	13

The same 4 items appeared in the two corpora, but with much higher use of *kuhusu* and *badala* in TC. As for *badala* and *baina* the almost only colligate is *-ya* in both corpora. As for *kuhusu*, in TC the most frequent colligates are demonstrative forms, whereas there is only

one co-occurrence with a demonstrative form in ZC (*kuhusu ile mvua*).

Arabic idomatic expressions (3 items)

Kadha wa kadha; Lila na fila; Sema wa taa

Only one item: *Kadha wa kadha*

TC idiomatic expressions: words 1, tokens 2 (only in Jozi)

ZC idiomatic expressions: words 1, tokens 7 (in Tata, Bwana Musa, Kiu)

29. *Aliusimamisha mchi katika kinu na kuuzungusha mara kadha wa kadha huku akimtaazama mama yake*

She took up a pestle in a mortar and turned it around and around several times as she looked at her mother'

30. *Hadija alisimama na kimbia nyimbo kadha wa kadha za taarabu.*

'Khadija stood up and sang several *taarab* songs'

La! Haya!

The contexts of these items have been checked thoroughly in order to avoid ambiguity with *la* (cl. 5 of *-a*) and *haya* (cl. 6 of *h-*)

TC – La! 12 occurrences ZC - La! 28 occurrences

TC - Haya! 18 occurrences ZC – Haya ! 29 occurrences

Adjectivals with concords

(7 items in BL: *-anisi, -faransa, -fidhuli, -halifu, -halisi, -lainifu, -tatai*).

A search for forms of the 7 lemma from BL, made with REGEX⁵, produced the following results:

ZC: lemma 2; words 4; tokens 13 (*kifaransa* 1; *kuhalifu* 1; *mhalifu* 6; *wahalifu* 5)

⁵ Regex search string used for retrieving adjectivals forms:

(m|mw|wa|w|mu|mi|ji|ma|ki|ch|vi|vy|n|ny|mb|nd|ku|kw|pa)(anisi|faransa|fidhuli|halifu|halisi|lainifu|tatai) (See App. 8)

TC: lemma 1; words 2; tokens 2 (*mhalisi* 1; *wahalifu* 1)

In order to have a first idea of the representation of grammatical items of Arabic origin in Swahili a counter check was performed on some adverbs, conjunctions and prepositions of Bantu origins. The items have been extracted from the list by Bertoncini-Zúbková (1973). The results (see App. 65 and 6) are as follows:

Bantu grammatical items: 123

Arabic grammatical items: 60

Bantu Adverbs – (pick list: 40 items – basic list: 31 items)

In the case of adverbs, from a list of 40 items (some of which are inflected forms), this were the results:

Tab. 7 – Bantu adverbs – Frequency list – 40 items					
TC: words: 39, tokens: 6591; ZC: words 39, tokens 5999					
Tanzanian corpus			Zanzibarian corpus		
1.pale	642		1.huku	659	
2.tu	593		2.juu	644	
3.hapa	521		3.tu	619	
4.juu	450		4.pale	395	
5.kisha	385		5.ndani	364	
6.huku	381		6.chini	353	
7.kwanza	343		7.mbele	325	
8.pamoja	336		8.kisha	304	
9.pia	314		9.hapa	252	
10.ndani	312		10.nje	242	
chini	254		nyuma	239	
nje	235		kwake	207	
mbele	213		pamoja	159	
kwake	191		kwanza	155	
kule	138		pia	149	
nyuma	127		pengine	121	
mbalimbali	106		kule	111	
jana	105		mbona	81	
pole	103		kwao	80	
kwao	94		jana	65	
mbona	86		punde	64	

siyo	82		lini	64
kwangu	71		mno	51
mapema	70		siyo	41
lini	61		kando	37
huenda	57		kwangu	34
mno	51		mle	31
kwetu	50		kwako	30
kando	32		humu	29
humu	31		mapema	20
kwako	29		mbalimbali	18
mle	27		kwetu	16
			pole	12
pengine	24		huenda	12
kwenu	23		kwenu	10
punde	21		yapata	4
polepole	18		pasipo	1
kienyeji	7		kienyeji	1
pasipo	6			
yapata	2			

Within the first 10 most frequent items, 7 items are the same.

An investigation on *hapa* produced *hapa na pale* as the most frequent sequence both in TC (17) and in ZC (9). As for *pale*, the most frequent colligates are:

TC: *pale kijijini* (22), *pale nje* (21, *pale mezani* (19), *pale mbele* and *pale pale* (17)

ZC: *pale pale* (29), *pale alipokuwa* (19) *pale chini* and *pame penye* (16)

The sequence *pale na* was not considered as it may belong to various syntagmatic constructions.

Bantu conjunctions – (pick list: 10 items – basic list: 9 items)

Tab. 8 – Bantu conjunctions frequency list (10 items)			
TC: words 9, tokens 13658; ZC: words 10, tokens 12125			
Tanzanian corpus		Zanzibarian corpus	
na	10154	na	10498
kwamba	1955	kuwa	677
kuwa	800	kwamba	375
kwani	453	kwani	174
kuliko	110	ingawa	161
ikiwa	88	kuliko	125
ingawa	79	ikiwa	97
iwapo	18	ijapokuwa	13
ijapokuwa	1	iwapo	4
		ijapo	1

The same 9 items appear in both lists, except for one occurrence of *ijapokuwa* in TC (13 occ. in ZC) and only one occurrence of *ijapo* in ZC (in Kiu)

As for *kwamba* we have:

TC: co-occurrences of *ni kwamba* (79), V(verb form) *wazi kwamba* (69), *kama kwamba* (29); *kwamba alikuwa* (88), *kwamba ni* (47)

ZC: *kama kwamba* (30), *ni kwamba* (29); *kwamba alikuwa* (10), *kwamba* PP(personal pronouns) (20)

Bantu prepositions – (pick list: 8 items- basic list: 7 items)

Tab. 9 – Bantu prepositions frequency list (8 items)			
TC: words 8, tokens 7155; ZC: words 8, tokens 5694			
Tanzanian corpus		Zanzibarian corpus	
kwa	4617	kwa	3571
katika	1406	katika	863
kwenye	404	kwenye	614
kutoka	377	mpaka	357
tangu	189	kutoka	151
mpaka	102	tokea	71
mnamo	59	tangu	51
tokea	1	mnamo	16

We have the same 8 items in the two corpora, but their frequency is quite different (ex.: *tangu* TC 189, ZC 51; *tokea* TC 1, ZC 71)

ZC: *tokea hapo* (20)

Here is the only occurrence of *tokea* in TC:

31. *Tazameni, hii adhabu yote tuliyoipata tokea jana hadi leo ni kwa sababu ya Sifuni.*

‘Look, all this punishment we are getting since yeasterday is because of Sifuni’

Conclusive remarks

Here follows a summary of the results of the search for words (W) and tokens (T) of Bantu (B) and Arabic (A) origins in Tanzanian (TC) and Zanzibarian (ZC) Swahili corpora made of contemporary prose texts.

Adverbs							
Bantu				Arabic			
TC		ZC		TC		ZC	
W	T	W	T	W	T	W	T
39	6591	39	5999	35	6375	35	4083

Conjunctions							
Bantu				Arabic			
TC		ZC		TC		ZC	
W	T	W	T	W	T	W	T
9	13658	10	12125	17	6929	18	5638

Prepositions							
Bantu				Arabic			
TC		ZC		TC		ZC	
W	T	W	T	W	T	W	T
8	7155	8	5694	4	818	5	461

While the occurrences of Bantu prepositions (8) double the occurrences of those of Arabic origins (4 and 5), the occurrences of

Bantu adverbs (39) are only slightly higher in number than the occurrences of adverbs of Arabic origins (35). As for conjunctions, the occurrences of items of Arabic origins (17 and 18) double the occurrences of those of Bantu origins (8).

From the above data we can state that, at least for adverbs and conjunctions, there is a significant representation of Arabic loans in Swahili contemporary prose texts. A first check of their distribution within Tanzanian and Zanzibarian corpora shows some differences in the use of some of the items. Although in most cases the first most frequent items do correspond, still there are differences in number of occurrences (see the cases of *sana* and *ala!*). Differences in the collocates of numerals also resulted in the marked differences in clausal patterns. Idiomatic expressions were also found (*ncha saba* ‘countless’, ‘endless’; *korja ya* ‘around twenty’). Differences in the two corpora as for distribution of colligates of *kama* and *kuhusu* have also been detected.

The resulting data of this search and counting cannot be taken as a base for final statements concerning incidence or Arabic grammatical loans and differences between mainland and Zanzibarian Swahili. They can rather be seen as indicators of possible further research. Analysis of collocates and colligates in fact cannot be done by considering only quantitative data. Attentive consideration to syntagmatic and semantic elements must be conscientiously given. Such a detailed analysis of contexts requires long time and considerable manual work. The results of this first work can give hints as to where to search for gathering more data based on information which can lead to more relevant and reliable conclusions.

References:

- Baldi, S., 1988, “A first ethnolinguistic comparison of Arabic loanwords common to Hausa and Swahili”, *AION XXXXVIII* 4, Supplement n. 57, 83 p.
- 2008a, “Arabic Loanwords in Swahili: Addenda”, *Lingua Posnaniensis*, 9-22.

- 2008b, *Dictionnaire des emprunts arabes dans les langues de l'Afrique de l'Ouest et en swahili*, Paris: Karthala.
 - 2009a, "Emprunts et mots voyageurs dans le bassin du Lac Tchad", in: *Actes du XIIIe colloque international du Réseau Méga-Tchad (Maroua, 31 octobre - 3 novembre 2005)*, Marseille: IRD Editions.
 - 2009b, "Hausa and Swahili: A Common Arabic Heritage", in: S. Baldi and H. M. Yakasai (eds.), *Proceedings of the 2nd International Conference on Hausa Studies: African and European Perspectives*, Napoli: Università degli Studi di Napoli.
 - 2009c, "Swahili", in: L. Edzard, R. de Jong (eds.), *Encyclopedia of Arabic Language and Linguistics*, vol. IV, Leiden: Brill.
 - 2011, "Swahili: a Donor Language", *Lingua Posnaniensis*, LIII, 7-24.
 - and Watimila H., 2014, "Ideophones in Swahili: a preliminary survey", in: Iwona Kraska-Szlenk and Beata Wójtowicz (eds.), *Current research in African Studies. Papers in Honour of Mwalimu Dr. Eugeniusz Rzewuski*, Warsaw: ELIPSA.
- Bertoncini-Zúbková, E., 1973, "A tentative frequency list of Swahili words", *AION* 33, 297-363.
- 1985, "Quantitative analysis of Swahili vocabulary", in: R.G. Schuh (ed.), *Studies in African Linguistics, Supplement 9: Précis from the 15th Conference on African Linguistics, UCLA, March 29-31, 1984*.
- Toscano, M., Aiello, F., 2002, "Supporting autonomous Kiswahili language learning: a first experience in transferring knowledge from research to teaching", in: P. Evangelisti, C. Argondizzo (eds.) *L'apprendimento autonomo delle lingue straniere – Filosofia e attuazione nell'università italiana*, Soveria Mannelli: Rubettino Editore.
- and Acquaviva G., Aiello F., 2012, "An on-line course for autonomous learning of Swahili through literature", in: C.

Argondizzo (ed.), *Creativity and Innovation in Language Education*, Bern: Peter Lang Publishing.

Corpus:

Tanzanian

Chachage, C.S.L., 2002, *Makuadi wa Soko Huria*, Dar es Salam: E & D Publisher Limited.

Kezilahabi, E., 1975, *Dunia uwanja wa fujo*, Nairobi: EALB.

--- 1979, *Gamba la nyoka*, Arusha-DSM: Eastern Africa Publications.

Mung'ong'o, C.G.M., 1980, *Njozi iliyopotea*, DSM: TPH.

Mtobwa, B. R., 1998, *Dar es Salaam usiku*, Dar es Salaam: East African Publishers.

Zanzibarian

Abdulla, M.S., 1960, *Mzimu wa watu wa kale*, DSM-Nairobi-Kampala: EALB.

Mohamed, M.S., 1972, *Kiu*, Dar es Salaam: East African Publishing House.

--- 1976, *Nyota ya Rehema*, London: Oxford University Press.

Mohamed, S. A., 1980, *Dunia mti mkavu*, Nairobi: Addison-Wesley Longman.

--- 1980, *Utengano*, Nairobi: Longman Kenya.

--- 1990, *Tata za Asumini*, Nairobi: Longman Kenya.

App. 1 List of grammatical items of Arabic origin from BL and BZL in grammatical order

The following is the list of items from BL and BZL, according to grammatical order. Numbers in the rank column mark the items from BZL. Graphic variants have been considered together and listed according to alphabetical criterion (e.g. in the case of *makusudi*, *kusuki*, it has been decided to insert *makusudi* as first item; the frequency in fact will be checked through the search in the text). Items with labels referring also to verbs and nouns have been maintained. Ex.: *-kabidhi* was labeled as *v./adj.*; the *v.* is not maintained. In case of multiple labels the sequence is the one adopted by the author.

Tab. 10 List of grammatical items of Arabic origin from BL and BZL, grammatical order, used as a base for searching in TC and ZC.

	rank	item	gram.	glossa
1.	27	kila	adj.	every
2.		a tisa	adj.	ninth
3.		anisi	adj.	pleasing
4.		faransa	adj.	French
5.		fidhuli	adj.	arrogant
6.		halifu	adj.	rebellious
7.		kabidhi	adj.	economical
8.		kaidi	adj.	obstinate
9.		lainifu	adj.	smooth; facile
10.		najisi	adj.	unclean
11.		safihi, safii	adj.	pure
12.		sharifu	adj.	honourable
13.		stahili	adj.	merit
14.		tatai	adj.	cunning
15.		aali	adj.	good
16.		adibu	adj.	decorous
17.		adili	adj.	right
18.		adimu	adj.	rare
19.		akali	adj.	a few of
20.		aula	adj.	better
21.		bahasa	adj.	cheap
22.		bora	adj.	excellent
23.		bulibuli	adj.	a white, embroidered school-cap
24.		dhahili	adj.	evident
25.		dhaifu	adj.	weak
26.		dhalimu	adj.	unjust
27.		duni	adj.	mean
28.		fani	adj.	worthy
29.		fasaha, fasihi	adj.	elegance
30.		ghali	adj.	scarce
31.		haba	adj.	little; rare
32.		hafifu	adj.	trifling

	rank	item	gram.	glossa
33.		haramu	adj.	forbidden
34.		hodari	adj.	able
35.		hututi	adj.	difficult
36.		ina	adj.	certainly
37.		kamili	adj.	perfect
38.		karimu	adj.	liberal
39.		laghai	adj.	deceitful
40.		madhubuti	adj.	precise
41.		mahiri	adj.	skilful
42.		mahsusi	adj.	particular
43.		mahututi	adj.	serious
44.		makeruhi, makuruhi	adj.	offensive
45.		maridhawa	adj.	in abundance
46.		muhimu	adj.	important
47.		nadhifu	adj.	clean
48.		nadra	adj.	uncommon
49.		nakawa	adj.	clear
50.		rahimu	adj.	merciful
51.		rahisi	adj.	cheap
52.		raufu	adj.	gentle
53.		saghiri	adj.	small
54.		sahala	adj.	light
55.		sahihi	adj.	true; signature
56.		salihi	adj.	good
57.		sunu	adj.	good
58.		swafi	adj.	clean
59.		tahafifu	adj.	trifling
60.		thabiti	adj.	firm
61.		thelathini	adj.	thirty
62.		yakini	adj.	truth
63.	135	awali	adj. adv.	(at) first
64.		ajabu	adj./adv.	wonder
65.		halisi	adj./adv.	real
66.		ajib, ajibu	adj./adv.	wonderfully
67.		kadha wa kadha	adj./adv.	uncertain number

	rank	item	gram.	glossa
68.		badiri	adj./n.	<i>applied to various celestial phenomena and omens</i>
69.		sawa	adj./n.	equal
70.		kusudi	adv.	(on) purpose
71.		abadan	adv.	never
72.		abadi	adv.	ever
73.	133	afadhali	adv.	better
74.		aghalabu, aghlabu	adv.	more often
75.		baada	adv.	after
76.	110	bado	adv.	not yet
77.	130	bure	adv.	gratis, vainly
78.	133	daima	adv.	constantly
79.		dawamu	adv.	perpetually
80.		dike	adv.	exactly
81.		fauka	adv.	more (than)
82.		foko	adv.	more (than)
83.		ghalibu	adv.	more often
84.		hadhara	adv.	in front of
85.		halafa	adv.	disobediently
86.	131	halafu	adv.	afterwards
87.	100	hasa	adv.	exactly
88.	132	hima	adv.	quickly
89.		hobelahobela	adv.	anyhow
90.		inshallah	adv.	oh yes, certainly
91.	115	kabisa	adv.	utterly
92.	123	kadhaa, kadha	adv.	uncertain number
93.	114	kadhali	adv.	likewise
94.		kasi	adv.	much
95.	133	labda	adv.	perhaps
96.	134	milele	adv.	always
97.		mkabala	adv.	in front (of)
98.		nomi	adv.	full, up to the brim
99.		nusura	adv.	almost
100.		salimini	adv.	safely
101.	21	sana	adv.	very much

	rank	item	gram.	glossa
102.	52	sasa	adv.	now
103.		sawia	adv.	then
104.		tasihili	adv.	quickly
105.		tike	adv.	exactly
106.		wahedu	adv.	alone
107.	133	walau	adv.	at least
108.	97	zaidi	adv.	more
109.	127	kusudi, makasudi	adv. n.	(on) purpose
110.	120	asubuhi	adv. n.	morning
111.	131	ghafila	adv. n.	suddenly
112.	70	mara	adv. n.	a time, at once
113.	104	zamani	adv. n.	time, past
114.	51	baadaye	adv. pr.	after
115.	109	karibu	adv. pr.	near
116.		hususa	adv./adj.	particular
117.		kadiri	adv./conj.	measure; whilst
118.		dahari	adv./n.	always
119.	135	aidha	conj.	moreover
120.	106	ama	conj.	either...or
121.	69	au	conj.	or
122.	127	bali	conj.	but
123.	80	ili	conj.	in order that
124.	10	kama	conj.	as, that, etc.
125.	42	lakini	conj.	but
126.		maadam	conj.	when
127.		madhali	conj.	while
128.		mathalan, mazalani, methalan, methalan	conj.	for instance
129.		mintarafu	conj.	concerning
130.		taraa	conj.	if
131.	123	wa	conj.	and
132.		waima, waina	conj.	if not
133.	86	wala	conj.	nor
134.		walakini	conj.	but
135.	110	yaani	conj.	that is
136.	76	tena	conj. adv.	that is
137.	118	jinsi	conj. n.	method; how

	rank	item	gram.	glossa
138.	119	kadiri	conj. n.	whilst
139.	134	mradi	conj. n.	plan; and so
140.	101	kabla	conj. pr.	before
141.	116	hadi	prep/conj./adv.	until, then
142.	38	hata	prep/conj./adv.	until, etc.
143.	96	ila	conj. prep.	except
144.		lila na fila	idiom.	for good and bad
145.		semaa wa taa	idiom.	hear and obey
146.	87	basi	int. conj. adv.	well
147.	130	ebu, hebu	inter.	well, then
148.	135	la	inter.	no
149.	133	ahasante, ahsante, asante	interj.	thanks
150.		ala!	interj.	<i>expression of annoyance</i>
151.		alaala	interj.	immediately
152.		Alhamdulillah!	interj.	Praise be to God!
153.		Amin	interj.	Amen
154.		ashakum	interj.	pardon me
155.		audhubillahi!	interj.	<i>example of impatience</i>
156.		bismillahi	interj.	in the name of God
157.		halahala	interj.	at once
158.		hamadi!	interj.	<i>used when a person stumbles</i>
159.		Hasha	interj.	certainly not
160.		hashakum	interj.	<i>vulgar</i>
161.		haya!	interj.	come on!
162.		hobe!	interj.	go!
163.		huss!	interj.	make less noise!
164.		labeka!	interj.	at your service!
165.		laiti!	interj.	oh that!
166.		lebeke!	interj.	at your service!
167.		marahaba	interj.	<i>used as a common rejoinder to the salute of an inferior</i>
168.		ole	interj.	exclamation of woe
169.		sefule	interj.	you vile person

	rank	item	gram.	glossa
170.		simile	interj.	make way
171.		wallahi! wallai!	interj.	by God!
172.		yaa!	interj.	oh God!
173.		amini	n./adj.	believe
174.		ashara	n./adj.	ten
175.		aushi	n./adj.	live, wear
176.		bahili	n./adj.	miser
177.		dhila, dhili	n./adj.	low
178.		hanamu	n./adj.	oblique
179.		huri, huria, huru	n./adj.	free
180.		imara	n./adj.	firmness
181.		jinaï	n./adj.	crime, criminal
182.		makini	n./adj.	quiet
183.		marehemu	n./adj.	the late
184.		wahedi	n./adj.	one
185.		wastani	n./adj.	middling
186.		siasa	n./adv.	politics
187.		lahaula	n./interj.	blasphemy
188.		hamsauishirini	num	twenty five
189.	135	hamsini	num	fifty
190.		hamstashara	num	fifteen
191.		arobaini	num.	forty
192.		asherini	num.	twenty
193.		edashara	num.	eleven
194.	125	elfu	num.	thousand
195.		hamsa	num.	five
196.	133	ishirini	num.	twenty
197.	124	mia	num.	hundred
198.		miteen	num.	two hundreds
199.	123	saba	num.	seven
200.		sabaini, sabini	num.	seventy
201.	115	sita	num.	six
202.		sitashara	num.	sixteen
203.		sitini	num.	sixty
204.		theluthi	num.	third part
205.		themani,	num.	eight

	rank	item	gram.	glossa
		themanya		
206.		themanini	num.	eighty
207.		themtnashara	num.	eighteen
208.		sumni, themuni, thummi	num.	eighth part
209.	132	tisa	num.	nine
210.		tisini	num.	ninety
211.	129	naam	particle interj.	yes
212.		baghairi	prep.	without
213.		bi	prep.	by
214.		fi	prep.	on, with
215.		laula	prep.	if not, unless
216.		min	prep.	from
217.		minghairi	prep.	without
218.	131	badala, badala ya	prep.	instead of
219.	130	baina, baina ya	prep.	between
220.	84	bila	prep.	without
221.	124	kuhusu	prep.	concerning
222.	34	wakati	v.	time

App. 2 Pick lists of grammatical items of Arabic origin, from BL and BZ

The pick list is based on Tab. 1. (242 items). The graphic variants have been separated for the sake of searching in the text with Concordance. In the countings, variants will be considered as single item.

2.1 Adjectives 83 items

- | | | | |
|----|-------|-----|----------|
| 1. | ali | 10. | anisi |
| 2. | adibu | 11. | aula |
| 3. | adili | 12. | aushi |
| 4. | adimu | 13. | awali |
| 5. | ajabu | 14. | badiri |
| 6. | ajib | 15. | bahasa |
| 7. | ajibu | 16. | bahili |
| 8. | akali | 17. | bora |
| 9. | amini | 18. | bulibuli |

- | | | | |
|-----|-----------|-----|-----------|
| 19. | dhahili | 52. | mahiri |
| 20. | dhaiifu | 53. | mahsusi |
| 21. | dhalimu | 54. | mahututi |
| 22. | dhila | 55. | makeruhi |
| 23. | dhili | 56. | makuruhi |
| 24. | duni | 57. | makini |
| 25. | fani | 58. | marehemu |
| 26. | faransa | 59. | maridhawa |
| 27. | fasaha | 60. | muhimu |
| 28. | fasihi | 61. | nadhifu |
| 29. | fidhuli | 62. | nadra |
| 30. | ghali | 63. | najisi |
| 31. | haba | 64. | nakawa |
| 32. | hafifu | 65. | rahimu |
| 33. | halifu | 66. | rahisi |
| 34. | halisi | 67. | raufu |
| 35. | hanamu | 68. | safihi |
| 36. | haramu | 69. | safii |
| 37. | hodari | 70. | saghiri |
| 38. | huri | 71. | sahala |
| 39. | huria | 72. | sahihi |
| 40. | huru | 73. | salihi |
| 41. | hututi | 74. | sawa |
| 42. | imara | 75. | sharifu |
| 43. | jinai | 76. | stahili |
| 44. | kabidhi | 77. | sunu |
| 45. | kaidi | 78. | swafi |
| 46. | kamili | 79. | tahafifu |
| 47. | karimu | 80. | tatai |
| 48. | kila | 81. | thabiti |
| 49. | laghai | 82. | wastani |
| 50. | lainifu | 83. | yakini |
| 51. | madhubuti | | |

2.2 Adverbs 53 items

- | | | | |
|-----|--------------|-----|------------|
| 1. | abadan | 28. | kadha |
| 2. | abadi | 29. | kadhaa |
| 3. | afadhali | 30. | kadhaliika |
| 4. | aghalabu | 31. | kadiri |
| 5. | aghlabu | 32. | karibu |
| 6. | asubuhi | 33. | kasi |
| 7. | baada | 34. | kusudi |
| 8. | baadaye | 35. | labda |
| 9. | bado | 36. | makasudi |
| 10. | bure | 37. | mara |
| 11. | dahari | 38. | milele |
| 12. | daima | 39. | mkabala |
| 13. | dawamu | 40. | nomi |
| 14. | dike | 41. | nusura |
| 15. | fauka | 42. | salimini |
| 16. | foko | 43. | sana |
| 17. | ghafila | 44. | sasa |
| 18. | ghalibu | 45. | sawia |
| 19. | hadhara | 46. | siasa |
| 20. | halafa | 47. | tasihili |
| 21. | halafu | 48. | tike |
| 22. | hasa | 49. | wahedu |
| 23. | hima | 50. | wakati |
| 24. | hobelahobela | 51. | walau |
| 25. | husususa | 52. | zaidi |
| 26. | inshallah | 53. | zaman |
| 27. | kabisa | | |

2.3 Conjunctions 28

- | | | | |
|----|-------|-----|-------|
| 1. | aidha | 6. | hata |
| 2. | ama | 7. | ila |
| 3. | au | 8. | ili |
| 4. | bali | 9. | jinsi |
| 5. | hadi | 10. | kabla |

11.	kama	20.	mintarafu
12.	lakini	21.	mradi
13.	maadan	22.	taraa
14.	madhali	23.	tena
15.	mathalan	24.	waima
16.	mazalani	25.	waina
17.	methalan	26.	wala
18.	minajili	27.	walakini
19.	minghairi	28.	yaani

2.4 Interjections 36 items

1.	ahasante	20.	labeka
2.	ahsante	21.	lahaula
3.	ala	22.	laiti
4.	alaala	23.	lebeka
5.	alhamdulillah	24.	marahaba
6.	amin	25.	marhaba
7.	asante	26.	masalala
8.	ashakum	27.	masalale
9.	audhubillahi	28.	masalkheri
10.	basi	29.	naam
11.	bismillahi	30.	ole
12.	ebu	31.	salala
13.	halahala	32.	sefule
14.	hamadi	33.	simile
15.	hasha	34.	wallahi
16.	hashakum	35.	wallai
17.	hebu	36.	yaa
18.	hobe		
19.	huss		

2.5 Numerals 30 items

1.	arobaini	4.	edashara
2.	ashara	5.	elfu
3.	asherini	6.	hamsa
		7.	hamsauishirini

- | | | | |
|-----|-------------|-----|--------------|
| 8. | hamsini | 20. | thelathini |
| 9. | hamstashara | 21. | theluthi |
| 10. | ishirini | 22. | themani |
| 11. | mia | 23. | themanini |
| 12. | miteen | 24. | themanya |
| 13. | saba | 25. | themntashara |
| 14. | sabaini | 26. | themuni |
| 15. | sabini | 27. | thumni |
| 16. | sita | 28. | tisa |
| 17. | sitashara | 29. | tisini |
| 18. | sitini | 30. | wahedi |
| 19. | sumni | | |

2.6 Prepositions – 8 items

- | | | | |
|----|----------|----|--------|
| 1. | baghairi | 5. | badala |
| 2. | fi | 6. | baina |
| 3. | laula | 7. | bila |
| 4. | min | 8. | kuhusu |

2.7 Idiomatic expressions list – 3 items

- | | | | |
|----|----------------|----|--------------|
| 1. | lila na fila | 3. | semaa wa taa |
| 2. | kadha wa kadha | | |

2.8 Separate search : la! Haya!

App. 3 Grammatical items of Arabic origin found in Tanzanian and Zanzibarian corpus

Arabic grammatical items – full list – found in TC and ZC (239 items)							
TC words: 128 tokens 16100 w/t 126 – ZC:							
words 135- tokens 11940 w/t 88							
Alphabetical order				Frequency order			
Tanzanian corpus		Zanzibarian corpus		Tanzanian corpus		Zanzibarian corpus	
adimu	2	adimu	4	kama	2109	kama	1394
afadhali	15	afadhali	25	lakini	1288	lakini	1201
ahsante	3	aghalabu	13	sasa	935	sasa	791

aidha	1	aghlabu	2	hata	899	tena	711
ajabu	80	ahsante	10	sana	758	au	647
ajibu	3	aidha	3	baada	694	hata	561
ala	2	ajabu	83	wakati	625	kila	521
ama	48	ajibu	1	mara	611	mara	455
amin	2	akali	4	tena	606	bila	336
amini	2	ala	16	kila	504	baada	327
arobaini	7	ama	55	zaidi	470	wakati	322
asante	34	amini	2	bila	435	wala	284
ashakum	2	arobaini	2	au	394	basi	240
asubuhi	187	asante	9	hadi	358	zaidi	236
au	394	asubuhi	76	ili	356	karibu	194
aushi	1	au	647	bado	337	bado	187
awali	40	aula	1	basi	311	hasa	177
baada	694	awali	5	kabisa	308	ili	175
baadaye	101	baada	327	karibu	249	labda	171
badala	136	baadaye	117	kabla	243	sana	162
bado	337	badala	26	kuhusu	238	kabisa	154
baina	9	badiri	1	jinsi	194	halafu	141
bali	87	bado	187	asubuhi	187	bali	137
basi	311	bahili	1	wala	147	jinsi	120
bila	435	baina	82	labda	141	baadaye	117
bora	56	bali	137	mradi	138	ila	111
bure	39	basi	240	badala	136	ghafila	103
daima	42	bila	336	ghafila	135	kabla	95
dhaifu	14	bora	75	hasa	112	bure	84
dhalimu	2	bure	84	halafu	104	ajabu	83
duni	12	daima	46	zamani	103	baina	82

ebu	25	dhaifu	14	baadaye	101	zamani	77
elfu	27	dhalimu	2	kadhalika	87	hadi	77
fani	6	dhila	3	bali	87	asubuhi	76
fasihi	4	duni	4	siasa	83	bora	75
ghafla	135	ebu	1	ajabu	80	hebu	61
ghali	1	edashara	1	mia	75	mia	59
haba	8	elfu	28	muhimu	72	ama	55
hadhara	2	fasihi	2	makini	66	sawa	51
hadi	358	fi	4	sita	65	sita	47
hafifu	24	ghafla	103	ishirini	56	daima	46
halafu	104	ghali	3	bora	56	yaani	38
halisi	27	haba	23	kamili	52	saba	35
hamsa	1	hadhara	2	sawa	51	kasi	35
hamsini	21	hadi	77	ama	48	naam	32
haramu	6	hafifu	15	rahisi	46	muhimu	31
hasa	112	halafu	141	kadhaa	43	ina	31
hasha	2	halisi	10	daima	42	kadha	30
hata	899	hamadi	3	saba	41	ishirini	30
hebu	38	hamsini	18	awali	40	hima	30
hima	12	haramu	20	bure	39	elfu	28
hodari	12	hasa	177	hebu	38	badala	26
huria	7	hasha	4	marehemu	37	afadhali	25
huru	20	hashakum	2	asante	34	rahisi	24
husususa	1	hata	561	kasi	33	marehemu	23
ila	15	hebu	61	yaani	32	makini	23
ili	356	hima	30	tisa	30	kamili	23
imara	6	hodari	22	halisi	27	haba	23
ina	16	huria	7	elfu	27	tisa	22

ishirini	56	huru	11	ebu	25	hodari	22
jinsi	194	ila	111	hafifu	24	sahihi	20
kabisa	308	ili	175	sabini	21	haramu	20
kabla	243	imara	11	hamsini	21	mradi	19
kadha	6	ina	31	laiti	20	kusudi	18
kadhaa	43	inshallah	1	huru	20	hamsini	18
kadhalika	87	ishirini	30	themanini	18	kadhalika	16
kadiri	10	jina	1	kusudi	16	ala	16
kama	2109	jinsi	120	ina	16	siasa	15
kamili	52	kabisa	154	ila	15	milele	15
karibu	249	kabla	95	afadhali	15	hafifu	15
kasi	33	kadha	30	sahihi	14	mkabala	14
kila	504	kadhalika	16	dhaifu	14	dhaifu	14
kuhusu	238	kadiri	8	minajili	13	kuhusu	13
kusudi	16	kama	1394	hodari	12	aghalabu	13
labda	141	kamili	23	hima	12	imara	11
laghai	2	karibu	194	duni	12	huru	11
laiti	20	karimu	1	thelathini	11	halisi	10
lakini	1288	kasi	35	kadiri	10	ahsante	10
madhubuti	3	kila	521	baina	9	thelathini	9
mahiri	1	kuhusu	13	yakini	8	sitini	9
mahsusi	4	kusudi	18	tisini	8	nusura	9
mahututi	6	labda	171	sitini	8	nadhifu	9
makini	66	labeka	1	haba	8	madhubuti	9
mara	611	laghai	2	naam	7	laiti	9
marahaba	6	lahaula	2	huria	7	asante	9
marehemu	37	laiti	9	arobaini	7	kadiri	8
maridhawa	2	lakini	1201	marahaba	6	wallahi	7

mia	75	madhali	3	mahututi	6	huria	7
milele	3	madhubuti	9	kadha	6	walakini	6
minajili	13	makini	23	imara	6	thabiti	5
mkabala	2	mara	455	haramu	6	nadra	5
mradi	138	marahaba	2	fani	6	awali	5
muhimu	72	marehemu	23	sawia	5	stahili	4
naam	7	mia	59	ole	5	sharifu	4
nadhifu	5	milele	15	nusura	5	hasha	4
nadra	1	mintarafu	1	nadhifu	5	fi	4
nusura	5	mkabala	14	mahsusi	4	duni	4
ole	5	mradi	19	fasihi	4	akali	4
rahisi	46	muhimu	31	milele	3	adimu	4
raufu	1	naam	32	madhubuti	3	wastani	3
saba	41	nadhifu	9	ajibu	3	sunu	3
sabini	21	nadra	5	ahsante	3	madhali	3
sahihi	14	najisi	1	yaa	2	hamadi	3
salimini	1	nusura	9	walau	2	ghali	3
sana	758	ole	2	theluthi	2	dhila	3
sasa	935	rahisi	24	mkabala	2	aidha	3
sawa	51	saba	35	maridhawa	2	yakini	2
sawia	5	safii	1	laghai	2	walau	2
siasa	83	sahihi	20	hasha	2	sawia	2
sita	65	salala	1	hadhara	2	ole	2
sitini	8	sana	162	dhalimu	2	marahaba	2
sunu	1	sasa	791	ashakum	2	lahaula	2
tena	606	sawa	51	amini	2	laghai	2
thabiti	1	sawia	2	amin	2	hashakum	2
thelathini	11	sharifu	4	ala	2	hadhara	2

theluthi	2	siasa	15	adimu	2	fasihi	2
themanini	18	sita	47	walakini	1	dhalimu	2
tisa	30	sitini	9	thabiti	1	arobaini	2
tisini	8	stahili	4	sunu	1	amini	2
wakati	625	sunu	3	salimini	1	aghlabu	2
wala	147	tena	711	raufu	1	tisini	1
walakini	1	thabiti	5	nadra	1	salala	1
walau	2	thelathini	9	mahiri	1	safii	1
yaa	2	tisa	22	hususu	1	najisi	1
yaani	32	tisini	1	hamsa	1	mintarafu	1
yakini	8	wakati	322	ghali	1	labeka	1
zaidi	470	wala	284	aushi	1	karimu	1
zamani	103	walakini	6	aidha	1	jinai	1
		walau	2			inshallah	1
		wallahi	7			edashara	1
		wastani	3			ebu	1
		yaani	38			bahili	1
		yakini	2			badiri	1
		zaidi	236			aula	1
		zamani	77			ajibu	1

App. 4 Full list of Bantu and Arabic grammatical items from E. Bertoncini list, ordered by rank

	rank	item	gram	glossa
1.	2	na	conj. pr.	and, with
2.	3	h-/h-o	dem. a.	this
3.	4	kwa	prep.	for, etc.
4.	5	katika	prep.	in, at
5.	7	-ake	pos. a.	his
6.	10	kama	conj.	as, that, etc.
7.	12	-angu	pos. a.	my

8.	13	moja	num.	one
9.	15	amba-	rel. pron.	who, which
10.	17	-ote	a.	all
11.	17	-le	dem. a.	that
12.	21	sana	adv.	very much
13.	22	-ingi	a.	many
14.	23	-ingine	a.	other, some
15.	24	hapa/huku/humu	adv.	here, etc.
16.	25	kwamba	conj.	that
17.	27	kila	a.	every
18.	29	-enye	a.	having
19.	31	-ao	pos. a.	their
20.	33	mimi	pron.	I
21.	34	wakati	v.	time
22.	36	ndi-	emphat. copula	
23.	38	hata	conj. pr.	until, etc.
24.	39	-etu	pos. a.	our
25.	40	-kubwa	a.	big
26.	42	lakini	conj.	but
27.	44	kuwa	conj.	that
28.	46	-ako	pos. a.	your
29.	47	-kuu	a.	great
30.	47	-wili	num.	two
31.	49	yeye	pron.	he, she
32.	51	baadaye	adv. pr.	after
33.	52	sasa	adv.	now
34.	57	tokea/(ku)toka	prep.	from
35.	58	-ema	a.	good
36.	58	juu	adv. pr.	above
37.	61	pia	adv.	too
38.	63	wewe	pron.	you
39.	65	kweli	adv.	truth
40.	67	-dogo	a.	small
41.	67	kwanza	adv.	first
42.	69	au	conj.	or
43.	70	mara	adv.	a time, at once
44.	72	tatu	num.	three

45.	76	tena	conj. adv.	then
46.	80	kwenye	prep.	in, at, etc.
47.	80	ili	conj.	in order that
48.	81	leo	adv.	today
49.	82	mbali(mbali)	adv.	far, distinct
50.	84	-o-ote	a.	whatever
51.	84	pamoja	adv.	together
52.	84	mpaka	prep.	limit; up to
53.	84	bila	prep.	without
54.	86	tu	adv.	only
55.	86	wala	conj.	nor
56.	87	basi	int. conj.	well
57.	88	kwangu,...	adv.	at my place, etc.
58.	91	-zuri	a.	nice
59.	91	mbele	adv. pr.	before
60.	93	-enyewe	a.	having
61.	96	ndani	adv. pr.	inside
62.	96	sisi	pron.	we
63.	96	ila	conj.	except
64.	97	chini	adv. pr.	down, under
65.	97	zaidi	adv.	more
66.	99	kati(kati)	adv. pr.	middle, among
67.	100	wao	pron.	they
68.	100	hasa	adv.	exactly
69.	101	kabla	conj. pr.	before
70.	103	kule/pale/mle	adv.	there
71.	104	jana	adv.	yesterday
72.	104	nini?	pron.	what?
73.	104	zamani	adv.	time, past
74.	105	-pya	a.	new
75.	105	kisha	adv.	afterwards
76.	106	ama	conj.	either...or
77.	107	gani	a. int.	what kind of
78.	107	-zee/mzee	a.	old
79.	107	ingawa	conj.	though
80.	107	pili	num.	the second
81.	109	-kali	a.	sharp

82.	109	mbio	adv.	running; fast
83.	109	karibu	adv. pr.	near
84.	110	nyuma	adv. pr.	behind
85.	110	bado	adv.	not yet
86.	110	yaani	conj.	that is
87.	112	-zima	a.	whole, sound
88.	112	kumi	num.	ten
89.	112	nne	num.	four
90.	113	kwani	conj.	because
91.	113	juzi	adv.	the other day
92.	114	mno	adv.	too much
93.	114	nje	adv.	outside
94.	114	tangu	prep.	since
95.	114	kadhalika	adv.	likewise
96.	115	tano	num.	five
97.	115	-enu	pron.	your (pl.)
98.	115	kabisa	adv.	utterly
99.	115	sita	num.	six
100.	116	nani?	pron.	who?
101.	116	hadi	conj. pr.	until, then
102.	117	-baya	a.	bad
103.	117	-chache	a.	few
104.	117	-geni/mgeni	a.	foreign(er)
105.	118	-refu	a.	long, tall
106.	118	-tukufu	a.	glorious
107.	118	pasipo	adv.	without
108.	118	jinsi	conj.	method; how
109.	119	ijapo(kuwa)	conj.	although
110.	119	kadiri	conj.	measure; whilst
111.	120	wazi(wazi)	a.	open, clear
112.	120	ikiwa	conj.	if
113.	120	kesho	adv.	tomorrow
114.	120	asubuhi	adv.	morning
115.	121	kuliko	conj.	than
116.	121	mnamo	prep.	at, etc.
117.	121	wapi?	pron.	where?
118.	122	pengine	adv.	perhaps

119.	122	ewe	pron.	you
120.	123	kadha(a)	adv.	uncertain number
121.	123	wa	conj.	and
122.	123	saba	num.	seven
123.	124	mia	num.	hundred
124.	124	kuhusu	prep.	concerning
125.	125	-gumu	a.	hard
126.	125	elfu	num.	thousand
127.	126	mgonjwa/-gonjwa	a.	sick (person)
128.	127	-eupe	a.	white
129.	127	(ma)kasudi	adv.	
130.	127	bali	conj.	but
131.	128	jani	a.	leaf, grass
132.	128	-zito	a.	heavy
133.	128	-pi?	a. int.	which?
134.	128	je	interrog. part	
135.	129	sawa(sawa)	a.	equal
136.	129	pole(pole)	a. adv.	mild; gently
137.	129	huenda	adv.	possibly
138.	129	ati	interj.	« I say »
139.	129	ninyi	pron.	you (pl.)
140.	129	naam	particle	yes
141.	130	-ovu	a.	bad
142.	130	kumbe!	inter.	what!
143.	130	jioni	adv.	evening
144.	130	bure	adv.	gratis, vainly
145.	130	(h)ebu!	inter.	well then!
146.	130	baina (ya)	prep.	between
147.	131	-ekundu	a.	red
148.	131	lini?	adv	when?
149.	131	kushoto	adv.	left-side
150.	131	milioni	num.	million
151.	131	halafu	adv.	afterwards
152.	131	ghafla	adv.	suddenly
153.	131	badala (ya)	prep.	instead of
154.	132	-eusi	a.	black
155.	132	-fupi	a.	short

156.	132	-tupu	a.	bare
157.	132	yapata	adv.	about
158.	132	nane	num.	eight
159.	132	hima	adv.	quickly
160.	132	tisa	num.	nine
161.	133	-jinga	a.	ignorant
162.	133	kando(kando)	adv.	aside
163.	133	mbona?	adv.	why?
164.	133	siyo	adv.	no
165.	133	iwapo	conj.	in case
166.	133	kimya	adv.	silence
167.	133	afadhali	adv.	better
168.	133	daima	adv.	constantly
169.	133	labda	adv.	perhaps
170.	133	walau	adv.	at least
171.	133	ahsante!	int.	thanks!
172.	133	ishirini	num.	twenty
173.	134	-ngapi?	a. int.	how many?
174.	134	milele	adv.	always
175.	134	mradi	conj.	plan; and so
176.	135	-epesi	a.	quick
177.	135	kienyeji	adv.	in native manner
178.	135	mapema	adv.	early
179.	135	punde	adv.	presently
180.	135	awali	a. adv.	(at) first
181.	135	aidha	conj.	moreover
182.	135	la!	inter.	no!
183.	135	hamsini	num.	fifty

App. 5 Pick list of Bantu grammatical items - From the list contained in E. Bertoncini's work (App. 4; see References)

Bantu grammatical items: 123.

Arabic grammatical items: 60.

5.1 Bantu adverbs pick list – 40 items

Duplicated items have not been inserted as separate form, except for polepole and pole; possessives in inflected locative forms (kwangu, etc.) have been listed separately.

polepole	jana
pole	kisha
lini	mno
hapa	nje
huku	pasipo
humu	pengine
pia	huenda
kwanza	yapata
mbalimbali	kandokando
pamoja	kando
tu	mbona
kwangu	siyo
kwetu	kienyeji
kwenu	mapema
kwake	punde
kwako	juu
kwao	mbele
kule	ndani
pale	chini
mle	nyuma

5.2 Bantu conjunctions pick list – 10 items

kwamba	na
kuwa	
ingawa	
kwani	
ijapokuwa	
ijapo	
ikiwa	
kuliko	
iwapo	

5.3 Bantu prepositions pick list – 8 items

mpaka
kwa
katika
tokea
kutoka
kwenye
tangu
mnamo

App. 6 Lists of Bantu grammatical items found in TC and ZC corpus separated by categories

Tab. 11 Bantu adverbs – 31 items

1.	129	pole(pole)	adv.
2.	131	lini	adv
3.	24	hapa/huku/humu	adv.
4.	61	pia	adv.
5.	67	kwanza	adv.
6.	82	mbali(mbali)	adv.
7.	84	pamoja	adv.
8.	86	tu	adv.
9.	88	kwangu,...	adv.
10.	103	kule/pale/mle	adv.
11.	104	jana	adv.
12.	105	kisha	adv.
13.	114	mno	adv.
14.	114	nje	adv.
15.	118	pasipo	adv.
16.	122	pengine	adv.
17.	129	huenda	adv.
18.	132	yapata	adv.
19.	133	kando(kando)	adv.
20.	133	mbona	adv.
21.	133	siyo	adv.
22.	135	kienyeji	adv.
23.	135	mapema	adv.

24.	135	punde	adv.
25.	58	juu	adv.
26.	91	mbele	adv.
27.	96	ndani	adv.
28.	97	chini	adv.
29.	110	nyuma	adv.
30.	58	juu	adv. pr.
31.	91	mbele	adv. pr.
32.	96	ndani	adv. pr.
33.	97	chini	adv. pr.
34.	110	nyuma	adv. pr.
35.	99	kati(kati)	adv. pr.

Tab. 12 Bantu Conjunctions – 9 items

1.	25	kwamba	conj.
2.	44	kuwa	conj.
3.	107	ingawa	conj.
4.	113	kwani	conj.
5.	119	ijapo(kuwa)	conj.
6.	120	ikiwa	conj.
7.	121	kuliko	conj.
8.	133	iwapo	conj.
9.	2	na	conj.

Tab. 13 Bantu numerals – 9 items

1.	13	moja	num.
2.	47	-wili	num.
3.	72	tatu	num.
4.	107	pili	num.
5.	112	kumi	num.
6.	112	nne	num.
7.	115	tano	num.
8.	131	milioni	num.
9.	132	nane	num.

Tab.14 Bantu prepositions – 7 items

1.	84	mpaka	prep.
2.	4	kwa	prep.
3.	5	katika	prep.
4.	57	tokea/(ku)toka	prep.
5.	80	kwenye	prep.
6.	114	tangu	prep.
7.	121	mnamo	prep.

Tab. 15 Bantu pronouns – 12 items

1.	33	mimi	pron.
2.	49	yeye	pron.
3.	63	wewe	pron.
4.	96	sisi	pron.
5.	100	wao	pron.
6.	104	nini	pron.
7.	115	-enu	pron.
8.	116	nani	pron.
9.	121	wapi	pron.
10.	122	ewe	pron.
11.	129	ninyi	pron.
12.	15	amba-	rel. pron.

Tab. 16 Bantu miscellanea

1.	3	h-/h-o	dem. a.
2.	17	-le	dem. a.
3.	36	ndi-	emphat. Copula
4.	7	-ake	pos. a.
5.	12	-angu	pos. a.
6.	31	-ao	pos. a.
7.	39	-etu	pos. a.
8.	46	-ako	pos. a.
9.	130	kumbe	inter.
10.	129	ati	interj.
11.	128	je	interrog. particle