

## Language Specific Peculiarities Document for SWAHILI as Spoken in KENYA

Swahili is a Bantu language (Contini-Morava, 1997) spoken in much of East Africa, including Kenya, Tanzania, Burundi, Mayotte, Mozambique, Somalia, South Africa, and Uganda (Lewis, 2009). It is an official language of Kenya and Tanzania, as well as a national language of the Democratic Republic of the Congo and Uganda (Lewis, 2009).

### 1. Special handling of dialects

The standardized form of Swahili is based on the *Kiunguja* dialect of Zanzibar, and is used in official and educational domains throughout Kenya and other countries in East Africa. Typical conversational Swahili in Kenya tends to reflect local dialects of Swahili, which vary by region (Lewis, 2009).

Dialect Region	Dialect name
Nairobi region	Kinairobi
Mombasa region	Kimvita
Other regions (generally rural and remote)	various minor dialects e.g. Kibajuni, Kiamu, Kipate

There are a large number of recognized Swahili dialects spoken in Kenya that are generally mutually intelligible. This mutual intelligibility contributes to Swahili's role as a vehicle for inter-ethnic communication in Kenya. Most of these dialects are spoken in sparsely populated rural areas. The largest urban centers in Kenya are Nairobi and Mombasa, each of which has its own recognized local dialect of Swahili and a highly literate speaker base (Kimalu, P., et al., 2001). General patterns of mutual intelligibility notwithstanding, each of these varieties has certain registers or code-switching which may render portions of speech between two individuals using the same variety less intelligible to an outsider from a different variety.

The variant spoken in the Nairobi region is commonly known as *Kinairobi* and is influenced by an urban vernacular known as *Sheng*, which has developed throughout Nairobi (and also other parts of Kenya) since its alleged origins among the youth of Nairobi's Eastlands (Githiora, 2002). Speech collected in the Nairobi region is influenced by *Sheng* and its idiosyncrasies (e.g. code-switching and argot word formation). Though the influence of *Sheng* is a recent sociolinguistic development, it is the subject of extensive and ongoing research (Githiora, 2002; Rudd, 2008; Ogechi, 2009), which is helpful in informing the collection and orthographic standardization of data in Nairobi.

The data collection includes the Nairobi regional dialect. Speech from Mombasa speakers were not collected, as there were issues of mutual intelligibility with the dialect spoken in the larger Nairobi region where code-mixing with *Sheng* was difficult to avoid or control. The remaining

regions were excluded both for issues of mutual intelligibility as well as lack of dialects with a substantial enough population size.

## 2. Deviation from native-speaker principle

Swahili is a lingua franca in East Africa and is spoken by upwards of 120 million people, though estimates vary as the figure is difficult to establish accurately. Of this population, less than 2 million people are native speakers according to conservative estimates (Wald, 1990). The vast majority of speakers in Nairobi speak the language as a lingua franca and not as native speakers (although they are exposed to Swahili essentially from birth). Therefore, the collection will include fluent non-native speakers.

## 3. Special handling of spelling

The *Standard Swahili Dictionary (Kamusi ya Kiswahili Sanifu)* of 2004 by the Institute of Kiswahili Research of the University of Dar es Salaam (TUKI 2004) is the main reference texts for Standard Swahili. However, readers are exposed to many standard lexical variations, e.g., *arobaini~arubaini* 'forty'<sup>1</sup>.

Many English loanwords (e.g. 'waste') are spelled as in English (rather than being nativized to Swahili orthographic practices). Note also that hybrid spellings are attested, with Swahili affixes plus an English root in English (rather than nativized) spelling. For example, the phrase 'kuburn mamovies' is pronounced /" k u . b \_ < A n # m A . " m u . v i s/; the phrase 'zinatake place' could plausibly be pronounced /z I n A " t E k # " p I E j s/ (pronunciations given in X-SAMPA).

## 4. Description of character set used for orthographic transcription

Swahili uses the Basic Latin Unicode range which is U+0041-U+007A. The letters X and Q do not exist in Swahili. The characters é (U+00e9) and á (U+00e1) are also used in some names originating from French or Spanish.

## 5. Description of Romanization scheme

None.

## 6. Description of method for word boundary detection

Word boundaries in the orthography are determined by localization of white spaces.

## 7. Table containing all phonemes in the stipulated notation

The phonemic transcription of the words in this database uses X-SAMPA symbols, which can be found at <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>. The total number of phonemes is 38. There are 33 consonants (including 2 semi-vowels and 4 foreign phonemes) and 5 vowels (Contini-Morava, 1997).

---

<sup>1</sup> native speaker example

### SWAHILI PHONE CHART

TYPICAL SWAHILI CORRESPONDENCE	IPA	SAMPA	EXAMPLE
<b>CONSONANTS</b>			
p	p	p	<b>paa</b>
b	b	b_<	<b>banda</b>
mb	<sup>m</sup> b	m_b	<b>mbeleko</b>
t	t	t	<b>tena</b>
d	d	d_<	<b>dada</b>
nd	<sup>n</sup> d	n_d	<b>ndimu</b>
k	k	k	<b>kaanga</b>
g	ɡ	g_<	<b>gome</b>
ng	<sup>ŋ</sup> g	N_g	<b>ngazi</b>
j	ʃ	J_<	<b>jana</b>
nj	<sup>n</sup> dʒ	n_dZ	<b>njia</b>
f	f	f	<b>fupi</b>
v	v	v	<b>viatu</b>
th	θ	T	<b>thelathini</b>
dh	ð	D	<b>dhahabu</b>
s	s	s	<b>saa</b>
z	z	z	<b>ziwa</b>
sh	ʃ	S	<b>shati</b>
ch	tʃ	tS	<b>chakula</b>
kh	x	x	<b>kheri</b>
gh	ɣ	G	<b>ghali</b>
h	h	h	<b>habari</b>
m	m	m	<b>maji</b>
n	n	n	<b>nuka</b>
ny	ɲ	J	<b>nyasi</b>
ng'	ŋ	N	<b>ng'ombe</b>
n	ɲ	n=	<b>nchi</b>
m	m	m=	<b>mtoto</b>
n(k, g)	ɲ	N=	<b>nge</b>
r	r	r	<b>rafiki</b>
l	l	l	<b>leo</b>
y	j	j	<b>yeye</b>
w	w	w	<b>wiki</b>
<b>VOWELS</b>			
i	i	i	<b>ljumaa</b>
e	ɛ	E	<b>wewe</b>

TYPICAL SWAHILI CORRESPONDENCE	IPA	SAMPA	EXAMPLE
a	ɑ	A	alfajiri
o	ɔ	O	moja
u	u	u	udongo

OTHER SYMBOLS	
“	primary stress
.	syllable break
#	word boundary

## Notes

- Primary stress almost always falls on the penultimate syllable, with the exception of some foreign loanwords.
- There is some variation between /r/ and /l/ in Swahili. They are merged for some speakers.

## 7.1 List of rare phonemes

The following phonemes occur infrequently in Swahili:

IPA	SAMPA
<sup>n</sup> dʒ	n_dZ
θ	T
ð	D
x	x
ɣ	G
ŋ	N
ɲ	n=
ɳ	N=

## 7.2 Regular phonemic variation

The following continuant phonemes, which occur primarily in Arabic loanwords, have regular variant realizations in Swahili (Contini-Morava 1997: 849). These “Swahilizations” are a distinctive feature of Nairobi slang “Sheng” Swahili, as well as other non-standard dialects of Swahili.

While the dorsal phonemes /x/ and /ɣ/ are in free variation with their “Swahilized” counterparts (see table below), the interdental fricatives, /θ/ and /ð/, derived from Classical Arabic emphatic consonants /t/ and /d/ are less stable and have a series of possible permutations (personal communication, Appen linguist-consultant). These are delineated in the table below.

“Swahilized” variant pronunciations of words that contain these fricatives have been automatically generated in the lexicon.

Arabic Fricative		Regular Variant	
IPA	SAMPA	IPA	SAMPA
ɣ	G	ɡ	g_<
x	x	h	h
ð	D	d ~ z	d_< ~ z
θ	T	t ~ s ~ ð	t ~ s ~ D

### 7.3 Foreign language influences

The data contains a large number of borrowings from English (increasingly used in urban contexts) and Arabic (from historical borrowings). The data collection has attempted to constrain the extent of code-mixing and code-switching (between Swahili and English, and potentially other regional languages). Nevertheless, the data does contain a very significant proportion of English loan words, both in isolation and in the context of Swahili sentences; we believe this to be symptomatic of Nairobi Swahili. Note that English loanwords (e.g., warrior, year) are generally pronounced as in British (non-rhotic) dialects. Note also that some orthographic words are ambiguous between native Swahili words and English loanwords. Such homographs can have very different pronunciations: for example, the word ‘lake’ could be pronounced /l a k E/ when it represents the native Swahili word, but /l E k/ when it represents the English loanword. The pronunciation lexicon is not guaranteed to give both pronunciations, nor is it guaranteed to give all pronunciations actually found in the audio data.

## 8. Other Language Items

### 8.1 Table of Digits

Digit	Word
0	sifuri/sufuri
1	moja
2	mbili
3	tatu
4	nne
5	tano
6	sita
7	saba
8	nane
9	tisa

## 8.2 Other Numbers

Digit	Word
10	kumi
20	ishirini
30	thelathini
40	arubaini/arobaini
50	hamsini
60	sitini
70	sabini
80	themanini
90	tisini
100	mia/mia moja
1000	elfu/elfu moja
10,000	elfu kumi
100,000	laki/laki moja
10 million	milioni kumi

## 8.3 Alphabet Pronunciations

Swahili letters are often pronounced as in British English. The table below gives both the British English-like pronunciation and the native Swahili pronunciation for each letter. (Note that the Swahili alphabet does not contain the letters Q and X; however, they may occur in borrowings.)

Letter	British-English Pronunciation	Letter	Native Swahili Pronunciation
A	E j	A	A
B	b_ < i	B	b_ < A
C	s i	CH	tS A
D	d_ < i	D	d_ < A
E	i	E	E
F	E f	F	f A
G	J_ < i	G	g_ < A
H	E j tS	H	h a
I	A j	I	i
J	J_ < E j	J	J_ < A
K	k E j	K	k A
L	E l	L	l A
M	E m	M	m= A
N	E n	N	n= A
O	O	O	O

Letter	British-English Pronunciation	Letter	Native Swahili Pronunciation
P	p i	P	p A
Q	k j u		
R	A r	R	r A
S	E s	S	s A
T	t i	T	t A
U	j u	U	u
V	v i	V	v A
W	d < A . b < i . l j u	W	w A
X	E k s		
Y	w A j	Y	j A
Z	z E d <	Z	z A

## 9. References

- Chuo Kikuu cha Dar es Salaam. Taasisi ya Uchunguzi wa Kiswahili (Institute of Kiswahili Research, University of Dar-es-Salaam) 2006, TUKI, kamusi ya Kiswahili-Kiingereza = TUKI, Swahili-English Dictionary—3rd ed. Taasisi ya Uchunguzi wa Kiswahili, Chuo Kikuu cha Dar es Salaam, Dar es Salaam.
- Contini-Morava, E. 1997, ‘Swahili Phonology’, in Alan S Kaye (ed), *The Phonologies of Asia and Africa: (including the Caucasus) Volume 2*, Eisenbrauns, Winona Lake, Indiana.
- Githiora, C. 2002, ‘Sheng: Peer language, Swahili dialect or emerging creole?’, *Journal of African Studies*, vol.15, no.2, pp.159-181.
- Kenya National Bureau of Statistics 2009, ‘Kenya 2009 Population and Housing: Census Highlights’. Online version: <http://www.scribd.com/doc/36672705/Kenya-Census-2009>
- Kimalu, Paul K., et al. 2001, ‘Educational Indicators in Kenya’, working paper, Kenya Institute for Public Policy Research and Analysis. Online version: [www.kippra.org/docs/WP4.pdf](http://www.kippra.org/docs/WP4.pdf)
- Kusters, C. W. 2003, ‘Swahili’, in *Linguistic complexity: the influence of social change on verbal inflection*, LOT, Utrecht.
- Lewis, M. Paul (ed.), 2009. *Ethnologue: Languages of the World*, Sixteenth edition. Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com>
- Ogechi, N. O. 2009, The language situation in Kenya, paper presented at symposium conducted as part of the IUPUI Fulbright Hays Group Projects at Moi University, Eldoret, Kenya, 9 July - 3 August.
- Rudd, W. 2008, *Sheng: The mixed language of Nairobi*, dissertation, Mucie, Indiana, Ball State University, viewed 15 August 2012, ProQuest Dissertations and Theses.
- TUKI (2004) *Kamusi ya Kiswahili Sanifu—2<sup>nd</sup> ed.* .Oxford University Press, Nairobi.

Wald, B. 1990, 'Swahili and the Bantu Languages', in Bernard Comrie (ed), *The Major Languages of South Asia, the Middle East, and Africa*, Routledge, London.