Internal and External Linguistic Affiliations of Central Kenyan Bantu

A Dissertation Project by Paul Starzmann

The Structure of the Thesis:

- Scientific Context: Genetic Inheritance vs. Areal Diffusion
- Central Kenya: Languages and Dialects, Geography, Social Structure, History, Oral Traditions
- **Quantitative Analysis (Dialectometry)**
- **Qualtitative Analysis**
- Conclusion

1. Scientific Context: Genetic Inheritance vs. Areal Diffusion

From the works of Diedrich Westermann (1927, 1949) to recent studies in areal linguistics (e.g. Dimmendaal 2001, Güldemann 2008, Kießling et al. 2008):
Areal linguistics as a 'critical answer' to the limited tree-model in language classification and the study of African languages

2. An Introduction to the Object of Research: Central Kenyan Bantu (CKB)

Language	Number of Speakers
Gikuyu	7 Mio.
Kamba	4 Mio.
Meru	2 Mio.
Embu/Mbeere	500.000
Tharaka	140.000
Chuka	70.000

Map 1: The Central Kenyan Bantu Languages in the Kenyan Highlands

3. Dialectometrical Analysis of Central Kenyan Bantu

3.1 The Method in a Nutshell

Geometry = measurement of Earth

Dialectometry = measurement of Dialects

The goal of dialectometry is to assess statistically to which extent different dialects share **phonological**, **lexical**, and **morphological features** (and to identify **dialect centers**) – it is a means of synchronic analysis and, therefore, does not by itself provide any (historical) clues to why certain linguistic traits are shared (see 4. Qualitative Analysis).

3.2 The Data

Data are drawn from the literature (e.g. Barlow 1951, Benson 1964, Möhlig 1974) as well as from previously unpublished archives (made accesible by courtesy of Wilhelm Möhlig).

The lexical data: 600 words, 117 locations – more to be elicited in the field between June and August 2012 (Gikuyu, Nida, Gichugu)

3.3 Phonological Dialectometry of CKB

STEP 1: Identification of regular correspondences

The basis for phonological dialectometry is **regular sound correspondence**, through which we can correlate the phoneme systems of all considered dialects,
e.g. the dia-phoneme *R:

Dia-Phoneme	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
*R	ɾ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	l

*Table 1: Phonetic realizations of the dia-phoneme *R*

If at least two cases of recurrent correspondences are identified, they are considered to be proof of regular correspondence in dialectometrical analysis, e.g.

Item	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
to remain	-i.kaɾa	-i.kaɽa	-i.kaɽa	-i.kaɽa	-i.kaɽa	-i.kaɽa	-i.kaɽa	-i.kaɽa	-ɪ.kala
pumpkin	i.ɾɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.ɽɛngɛ	ɪ.lɛngɛ

Table 2: Items 082 and 473 – attestation of regular correspondence

A number of cases show **multiple matches / multiple correspondences,** again the series of the dia-phoneme *R:

Item	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
charcoal	i.kaɾa	ɪ.kaɽa	ɪ.kaɽa	ɪ.kaɽa	ɪ.kaɽa	ɪ.kaɽa	ɪ.kaɽa	ɪ.kaɽa	ma.kaa
to bite	-ɾuma	-ɽuma	-ɽuma	-ɽuma	-ɽuma	-ɽuma	-ɽuma	-ɽuma	-uma

Table 3: Items 214 and 311 – attestation of regular correspondence

➔Recognition of multiple matches:

Dia-Phoneme	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
*R	ɾ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	**l, ø**

Table 4: A modification of table 1 recognizing multiple correspondences

STEP 2: Setting up a system for comparison

In order to compare the phonetic differences between dialects, an adequate system of comparison has to be set up, e.g. *MP

Dia-Phoneme	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
*MP	mb	mb	mb	mp	mp	mp	mp	mp	mb
'cat'	**mb**aka	**mb**aka	**mb**aka	**mp**aka	**mp**aka	**mp**aka	**mp**aka	**mp**aka	**mb**aka
'maize'	**mb**ɛ**mb**ɛ	**mb**ɛ**mb**ɛ	**mb**ɛ**mb**ɛ	**mp**ɛ**mp**a	**mp**ɛ**mp**ɛ	**mp**ɛ**mp**ɛ	**mp**ɛ**mp**ɛ	**mp**ɛ**mp**ɛ	**mb**ɛ**mb**a

*Table 5: Phonetic realization of *MP (attested by items 291 and 406)*

Table 5 shows that *MP in CKB is realized as

mp prenasalized, **voiceless**, bilabial plosive

mb prenasalized, **voiced**, bilabial plosive

The two sounds are only distinguished by the feature [+/- voice]

Feature	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
	mb	mb	mb	mp	mp	mp	mp	mp	mb
voice	+	+	+	-	-	-	-	-	+

*Table 6: Contrastive feature analysis for dia-phoneme *MP*

STEP 3: Counting concurrences

All concurrent features are counted in each series and written down in a distance matrix of the following type:

Embu/Mbeere	1							
Chuka	1	1						
Mwimbi-Muthambi	0	0	0					
Igoji	0	0	0	1				
Miutini	0	0	0	1	1			
Imenti	0	0	0	1	1	1		
Tharaka	0	0	0	1	1	1	1	
Kamba	1	1	1	0	0	0	0	0
	Gikuyu	Embu/Mbeere	Chuka	Mwimbi-Mu.	Igoji	Miutini	Imenti	Tharaka

*Matrix 1: Dialectometrical result for series *MP*

The sum of all matrices renders the dialectometrical overall result:

Ndia A	100																	
Ndia B	96	96																
Gichugu A	99	100	96															
Gichugu B	96	96	100	96														
Embu	64	64	66	64	66													
Mbeere	64	66	68	64	66	99												
Chuka	70	72	72	72	72	78	84											
Muthambi	66	68	68	68	66	73	72	86										
Mwimbi	58	59	58	59	58	70	73	81	92									
Igoji	54	59	54	55	54	68	69	81	91	97								
Miutini	57	58	58	58	58	72	73	77	84	92	93							
Nkubu	61	64	62	64	65	73	73	85	86	84	86	76						
N-Imenti	57	58	59	58	59	65	62	76	81	74	76	76	91					
Tharaka-E	68	70	69	69	69	69	69	80	88	80	78	82	88	88				
Tharaka-W	65	66	66	65	66	65	66	81	85	85	81	81	86	82	96			
Masaku	72	70	72	70	73	70	72	72	69	72	74	73	70	64	74	70		
Yatta	69	69	72	70	72	72	72	72	69	72	72	74	70	62	72	70	100	
Kitui	68	69	70	69	70	69	72	72	68	70	70	73	70	62	70	70	99	100
	Gikuyu	Ndia A	Ndia B	Gichugu A	Gichugu B	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	Tharaka-E	Tharaka-W	Masaku	Yatta

Matrix 2: Preliminary phono-dialectometrical result for all dialects of CKB in %

3.4 Lexical Dialectometry of CKB

The basis for the lexical analysis is a wordlist with 600 items, elicited in a total of 117 locations in Central Kenya = more than 70.000 items – and there are more to come!

STEP 1: Setting up the Excel-database - Substituting word forms with characters

In general, word forms are either **identical**, **partially (i.e. phonologically and/or morphologically) divergent**, or **fully divergent**. Each word form is assigned a symbol (A1, A2, B etc.) when setting up the data base:

Loc.	1a	1b	2	3a	3b	4	5	...	96	97
015	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua		ka.ɲwa	ka.ɲwa

Table 8: Raw data for item 015 'mouth' (excerpt)

Item #015 'mouth'	1.	ka.ɲua	A1
	2.	ka.nua	A2
	3.	ka.ɲwa	A3

A1:A2 = phon. divergence
A1:A3 = phon. divergence
A2:A3 = phon. divergence

Item #025 'left hand'	1.	ʊ.məðɔ	A1
	2.	kɪ.məðɔ	A2
	3.	kw.aka	B

A1:A2 = morph. divergence
A1:B = full divergence
A2:B = full divergence

Item #073 'blister'	1.	kɪ.a:ru	A
	2.	gɪ.tɔ:yɔ	B1
	3.	gʊ.tɔ:ya	B2
	4.	yau	C

A:B1 = full divergence
A:B2 = full divergence
A:C = full divergence
B1:B2 = acc. divergence
B1:C = full divergence etc.

Loc.	1a	1b	2	3a	3b	4	5	...	96	97
015	A1	A1	A1	A1	A1	A1	A1		A3	A3

Table 9: Rendered data (excerpt)

STEP 2: Recoding with R: LexMatrix$_A$

	1a	1b	2	3a	3b	4	5	...	96	97
1a	0	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1		A1:A3	A1:A3
1b	A1:A1	0	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1		A1:A3	A1:A3
2	A1:A1	A1:A1	0	A1:A1	A1:A1	A1:A1	A1:A1		A1:A3	A1:A3
...				0						

Matrix 3: Example of LexMatrix$_A$ for item 015 'mouth' (excerpt)

STEP 3: Automatic replacement with identity values: LexMatrix$_B$

In dialectometry, lexical identity and divergence are rated accoring to the following scale:

Identity	= 4 Points	e.g. A:A, B1:B1
Morphological Divergence	= 3 Points	e.g. A1:A2, B1:B2
Phonological Divergence	= 2 Points	e.g. A1:A2, B1:B2
Accumulated Divergence	= 1 Point	e.g. A1:A2, B1:B2
Full Divergence	= 0 Points	e.g. A:B, B1:C1

In the case of 015 'mouth': A1:A1 = identical (4); A1:A3 = phon. divergent (2)

	1a	**1b**	**2**	**3a**	**3b**	**4**	**5**	**...**	**96**	**97**
1a	0	4	4	4	4	4	4		2	2
1b	4	0	4	4	4	4	4		2	2
2	4	4	0	4	4	4	4		2	2
3a	4	4	4	0	4	4	4		2	2
3b	4	4	4	4	0	4	4		2	2
4	4	4	4	4	4	0	4		2	2
5	4	4	4	4	4	4	0		2	2
...								0		
96	2	2	2	2	2	2	2		0	2
97	2	2	2	2	2	2	2		2	0

Matrix 4: Example of LexMatrix$_B$ for item 015 'mouth' (excerpt)

Again, the sum of all LexMatrices$_B$ renders the overall dialectometrical result on the lexical level!

'Unsuitable concepts' - a challenge for statistical robustness!

For some entries in the 600-wordlist (e.g. 062 'deaf', 071 'to get hurt', 296 'arrow') there are no or only partial data available - these case have to be cleared from the database.

In other cases, there are data missing for specific locations - plausible vaules (arithmetic average) have to be determined for these missing items in order to keep the result statistically robust.

3.5 Morphological Dialectometry

Dialectometrical calculations on the morphological level follow the method of lexical dialectometry, except that it is only distinguished between

Identity = 2 Points
Phonological divergence = 1 Point
Full divergence = 0 Points

The morphemes under consideration (some still to be verified in the field) are: nounclass marker, adjective marker, pronoun marker, subjectconcord, objectconcord

3.6 Expected Results

The dialectometrical method provides specific results (distance matrices) for each linguistic level, phonology, lexicon, and morphology.

With the statistical method of **cluster analysis**, certain dialects / locations can be grouped together according to specific threshold values. These groups constitute **dialect clusters** / **centers of innovation** (areas of relatively **low linguistic variation**) of different 'qualities':

'Quality'	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Phonological proximity	✓	✓	✓	×	✓	×	×
Lexical proximity	✓	✓	×	✓	×	✓	×
Morphological proximity	✓	×	✓	✓	×	×	✓

Table 10: Possibile 'qualities' of dialect centers

Based on the literature and control samples, it is safe to assume that inter alia the following language varieties constitute relatively 'strong' clusters:

GIKUYU-CLUSTER	EMBU/MBEERE-CLUSTER	MERU-CLUSTER	THARAKA-CLUSTER	KAMBA-CLUSTER
Kiambu Murang'a Nyeri Mathira Ndia (?) Gichugu (?)	Embu Mbeere	North-Imenti Nkubu-Imenti	Tharaka-E Tharaka-W	Machakos Kitui Yatta
		MERU-THARAKA-CLUSTER (?)		
...			THARAKA-KAMBA (Machakos)-CLUSTER (?)	

Level of Relation

Table 11: Expected clustering of the dialects in Central Kenya (estimate)

4. Qualtitative Analysis

Note: Dialectometry is a means of synchronical analysis - it shows a synchronic picture of dialectal relations but does not provide any insight into historical processes, i.e. how dialect clusters came into being.

➔ Which factors played a major role in the growth of dialect centers such as the ones shown in table 11?

4.1 Internal Convergence

Haspelmath/Tadmor (2009) show that the tendency to lexical borrowing correlates with semantic domains:

Semantic field	Loanwords as % of total
Religion and belief	41,2
Clothing and grooming	38,6
The house	37,2
Law	34,3
...	...
Kinship	15,0
The body	14,2
Spatial relations	14,0
Sense perception	11,0

Table 12: Excerpt of the loanword-typology (Tadmor 2009: 64)

Control Sample: Kamba (Mumoni, Kitui, Machakos) - Mbeere - Tharaka

Mumoni	845			
Kitui	934	736		
Mbeere	558	343	394	
Tharaka	258	243	194	400
	Machakos	**Mumoni**	**Kitui**	**Mbeere**

Matrix 5: The Body

934 Masaku : Kitui
845 Masaku : Mumoni
763 Mumoni : Kitui
558 Masaku : Mbeere
400 Mbeere : Tharaka
394 Mbeere : Kitui
343 Mbeere : Mumoni
258 Masaku : Tharaka
243 Mumoni : Tharaka
194 Kitui : Tharaka

Mumoni	771			
Kitui	792	845		
Mbeere	546	626	695	
Tharaka	464	604	655	385
	Machakos	**Mumoni**	**Kitui**	**Mbeere**

Matrix 6: The House

845 Mumoni : Kitui
792 Masaku : Kitui
771 Masaku : Mumoni
695 Mbeere : Kitui
655 Kitui : Tharaka
626 Mbeere : Mumoni
604 Mumoni : Tharaka
546 Masaku : Mbeere
464 Masaku : Mbeere
385 Mbeere : Tharaka

4.2 External Convergence

Example 1: Swahili loans

198 'wall'	Swahili *ukuta* >	**ʊkʊta**	Machakos
		ʊvai	Kitui
		rʊðingɔ	Gikuyu, Embu/Mbeere, Chuka, Meru
226 'water pot'	Swahili *mbisu* >	**mbisʊ**	Kitui
		mbisɔ	Machakos
		mʊtʊngi	Yatta, Kitui
		ɲʊngʊ	Gikuyu, Embu/Mbeere, Chuka, Meru, Machakos
265 'field'	Swahili *shamba* >	**samba**	Kitui
	Swahili *kiwanja* >	**kɪwanza**	Machakos, Kitui
		kɪgwanja	Muthambi
		kɪɛːni	Meru, Tharaka
		kɪbaːrɔ	Chuka, Embu/Mbeere

Table 13: Swahili loans in CKB

Example 2: **Maasai loans**

043 'blood'	Maasai -*sàrgɛ́* >	**ðarikɛ**	Meru (Imenti)
		ðarigi	Gikuyu
		ndamu (Sw.)	Miutini, Igoji, Mwimbi-Muthambi
		nðakamɛ	Mwimbi-Muthambi, Chuka, Embu/Mbeere, Tharaka, Kamba
052 'to bathe'	Maasai *a-él* >	**-ɪːria**	Meru (Imenti)
		-ɪːcɪria	Tharaka
		-ðamba	Embu/Meere, Chuka, Mwimbi-Muthambi, Igoji, Miutini, Kamba
		-ɪðamba	Gikuyu
273 'to pluck'	Maasai a-soló >	**-ðʊrania**	Meru (Imenti)
		-tua/-twa	Embu/Meere, Chuka, Mwimbi-Muthambi, Igoji, Miutini, Tharaka, Kamba

Table 14: Maasai loans in CKB

4.3 School Languages

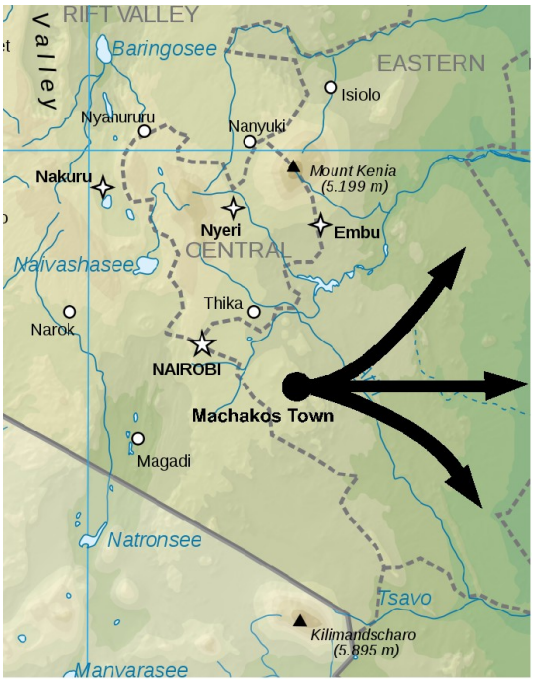
The Kamba dialect of Machakos constitutes a relatively 'strong' lexical cluster. The people of the area were among the first to recieve state-sponsored education in 1914 (Sheffield 1971).

➔Did school education in Kamba contribute to the emergence of the Machakos-dialect?

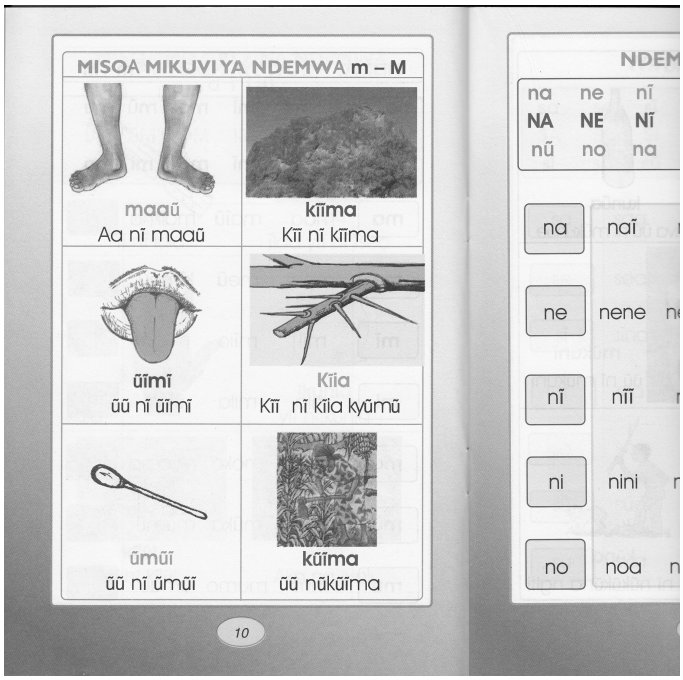
Some items suggest that there has been lexical diffusion from Machakos throughout the whole Kamba language area, e.g. 017 'tongue':

Dialect	Wordform	Occurrence
Machakos	ʊ.ɪmɪ A1	90%
	ʊ.ɪmɛ A2	10%
Yatta	ʊ.ɪmɪ A1	50%
	ʊ.ɪmɛ A2	50%
Kitui	ʊ.ɪmɪ A1	54%
	ʊ.ɪmɛ A2	21%
	w.ɪmɛ A3	17%
	w.ɪmi A4	8%

Table 15: Distribution of 017 'tongue' in Kamba



Map 2: Possible spread of 'ʊ.ɪmɪ' from Machakos into the rest of Ukambani through school education



'tongue' in a Kamba school book (Watuma 2004)

4.4 Genetic Inheritance?

The question remains: How important is genetic inheritance for the emergence of dialect clusters? And how can we relate CKB to the wider Bantu context?

CB	Gikuyu	Embu/ Mbeere	Chuka	Mwimbi-Muthambi	Igoji	Miutini	Imenti	Tharaka	Kamba
*-pákà CS 1420	**mb**aka	**mb**aka	**mb**aka	**mp**aka	**mp**aka	**mp**aka	**mp**aka	**mp**aka	**mb**aka
*-pémbá CS 1475	**mbɛmbɛ**	**mbɛmbɛ**	**mbɛmbɛ**	**mpɛmpa**	**mpɛmpɛ**	**mpɛmpɛ**	**mpɛmpɛ**	**mpɛmpɛ**	**mbɛmba**

Table 16: Items 291 'cat' and 406 'maize' in relation to Common Bantu CS 1420 and CS 1475

5. References

Barlow, A. Ruffel (1951). *Studies in Kikuyu grammar and idiom*. Edinburgh: Blackwood & Sons.

Benson, Thomas Godfrey (1964). *Kikuyu-English dictionary.* Oxford: Clarendon Press.

Dimmendaal, Gerrit (2001). Areal Diffusion and Genetic Inheritance: An African Perspective. In: *Areal Diffusion and Genetic Inheritance – Problems in Comparative Linguistics*, ed. by Alexandra Aikhenvald and Robert Dixon. Oxford: Oxford University Press. 358-392.

Güldemann, Tom (2008). The Macro-Sudan belt: towards identifying a linguistic area in northern sub-Saharan Africa. In: *A Linguistic Geography of Africa*, ed. by Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press. 151-185.

Haspelmath, Martin & Uri Tadmor (2009). *Loanwords in the World's Languages: A Comparative Handbook.* Berlin: Mouton de Gruyter.

Kießling, Roland, Maarten Mous and Derek Nurse (2008). The Tanzianian Rift Valley area. In: *A Linguistic Geography of Africa*, ed. by. Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press.186-227.

Möhlig, Wilhelm J.G. (1974). *Die Stellung der Bergdialekte im Osten des Mt. Kenya – Ein Beitrag zur Sprachgliederung im Bantu.* Berlin: Reimer Verlag.

Sheffield, James (1971). *Education in the Republic of Kenya*. Washington: US Dept. of Health, Education, and Welfare.

Watuma, Blanche 2004. *Kikamba kusoma kwa mbee*. Nairobi: Sengani Publishers.

Westermann, Diedrich (1927). Die westlichen Sudansprachen und ihre Beziehungen zum Bantu. Beiheft zu *MSOS*, Jahrgang 30.

Westermann, Diedrich (1949). Sprachbeziehungen und Sprachverwandtschaft in Afrika. *Sitzungsberichte der Deutschen Akademie der Wissenschaften zu Berlin*, Jahrgang 1948, Phil-hist. Klasse Nr. 1. Berlin.