HU Berlin
Institut für Asien- u. Afrikawissenschaften
Seminar für Afrikawissenschaften

Linguistik-Kolloquium
28. Januar 2014
Paul Starzmann

A Dialectology of Central Kenyan Bantu: Quantitative and Qualitative Analysis

0. Introductory Remarks: The PhD-Project in a Nutshell

<u>"Internal and External Linguistic Affiliations of Central Kenyan Bantu"</u>

- Full dialectological survey of Central Kenyan Bantu ⇨ Identifiying 'dialect clusters'

- Historical interpretation ⇨ Explaining the emergence of dialect clusters

- Connecting linguistic and extra-linguistic evidence ⇨ Towards a 'grand scenario'

  In short: **Where** is there little variation? And **why** is there little variation?

CENTRAL KENYAN BANTU (CKB)

Gikuyu	Kamba	Meru	Embu/Mbeere	Tharaka	Chuka
Kiambu	Masaku	Imenti	Embu	Tharaka-East	Chuka
Murang'a	Yatta	Nkubu	Mbeere	Tharaka-West	
Nyeri	Kitui	Miutini			
Mathira		Igoji			
Ndia		Mwimbi			
Gichugu		Muthambi			

The outline of the thesis:

1. Introduction: The Scientific Context

2. The Extra-Linguistic Evidence

3. **Quantitative Analysis**

4. **Qualitative Analysis**

5. Conclusion

The outline of this talk:

1. Scientific & Historical Context

2. Quantitative Analysis

   - Method & Data

   - Phonology

   - Noun Morphology

   - Lexicon

3. Qualitative Analysis

   - Across categories

   - Phonology

   - Noun Morphology

   - Lexicon

4. Summary & Outlook

1. Scientific and Historical Context

| Linguistic Congruence in Historical Linguistics	
Divergence	**Convergence**
Genetic Inheritance	(Areal) Diffusion
Linguistic congruence is due to shared innovation / retention, e.g. the family-tree model	Linguistic congruence may be due to language contact, e.g. the stratification model

▸ Especially in Bantu history, language contact has played a major role (Möhlig 1979, 1981).

▸ In order to shed light on this history, any model and method applied need to take linguistic convergence into account.

The Extra-Linguistic Evidence: The History of Central Kenya

The oral traditions of the region suggest a classical contact scenario:
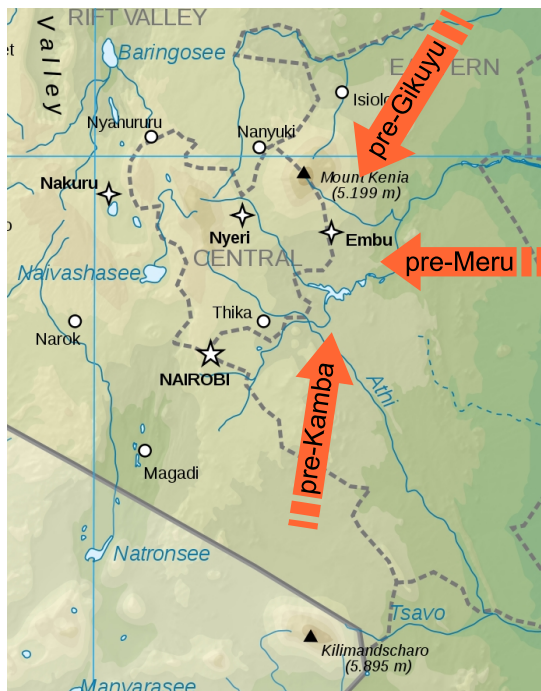
Map 1: The three major migration routes into CK
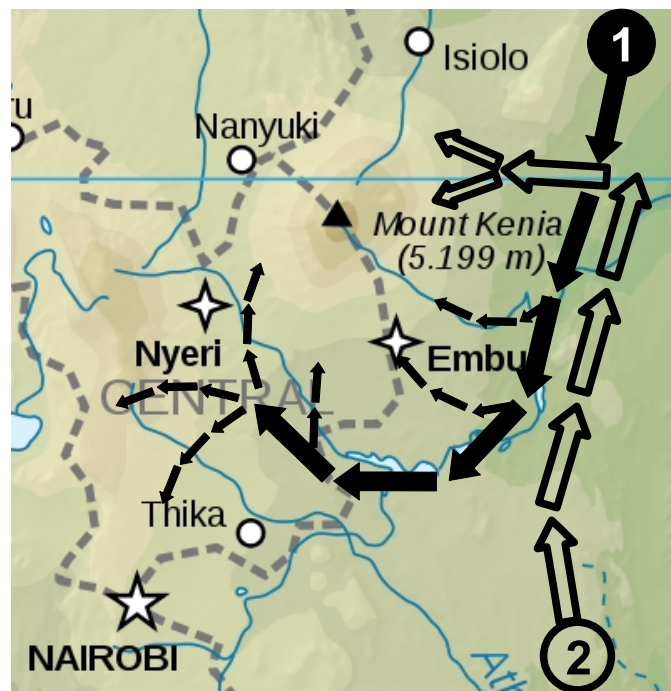
Map 2: Pre-Gikuyu (1) and Pre-Meru (2) migration within the Kenyan Highlands (ca. 1500-1900 AD)

Note: At the time of initial immigration, there was **no ethnic identity** among the early pioneers as we know it today. The movements were spearheaded by small groups on the **family level**. Throughout time, the different sections of population engaged in **trading** and **marriage relations** as well as **military conflicts** as different social, economic, and military alliances were established throughout the centuries.

▸ Oral Traditions paint a picture of **social and cultural interdependence > convergence!**

2. Quantitative Analysis

2.1 Method and Data

The Method of Dialectometry = measurement of dialects

= statistical assessment of the phonological, lexical, and morphological **proximity** between dialects on the synchronic level carried out through **pair-comparison**, e.g.:

Dialect A : Dialect B Dialect A : Dialect C Dialect A : Dialect D	Dialect B : Dialect C Dialect B : Dialect D	Dialect C : Dialect D

For example, the fictitious dialects A, B, C, and D are compared in regard to a feature x:

	Dialect A	Dialect B	Dialect C	Dialect D
feature x	+	-	+	-

Table 1: Distribution of feature x *in the dialects A, B, C, and D*

If two dialects concur (both show either + or -), they are counted as 1; if they disagree, the relationship between two dialects is counted as 0 ⇨ a similarity matrix can be set up:

Dialect A	0			
Dialect B	0	0		
Dialect C	1	0	0	
Dialect D	0	1	0	0
	Dialect A	Dialect B	Dialect C	Dialect D

Matrix 1: Similarity Matrix showing the affiliations between A, B, C, and D in regard to feature x

▸ The sum of all similarity matrices renders the overall dialectometrical result.

Note: In the above example, it is assumed that linguistic variation is **binary**. This holds for phonological differences, while morphological and lexical variation may be **gradual** ⇨ in the latter two, it is genearlly distinguished between (1.) identity, (2.) partial divergence, and (3.) full divergence (see below).

- published (Möhlig 1974) and archival[1] material as well as my own elicitations (conducted in the field in the summer of 2012)
- Elicitation of a 600-wordlist in a total of 127 locations in Central Kenya since 1970; 104 entries have proven to be unusable for comparison > 496 lexical items compared
- The lexical data base comprises almost 63,000 tokens

  = 110 pages or more than 8m^2 of data!

Data-Mining: Multidimensional Scaling (MDS)

Dialectometrical results are represented in a **similarity matrix** (see Matrix 1 above) that depicts the proximity between dialects, not unlike a distance[2] matrix commonly known from geographic road maps, e.g.:

Berlin	0				
Frankfurt	548	0			
Hamburg	289	493	0		
Köln	576	195	427	0	
München	586	392	776	577	0
	Berlin	Frankfurt	Hamburg	Köln	München

Matrix 2: Distances between five German cities (in km)

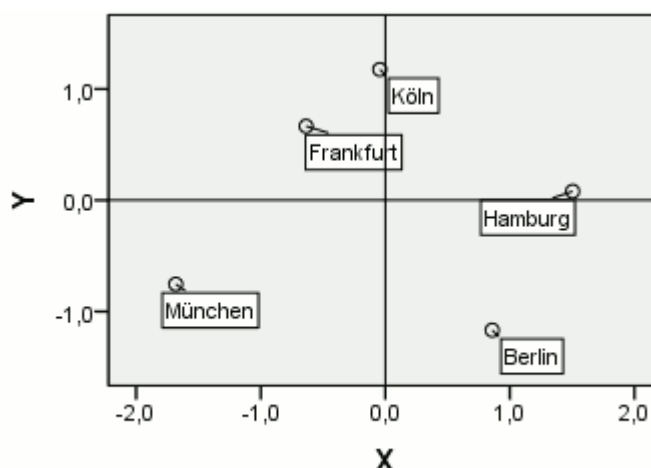By means of multidimensional scaling, the distances above can be represented in a two-dimensional space:

Figure 1: Multidimensional Scaling of Matrix 2 (diagram licensed under public domain)

---

1   The Kamba data are provided by courtesy of Wilhelm Möhlig (University of Cologne), who kindly granted me access to his archives.

2   In a distance matrix, high values represent low distance, while low values represent high distance; in a similarity matrix, on the other hand, high values represent low distance. The latter may be converted into the former by substituting reciprocal values (a number which yields 1 when multiplied by *x*; reciprocal values are written as 1/x).

2.2 Phonological Dialectometry: Measuring phonological distance

Phonological dialectometry measures the **phonetic differences** between dialects by applying the method of *feature analysis* (Jakobson et al. 1952, Chomsky & Hall 1968).

MERU	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Voiceless stops			/t/			/k/	
Voiced stops	/b/					/g/	
Prenasalized voiced stops	/mb/		/nd/			/ng/	
Prenasalized voiceless stops	/mp/		/nt/			/nk/	
Affricate			/c/				
Fricatives		/ð/	/j/				/ɦ/
Prenasalized voiced fricatives		/nð/	/nj/				
Prenasalized voiceless fricatives			/nc/				
Flap				/r/			
Nasals	/m/		/n/		/ɲ/	/ŋ/	

Table 2: The consonant system of Meru (Möhlig 1974: 77)

'Phoneme decay'

EMBU/MBEERE	Labial	Dental	Alveolar	Retroflex	Palatal	Velar	Glottal
Voiceless stops			/t/			/k/	
Voiced stops	/b/					/g/	
Prenasalized stops	/mb/		/nd/			/ng/	
Affricate			/c/				
Fricatives		/ð/					/ɦ/
Prenasalized fricatives		/nð/	/nj/				
Flap				/r/			
Nasals	/m/		/n/		/ɲ/	/ŋ/	

Table 3: The consonant system of Embu and Mbeere (Möhlig 1974: 81)

		Meru (Imenti-Dialect)	**Embu / Mbeere**
/c/	realized as	dʃ = voiced alveo-prepalatal affricate	ʃ = voiceless prepalatal fricative
/c/_/i, u/	realized as	dʃ = voiced alveo-prepalatal affricate	tʂ = voiceless addental postalveolar affricate

Table 4: Two examples of phonetic differences between Meru and Embu / Mbeere

▸ For the purpose of systematic comparison, all phoneme systems under scrutiny are correlated through **regular sound correspondence**, e.g.

020 'neck'	**nk**iːngɔ	(Chuka, Meru, Tharaka)
	ngiːngɔ	(Gikuyu, Embu, Mbeere, Kamba)

045 'heart'	**nk**ɔrɔ	(Chuka, Meru, Tharaka)
	ngɔrɔ	(Gikuyu, Embu, Mbeere)
	ngɔɔ	(Kamba)

Table 5: 'neck' and 'heart' in Central Kenyan Bantu

If at least two cases of recurrent correspondence are identified, they are considered proof of regular correspondence in dialectometrical analysis > a **dia-phoneme-series** can be constituted, e.g. *NK.

Table 4 shows that in CKB the dia-phoneme *NK is realized as

nk   prenasalized, **voiceless,** velar plosive

ng   prenasalized, **voiced**, velar plosive

Dia-Phoneme	Feature	GIKUYU						EMBU/MBEERE			MERU						THARAKA		KAMBA		
		Kiambu	Muaraŋa	Nyeri	Mathɪra	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
*NK	[voice]	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	+

*Table 6: Feature Analysis of dia-series *NK*

Note: The method of dialectometry is a strictly **synchronic** analysis. Therefore, **'multiple matches'** must be treated accordingly, e.g. Tharaka vs. Kamba:

*R₁ > /Ø/ in Kamba		
019 'throat'	mʊ.mɛrɔ	Tharaka
	mʊ.mɛɔ	Kamba
021 'shoulder'	gɪ.turɔ	Tharaka
	kɪ.tuɔ	Kamba

*Table 7: Attestations of *R₁*

*R₂ > /l/ in Kamba		
016 'lip'	mʊ.rɔmɔ	Tharaka
	kɪ.lɔmɔ	Kamba
082 'to remain'	-kara	Tharaka
	-ɪ.kala	Kamba

*Table 8: Attestations of *R₂*

Dia-Phoneme	Feature	GIKUYU						EMBU/MBEERE			MERU						THARAKA		KAMBA		
		Kiambu	Muaraŋa	Nyeri	Mathɪra	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
*R₁		ɾ	ɾ	ɾ	ɾ	ɾ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	Ø	Ø	Ø
	back	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	na	na	na
*R₂		ɾ	ɾ	ɾ	ɾ	ɾ	ɾ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	ɽ	l	l	l
	stop	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-
	back	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	-	-	-

*Table 9: Dia-Series *R₁ and *R₂ in Central Kenyan Bantu*

- ▸ A total of 42 *dia-series* has been established
- ▸ 12 of these series show no variation and are considered non-diagnostic > they have been disregarded in the dialectometrical calculations
- ▸ 95 *feature series* are compared (i.e. the phonological database comprises 95 rows)

STEP 1: Recoding (converting the data table into matrices)

Dia-Phoneme	Feature	Kiambu	Muaɲa	Nyeri	Mathira	Ndia	Gichugu	
*NK	[voice]	+	+	+	+	+	+	...

*Table 9: Example of raw data (Excerpt: *NK in Gikuyu)*

Kiambu	0	+ : +	+ : +	+ : +	+ : +	+ : +	
Muraŋa	+ : +	0	+ : +	+ : +	+ : +	+ : +	
Nyeri	+ : +	+ : +	0	+ : +	+ : +	+ : +	
Mathira	+ : +	+ : +	+ : +	0	+ : +	+ : +	
Ndia	+ : +	+ : +	+ : +	+ : +	0	+ : +	
Gichugu	+ : +	+ : +	+ : +	+ : +	+ : +	0	
...							
	Kiambu	Muraŋa	Nyeri	Mathira	Ndia	Gichugu	...

*Matrix 3: Recoded data for *NK [+/- voice] (Excerpt: Gikuyu)*

STEP 2: Evaluation of concurrences

$$+ : + \quad = 1$$
$$- : - \quad = 1$$
$$+ : - \quad = 0$$

Kiambu	0	1	1	1	1	1	
Muraŋa	1	0	1	1	1	1	
Nyeri	1	1	0	1	1	1	
Mathira	1	1	1	0	1	1	
Ndia	1	1	1	1	0	1	
Gichugu	1	1	1	1	1	0	
...							
	Kiambu	Muraŋa	Nyeri	Mathira	Ndia	Gichugu	...

*Matrix 4: Similarity matrix for *NK [+/- voice] (Excerpt: Gikuyu)*

STEP 3: Adding all matrices and tracking frequency

Kiambu	0	78	95	87	71	
Muraŋa	78	0	78	86	88	
Nyeri	95	78	0	87	71	
Mathira	87	86	87	0	79	
Ndia	71	88	71	79	0	
...						
	Kiambu	Muraŋa	Nyeri	Mathira	Ndia	...

Matrix 5: Sum-matrix showing absolute similarities of the Gikuyu dialects (excerpt)

Kiambu	95	95	95	95	95	
Muraŋa	95	95	95	95	95	
Nyeri	95	95	95	95	95	
Mathira	95	95	95	95	95	
Ndia	95	95	95	95	95	
...						
	Kiambu	Muraŋa	Nyeri	Mathira	Ndia	...

Matrix 6: Frequency matrix showing numbers of occurrences

‣ The sum matrix divided by the frequency matrix yields the overall result showing

relative similarities:

West

East

		Kiambu	Muraŋa	Nyeri	Mathira	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
Gikuyu	Kiambu	0																			
	Murang'a	0,82	0																		
	Nyeri	1	0,82	0																	
	Mathira	0,92	0,91	0,92	0																
	Ndia	0,75	0,93	0,75	0,83	0															
	Gichugu	0,87	0,88	0,78	0,79	0,83	0														
Embu/ Mbeere	Embu	0,6	0,55	0,6	0,56	0,62	0,62	0													
	Mbeere	0,62	0,57	0,62	0,58	0,64	0,64	0,98	0												
Chuka	Chuka	0,67	0,58	0,67	0,63	0,63	0,63	0,78	0,8	0											
	Muthambi	0,66	0,48	0,66	0,58	0,54	0,58	0,66	0,68	0,84	0										
	Mwimbi	0,65	0,52	0,65	0,57	0,57	0,61	0,65	0,67	0,81	0,97	0									
Meru	Igoji	0,63	0,56	0,63	0,59	0,61	0,61	0,67	0,69	0,83	0,88	0,92	0								
	Miutini	0,57	0,54	0,57	0,57	0,59	0,55	0,59	0,61	0,71	0,76	0,79	0,87	0							
	Nkubu	0,65	0,58	0,65	0,61	0,63	0,56	0,69	0,72	0,85	0,82	0,79	0,83	0,75	0						
	N-Imenti	0,65	0,52	0,65	0,61	0,57	0,57	0,63	0,65	0,79	0,82	0,79	0,83	0,75	0,92	0					
Tharaka	E-Tharaka	0,65	0,47	0,65	0,57	0,53	0,57	0,61	0,63	0,75	0,8	0,77	0,75	0,71	0,81	0,81	0				
	W-Tharaka	0,59	0,52	0,59	0,55	0,57	0,57	0,63	0,65	0,83	0,74	0,71	0,75	0,66	0,85	0,77	0,89	0			
Kamba	Masaki	0,47	0,55	0,47	0,52	0,62	0,52	0,52	0,54	0,51	0,41	0,44	0,45	0,47	0,46	0,38	0,37	0,41	0		
	Yatta	0,47	0,55	0,47	0,52	0,62	0,52	0,52	0,54	0,51	0,41	0,44	0,45	0,47	0,46	0,38	0,37	0,41	1	0	
	Kitui	0,47	0,55	0,47	0,52	0,62	0,52	0,52	0,54	0,51	0,41	0,44	0,45	0,47	0,46	0,38	0,37	0,41	1	1	0

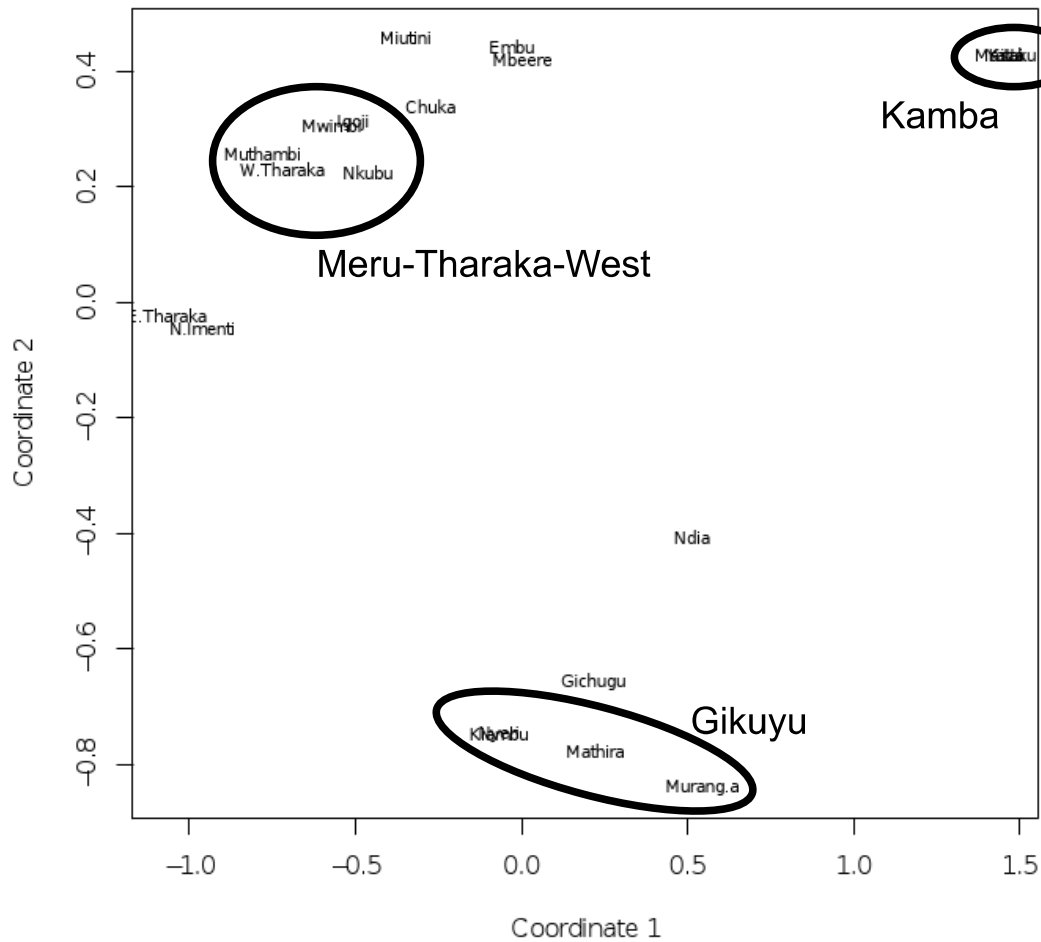*Matrix 6: Relative phonological similarities between all dialects of CKB (overall result)*

Figure 2: Phonological distances between the dialects of CKB

Summing up: What is measured by phono-dialectometry?

(a) **Phonetic differences**, e.g. [+voice] versus [-voice]

(b) **Phonological differences**: 'Phoneme decay'

Items 020 'neck' and 045 'heart' are attestations of dia-series *NK:

*NK > nk (Chuka, Meru, Tharaka)

> ng (Gikuyu, Embu, Mbeere, Kamba)

Items 030 'back' and 475 'many' are attestations of dia-series *NG:

030 'back'   -(g)ɔ**ng**ɔ   (all of CKB)

475 'many'   -i**ng**ɪ   (all of CKB)

Dia-Phoneme	Feature	GIKUYU						EMBU/ MBEERE			MERU						THARAKA		KAMBA		
		Kiambu	Muanja	Nyeri	Mathira	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
*NK	[voice]	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	+
*NG	[voice]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

*Table 10: The merger of *NK and *NG in Gikuyu, Embu/Mbeere, and Kamba*

9

(c) Rule-based differences

Möhlig (1974: 81) states that in Embu the dia-phoneme *MB is realized as [mv] before /i/ and /u/. The rule *MB/_/i, u/ > [mv] sets Embu apart from all other CKB-dialects:

Dia-Phoneme	GIKUYU						EMBU/ MBEERE			MERU						THARAKA		KAMBA		
	Kiambu	Muaŋa	Nyeri	Mathıra	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
*MB/_/i, u/	mb	mb	mb	mb	mb	mb	**mv**	mb	mb	mb	mb	mb	mb	mb	mb	mb	mb	mb	mb	mb

*Table 11: Dia-Series *MB/_/i, u/*

2.3 Morphological Dialectometry: Noun Morphology

The measurement of morphological differences follows the dialectometrical principles described above. In this study, the dialectal differences in the following systems are measured:

- nominal markers
- adjective markers
- subject markers
- object markers
- pronoun markers

In contrary to phonological dialectometry (binary differences), the evaluation of morphological differences requires a more elaborated scale (tertiary differences). It is generally distinguished between (1.) **identity**, (2.) **partial divergence**, and (3.) **full divergence**, e.g. Class 2 in Chuka and Mwimbi (Meru):

	Noun	Adjective	Subjectmarker	Objectmarker
Chuka	a-	a-	ma-	-ma-
Mwimbi	a-	ba-	ba-	-ba-
	identical	**partially div.**	**partially div.**	**partially div.**
	2 Points	1 Point	1 Point	1 Point

Table 12: Class 2 in Chuka and Mwimbi

Note: Any differences in the noun class system that are based on (regular) phonological differences are disregarded in morphological dialectometry in order to avoid 'data inflation':

Dia-Series 5a. *R₁ /_/a, ε, ɪ, ɔ, ʊ/ > /Ø/ in Kamba

      Class 11   *rʊ-*   all of CKB except Kamba

      Class 11   *ʊ-*   all of Kamba

   ▸ Class 11   *Rʊ-*   all of CKB

2.4 Lexical Dialectometry

Again, lexical dialectometry follows the principles described above. It is distinguished between

- **identity**

- **partial divergence**

  (a) **morphological divergence**

  (b) **phonological divergence**

  (c) **morphological _and_ phonological divergence**

- **full divergence**

STEP 1: Converting raw language data

Loc.	1a	1b	2	3a	3b	4	5	...	104	105
	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua	ka.ɲua		ka.nua	ka.nua

Table 13: Raw data for item 015 'mouth' (excerpt)

Item #015 'mouth'	1.	ka.ɲua	A1	**A1:A2 = phon. divergence**
	2.	ka.nua	A2	**A1:A3 = phon. divergence**
	3.	ka.ɲwa	A3	**A2:A3 = phon. divergence**

Item #025 'left hand'	1.	ʊ.mɔðɔ	A1	**A1:A2 = morph. divergence**
	2.	kɪ.mɔðɔ	A2	**A1:B = full divergence**
	3.	kw.aka	B	**A2:B = full divergence**

Item #073 'blister'	1.	kɪ.aːru	A	**A:B1 = full divergence**
	2.	gɪ.tɔːyɔ	B1	**A:B2 = full divergence**
	3.	gʊ.tɔːya	B2	**A:C = full divergence**
	4.	yau	C	**B1:B2 = acc. divergence**
				B1:C = full divergence etc.

Loc.	1a	1b	2	3a	3b	4	5	...	104	105
	A1	A1	A1	A1	A1	A1	A1		A2	A2

Table 14: Rendered data for item 015 'mouth' (excerpt)

STEP 2: Recoding with R > **LexMatrix$_A$**

	1a	1b	2	3a	3b	4	5	...	104	105
1a	0	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1		A1:A2	A1:A2
1b	A1:A1	0	A1:A1	A1:A1	A1:A1	A1:A1	A1:A1		A1:A2	A1:A2
2	A1:A1	A1:A1	0	A1:A1	A1:A1	A1:A1	A1:A1		A1:A2	A1:A2
...				0						

Matrix 7: LexMatrix$_A$ für item 015 'mouth' (excerpt)

STEP 3: Evaluationg lexical differences > **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: A2 = phonologically divergent (2)

	1a	1b	2	3a	3b	4	5	...	104	105
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
...								0		
104	2	2	2	2	2	2	2		0	2
105	2	2	2	2	2	2	2		2	0

Matrix 8: LexMatrix_B for item 015 'mouth' (excerpt)

Note: Again, differences in the lexicon that are based on regular phonological and / or morphological differences are disregarded in order to avoid 'data inflation', e.g.:

Dia-Series 5a. *R_1 /_/a, ɛ, ɪ, ɔ, ʊ/ > /Ø/ in Kamba

Item 137 'to cry' -rɪra all of CKB except Kamba

-ɪa Kamba — both forms are treated as regular / identical

STEP 4: Adding all LexMatrices_B and tracking frequency

13	0	2025	1984	1933	1912	
14	2025	0	2005	1924	1911	
15	1984	2005	0	1926	1925	
16a	1933	1924	1926	0	2013	
16b	1912	1911	1925	2013	0	
...						
	13	14	15	16a	16b	...

13	496	496	496	492	492	
14	496	496	496	492	492	
15	496	496	496	492	492	
16a	492	492	492	496	492	
16b	492	492	492	492	496	
...						
	13	14	15	16a	16b	

Matrix 9: Sum matrix showing the absoulte similarities between locations 13 - 16b (Igoji)

Matrix 10: Frequency matrix showing the number of occurrences (i.e. number of compared items)

▸ The frequency matrix allows us to maintain **statistical robustness** in spite of 431 missings in the raw data base, e.g., in the case of 16a : 16b only 492 out of 496 items can be compared due to 4 missing entries in 16b.

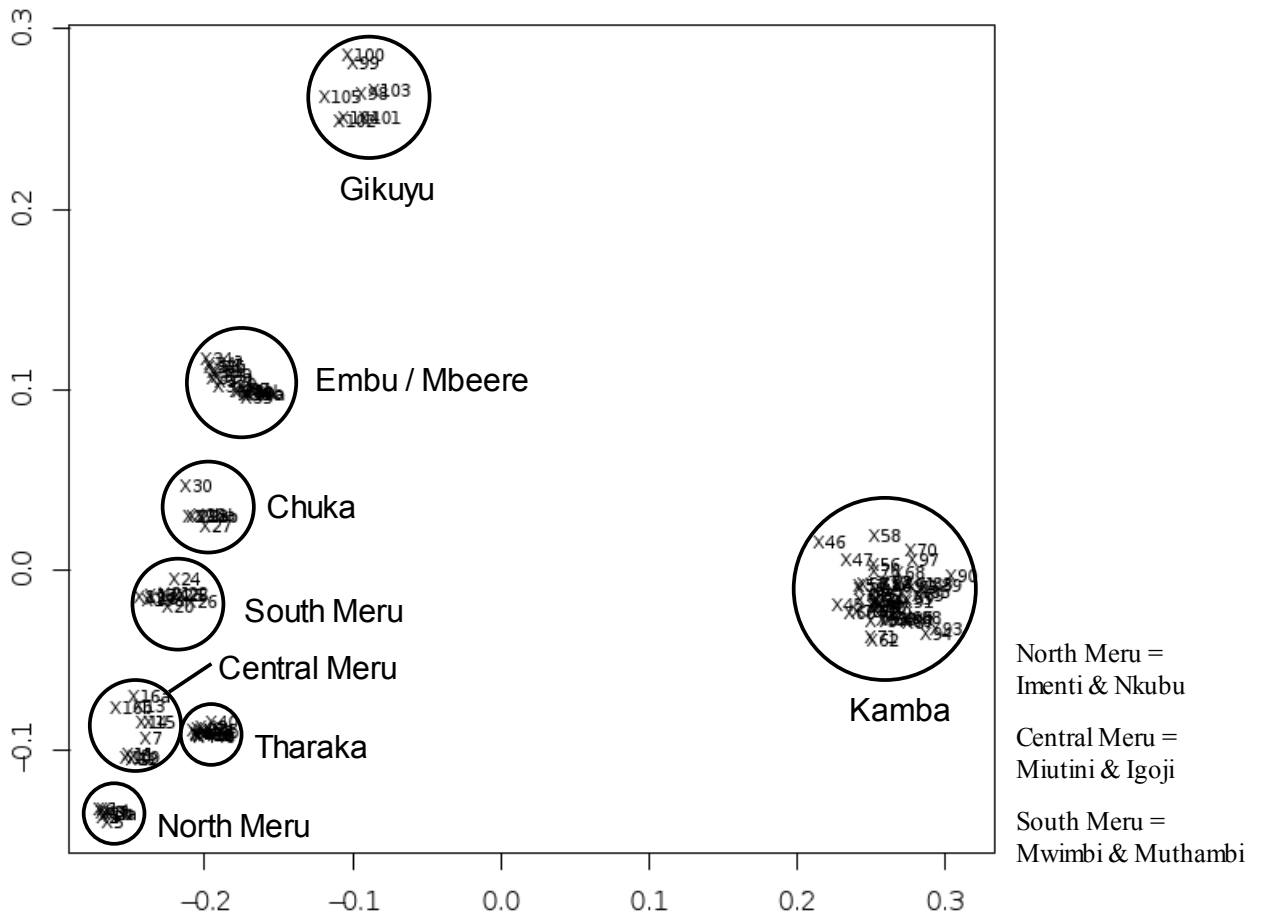▸ The sum matrix divided by the frequency matrix yields the overall result (**rel. similarity**).
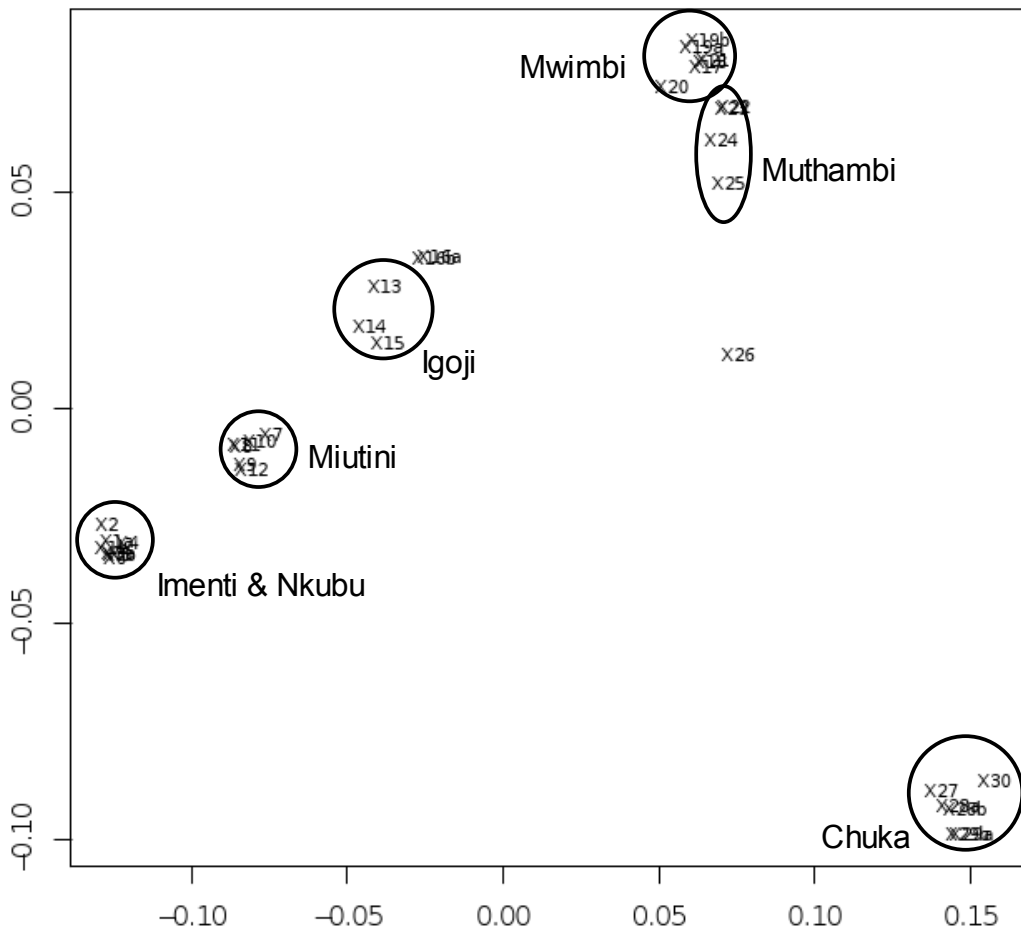
Figure 3: Lexical distances of CKB

Figure 4: Lexical distances of Meru and Chuka

3. Qualitative Analysis

‣ The procedures described above yield **synchronic results** ('linguistic snapshot') – in order to deduct **historical claims** from this data, a **qualitative analysis** is required.

‣ The dialectometrical results show the **linguistic distances** between the dialects of CKB – little or no synchronic variation (= low distances) may historically be due to

- chance
- universal tendencies
- **genetic inheritance**
- **language contact**

3.1 Comparison of linguistic distances across categories
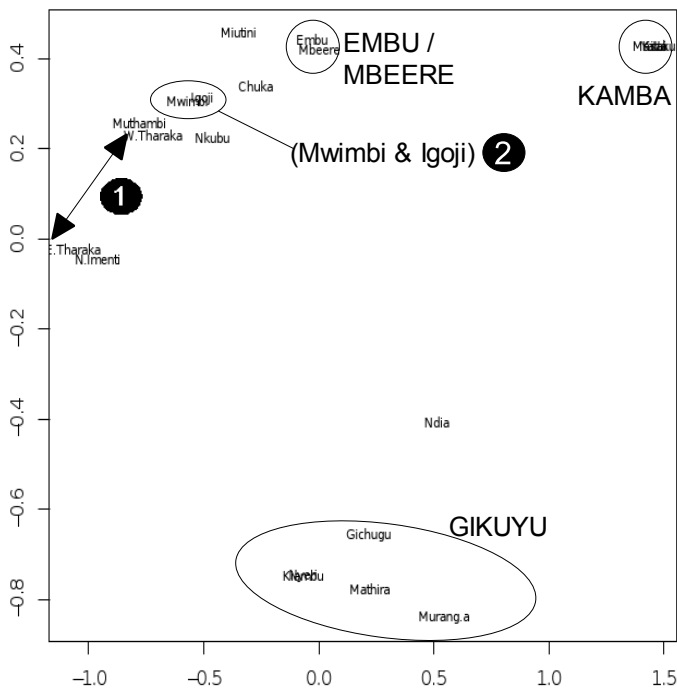
Figure 5: Phonological distances in CKB
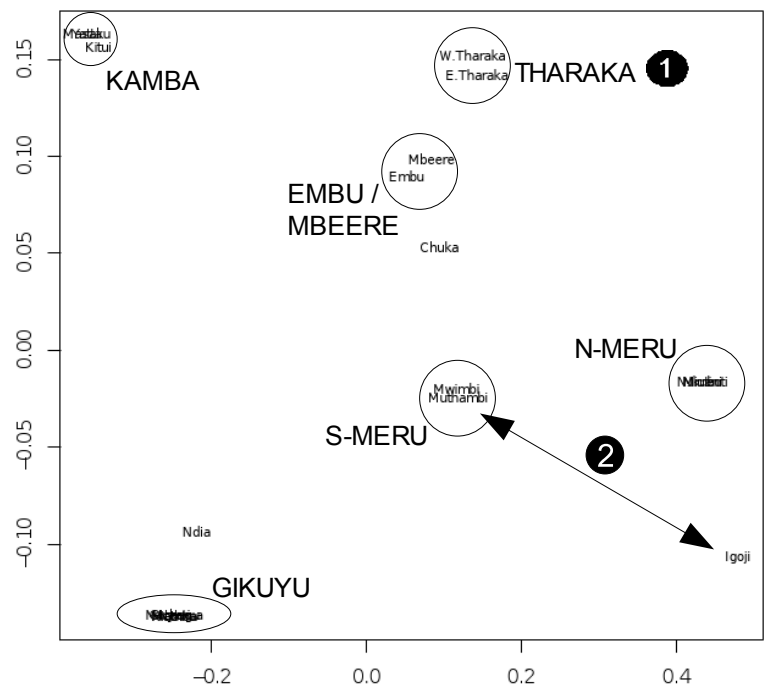
Figure 6: Nominal-morphological distances in CKB

Q: Is there any diagnostic value in the 'lineup' of phonology and morphology?

	Phonology	Nominal Morphology
CASE ❶ **Tharaka**	in the vicinity of the Meru dialects	in the vicinity of Embu / Mbeere
	W-Tharaka affiliated w/ Muthambi; E-Tharaka affiliated w/ Imenti	relatively low distance between East- and West-Tharaka
CASE ❷ **Igoji**	almost identical w/ Mwimbi	relatively high distance between Igoji and Mwimbi-Muthambi

Table 15: Phonology vs. Nominal Morphology in two exemplary cases

"Is there any 'hierarchy' with respect to which categories are more, and which are less, borrowable?" (Aikhenvald & Dixon 2001: 14)

14

3.2 Phonology

Q: What is diagnostic in diachronic phonology?

▸ Dia-Series that show 'simple' (i.e. binary) variation need to be considered **non-diagnostic** as the possibility of universal tendencies cannot be ruled out, e.g. *NK

	Gikuyu, Embu / Mbeere, Kamba	Meru, Chuka, Tharaka
*NK >	**ng**	**nk**
	[+ voice] ⟶	**[- voice]**

*Table 16: Dia-Phoneme *NK and its phonetic realizations*

The variation above may be explained by a **'natural process'** (Stampe[4] 1973: 1):

Voiced stops are relatively difficult to articulate > this is often overcome by **devoicing**

The devoicing of other prenasalized plosives in Meru, Chuka, and Tharaka (e.g. /nd/ > /nt/, /mb/ > /mp/) can be explained by the fact that natural processes affect **natural classes** (Stampe 1979: 137)

⇨ if 'simple' dia-series are to serve as a diagnostic tool, additional information is required, e.g. in 'multiple matches':

Dia-Series $*R_1$ shows weakening (lenition) in Kamba, a natural process that can be described as C → Ø / _V (Mayerthaler[5] 1982: 230).

Dia-Series $*R_2$ shows a realization as [l] in Kamba, while it is realized as [ɾ] and [ɽ] respectively in all other CKB dialects:

	Gikuyu	Embu / Mbeere Chuka Meru	Kamba
$*R_1$	ɾ ⟶ [+back]	ɽ ⟶	Ø
$*R_2$	ɾ ⟶ [+back]	ɽ ⟶ [-back], [-stop]	l

*Table 17: Dia-Series *R₁ and *R₂ in Central Kenyan Bantu*

Additional information: Dia-Series $*R_1$ is attested by 56 lexical items ⎤
Dia-Series $*R_2$ is attested by 21 lexical items ⎦ — 8 : 3 ratio

Interestingly, four out of the items attesting $*R_2$ in Kamba are clearly Swahili loans:

003 brain	akili (Swahili)	>	akili (Kamba)
349 cheap	rahisi (Swahili)	>	laisi (Kamba)
457 road	barabara (Swahili)	>	βalaβala (Kamba)
514 line	mstari (Swahili)	>	mᴜ.sitali (Kamba)

▸ Possibly, dia-series *R$_1$ points towards genetic inheritance while *R$_2$ points towards language contact.

▸ In general, **marked variation** is most promising when it comes to ruling out chance and universal tendencies, e.g. *MP$_1$

Dia-Phoneme	GIKUYU						EMBU/MBEERE			MERU						THARAKA		KAMBA		
	Kiambu	Muarja	Nyeri	Mathira	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
*MP$_1$	ɦ	ɦ	ɦ	ɦ	mb	ɦ	mb	mb	mb	mp	mp	mp	mp	mp	mp	mp	mp	mb	mb	mb
anterior	-	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
voice	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	+	+	+
stop	-	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
prenasal	-	-	-	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ɦ = voiced glottal approximant; mb = prenasalized voiced bilabial plosive; mp = prenasalized voiceless bilabial plosive																				

*Table 18: Dia-Series *MP$_1$*

The variation [mb] vs. [mp] can be explained by the natural process of devoicing.

The variation [mp] vs. [ɦ] is, however, **unnatural** (i.e. more than one feature is affected), rendering universal tendencies a rather implausible explanation in this case:

[mp] ◄──────────► [ɦ]
[anterior]
[voice]
[stop]
[prenasal]
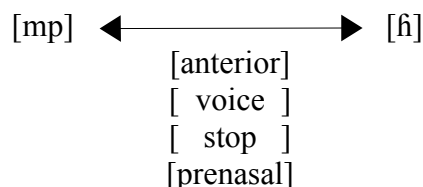
Q: How can we distinguish between internal and external phonological change (especially if no additional information is available)?

3.3 Noun Morphology

▸ Again, only marked variation in the noun class systems can be considered diagnostic in historical terms.

	GIKUYU						EMBU/MBEERE			MERU						THARAKA		KAMBA		
	Kiambu	Muarja	Nyeri	Mathira	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
	a-	a-	a-	a-	a-	a-	ma-	ma-	ma-	ba-	ba-	ba-	ba-	ba-	ba-	ba-	ba-	a-	a-	a-, ma-
	A1	A1	A1	A1	A1	A1	A2	A2	A2	A3	A3	A3	A3	A3	A3	A3	A3	A1	A1	A1, A2

Table 19: Unmarked variation in class 2 (subject markers)

By far the quirkiest variation is the **double prefixing** in the northern Meru dialects Igoji, Miutini, Nkubu, and Imenti described by Möhlig (1974), e.g. class 6 adjective markers:

GIKUYU						EMBU/ MBEERE			MERU						THARAKA		KAMBA		
Kiambu	Muarŋa	Nyeri	Mathɩra	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
mu-	mu-	mu-	mu-	mu-	mu-	mu-	mu-	mu-	mu-	mu-	ju:mu-	ju:mu-	ju:mu-	ju:mu-	mu-	mu-	mu-	mu-	mu-
A1	A1	A1	A1	A1	A1	A1	A1	A1	A1	A1	**A2**	**A2**	**A2**	**A2**	A1	A1	A1	A1	A1

Table 20: Marked variation between N-Meru and the rest of CKB in class 6 (adjective markers)

Another example of marked morphological variation, class 8 adjective markers:

GIKUYU						EMBU/ MBEERE			MERU						THARAKA		KAMBA		
Kiambu	Muarŋa	Nyeri	Mathɩra	Ndia	Gichugu	Embu	Mbeere	Chuka	Muthambi	Mwimbi	Igoji	Miutini	Nkubu	N-Imenti	E-Tharaka	W-Tharaka	Masaku	Yatta	Kitui
N-	*N-*	*N-*	*N-*	*N-*	*N-*	*i-, ci-*	*i-, ci-*	*i-, ci-*	*i-, bi-*	*i-, bi-*	*bi:bi-*	*bi:bi-*	*bi:bi-*	*bi:bi-*	*i-, bi-*	*i-, bi-*	*i-*	*i-*	*i-*
A	A	A	A	A	A	B1, B2	B1, B2	B1, B2	B1, B3	B1, B3	B4	B4	B4	B4	B1, B3	B1, B3	B1	B1	B1

Table 21: Marked variation in CKB in class 8 (adjective markers)

Table 21 represents five isoglosses dividing CKB into the following groups:

Group 1: Gikuyu

Group 2: Embu, Mbeere, Chuka

Group 3: Mwimbi, Muthambi, Tharaka

Group 4: Igoji, Miutini, Nkubu, Imenti

Group 5: Kamba

Groups 2, 3, and 5 are affiliated by common form B1 /i-/.

Q: How can we distinguish between internal and external morphological change?

3.4 Lexicon

Again, the big question is: How can we distinguish between inheritance and contact?

A possible solution to the problem: **The loanword typology** (Tadmor et al. 2010)

The loanword typology project = quantitative study of **loanwords** in 41 languages worldwide aiming at the identification of (groups of) meanings that are generally **borrowing-resistant**.

▸ Differences in word classes: nouns > verbs > adjectives and adverbs

→ borrowability

▸ Differences in **semantic fields**:

SEMANTIC FIELD	LOANWORDS AS % OF TOTAL
Religion and belief	41,2
Clothing and grooming	38,6
The house	37,2
Law	34,3
Social and political relations	31,0
Agriculture and vegetation	30,0
Food and drink	29,3
Warfare and hunting	27,9
Possession	27,1
Animals	25,5
Cognition	24,2
Basic actions and technology	23,8
Time	23,2
Speech and language	22,3
Quantity	20,5
Emotions and values	19,9
The physical world	19,8
Motion	17,3
Kinship	15,0
The body	14,2
Spatial relations	14,0
Sense perception	11,0

Table 22: Semantic fields ranked by loanword percentage (Tadmor et al. 2010: 232)
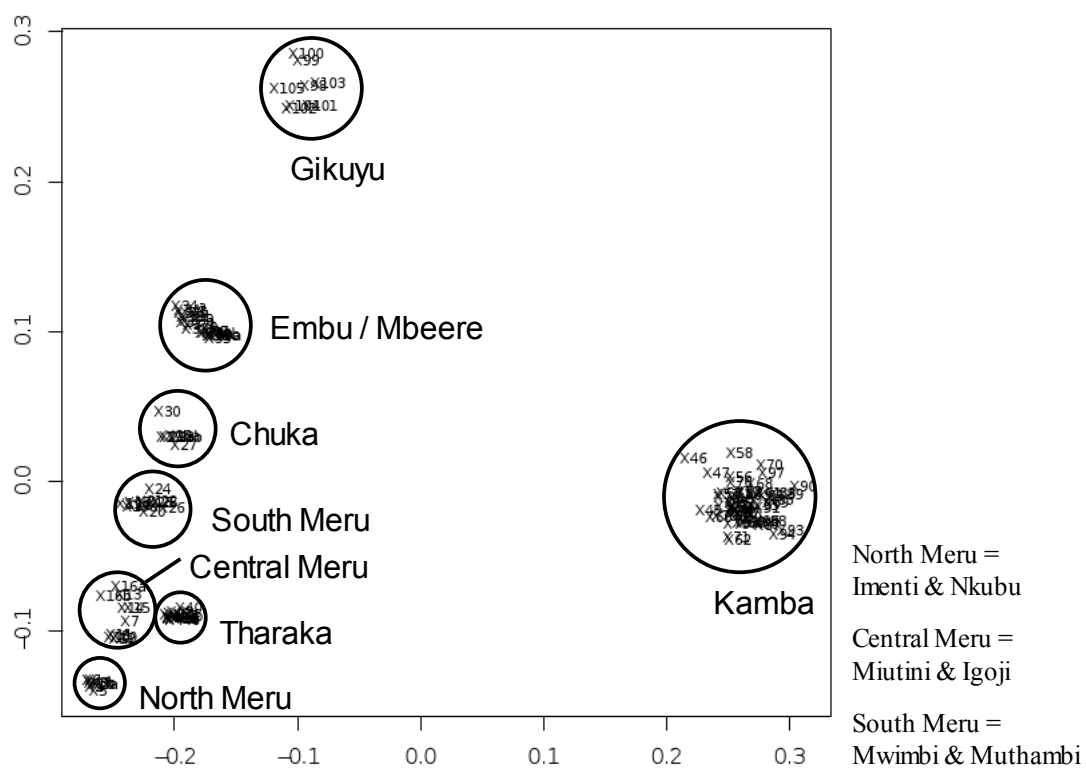
Figure 7: Lexical distances of CKB
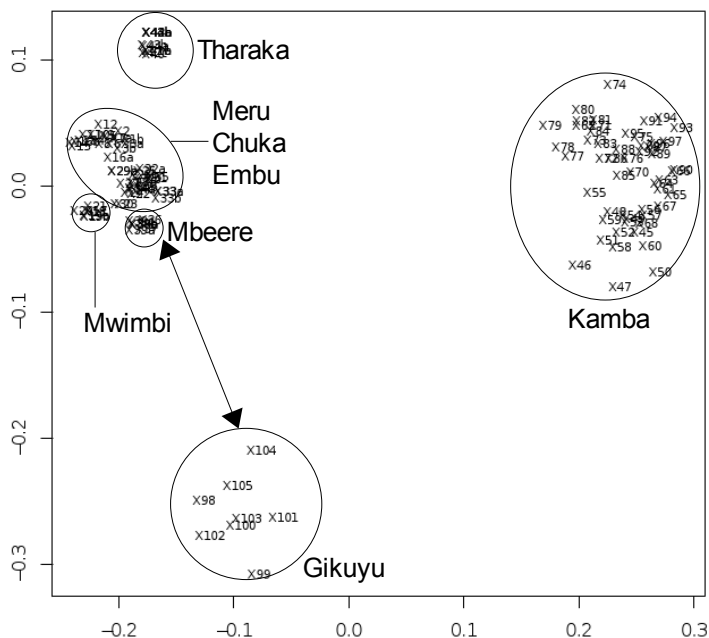
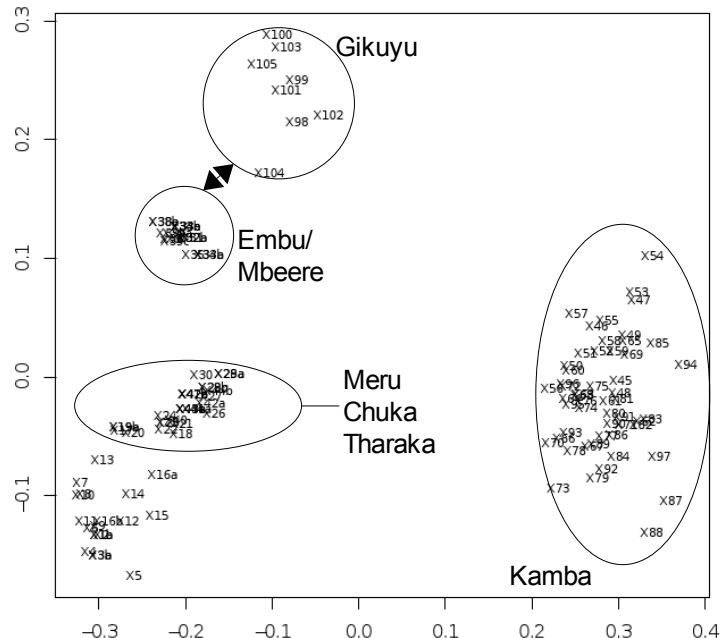*Figure 8: Lexical distances in CKB (the body)*

Figure 9: Lexical distances in CKB (the house)

Two exemplary cases:

	The body	The house
Gikuyu : Embu	relatively **high** distance	relatively **low** distance
Tharaka : Meru	relatively **high** distance	relatively **low** distance

Table 23: Lexical distance of selected varieties of CKB according to different semantic domains

▸ high distance in **core vocabulary** = weak *genetic* affiliation?

▸ low distance in **cultural vocabulary** = strong *contact* affiliation?

Interestingly, at least seven words out 41 compared in the semantic field 'the house' are clearly borrowed from Swahili:

	Swahili	**Embu**	**Gikuyu**
200 window	dirisha	ndiri:ca	ndiri:ca
201 door	mlango	mʊ.rangɔ	mʊ.rangɔ
243 chair	kiti	gɪ.tɪ	gɪ.tɪ
246 basket	kikapu	gɪ.kabʊ	gɪ.kabu
247 bottle	chupa	mʊ.cu:ba	cuba
250 matchet	panga	kɪ.banga	banga
257 lamp	taa	ta:wa	tawa

Table 24: Swahili loans in Embu and Gikuyu

▸ Embu and Gikuyu are quite distant from each other in terms of phonology, noun morphology, and lexicon. As far as terminology in the semantic domain 'the house' is concerned, the distance is, however, relatively low - this is possibly due to a common influence from Swahili.

4. Summary and Outlook

▸ Summary of the quantitative analysis

Dialectometry measures the synchronic proximity between dialects on the following linguistic levels:

phonology - variation in phonetic realization, phonological systems, and phonological rules

noun morphology - formal variation in the noun class system

lexicon - phonological and morphological variation in the vocabulary

Multidimensional Scaling depicts the linguistic distances between the varieties of CKB and enables us to identify **dialect clusters** (areas of low linguistic variation); additional investigation by means of *cluster analysis* still pending.

▸ Summary of the qualitative analysis

Dialect clusters may have come into being due to (1.) chance, (2.) universal tendencies, (3.) genetic inheritance, and (4.) language contact.

The concepts of **naturalness** / **markedness** enable us to rule out chance and universal tendencies > the challenge, then, is to distinguish **genetic inheritance** from **contact**!

> "Contact is a source of linguistic change if it is less likely that a particular change would have happened outside a specific contact situation." (Thomason 2010: 32)

▸ Outlook: Connecting linguistic and extra-linguistic evidence

Example 1: Mbeere : Embu : Kamba

Linguistic findings	Extra-linguistic evidence
- Embu and Mbeere are almost identical linguistically - Concerning phonology and morphology, Mbeere is closer to Kamba than any other dialect of CKB - Lexically, Embu / Mbeere are closely affiliated w/ Meru	The Mariguuri legend: The Mbeere migrated into CK with the Embu to their right and the Kitui-Kamba to their left > they consider both groups to be their relatives (Mwaniki 1973: 22 f.)

Example 2: Chuka

Linguistic findings	Extra-linguistic evidence
The Chuka are the 'odd guys out' in linguistic terms (phonologically, morphologically, lexically).	Orde-Brown (1925: 20) reports that the Chuka consider themselves to be the original inhabitants of their territory

Note: It is very likely that more than one historical development is responsible for the emergence of a particular cluster. If contact, in a specific case, is a plausible explanation, the type of contact / the direction of borrowing need to be specified.

References:

Aikhendvald, A. & R. Dixon (2001). *Introduction.* In: Areal Diffusion and Genetic Inheritance, ed. by A. Aikhendvald and R. Dixon. Oxford: OUP. 1-26.

Chomsky, N. & M. Halle (1968). *The Sound Patterns of English.* New York: Harper & Row.

Jakobson, R. et al. (1952). *Preliminaries to Speech Analysis. The Distinctive Features and their Correlates*. Technical Report, Acoustic Laboratory, Massachusetts Institute of Technology, 13.

Krefeld, T. (2010). *Phonologische Prozesse.* In: Language Typology and Language Universals, ed. by M. Haspelmath et al. (HSK, Vol. 20.2). Berlin: Mouton de Gruyter. 1336-1347.

Mayerthaler, W. (1982). *Markiertheit in der Phonologie.* In: Silben, Segmente, Akzente, ed. by T. Vennemann. Tübingen: Niemeyer. 205-246.

Möhlig, W. (1974). *Die Stellung der Bergdialekte im Osten des Mt. Kenya.* Berlin: Reimer.

Möhlig, W. (1979). *The Bantu nucleus: its conditional nature and its prehistorical significance*. SUGIA 1. 109-104.

Möhlig, W. (1981). *Stratification in the history of the Bantu languages*. SUGIA 3. 251-294.

Mwaniki, H. (1973). *Embu historical texts*. Kampala: East African Literature Bureau.

Orde-Brown, G. (1925). *The vanishing tribes of Kenya*. London: Seeley, Service & Co.

Stampe, D. (1973). *A Dissertation on Natural Phonology*. New York: Garland.

Stampe, D. & P. Donegan (1979). *The Study of Natural Phonology.* In: Current Approaches to Phonological Theory, ed. by D. Dinnsen. Bloomington: Indiana University Press. 126-172.

Tadmor, U. et al. (2010). *Borrowability and the notion of basic vocabulary.* Diachronia 27,2: 226-246.

Thomason, S. (2010). *Contact Explanations in Linguistics.* In: The Handbook of Language Contact, ed. by R. Hickey. Malden: Blackwell. 31-47.