

Bantoid lexical diversity from an individual-based perspective

Jeff Good
University at Buffalo
jcgood@buffalo.edu

Acknowledgments

- Wordlist data collected by Nelson C. Tschonghongei,
University of Yaoundé I
- Buffalo-based researchers: Pierpaolo Di Carlo, Clayton Hamre
- Local project manager: Achuo Christopher Ikom
- Local data manager: Charles Nyoh Abang
- The work is supported by funding from the US National
Science Foundation: Documenting Endangered Languages,
Linguistics, and Geographic and Spatial Sciences Programs



KPAAM-CAM



- Key Pluridisciplinary Advances in African Multilingualism (KPAAM-CAM)
- Collaboration between U. Buffalo, U.Yaoundé I, U. Buea, U. Bamenda, CATUC Bamenda, and U. Dschang
- Ongoing project: Multilingualism and socio-spatial networks
- New projects
 - Training materials for language documentation
 - The impact of displacement on multilingualism and language endangerment

Bantoid diversity

- The entire Bantoid area is characterized by a high degree of linguistic diversity—it is a global linguistic “hotspot”
- What sociolinguistic dynamics have led to this diversity?
- What role does language-internal and dialect-internal variation play in shaping the larger Bantoid picture?

Bantoid lessons

- Linguistic variation in Bantoid area can inform us about
 - How small languages maintain their vitality
 - The sociolinguistic mechanisms of linguistic diversification
 - The relationship between linguistic knowledge and identity
 - How linguistic varieties become conceptualized as a single code, and what it means for a code to be a “language”
- Approaches to linguistic analysis derived from Western contexts are insufficient for the exploration of these issues

Evolution of my own Bantoid work

One language (2004)

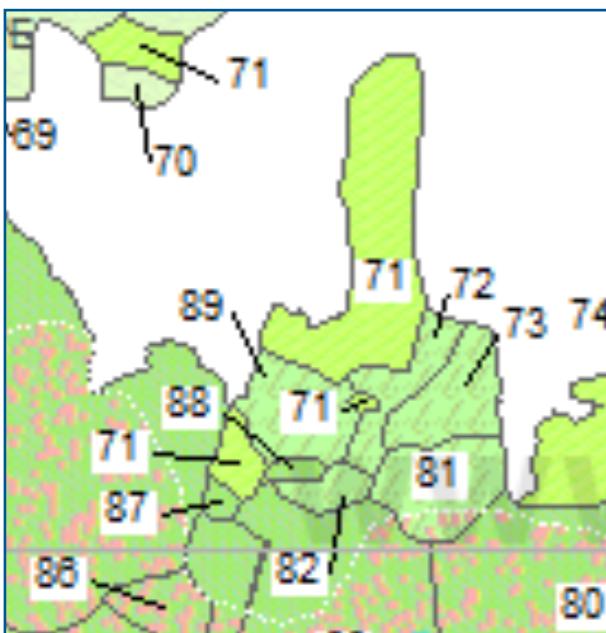
Many languages (2005–2008)

Entire linguistic area (2009–2013)

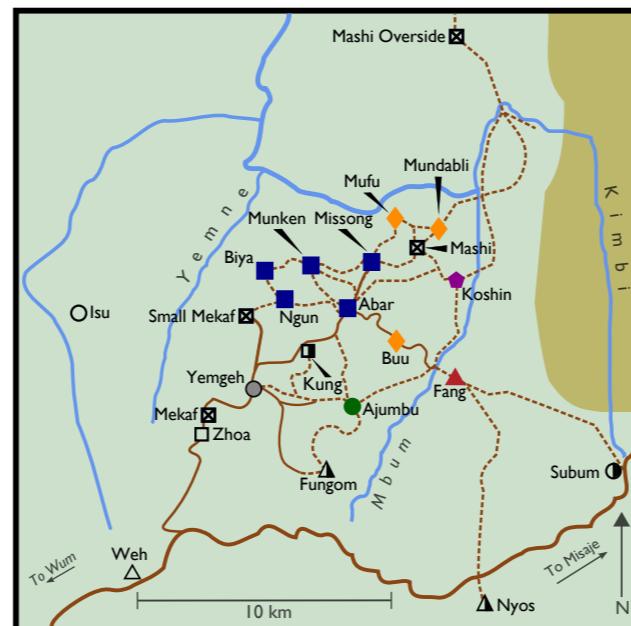
Multilingual individuals living in an area (2014–2018)

Multilingual individuals and social contexts (2019–Pres.)

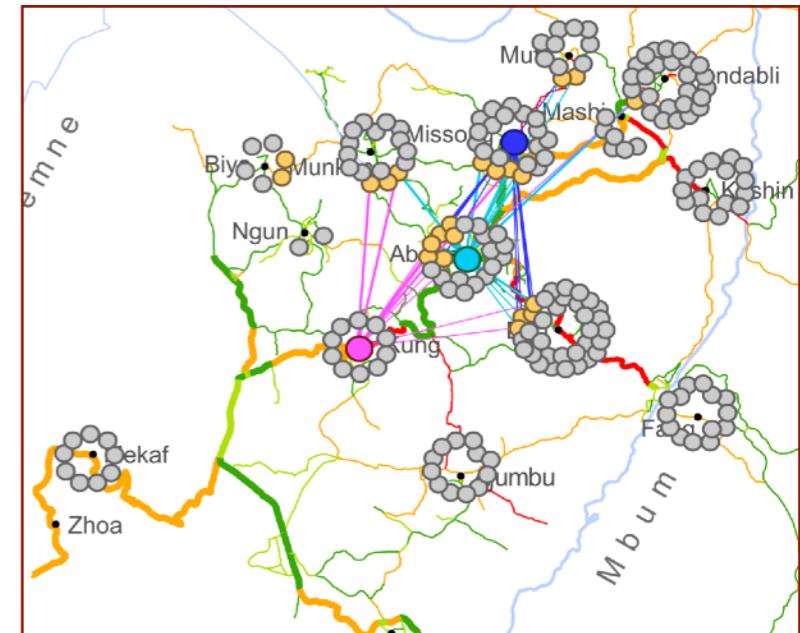
Increasing incorporation of diversity



Map source: Ethnologue



6

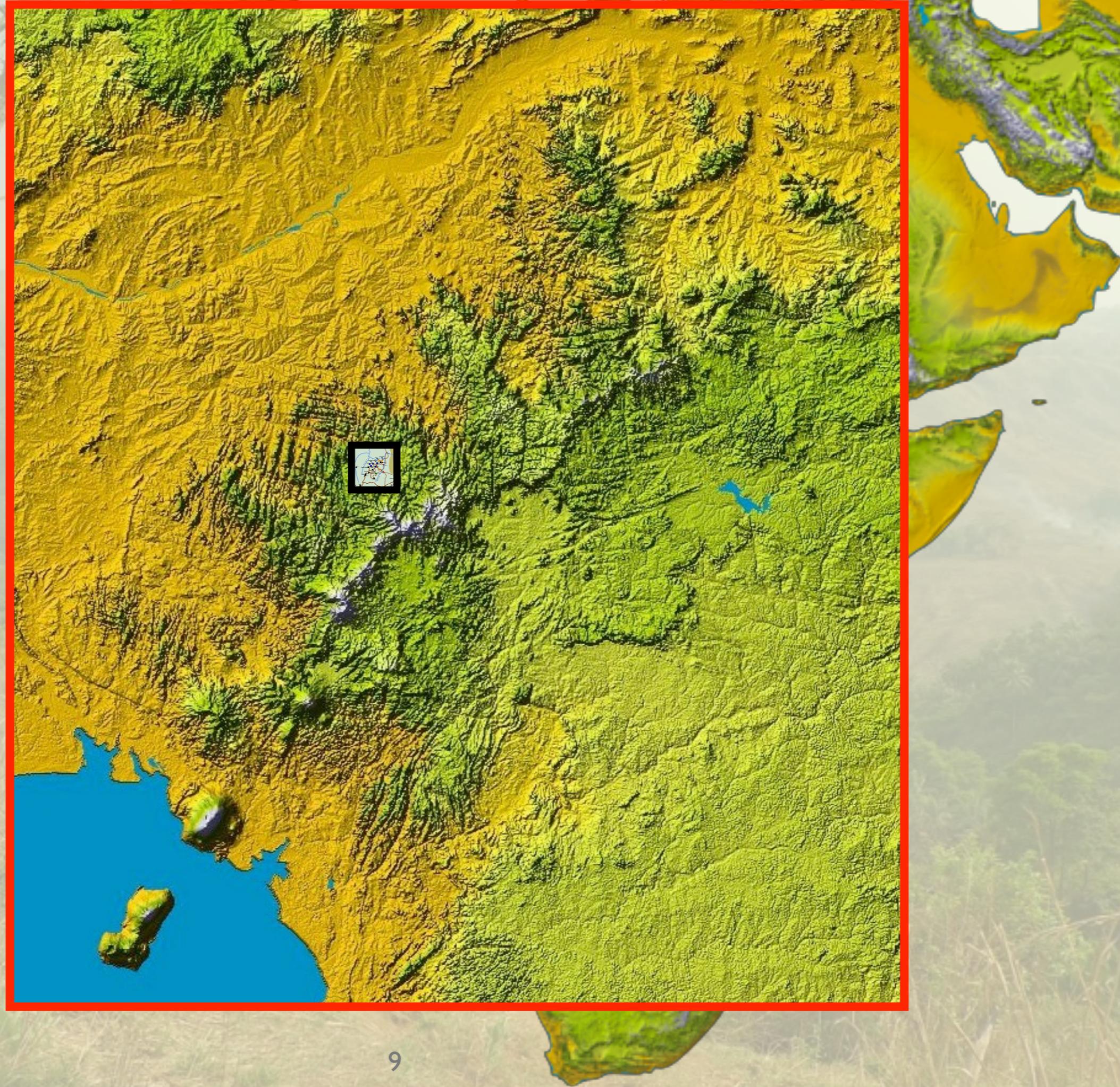




Lower Fungom

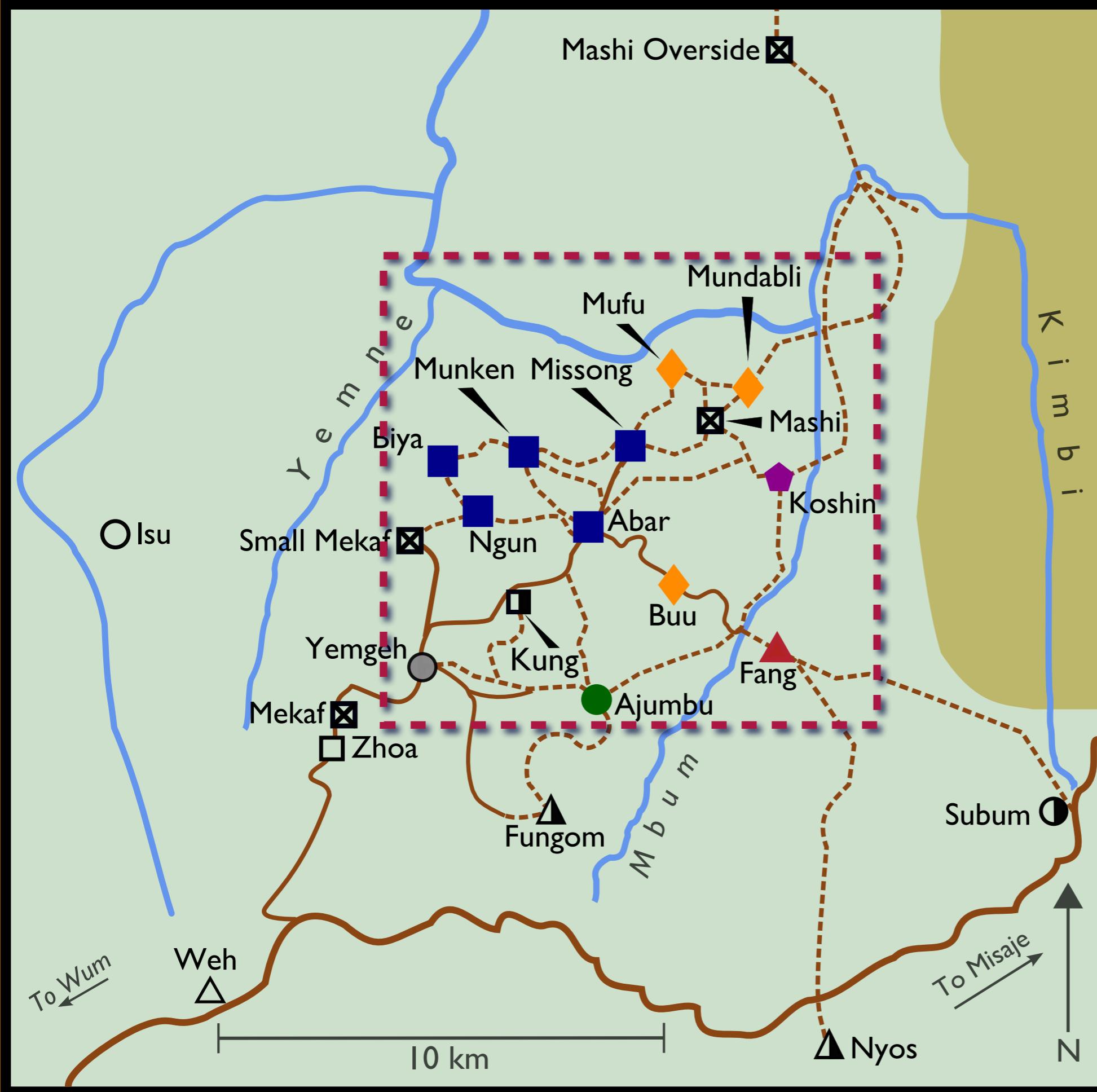
At the northern edge of the Grassfields



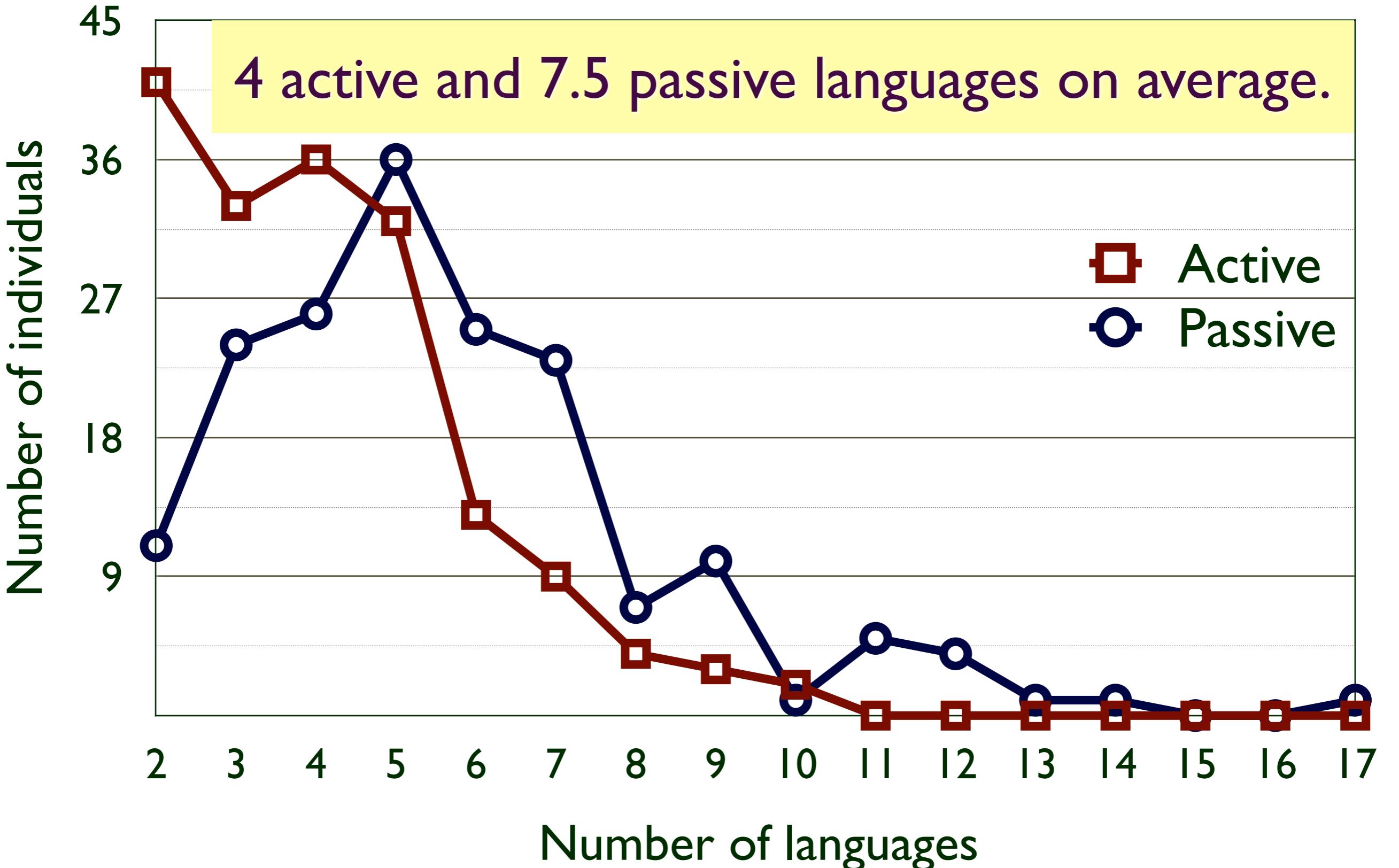


Lower Fungom

- 13 villages
- 13 named lects
- 7–9 “languages”
- 2 dialect clusters
- Bantoid subgroup
- 12,000(?) people
- Rural economy
- Localist attitudes
- Multilingualism/
multilectalism
pervasive



Reported number of languages spoken ($n = 174$)





In sociolinguistic terms, Bantu speakers have long lived in a multilingual continuum, where many speakers master not just their own variety of speech but also those of their neighbors. Linguistic differentiation and convergence are actively pursued, one serving to establish group identities, the other to forge alliances and to foster good friendship.

—Schadeberg (2003: 158–159)

Language and political independence



[The sub-chief of Munken Down] wanted to split off from Munken, make Munken Down independent, and become its chief. But, you know, it was just a crazy plan. He didn't even have a separate language! [laughs]

–Kum Nixon, Munken [abar1238], 8 March 2010

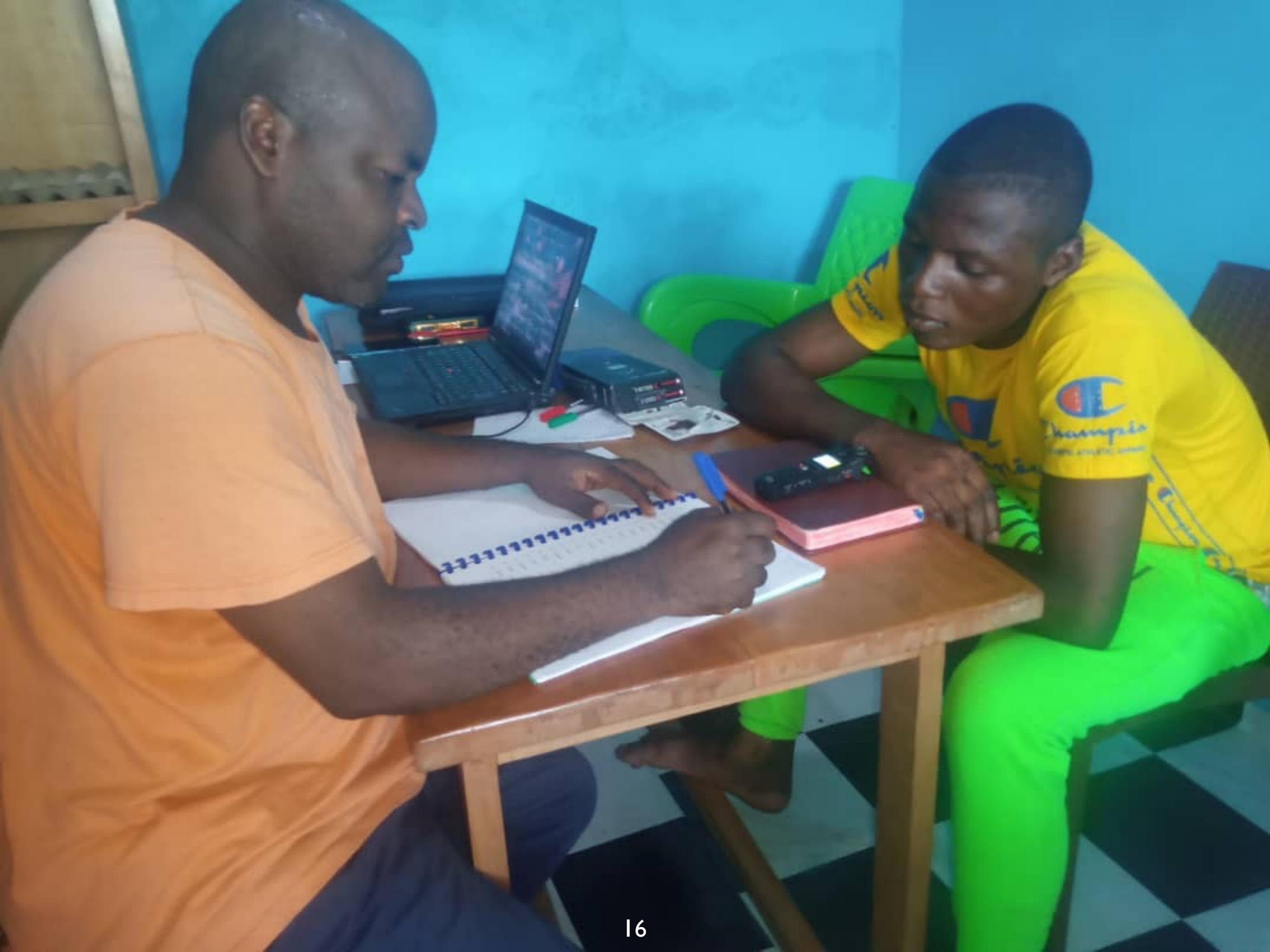
From varieties to languages



- Linguistic diversity requires the formation of new languages
- Linguistics can “find” languages by using tools such as mutual intelligibility and lexicostatistical calculations
- Individuals construct languages by seeing some kinds of variation as sociolinguistically distinctive and linking it to political units

Individual-based word lists

- Based on methods first developed by Angela Nsen Tem (see Mba & Nsen Tem 2020)
- Wordlists collected by Nelson C. Tschonghongei in individual sessions, without standardization/harmonization
- More than 16,000 entries currently across 49 speakers covering all thirteen Lower Fungom varieties
- Detailed sociolinguistic information collected for each speaker to help analyze variation



Research questions

- Individual-level questions
 - How extensive is individual-level lexical variation in these languages?
 - Do some varieties show more individual-level variation than others?
 - To what extent can individual-level variation be tied to a speaker's multilingual repertoire?
- Area-level questions
 - What are the patterns of lexical similarity across Lower Fungom varieties and how clear-cut are “language” boundaries?
 - What lexical items may be emblematic of a specific variety?

Wordlist analysis

- Standardized concept lists used (in two different versions)
- Processed using CLDFBench toolkit (Forkel & List 2020) for cleaning and IPA mapping
- LingPy (List et al. 2018) were used for further analysis
- LingPy was developed for historical linguistic analysis and, in particular, cognate detection

-	ʊ	ɔ	t	o	ɔ	m	-	-	-
-	ʊ	ɔ	t	a	ɔ	m	-	-	-
k	ə	ɔ	t	ø	ɔ	-	-	-	1
-	ʊ	ɔ	t	ɔ	ɔ	m	-	-	-
k	ə	ɔ	t	ʊ	ɔ	k	ə	1	1
k	ə	ɔ	t	ʊ	ɔ	k	ə	1	1
k	ə	ɔ	t	ʊ	ɔ	k	ə	1	1
-	ʊ	ɔ	t	o	ɔ	m	-	-	-
-	ʊ	ɔ	t	ʊ	ɔ	m	-	-	-
-	o	ɔ	t	o	ɔ	m	-	-	-
-	ʊ	ɔ	t	ʊ	ɔ	m	-	-	-
k	ə	ɔ	t	ʊ	ɔ	k	ə	1	1
-	ʊ	ɔ	t	o	ɔ	m	-	-	-
k	ə	ɔ	t	o	ɔ	-	-	-	1

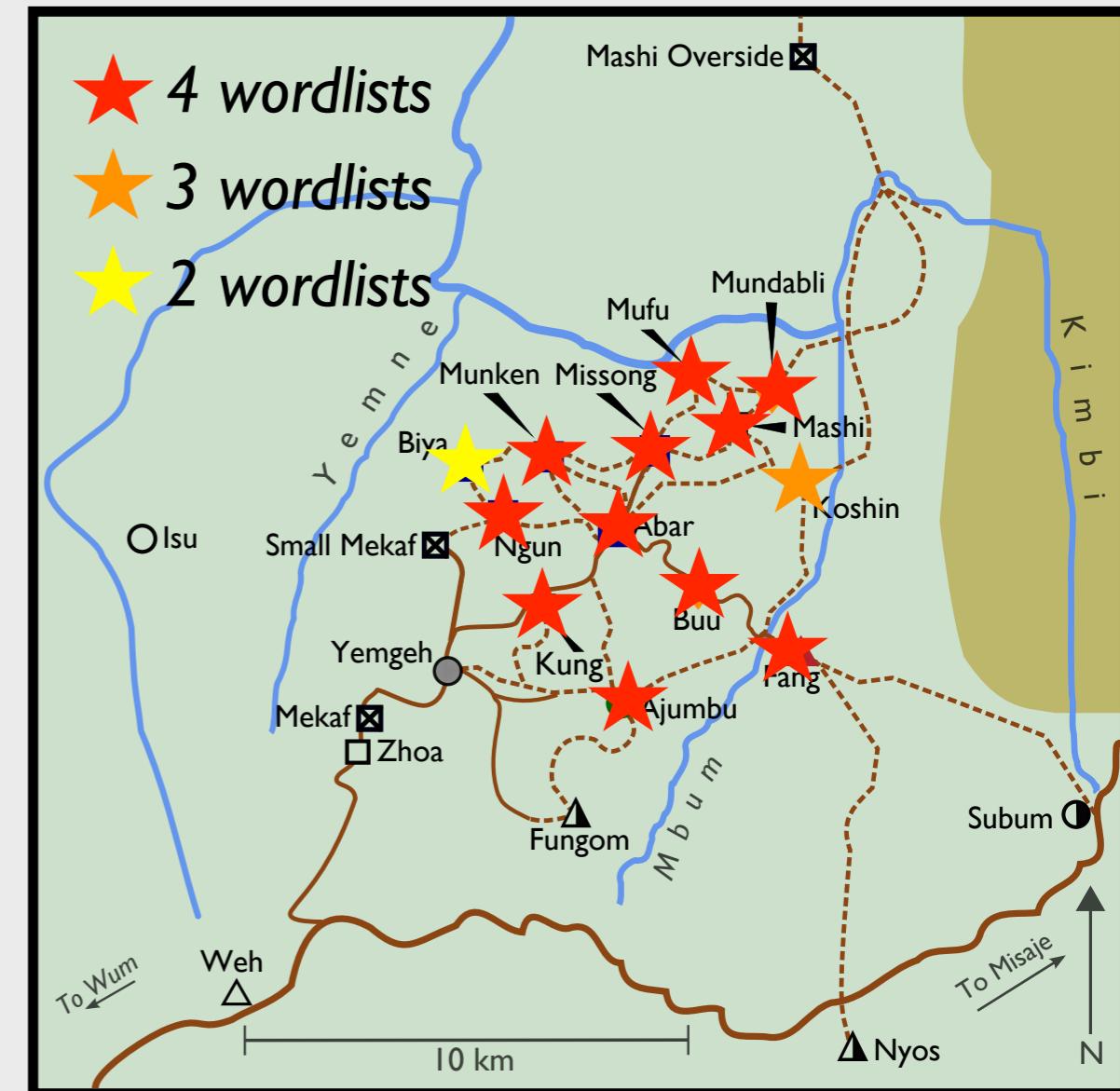
Similarity set for ‘compound’

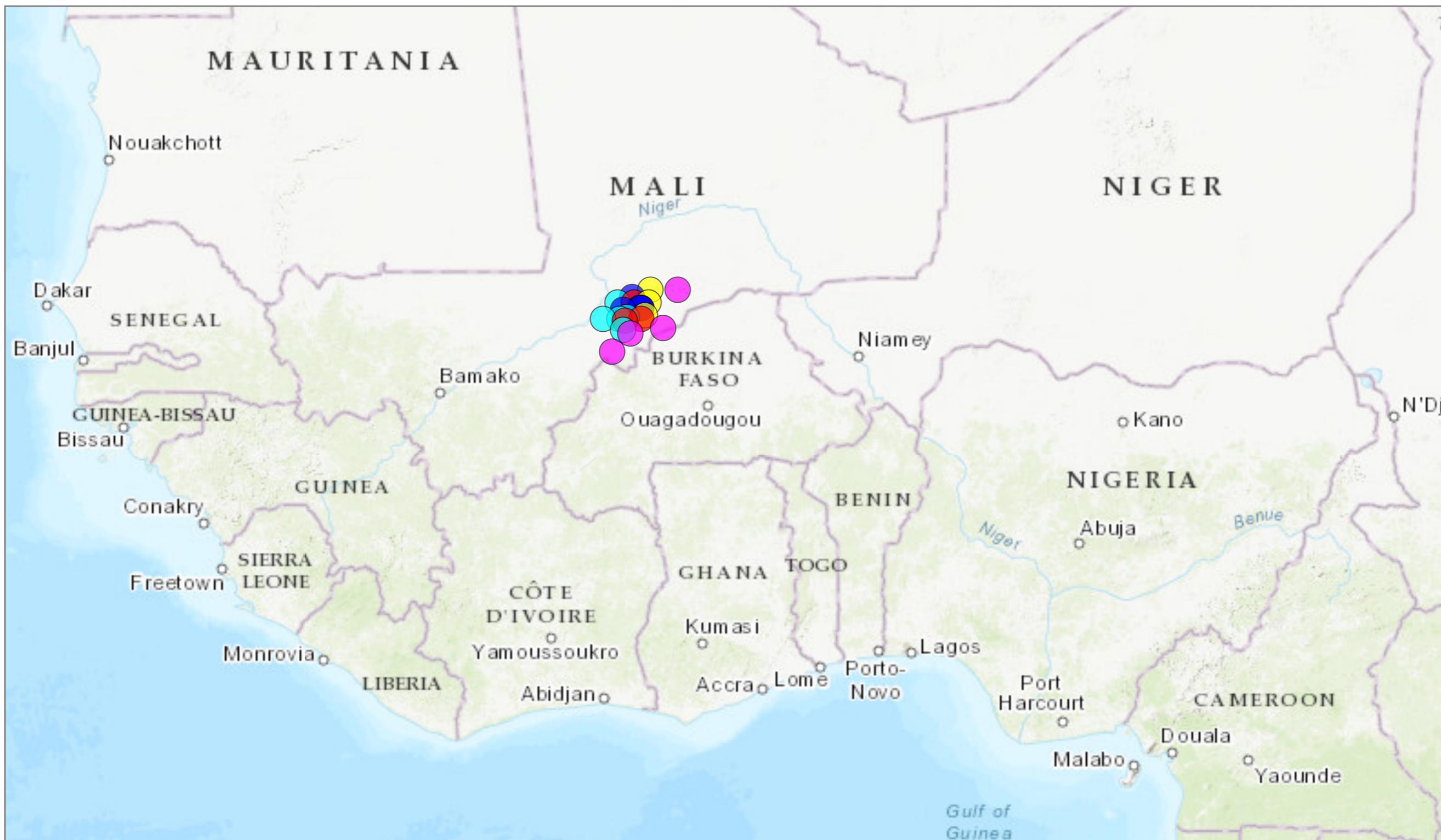
Synchronic cognate analysis

- LingPy uses sound-based principles to assess the similarity of sets of words
- Our primary goal in using it was to gain a detailed view into lexical variation in the local, synchronic space
- Adapting LingPy for synchronic sociolinguistic analysis is straightforward, though this appears to be a new use for it

Data overview

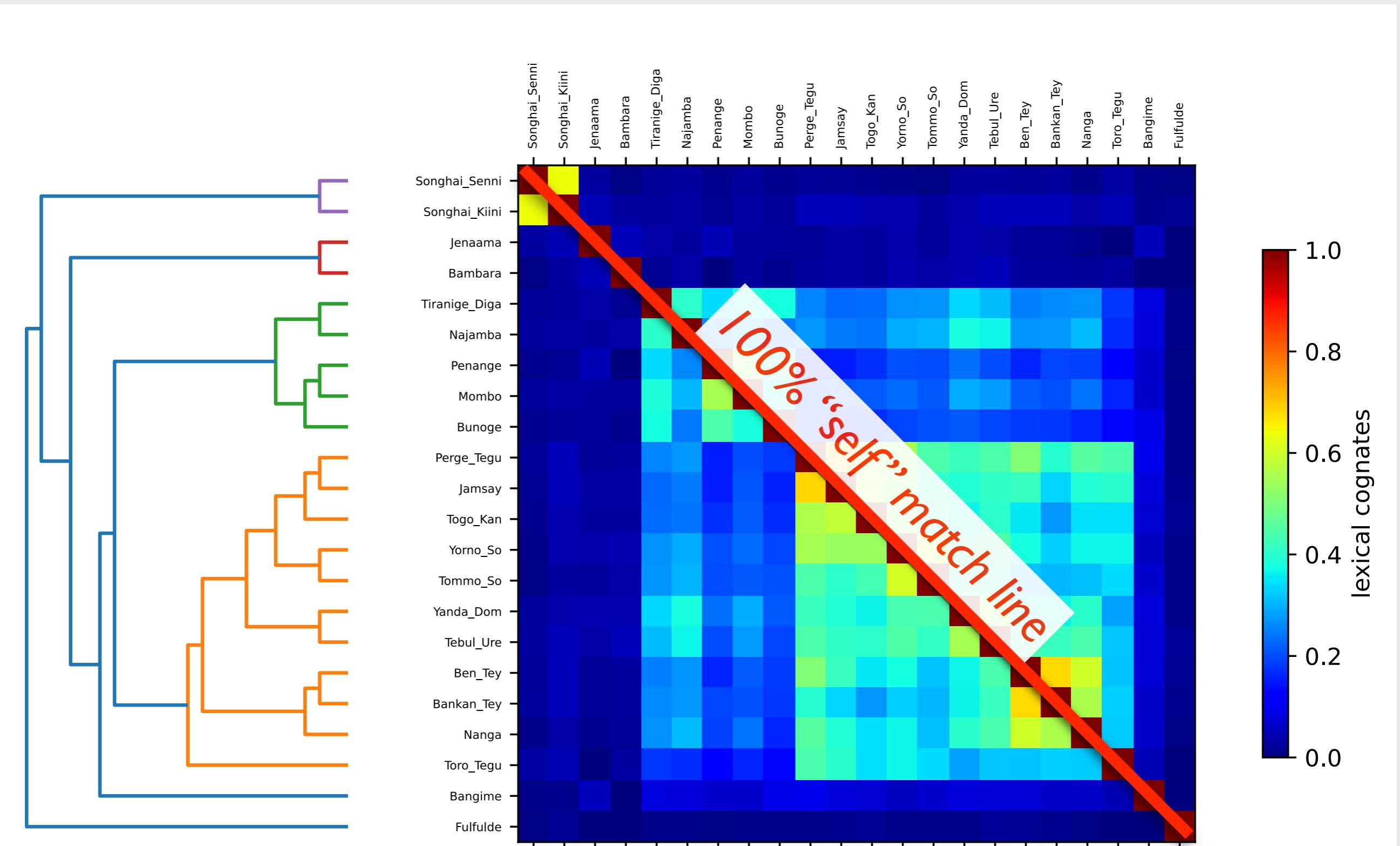
- 49 words lists
 - 4 for 11 of the varieties
 - 3 for 1 variety
 - 2 for 1 variety
- Collected in two phases, with some changes to the concept list
- Plans to collect more wordlists
- We are not aware of any comparable dataset





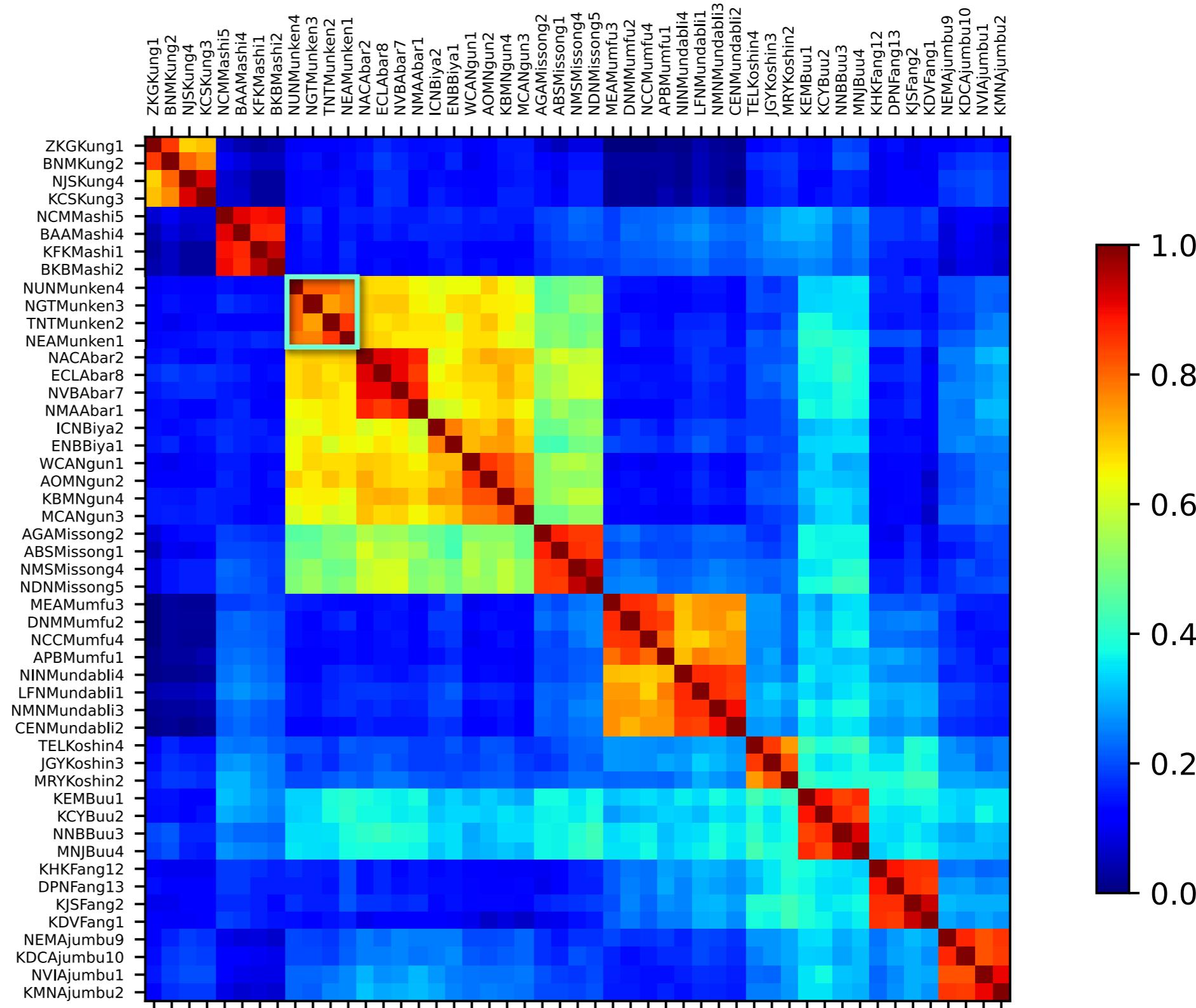
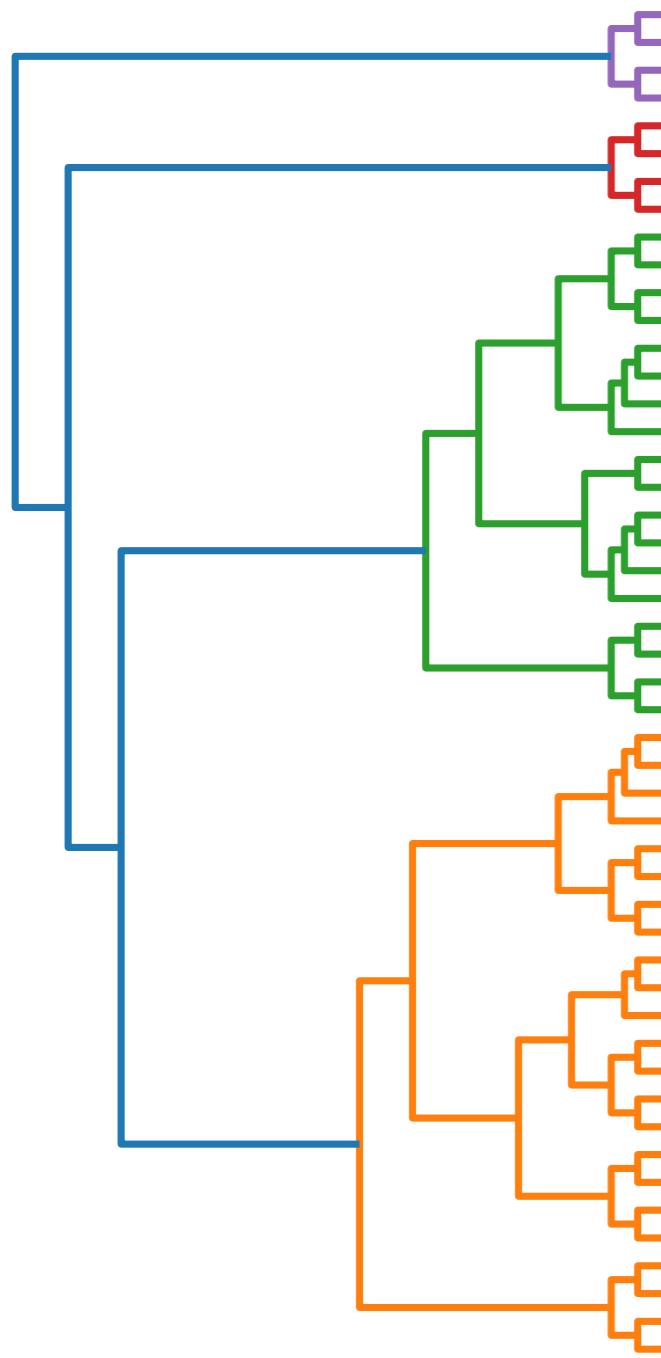
The Dogon languages

A heat map for cognate similarity in a language-based wordlist for languages in the Dogon area as a contrast with the Lower Fungom data

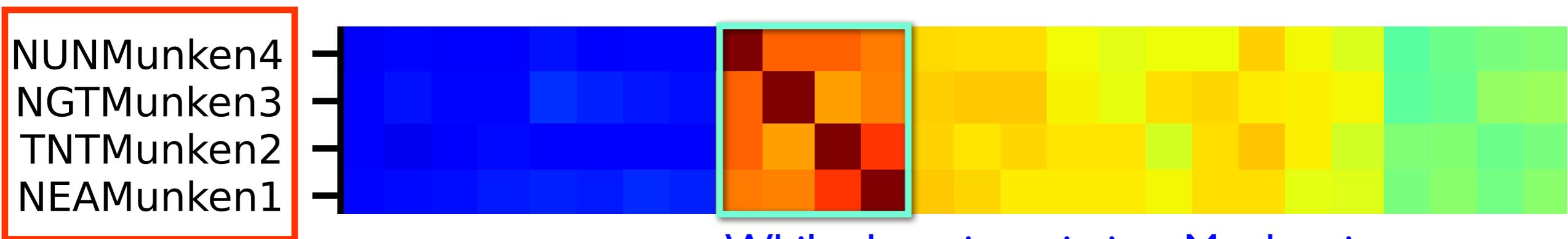


Hantgan & List (to appear, supplemental materials)

49 Lower Fungom wordlists, concepts with 75% (37/49) coverage (6600 words) LingPy Sound-Class-Based Phonetic Alignment (SCA)



Similarity sub-matrix for four individual-based wordlists collected on the Munken variety



Individual doculects
Speaker Initials
Variety Name
ID number

While there is variation, Munken is clearly visible in the heat map.
Its close relations to other Mungbam varieties are also visible.

Calculated distances for Munken varieties

NUNMunken4	1.00	0.83	0.83	0.80
NGTMunken3	0.83	1.00	0.77	0.80
TNTMunken2	0.83	0.77	1.00	0.87
NEAMunken1	0.80	0.80	0.87	1.00

Concept: 574_gardenegg (ID: 100)

CogID	Language	Entry	Aligned Entry
2402	BNMKung2	kə ⁵ ŋa ⁵ ?a ¹	k ə 5 n - - a 5 ? a 1
2402	ENBBiyal	fɪ ¹ ndʒa ¹ ha ⁵	f i 1 n d ʒ a 1 h a 5
2402	ICNBIya2	fɪ ¹ ndʒɛ: ¹ ⁵	f i 1 n d ʒ ε: 1 - - 5
2402	KBMNgun4	fɪ ¹ ndʒa ¹ ha ⁵	f I 1 n d ʒ a 1 h a 5
2402	KCSKung3	kə ¹ ŋa ⁵ ?a ¹	k ə 1 n - - a 5 ? a 1
2402	MCANgun3	ndʒa ¹ ha ⁵	- - - n d ʒ a 1 h a 5
2402	NJSKung4	kə ⁵ ŋa ⁵ ?a ¹	k ə 5 n - - a 5 ? a 1
2402	WCANgun1	fɪ ¹ ndʒa ⁵ ha ⁵	f i 1 n dʒ - a 5 h a 5
2402	ZKGKung1	kə ¹ ŋa ¹ ?a ¹	k ə 1 n - - a 1 ? a 1
2403	APBMumfu1	gχɔ ¹ ŋ	g χ ɔ 1 - ŋ
2403	DNMUMMumfu2	gjɔ ¹ ŋ⁵	g j ɔ 1 - ŋ⁵
2403	MEAMUMMumfu3	gχɔ ⁵ ¹ŋ	g χ ɔ 5 1 ŋ
2403	NCCMUMMumfu4	gχɔ ¹ ŋ	g χ ɔ 1 - ŋ
2433	NEAMunken1	ʃi ¹ ndʒa ⁵	ʃ i 1 n dʒ - a 5
2433	NGTMunken3	ʃi ³ ndʒa ⁵	ʃ i 3 n dʒ - a 5
2433	NUNMunken4	ʃi ⁵ ndʒa ⁵	ʃ i 5 n dʒ - a 5
2433	TNTMunken2	ʃi ¹ ndʒja ⁵	ʃ i 1 n dʒ j a 5

Concept: 1167_seed (ID: 17)

CogID	Language	Entry	Aligned Entry
384	NMAAbar1	i ⁵ kp <u>U</u> ¹ ⁵	i 5 kp u 1 - 5
384	NUNMunken4	I ¹ kpo ¹	I 1 kp o 1 - -
384	NVBAbar7	I ¹ kp <u>U</u> ¹	I 1 kp u 1 - -
384	TNTMunken2	i ¹ kp <u>U</u> ¹	i 1 kp u 1 - -
387	NDNMissong5	I ¹ ʃa ⁵ m	I 1 ʃ a 5 - m - -
387	NGTMunken3	I ⁵ ʃa ⁵ m	I 5 ʃ a 5 - m - -
387	NINMundabli4	sa ³ hm	- - s a 3 h m - -
387	NMNMundabli3	sa ⁵ m	- - s a 5 - m - -
390	BNMKung2	I ¹ lə ⁵ m	--
399	KCSKung3	fə ¹ sa ⁵ ŋ	f ə 1 s a 5 ŋ
399	KDCAjumbu10	ʃɛ ⁵ n	- - - ʃ ɛ 5 n
399	KMNAjumbu2	ʃɛ ⁵ n	- - - ʃ ɛ 5 n
399	NEMAjumbu9	ʃɛ ⁵ n	- - - ʃ ɛ 5 n
399	NJSKung4	fə ¹ sa ⁵ ŋ	f ə 1 s a 5 ŋ
399	NVIAjumbu1	a ⁵ ʃɛ ⁵ n	- a 5 ʃ ɛ 5 n
416	NEAMunken1	ŋg <u>U</u> ¹	--
431	ZKGKung1	i ⁵ z <u>U</u> ⁵ ?	--

49 wordlists, concepts with 75% (37/49 wordlists) coverage (6600 words)

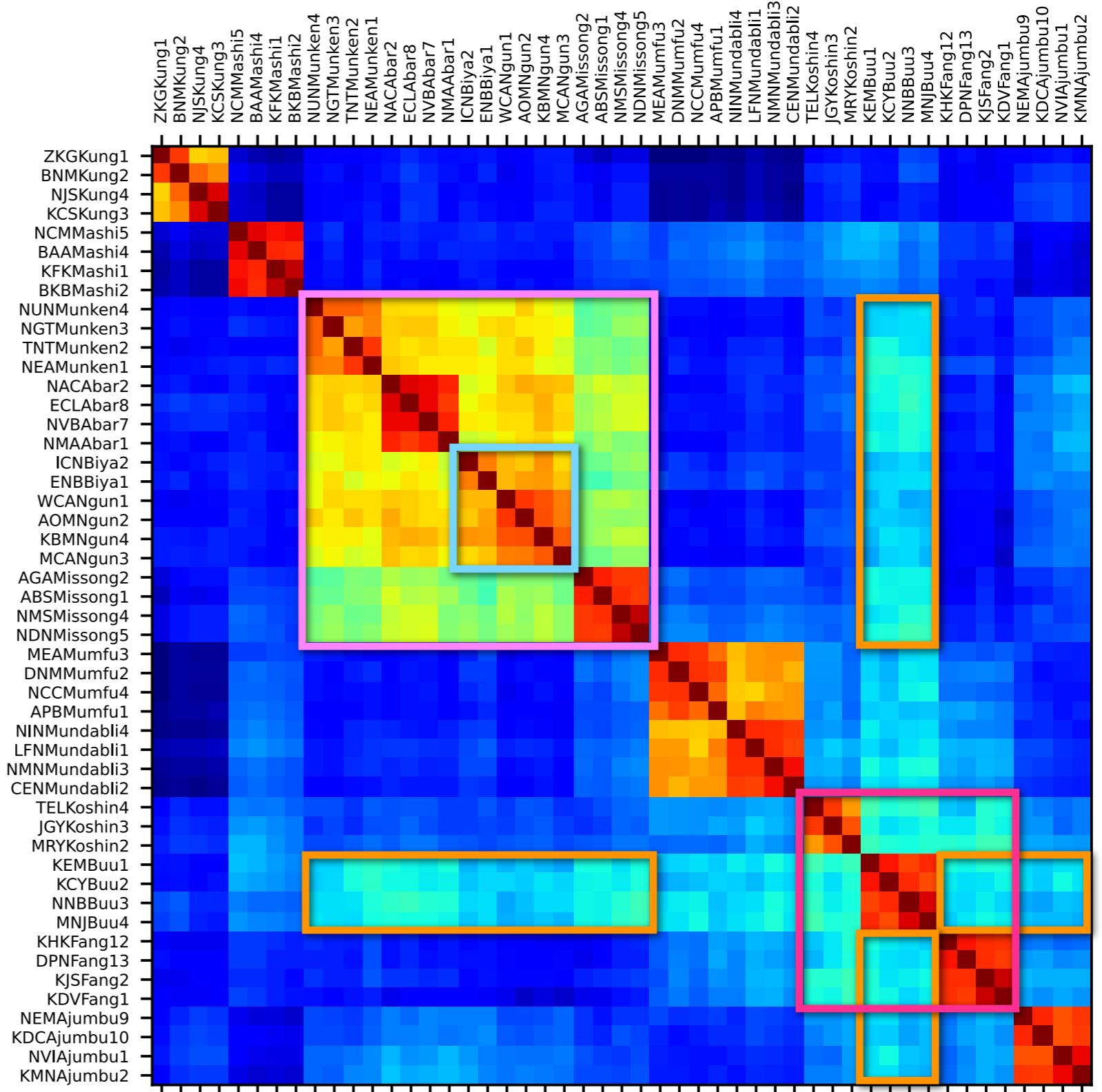
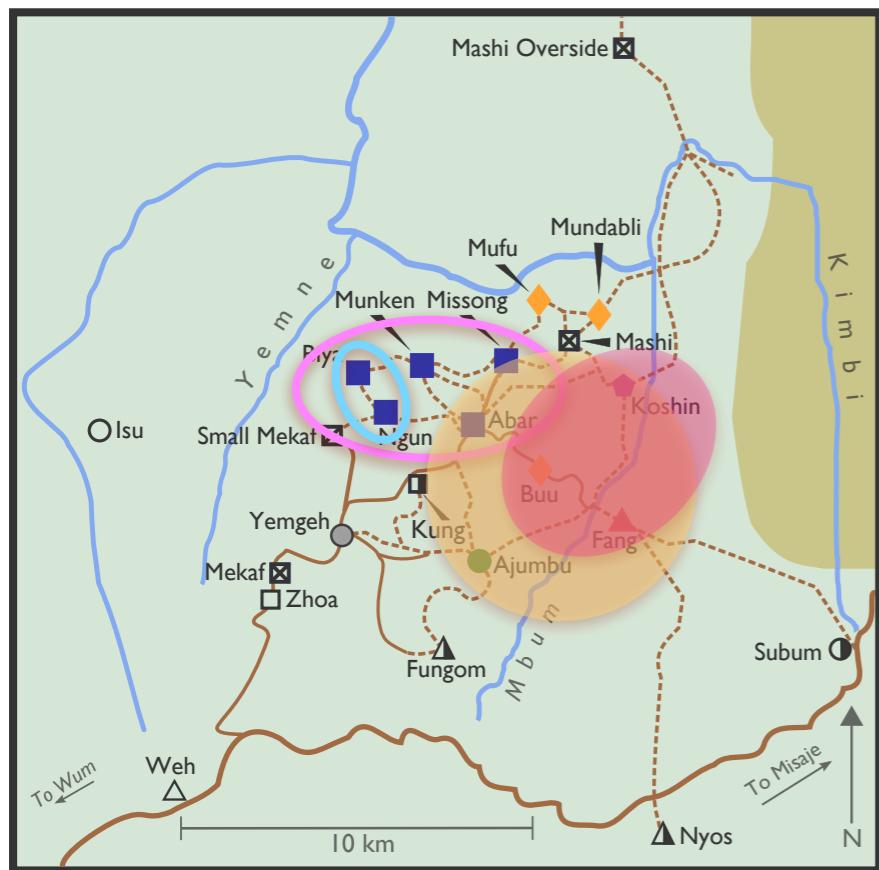
LingPy Sound-Class-Based Phonetic Alignment (SCA)

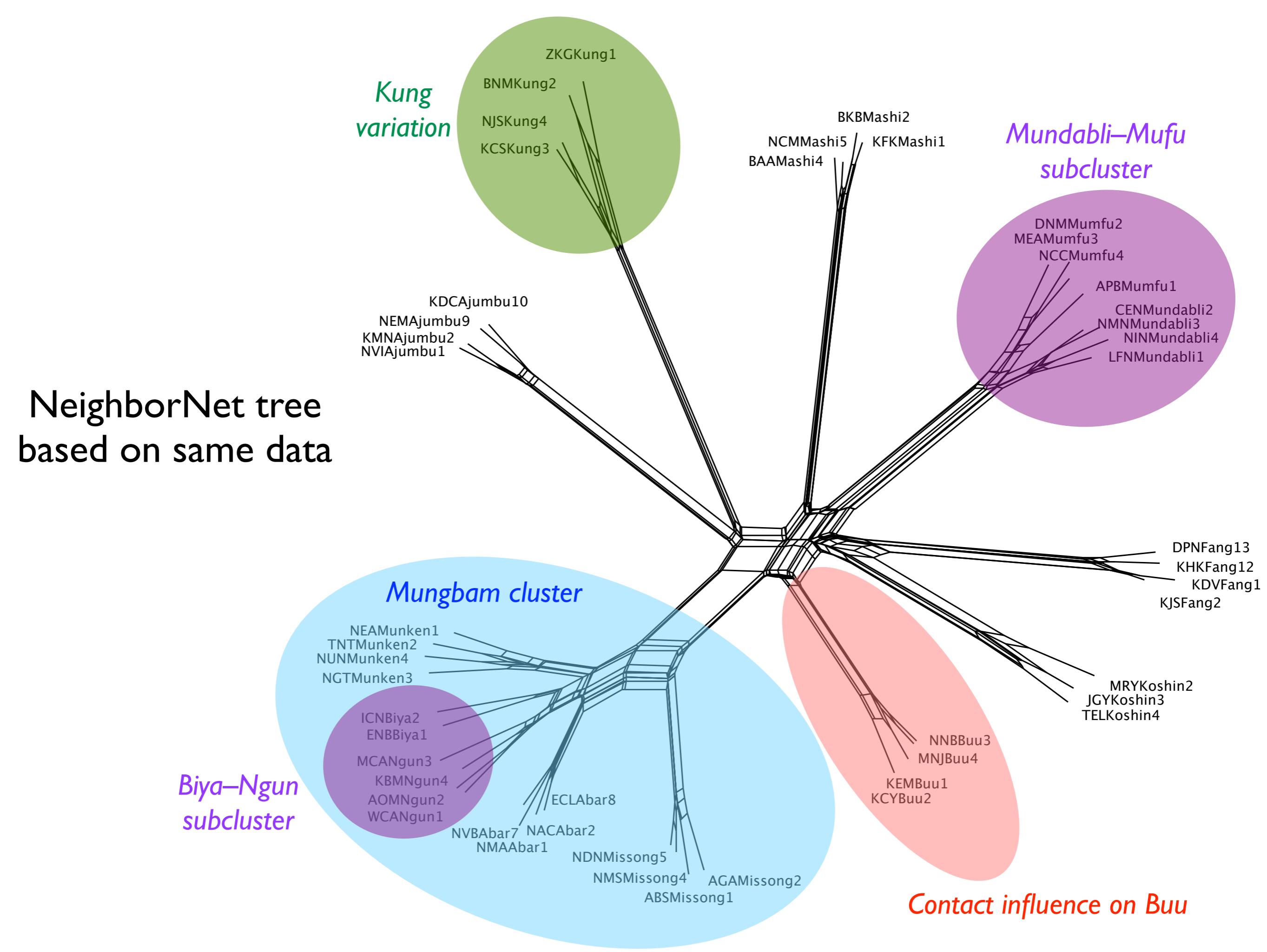
Mungbam cluster

Biya–Ngun subcluster

Buu contact relationships?

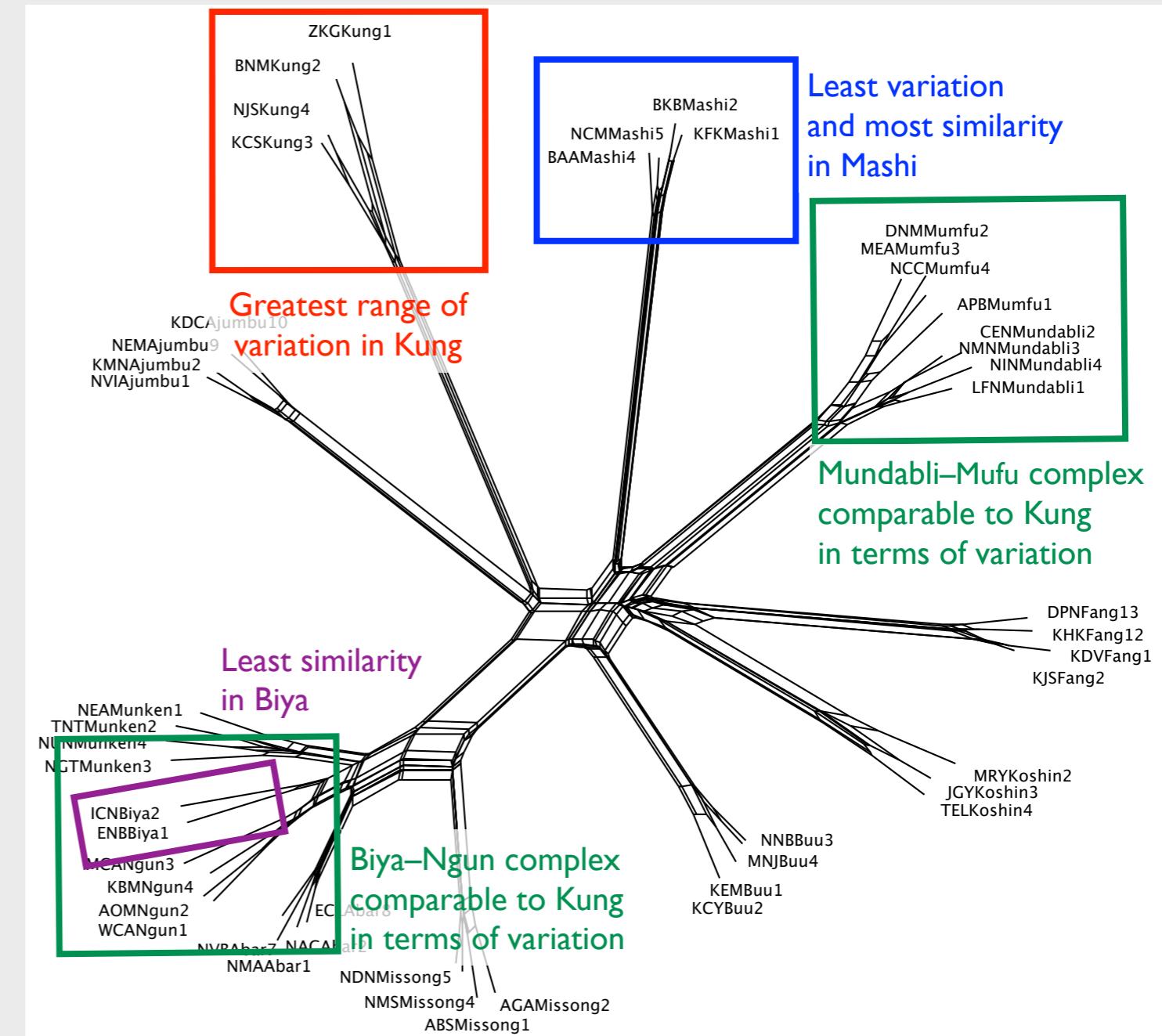
Southeastern contact area?





Patterns of individual variation

- Similarity within a variety
 - Maximum: 0.91 (Mashi)
 - Minimum: 0.80 (Biya)
 - Average: 0.85
- Range of variation in a variety
 - Maximum: 0.28 (Kung)
 - Minimum: 0.12 (Mashi)
 - Average: 0.20
- Variation across Biya and Ngun and Mundabli–Mufu similar to variation within Kung
- We are not aware of a “baseline” comparison to assess the results



Finding emblematic differences



- In a highly multilingual space, what lexicogrammatical distinctions become markers of different “languages”?
- Watson (2018) explores this for the Casamance region of Senegal, looking especially at phonological factors
- This dataset allows us to explore which meanings are associated with more or less “cognate” variation

Concept: 381_day (ID: 75)

CogID	Language	Entry	Aligned Entry					
1884	APBMumfu1	tɔɪ³	t	ɔ	i	3	-	
1884	CENMundabli2	to³ʊ³	t	o	3	ʊ	3	
1884	DNMUMumfu2	tɔ⁵	t	ɔ	5	-	-	
1884	LFNMundabli1	toɪ³	t	oɪ	3	-	-	
1884	MEAMumfu3	tɔ³	t	ɔ	3	-	-	
1884	NCCMumfu4	tɔɪ⁵	t	ɔɪ	5	-	-	
1884	NMNMundabli3	to³	t	o	3	-	-	
1885	BAAMashi4	a⁵na³ŋ	a	5	n	a	3	ŋ
1885	BKBMashi2	a⁵na⁵ŋ	a	5	n	a	5	ŋ
1885	KFKMashi1	a⁵na⁵ŋ	a	5	n	a	5	ŋ
1885	NCMMashi5	a⁵na⁵ŋ	a	5	n	a	5	ŋ
1889	DPNFang13	ka⁵lə³	k	a	5	l	ə	3
1889	KDVFang1	ka¹lə¹	k	a	1	l	ə	1
1889	KJSFang2	ka⁵lə¹	k	a	5	l	ə	1
1889	MNJBuu4	kə⁵tə⁵	k	ə	5	t	ə	5
1891	ENBBiyal	m¹fʷɪ⁵	m	1	fʷ	i	5	
1891	ICNBIya2	m¹fʷɪ⁵	m	1	fʷ	i	5	
1893	JGYKoshin3	kpɔ⁵₁ŋ	kp	ɔ	5	1	ŋ	
1893	MRYKoshin2	kpɔ⁵₁	kp	ɔ	5	1	-	
1893	TELKoshin4	kpɔ⁵₁	kp	ɔ	5	1	-	

Concept: 348_cowcattle (ID: 69)

CogID	Language	Entry	Aligned Entry
1807	BAAMashi4	nə ¹	- - n ə 1
1807	BKBMashi2	n ¹ nə ¹	n 1 n ə 1
1807	KFKMashi1	nə ¹	- - n ə 1
1809	ABSMissong1	mbɔ ¹ ŋ	- - m - b ɔ 1 - - - - - ŋ
1809	AGAMissong2	mbʊ ¹ ŋ	- - m - b ʊ 1 - - - - - ŋ
1809	AOMNgun2	mbɔ ¹ ŋ	- - m - b ɔ 1 - - - - - ŋ
1809	APBMumfu1	mbɔ ¹ ŋ	- - m - b ɔ 1 - - - - - ŋ
1809	BNMKung2	mbɔ ¹ lə ⁵ ?	- - m - b ɔ 1 l ə 5 ?
1809	CENMundabli2	mbɔ ⁵ ŋ	- - m - b ɔ 5 - - - - 1 ŋ
1809	DNMMumfu2	mbɔ ¹ ŋ	- - m - b ɔ 1 - - - - - ŋ
1809	DPNFang13	nɔ ¹ lə ⁵	- - n - b ɔ 1 l ə 5 -
1809	ECLAbar8	nɔ ¹ ŋ	- - n - b ɔ 1 - - - - - ŋ
1809	ENBBiya1	nɔ ¹ ŋ	- - n - b ɔ 1 - - - - - ŋ
1809	ICNBIya2	nɔ ¹ ŋ	- - n - b ɔ 1 - - - - - ŋ
1809	JGYKoshin3	nɔ ¹ ŋ	- - n - b ɔ 1 - - - - - ŋ
1809	KBMNgun4	nɔ ¹ ŋ	- - n - b ɔ 1 - - - - - ŋ
1809	KCSKung3	mbɔ ¹ ŋ	- - m - b ɔ 1 - - - - - ŋ
1809	KCYBuu2	mbo ¹ lə ⁵	- - m - b o 1 l ə 5 -
1809	KDVFang1	mbɔ ¹ lə ⁵	- - m - b ɔ 1 l ə 5 -
1809	KEMBuu1	mbo ¹ lə ⁵	- - m - b o 1 l ə 5 -

Most and least stable concepts
based on those found in at least 37
wordlists (136 concepts in total)

Calculated as normalized
entropy based on cognate sets

Concept	Homogeneity
<i>ear</i>	0.97
<i>cowcattle</i>	0.94
<i>tongue</i>	0.93
<i>breast</i>	0.93
<i>child</i>	0.93
<i>mother</i>	0.91
<i>chief</i>	0.89
<i>grave</i>	0.89
<i>axe</i>	0.89
<i>fowl</i>	0.87
<i>bag</i>	0.86
<i>book</i>	0.84
<i>head</i>	0.84
<i>sheep</i>	0.84
<i>gong</i>	0.83
<i>sand</i>	0.83
<i>headpad</i>	0.83
<i>horse</i>	0.82
<i>rope</i>	0.82
<i>bird</i>	0.82
<i>medicine</i>	0.81
<i>sieve</i>	0.81
<i>heart</i>	0.80
<i>raffia bamboo</i>	0.80

Concept	Homogeneity
<i>termite</i>	0.37
<i>rainbow</i>	0.41
<i>trap</i>	0.47
<i>stormwind</i>	0.47
<i>star</i>	0.48
<i>garden egg</i>	0.48
<i>pot</i>	0.49
<i>compound</i>	0.50
<i>day</i>	0.50
<i>elephant stalk</i>	0.50
<i>story</i>	0.50
<i>mushroom</i>	0.51
<i>fish</i>	0.51
<i>grasshopper</i>	0.52
<i>wingless termite</i>	0.52
<i>zinc</i>	0.52
<i>banana</i>	0.52
<i>gizzard</i>	0.54
<i>cap</i>	0.54
<i>cup</i>	0.54
<i>spider</i>	0.54
<i>pepper</i>	0.54
<i>blood</i>	0.54
<i>knife</i>	0.55

Concluding questions

- How can we bring individual-level sociolinguistic data more directly into the analysis?
- How can we include data on grammatical patterns?
- To what extent are the Lower Fungom patterns representative of the rest of Bantoid?

