

WHAT KIND OF LANGUAGE IS SWAHILI?*

THOMAS J. HINNEBUSCH

Introduction

Recently we have seen the appearance of an interesting and provocative book on the Swahili. This book, by Ali Amin Mazrui and Ibrahim Noor Shariff (1994), takes a serious look at the question of Swahili identity and origins.¹ This paper has at least two goals. One is to help define the nature of the debate about origins, and in so doing I will explicate and critique the Mazrui and Shariff hypothesis. The second is to reiterate the theme of the study of Swahili by Derek Nurse and the present author (1993), entitled *Swahili and Sabaki: A Linguistic History* (hereafter N&H). The linking of Swahili and Sabaki in the title was deliberate: the history of Swahili is inextricably intertwined with that of Sabaki and we cannot speak of the former without direct reference to the latter.

The paper is divided into several sections. The first reviews the position taken by Mazrui and Shariff, the second discusses the view of N&H, implicit in their work on Sabaki, that Swahili is an integrated development from its African heritage, the Sabaki languages. Finally, a critique of the Mazrui and Shariff hypothesis will conclude the paper.

The Mazrui/Shariff Hypothesis.

Mazrui and Shariff (hereafter M&S) present an intriguing proposal for the origin of Swahili. They speculate, based on their reading of the Periplus of the Erythraean Sea, a 1st century document written by an early Greek traveler to East African territories (see Appendix 1 for the relevant text), that the Swahili language developed from a pidgin of Arabic, that is, one whose basic vocabulary was drawn from Arabic. Their reconstruction can be summarized as follows:

* Since much of the discussion in this paper rests on Nurse and Hinnebusch (1993), I want to acknowledge Derek Nurse's contribution to this paper. Much of his work and analysis is incorporated here and I am indebted to him. The present discussion and interpretation are my own.

¹ This book is one of several that have recently appeared that deal with Swahili history and identity and which to some extent discuss the role of language and/or linguistics to shed light on the question of identity. Two others are: de Vere Allen (1993); Middleton (1992).

- 1 By approximately 100 AD Arab traders are present on the East African Coast.
- 2 In order to facilitate communication with inhabitants along the coast an Arabic-based pidgin developed, but the acrolect (Arabic) only provided lexicon and no grammar; "its grammar was based predominantly on primary universal patterns" (67). They are unclear to what extent the basilect(s) (Bantu languages) provided non-lexical structure.
3. Arab-speaking traders begin to intermarry and produce offspring who initially learn the pidgin which then in the normal course for many pidgins develops into a creole.
4. Decreolization occurs as a function of contact and association with substratum languages (Bantu languages); Swahili, as we know it today, develops.

They recognize that it is very difficult to adduce any "linguistic evidence for this Swahili genesis..." (67). One outcome of decreolization is the effect that "a decreolized language will have lost all traces of its pidgin and creole genesis" (65). Subsequent linguistic methodology and the Comparative Method will identify the decreolized language with one or more of the substratum languages and would be unable to detect its pidgin/creole origins. As evidence, they state that "the hypothesis would certainly explain Swahili's substantial proportion of Arabic words, its lack of a tonal system that is characteristic of virtually all other neighboring African languages, and the fewer distinctions in its concord system" (67). We will return at the end of the paper to examine some of these claims in light of the following presentation of aspects of Swahili grammatical structure.

Swahili, a "different kind" of Bantu language? While observers have recognized that Swahili is a Bantu language², this recognition has usually developed as a consequence of long distant comparisons between Swahili and a general notion of what constitutes a Bantu language typologically. Some have remarked on the fact that Swahili has absorbed huge amounts of foreign loans often concluding that Swahili is lexically quite unlike other Bantu languages. The result has been a string of quite odd and ambivalent characterizations of Swahili, some of which are critiqued by Mazrui and Shariff (also see N&H:36-37). The focus in this section is to show that Swahili is not atypical either linguistically or sociolinguistically; moreover, that it developed out of a very specific Bantu matrix. Any discussion of Swahili origins that purports to contribute to the debate of what kind of language Swahili is, or who the Swahili are, must take this into account. This was indeed the implicit, if not overt, conclusion of Nurse and Hinnebusch (N&H 1993:308): "...most Swahili dialects are no more different from other Sabaki languages than the latter are from each other." The view espoused there and in this article is that Swahili is only a different kind of Bantu language to the extent that the Sabaki languages (for a list see Appendix 2) and their matrix, the Northeast Coast Bantu (NEC) languages - all Swahili's nearest linguistic neighbors - are different kinds of Bantu languages. Swahili, beyond the specific character of its lexicon, is no more a special case than other Sabaki and NEC languages. Even with its lexicon, Swahili is as typical as any other

² Mazrui and Shariff recognize that Swahili is classified as a Bantu language: "less because of its vocabulary, but more because of its grammatical structure" (67).

language that has a history of contact with other languages; examples are many but I mention just three: Persian and Turkic languages with massive Arabic loaning, and the comparable case of English with Romance and other loans.

The Sabaki Matrix.

The languages and how we distinguish them. In N&H the evidence and reasoning for reconstructing proto Sabaki is laid out as well as the justification for considering Swahili a Sabaki language. Both lexicostatistics (hereafter LS) and the Comparative Method are relied on to reconstruct the proto matrix from which Swahili and its other sister languages evolved. Earlier work, and partially N&H, establish the linguistic context from which Sabaki itself developed, namely, the Northeastern Coastal Bantu (NEC) languages. It is out of this group that Sabaki along with Seuta (Shambala, Bondei, etc.), and Ruvu (Luguru, Zaramo, Gogo, etc.), plus Pare, developed.³ Lexicostatistics and lexical studies (see Chapter 3 in N&H) have been important for characterizing in specific, measurable and testable terms, the lexis of Swahili. CB has been crucial for specifying in precise terms how Swahili developed and evolved from its matrix.

Within Sabaki we can demarcate several distinct groups as well as individual languages which make up the division. Some of the Sabaki groups, such as Swahili, Mijikenda, Comorian, and Pokomo are either dialect clusters or dialect continua. The other members, Mwani, Mwiini, and Ilwana (Elwana), are single languages with minor dialect divergence if any. For greater detail see Appendix 2 or the reader can consult N&H (1993:266-285).

The Swahili continuum. Swahili is spoken along a 1000 mile East African coastal strip and offshore islands, etc. It is subclassified into two groups. There is a northern dialect group (hereafter ND) consisting of Tikuu (Ti) - also known as Bajuni, Siu (Si), Pate (Pa), Amu (Am), mainly spoken in the Lamu archipelago and surrounding areas of northern Kenya; another part of ND includes Mvita (Mv), Jomvu (Jo), Chifundi (Chi), and some poorly studied, marginalized dialects; all of the latter are spoken north and south of Mombasa or in Mombasa itself. The other Swahili group, a southern set (hereafter SD), is made up of the dialects of Zanzibar and Pemba and of the Tanzania coast: Unguja (Ung), Makunduchi (Mak), Tumbatu (Tu), Pemba (Pe), Vumba (Vu), and Mtang'ata (Mt).⁴ It should be emphasized that while the Swahili dialects do not consist of an unbroken, geographically uninterrupted, chain - although parts are, the various communities have maintained connections mainly through

³ Other East African groups to which NEC most closely fits genetically, such as Southern Highlands (Hehe, Bena, etc.), etc., will not concern us in the discussion.

⁴ Also presumably part of the Southern Dialects of Swahili would be the Swahili spoken in Swahili communities along the southern Tanzania coast up to but not including the parts of northern Mozambique where Mwani is spoken, namely, those in Kilwa and on Mafia Island. These forms of Swahili speech are not addressed here because of a paucity of data (e.g., for Mafia) or because either a form of Kiunguja or Standard Swahili has replaced the original dialect (e.g. Kilwa). For more detail see N&H:12-14

trade and social intercourse via the sea lanes.⁵ Further, because of how they are strung out over such a large area various subparts of the chain have had different contact experiences, some of which can be seen in the kinds of non-Arabic loan words present in the different subparts.

Mwiini (Mw), spoken in southern coastal Somalia centered on the town of Brava, can either be considered as part of the Swahili subgroup, thus a Swahili dialect, or as a distinct language, depending on one's interpretation of the lexicostatistical results (N&H 1993:278-281). It is, however, not mutually intelligible with other Swahili dialects and in this paper we treat it as a separate language, albeit one that is very closely related to Swahili and that on other than lexicostatistical criteria, it could be argued, places it within Swahili. Mwiini, spoken as it is in southern Somalia, is situated just north of the Swahili dialects of the Lamu area and shares with those dialects high percentages of basic lexis and other linguistic material.

How Swahili is Sabaki. Beginning with European scholars' scrutiny of Swahili and other languages of the East African coast, disproportionate attention has always been paid to the lexical content of Swahili. Let's look at the facts.

1. Lexicostatistical studies place Swahili squarely within Sabaki (N&H:266-281). There are no other languages or groups of Bantu languages in Eastern Africa to which Swahili is related in this way.

2. Within Sabaki, Swahili has the same LS status as most of the other dialectally complex major subgroups, namely Pokomo and Mijikenda (see Fig. 1). Swahili overall has retained about the same percentage of basic Bantu vocabulary in a 100-word list as Mijikenda and Pokomo, but somewhat less than has Comorian:

Figure 1. Sabaki Internal Averages⁶

Comorian	81%	Pokomo	73%
Mijikenda	73%	Swahili	74%

3. As a language, the Swahili dialect continuum is as cohesive lexically as the Pokomo dialect chain or as the Mijikenda cluster of dialects, but not as cohesive as Comorian (Fig. 1). By "cohesive" is simply meant that a group, in general, among its various member languages

⁵ For greater detail and explication on each of dialects of the Swahili continuum see N&H 1993:5-14. While this paper and the research on which it is based focuses on the dialects named here, there are other named dialects which we do not discuss, mostly because of a lack of data.

⁶ An internal average is the average of the percentages of common vocabulary (say of a 100 item basic list) shared among each member of the group with each other member of the group, apart from what percentages of commonly shared vocabulary members of the group might share with languages outside the group. Thus, for example, the Comorian 81% is the average of each Comorian dialect comparison of commonly shared vocabulary with other Comorian dialects: Ngazija-Mwali 81, Ngazija-Nzuani 77, Ngazija-Maore 80, Mwali-Nzuani 83, Mwali-Maore 84, Nzuani-Maore 83.

is "lexically uniform." We can say that the higher the internal average the more uniform the language group or dialect cluster is, and conversely, lower figures indicate greater lexical diversity among the members of a group. Thus, lexically speaking, the Comorian dialects are more internally similar than any of the other Sabaki groups including Swahili.

4. Parts of Swahili are more cohesive than others (see Fig. 2) ⁷:

Figure 2. Swahili internal averages

ND (Am, Pa, Si, Ti) 89%	SD (Ung, Pe, Vu, etc) 72%
ND + Mombasa area 82%	Zanizbar and Pemba dialects 72%

Instead of "cohesiveness" we could also look at these figures as measures of "conservatism." That is, some parts of the Swahili continuum have undergone less erosion of their core Bantu vocabulary and have retained more core vocabulary. There are two points to be made here: a) Swahili is not a monolith; it is dialectally diverse and its parts have different histories - a fact that has to be taken into account in any discussion of origin. b) Parts of Swahili are comparable to subparts of Sabaki which also retain high percentages of core Bantu vocabulary. For example, the Northern Swahili figure of 82% is comparable to the Comorian figure of 81%.

Lexical content of Swahili and Sabaki languages. Apart from the above, there is insight to be gained by looking at the nature of the basic core vocabulary of Swahili and other Sabaki languages and its content.

Swahili lexis is characteristically Sabaki in respect to lexical origins. One test of this can be seen in the 100 basic word list that was used to calculate the lexicostatistical similarity indices (N&H 1993:274). Basic word lists (Swadesh 1952) tend to be stable and less vulnerable to borrowing or replacement than other sorts of semantic-based sets of vocabulary, thus if Swahili is inherently a Sabaki language then it should show characteristics similar to the other members of the group. This is what we find:

⁷ It should also be emphasized here, that while the ND set stands out lexicostatistically, the SD set does not. Lexicostatistically Swahili consists of a core cluster of ND dialects with the others chained incrementally to the core in a typical continuum. Thus, it is only ND that is positively subgrouped while SD is not, although parts of SD do subgroup, e.g., Pemba, Tumbatu, Makunduchi, and Unguja. However, the distinction of ND and SD depends only partially on LS; there are other confirming linguistic criteria (N&H 1995: *passim*)

Figure 3. Sources of Lexis in Sabaki Languages, based on Nurse (N&H 1993:324ff.)

	CB	PSA	Loans	Other
Swahili ⁸	72	9	9.5/5 (14.5)	3
Comorian	81	6	7/5 (12)	1
Mwiini	77	10	6/5 (11)	2
Mwani	69	9	15/3 (18)	4
Mijikenda	74	6	13/0	7
Pokomo	76	6	15/0	3
Ilwana	69	7	19/0	5
Average	74	7.5	12/3 (15)	3.5

CB = Common Bantu; PSA = Proto Sabaki (includes 1-2% Northeast Coast or Swahili innovations) "Other" includes items whose origins are unknown. In the "Loans" column the split figures are local loans versus Arabic (or other Indian Ocean languages) respectively.

Figure 4 Sources of Lexis in Swahili Dialects, based on Nurse (N&H 1993:324)

	CB	PSA	Loans	Other
Northern dialects	80	11	4.5/2	3
Mombasa dialects	72	9	10/5.5	3
Southern dialects	68	8	14/5.5	4

These figures show that Swahili is lexically similar and comparable to other languages of its matrix:

1. Swahili retains Bantu lexis in percentages that parallel other Sabaki languages; the 72% of retained lexis matches closely the 74% average for the rest of Sabaki (Fig. 3).
2. Swahili's attestation of innovative Proto Sabaki lexis is similar to other Sabaki languages (Fig. 3).
3. Loan material in Swahili from all sources, including Arabic, is not appreciably nor significantly different than the amount of loan material we find in other Sabaki dialects. The 14.5% figure of total loans is in line with the 15% average for Sabaki (Fig. 3).
4. Within Swahili, SD (dialects of Zanzibar, Pemba, and adjacent Mrima coast) show greater lexical intrusion from loaning languages, whereas ND of the Lamu archipelago shows 50% less intrusion. (Fig. 4).⁹

⁸ The figure for Swahili here is the average of figures for Ti, Si, Pa, Am, Mv, Jo, Chi, Vu, Mt, Pe dialects, Ung, Tu, and Mak. Since it is an average it conceals the high averages of Bantu-lexis retention for the Lamu cluster and the somewhat lower figure for the Zanzibar cluster. See Table 2 (where the breakdown is given).

5. In the Loan category an interesting contrast exists when we compare coastal Sabaki languages (Sw, Com, Mw, and Mn) versus hinterland Sabaki languages (MK, Po, El). Coast Sabaki languages have about 5% foreign loans (mostly Arabic in origin) while the hinterland group has none in the 100 word list (Fig. 3). The hinterland group also stands out in having larger percentages of local loans, either from other Sabaki languages or the area's other Bantu languages. Swahili's pattern here is closer to the hinterland group than to either Comorian or Mwiini which are located on the periphery of the Sabaki zone and thus less likely to be influenced by local loaning whereas they presumably were subject to the same forces that led to the borrowing of Arabic loans.

6. Basic vocabulary has only moderately been affected by foreign influences as seen in the study of the 100 word list. In a similar test applied by Nurse to a larger 1000 item list of relatively culturally neutral vocabulary for a selection of Swahili dialects (Ti, Pa, No, Pe, and Ung; N&H 1993:327) there are somewhat complimentary results (see Fig. 5 below). The majority of lexical items in the four dialects, ranging from 66% to 69%, is inherited lexis, but there is a larger number of loans from 17% to about 22%, but there is also a much larger set of items of unknown provenance (11 - 14%); among this latter set could, of course, be loans and this would increase the percentages of loans, both local and foreign. Important here, however, is the fact that foreign loans (Arabic, etc.) are still a smaller figure relative to the total set (usually about 50% of the total number of loans). Moreover, these percentages are in line with what we know of other Sabaki languages and other East African Bantu languages (N&H 1993:327). Swahili is not unique in borrowing and we do not have to go far afield to document that fact, e.g. English (Fig. 5). It's only the particulars that are different, dependent as they are on a different cultural-historical mix of variables. For example, Swahili communities, and communities of Comorian, Mwani, and Mwiini speakers, have for a long time been in contact with non-East Africans; whereas, in contrast, other Sabaki groups have had contact both with speakers of hinterland languages, as have Swahili speakers. For example, Ilwana has large loan sets from Cushitic languages, and Mijikenda from Central Kenyan Bantu languages (N&H 1993:332)

⁹ This certainly cannot be because ND has come under less external Arabic pressure than the others. The influence of Islam, for example, is as strong if not stronger and the northern islands have long been recognized by scholars as the heartland of Swahili classical literary tradition. Both have been long term operators in creating opportunities for, especially, Arabic loan intrusion. Bertoni's study (1973) of lexical frequency is clear on this matter

Figure 5. Sources of lexis in 1000-word list

	ND		SD		
	Tikuu	Pate	Pemba	Unguja	(English)
Inherited	68	69	66	65.5	(83)
African loans	9	9	8?	7.5?	
Sw dialect loans	-0.5	2?	4.5?		
Foreign loans	7.5	7.5	10	10	
Total loans	16.5	17	20?	22?	(15)
Swahili-specific	22	1.5	2		
Unknown	13.5	12	12.5	10.5	(2)

NB: Question marks indicate some doubt about the identification of some items in the list.

Arabic Loans in Swahili. At least two sets of Arabic loans can be distinguished in Swahili (N&H (1993:315ff)). One set, by far the majority, is relatively recent in origin having entered the language during the period of Omani suzerainty which started in the seventeenth century (also see Zawawi 1979:3). The other set is an older body of loan material also from Arabic (some ultimately from Persian). We are able to make the distinction because:

1) The recent Omani set has a high degree of semantic and phonological identity between the Omani source and synchronic form of the loan in Swahili (making adjustments for adaptive phonological processes). Furthermore, this set does not exhibit earlier phonological processes that are attested in Sabaki.

2) The older pre-Omani set of vocabulary does, on the other hand, exhibit regular Sabaki shifts with a wide distribution of reflex items throughout Sabaki, and thus looks either like Proto Sabaki or Proto Swahili items.¹⁰ Some examples are *-kant- 'cut' (< Arabic/Indian languages), *lwavu 'fishing net' (< Persian/ Indian languages), *mpula 'steel' (< Persian), *mtungi 'waterpot' (< Persian), *ngamila 'camel' (< Arabic), etc. Others are likely to have entered Proto Swahili even earlier; some examples are *lubani 'incense' (Arabic), *musikiti 'mosque' (Arabic), *kanzu 'men's garment' (Arabic), etc. (See Nurse in N&H, 316ff).

¹⁰ Any irregularity is comparable to the irregularity seen in nonborrowed inherited lexis and can be attributed to later interdialectal borrowing.

The Sabaki character of Swahili.

When we look at other aspects of the grammar of Swahili, it is apparent that much of what we think of as being "Swahili" is not unique to the language; many of its features are Sabaki features and shared with neighboring Sabaki languages. Many of these features are inherited from even earlier intermediate proto periods, from NEC, or even earlier stages. Typically, others are due to historical events involving social interaction among various Sabaki-speaking groups.

Phonology. Some of these features in the area of phonology are:

1. **The reduction of 7 Bantu vowels to 5.** The shift from a Bantu seven-vowel system to a five-vowel one in Swahili is part of a general areal phenomenon in the region and is widely attested in NEC and beyond in adjacent groups to the south. There is no evidence that the reduction is due to non-East African external influence (N&H:176-177).

2. **Reduction of tonal distinctiveness.** Swahili has lost tone. Philippson argues (in N&H:248-265) that Swahili simply represents an endpoint of an areal development that other Sabaki languages are continuing today.

3. **Aspirated voiceless stops.** Swahili contrasts a series of voiceless stops with voiceless aspirated stops (p, t, č, k and p^h, t^h, č^h, k^h). The aspirated series derives from sequences of nasal plus voiceless stop as they do in Mwiini, Mijikenda and Pokomo and parts of NEC (for details see N&H 1993:155-162).

4. **The loss of *l** in certain environments in Swahili is well documented (e.g., *-ikala > -kaa 'live, sit, stay'). The initial stage of *l-loss (before *u) is attested not only by Swahili but by Lower Pokomo and Comorian, whereas the other Sabaki languages (Mw, UP, El, MK, and Mn) preserve *l before *u, e.g., -uma versus -luma 'bite' (< *-luma).

There are further phonological changes that only parts of the Swahili continuum share with other Sabaki languages. Some examples of these are:

1. **PSA *f and *v.** In the Swahili dialects of the Lamu area PSA *f and *v before the PSA H(igh)V(owel) *i shift to /s/ and /z ~ ɖ/. This is a change that the Northern Swahili dialects share with Comorian respectively while in the Southern dialects we find /f/ and /v/ preserved. A few examples suffice: PSA *-fika 'arrive' : -fika (Ung), -sikilia (Am), -siha 'come down' (Ng); PSA *-vil- 'be spoiled' : -via (Ung), -zia (Am), -zila (Ma) (see N&H 1993:118-119).

2. **Strengthening.** In the Southern Swahili dialects voiced stops are attested where other Swahili dialects attest nonstops in the same environment following a reconstructed front HV *i; thus: -iŋa 'steal' (Ung, Mak, Vu, etc.); -iwa (Am) < PSA *-iwa, CB *-yŋb-; domo (SD) 'large lip'; omo 'prow' (Am) < PSA *ilomo 5, CB *dòmò. The attestation of /b/ and /d/ is due to the effect of the preceding front HV *i (N&H:133-145). A general form of the shift is

attested in Comorian; it applies to all PSA consonants preceded by the front HV,¹¹ e.g., -hiḃa 'steal' (Nz) (cf. -iḃa 'steal' in Ung, Mak, Vu, etc.); -iwa (Am) < PSA *-iwa, CB *-yīb- and ḃomo/malomo 'lip' (Ng, Ma).

3 **Lenition.** The Mijikenda languages, Pokomo, and Comorian attest a change in which *p and *t are usually attested as /h, ɸ, β, or v/ (< *p) and /h, ɾ or ɾ/ (< *t) respectively in intervocalic position; in Comorian *k is also weakened becoming /h/, e.g., PSA *-pangul- 'wipe': ɸangula- (UP), -ɸanguya (LP), -hangusa (Gi), -βangua (Ng), -vangua (Nz) and PSA *-tapik- 'vomit': -ɾaɸika (UP), -haɸika (LP), -βaβika (Du), -raβiha (Ng), -raviha (Nz). The Swahili dialects which share features of lenition are Makunduchi, in which only *p is lenited, and Vumba and Chifundi in which both *p and *t have lenited reflexes: for example: -raβika 'vomit' (Vu), -tavika (Mak); cf. -tapika (Am, Mv, Ung); mfuḃa 'bone' (Chi, Vu); cf. mfupa (Am, Mv, Ung); and mri ~ muri 'tree' (Chi), mri (Vu).

4. **PSA *c > [-palatal].** In part of the Sabaki cluster PSA (and CB) *c becomes [-palatal] while the rest of the group, for the most part, preserves the palatality of the proto segment. In Mijikenda, Comorian, and Lower Pokomo, *c > ts ([+dental, -palatal]). Similarly, the Northern Dialects only, and Mwiini, share the shift of *c to a non-affricated dental, thus *c > ɸ ([+dental, -palatal, -affricate]). Thus, we conclude ND innovates along parallel lines with Mijikenda, Comorian, Lower Pokomo, and Mwiini while the southern Swahili dialects, Elwana, and UP are conservative in preserving the palatality of the proto consonant. Examples: PSA *-wicɪ- 'raw': -witsi (LP), -itsi (Gi, Du, Ng, Nz, Ma), wiṭi (Mw, Am), -viti (Ti) but -biči (Ung); and PSA *ijico 'eye': dzitso (LP, MK, Com), iṭo (Am, Si, Pa, Mw), ɟiṭo (Mv) but ɟičo (Ung), zičo ~ ɟičo (Pe), dičo ~ jičo (Mak).¹²

5. **PSA *nj > [-palatal].** There is a parallel change affecting the same languages in which PSA *nj > /ndz/ in Lower Pokomo, Mijikenda, and Comorian, and /nɟ/ in ND and Mwiini:

PSA *-vunja 'break'	-vundza (MK, Com), -vunḃa (Am, Mv), -vuunḃa (Mw) but -vunja (UP, Ung, Mn)
PSA *njala 'hunger'	ndzaa (LP), ndzala (Gi, Du), nda(y)a (Ng), ndza (Nz, Ma), nḃaḃa (Mw), nḃaa (Am, Mv) but njaa (Ung), njaya (Tu, Pe), njala (Mn)

6. **Behavior of PSA *g.** The Northern Swahili dialects show the loss of *g in stem-initial and intervocalic environments in a large number of cases, a change it shares with Comorian. Elwana, Mwiini, and Mwani are regular in losing *g. Though this is not a completely regular

¹¹ The back HV also conditions strengthening in Comorian: mvuḃa 'bellows' (PSA *muvuwa, CB *-gùba, cf. muvowo in Gi); PSA *g is preserved rather than lenited: goho (Ng) 'hollow trunk' (PSA *igogo, CB -gògò) but -haha (Ng) 'go bad' (CB *-gàg-); and the PSA voiceless stops are preserved rather than weakened: paa/maḃaa (Ng) 'roof' (PSA *ipala), trumbo/marumbo (Ng) 'belly' (PSA *itumbo), kapwa/mahapwa (Nz) 'armpit' (PSA *ikwapa). For details and other examples see N&H:133-145.

¹² In Mwani, PSA *c > s. This is the result of an areal change that all languages in the region attest. We assume that earlier stages of Mwani attested /č/.

shift in the Northern Swahili dialects, the Southern dialects, as a rule, preserve a voiced stop reflex of *g. Examples:

PSA *-gawanya 'share'	-aawanya (Mw), -avanya (Ti), -awanya (Si, Pa, Am, Mn), -gawanya (Mv, Ung), -(w)anya (Ng), -anyisa (Ng)
PSA *-caagula 'choose'	-tsagula (Gi), -ɽeua ~ -ɽaua (ND), -čagua (Ung), -ɽaɽa (Mw), -tsa(h)ua (Ng, Nz, Ma)

Morphology. There are two areas from morphology which will reinforce the theme of this section, that of tenes/aspect and the noun class system.

Tense/Aspect. There is considerable diversity in the tense/aspect system of Swahili from dialect to dialect. This is also true of Sabaki and NEC. Moreover, if we only speak of Standard Swahili we miss the richness and the variety of the system. Our goal here will be to show how some aspects of the system - our treatment is not comprehensive - either derive from an older geographically extensive system or are similar, if not identical, to the system in Sabaki or NEC languages.

Swahili exhibits the follow characteristics; we look at some examples as both Swahili features and as Sabaki or NEC features:

1. **Future.** Swahili dialects express the future using a reflex of *-caka 'want' (ni-*ta*-songa mbele 'I will forge ahead', ni-*taka*-vyo-songa mbele 'as I will forge ahead', ru-*cha*-cheka (Vu) 'we will laugh', ri-*tso*-tseka (Nz) 'we will laugh'). This is a pattern that is widely distributed throughout NEC languages involving the use of auxiliaries (the use of reflexes of the verbs for 'go, come, want'), but is not the only pattern.¹³

2. **Nonpast -a-** (< PB *-à-). This is found in the Northern Swahili dialects (Amu: tw-*a*-penda 'we love'; Mvita: yu-*a*-ja 's/he comes/is coming'). It is widespread throughout NEC (Upper Pokomo: mw-*a*-yomba 'you ask for/will ask for'; Luguru t-*o*-ghula (o < a + u) 'you buy/will buy/are buying'; Shambala mw-*a*-dika 'you cook'; etc. N&H:384.

3. **Narrative *-ka-**. This is common throughout NEC to mark narrative and is found in Mijikenda, Pokomo, most Swahili dialects, Pare, and many Ruvu languages (e.g., St. Swahili ni-*ka*-ongea naye 'and (then) I chatted with him/her'; N&H:388)

4. **Past** in NEC languages is complex. There are several ways of expressing past that can be reconstructed for NEC and Sabaki; Swahili makes use of three of them: reflexes of *-á-, the *-V(owel) H(armony) suffix, and the *-ile suffix.¹⁴ Examples:

¹³ Some NEC languages have no overt future and use adverbials or auxiliaries (e.g. Pokomo, Mwani, and most Ruvu languages); the others have discrete future markers, in some languages, but not many, more than one. Apparently all morphemes marking 'future' in Sabaki and NEC languages are innovations.

¹⁴ The fourth is *-Ø-, e.g., Upper Pokomo hu-Ø-cheza 'we played', Mwani: ka-Ø-n-singana 's/he met her/him'; this is not attested in Swahili.

a. *-á- is found in the Northern and Southern Swahili dialects: *tw-a-li-teka mai* (Am) 'we drew water' (< *-a-li- 'past + aux -li- 'be'); *chwe-e-kuya yana* (Si) 'we came yesterday' (-e- < *-a- + (l)i-; *tu-li-fukuzana* (Ung) 'we chased each other' (-li- < *-a- + aux -li- 'be'); *k^ha-rw-a-li-gombana* (Chi) 'we didn't quarrel'; *s-a-fika* (Vu) 'I have just arrived'.

Parallels exists in Sabaki and other NEC languages; for example (N&H:389): *f-a-gula* (MK) 'we bought (far past)'; *w-a-ti-ku-beete* (Mw) 'he had got lost'; *r-a-lima* (Com:Ng) 'we cultivated'; *kw-a-gula* (Gogo) 'we bought (middle past)'; *n-a-dika* (Shambala) 'I cooked (far past)'; *n-a-long-aga* 'I spoke (far past)'.

b. *-VH (Vowel Harmony) Suffix occurs in SD (but not Unguja nor Standard Swahili) and Comorian with the meaning 'past' in SD and 'past/perfect' in Comorian; the tense is marked by a copy of the root vowel in wordfinal position, and in SD by a special set of subject prefixes, *ku-* and *ka-*, for 2nd and 3rd persons respectively. It marks positive past in SD and past/perfect in Comorian:

Southern Dialects of Swahili		Comorian	
<i>kw-Ø-aw-a</i>	'you came out'	<i>tsi-Ø-law-a</i>	'I have come/came out'
<i>kw-Ø-end-e</i>	'you went'	<i>and-e</i> (/a-Ø-end-/)	'he went/has gone'
<i>ku-Ø-pik-i</i>	'you cooked'	<i>a-Ø-him-i</i>	's/he (has) stood up'
<i>ka-Ø-ni-ngo-j-o</i>	'he waited for me'	<i>tsi-Ø-mw-on-o</i>	'I saw/have seen her'
<i>ka-Ø-uz-u</i>	's/he sold'	<i>a-Ø-hul-u</i>	's/he (has) bought'

c. *-ile 'general perfect/past' (N&H:372-373, 393-394). Although reconstructed for CB, NEC, and Sabaki, its reflexes are only attested in northern Sabaki languages including the Lamu area dialects of Swahili, Mwiin, Ilwana, Pokomo, and Mijikenda; while loosing ground in modern ND, it is widely attested in Swahili poetry: *u-Ø-gur-ie* (Pa) 's/he moved home'; *u-Ø-elew-ee* (Pa) 's/he understood'; *k^h-u-Ø-las-ile* (Sw. poetry) 'you never ceased' (-lat-) (Miehe 1979:190); *Ø-peenz-ele* (Mw) 's/he liked' (-penda); *Ø-pis-ile* (Mw) 's/he passed' (-pita); *u-Ø-jar-ie* (El) 'you have gotten old'; *a-Ø-som-ee-o* (UP) 's/he who has read'; *k^ha-fu-Ø-gul-ire* (Gi) 'we didn't buy'; *yu-Ø-rer-e* (Gi) 's/he is asleep = 's/he has slept' (-lala).

The diachronic (and synchronic) morphophonemics of this suffix are interesting. Some of the Sabaki dialects have lost *l (e.g. Pate), and the vowel *i triggers stem-final consonant mutations, e.g., *u-Ø-pis-ie* 'he passed (Ti)' < -pit-; *u-Ø-vush-ie* 'he crossed (Si)' < -vuk-; *u-Ø-fund-ie* 'he closed (Am)' < -fung-; (N&H:372). Though restricted in Swahili, there are accounts that it once was present in the Mombasa area dialects. It is curious that the only Bantu languages in East Africa that do not attest the suffix are Comorian, Chifundi, and SD including Standard Swahili.

Noun Class System. We can posit 17 noun classes for NEC and Sabaki languages (see Appendix 3). Swahili aligns nicely with the other Sabaki languages and the characteristics of its noun class system match closely those of Sabaki and many NEC languages. It has lost the preprefix (still attested in Comorian and Lower Pokomo and most dialects, including Comorian, have lost the Class 12 *ka- diminutive, but reflexes still exist in the ND and in

most other Sabaki languages. Thus, the noun class system of Swahili is not atypical or unusual when viewed in the context of its matrix.

The Sabaki/Swahili Nexus.

If we plotted all these similarities and differences on a map it would be clear immediately that Swahili as a language, and therefore as a society and culture, is inextricably tied in with its Sabaki relatives in complex and intricate ways. A map of this sort can be found in Appendix 4. Swahili is not a simple monolithic entity. Quite the contrary. It exhibits all the richness and subtlety of a natural language that has followed normal linguistic progression and change through time and space and has been subject to the same sociocultural historical forces as its nearest relatives and neighbors. In just these few examples from phonology and morphology, the diversity that we find in Swahili is paralleled by diversity in Sabaki. The shift of PSA *c > [-palatal], the shift of PSA *f and *v to /s/ and /z/, and the behavior of PSA *g distinguishes ND from Zanzibari/Pemba and other SD. These same shifts are characteristics of parts of neighboring Sabaki languages. Other shifts, such as the aspirated/non-aspirated contrast in all of Swahili is also manifest in major portions of other Sabaki languages.

In visualizing the isogloss map (Appendix 4) representing these facts, one should try and imagine that the map represents the result of historical forces that happened at different times and over time. The map represents strata that have temporal and spatial dimension. Such a map is not the product of a single event, but a series of events through time, and each event has its own history (sometimes opaque, sometimes not). For example, to explain the widespread attestation of aspiration in Swahili and other Sabaki languages we have to assume that the change innovated at a time when the Swahili community and the Mijikenda and Pokomo communities were interacting in ways that allowed the shift to spread, or possibly were even part of the same intermediate proto community. Similarly, to explain a change, such as PSA *c > [-palatal] in ND (but not SD) and in Mijikenda, Lower Pokomo, and Comorian, or the shift of PSA *f and *v to s and z in ND and Comorian points to strong interaction among some Swahili and some Sabaki communities after the Swahili group became dialectally complex or at a time when the Swahili or the pre-Swahili had spread, but were not completely distinct from their Sabaki neighbors. The behavior of PSA *g suggests strong links between Comorian communities and Northern Swahili communities but not with Southern Swahili communities, but on the other hand, Strengthening which SD shares with Comorian attests to links at other times. Finally, when Lenition affected all of Mijikenda, Pokomo, and Comorian, Swahili did not participate except for a small subset of socially marginal dialects in the Vanga area (see Nurse and Walsh 1992) which have recently been impacted by neighboring Mijikenda languages. At times we see blending and merging, at other times we see the opposite.

From an examination of this and other data we conclude that the connection between Swahili as a whole and the other Sabaki languages, on the one hand, and interactions among the various parts of Swahili, on the other, is the result of a complex linguistic and sociohistorical interplay that has been ongoing for centuries, the roots of which are to be found partly in the Bantu/Sabaki matrix, partly within Sabaki itself, and partly within the Swahili continuum.

The Decreolization Hypothesis -- the outlines of a critique

I do not intend to refute the M&S hypothesis point for point. However, I do want to develop the general framework of an argument.

1) First, we need to repeat, following M&S themselves, that there is no linguistic evidence either adduced by them, or any that I am aware of, that supports a hypothesis that Swahili was ever an Arabic-based pidgin.

2) While M&S describe this early pidgin as *Arabic-based* lexically, they are silent about the nature of the pidgin's other components, about the nature of the basilect (presumably Bantu). They only speak of substrate languages in the context of decreolization.

3) Their scenario is based on a reading of the *Periplus of the Erythraean Sea*, a document describing 2nd century East Africa. While an important document, it should not be the only basis for reconstructing Swahili history since, importantly, it never identifies the ethnicity nor linguistic affiliations of the peoples inhabiting the coast at that time. In my view, an African-based pidgin is just as likely a reconstruction based on a reading of the *Periplus*. But there is so little information given in that document that at best we are guessing about such matters. Further, there is a great deal more about the history of East Africa that is known and must be taken into account in doing such a reconstruction (see Nurse and Spear 1985).

4) Other than citing the date of 100 AD for the presence of the Arabic-based pidgin and the speculation that the creole developed about the same time, M&S provide little detail on the chronology of the creole continuum after that date or how this might coordinate with what we know from archaeology and other sources of historical information.

5) The M&S scenario would presuppose that the Arab merchant traders were socially dominant and/or were trade initiators. This would be a sociocultural requirement for the appearance of an Arabic-based pidgin; however, it is more likely the other way around: Cushitic speakers, Bantu speakers, Sabaki speakers, indeed, pre-Swahili communities, all major players in East African history, depending on the time frame, have always been the socially and commercially dominant element in East African history, already in control of coastal trade, the founders of coastal settlements, and the major players in all domains. In such a context Arabic is an unlikely candidate for creole status, let alone pidginization, because the dominant core of East African society has always been African, and not Arabic. Normal

unbroken linguistic transmission from generation to generation would be the norm and would adhere to similar models elsewhere. South Arabians, Persians, and others, were simply other layers superimposed on pre-existing trade associations.

6) M&S claim that "the overwhelming evidence of an Arabic lexicon" is evidence for an "Arabic-based" pidgin. The Arabic content of Swahili can more simply be explained by ordinary borrowing mechanisms. We do not need to appeal to a complex hypothesis of pidginization and decreolization to explain Arabic lexis in Swahili. The major flaw in their argument is the fact that the basic core vocabulary of Swahili is Bantu in origin. All cases of X-based pidgins or creoles that have been described in the literature, that I am familiar with, preserve the "base" or "core" vocabulary in the developmental continuum from the acrolect through decreolization. Their hypothesis says nothing about the relexification of the core vocabulary from Arabic to Bantu in content.

7) M&S explain away the lack of Arabic grammar elements in the hypothetical creole by claiming that the grammar of the creole "was based predominantly on primary universal patterns." In all cases of well documented creoles there is no creole which does not exhibit morphological and syntactical elements from the acrolect. If X-based pidgins are lexically oriented in a certain direction we can expect there to be morphological and grammatical elements from the same source. Creoles simply do not develop only in terms of "primary universal patterns." The literature is fairly clear on this point and we don't see a simple transference of features from acrolects to basilects (see Bickerton's work on decreolization and the creole continuum, 1980).

8) The predominantly culturally oriented nature of Arabic lexis in Swahili, can more easily be explained by normal borrowing transfer mechanisms. While there is a relatively large amount of Arabic lexis, none of the Arabic morphological apparatus that was borrowed along with lexical items is productive (e.g., the *ma*, *mi*, and *mu* elements of Arabic loans) is productive, and, moreover, has been relexified in terms of Bantu/Sabaki morphological structure (see Zawawi 1979:39ff). Furthermore, the frequency of Arabic vocabulary in present-day written Swahili corresponds to a continuum of literary type ranging from informal textual material of a journalistic character, through modern texts (novels, short stories, translations, etc.) to ancient texts (poetry, epics with Arabic or Islamic themes). The highest percentages of loans from Arabic are found in the later category and the lowest in journalistic texts (Bertoncini 1973:302).¹⁵ This gradation is characteristic of loaning, not pidginization nor decreolization.

9) M&S cite the presence of massive numbers of Arabic loans in Swahili as evidence of an earlier Arabic-based pidgin. However, most loans date from relatively recent times, long after

¹⁵ In Bertoncini's study of frequency in Swahili lexis her results show that 15% of the one hundred most frequently used words are "Oriental" in origin; the percentage goes up to about 37% in the one thousand most frequently used set of vocabulary. Her study was based on a list of 1,143 entries taken from a count of 40,000 occurrences in 100 sources.

Swahili attained its present form and structure. Only a very small part of that material can be dated earlier, and that material can be explained by normal borrowing.

10) M&S see a "reduced" noun class system in Swahili, and the loss of tonal distinctiveness, as best explained by their hypothesis. However, the Swahili noun class system is virtually identical to the Sabaki and NEC systems. The loss of tonal distinctiveness is an areal phenomenon, related to the general areal shift throughout Eastern Bantu involving the shift of "pure" tone systems to pitch-accent systems. These aspects of Swahili's grammar can best and more simply be viewed as "spontaneous" change and we need not appeal to decreolization of an Arabic pidgin to explain them.

11) If creolization from an Arabic pidgin played any role in the development of Swahili then we would see significant chunks of Arabic structure along with Bantu chunks making up Swahili (see the Mbugu/Ma'a case in Tanzania). Other than some prepositions and conjunctions (*kabla ya*, *baada ya*, etc., which again can be explained by normal lexical borrowing (not morphological or syntactic), we don't find structural evidence of such a sort.¹⁶

12) Swahili, as seen above, is a structurally complex Bantu language that fits in with its matrix in equally complex ways that cannot be simply explained away by a (de)creolization process. Swahili shows a complex interplay of features with other languages of the area. Some of these may have been generated in non-Swahili communities and spread into Swahili (e.g. *l-loss, nasal-rules, etc.) and others have been generated in Swahili and radiated out to others (e.g., loss of tonal distinctiveness), and some Sabaki processes have only marginally or non-uniformly affected Swahili (e.g. *p- and *t-lenition). Such complex dialectal facts cannot be explained by a simple (de)creolization process. A more appropriate model, or set of models, are normal generation by generation historical change, areal change, sociolinguistic interaction, and borrowing. Hymes (1971:5) has remarked that pidginization and creolization represent the extreme to which social factors can go in shaping the transmission and use of languages. In our case here, extreme explanations throw little light on the question of Swahili origins.

While Arab traders were surely present on the East African coast, what is the evidence that they ever constituted more than a handful of adventurers, settlers, or traders in the time of *Periplus*? (It was only in the late centuries of the current era that we can speak of Arab suzerainty on the coast, much too late to help support the M&S hypothesis). Thus were social conditions present for the development of an Arabic-based pidgin in the early centuries of the first millennium? Not likely; nor could Islam have been a factor at the time of the *Periplus*. The first mosques only began to appear around 1000 AD at about the time the earliest pre-Swahili communities were developing. Even supposing the development of an Arabic-based pidgin, what we know of the development of the "creole continuum" elsewhere would predict

¹⁶ Lest we over-simplify here, pertinent to the discussion, especially in regard to the role of Arabic contact in shaping Swahili, is Wald's work on substratal and superstratal influences on Swahili (Wald, to appear).

creolization in the direction of the acrolect (viz. Arabic; also see the cases of French-, Portuguese- and English-based pidgins)) and not the basilects (viz. Bantu). If as Nurse and Spear (1985) argue in their history of the Swahili that the coastal-dwelling Sabaki-speaking populations (from which the nascent pre-Swahili and then identifiable Swahili communities developed) were already engaged in trade, even assuming Arab participation, social conditions would have presupposed the development of a Sabaki-based pidgin, not an Arabic-based pidgin. Modern, or pre modern, cases of Arab or other settlement on the coast that we know most about are the Omanis, Mazruis, and Nabhanys, and even earlier, the Shirazi. All were linguistically absorbed into pre-existing Swahili communities. None of these communities exercised the linguistic dominance to prevent absorption, let alone provide the conditions for the development of an Arabic-based pidgin. Arabic speakers learned the local dialects and were absorbed. Why would the earlier times about which we know considerably less call for a different model? Other modern models are equally enlightening. English-African interaction in East Africa produced Swahili-based pidgins as did Indian-African interaction, not English-based nor Hindi-based pidgins. Even assuming an early Arabic-based pidgin, it could not have been long lived. It disappeared as it was aborning and surely could have played no role in explaining what kind of language Swahili is today.

References

- Allen, James de Vere. 1993. *Swahili origins: Swahili culture and the Shungwaya phenomenon*. London: James Currey; Nairobi: E. A. E. P.; Athens: Ohio University Press.
- Bertoncini, Elena. 1973. "A tentative frequency list of Swahili words." *Annali dell'Istituto Orientale di Napoli* 33: 297-363.
- Bickerton, Derek. 1980. "Decreolisation and the Creole continuum." *Theoretical orientations in Creole studies*, ed. by Albert Valdman and Arnold Highfield. pp. 109-127. New York: Academic Press.
- Hymes, D. ed. 1971. *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Mazrui, Alamin M. and Ibrahim Noor Shariff. 1994. *The Swahili: Idiom and identity of an African people*. Trenton, New Jersey: Africa World Press.
- McCrindle, John W. 1879. *The commerce and navigation of the Erythraean Sea*. Amsterdam: Philo Press (Reprint 1973).
- Middleton, John. 1992. *The world of the Swahili: An African mercantile civilization*. New Haven and London: Yale University Press.

- Miehe, G. 1979. *Die Sprache der älteren Swahili-Dichtung*. Berlin: Dietrich Reimer.
- Nurse, Derek. 1985. "Dentality, areal features, and phonological change in Northeastern Bantu." *Studies in African Linguistics* 16,3: 243-279.
- Nurse, Derek and Thomas J. Hinnebusch. 1993. *Swahili and Sabaki. A linguistic history*. Berkely, Los Angeles, London: University of California Press.
- Nurse, Derek and Tom Spear. 1985. *The Swahili. reconstructing the history and language of an African society, 800-1500*. Philadelphia: University of Pennsylvania Press.
- Nurse, Derek and Martin Walsh. 1992. "Chifundi and Vumba: partial shift, no death." *Language death, factual and theoretical explorations with special reference to East Africa*, ed. by Matthias Brenzinger. Berlin: Mouton de Gruyter.
- Rzewuski, E. 1991. "Mother-tongue/father-tongue convergence: on swahilization and deswahilization in Mozambique." *Akten des 7. Essener Kolloquiums über "Minoritätensprachen / Sprachminoritäten"*, ed by J. R. Dow and T. Stolz, pp 267-305. Bochum: Brockmeyer.
- Schoff, Wilfred H. 1912 *The Periplus of the Erythraean Sea*. New York: Longmans, Green. (2nd edition 1974, Delhi: Oriental Books Reprint Corporation)
- Swadesh, Morris 1952. "Lexicostatistic dating of prehistoric ethnic contacts." *Proceedings of the American Philosophical Society* 96:452-463.
- Wald, Benji. n.d. "Substratal and superstratal influences on the evolution of Swahili syntax: central East Coast Bantu and Arabic." *SUGIA* [forthcoming].

Appendix 1

Text from the *Periplus* (from Schoff 1912)

"15. Beyond Opone, the shore trending more toward the south, first there are the small and great bluffs of Azania; this coast is destitute of harbors, but there are places where ships can lie at anchor, the shore being abrupt; and this course is of six days, the direction being south-west. Then come the small and great beach for another six days' course and after that in order, the Courses of Azania, the first being called Sarapion and the next Nikon; and after that several rivers and other anchorages, one after the other, separately a rest and a run for each day, seven in all, until the Pyrala³/₄ islands and what is called the channel; beyond which, a little to the south of south-west, after two courses of a day and night along the Ausanitic coast, is the island Menuthias, about three hundred stadia from the mainland, low and wooded, in which there are rivers and many kinds of birds and the mountain tortoise. There are no wild beasts except the crocodiles; but there they do not attack men. In this place there are sewed boats, and canoes hollowed from single logs, which they use for fishing and catching tortoise. In this island they also catch them in a peculiar way, in wicker baskets, which they fasten across the channel-opening between the breakers

16. Two days' sail beyond, there lies the very last market-town of the continent of Azania, which is called Rhapta; which has its name from the sewed boats (*rhaptôn ploiarion*) already mentioned; in which there is ivory in great quantity, and tortoiseshell. Along this coast live men of piratical habits, very great in stature, and under separate chiefs for each place. The Mapharitic chief governs it under some ancient right that subjects it to the sovereignty of the state that is become first in Arabia. And the people of Muza now hold it under his authority, and send thither many large ships; using Arab captains and agents, who are familiar with the natives and intermarry with them, and who know the whole coast and understand the language.

17. There are imported into these markets the lances made at Muza especially for this trade, and hatchets and daggers and awls, and various kinds of glass; and at some places a little wine, and wheat, not for trade, but to serve for getting the good-will of the savages. There are exported from these places a great quantity of ivory, but inferior to that of Adulis, and rhinoceros-horn and tortoise-shell (which is in best demand after that from India), and a little palm-oil

18. And these markets of Azania are the very last of the continent that stretches down on the right hand from Berenice; for beyond these places the unexplored ocean curves around toward the west, and running along by the regions to the south of Aethiopia and Libya and Africa, it mingles with the western sea."

According to Schoff (92ff): Courses of Azania are "strips of desert coast extending below the equator." Sarapion "may be the modern Mogdishu"; Nikon "is perhaps the modern

Barawa: Rivers and anchorages refers to "the modern *El Djesair* or 'coast of islands'" (=the Bajuni Islands (TJH)). Pyralæ Islands "are evidently Patta, Manda, and Lamu". Ausanitic Coast: "Ausan was a district of Kataban in South Arabia". Menuthias: "The first island south of Manda is Pemba.... But the topographic description is perhaps truer to Zanzibar, and the name seems perpetuated in the modern Monfiyeh (= Mafia (TJH))." (Schoff thinks this part of the text is corrupt with whole passages missing). Raphta: the identification depends on the identity of Menuthias, thus it might be Pangani, Bagamoyo, or Kilwa. Muza: a town in modern Yemen on the Red Sea coast; Adulis is the modern town of Massaua, Eritrea. Berenice: Southern Egyptian town, now apparently in ruins, just north of the Sudanese border on the Red Sea.

McCrindle's translation (1879) of the last several lines of 16 are: "The whole territory is governed by the despot of Mopharitis, because the sovereignty over it, by some right of old standing, is vested in the kingdom of what is called the First Arabia. The merchants of Mouza farm its revenues from the king, and employ in trading with it a great many ships of heavy burden, on board of which they have Arabian commanders and factors who are intimately acquainted with the natives and have contracted marriage with them, and know their language and the navigation of the coast." (1879:72-73)

Appendix 2

The Sabaki Languages (and Key to Abbreviations of Language Names)

1. Elwana (El; Kenya: Tana River north of Hola)
2. Pokomo (Po; Kenya: Tana River)
 - Upper Pokomo (UP): Milalulu, Zubaki, Ndura, Kinakomba, Gwano, Ndera
 - Lower Pokomo (LP): Mwina, Kulesa, Ngatana, Dzunza, Buu, Kalindi
3. Mijikenda (MK; southern Kenya and northern Tanzania coastal hinterland)

Chonyi (Ch)	Digo (Di)	Giryama (Gi)	Kauma (Ka)
Duruma (Du)	Jibana (Ji)	Kambe (Ka)	Rabai (Ra)
Ribe (Ri)	Seguju ¹⁷		
4. Comorian (Com; Comoro Islands)

Ngazija (Ng; Grande Comore Island)	Maore (Ma; Mayotte Is.)
Nzuani (Nz; Anjouan Is.)	Mwali (Mh; Moheli Island)
5. Swahili (Sw)
 - Mwiini (Mw; Brava, and Brava area of Somalia)
 - Northern Dialects (ND)

Amu (Am; Lamu archipelago)	Tikuu (Ti; so. Somalia, Lamu archipelago)
Pate (Pa; Lamu archipelago)	Siu (Si; Lamu archipelago)
Malindi (Ma; Malindi, Kenya)	Mvita (Mv; Mombasa and environs)
Jomvu (Jo; Mombasa environs)	Ngare (Nga; Mombasa environs)
Chifundi (Chi; Funzi Island and opposite mainland, eastern Wasini Island)	
 - Southern Dialects (SD) ¹⁸
 - Pemba (Pe; Pemba Island)
 - Unguja (Ung; Zanzibar Town)
 - Makunduchi (Mak; so. and e. Zanzibar Is.)
 - Tumbatu (Tu; Tumbatu Is, no. Zanzibar Is. and so. Pemba)
 - Vumba (Vu; w. Wasini Island, Shimoni peninsula, Vanga area)
 - Mtang'ata (Mt; coastal area no. and so. of Tanga, Tanzania)
 - Mafia (Mafia Is.)
 - Kilwa (=Unguja/St. Sw.; Kilwa Is.)
6. Mwani (Mn; no. Mozambique coast, Kerimba Islands)

¹⁷In addition to the nine traditionally recognized dialects Mijikenda includes Segeju, originally a Central Kenya language, now linguistically, but not necessarily socially, assimilated by the Digo.

¹⁸ There is a SD variant, very similar to Unguja, spoken on Madagascar in the village of Marodoka on the small island of Nossé-Be. South of the Mwani area along the Mozambique coast, reports and other evidence indicate the historical presence of Swahili-speaking communities (Rzewuski 1991)

Appendix 3

Noun Classe System: Sabaki and NEC (N&H 1993:338)

	PNEC	LP	MK	ND	SD	Com	Luguru
1.	*u+mu-	o+mu-	m-	M-	m-	u+mu-	i+mu-
2.	*a+βa-	a+wa-	a-	Ua-	wa-	u+wa-	i+wa-
3.	*(g)u+mu-	o+mu-	m-	M-	m-	u+mu-	yu+mu-
4.	*I+mi-	e+mi-	mi-	mi-	mi-	i+mi-	i+mi-
5.	*li+i-	edi+Ø-	Ø-	i-	Ø-	li+Ø-	li+Ø-
5a.	*i+jj-	edi+dz-	dz-	i-/j-	j-	li+dz-	li+ts-
6.	*(g)a+ma-	a+ma-	ma-	ma-	ma-	ya+ma-	ya+ma-
7.	*I+ki-	e+ci/ki-	ki-	ki-	ki-	i+jj-	i+ki-
8.	*i+vi-	e+vi-	vi-	zi-	vi-	i+zi-	i+pfi-
9.	*i+N-	e+N-	N-	N-	N-	i+N-	i+N-
10.	*zi+N-	e+N-	N-	N-	N-	zi+N-	tsi+N-
11.	*u+lu-	o+yu-	lu-	u-	u-	u+u	u+lu-
12.	*a+ka-	a+ka-	ka-	ka-	-	-	i+ka-
14.	*u+βu-	u-	u-	= 11	= 11	= 11	u+u-
15.	*u+ku-	ku-	ku-	ku-	ku-	?+u-	u+ku-
16.	*a+pa-	(Φa-)	(ha-)	(pa-)	(pa-)	(?+βa-)	i+ha
17.	*u+ku-	(ku-)	(ku-)	(ku-)	(ku-)	(?+h-)	u+ku-
18.	*u+mu-	(mu-)	(m-)	(M-)	(m-)	(?+mu-)	u+mu-

Appendix 4

A Selection of Swahili/Sabaki Isoglosses



