Notes on the Morphosyntactic Bias of Verbal Constituents in Sheng Texts

Daisuke SHINAGAWA

0. Introduction

Sheng is an urban mixed code emerged supposedly in 50's to 70's in downtown area of Nairobi as a peer language of the youth and currently spreading to major local centers in Kenya and even to some urban capitals in other East African countries. In terms of its linguistic structure, Sheng is said to be "based primarily on Swahili structure, with the lexicon drawn from Swahili, English and the various mother-tongue languages (Abdulaziz and Osinde, 1997: 43)". That is, Sheng can be called as a mixed language in a wider sense, but specifically it may well be referred to as a "trilingual mixed language", following Thomason (2003)'s terminology "bilingual mixed language[1]", since most of the speakers have a certain, if not sufficient, command of both Swahili and English besides their mother tongue, i.e. any one of various vernacular languages. Though it is true that the basic morphosyntactic frame of Sheng is largely equivalent to that of (structurally simplified) Swahili, we can also point out some uniqueness in Sheng's grammatical structure that differs from Swahili itself, as suggested by Ogechi (2005) etc. Along this line, Shinagawa (2006) investigates its formal particularities, claiming that what brings the morphosyntactic specificity of Sheng can be regarded, at least partly, as abstract features shared, beyond their surface diversity, by local vernacular (mostly Bantu) languages rather than as a mere influence from a particular single language. As noted by Bosire (2006: 192), a number of the preceding studies on Sheng have devoted to its sociological or sociolinguistic aspects and consequently exploration to describe such structural hybridity has not been done

1 "The fundamental division [between pidgins and creoles, and bilingual mixed languages] has to do with the process by which mixed languages emerge: imperfect learning plays a significant role in the genesis of pidgins and creoles, but not in the genesis of bilingual mixed languages. In bilingual mixed languages, particular structural and lexical subsystems are adopted intact from each source language, with a small amount of asymmetrical distortion or adaptation to the structure of the other language" (ibid.: 22).

robustly. This article, thus, aims to shed light on its linguistic structure and take a step forward to investigate the formal specifics of this mixed code, focusing on the morphosyntactic frame of the verbal constituents that agglutinatingly composed by various kinds of morphemes from different source languages.

1. Background

1.1 Morphological structure of the Swahili verb

As noted above, the fundamental structure of the Sheng verb basically follows that of Swahili, which is illustrated in (1).

(1) Basic morphological structure of the Swahili-type verb
VERB {INFL-P [SM-TAM-(RM-)(OM-)] ≠ STEM [Base(-DSuf)] ≠ INFL-S [-FV]}

Among these morphemes, subject marker (SM), verbal base (Base), and final vowel (FV) are obligatory in the finite verb. SM, object marker (OM) and relative marker (i.e., relativizer, RM) show the noun class concord with the subject, object and antecedent NPs respectively. Together with tense and aspect marker (TAM), they form a morpheme cluster INFL-P (inflectional prefixes). Base can optionally take one or more derivational suffixes (DSuf), whose basic function is to add or reduce the number of nominal argument or to change its thematic role, i.e., voice marking. These two morphemes construct a STEM, i.e., lexical core of the verb. FV assigns the syntactic profile as a verb to the constituent and indicates the mood distinction. Hence, FV itself is regarded as a separate morpheme cluster INFL-S (inflectional suffix). There are two morphemes classified as FV, namely indicative (INDC) *-a* and subjunctive (SUBJ) *-e*. Simple examples are as follows; INDC: ***nitakuandikia*** *barua* {ni-ta-ku ≠ andik-i ≠ a} (= INFL-P [1sgS-FUT-2sgO-] ≠ STEM [write-APPL] ≠ INFL-S [-INDC]) "I will write a letter to you"; SUBJ: ***nikuandikie*** *barua*? {ni-ku ≠ andik-i ≠ e} (= INFL-P [1sgS-2sgO-] ≠ STEM [write-APPL] ≠ INFL-S [-SUBJ]) "Shall I write a letter to you?"; Relative form: *(Yeye)* ***aliyeandika*** *barua* {a-li-ye ≠ andik ≠ a} (= INFL-P [3sgS-PST-1sgR-] ≠ STEM [write] ≠ INFL-S [-INDC]) "(S/he) who wrote a letter" etc.

Another thing to be noted here is about an obvious surface feature of the mixed verbal complex. As illustrated in (2), the structure including a non-Swahili Base lacks INFL-S (FV *-a/ -e*). This feature itself is not unusual in Swahili since the same is applied to the verb stems of Arabic origin; e.g. *a-me ≠ rudi nyumaba-ni* {1S-PERF-return(v.i.) home-LOC} "S/he (has) returned home", where the Base *rudi* is used without suffixation of FV, while *a-me ≠ rudi-sh ≠ a ki-tabu* {1S-PERF-return(v.i.)-CAUS-FV CPx7-book} "S/he has returned a book", where a FV appears because of the presence of a DSuf *-(i)sh*. This can be generally explained that FV is said to be functionally redundant because INFL-P implies its syntactic property of the constituent

as a verb[2].

(2) Alivuta sigara kwa makini kama **anathink** deeply. [W]
"He smoked a cigarette consciously as if he **was thinking** deeply."
{kama a-na ≠ *think* *deeply*}
as 3sgS-CONT ≠ think deeply

(3) **Watanithinkia** nini watu nikianza kusema opposite? [W]
"What will they **think of me**, if I start to say something opposite (to what I said)?"
{wa-ta-ni ≠ *think*-i ≠ a nini}
3plS-FUT-1sgO ≠ think-APPL ≠ FV what

In (2), mixed verb *a-na ≠ think* consists of INFL-P *a-na-* and an English Base *think* but lacks INFL-S, while in (3) *wa-ta-ni ≠ think-i ≠ a* takes a FV because of the suffixation of applicative *-i*. This means that if STEM takes any DSuf, then a FV must be affixized for not violating the well-formedness of the constituent, i.e. *{Base-DSuf#} cannot be allowed. The structure illustrated in (1) can thus be rewritten as follows.

(1') Basic morphological structure of the verb
VERB {INFL-P ≠ STEM [Base(-DSuf)*] ≠ (INFL-S)*}
* In the case of having non-Swahili Base, if there is no element in DSuf slot, then FV should be deleted.

1.2 Theoretical background: 4-M model and the Uniform Structure Principle (USP)

A so-called mixed language is generally regarded as a code whose grammatical frame and its lexicon are derived from different source languages (Thomason, 2003: 21). Myers-Scotton (1997) terms a language that provides the grammatical frame as Matrix Language (ML) and that supplies the lexical items as Embedded Language (EL) under her theoretical model called Matrix Language Frame (MLF) model. Thus in the composition of Sheng, it is obvious that Swahili, English and various local vernaculars are involved as ELs. However, its ML cannot be determined with such clarity, since the morphosyntactic structures of Sheng reflect composite nature made up by Kenyan Swahili with some influence from local vernacular languages as illustrated in (4).

(4) **unasemekengo** fitu singine tu chamani!
"You're just talking about other things, man!"

2 In addition, the mood distinction is also marked, implicitly though, by the arrangement of TAM elements in Swahili, i.e., INDC basically needs to take a morpheme in TAM slot, while SUBJ has no marker in the slot.

{u-na-sem-ek-engo fi-tu si-ingine tu chamani}
2sg.S-PRES-say-NEUT-EMPH CPx8-being CPx10-other just INTERJ
Sw) **Unasema** vitu vingine tu jamani!

A verbal suffix *-engo* that indicates a kind of "emphatic" modality is not traced back to any Swahili varieties, though all the resting grammatical morphemes (Grams) are provided from Swahili (with regular phonetic distortion and simplified concord). This morpheme can be further divided into *-eng* and *-o*, the former of which is seemingly from (a dialect of) Luhya, a Bantu language spoken in western part of Kenya (and is traced back to Proto Bantu **-aga*), and the ending *-o* seems to be introduced from Luo, a major Nilotic language (for detailed description see Shinagawa, 2006: 132-133).

In order to analyze such heterogeneity found in the grammatical subsystem of Sheng, needless to say, detailed description of its grammatical elements must be necessary and especially a precise classification method for various Grams is needed. The 4-M model proposed by Myers-Scotton and Jake (2000) provides a descriptive framework that seems quite suitable for such an aim, though the primal intention of this model is to capture the relation between linguistic forms and the process of speech production[3]. This model, whose adequacy is grounded by various evidence from aphasia, second language acquisition etc., classifies morphemes into four groups, namely 1) Content morpheme, 2) Early System morpheme, 3) Late Bridge System morpheme and 4) Late Outsider morpheme, the latter two of which are subcategories of the integrating upper category called Late System morpheme.

Table 1: 4-M model and its application to the verbal Grams of Sheng[4]

(cf. Myers-Scotton and Jake, 2000: 1062)

	[+conceptually activated]		[-conceptually activated]	
	[+thematic role A/R]	[-thematic role A/R]	[-outside MP]	[+outside MP]
	1) Content M.	2) Early Sys-M.	3) Late Bridge Sys-M.	4) Late Outsider Sys-M.
	Base	(DSuf)	TAM, FV, PreF	SM, OM, RM, DSuf
A	content	system		
B	lemma (mental lexicon) level		formulator level	
C	(head)	within the maximal projection (MP)		outside the MP

Myers-Scotton (2003: 91) defines mixed languages ("split languages" in her terminology) as follows; "all split languages show a composite structure *that goes beyond*

3 "it is primarily a model of how morphemes are accessed [on speech production] (Myers-Scotton and Jake, 2000: 1069)".

4 Abbreviations are as follows; [(thematic role) A/R] = assign or receive a thematic role, (in the vertical column) A = thematic role distinction, B = access level distinction, C = syntactic relational distinction; MP = maximal projection.

a composite at the level of lexical-conceptual structure. ... This means changes in content morphemes and other conceptually-based elements are ruled out as sufficient evidence of a split language". In order to measure the syntactic hybridity, thus it is important to describe the elements classified as Late System morphemes, and for Sheng to be called a true mixed language, there must be some influence on these morphemes from languages other than the basic ML, Swahili.

Regarding the dynamic process of creating a mixed language, we should refer to a principle called Uniform Structure Principle (USP). The definition of the principle is as follows (quoted from Myers-Scotton, 2003: 100): "A given constituent type in any language has a uniform abstract structure and the requirement of well-formedness for this constituent type must be observed whenever the constituent appears (Myers-Scotton, 2002: 121)", i.e., this can be paraphrased as a kind of structure/pattern-preserving constraint applied to a certain constituent or phrase. This notion, by its definition, implies contradicting directions of change, i.e., it predicts both resistance against influx of the outer (i.e. non-ML) elements and conseration of ML's grammatical system on one hand, and overall systematic transfiguration from the system of ML to that of another participating language on the other, if ML once permits systematic invasion from non-ML. Thus the USP can be seen as a key notion for the emergence of a mixed language. In the following discussion, we shall focus on it as a pattern-preserving constraint and how it works in the verbal constituent.

1.3 Sheng and Kenyan Pidgin Swahili

At the end of this section, we should notice on the difference between Sheng and Kenyan Pidgin Swahili (PiS). As mentioned in Shinagawa (2006), PiS is practically a synonym for Kenyan Upcountry Swahili, which is a cover term for local varieties of Swahili created by mixing of Coastal Swahili and local vernaculars for inter-ethnic communication. Swahili is one of the varieties, which in turn provides Sheng with its grammatical basis. That is, the PiSs are basically grounded on the location (different from region to region) and emerged supposedly in the dawn of the 20th century when Swahili started to spread to the "upcountry". Thus it is important to note here that a crucial point investigated in the linguistic study of Sheng is the grammatical difference between Sheng and PiS, not Standard Swahili. In other word, for Sheng to be regarded as a stable and independent mixed language, the systematic difference between them must be existed.

2. Texts investigated

The texts investigated in this study are from two different sources. One is a macaronic (i.e. composed by mixing of several linguistic codes, mainly Swahili and English in

this case) short story titled *Without Kiinua Mgongo*[5] (lit. "Without gratuity", abbreviated as [W]) by David Maillu, first published in 1989 in Nairobi. From this text (75 pages in total), the data including 401 samples of mixed verbal constituents, namely predicate (indicative) verbs, subjunctive verbs, relative verb forms etc, are obtained. Note that those constructed by a single source language in terms of both EL and ML are not included in the data. While this is one of a few well-known published texts which said to be written in Sheng, however, we should notice that the work may contain deliberate stylistic expressions and intentionally unnatural phrasing for literary purposes, hence the expressions in the text could be more or less distant from current natural conversation. The data are thus made use of especially in the statistic analyses.

The other source is a literary magazine *Kwani?*[6] which assumes (young) urbanites as potential readers and started to be issued in 2003 as a "Sheng speaking" magazine. The text investigated is from the article titled "Sheng Interviews" (abbreviated as [S]) in *Kwani?-03* published in 2005, in which two young hip-hop artists, both in their early twenties of age, interview with an old man (almost 60 years of age) and talk about the colonial and post-colonial history of Nairobi including the genesis of Sheng. This being recently published, i.e., reflected current characteristics and apparently less affected in terms of stylistic matters (unlike the case of [W]), the examples referred to in the following discussion are mainly quoted from this text.

3. Morphosyntactic frame of mixed verbal constituents

This section presents a list of morphosyntactic features found in descriptive and statistic data extracted from [W] and [S]. Following some notes for the general grammatical tendencies in 3.1, descriptions for the verb consisting morpheme clusters namely INFL-P (3.2), STEM (3.3) and INFL-S (3.4) are provided. In addition, morphosyntactic features of the relative constructions, in which the formal differences among Standard Swahili, PiS and Sheng are somewhat salient, are dealt with in 3.5.

3.1 General tendencies: Simplified concord and Isolation/ Analyzation

There are two processes to be noted as basic tendencies in overall structure of the Sheng verb, namely simplified concord and morphological isolation. Simplification in grammatical agreement is said to be one of the remarkable syntactic features of the mixed languages in general. In the case of PiS, Heine (1979) lists some basic formal characteristics including morphosyntactic and phonological simplification.

5 Appreciation goes to Dr. Michael Gromov (Dept. of Literature, University of Nairobi) who kindly allowed the author to access his collection of books including [W] and gave significant suggestions for the study of Sheng.

6 The author's thanks are also heading to the editing team of *Kwani?* for their permission to cite parts of the articles and support for data collection.

Simplification process is also clearly found in Sheng (Shinagawa, 2006) and simplified agreement seems to be a norm in the noun-class concord (3.2.1, Table 2).

Isolation process is the other conspicuous feature of especially the mixed verbal constituents. As Heine (ibid: 91) points out, the value of synthesis index, i.e. morphemes per word ratio, of verbal constituents is considerably low in PiS[7] and this tendency gives rise to the morphologically isolating nature of words (see also Myers-Scotton, 1979 for descriptive data). In the written texts of Sheng, it is sporadically observed that some of verbal and nominal complexes are divided into morpheme clusters by spacing (or hyphenation) as if they were independent words. It can be said that this orthographic characteristic apparently reflects writers' meta-linguistic understanding on the boundary of linguistic forms.

(5) **Ma pick-pocket** walikuwa [S]
"Pick-pockets were there."
{ma pick-pocket wa-li-ku-wa}
CPx.6 pick-pocket 3plS-PST-INF-be
Sw) Wanyakuzi walikweko.

(6) wanashinda pamoja, na kizungu **wana jua jua**. [S]
"They live together, and they know English well"
{wa-na-shind-a pamoja, na ki-zungu wa-na jua jua}
2S-PRES-INF-live-INDC together, and CPx7-European 2S-PRES know
Sw) Wanashinda pamoja, na kizungu **wanakijua** sana.

It can also be said that this kind of isolation process contributes to making STEM "stand-alone" and to deviation from the rigid syntactic frame of the ML. It should be also mentioned that this analytic nature allows the mixing of elements at the word level and permits the morpheme cluster, not the whole word or phrase, to be a target domain of USP.

3.2 INFL-P

Among the morphemes in the INFL-P cluster, SM, OM and RM, i.e. the concordance markers, are categorized into Late Outsider Morpheme since they indicate the relationship with argument NPs, whereas TAM could be classified as Late Bridge Morpheme because they have no syntactic reference to any element outside its maximal projection.

3.2.1 Simplified concord

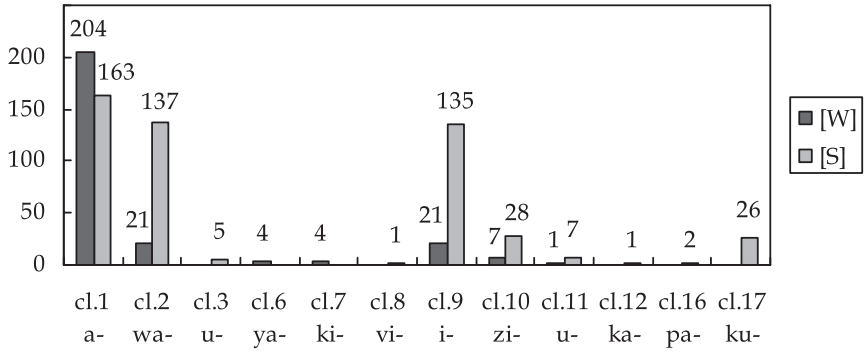
As numerous studies have pointed out, simplification of the grammatical agreement is

7 Heine (1979: 91) regards the morpheme/word ratio in PiS as less than 1.50. This value is said to be quite low when we think of the basic morphology cited in (1') (cf. the noun has also {Prefix-stem} structure).

remarkably seen in Sheng as well as PiS (Heine, 1979 in PiS, Ferrari, 2004 in Sheng etc). Few examples would suffice in this point: a) *kumbe mwenyewe* ***a-ko*** *nyuma* (Sw: *kumbe yeye mwenyewe* ***yu-ko*** *nyuma*), "Hey, he himself is behind", b) *Magari* ***i-li-ku-wa*** *ya wazungu* (Sw: *Magari* ***ya-li-ku-wa*** *ya wazungu*), "Cars were of the Europeans (There were no cars for the Africans)", c) *shauri Kiswahili si watu wengi* ***wa-li-i-jua*** (Sw: *kwa sababu Kiswahili si lugha ambayo watu wengi* ***wa-li-ki-jua***), "Because Swahili is not the language that many people know". Example a) reflects the fact that the formal distinction of cl. 1 SM, namely between *a-* (used with general verbs) and *yu-* (with locative predicates), has been almost entirely lost, while b) shows a typical deviation from the norm of Standard Swahili concord, i.e., the class 6 noun *Magari* is grammatically agreed with the class 9 SM *i-*. Example c) shows that OM is also subject to simplified concord (NP = cl. 7 *Kiswahili*, but SM = cl. 9, see also Table 3).

The systematic erosion of the noun class agreement system in Sheng[8], i.e., simplification process from the full class concord in Swahili type to the concord based on the two-way distinction, where [+human, sg/pl] NPs are agreed to cl. 1/2 and [-human, sg/pl] NPs to cl. 9/10, is clearly confirmed in Table 2[9]. In addition to those two gender pairs, cl. 17 SM *ku-* is also used especially in [S], but it appears limitedly with possessive stem *-na*, i.e., *ku-na* (NP) "there is/are (NP)", *ku-li-ku-wa-na* (NP) {17S-PST-INF-be-POSS}, "there was/were (NP)" etc.

Table 2: Number of class agreement SMs appeared in the texts



One of the characteristics of the nominal morphology mentioned in the previous studies is the relatively frequent use of cl. 12 prefix *ka-* that does not exist in the norm of Standard Swahili (Myers-Scotton, 1979: 122-123, Shinagawa, 2006: 127). On the contrary, the SM of cl. 12 is hardly appeared in the texts as Table 2 shows. However, that does not mean that the cl. 12 NPs themselves are not used in Sheng but that the noun class distinction is only limitedly reflected in INFL-P of the verb as explained above. The same is applied to the case of cl. 6 NPs (*ma*-NP), i.e., while NPs morphologically marked as cl. 6 are numerous, the agreement marking on the verb is

8 For detailed description, see Shinagawa (2006).

9 Cl. 4 and cl. 5 SMs are omitted in the table simply because they are not appeared at all.

almost entirely done by the cover class of [-human], i.e., cl. 9/10.

3.2.2 Use of OM

Suppose that Sheng follows the tendency toward simplification referred to in 3.1, it is naturally assumed that the verb tends to reject object marking since OM is not syntactically obligatory in the structure. However, it can be said from the data that at least OM itself is not entirely excluded from the structure as illustrated below.

(7) Na zote **umezipiga** ndonyo. [S]
"and you sold everything (all the cows you inherited)."
{u-me-zi-pig-a ndonyo}
2sgS-PERF-10O-beat-INDC market
Sw) Na wote **umewauza** (sokoni).
{u-me-wa-uz-a}
2sgS-PERF-2O-sell-INDC

(8) '... **utaya-understand** yote eventually.' [W]
"... you will eventually understand everything."
{u-ta-ya-*understand*}
2sgS-FUT-6O-*understand*
Sw) **utayaelewa** (mambo) yote mwishowe.
{u-ta-ya-elew-a}
2sgS-FUT-6O-understand-INDC

The data presented in Table 3[10] shows that in [W] the frequency of OM (68.7%) is perhaps rather higher than it is assumed in terms of the simplification bias, while the value drawn from [S] (25.8%) is relatively low, though not scarce at all. Needless to say, such numerical discrepancy results from the difference of text genre[11] on one hand, but on the other hand, it can also be said that the declination of the percentage parallels the "synthetic to analytic" change, i.e., the general drift of isolation found in the morphosyntactic structure of Sheng.

According to the data, it is also suggested that the reflexive markers (REF) are stably used, though not quite many in number, comparing to the other normal class OMs. This may be partly because REF is not a pure Late Outsider since its deictic referent is automatically determined (i.e., the subject) and therefore can be regarded rather as a STEM consisting element that assigns the relational notion ("the action is taken to self/ by oneself" etc.) to the Base.

10 OM frequency is the ratio of the number of OMs to that of the verbs appeared in the whole text. The number of verbs, a denominator, excludes intransitive verbs, copulative verbs and transitive verbs that take a clausal complement, i.e., the verbs syntactically unable to take an OM.

11 That is, there is a tendency that sentences with formal or literal style prefer to take OM, while it is omitted in the oral expression.

Table 3: Number of OMs appeared in the texts

		[W] (V=252)		[S] (V=233)
OM total		173 (68.7%)	57 (w/o cl. 1/2)	60 (25.8%)
REF		13 (7.5)	(22.8)	13 (21.7)
Person	1 sg/ pl	10 (5.8)	(17.5)	12 (20.0)
	2 sg/ pl	12 (6.9)	(21.0)	10 (16.7)
	3 sg/ pl (=cl. 1/2)	116 (67.0)	–	14 (23.3)
Class	cl. 9/ 10	19 (11.0)	(33.3)	11 (18.3)
	others	3 (1.7)	(5.2)	0 (–)

(9) Unacheki vile sasa **tumejibanga**. [S]
"You see, we smoke bhang like that."
{tu-me-ji-bang-a}
1plS-PERF-REF-bhang/marijuana-INDC
Sw) Unaona, kama vile tunavuta bangi.

The verbal stem of *tumejibanga* is from Swahili *bangi* "bhang, marijuana", i.e., the form *bang-a* is a denominative verb stem coined in Sheng. In this case, REF can be seen as a kind of (part of) syntactic intransitivizer (See also example (12)). More on the relation between frequency/stability of Grams and their affiliations in terms of the morpheme cluster shall be referred to in section 4.

3.3 STEM

In Swahili-type structure formulized in (1), STEM consisting morphemes are Base, a Content morpheme, and DSuf, which can be classified as either Late Outsider or Early System Morpheme. It is regarded as Late Outsider because its basic function is voice marking, i.e., the grammatical expression indicating the case relation with syntactically relating nominal argument(s). It can also be seen as Early System Morpheme because some DSufs are used to modify the lexical meaning of Base rather than to control the case relation of the argument. This classificatory bifurcation can be explained as being brought about by lexicalization process from the former to the latter. As illustrated in (4), DSuf is highly excluded from STEM in the case of the mixed STEM.

3.3.1 Restriction of taking DSuf in mixed VP

There is a somewhat clear tendency that DSufs are generally excluded from the mixed verbal constituents.

The data from [W] shows there are only 9 examples out of 401 samples (2.2%) that take any DSuf in the structure. Regarding [S], 156 constituents (31.2%) affixize DSuf in total, but the number is restricted to 5 (15.2%) in the case of STEMs that

consist of non-Swahili Base. Here again is a significant difference between the percentage in [W] and that in [S].

Table 4: Number of DSufs appeared in the texts

	[W] (mixed)	[S] total	[S] mixed
DSuf total	9/ 401 (2.2%)	156/ 501 (31.2%)	5/ 33 (15.2%)
Passive *-w* etc.	3	91 (54.5%)	3
Applicative *-i* etc.	6	28 (16.8%)	–
Causative *-iz, -ish* etc.	–	19 (11.4%)	2
Stative *-ik* etc.	–	6 (3.6%)	–
Reciprocal *-an* etc.	–	23 (13.8%)	–

One possible factor that gives rise to the difference of the acceptability of DSuf between the texts may relate to phonological difference between pure borrowings and lexicalized forms, the latter of which has clear formal adjustment to the phonological system of Sheng, while the former receives little or no such modification.

(10) angeanza namna gani **kumconfessia** Nzuki [W]
"How would she start to confess (the fact) to Nzuki?"
{ku-m-*confess*-i-a}
INF-3sgO-*confess*-APPL-INDC
Sw) Namna gani angeanza **kumkiria** Nzuki
{ku-m-kiri-i-a}
INF-3sgO-confess-APPL-INDC

(11) Kimathi alikuwa **amemadwa**. [S]
"Kimathi had been killed/ murdered."
{a-li-ku-wa a-me-mad-w-a}
3sgS-PST-INF-be 3sgS-PERF-murder-PASS-IND
Sw) Kimathi alikuwa **ameuawa**.
{a-me-ua-w-a}
3sgS-PERF-kill-PASS-IND

In (11), the base *mad-a* is coined from English *murder* with clear phonological and morphological (assignment of FV) adjustment and almost all the mixed constituents with non-Swahili bases found in [S] are of this type, whereas the base in (10) *confess* seems to be adopted comparatively directly from an EL (i.e. pure borrowing) and generally the mixed verbs in [W] are "non-lexicalized". It follows that the tight restriction of the DSuf acceptability in [W] may well be grounded on this fact. This means also that the lexicalized verb bases are relatively capable of taking DSuf as in (11) and (12) below, in which causative DSuf *-iz* verbalizes the nominal stem *bangi*.

(12) **tuna bangaiza** hivi kiasi [S]
"We (let ourselves) smoke bhang a bit like this."
{tu-na banga-iz-a hi-vi kiasi}
1plS-PRES bhang/marijuana-CAUS-INDC DEM(near)-8 a little
Sw) **tuna(zi)vuta bangi** hivi kiasi
{tu-na-zi-vut-a bangi}
1plS-PRES-(9O-)smoke(lit. pull)-INDC bhang/marijuana

3.3.2 Isolation ("self-standing" STEM)
As illustrated in (6) in 3.1 and found also in (12) above, Sheng verbs are sometimes spelled with spacing between INFL-P and STEM, supposedly reflecting (speakers' understanding of) the basic morphosyntactic frame.

(13) Enyewe mtu mjanja sana **anaeza chanuka** [S]
"One who is a very tricky person can outwit (others)."
{a-na-ez-a chanuk-a}
3sgS-PRES-be able-FV be agile/outwit-FV
Sw) Aliye mjanja sana **anaweza kushinda** kwa akili
{a-na-wez-a ku-shind-a}
3sgS-PRES-be able-FV INF-win-FV

anaeza "S/he is able, S/he can" is an equivalent of Swahili *anaweza*, which should take an infinitive form, *ku*-STEM-*a*, as its verbal complement in the grammatical norm. In (13) however, it takes a bare STEM *chanuk-a*, which lacks the infinitive marker *ku*-. That is, the STEM is getting to be used as a self-standing item under the influence of the morphologically isolating tendency. This is also confirmed in the constituents with a RM such as (14).

(14) **walio kaa** Maringo walikuwa wanajiweza. [S]
"People who lived in Maringo was able to make their own living."
{wa-li-o kaa Maringo}
3plS-PST-3plR stay M
Sw) **waliokaa** Maringo walikuwa wanajitegemea/ wanaweza kujitegemea.
{wa-li-o-kaa}
3plS-PST-3plR-stay

3.4 INFL-S

In the Swahili-type verbal structure, FV is the only morpheme classified in INFL-S. However in vernacular Bantu languages, there is another morpheme categorized into this cluster, i.e., Prefinal (PreF). PreF is reconstructed in Proto-Bantu by Meeussen (1967) as **-ag* and its corresponding forms in present Bantu languages denote grammatical functions such as habitual aspect, imperfective aspect, emphatic modality

etc. varying from language to language.

3.4.1 Non-Swahili element *-anga*

There are some Grams of Sheng which are not traced back to Standard Swahili. INFL-P *-anga* (<**ag*(PreF)-*a*(FV)) is typical of such morphemes in the verbal constituent though it appears also in PiS frequently (Shinagawa, 2006: 131-133). In the texts investigated, there are not so many samples with this element but it can be said that it is constantly used in various syntactic environments such as with a copulative verb (15), with a general verb (16) and in a relative construction (17).

(15) **alikuwanga** huko. [S]
"He really was there."
{a-li-ku-w-anga hu-ko}
3sgS-PST-INF-be-EMPH DEM(middle)-17
Sw) **(Kweli) Alikuwa** huko.

(16) Na mugithi, ilikwa siko hizo **zinachezanga** sana? [S]
"And (about) Mugithi, were they (their songs) played very much in those days?"
{zi-na-chez-anga}
10S-PRES-play-EMPH
Sw) Na (kuhusu nyimbo za) Mugithi, zilikuwa **zinacheza** sana siku hizo?

(17) Ni **vile unaskianga** piracy. [S]
"(It is exactly) like how you listen to pirated CDs."
{u-na-ski-anga}
2sgS-PRES-listen to-EMPH
Sw) Ndivyo **kama unavyosikia** muziki wa uharamia.

Regarding its co-occurrence with non-Swahili bases, as assumed from the discussion in 3.3.1) above, there are no example found in the mixed verbal constituent in both data.

3.5 Relative clause

Swahili has three distinctive forms of the relative construction; e.g. a sentence like "a person who is standing" can be differently constructed as follows; 1) analytic structure; *m-tu amba-ye a-na-simam-a*, where *amba-ye* {*amba*-RM} functions as a relativizer, 2) synthetic structure with TAM and RM; *m-tu a-na-ye-simam-a*, 3) synthetic structure with suffixized RM without specific tense marking; *m-tu a-simama-ye*.

3.5.1 Zero marking and DEM as a relativizer

As shown in Table 5 below, half of the relative clauses found in [S] are marked by exponents different from those in the Standard Swahili norm, i.e., relatives without

any Grams (zero marking) and marked by the demonstrative (DEM) *-le* "that, those". Emergence of the former is naturally assumed from the simplification process as a basic tendency in Sheng and the latter can be seen as a reflection of the analytic bias of its morphosyntactic frame. The use of DEM as a relativizer is well attested in various North-Eastern Bantu languages (Shinagawa, 2006: 134), thus it is also possible to say that the DEM-relative construction can be seen as a result of syntactic influence from mother-tongue languages of the Sheng speakers[12]. (18) is an example of zero marked relative and (19) and (20) are of DEM-relative.

(18) kuliletwa **mabasi inaitwa** mang'oro. [S]
"The buses called mang'oro were introducued (in Nairobi)."
Sw) Kuliletwa **mabasi yaliyoitwa** mang'oro.

(19) Si hao ni **wale wasee ulikuwa unaniambia**. [S]
"Ain't they the ones (old men) whom you were telling me about?"
Sw) Hao si (wale) **wazee ambao ulikuwa unaniambia**.

(20) sana sana **ile sisi tulikuwa tunaishi** ilikuwa mia moja na themanini, [S]
"Usually that (apartment) we lived in was (rent for) 180 shillings,"
Sw) Sana sana **nyumba ambayo sisi tulikuwa tunaishi,** (kodi yake ni) ilikuwa (shilingi) mia moja na themanini,

3.5.2 *-enye* instead of *amba-*

One clear characteristic found in the relative construction in the texts is apparent absence of the construction type 1) that is usually called "*amba-* relative". Myers-Scotton (1979) discusses that if relative constructions are used in PiS, then the structure may well be either zero marking or "*amba-* relative"[13]. This claim is quite plausible when we think about isolating tendency of the verbal structure mentioned in 3.1. The facts found in Sheng texts, however, are not necessarily so. Surprisingly, there is no single example that takes *amba-* in the relative constructions in both data.

Instead of *amba-*, there appears another relative marking Gram *-enye*, which originates from Swahili "possessive relative" indicating the meaning of "(NP) having sth". However, as illustrated in (21) and (22) below, this morpheme is used as a simple relativizer, i.e., *-enye* abstracts its original function and used as if it took over the role of *amba-* in PiS.

12 DEM-relatives have also been recognized broadly in mixed languages in general. At this point, we should notice that this construction itself might be seen not merely as a Sheng specific phenomenon but as realization of somewhat universal tendency of mixed languages.

13 "If the relative occurs at all in Kenyan up-country Swahili, it is more likely to appear in the *amba* construction. This preference is in keeping with the favoring of isolating-analytic forms in the up-country variety (Myers-Scotton, ibid.: 120)".

Table 5: Number of relative marking Grams in [S]

zero marked		16	(25.8)
marked by	RM	21	(33.9)
	DEM	15	(24.2)
	-enye	10	(15.1)
	amba-	0	(–)
total		62	(100 %)

(21) hao walikuwa **ma-askari wenye walikuja** wakiwa na wazungu. [S]
"They were guards who came (to Nairobi) with the Europeans."
{ma-askari wa-enye wa-li-ku-j-a}
CPx6-guard 2PPx-having 2S-PST-INF-come-FV
Sw) Hao walikuwa **askari ambao walikuja** na wazungu.
{ø-askari amba-o wa-li-ku-j-a}
CPx10-guard REL-2RM 2S-PST-INF-come-FV

(22) Kariuki Chotara **mwenye alikuwa** pande ya Nakuru [S]
"Kariuki Chotara who was in the vicinity of Nakuru"
{(K.C.) mu-enye a-li-ku-w-a}
(K.C.) PPx1-having 1(=3sg)S-PST-INF-be-FV
Sw) Kariuki Chotara **ambaye alikuwa** upande wa Nakuru
{(K.C.) amba-ye a-li-ku-w-a}
(K.C.) REL-1RM 1(=3sg)S-PST-INF-be-FV

4. Morphosyntactic bias in Sheng verbal constituents

As assumeed in 3.1, simplification and isolation processes as general tendencies in language mixing are largely confirmed in the data. The former is well attested in 3.2.1, 3.2.2 and zero marking in 3.3.2, while the latter is reflected in such phenomena as 3.3.1, 3.3.2, the use of demonstratives as a relative marker in 3.5.1 and 3.5.2. However, we can also point out other principles from the analyses discussed in section 3.

(23) Morphosyntactic bias in Sheng verbal constituents
a) Simplification process in grammatical agreement: 3.2.1, 3.2.2, 3.5.1
b) Isolation process: 3.3.1, 3.3.2, 3.5.1, 3.5.2
c) Suffixation to Prefixation (Structural leveling): 3.5.1, 3.5.2
d) "Habitats" effect avoiding mixture of elements
=USP applied at the morpheme cluster level: 3.4.1

4.1 Pattern-preserving bias

Additional principle c) is on the morphological preference of {Prefix-STEM} structure rather than {STEM-Suffix} one, where both affixes mark the grammatical agreement (AG). This structural tendency is basically applicable to other various words obligatorily marked by an agreement affix such as nouns, demonstratives, adjectives, associatives etc. More importantly, this bias is much more consistently applied to (North Eastern) Bantu in general than to Standard Swahili. For instance, what should be considered in the phenomenon described in 3.5.2 is why "*amba-*" relative which is readily assumed to be utilized in Pis (see footnote 13) has been practically out of use and why "*-enye*", originally a possessive relative marker, can take over its position. This seeming anomaly can be explained by the principle c), i.e., {AG-*enye*} overcomes {*amba*-AG} supported by the general structural preference. Similarly it can be said that this bias supports the stable use of DEM {AG-*le*} and RM which is included in INFL-P cluster (see Table 5 in 3.5.2). It is also clear that principle c) is relevant to the problem on relatively unfluctuating affixation of OM which is not structurally compulsory and thus inconsistent with the simplification tendency (see 3.2.2).

4.2 "Habitats" of Grams

What is linguistically crucial in describing a mixed language is, as mentioned in 1.2, to capture the compositeness found in the abstract grammatical structure, i.e., the level of Late System Morphemes (see Table 1), namely INFL-P and INFL-S. The composition of the verb consisting morphemes and their sources can be schematized as in (24) and (25). As referred to in 3.3.1), acceptability of DSuf varies depending on the degree of lexicalization of Base, i.e., whether it is a purely borrowed item or a lexicalized one.

(24) Morphological structure and acceptability of outer items: pure borrowing

Structure:	INFL-P [SM-TAM-(RM-)(OM-)]	≠ STEM [Base]	(≠ INFL-S [-DSuf-PreF-FV])
Source:	↑ Kenyan Swahili	↑ borrowed	↑ Kenyan Swahili/ North Eastern Bantu
Function:	Syntactic (Concord)	Lexical	Pragmatic
Access L.:	Formulator	Lemma	Lemma/Formulator

(25) Morphological structure and acceptability of outer items: lexicalized

Structure:	INFL-P [SM-TAM-(RM-)(OM-)]	≠ STEM [Base(-DSuf)]	≠ INFL-S [-PreF-FV]
Source:	↑ Kenyan Swahili	↑ Sheng (coined/lexicalized)	↑ Kenyan Swahili/ North Eastern Bantu
Function:	Syntactic (Concord)	Lexical	Pragmatic

Access L.:	Formulator	Lemma	Lemma/Formulator

Basically the abstract overall structure and the Grams are inherited from Swahili; it is quite clear that the Grams in INFL-P are provided almost totally from Swahili with radically simplified concordance, while INFL-S cluster is slightly open to take a non-Swahili exponents, i.e., PreF *-anga* or *-engo* etc. (see (4) in 1.2)[14].

Although, at least according to the written texts investigated in this study, the composite nature in verbal morphosyntax of Sheng is only scarce at present, it can be pointed out that the Grams "inhabit" their own morpheme cluster depending on their origin and do not invade the others' "habitat" as illustrated in (24-25). This is nothing but a paraphrase of the USP effect at the morpheme cluster level and, as mentioned in 3.1 and 3.3.2), what makes it possible is the isolation process as a general bias in making up a mixed language. This bias, additional principle d) in (23), can be seen as providing ground for excluding DSuf from the mixed STEM described in 3.3.1). From this point of view, it is interesting to note that the reflexive (REF) *ji-* persists in the structure and is actively utilized comparing to the other OMs (see 3.2.2), since this can be seen as a result of both d) "habitat" effect and c) pattern-preserving bias.

5. Conclusion

Through the description presented in this paper, it is confirmed that there is a set of dynamic bias that underlies the abstract morphosyntactic structure and it may provide Sheng with grammatical compositeness, although at present there are shallow indications that prove any clear systematic difference between Sheng and Swahili (especially PiS) at least in the verbal constituents. However, it is worth pointing out that the formal characteristics found in relative constructions, for example, may suggest a possibility for Sheng to obtain abstract compositeness which is contributed by (interaction of) the bias summarized in (23).

Regarding the acceptability of the outer Grams, it can be said that the most likely to be incorporated would be the PreF-FV sequence in the verbal complex. As mentioned in section 4, this predication is grounded by the "habitat" effect, but it should be also mentioned that PreF (*-anga* etc.) in Sheng denotes a kind of pragmatic notion rather than syntactic one, i.e., functionally analogous to, say, a discourse marker, which has significant relevance with Content Morpheme in terms of "cognitive dominance (Tomasello, 2003)" and is, in turn, well likely to be borrowed. Thus it may be worth attempting to investigate the process in which Sheng may (or

14 In this sense, the Gram which is the most likely to be accepted would be the PreF-FV sequence and, if change towards grammatical compositionality goes on, there may be a possibility that other INFL-S items found in North Eastern Bantu in general such as (exponents of) *-ile* of the past tense/anterior aspect marker might be a next candidate to be introduced in Sheng Gram system. However, it should be also mentioned that the incorporation of PreF *-anga* in the structure may well be grounded by the fact that it denotes pragmatic (rather than inflectional) function as a kind of discourse marker and if it is so, inflectional *-ile* should be regarded as an unsuitable element to be involved.

may not) obtain more grammatical hybridity by functional (or so-called "usage-based") approach. Whichever approach may be taken, there is still a considerable need for robust descriptions on linguistic structure especially of its noun-relating morphosyntax and its various sentential constructions etc., which are left to be described so far in the study of Sheng.

Abbreviations

1sg, 2pl ...	: Person + Number	EL	: Embedded Language
1, 2, 3...	: Noun Class	ML	: Matrix Language
SM (S in gloss)	: Subject Marker	PST	: Past
OM (O in gloss)	: Object Marker	PRES	: Present
TAM	: Tense and Aspect Marker	FUT	: Future
RM (R in gloss)	: Relative Marker	PERF	: Perfect (Anterior)
INF	: Infinitive Marker	CONT	: Continuous
DSuf	: Derivational Suffix		
PreF	: Prefinal	APPL	: Applicative
FV	: Final Vowel	CAUS	: Causative
CPx	: Class Prefix	NEUT	: Neuter
PPx	: Pronominal Prefix	PASS	: Passive
LOC	: Locative Marker	RECIP	: Reciprocal
DEM	: Demonstrative		
REL	: Relative Stem		

References

Abdulaziz, M. and K. Osinde. 1997.
Sheng and Engsh: development of mixed codes among the urban youth in Kenya /in/ *International Journal of the Sociology of Language* 125, pp. 1–21.

Bosire, M. 2006.
Hybrid Languages: The case of Sheng, /in/ Olaoba, F. Alasanyin and M. Pemberton (Eds.) *Selected Proceedings of the 36th Annual Conference on African Linguistics*, pp. 185–193, Cascadilla Proceedings Project

Ferrari, A. 2004.
Le Sheng: Expansion et Vernacularization d'une Variété Urbaine Hybride à Nairobi /in/ Akinlabi, A. and O. Adesola (Eds.) *Proceedings of 4th World Congress of African Linguistics*, pp. 479–495, Rüdiger Köppe

Githiora, C. 2002.
Sheng: peer language, Swahili dialect or emerging Creole? /in/ *Journal of African Cultural Studies*, vol. 15, No. 2, Routledge

Heine, B. 1979.
Some Linguistic Characteristics of African-based Pidgins /in/ I. F. Hancock (Ed.) *Readings in Creole*

Studies, E-Story Scientia
Mbaabu, I. 1996.
Language Policy in East Africa, Education Research and Publications (ERAP), Nairobi
Mbaabu, I. and K. Nzuga. 2003.
Sheng-English Dictionary: Deciphering East Africa's Underworld Language, TUKI, Dar es Salaam
Meeussen, A. E. 1967.
Bantu Grammatical Reconstructions /in/ *Africana Linguistica III*, Musee Royal de L'afrique Centrale, Tervuren
Mugane, J. 2006.
Necrolinguistics: The Linguistically Stranded /in/ Mugane et al (Eds.) *Selected Proceedings of the 35th Annual Conference on African Linguistics*, pp. 10–21, Cascadilla Proceedings Project
Myers-Scotton, C. 1979.
Nairobi and Kampala varieties of Swahili /in/ I. F. Hancock (Ed.) *Readings in Creole Studies*, E-Story Scientia
Myers-Scotton, C. 2003.
What lies beneath: Split (mixed) languages as contact phenomena /in/ Matras, Y. and P. Bakker (Eds.) *The mixed language debate: theoretical and empirical advances*, pp. 73–106, Mouton de Gruyter
Myers-Scotton, C. and J. Jake. 2000.
Four types of morpheme: evidence from aphasia, code switching, and second-language acquisition /in/ *Linguistics* Vol. 38, pp. 1053–1100.
Ogechi, N. O. 2005.
Does Sheng have a Kiswahili grammar? /in/ Brenzinger, M. and A. Mietzner (Eds.) *Annual Publication in African Linguistics*, *No. 3*, pp. 5–25, Rüdiger Köppe
Shinagawa, D. 2006.
Particularities of Sheng in written texts /in/ *Studies of Integrated Text Science (SITES) Vol. 4*, *No. 1*, pp. 119–137, Graduate School of Letters, Nagoya University
Spyropoulos, M. 1987.
Sheng: Some Preliminary investigation into a recently emerged Nairobi street language /in/ *Journal of the Anthropological Society of Oxford* 18: 125–136.
TUKI (Taasisi ya Uchunguzi wa Kiswahili). 2001.
Kamusi ya Kiswahili-Kiingereza, Toleo la Kwanza (Swahil-English Dictionary, First Edition), TUKI, Dar es Salaam
Thomason, S. G. 2003.
Social factors and linguistic processes in the emergence of stable mixed languages /in/ Matras, Y. and P. Bakker (Eds.) *The mixed language debate: theoretical and empirical advances*, pp. 21–39, Mouton de Gruyter
Tomasello, M. 2005.
Constructing A Language: A Usage Based Theory of Language Acquisition, Harvard University Press
Yukawa, Y. 1980[a]. [in Japanese]
On Kikuyu Verbs, /in/ *Asian and African Linguistics 9*, pp. 43–82, ILCAA, Tokyo
Yukawa, Y. 1980[b]. [in Japanese]
On Kamba Verbs, /in/ *Asian and African Linguistics 9*, pp. 83–118, ILCAA, Tokyo

Texts investigated

Maillu, David. 1989.
Without Kiinua Mgongo, Maillu Publishing House, Nairobi
Wiji, G. and Kitu Sewer. 2005.
Sheng Interviews, /in/ Binyavanga Wainaina (Ed.) *Kwani? 03*, pp. 170–190, Kwani Trust, Nairobi