The morphologization of an Arabic creole

Jonathan Owens
University of Bayreuth

East African Nubi has classic attributes of a creole — it was formed in a short period of time and its structure diverges dramatically from its lexical source, Egyptian and Sudanic Arabic — yet it differs from most creoles as well in that it has a fairly robust morphology (Owens 2001). One could call it a morphologically rich creole, even if its morphology is much simpler than that of Arabic. Understanding why this happened in Nubi presupposes having a solid descriptive historical linguistic account of how this came about. While concentrating on this latter issue, this paper lays the groundwork for understanding the ‘why’ by examining Nubi relative to current theories of creole genesis, including recent SLA models, and by showing that discourse embeddedness played an important role in guiding its development.

Keywords: Arabic, creoles, morphological change, morphological restructuring, theories of language change

1. Introduction

The formal origins of Nubi morphology are mostly quite transparent. They derive either from refunctionalization of Sudanic or Egyptian Arabic morphemes, or in one important set of cases as will be argued here, from the refunctionalization of phonemic material. Additionally, there is a certain amount of frozen morphology, material which in the Arabic source is morphemic but which in Nubi is simply part of a complete lexeme. A problem with a less obvious answer is why Nubi morphology emerged as it did. While the latter question is the more interesting one, addressing it requires having a good descriptive account of the former. This paper thus concentrates on the first issue, while developing a number

---

1. Abbreviations are as follows. BV basic variety, CA Cairene Arabic, EA Egyptian Arabic, KA Khartoum Arabic, NA Nigerian Arabic, SA Sudanic Arabic, TL Target Language, WSA western Sudanic Arabic

ISSN 0920–9034 / E-ISSN 1569–9870 © John Benjamins Publishing Company
of historical and analytic constructs which can contribute to an understanding of the second.

For this latter task, creole linguistics has a considerable history of developing explanatory models for the genesis of creole languages. As Nubi has been largely ignored in this debate (apart from those few linguists who deal in Arabic creoles), I will briefly deal with approaches which are argued to be of little interest to understanding ancestral Nubi — the Bioprogram, foreigner talk, feature pool, and prior pidginization theories — and then discuss in greater detail the newer attempts to understand creoles in terms of SLA.

This paper is divided into 5 parts. Section 2 summarizes key background aspects of Nubi — when and under what circumstances the creole came into existence, the Arabic dialectal origins of Nubi, the lack of substratal influence in the resulting creole, and the methodology used to explore its morphological genesis. Section 3 is a descriptive and historical account of the origin of morphemes in Nubi, tracing their fate as they moved into Nubi, looking both at those which became frozen into non-morphemic components of Nubi lexemes, and at those which maintained a morphemic integrity in Nubi. In Section 4, as in Owens (2001 and elsewhere) it is argued that overall reductionist models explaining Nubi genesis capture at very best a part of the development of Nubi. An alternative explanatory perspective is offered in which four independent variables are identified which are pervasive and central to understanding the emergence of Nubi morphology. In Section 5 the special status of Nubi vis a vis both creoles and Arabic is reiterated.

2. Background to Nubi

2.1 History

2.1.1 Pre-southern Sudan
The backdrop to the origin of Nubi comes from a colonial enterprise initiated by Mohammad Ali, the Khedive of Egypt and effectively the most powerful ruler of the Ottoman Empire. In 1820 he ordered the invasion of the northern Sudan, whose Funj rulers were quickly subdued.

There is no reliable information about the language of this conquering army. However, there is no reason to assume, as does Kihm (2011:46), that it would have been a pidgin or a creole Arabic. The Sudanic area (see 2.2 below) had long since supported a large Arab population. Beginning in 1214 it is reported that Arab tribes began leaving southern Egypt and moving into what is today northern Sudan. Until 1315 the northernmost Nubian Kingdom of Dongola kept this population at bay, though in that year, at the instigation of the Egyptian Mameluke
Sultan Baybars there was a large-scale invasion of Arabs which conquered the kingdom. In the wake of this invasion, Arab populations moved westward into the Lake Chad area, so that it is reported in 1390 that there was already a large population of Arabs in that region.

In the aftermath of this invasion, Arabic tribes regularly assimilated non-Arabic elements into their society. Braukämper (1994) reports that a particularly important encounter in about 1635 was between Fulani groups moving eastward with Arab tribes in eastern Chad. Both groups were nomads, the Arabs camel (and sheep) nomads, the Fulani cattle nomads. As camels, unlike cattle, were unable to thrive south of the 13th parallel, Arabs adopted the Fulani cattle culture, thus forming what is today known as the Baggara (< baggaara ‘one who works with cattle’) culture. The baggara culture corresponds fairly precisely with what has been identified as the western Sudanic dialect area (WSA, see 2.2). Braukämper postulates that this culture formed originally around Waddai in eastern Chad, and spread both eastwards into Darfur and Kordofan and westwards into what is today NE Nigeria.

As Braukämper concludes (1994: 25), in the Fulani-Arab cultural encounter a large number of Fulani were arabicized — today there is only a negligible Fulfulde-speaking population in Chad. As far as the native Arabic of the WSA, there are no indications in it of pidginization or creolization. The Arabic dialects of this region can be integrated on a dialectal basis as ones successively more like those of Upper Egypt as one moves from northeastern Nigeria across Chad, into northern Sudan and on into Upper Egypt (Owens 1994).

Extrapolating from this brief history, there is no a priori expectation that a large-scale invasion of the Sudanic region in 1820 from Egypt would have produced a creole form of Arabic. To the contrary, the invading armies would have

---

2. Kihm (2011: 46–7) states that the origin of Nubi was 1820, citing Wellens (2005) as the source for his position. There are two problems in this, however. First, Wellens (2005: 30) accepts the position of Owens (1997) that creole formation did not start until the opening of the southern Sudan after 1851. Secondly, Wellens does suggest that the 1820 invasion would have produced a form of pidgin Arabic (2005: 26). While this may well have been the case, (1) there is no documentary evidence for this, as Wellens herself notes, and (2) much more importantly, there is equally no evidence for a connection between such a possible variety and Nubi or Juba Arabic.

I believe Kihm and other creolists might have an interest in extending the origin of Nubi back to 1820. Starting Nubi in 1820 gives it a considerably longer gestation period than does 1851. The latter date, yielding effectively less than 37 years (1851–1888, see below), sets ancestral Nubi within the period of formation usually attributed to the radical Creoles. Radical creoles, however, have come to be associated with important linguistic attributes, in particular their heavy reliance on substratal features. Nubi, however, is a ‘temporal’ radical creole with negligible substratal influence. If ancestral Nubi is dated to 1820, troubling structural issues can be avoided, since a longer formation period is thereby implied.
met in part the ‘same’ Arabic-speaking populations, and to the extent that any language switch in favour of Arabic was thereby engendered, there is no reason to suppose that anything else happened than had been happening in the area since 1400, namely that new speakers of Arabic would have learned Arabic perfectly. Indeed, assuming that the situation which Mohammad Ali’s army met in the northern Sudan was similar to that today, the expectation would have been for minority groups to adapt linguistically to the dominant Arabic and learn Arabic as their major L2 lingua franca. Extrapolating further, assuming that the northern Sudanese and Egyptians who later moved into the southern Sudan and became the learning model for ancestral Nubi applied these sociolinguistic language-learning norms to their usage in the southern Sudan, the conditions would have been met for the use of a normal (not simplified, see 4.1 below) Arabic. As will be seen in Section 3 below, the linguistic testimony substantiates this assumption.

2.1.2 The southern Sudan (today South Sudan)

The next, and for the history of Nubi, crucial step in the Egyptian colonial expansion was to move into the southern Sudan. In fact, the origins of Nubi can be dated with considerable precision. It began in 1849 or 1851 and had achieved its current form, broadly, by 1888. 1849 is the date of the first trading expedition, led by Major Selim, which pushed into the southern Sudan, then made largely impassable by the Sud swamp, a natural barrier that had cut off the area from the North. This expedition left the South the same year. In 1851 the first permanent settlement was established in the area of Wau, a mission supported by Austrian missionaries. From then on settlement from the north proceeded quickly. These persons were supported by and the agents of, the government of the Anglo-Egyptian Sudan, nominally under the control of the Ottoman empire. By 1870 Schweinfurth (1918), who surveyed the westerly quadrant of the southern Sudan (Mahmud 1983, Owens 1990) reports that there were trading camps established every fifteen miles. There had already by this time developed a quadripartite social organization in the southern Sudan. On the one hand there were indigenous villagers who

---

3. It should not be supposed that the only direction of language switch would have been in the direction of Arabic. As is the nature of language contact, shift can go in many directions. In northeastern Nigeria, for instance, the dominant trend historically was in the direction of Kanuri (including assimilated Arabs). Overall, Arabic is but one of many languages in the region, and hardly always the dominant one.

As far as the attested recruitment of probably non-Arabic mother tongue speakers into Mohammad Ali’s army, there is again no presumption that such speakers would have in the long term spoken a pidgin Arabic. To the contrary, those sent to Egypt for training would probably have learned Egyptian Arabic, which by the nineteenth century was all-dominant.
Jonathan Owens

were often the object of the slaving agents in the camps, but also served as the basic recruits for the soldiers in the camps. Within the camps three groups emerged. At the top were the commanders, a variegated group of adventurers from different Arab and European countries. Among these was Emin Pasha (Edouard Schnitzer), a Silesian by birth and governor of Equatoria province, who, as will be seen, figured importantly in the endgame of the emergence of Nubi. A second tier of camp inhabitants were the soldiers of the Egyptian government, recruited apparently largely from the northern Sudan, both Arabs and Nubian speakers (Dongolawi). At the bottom of the social hierarchy in the camps were the recruits from African villages and their many wives and acolytes. These constituted the largest percentage of the three groups, estimated to constitute at least 80% of the camp population (Owens 1990: 222). They were consciously recruited from across the array of 25–30 different ethnic groups, similar to the logic of the Atlantic slave trade, so as to minimize the possibility that any one group become dominant within the camps. This large mixture of ethnic southern Sudanese groups is certainly an important reason for (1) the emergence of Nubi and (2) the lack of any dominant southern Sudanese-based lingua franca.

It can be surmised from many late nineteenth century eyewitness sources that there was considerable tension among the three different groups within the camps. These differences were particularly crucial beginning in 1886. This is when the Mahdi, Mohammad Abdallah, who led the successful uprising that resulted in 1885 in the conquest of Khartoum and all of northern Sudan and the temporary end of Egyptian-Ottoman rule in that region, attempted to conquer the last bastion of nominal Egyptian rule in the Sudan, Equatoria Province in the southern Sudan. When the Mahdi’s attacks began, a good number of the medial group, the northern Sudanese, defected to the Mahdi. Still, Equatoria held on, with the Equatoria governor Emin Pasha ruling from Rejaf, near the contemporary South Sudan capital of Juba. Emin Pasha’s plight excited the interest of the West, transfixed by the death of the governor of Sudan, General Charles Gordon, and loss of Khartoum, and an Emin Pasha relief expedition was quickly organized and entrusted to the American explorer and self-publicist, Henry M. Stanley. In a daring but also sanguinary march across the contemporary Congo Republic, he found Emin Pasha and the remnants of his command at Wadelai on the northern shores of Lake Albert in 1888. Stanley took Pasha along with most of his Egyptian officers to the (now) Tanzanian coast. Much of his army, however, had no interest in rescue. These were, in fact, not Egyptians, but rather the southern Sudanese recruits and their families and followers. These stayed on in Uganda initially and became the ancestral population of the contemporary East African Nubi.

There is one more important socio-demographic note. Emin Pasha’s army consisted of two army corps. Only one went with him to Wadelai and thence to
East Africa. The other stayed on in the southern Sudan, where eventually they were defeated by the Mahdi’s army. However, it may be tentatively concluded that this population was not entirely wiped out. Until today Juba Arabic and Nubi are mutually intelligible, and nearly all of the basic morphological observations made for Nubi below also pertain to Juba Arabic (see Watson & Biajo Ola 1986; Smith & Ama 1985: 205–19; Tosco & Manfredi 2013; Miller 1993; Wellens 2005: 31 for descriptions and summaries). These close similarities are explained by the assumption that the structures we know today for Nubi were already in place by 1888, and hence survived the demographic split which occurred in that year.

For comparative creole purposes this survey has highlighted the following key points.

1. The gestation period for Nubi was no more than 37 years, putting it diachronically among the radical creoles.
2. Ancestral Nubi society consisted of a demographic majority who spoke a great many languages, no one of which was dominant in the era prior to the Sudanese Egyptian colonial era. Arabic, or a target thereof, as the language of the colonizing power was a custom-made lingua franca.
3. Ancestral Nubi society was socially stratified, with antagonisms among the different social groups. Tellingly, there was in particular a lack of trust between the northern and southern Sudanese. It can be postulated that this social cleavage would have been one more factor inhibiting the learning of a normal Arabic by the southern Sudanese cohorts.

All in all, therefore, conditions were excellent for the emergence of a new language, a creole.

In this article I use Nubi as a stand-in for the two basic varieties of Creole Arabic, Juba Arabic and East African Nubi. The latter is not only the better described, thanks to Wellens (2005), but also the variety better known to the present author (particularly Nairobi Nubi).

As a terminological note, in this article I will contrast Arabic with Nubi, i.e. it is assumed axiomatically that Nubi is not a form of or variety of Arabic, even if it historically derives from it. Arabic in this context means particularly the Arabic of Egyptian and the Arabic-speaking Sudanic region (= northern Nigerian, northern Cameroon, Chad, and Sudan), though typologically the differences which Nubi shows against Egyptian and Sudanic Arabic are largely replicated for all varieties of Arabic, including Classical Arabic (see, e.g., n. 48 for the striking similarity between Emirati and Nigerian Arabic). ‘Sudanic Arabic’ encompasses the entire Arabic-speaking region from Nigeria to the Red Sea. ‘Sudanese Arabic’ is the Arabic of the state of Sudan. The term ‘western Sudanic Arabic’ is explained in the next section. ‘Ancestral Nubi’ refers to the ancestor of Nubi postulated to have
arisen in the second half of the nineteenth century, or to the pre-1888 population of southern Sudan among whom Creole Arabic emerged. Given what has been said in this section, modern Nubi, Nubi as we know it today, is assumed to have existed by 1888. 'Southern Sudan' in this paper refers to the area in which Nubi arose in the nineteenth century. This includes what has become today the new nation of the South Sudan, as well as areas bordering in northern Uganda, NE Congo Republic, and the extreme east of the Central African Republic. Finally, except where it is relevant to understanding the development of morphology, no special attention will be given to the origins of Nubi phonology. I will conventionally represent the etymological origin of Nubi words, where relevant, with corresponding Cairene, Khartoum forms, even if it is possible that the exact origin is different (see Owens 1985; Kaye & Tosco 2001: 78; Wellens 2005: Chapter 2 for more details and discussion).

2.2 Arabic dialect sources

Depending on the perspective, two or three dialect areas are represented in Nubi. There is a broad distinction between Cairene or northern Egyptian Arabic (EA, approximately the area north of Minya and Bani Swayf) on the one hand, and Sudanic Arabic (SA) on the other. Sudanic Arabic includes all of contemporary Sudan, Chad, northern Cameroon, and northeastern Nigeria. On the other, the Sudanic region can be divided into a number of sub-areas. Two are of interest here. One is that of the Nile Valley, for which Khartoum Arabic (KA) will be eponymously used for illustration, and western Sudanic Arabic (WSA). Geographically WSA begins in Kordofan (see Manfredi 2010) and includes Chadian, northern Cameroonian and Nigerian Arabic (Owens 1994). As noted above, according to Braukämper (1994) its historical locus was eastern Chad.

These three dialectal sources are readily discernible in Nubi. As ever in the case of dialects, there are also many features which either two or all three of these dialects share. This paper is not about the dialectal origins of Nubi, and therefore the dialectal differences will be mentioned only to the extent that they are relevant to understanding the origins of Nubi morphology. Still, it is appropriate to illustrate how the different dialectal origins are discernible.

4. The three dialectal sources correspond to documented origins of Arabic-speakers present in the southern Sudan of the second half of the nineteenth century. Egyptian/Cairene Arabic would have been the dialect of many officers, Nile Valley (Khartoum) Arabic the Arabic of the medial group described in the previous section, while the western Sudanic variety, if not also spoken by some of the northern soldiers, was brought by itinerant traders, jallaaba as they are called, who came into the southern Sudan seasonally.
Arabic ‘jim’. The sound which corresponds to the letter ‘j’ or ‘jim’, a sound whose proto-value in Arabic is itself a matter of ongoing debate (see Owens 2013a for recent summary), is realized in Nubi as either /g/ or /j/. /j/ is by a very large margin the more common reflex in Nubi.

(1) Cairene vs. Sudanic
   a. /g/  
      rāgi ‘man’, < raagīl
      gildū ‘skin’ < gild-u
   b. /j/ = [dž] or [j]
      rījāl ‘men’ < rījāal
      jēre ‘run’ < jara
      jisīm ‘body’ < jisīm
      ja ‘come’ < ja

The /g/ reflex goes back to Cairene /g/, whereas the /j/ is common throughout the Sudanic region, including both Khartoum and western Sudanic Arabic.

3msg object suffix (see 3.1.1). This morpheme appears in two forms, -u and -a. The -u variant is that of Khartoum and Cairene, the -a of WSA.

(2) 3msg
   a. gild-u ‘his skin’ = Nubi gildū ‘skin’
   b. ragābt-u ‘his neck’ = ragābtu ‘neck’ vs.
   c. id-a ‘his hand’ = ida ‘hand’
   d. batn-a ‘his stomach’ = batna ‘stomach’

Two general observations are relevant about the dialectal origins. First, as far as variants go, the Khartoum dialect is a medial dialect as it were. It shares features of Cairene, as in (2a, b), as well as with WSA, as in (1b). Secondly, in general it is

5. Unless otherwise stated, the meaning of the Arabic word is the same as glossed for Nubi (or vice versa).

6. A reviewer suggests that these could originally be epenthetic vowels, not 3msg object pronouns. A full discussion of this issue belongs to the historical phonology of Nubi, which is outside the scope of this paper. However, as one observation, typically in nouns and adjectives a final CV syllable will be created out of an Arabic C*-final form, by adding a postthetic [i] after a dental consonant, e.g. tini ‘mud’ < tiin (or < tiin). This final -i, moreover, may be targeted for deletion under conditions which still require phonological specification. A final -u and especially, a final -a as in (2) has no general phonological explanation and unlike a postthetic -i, these are not deleted variably. At the same time, the forms perfectly match the morphemic sources assumed in this paper.

7. The dialectal situation of Egypt is more complicated than that presented here. Cairene is hardly representative of Egypt as a whole, though it clearly does provide one distinctive input
the features of Sudanic Arabic — either WSA singly or Sudanic Arabic in general, which predominate in Nubi. This is recognizable in two ways.

First, on a token basis, where the Nubi displays a contrast between Sudanic Arabic and Cairene sounds, there are more lexical tokens of the Sudanic Arabic forms than the Cairene. (1) above is typical in this respect — there are far more /j/ SA reflexes of ‘jiim’ than there are Cairene /g/.

Secondly, certain sounds distinctive of Cairene Arabic do not appear at all in Nubi. For instance, Cairene Arabic has a reflex */ʔ/ corresponding to the Arabic letter ‘qaaf’ (again, the proto-quality is problematic). Thus, corresponding to (2b), in Cairene one has:

(3) raʔabt-u ‘his neck’

In this case it could be observed that the reflex /ʔ/ in general has been lost in Nubi, so nothing can be deduced from this instance. However, in the WSA area there is a reflex /ʔ/ which derives not from ‘qaaf’ but rather from ‘ain’, */ʕl/, and this may surface as a semivowel intervocalically. Nubi for, instance, has gayi ‘stay, stand’ which derives from < gaaʕid or gaayid, ultimately < *gaaʕid. While the /ʔ/ has been lost in Nubi, its historical presence can be discerned in a /y/ which results from its loss. No such correspondences from Cairene /ʔ/ ‘qaaf’ are found.

In the case of Cairene ‘qaaf’, on the other hand, it is always the Sudanic value /g/ of ‘qaaf’ which is chosen, as in (2b) ragabtu above. Thus it is not only that Sudanic values are dominant on a token basis; in some instances as well, on a type basis it is the Sudanic type value which categorically appears in Nubi. In the following description a number of instances will be pointed out attesting to the dominance of SA-origin forms in Nubi.

2.3 Corpus, background tracking

Before moving on, it is necessary to explain that at various points in this paper reference will be made to inferences which can be drawn from discourse as it might have appeared to ancestral Nubi. In order to ascertain what nineteenth century discourse might have looked like, modern day spoken Arabic corpora will
be adduced, much in the way historical linguistics uses contemporary forms to reconstruct past ones. These will be termed ‘background tracking data’, as they attempt to recreate the discourse environment of ancestral Nubi. For this, two data sets are used, as described in n. 10 as well as below. The evidential value of this exercise is to determine whether the suggestion that such and such a morpheme could have been a source for a Nubi morpheme passes what can be termed the ‘discourse realism test’. Would ancestral Nubi in fact actually have been exposed to the proposed source? In one case, as will be seen (3.5.1), background tracking provides important evidence for determining the greater plausibility of one proposed historical derivation over another.

Background tracking is akin to Labov’s (1994: 580) probability matching. In probability matching, populations are distributed as a function of the stimulus they are presented: if ten fish in a pond are fed from two sources, with the amount of food apportioned at each of the two sources on a 4:1 basis, in a short period of time, eight fish will be feeding from one source and two from the other. Labov has used probability matching to explain the gradual change or lack thereof in vowel systems and Spanish plural -s deletion. Background tracking as used here has, necessarily, a more modest aim, namely to show that a form would have been frequent enough to have been used in one capacity or another by ancestral Nubi. For reasons discussed in 4.3.2 below, discourse frequency, unlike fishfeed, is in and of itself not cause for the emergence of Nubi structures, and unlike Labov’s examples which assume an overview over the entire linguistic environment (or a representative sample thereof), an important variable is missing from the current statistics, namely the interlanguage Nubi forms which ancestral Nubi would have heard and created, along with the Arabic. Arabic frequency does, however, play an important facilitating role, and indeed, in some measures, for example those presented in Table 1 below, the Nubi results do in fact track the Arabic model quite faithfully.9

Given the importance of Sudanic Arabic as a source of Nubi, a corpus of Sudanic Arabic from Nigerian Arabic will be the main corpus used, though at one point Emirati Arabic is adduced as well.10 Though a source further east in the Sudan would have been better, in these cases one makes do as best as one can with what one has at one’s disposal. It might be objected here that the relevance of a contemporary Nigerian Arabic corpus to the 19th century Sudan is tenuous. From

---

9. While background tracking bears a resemblance to the ‘feature pool’ approach developed by Mufwene (e.g. 2010) and Siegel (2008: Chapter 8), in fact it is drawn directly from variationist sociolinguistics, including koineization research (e.g. Thelander 1982).

10. Two sets of texts are used, designated set 1 and set 2. Set 2 has two texts, GR21AEJX and IM107. Set 1 is larger consisting of 15 texts. Both set 2 texts are available online, from set 1 the following are available online: IM23, IM38Habab, GR22AEJX.
the perspective of Arabic dialectology and language history, however, the matter is otherwise. First, the WSA dialect area, which includes Nigerian Arabic at its western extreme and Kordofanian and Darfurian Arabic at its eastern, is a remarkably homogeneous and very distinctive dialect area among Arabic dialects. As described in 2.2 above, and illustrated elsewhere (e.g. n. 14), there are incontrovertible and specifically WSA morphological traces that are found in Nubi. Secondly, and related to this point, there is little evidence for dramatic change within this area (homogeneity implies this) since at least 1635, so that it is quite plausible to take this Arabic variety as one model for what ancestral Nubi would have heard. Indeed, given the relatively short time span separating the formation of Nubi from the present, one hopes that continuing research collects larger corpora, both from Nubi and SA, to allow for more such comparisons.

2.4 Nubi and lack of substratal influence

A considerable amount of research over the last 20 years has revolved around the issue of the role of the substrate in creole languages. Plag (2008: 116) notes the close relationship between discussion of substrate and SLA processes. Arguing for a strong substratal role in creole development, Siegel (2008: 81) asserts that ‘The point must be stressed, however, that most of the morphology in a pidgin or a creole is not derived from the lexifier in terms of both form and function.’ Others (Mufwene 1990: 9, 13, Smith, 2006: 60, Winford 2006, and Kouwenberg 2006: 207) caution that the origin of creole structures may be heterogeneous and do not preclude significant lexifier grammatical influence, a position echoed in McWhorter’s assertion (2012: 171) that ‘the basic hybridity of creoles is, in itself, obvious to all.’ The case for creoles being largely classed with their lexifier is made most forcefully by Chaudenson (1992), though this seems to be a minority position among creolists.

Without entering into the fraught issue of the classification of creoles, it is interesting to determine to which ‘stratist’ layer Nubi belongs, and in the current description it will emerge that most of the morphemes have a transparent Arabic origin, even if they are transformed to a lesser or greater degree in Nubi. The role of the substrate languages is secondary, to non-existent. Owens (1991a) explicitly looked for substratal traces from phonology, morphology, and syntax in Nubi from two languages, Mamvu and Bari, whose speakers are known to have been part of the ancestral Nubi substrate population, and it was argued that Arabic played the dominant role in accounting for the 20 structures examined.
This is not to say that structural substratal features cannot be identified in Nubi. However, particularly in regards to morphology, it is safe to say that substrate influence is negligible at best.

I finally note here that I do not deal with the Bantu strata of Nubi, Swahili and Luganda (see, e.g., Khamis 1994).

3. Origins of Nubi morphology

The origins of Nubi morphology are best understood through the prism of their origins in Arabic. Without exception, as will be argued here, all Nubi bound morphemic material can be explained to the largest degree as a transformation or re-functionalization, in some cases a radical transformation of original Arabic morphemes. In two instances Nubi maintains the original Arabic morphemic value on a one-to-one basis.

The transformations can be represented on a sextepartite schema, as follows.

(4) From Arabic to Nubi:
   a. Arabic morphemes become frozen morphemic traces
   b. Arabic morphemes incorporated into category indicators
   c. Arabic morphemes functionalized into corresponding Nubi morphemes
   d. Arabic morphemes reinterpreted into related, but often radically different Nubi morphemic values
   e. Arabic phonological material transformed into Nubi morphemic
   f. Arabic morphemes lost in Nubi

These cases will be discussed in turn in the rest of this section, (4f) in Section 4.1.

3.1 Arabic morphemes become frozen morphemic traces

The first case concerns the fate of Arabic morphemes which although recognizable in Nubi, nonetheless have no function, having been assimilated as historical relics in newly formed Nubi lexemes. While these frozen traces have no discrete morphemic value in Nubi, nonetheless there are three prominent cases which have been assimilated in such a regular way that they can be considered to have constituted new formal declensions or conjugations in Nubi. In addition there are

---

11. A good candidate is the distinction between alienable and inalienable possession, which is explicitly marked by syntactic means (Wellens 2005:112, Heine 1982:42). As Arabic has no formal marking of the distinction, this could well be due to substratal influence. Another are various phonological features, such as vowel harmony and the loss of final obstruents.
number of single morphemes frozen into high-frequency lexemes which can be noted.

3.1.1 The Arabic definite article al-, possessive/object pronouns
Two major classes of Arabic morphemes occur which have been lexicalized in Nubi and frozen into the new Nubi lexeme. The larger class involves the definite article al-.12

Briefly, in Arabic the definite article marks definiteness, which is established by genericness, previous mention, or by frame-based context (for discussion, Owens 2009). Lack of the definite article indicates existence, implying either irrelevant or discourse-relevant reference: beet ~ al-beet ‘a house, any old house’ ~ ‘the-house’.

Al- as a definiteness marker has been lost in Nubi. Still, its trace is found consistently in one context, namely:

(5) al-*h/s-V > IV

That is, the definite article as a form is regularly maintained if the original lexeme began with a pharyngeal consonant followed by a vowel. The vowel can be any of /a/, /u/, or /i/.

(6) lufár ‘hole’ < *al-hufra
    lobú ‘wind’ < *al-habuub
    lasaya ‘stick’ < *al-sásaaya13

This diachronic rule is specific to this combination. Non-initially a *h or *s is either simply dropped, or *h may surface as h.

(7) amer ‘red’ < ahmar
    bahar ‘river’ < bahar
    laam ‘meat’ < laham

Word initial /a/, on the other hand, usually does not occur in Nubi with the definite article trace.

12. Al- is Sudanic Arabic, il- Egyptian. In both dialects, the initial vowel is frequently elided in discourse, and in Nubi in most tokens only l- survives. In the few cases where the vowel survives in Nubi it is a-, e.g. azól (or ajól)’person’ < az-zool ‘the person’, with z-z > z (Nubi regularly degeminates). I assume on this basis that the Sudanic definite article variant al- is the source.

The morphologization of an Arabic creole

(8) arnam ‘rabbit’ < arnab
asili ‘origin’ < ašil
akili ‘food’ < akil
awa ‘father’s sister’ < ? (non-Arabic source)

The retention of the definite article in Nubi thus contrasts for instance with the retention of a frozen definite article in French-based creoles before a vowel-initial noun (Taylor 1961: 85). In Nubi the context specifically targets the two initial pharyngeal sounds.

There are only a few exceptions to these rules. In three *ʕ words, no definite article surfaces.

(9) ajusi ‘old person’ < *rajuuz
aduy ‘enemy’ < sadu
afiya ‘life’ < ʕaafiya ‘health’

On the other hand, in four tokens the definite article surfaces in non-*ʕ-words.

(10) lifil ‘elephant’ < al-fīl
lungara ‘drum’ < *al-nugura (?)
larda ‘ant’ < al-arḍa
lufás ‘axe’ < al-faas

In general, however, the very specific and definable context where *al- remains in Nubi is striking.

The second set of morphemes are body parts and kinship terms, which often surface in Nubi with a frozen first or third person singular possessive pronoun. This is 1SG -i or -i and 3MSG -u or -a (see (2) above for differing dialectal origins) in the Arabic source. The occurrence of the 1SG vs. 3MSG is regular, the former for kinship terms or persons, the latter for body parts.

(11) ragabtu ‘neck’ < ragabt-u ‘his neck’
gildu ‘skin’ < gild-u ‘his skin’
nyangartu < munxurt-u ‘his nose’ (see Kaye 1987 for etymology)
batna ‘stomach’ < batn-a ‘his stomach’
galba ‘heart’ < galb-a ‘his heart’
ida ‘hand’ < ūd-a ‘his hand’
kasuma ‘mouth’ < yašum-a or xašum-a ‘his mouth’
kitfa ‘shoulder’ < kitf-a ‘his shoulder’
akti ‘sister’ < axt-i ‘my sister’
amati ‘sister-in-law’ < ūammt-i ‘my aunt’, or ammaah-aat-i ‘my mothers’
In one token a kin term has the 3msg suffix.

(12) akú ‘brother, sister’ < axú ‘his brother’

Outside of these two semantic domains, few Nubi words surface with these suffixes. Ebu ‘shame’ < ħeeb-u ‘his shame’ is one of the few. It should also be noted, for comparative purposes below (3.1.2), that the 3msg object pronoun in EA and KA is also -u, masak-u ‘he held it.m’, b-i-msuk-u ‘he holds it.m’ (homophonous here with ‘they hold’), this corresponding to -a in WSA, b-i-msuk-a ‘he holds-it.m’.

Frequency

As seen in (11, 12), Nubi kin terms and body parts largely to the exclusion of other semantic classes contain frozen traces of Arabic possessive suffixes. It is interesting here to adduce the first background tracking with natural texts from Nigerian Arabic. Not surprisingly, body parts and kinship terms occur to a high frequency in possessed position. The corpus survey of Nigerian Arabic texts (set 1) revealed the following statistics for tokens attested more than three times. Without dwelling on the statistics, it is clear that in the background tracking data, kinship terms and body parts occur more often possessed by possessive pronouns than do other words (bagar ‘cattle’, akil ‘food’, darb ‘road’, beet ‘house’).

The occurrence of the frozen pronoun suffix traces in kin terms thus tracks in a general way the higher frequency in the semantic classes in which they occur in natural discourse.

Before moving on to the cases where Arabic morphemes do maintain a morphemic value in Nubi, it can be noted in passing that this paper does not deal with instances of two Arabic words coalescing into single Nubi lexical items, cases like ásede ‘now’ < hassa da ‘now’, lit. ‘now this’, úwede ‘this’ < hu da lit. ‘he this’. In a not insignificant number of cases the Arabic demonstrative da ‘this’, as in these examples, coalesces with a deictic word. Again, Nubi closely replicates frequently-occurring

---

14. The stressed /í/ in this and other kin terms in this list is specifically a WSA reflex, cf. WSA jidd-i ‘my grandfather’ vs. KA jidd-i (Manfredi 2010: 70). Muku ‘brain’ < muxx-u ‘his brain’ and roho ‘spirit’ < ruuḥ-u ‘his spirit’ are also possible examples for this pronominal remnant, though the final vowel could be postthetic with harmonic value.

15. Combining all lexemes into two lexical sets, kin/body parts vs. non-kin/non-body parts, the numbers represent a significant statistical difference (kin + body part vs non-kin x possessed vs non-possessed), df 3, chi sq. — 315, p < .000.
patterns in Sudanic Arabic. In the background tracking set 1, for instance, the independent pronoun *hu* ‘he’ occurs 411 times, of these, in combination with *da* = *hu da* ‘this one’, 80 times (20%). Sudanic Arabic, in turn, follows an areal pattern in which a demonstrative marks deictic words and dependent clause boundaries.

### 3.1.2 Arabic morphemes incorporated conspiratorially into category indicators

The most complex morphemic transition from Arabic to Nubi concerns the verb. The Nubi verb consists of a segmentally invariable stem, quite unlike its Arabic counterpart where perfect vs imperfect is marked by a combination of stem ablaut and inflectional affixes. The Nubi verb paradigm is illustrated in (36) below. Since reference will be made to the Arabic verb paradigm at a number of places in this article, in Table 2a a complete paradigm, perfect, imperfect, and imperative, is given from Sudanic Arabic, which can be taken as representative of any variety of Arabic.\(^{16}\) The key elements are the differentiated person, number and gender components. In the paradigm, the *b*- prefix in the imperfect is an indicative marker (see 3.3.2 below), while the initial *á* or *í* is termed the ‘preformative vowel’.

---

16. See Owens (2009: 29) on this point. The Arabic verb paradigm emerges intact in proto-Arabic, so far as reconstruction allows us to establish, and continues basically unchanged throughout Classical Arabic up to the present, in all Arabic dialects. The paradigm, in fact, is traceable to proto-Semitic, well over 5,000 years old. Table 2a is adapted from Trimingham (1959: 73–4).
Jonathan Owens

Table 2a. Sudanic Arabic verbs paradigms, masak ‘grab, hold, keep’

<table>
<thead>
<tr>
<th></th>
<th>imperfect</th>
<th></th>
<th>perfect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SG</td>
<td>PL</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>b-á-msuk</td>
<td>bi-n-ámsuk</td>
<td>masak-t</td>
</tr>
<tr>
<td>2m</td>
<td>bi-t-ámsuk</td>
<td>bi-t-ámsuk-u</td>
<td>masak-t</td>
</tr>
<tr>
<td>2f</td>
<td>bi-t-ámsuk-i</td>
<td>bi-t-ámsuk-an</td>
<td>masak-t-i</td>
</tr>
<tr>
<td>3m</td>
<td>b-í-msuk</td>
<td>b-í-msuk-u</td>
<td>masak</td>
</tr>
<tr>
<td>3f</td>
<td>bi-t-ámsuk</td>
<td>b-í-msuk-an</td>
<td>masak-at</td>
</tr>
</tbody>
</table>

Imperative

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>á-msuk</td>
<td>á-msuk-u</td>
</tr>
<tr>
<td>F</td>
<td>á-msuk-i</td>
<td>á-msuk-an</td>
</tr>
</tbody>
</table>

In addition to these basic inflectional categories, the Arabic verb is characterized by a rich derivational system, and contains a number of sub-classes of basic verbs, based usually on the presence of a semivowel or long vowel in the stem, as will be introduced as relevant below, and can be briefly illustrated in Table 2b.

Table 2b. Partial table of verb derivations, with traditional numbering of derived stems

II, $C_1VC_2C_2VC_3$, masak + $C_2$ gemination = causative or intensive, massak ‘he had someone grab’

V, ta-$C_1VVC_2VC_3$ = reciprocal, ta-masak ‘cohere, stick together (like dough)’

VI, ta-$C_1VC_2C_2VC_3$, = stative to intensive, ta-massak ‘he kept to his opinion’

VII, an-$C_1VC_2VC_3$ = passive, an-masak ‘he was grabbed’

‘Weak verbs’, i.e. those that do not have the canonical $C_1VC_2VC_3$ stems, characterized by morphophonologically-determined stem changes.

CVVC: gaal ‘he said’, gaal-at ‘she said’, gul-ǝ ‘I said’, ba-guul ‘I say’

CVC$_2$C$_2$: radd ‘return’, radd-at ‘she returned’, raddee-t ‘I returned’, ba-rudd ‘I return’

CVCV: maša ‘he went’, mašee-t ‘I went’, ba-mši ‘I go’

The general outcome of the segmental form of the Nubi verb can be understood in the following terms. First, virtually no derived stems survive in Nubi. The only one attested in the data base which will be introduced below is stenu ‘wait’ < as-tanna. Otherwise, derived verbs formed in Arabic with various morphemes ($t-$, $n-$, $l-$ etc.) are non-existent.17 Similarly, none of the stem alternations associated with the weak verbs are attested in Nubi either.

17. As can be seen in Table 2b, a number of Arabic derived stems are marked by vowel or consonant lengthening. Since long vowels and geminate consonants are lost in Nubi, these derived verb classes are effaced on phonological grounds.
Secondly, there is a very strong tendency for a verb to end in a vowel. If the Arabic verb source stem itself ends in a vowel, this vowel will usually be retained in Nubi. For instance, there is a sub-class of basic Arabic stems which end in \textipa{-a} or \textipa{-i}, verbs like \textipa{aba} ‘refuse’ or \textipa{aški} ‘complain’. These surface in Nubi as \textipa{aba} ‘refuse’ and \textipa{asiki} ‘accuse’. In a number of cases the Arabic has a final /s/ or /h/, consonants which in Nubi are usually lost (see 3.1.1 above), leaving an \textipa{-a} final verb, \textipa{sara} ‘herd cattle’ (< \textipa{saraḥ}), \textipa{rada} ‘suckle’ (< \textipa{raḍa}), \textipa{nafa} ‘be useful’ (< \textipa{nafa}).\footnote{Though as pointed out in Owens (1985), final laryngeals in Sudanic Arabic tend to weaken and are often lost (Owens 2007), so determining which Arabic variety provided the Nubi source is a significant issue.} Otherwise, if the Arabic stem ends in a consonant, there is a very strong tendency for Nubi verbs to end in \textipa{-u}, as will be elaborated on in 3.1.2.3 below.

Thirdly, there is strong tendency for Nubi verbs to be based on the imperative verb.

The last two points are relevant to an account of Nubi morphology, and will be treated in the following three sections. In this section use is made at a number of points to a list of 244 verbs from an unpublished vocabulary of Owens dating from 1978 that will serve as illustration and for statistical treatment.\footnote{It is intended to put this list online. The data is taken over verbatim from the list. In some words my renditions differ from Wellens. The reason for the differences, e.g. Kenyan Nubi (Owens) vs. Ugandan (Wellens), or otherwise need to be elaborated upon elsewhere.}

3.1.2.1 Arabic stems

Imperative

One of the most striking and remarked upon aspects of the historical development of Nubi is the origin of the verb. The position advanced in Owens (1985) was that in the majority of cases — one can estimate between 60–65% of all lexical types — it derives from what in Sudanic Arabic is the mpl imperative verb. Typical examples are as follows.

\begin{equation}
\text{amsuku} < a\text{-msuk-u} \ 'catch\text{-mpl}' \ (or \ -3\text{MSG object suffix}) \\
\text{ašurubu} < a\text{-šrab-u} \ 'drink\text{-mpl}'
\end{equation}

In Sudanic Arabic, the initial \textipa{a-} is what is termed in Arabic dialectology a preformative vowel, while the \textipa{-u} is the plural marker, more on which is in 3.2.1.3 below. That this form derives from Sudanic Arabic specifically is evident from the fact that the preformative vowel is \textipa{a-}. In Egyptian Arabic, indeed in most Arabic dialects, the preformative vowel is \textipa{i-}, e.g. \textipa{i-msik-u}, etc.

A second class of original imperative verb pertains to the so-called medial weak and geminate verbs. These have the canonical shape CVVC or CVCC. After
Jonathan Owens

the regular historical processes of vowel shortening and degemination in Nubi, as well as the addition of final -u (see below), these give correspondences as follows.

(14) Weak medial and doubled verbs (see Table 2b)

<table>
<thead>
<tr>
<th>Arabic</th>
<th>Nubi</th>
</tr>
</thead>
<tbody>
<tr>
<td>fuur 'boil'</td>
<td>furu</td>
</tr>
<tr>
<td>šiil 'carry'</td>
<td>silu or šilu</td>
</tr>
<tr>
<td>sidd 'close'</td>
<td>sidu</td>
</tr>
<tr>
<td>gumm 'get up'</td>
<td>gum</td>
</tr>
</tbody>
</table>

A third set of imperative verbs derive from degeminated, original derived verbs. In Arabic, the class 2 derived forms have the canonical shape CVCCVC (see Table 2b). The medial CC gets degeminated regularly in Nubi, giving the following:

(15) fasulu 'explain' < faṣṣul-u 'explain-mpl'
gofulu 'lock' < ɣafful-u 'lock-mpl'

In the 244 verb database, 147 (65%) of all Nubi verbs derive from these imperative sources.20

Perfect verb

The only other major Arabic form class discernible in the Nubi verb derives from the 3M perfect verb, some 36 (16%) in the data base.

(16) dakalu 'enter' < daxal-u 'they entered'
raham 'show mercy' < raḥam 'he showed mercy'
sara 'herd cattle' < sarah 'he herded, pastured cattle'

Other verbs derive from a rather mixed origin, as will be summarized in the next section, 3.1.2.2.21

20. That is, taking as the total number of verbs 244 – [17 verbs of uncertain imperative or non-imperative origin = 227]; see n. 21.

21. In addition, in some 17 verbs (7% of the total), while a verbal source is clear, it is not clear whether the source is an imperative or perfect verb. Typical examples are tolu 'remain, wait' and seretu 'tear up'. The former derives either from tawwul-u 'stay long-mpl' or tawwal-u 'they stayed long'. The Arabic difference resides in the final stem vowel, but this contrast is lost in the contraction in Nubi of /awwV/ to /o/. seretu derives from šarrut-u 'tear up-mpl' or šarrat-u 'they tore up'. All Arabic consonants degeminate in Nubi and š > s irregularly. There is a fairly general rule which raises an /a/ to /e/ in the context of /r/, which would yield *serVtu. There furthermore is a strong tendency towards stem-internal vowel harmony, based on a front/back parameter, so once the first vowel is /e/, /e/ is favored in the second as well on strictly phonological grounds (Owens 1985 for details). This leaves the precise vocalic source of the second vowel opaque.
3.1.2.2 Other patterns. About 15% of the verbs in the sample are filled out by a somewhat unpredictable melange of forms.

Adjective to intransitive verb
A number of intransitive verbs derive from Arabic adjectives.

(17) Adjective origin Nubi verbs

haraghan ‘sweat’ < haragaan ‘sweaty’
semín ‘become fat’ < samiin ‘fat’
bári ‘be cold’ < baarid ‘cold’

A number of derivations can be noted, based either on patterns which are otherwise attested only in a limited number of cases in Nubi, or based on nonce tokens of inflected verbs.

(18) Odd patterns

maši ‘go’ < maaši ‘going’ (active participle)
faḥim ‘understand’ < faaḥim ‘understand’ (active participle)
azu ‘want’ < taayiz-u ‘want it’ (active participle, CA)
nongusu ‘reduce’ < na-ngus-u ‘we reduce’ (WSA 1pl form)
nisitu ‘forget’ < nisii-tu ‘you.mpl forgot’ (SA, cf. CA nisee-tu)
ligó ‘get < liq-ó ‘they got it’ (SA, cf. CA liqú)
ninzili ‘descend’ < ni-nzil ‘we descend’ (CA or KA 1pl form)
telim ‘dream’ < ta-halim ‘you dream’

3.1.2.3 Final -u, category indicator. Over 85% of Nubi stems replicate well-profiled verbal form classes in Arabic, the dominant one being the imperative verb. A second element that needs accounting for is the overwhelming presence of a final -u on the verbs. This -u is not phonologically required. In general, Nubi stems can end in either a -C or a -V, even if the latter is considerably more common.

The final -u has a morphological origin. Two alternative accounts of the origin of these verbs have been proposed. In Owens (1985), as already stated, the origin is the mpl imperative. Versteegh (1984: 124) on the other hand, proposes that the segmentation is as follows, with -u being the 3msg object suffix

(19) amsuku < *a-msuk-u ‘grab-it.m’

A good summary of the contrasting positions is found in Wellens (2005: 337–45).

Since I will give a new interpretation to the fate of Arabic *-u in Nubi below, I will not go into all the arguments justifying each position, though do note that another argument in favour of the mpl origin is developed in 3.5.1 below. Here it is relevant to briefly note that counterexamples to a categorical interpretation of either position can be found. Against Owens, there clearly are verbs which derive
from a form with singular suffixed object pronoun, whose special morphophonetics is discussed in 3.4 and 3.5.1 below. These are verbs which end in a final stressed vowel, such as wedí ‘give’ < waddí ‘send it.m’ and nadi ‘call’ < naadí ‘call him’ (see n. 40).\(^{22}\)

Against Versteegh, there are intransitive verbs which have -\(u\), e.g. mutu ‘die’, nigitu ‘become cooked, ripen’, so -\(u\) cannot categorically be glossed as ‘transitive’ in contemporary Nubi (as in Versteegh 2008: 170).\(^{23}\)

It is one thing to observe formal identity of Nubi -\(u\) and putative Arabic sources, but another to determine whether they in fact would have been in the discourse which ancestral Nubi heard. Therefore it is relevant to again look at the tracking data, to see whether the two proposals pass the discourse realism test. The data for this comes from text 107SHRIF in NA set 2.

To carry out this count, an interpretation had to be made in respect of Nigerian Arabic, to count all 3msg object suffix forms as if they were -\(u\). That is, as noted in (11) above, in the Sudanic dialect region there are two 3msg object suffix forms, -\(u\), in KA and -\(a\) in WSA. As can be seen from (11), both are actually attested in the Nubi data. Versteegh’s argument that the -\(u\) derives from the 3msg object suffix is interpreted here as a surface realization. Ancestral Nubi would have heard -\(u\) (of course, in addition to -\(a\)). Since Nigerian Arabic lies in the WSA area, the texts only have the -\(a\) realization. In order to give a tracking account, the assumption is made that comparable texts from KA would yield a similar statistic, except the 3msg would be realized as -\(u\), rather than -\(a\), i.e. NA serves as a stand-in for the same morpheme in different allomorphic form in KA. For instance, the textually

\(^{22}\) Note that I did not claim (as Wellens 2005:339 seems to understand) that all verbs come from the imperative plural, only that the final -\(u\) is originally a mpl. morpheme.

It is relevant to point out that the descriptive data in Owens 1977 (Nairobi Nubi) as well as Luffin (2005:267 Mombasa Nubi) differs to a degree from Wellens (2005 Uganda Nubi, Kampala, Bombo) as to the synchronic functioning of final -\(u\). In these studies the link between -\(u\) and transitivity is even less clear than in Wellens. While relevant, discussion of these differences is beyond the scope of this paper.

\(^{23}\) I would add that I find the argument advanced in Versteegh (1984:124) that Melanesian Pidgin (MP) has an analogous source to be a weak one. Versteegh observes an analogy between the MP transitive marker -\(im\) and his proposed -\(u\) as third person object marker. To be fair to Versteegh, research since his proposal (particularly Siegel 2008) has emphasized the role of transfer from substratal languages in furthering the transitive construction (as well as others) in MP. No comparable substratal influence is discernible in the case of Nubi. Note that proponents of transfer/substratal influence are careful to show that the substratal structures would indeed have been present in the population of incipient creole speakers.

To argue in a different direction, for a creole universal of object pronoun transfer, would be circular, and in any case would not correspond to developments in the vast majority of creoles.
attested šif-t-a ‘I saw him’ is taken as a stand-in for KA šif-t-u. For the mpl the problem is not so acute, since in the majority of cases NA -u = KA -u, i.e. the Sudanic region is fairly uniform in representing the mpl subject suffix as -u. There is one difference, however, and that is that in the 3mpl perfect verb, NA (as in the entire WSA region, including Kordofan, Darfur etc.) has -o, whereas KA has -u. Here again NA will be assumed as a stand-in for KA. (20) gives the tracking results.

(20) Tracking results, number of tokens mpl -u and 3msg object/possessor -u, from IM107SHRIF

<table>
<thead>
<tr>
<th>Suffix</th>
<th>Tokens</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>No suffix or other suffix</td>
<td>63</td>
<td>38%</td>
</tr>
<tr>
<td>-u = 3msg object/possessor</td>
<td>29</td>
<td>18%</td>
</tr>
<tr>
<td>-u = mpl</td>
<td>61</td>
<td>37%</td>
</tr>
<tr>
<td>-o = mpl</td>
<td>11</td>
<td>7%</td>
</tr>
<tr>
<td>(Total mpl)</td>
<td>72</td>
<td>44%</td>
</tr>
</tbody>
</table>

In this text segment there are 164 verbs in total (see discussion in 3.5.1, below). Nearly 62% end in an -u or its equivalent, with the mpl value constituting the single largest value, 44% of all tokens. The 3msg value, however, is also well-represented in token terms. The numbers in (20) are therefore quite robust, and one can accept that both hypotheses pass the discourse realism test, i.e. ancestral Nubi would indeed have been exposed to many tokens of both -u verbal values, so both serve as possible candidates for a source of contemporary Nubi -u.24

There are therefore two interesting issues regarding the segmental forms of Nubi verbs, the -u itself, and its derivation from Arabic imperatives.

More information, and the basis for a new interpretation of the -u itself, can be gathered by examining the 244 verb list statistically. Taking the occurrence or non-occurrence of -u on a Nubi verb as the dependent variable, those factors described in 3.1.2.1 and 3.1.2.2 will be defined as independent variables. The factor list is as follows. Raw frequencies for each category can be found in the appendix.

(21) Variables in analysis, dependent variable = -u verbs in Nubi

- Imperative or not (3.1.2.1)
- Transitive or intransitive (as per discussion above, this section)
- Perfect verb or not (see (16) above)
- Verbal source (3.1.2.2)
- Vowel-final in Arabic

24. Certainly in quantitative terms the mpl source is favored, and I believe it is the stronger source, inter alia for reasons discussed in 3.5.1 below. However, the statistics alone do not allow for a decision in this respect.
The last two variables require comment. As noted in 3.1.2.2, (17), some Nubi verbs have a non-verbal source. Since the \(-u\) is a verbal category, whether interpreted as plural marker or bound object marker, it is expected that a +verbal source in Arabic would correlate with final \(-u\). ‘Verbal source’ therefore codes whether the source in Arabic is verbal or not.

‘Vowel final in Arabic’ is another variable which codes the Arabic source language for a given verb. Looking at Arabic, most verbs which end in \(-V\) do not appear with a final \(-u\) in their Nubi reflex, though grammatically an \(-u\) does occur in their paradigms, i.e. at any position an inflectional \(-u\) occurs in Table 2a, these verbs also have an \(-u\). The following two verbs maintain their original Arabic vowel in Nubi.

\((22)\quad aba < aba \quad 'he refused'\)
\(asiki 'accuse' < aški 'complain!'\)

In theory, these verbs could have surfaced in Nubi in a final \(-u\) form, deriving from a plural \(-u\) source, e.g \(ašk-u \quad 'complain-mpl'\). There are in fact three original \(-V\) final verbs which in Nubi have been taken over in the \(-u\) form, so this possibility is realized.

\((23)\quad stenu < stann-u \quad 'wait-mpl' = stanna + u\)
\(abinu < abin-u \quad 'build-mpl' = abini + u\)
\(abidu < abd-u \quad 'begin-mpl' = abda + u\)

The statistical test simply asks what the origin of the final \(-u\) is, to the extent that this can be tested from contemporary data. Note that whereas transitivity can be included as an independent variable, since contemporary Nubi verbs are transitive or intransitive independently of whether or not they have a final \(-u\), one variable that cannot be tested is whether the final \(-u\) can be related to the MPL suffix, since contemporary Nubi has no MPL morphemic category.

Given that there are multiple independent variables, a method similar to multivariate analysis is used. The method applied is a conditional inference tree, which uses a binary recursive splitting algorithm in order to find maximally homogeneous subgroups of data.\(^{25}\) A conditional inference tree looks at all variables simultaneously (rather than binarily), ranks them according to greatest degree of influence on the dependent variable, and calculates significance values for each variable. The results are represented on a hierarchical graph, with the strongest determining variables represented at the highest nodes. The actual token counts on which the statistics are based are represented in bar graphs at the bottom of the chart.

\(^{25}\) I am greatly indebted to Robin Dodsworth, who consulted on and carried out the statistical analysis.
In Figure 1, the strongest factor is represented in the highest node, i.e. the presence of \(-u\), the dependent variable, is most influenced by whether the Arabic source is imperative or non-imperative. A look at the bar graphs shows that \(-u\) final occurs overwhelmingly on verbs with an imperative origin in Arabic. The lower nodes represent significant effects that operate within either the value ‘imperative’ or ‘non-imperative’ Arabic source. For instance, on the non-imperative side, unsurprisingly, Arabic source verbs with a final vowel (*aba ‘refuse’) are most likely to surface in Nubi as vowel final.\(^{26}\) On the imperative side, a subgroup is again formed by this parameter, and the same pattern emerges, with original vowel final verbs tending not to take a final \(-u\). Among verbs which end in a consonant (non-vowel-final), and only in this group, transitivity is a significant factor, with transitives favoring \(-u\).

What the tree indicates overall is that the observed, contemporary final \(-u\) on Nubi verbs is most strongly due to an origin from an Arabic imperative, with the Arabic imperative source favoring \(-u\). The next strongest factor is whether the original Arabic verb ends in a vowel or not, while the weakest of the significant factors,\(^ {27}\) transitivity, is operative only within a single subgroup of the other significant factors.

Two points emerge from this statistical examination of the data. First, it confirms that the Arabic imperative source is the most important one in the

\(^{26}\) That is, non-imperative Arabic source verbs which end in a consonant are more likely (\(p = .048\)) to attach a final \(-u\) than those that end in a vowel (which in fact are categorically without \(-u\)).

\(^{27}\) Whether or not the Arabic source is a perfect verb, and whether or not it has a verbal source are non-significant variables.
restructuring into the Nubi verb. Secondly, it suggests that transitivity is not an overall applicative variable, and so by inference the statistics do not support the strong claim that the final -u derives from the 3msg object suffix. Neither does it falsify my claim that the -u could have been originally the mpl -u, though as noted, this factor is not testable in the present statistical environment.

Given these considerations, I would provisionally suggest that while the final Nubi -u is to be regarded as having a morphemic source, ancestral Nubi perhaps did not distinguish categorically between a -u on a verb which represented mpl (see paradigm in Table 2a), and a -u which represented an object suffix, i.e. that they generalized -u₁ = plural subject and -u₂ = object simply to -u, ‘marker of verbalness’.²⁸ What this -u developed into in Nubi can be termed a ‘category indicator’. In this interpretation, original morphemic material was demorphologized, in that its original values, mpl and 3msg object marker were lost as discrete morphemes, but the ‘new’ -u still functioned to set off the majority of all verbs. The category indicator has no morphemic value in a traditional sense — it does not represent a discrete semantic value — but it does serve to mark more than half of all Nubi verbs and thus indicates a grammatical category. By the same token, the Arabic imperative verb (with canonical shape a-CCVC) was equally refunctionalized into a category indicator, which also became a mark of the category ‘verb’.²⁹ Somewhat paradoxically, the Nubi verb thus consists of statistically relevant morphemic elements, which are not morphemes.

3.1.3 Other frozen morphemes
To round off the discussion of this section, further significant frozen morphemes will be mentioned.

*Mata*, ‘negative imperative’. The negative imperative is *mata*.

(24) mata ášurubu
    don’t drink
    ‘Don’t drink.’

(25) mata-kum ášurubu
    don’t-pl drink
    ‘Don’t ye drink.’

²⁸. As mentioned above, the statistical analysis does not test a mpl source of -u.

²⁹. There is another significant statistical correlation worth pursuing, and that is the interaction between original vowel final verbs (like aba) and the lack of a final -u in Nubi, an issue that will not be gone into here for lack of space.
This derives from the Sudanic Arabic negative *ma* ta-.  

(26) *ma* ta-šrab
   not you-drink
   ‘Don’t drink.’

Nubi has taken the negative *ma*\(^{31}\), the *t*- of the second person and the Sudanic Arabic /a/ of the preformative vowel and melded them into a single lexeme, *mata* ‘negative imperative’.

‘Have’. Nubi has an inflected verb, *endi* or *endisi* ‘have’. This comes from the Arabic expression of possession, the preposition *ʕand* or *šind*- + object of preposition, the object representing the possessor, i.e. the Arabic construction = ‘at-X, *šind-i*, ‘at-me = mine’, *šind-ak* ‘at-you.m = yours’, etc. The Nubi verb derives from the 1sg expression of possession, *šind-i*. In Uganda Nubi (not Kenyan), an alternative form is *endisi* which derives from the Cairene negative *šind-ii-š* ‘I don’t have’ (as noted, often *š > s* in Ugandan Nubi). Nubi *endi, endisi* is itself another example of significant morphological reinterpretation: in Arabic *šind-* is morphologically a preposition. In Nubi *endi(si)* is an inflectable verb:

(27) *ana* bi-endi/endisi
   I    ft-have
   ‘I will have.’

### 3.2 Arabic morphemes functionalized into corresponding Nubi morphemes

The previous section outlined the main ways in which original Arabic morphemic material was historically demorphemicized and sometimes refunctionalized, while still leaving its trace as a category indicator. In this section I look at how Arabic morphemic material is maintained in Nubi, either in a one-to-one relation with Arabic, or in a refunctionalized capacity.

#### 3.2.1 Nominal plural

Both Arabic and Nubi have nominal plurals among nouns, adjectives, and demonstratives. For nouns and adjectives, Wellens (2005:75–7) lists no less than three types (suffixation, suppletion, ablaut), each with sub-classes, for instance six different suffix classes.

The most common is the addition of a stressed -á to a singular stem.

---

30. In Sudanic Arabic the negative morpheme used in the negative imperative may, depending on variety, be *ma*, *la*, or *ya*.

31. Which is also the normal Nubi negative, *uwo ma ašrubu* ‘he didn’t drink.’
This derives by a semi-regular process from the Arabic feminine plural suffix -áat (e.g. šadar-aat ‘trees’) as follows.

Formation of Nubi plural suffix -á.
-át input
-át regular vowel shortening
-á loss of final stop (common, though not wholly regular process in Nubi historical phonology)

Stress shift has generalized in Nubi so that shift alone, without the addition of -á, can also mark plurality.

Demonstratives also distinguish number, as in Arabic.

‘this’ ‘these’

There can occur, moreover, plural agreement between noun and adjective, either attributively or predicatively, both items showing marked inflectional plurals.

Certainly the Nubi plural system is reduced vis a vis Arabic in two important ways. First, it has a far smaller set of plural markers than does Arabic. Nigerian Arabic, for instance, has something in the range of 15 different plural classes, counting both internal and suffix plurals. Secondly, in Nubi, plural marking is not regular. It occurs more frequently with humans and animates than with non-animates, and may occur on the noun alone, but not on the adjective (Wellens 2005:78). This contrasts with Sudanic and Egyptian Arabic, where plurality is always marked wherever more than one (or two) discrete objects are referenced. Still, Nubi does have a functioning system of morphemic plurality, which derives transparently from Arabic whose basic morphological and syntactic properties it maintains.

Relative clause clitic

A second one-to-one correspondence is that of the Nubi relative clause clitic al-, corresponding to the SA relativizer al-. This is discussed and exemplified in 3.3.3, since it is in free variation with a morpheme which was refunctionalized into the same relative clause clitic function.
3.3 Arabic morphemes refunctionalized into ‘new’ Nubi morphemes

A number of Nubi morphemes have a transparent origin in an Arabic morpheme, but in Nubi the morphemic value has been altered to one degree or another.

3.3.1 The second person plural

It was argued above that the 2pl has a special place in defining the outcome of Nubi verb morphology. The 2mpl appears in another originally bound morpheme, this time refunctionalized, but maintaining the 2mpl value. This is the Arabic morpheme -kum, beet-kum ‘your mpl house’, šuf-tú-kum ‘I saw you mpl’. In EA -kum is a common plural, in KA it is 2mpl, opposed to 2fpl -kan. In Nubi -kum has four distributions, all with a common meaning of ‘2pl’.

(33) Distribution of -kum in Nubi
pluralizer in independent pronoun: ita ‘you sg’, ita-kum ‘you pl’
Suffix added to imperative verb: gum ‘get up’, gum-kum ‘get up ye’
Suffix added to negative imperative verb: mata ruwa ‘don’t go’, mata-kum ruwa ‘don’t ye go’
Suffix added to possessive marker ta-, ta-kum ‘yours pl’

By way of comparison, a representative listing of comparable Arabic forms is as follows.

(34) 2pl morphemes in Arabic
a. inta/itta ‘you m sg’ vs. intum/intu ‘you m pl’
b. gum ‘get up’, gumm-u ‘get up mpl’
c. la/ma/ya taguum ‘don’t sg get up, la/ma/ya taguum-u don’t ye get up

d. bitaa-kum ‘of yours mpl’

As can be seen, only (34d) shares the -kum form with Nubi. Nubi has thus maintained the basic 2pl meaning from Arabic, but given it a different, previously unknown distribution.

All in all, a conspiracy of original second person forms can be identified in Nubi, some refunctionalized from original morphemic material and distributed in new Nubi lexical material, one morpheme maintaining its meaning from Arabic, but with a radically new distribution. The list is as follows.

(35) The second person conspiracy
imperative verb > category indicator (Section 3.1.2.3)
second person imperfect prefix ta- (frozen) (3.1.3)
2mpl -kum possessor/object suffix, refunctionalized in various ways as 2pl pronoun (this section)
-u partly ambiguous as < mpl 2nd person imperative suffix (3.1.2.3, category indicator)
3.3.2 Tense, mode, aspect markers

Verb affixes bi-, gi-
Nubi has two verb prefixes, bi- marking ‘future’ and gi- ‘progressive.\(^{32}\) Bi- and gi- occur in combination, alone with the verb, or not at all. The bare stem gives a past meaning with action verbs, a durative meaning with statives. The structure of the Nubi verb is thus.

\[(36) \text{bi-gi-verb stem}\]

Exemplifying this with all possible combinations:

\[(37) \text{Nubi verb paradigm}\]

\begin{align*}
\text{uwo amsuku} & \text{ ‘He held.’} \\
\text{uwo bi-amsuku} & \text{ ‘He will hold.’} \\
\text{uwo gi-amsuku} & \text{ ‘He is holding.’} \\
\text{uwo bi-gi-amsuku} & \text{ ‘He will be holding.’}
\end{align*}

\text{Bi- transparently comes from the Sudanic/Egyptian bound indicative prefix bi-}.

\[(38) \text{bi-msuk-u} \text{ ‘They hold, are holding.’ (SA)}
\text{bi-timsik-u} \text{ ‘You.pl hold, are holding.’ (EA)}
\]

That bi- is an indicative prefix is shown by the fact that there are contexts (differing to a degree among the EA and SA varieties) where bi- cannot occur, e.g. in the negative imperative (see 34c above).

\[(39) \text{la ta-msuku} \\
\text{not you-hold} \\
\text{‘Don’t.2mpl hold’ (cf. *la bi-tamsuku)}
\]

\text{Bi- is a grammaticized, indicative prefix in these varieties, not a future as in Nubi, though there are hints in some dialects of the Sudan that bi- does mark a future meaning (Owens 1991b). It can thus be regarded as a refunctionalized verbal prefix, having shifted from indicative to future.}

\[^{32}\text{Kihm (2011: 48) does not consider these prefixes, though his only implicit argument against this classification is that he formally sees them as members of AUX, which by definition presumably must be independent syntactic elements. Morphologically, both bi- and gi- are bound forms in that they never occur as self-standing elements and only immediately pre-verb. They may undergo front-back vowel harmonic alterations, bu-ruwa ‘will go’ vs. bi-biyo ‘will sell’ and they can lose their vowel before a V-initial stem, b-arija ‘will return’. If they co-occur their order is always bi-gi, not *gi-bi-.}\]
Gi- is generally regarded as deriving from the participle gaasid, which in Sudanese and Chadian Arabic serves as a progressive con-predicate indicating that an action is in progress at the time of speaking/narrative (Trimingham 1959:71; Miller & Mustapha 1986; Roth 1979:56).

(40) gaasid b-i-msuk-u
    stay INDC-PV-hold-MPL
    ‘They are holding.’ (right now)

It generally is not used in Nigerian Arabic and in Egyptian Arabic the counterpart qaasid exists, though with a more constrained and with a far lower frequency of usage than in Sudanic Arabic.

Assuming the derivation from gaasid, Nubi maintains a one-to-one meaning with the Arabic source, in that in Nubi it only signals an action occurring at the time of speaking (or narrative time). With a future time adverbial, for instance, only bi- is possible:

(41) uwo bi-amsuku lamgaba sabá (*gi-amsuku)
    he FT-catch animal tomorrow’
    ‘He will hunt tomorrow.’

Given these derivations, it is remarkable that the sequence in Nubi is the opposite from that in the Arabic, that an originally independent word interpolated (*)gaasid or *qaasid) itself in front of a bound morpheme (b-).

3.3.3 Others
To round out the refunctionalized Arabic morphology the remaining morphemes may be listed. They will not figure in the further discussion.

Arabic participles of derived verbs prefix mi- or mu-, kammal ‘he finished’, mu-kammal ‘finished’ (passive participle). In addition, on an irregular basis certain nouns (place) prefix ma- ma-ktab ‘library’. This however, is in most dialects, including the Sudanic dialects, largely an unproductive, lexicalized element.

Nubi has the prefix ma-, which when added to the verb stem produces an adjectival passive participle. Wellens (2005:182) has noted 34 such forms.

(42) seretu ‘tear’ ma-seretu ‘torn’

Nubi has two relative clause clitics which are attached at the beginning of the relative clause, al- and abu-, each with various allo-forms.

33. Cf. Bongor pidgin Arabic in Chad, with the multi-functional particle gayi < gaayid < gaazid, probably reflecting a pre-grammaticized stage (Luffin 2007:636).
Jonathan Owens

(43) kalaam al-ita aju nade
thing which-you want that
‘That thing that you want.’

(44) ajol ab-endis sana milan
person who-has year many
‘A person who is old.’
(Wellens 2005: 128–9)

Al- is simply the SA relative clause marker al-. Here is a second case of one-to-one Arabic-Nubi morphological correspondence (3.2.2), though it should be noted that unlike SA al-, Nubi al- does not assimilate to a following dental or alveolar consonant. Abu, ab, etc. is another case of an interesting refunctionalization. Abu in SA and other varieties of Arabic is used in two ways. (1) It serves as a compound builder for various closed lexical classes, like names of animals and diseases, as in NA, abu naaṛ ‘firefly’, abu tassaas ‘heartburn’, and (2), as an attributive marker meaning ‘having, possessing’, naadum abu bagaran katiiraat ‘a person having many cattle’. Nubi does maintain a few of the ‘abu’ nominal compounds, e.g. ‘ab darag ‘cobra’,34 but most likely, the Nubi relative abu derives from the second function.

3.4 From phonological value in Arabic to morphemic in Nubi: The refunctionalization of stress

One of the striking aspects of Nubi is its use of morphemic stress and in one case (perhaps), tone. Four distinctive morphemes have been distinguished.

(45) Nubi morphemic stress and tone
Stem 1. Initial stress = imperative, past, ášurubu ‘drink!’
Stem 2. Final stress = unspecified/impersonal agent ašurubú ‘drunk’, termed ‘passive’ by Wellens
Stem 3. Penultimate stress = verbal noun (termed ‘gerund’ by Wellens):
ašurúbu ‘drinking’
Stem 4. HHL# = infinitive, ašúrúbu ‘drinking’35

34. In Sudanese Arabic, abu darag (Roth-Laly 1969: 17, citing Hillelson). This is a curious example for stability in language. In Classical Arabic daraq, SG. daraqa is a type of shield made of leather (Ibn Manēṣur, Līsaan 10:95, thirteenth century dictionary). One assumes the original designation pertains to the cobra’s hood, that looks like a shield.

35. Gussenhoven (2006) reduces the basic stress (or accent types in his framework) types to stems 1, 2 and 3, deriving stem 4 from stem 3 by a phrasal deaccenting rule. This is an interesting perspective which I tend to agree with, and could dovetail with the suprasegmental dynamics of inalienably possessed nouns, whose deaccentuation pattern was noted in Heine (1982:42).
Without going into all details, the complete set of forms requires trisyllabic verbs, of which there are many. Bisyllabic verbs distinguish 1, 2, and 4.

(46) a. *kútu* ‘put, putting’
    b. *kutú* ‘put’ (unspecified subject)
    c. *kútú* ‘putting’ (infinitive)

In this case, (46a) could be either stem 1 or 3.

Whereas most verbs have up to four suprasegmental forms, according to syllable shape, there is a small class containing some high frequency verbs which have a single suprasegmental shape. These are discussed further in 3.5.2 below.

Stems 1 and 2 are limited to predicative position. Stem 2 is designated as an impersonal subject, rather than a passive, even if in most cases it will have a passive meaning (and hence is termed ‘passive’ by Wellens).

(47) *ruwá* ‘someone went, they went’
    *ašurubú úwo* 
    drunk it
    ‘It was drunk.’

    *bi-dakal-ú sa yatu*
    *fτ-enter hour which*
    ‘At what time will be entered.’ (‘will there be entered’ in Wellens’ translation, 2005: 387)

Stem 3 is a verbal noun, with the distribution of nouns.

(48) *ašurúbu meríse katír batál*
    drinking alcohol much bad
    ‘Drinking a lot of alcohol is bad.’

In one construction, even while maintaining a position as object of the preposition *fί*, this stem is used as a present progressive when it occurs as complement to the existential morpheme *fί* ‘there is’.

(49) *uwo fί  fί ašurúbu meríse*
    he exist at drinking alcohol
    ‘He is (currently) drinking alcohol.’

Gussenhoven’s article also has interesting data on monosyllable stem 2, which appears to show a length component as well (2006: 199). He does incorrectly limit stress to the last three syllables of a stem (2006: 197), whereas Stem 1 simply has initial stress, which can be up to four syllables from the end of a word.
As Wellens explains (188–193), the verbal noun has the distribution and takes the modifiers of any noun. The infinitive occurs in complement position of certain verbs and prepositions, though also may overlap in distribution with the verbal noun. Both, for instance, occur as objects of prepositions.

(50) \textit{marya de fi kurúju to (verbal noun)}

woman the at digging her
‘The woman is busy working at the field.’ (2005: 192)

(51) \textit{ita gum fi kúrúju kurúju ta sabún}

you got up at dig.inf dig.vn of soap
‘You got up for tilling (the field) the working for soap’ (i.e. to get money to buy soap).

From this description it can be noted that Nubi distinguishes finite from non-finite verbs formally: stems 1 and 2 are finite, only occurring in predicative position, whereas stems 3 and 4 are non-finite.

3.5 Origin of suprasegmental morphemes

3.5.1 Stress

To this point the origin of Nubi morphemic material is straightforward since it corresponds closely to an analogous form in Arabic. This is not the case with the supra-segmental morphemes. In Arabic, stress is not morphemic, so with these four (or three, see n. 35) morphemes there is a case of non-morphemic material being converted into morphemic in Nubi. Developing Owens (2001: 362–4 and 2008), it may be suggested that the main impetus for the development of a stress morpheme was the recognition among ancestral Nubi that Arabic has frequent shifts of stress. In most cases these are purely phonologically determined. In one interesting case morphemic considerations also come into play, as discussed below.

(52) illustrates the basic Arabic stress determinants. The following four forms in Arabic are all accountable by the same phonological stress rule, for instance, namely that the first heavy syllable from the end be stressed (heavy syllable = VCC or VVC(C))

The following examples use the imperative forms of the verbs \textit{ašrub} ‘drink’, and \textit{faṣṣul} ‘separate’, the (super-)heavy syllables being marked in bold.

(52) Variable Arabic stress placement

a. stress first heavy syllable from end, if there is one

\textit{ášrub/fáṣṣul} ‘drink/separate’
ášrub-u/fáṣṣul-u:36 `drink-mpl'/separate-mpl', `drink-ye'
b. ašrub-ha/faṣṣul-ha: stress first heavy syllable from end, if there is one, `drink-it.f/separate-it.f'
c. ašrub-úu-ha/faṣṣul-úu-ha: stress first heavy syllable from end, if there is one `drink-mpl-it.f', `drink-ye it/separate-ye it'
d. ašrub-ú/faṣṣul-ú: see discussion below `drink-mpl-it.m', `drink-ye it/ separate-ye it', i.e. here the final -ú realizes a complex portmanteau that is abbreviated as mpl.3msg, i.e. masculine plural -u + a 3msg object suffix realized as the stress shift to -ú.

As can be seen, the stress shifts purely as a function of syllable weight changes. (52a–d) are the 'same' verb, but have different stress patterns. Only (52d) verges on morphemicity in Arabic. This may be explicable on historical grounds. The final stressed 3msg object suffix -u can be analyzed as realizing an underlying -hu. In Cairene, for instance, there is an alternation between:

(53) išrāb-u
drink-pl

and

(54) la išrab-u-húu-š
don't drink-pl-it.m-NEG
`Don't drink it.'

In (54) the allomorph -hu, lengthened regularly to -huu, of the 3msg object suffix does appear when it occurs before a negative (or any further) suffix. Its underlying existence may be postulated on the basis of its appearance before the negative. Given this, one could conceivably postulate for (52d) an underlying form such as ašrub-úu-hu, with some sort of contraction of -úu-hu to ú. In any case, there is at most one problematic counterexample to the observation that stress shifts in Arabic are phonological, not morphological as in Nubi. Still, this interpretive problem arises only from a synchronic perspective, since (52d) ašrub-ú does not have a surface heavy syllable. From a Nubi historical perspective, it will be argued below that the unusual final stress in fact served as a partial morphemic clue in the development of Stem 2 (45).

In addition to these phonologically-determined alternations, in dialects of southern NE Nigeria, Cameroon, and Chad, CVCV(C) forms, which in the other areas are stressed on the first syllable, are stressed finally. Thus ancestral Nubi could have been confronted with lexical stress variation of the type:

---

36. EA would stress the penultimate syllable here, faṣṣulu (Hellmuth 2013 for overview of stress in Arabic). This is another feature where Nubi derives from SA rather than EA.
Stress shifts are thus central to Arabic phonology. The stress shifts also are not infrequent. To ascertain this a very basic background tracking count was made of IM107SHRIF in NA set 2, a conversational text of about 1,550 words. All multisyllabic verbs were marked as one of three classes. The first can be termed the basic stem. The basic stem consists of bound person, number, and gender subject markers, and nothing else. The paradigm in Table 2a above is an illustration of what these subject markers look like in the basic stem in the imperfect. Note in the paradigm that the stress always falls on the preformative vowel. Any verb minimally must appear in basic stem form (excluding the optional b- indicative prefix). A stress-shifted stem is the basic stem plus added gender or object suffixes which force a stress change in the basic stem. The following two verbs, for instance, appear in the text.

(56)  
\[
\begin{array}{ll}
\text{gáal-at} & \text{said-3fsg} \\
\text{‘She said.’}
\end{array}
\]

(57)  
\[
\begin{array}{ll}
\text{gaal-óo-ha} & \text{said-3mpl-3fsg.obj} \\
\text{‘They.m said it.f.’}
\end{array}
\]

(56) has the stress of a basic stem. (57) adds the 3mpl subject morpheme -o, which in and of itself would not effect a change in stress (gáalo ‘they said’), but the addition of the 3fsg object suffix -ha requires that the vowel -o be lengthened, and the lengthened vowel draws stress, as a heavy syllable. Any stem whose stress has been shifted by the addition of affixes will be termed a ‘stress-shifted’ stem. (57) is a stress-shifted stem. Finally, a separate class was made which consists simply of a basic stem, but those basic stems which in the text are paired with another token of the same stem, but in stress-shifted form. For instance, the following two verb tokens occur in the text, both the perfect form of the verb meaning ‘repair’.

(58)  
\[
\begin{array}{ll}
a. & \text{wáasa l-baab} \\
& \text{repaired def-door} \\
& \text{‘He repaired the door.’}
\end{array}
\]

\[
\begin{array}{ll}
b. & \text{waasá} \\
& \text{‘He repaired it.’}
\end{array}
\]

(58b) is a stress-shifted stem, its stress shifted by the addition of the 3msg object suffix (see discussion around (54) above). (58a) is the ‘same’ stem, but in basic form, i.e. with no stress-changing suffix added. Similarly (56) is included among
the basic stems which in the text are paired with a token which has the stress-shifted stem form (57). The statistics in (59) show the three classes.

(59) Basic and stress-shifted stem count, NA texts, set 2
   a. Basic stem: 82, 50%
   b. Basic stem with paired stress-shifted stem: 41, 25%
      (total basic stems = 123, 75%)
   c. Stress-shifted stem: 41, 25%

There are 164 multisyllabic verb tokens (see (20) above, which used the same data set). The basic stem is the dominant one with 75% of all stems, but there are an appreciable number of stress-shifted stems as well = (59c) 25% of all tokens. Class (59b) has been distinguished to give an idea about how often in a given span of discourse (about 1,550 words in this case), a listener might be exposed to the same stem in two stress forms, basic stem, plus stress-shifted stem, examples like (58a) and (58b). (59b) counts how many tokens a basic stem verb like wáasa is paired at least once with an ‘identical’ stress-shifted stem like waasá. That is, speakers would have to have been able to identify the ‘same’ stem differing only in stress in 25% of the verbs. The overall point of these statistics is that the proposed explanation for the origin of Nubi suprasegmental morphemes passes the realism test — it may be assumed that stress shifts on the same lexeme were quite common in the Arabic which ancestral Nubi heard. Realistically, ‘common’ can be given a precise temporal face. The sample used here covers nine minutes and thirty seconds. Clearly it may be supposed that ancestral Nubi would have been exposed to thousands upon thousands of minutes of Arabic or Arabic-like discourse, giving them ample opportunity to extract regularities of stress shift. The role of frequency and stress is taken up further below.

A fourth argument for the morphologization of stress comes from plural marking. This pertains to the dominant plural suffix -á, which as seen in (29) derives from the Arabic feminine plural *-áat. In many nouns, the only difference between singular and plural is stress:

(60) gidída ‘chicken’ gidiidá ‘chickens’ (Wellens 2005:75)

Recalling 3.2.1, in Arabic the stress on the FPL nominal suffix -áat is phonologically determined, since áat forms a heavy syllable. In Nubi on the other hand the stress on -á is a part of the morphemic identity of -á, so here is an incontrovertible instance of stress morphologization. Whether the plural -á could have been a model for interpreting a morphemic value to verbal stress is a question that will

37. For the purposes of the brief illustration given here, it should be noted that four of the stress-shifted stem tokens are lexical variation of the type illustrated in (55).
not be addressed here. What the example does show is that the morphologization of phonological stress proceeded along a broad front in Nubi.

Four arguments therefore emerge favoring the idea that the Nubi suprasegmental morphemes derive from originally phonological alternations. First, the idea is plausible in that the different stress morphemes in Nubi correspond to different phonologically-determined stress positions in Arabic. Secondly, it can be assumed on the basis of a tracking comparison, that such shifts would have been frequent in ancestral Arabo-Nubi discourse. There would have been ample stimuli for ancestral Nubi to make new sense of the different stress positions by assigning them a fixed meaning. Thirdly, note that any stress-shifted stem adds morphemic material, in (57) and (58b), for instance, an object pronoun is added. It can be suggested that even if the pronouns themselves were not ‘correctly’ parsed, speakers could have been aware that the stress shifts were being made for a morphemic reason, the addition of a further morpheme. Fourthly, stress was morphologized in the Nubi plural marking.

A fifth perspective can be added as well, and this is to observe that all of the southern Sudanese substrate languages are tonal to one degree or another, in one way or another.38 While no southern Sudanese language has been found with shifts comparable to what is found in Nubi — as already shown above, the model is most likely the Arabic phonological stress shift — the idea of using suprasegmentals as morphemic distinctions can be suggested to have come from, or at least been strongly reinforced by southern Sudanese languages. Without explicit morphological cognates, this suggestion remains speculative, but plausible.

Assuming the correctness of this explanation, and anticipating the discussion in 3.5.2, it needs to be noted that Nubi itself has neither geminate consonants nor long vowels, and so has no concept, as it were, of a heavy syllable. What is argued for here is that the ancestral Nubi perception of the stress shifts would have been enough to motivate the morphologization of Arabic stress.

Before moving to the invariable verbal stems, it is necessary to compare the current account with the only other detailed alternative, a very different one of Wellens (2005: 353–5) that concentrates on the Nubi impersonal (= passive, (45b)) alone. Note that her account has been taken over uncritically (Kihm 2011: 69). In Wellens the development of the impersonal proceeds as follows, beginning with the hypothetical (62a).

38. For example, Zande, spoken in the SW quadrant of the southern Sudan, has lexical tone, zirá ‘leprosy’, zire ‘bird’, ziré ‘a lie’, as well as grammatical tone, m-a-gúmbá ‘I shall speak’ vs. mí-a-gúmbá ‘I spoke’ (low tone unmarked, Tucker & Bryan 1966: 143). Luanyjang Dinka, spoken in the NW quadrant of the southern Sudan, has a complicated tonal and allotonic system, with four distinctive tones, e.g. kóoot (HL) ‘acacia tree’ kóoot (LH) ‘care forːPST.PASS’ (Remijse & Ladd 2008: 174).
(61) a noun was left-topicalized, leaving behind a resumptive pronoun in its 
extraction site:

(62) a. *nyerekú de kan dúrubu uo kala
child the was hit him finished
‘...the child when [they] had already shot it/when it had already
been shot.’

The pronoun then merged with the root

(62) b. *nyerekú de kan dúrub(ú) kala
... ‘when the child had already been shot’

At this point, the former pronoun -u(o) becomes part of the stem, and the con-
struction is no longer a left topicalized one. Implicitly, though Wellens does not
describe how or why this came about, the initial stress on the verb was dropped.

(62) c. *nyerekú de kan durub(ú) kala
Finally the subject noun was returned from topic position to its post-verbal posi-
tion, giving the normal Nubi impersonal form:

(62) d. durubú nyerekú

This account has multiple problems, both general and specific ones. First, the only
stage which conforms to what we know of present-day Nubi is the last, (62d). The
other three stages have no comparative or textual support. They are there simply
to enable a convoluted derivation to work. Looking at the specific stages, (62a)
is argued to derive from a topicalization construction, which indeed does occur
in Nubi. As in Arabic, a noun is left-fronted to topic position, and a resumptive
pronoun (uo) marks its extraction site. However, both Arabic and Nubi expect an
independent expression of subject — in (62a) nyerekú is the object — which does
not occur.

Step 2, (62b) is equally unsupported. In general in Nubi object pronouns do
not cliticize to a verb, unless the verb is one which carries exceptional final stress
(e.g. ligó in (67) below). Wellens certainly would not be thinking of an Arabic
bound object pronoun, since as seen in 3.1.1, these failed to morphologize in Nubi.
In Arabic the 3MSG object suffix is never specified to bear independent stress in
any variety.39 The resulting form dúrubú is problematic, since the Nubi object pro-
noun is uo (or uwo), and there is no general process in Nubi deleting a final /o/.

39. The only variety of Arabic which could produce a final stressed -u in a surface realization
is North African Arabic, which is of no interest as a contributing dialect to Nubi. In this dia-
lect there is a strong tendency for a final utterance (not lexical) pattern with rising intonation.
Should an -u stand in utterance-final position, it would bear a rising intonation pattern.
Step 2 produces an intervening HLH tone pattern, which is otherwise unattested in Nubi, or Arabic, so as before, this form fails the realism test. Step 3, (62c) therefore arbitrarily lowers the tone or destresses the first syllable. The need to postulate a destressing process arises simply by the previous creation of the non-occurring HLH pattern.

New problems are added at step 4, (62d). The first is that impersonals, or passives, like (62d) are extraordinarily rare among the world’s language. Keenan (1985:251) a number of years ago noted the following universal, ‘… we know of no language which forms passives by merely reduplicating or changing tonal marking on transitive verbs’, i.e. the derived subject stays in situ, and no auxiliaries or segmental inflections are added. To my knowledge, Nubi, (and Juba Arabic) are the only languages in the world which have this property. Step (62d) thus asks that ancestral Nubi create a unique, new structure, by moving an argument out of a canonical subject position (Nubi is otherwise SVO) and creating this new hybrid. Secondly, it can be noted that Juba Arabic also has an impersonal passive (Watson & Biajo Ola 1986:37), evidence again for the close historical affinity of these two varieties.

Beyond avoiding the unsupported convolution of this account, I would adduce two further pieces of evidence for my own explanation. First, Wellens requires a final -u in her initial account, and because she follows Versteegh’s explanation for the origin of -u as a 3msg object marker, she needs to add an extra deus ex machina, an otherwise unattested stress on this suffix. In my alternative explanation (discussed in 3.1.2.3) the final -u is a plural subject marker, and as noted above (compare (52a) with (52c)), this suffix automatically becomes stressed whenever an object suffix is added after it. Secondly, Arabic so consistently uses a 3mpl form ending in -u as a passive, that in a study of Emirati Arabic which is introduced for comparative purposes in 4.1 below, we recognized a special category of ‘plural passive’ (termed ‘new referent’ in Table 3 below) among our null subject types (see Table 3, Owens et al. 2013). (63) is an example.

(63)  
\[ i-samm-úun-a dibs \]
\[ \text{3-call-mpl-it molasses} \]
‘They call it molasses = it is called molasses.’

(63) is taken from an Emirati Arabic text. Corresponding to this form in SA would be:

(64)  
\[ bi-samm-ù qalla \]
\[ \text{IND-call-mpl.it.m grain} \]
‘They call it grain = ‘it is called grain.’
(64) of course gives exactly the final stress one needs to understand the Nubi impersonal, and it does so by providing the morphological shadow form, the final -u which is so common in Nubi (3.1.2.3). It can be added here that background tracking data from set 2, the same set as cited in (20, 59) above, gives no less than 29 tokens (out of 164 verbs, 18%) in which a final -ú realizes the portmanteau morpheme, -mpl.3msg.

It needs to be emphasized that only a mpl -u regularly draws stress; the 3MSG object suffix -u, as noted above, by itself never does.40

In the case of bitransitive verbs, (64) would even provide a post-verbal slot for a second object.

(65) bi-samm-ú dagiig

IND-call-mpl.3msg flour

‘They call it ‘flour’ = It is called ‘flour.’

Whereas Wellens’ account essentially consists of a series of steps whose only justification is the need to provide an input into the final form, in the current account no new structure at all is needed that cannot be plausibly assumed to have been part of the ancestral Nubi discourse environment.

Admittedly, one final part of the puzzle is left unresolved. That is to account for the post-verbal position of the subject argument, when the verb is not bitransitive as in (65). Ancestral Nubi would have heard either (66a) or (66b), but not (66c), which would give the correct Nubi word order.

(66) a. ḏarab-ú

hit-mpl.him

’They hit him = He was hit.’

b. ḏárab-u al-walad

hit-3mpl def-child

’They hit the child = the child was hit.’

c. * ḏarab-ú al-walad

hit-3mpl def-child

40. Moreover, it is not only a mpl subject suffix which draws stress when a 3MSG object suffix is suffixed to a verb. Any verb stem that ends in a vowel does so. The same process which explains a model for the Nubi impersonal stem 2 from -mpl.3msg object -ú also explains the final stress on Nubi nadi ‘call’ and wedi ‘send’. Wedi, for instance, derives from wāddi ‘send msg’ + msg object suffix, which is realized by the same stress shift to a final vowel, waddí ‘send it.m’. In these cases ancestral Nubi generalized the complex Arabic verb forms into the new lexical verb stems. To create Stem 2 they generalized the -mpl.3msg into a new grammatical form.
Except for the probably infrequent bitransitive construction (65), there is no Arabic model analogous to (66c).41 This syntactic reinterpretation analogous to a number of instances of morphological reanalysis discussed above in this article, requires separate treatment.42

A final argument against Wellen’s explanation is that it says nothing about the Nubi verbal noun (45, Stem 3). While more needs to be said about the emergence of the verbal noun, the current account seeks a common factor in the development of the impersonal and the verbal noun, namely the perception that the shifting stress of Arabic could be interpreted morphologically.

A detailed account of how the quite odd Nubi impersonal verb came into existence thus further strengthens the argument advanced here that ultimately the different Nubi stem morphemes derive from stress morphemicization. In this particular aspect of the development of morphemicized stress in Nubi, a morphemic element does in fact come into play in the proposed derivation, since a final stressed -ú in Arabic implies not only mpl, but also a 3msg object suffix.

3.5.2 Invariable stems and proto-Nubi stress

Before moving to more general matters, it is relevant to discuss a small class of Nubi verbs whose stress is invariable. These always have non-initial stress, and are listed among (18) above.

(67) nisítu < nisíí-tu you.pl forgot
     ligó < lig-ó ‘got-they.it.m

The origin of the stress is transparent. Nisítu, for instance, derives from the 2mpl form of ‘forgot’, which has a heavy syllable /síi/ that attracts stress by the heavy syllable rule. What is interesting in these forms is that they have no distinctive verbal noun and unspecified subject forms, nisítu = ‘forget, forgotten, forgetting’.43

One possible explanation for the stress invariance of these verbs is an historical one, which may shed light on how the present system developed.

41. A reader correctly points out that this could be interpreted as ‘they hit him, the child’ with a cataphoric pronoun referencing the object. I doubt if this construction would pass the frequency test in background tracking as an obvious source for this word order.

42. I would only note here that the development probably involves the construction evolving from an in-situ post-verb object. I.e. the new passive construction preempts the post-verb object, merging it into the subject function. There is no need, as in Wellens, to posit intermediate steps via pre-verb topicalization followed by movement of the topicalized noun back to post-verb position.

43. Wellens (2005: 184) notes that they do have distinctive infinitive forms with high tone, nisítu ‘to forget’.
Stem 1, the imperative and past stem (see (45 Stem 1)), is also the stem which derives from what in (59) was termed the basic stem. Stem 1 appears to be the most frequent in actual Nubi discourse. A look at the first of Wellen’s texts, a story of about 1,530 words (2005: 381–9) bears this out. A count of verbs in this text reveals 28 impersonal Stem 2 tokens, all other verbs (ignoring verbal nouns) being Stem 1 active verbs (some 144 tokens).

As seen in Section 3.5.1, its corresponding stem in Arabic is also the most frequent stem in the text sample. Anticipating the discussion in 4.2, one imagines a point in ancestral Nubi where neither stress nor segmental stem form had yet been fixed. The word for ‘forget’ could have varied among different forms of the same stem. Evidence for this proto-variation can be found in present-day Nubi, where one still finds a few doublets, for instance ‘understand’ = *afham* < imperative stem and *fahim* < *faahim* active participle. Frequency effects would favor as a phonological source the stem which was most common in Arabic, which as seen in 3.5.1 (59) is the basic stem. In most Arabic verbs the stress on these basic forms falls in the initial syllable, which in fact is the stress on Nubi Stem 1. Thus, the Nubi Stem 1 can be interpreted as deriving its stress from the most common Arabic basic stem stress pattern. Still, exceptional non-initial stress could have developed and lexicalized in verbs of high frequency, as in (67).

At a second stage of ancestral Nubi, speakers, still in contact with Arabic, would have begun experimenting with moving stress to express grammatical differences, under the influence of alternations discussed in 3.5.1. This eventually jelled into the paradigm in (45). In words like *ligó*, however, they would have found the final stress already lexicalized on the final syllable. The emerging morphemic stress alternations would have no place to express itself on these forms, so they remained as they were originally lexicalized.

In this perspective, Nubi verbs would have developed from a single stress form, usually, but not always, initial.

4. General models for interpreting the genesis of Nubi

After the presentation of the origins of Nubi morphology, I now embed it in a general account of general creole linguistic issues. Creole studies have been marked

---

44. Verbs dependent on another verb, such as the complement of *agderi* ‘be able’ are left off the count.

45. This explanation also implies that the infinitive would have been a later development, since it does apply to verbs like *ligó* and *nisítu*. This assumption is strengthened by the observation that until today the infinitive has not been reported in Juba Arabic, so it can be viewed as an innovation in East African varieties.
by a number of interesting proposals seeking to generalize development of creoles across many individual creole languages. In many ways, these proposals are constitutive of the study of pidgin and creole languages, distinguishing creoles as a class of languages, rather than, say only grouping them within the historical purview of the lexifier languages they derive from. This perspective is entirely justified, given the fact that creoles by and large do not show developments typical of languages within a language family, or the historical changes which characterize single languages, Old to Middle to Modern French, for instance, or western Sudanic Arabic from wherever its ancestor is to be placed.

Given the rather benign neglect of Nubi within creole studies, outside of scholars who have worked on Creole Arabic itself, the applicability of the general models have either not been tested, or do not figure in the broader creole discussion. In Section 4.1 I would like therefore to play catch up.

4.1 Old models and those easily disposed of

First of all it should be clear from Section 1.3, as well as the descriptive Section 3, that one very important strand of creole studies, that which traces the influence of substratal languages on an emerging creole in terms of transfer (Siegel 2008) or relexification (Lefebvre 1999, Lumsden 1999), will be of little direct relevance. Whether one can or should interpret transfer theory as transfer from the superstrate language Arabic is a plausible perspective but not one I will examine here. Turning to theories which either have been suggested are relevant to the genesis of Nubi, or which should be, given their claims, one of the more important explanatory models in creole genesis was Bickerton’s Bioprogram (1981 etc.). As is well known, Bickerton argued that creoles, or a ‘radical’ sub-class thereof had such an impoverished learning input that speakers of the emergent creoles needed to resort to their innate language learning capacities, their bioprogram. In Owens (1980, 1985, 1990, 1991a) I argued that although the social and historical conditions for the emergence of Nubi matched those of the radical creoles (see summary in 2.1 above), Nubi lacked too many of the expected bioprogrammatic traits to make this explanation for Nubi-genesis plausible.

A second model, Mufwene’s ‘feature pool model’ (2009), sees creoles as emerging from a pool of features of potentially disparate language origin, namely the languages present in the pre-creole contact situation, possibly with influences from demographically dominant later arrivals (Mufwene 1996). Note that Mufwene himself has not suggested, to my knowledge, that this model should be applicable to Nubi, though it can be adduced here to the extent that the model is applicable to any creole language. It should equally be emphasized that echoing one strand of description in the current paper, Mufwene has consistently seen a significant role
of the lexifier at all levels of creole grammar evolution (e.g. 2009: 377, 2010: 363). At the same time, his feature pool model necessarily assumes substrate-lexifier hybridity (2009: 374, 390). Given that this model needs to find linguistic elements of diverse language origin in the creole, and given that Nubi morphology (and lexicon) is by and large of transparently Arabic origin to the exclusion of potential substrate languages, the model does little to elucidate the structural evolution of Nubi.

A third model of creole development sees a stage of pidginization as a necessary prerequisite for creole genesis (McWhorter 2002). In this context, a related discussion concerning the assumed simplification of Arabic dialects (a much argued proposition) in a foreign-talk register can be mentioned (Sharkawi 2010).

Beginning with the foreigner talk proposal, this model sees a significant simplifying input in a register shared between native and second language speakers, in which native speakers play a key role in defining the simplified variety. Sharkawi’s main purpose is to motivate an explanation for an alleged simplification of Arabic dialects in the early Islamic era, but he also relates it to the formation of Creole Arabic (2010: 239–244). Although Sharkawi gives no detailed developmental scenario describing how foreigner talk would have led to a creole, since he suggests that it did play a role, it is relevant here to point out a major, I would say fatal, problem with this model. This is that Nubi shows so much original Arabic bound morphology when one adds together the frozen traces and the refunctionalized morphology that it can be assumed that ancestral Nubi speakers were confronted with, and in one way or another absorbed, normal Arabic discourse, not a simplified variety.46 This assumption is further strengthened by the good correspondence between what morphemes did in fact emerge, and the frequencies of the cognates of these morphemes in the tracking data.

McWhorter’s pidginization model I think can be approached from two perspectives, depending on which work of McWhorter is referenced. McWhorter (2002) explains the widely-observed loss of inflections in creoles as due to a stage of prior pidginization. In pidginization, only communicatively necessary features are maintained (pp. 5, 28), while the rest, the ‘ornamental’ ones (pp. 4, 5, 6 etc.) are filtered out. McWhorter closes the logical circle by asserting that inflection is ornamental.

McWhorter here makes significant claims, whose detailed discussion is beyond the scope of this article. One, however, is immediately relevant. While the presence of a pidgin Arabic at some point in the early history of Nubi is quite plausible on socio-historical grounds, there is no direct link from this assumption

46. Similarly in some contemporary Arabic pidgins a great deal of full (if frozen) inflectional morphology is evident (Bizri 2010).
to the specific structure of Nubi. It is not clear how one would identify the Nubi pidgin ancestor. One might look for a model in a contemporary Arabic pidgin, such as Gulf Pidgin Arabic. Here the verb is selected on a combination of phonological (search for initial CV forms) and lexical criteria (use 3MSG imperfect verb) (Bakir 2010). Neither of these attributes, however, is discernible in the Nubi verb, so some reason and mechanism for a shift to the current structure, all of this in an extremely short time period, would need to be postulated. Alternatively, to say, for instance, that the early Sudanese pidgin Arabic verb was based on the complex 2mpl imperative verb form is circular against the ultimate outcome. Thus, either an early pidgin variety was replaced by a more complex, more highly morphologized variety which effaced the early pidgin traces, or one needs to reconstruct the structure of the pidgin to conform to the observed later developments.

In later work, McWhorter (2012) continues to advocate the prior-pidginization perspective, and indeed develops highly specific and well-argued structural effects which follow from this assumption, but he recognizes that it cannot be applied across the board to any creole language. Still, McWhorter is certainly correct in focussing on the loss of inflection as a hallmark of creole languages, for ultimately it is one of the key features which will figure in any debate about whether indeed there are any common traits which justify speaking of creoles as a delimitable linguistic (as opposed to socio-historico-linguistic) class.

I would like therefore to inquire into the dramatic loss of verbal inflectional material which is witnessed in the transition from Arabic to Nubi (compare Table 2a vs. (37)), approaching the issue from a functional perspective. To do this it is necessary to go beyond the simple structural juxtaposition of paradigms — Arabic vs. Nubi — and ask what the function of the Arabic person, number, gender inflections given in Table 2a above is. In a nutshell, one of the most important functions is to mark reference, either situational or text-anaphoric. To see this another background tacking set of data will be adduced.

In this case I look at a relevantly tagged corpus-based study of an Arabic dialect, namely Emirati Arabic (Owens et al. 2009, Owens et al. 2010, Owens et al. 2013). The following Table 3 shows how many subjects are overtly expressed,

47. And indeed, in his 2002 article he qualifies the universality of his prior pidginization perspective, finding it applies to ‘...the birth of many creoles’. However, he does not attempt the daunting task of delineating where one expects a line to be drawn among creoles.

48. Here one would have liked to have had a corpus from the immediate Egyptian-Sudanic dialect area, though the state of Arabic corpus linguistics does not offer the luxury of choosing from the tagged corpus which one immediately may need. To our knowledge, the Emirati corpus is the only one available with the relevant tagged (hence quantifiable) categories. Moreover, theoretical and methodological issues pertaining to subject reference have been explained,
whether nouns or (much more rarely) pronouns, and how many are Ø. It also shows where the last referent to the Ø subject lies. ‘Available in discourse’ means that it refers to an already-mentioned discoursal referent, but not one in the immediately preceding clause; ‘in previous clause’ means that it refers to an argument (subject, object, indirect object) in the previous clause, ‘situational’ means the speaker/addressee (I/you) is the subject. ‘New referent’ is an argument introduced for the first time. Null subject new referents are the plural passives alluded to in (63, 64) above.

Table 3. Null and overt subjects in Emirati Arabic (from Owens et al. 2013, Table 1)

<table>
<thead>
<tr>
<th>Previous mention</th>
<th>null</th>
<th>SV</th>
<th>VS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>new referent</td>
<td>60</td>
<td>78</td>
<td>36</td>
<td>174</td>
</tr>
<tr>
<td>available in discourse</td>
<td>256</td>
<td>108</td>
<td>24</td>
<td>388</td>
</tr>
<tr>
<td>in previous clause</td>
<td>818</td>
<td>152</td>
<td>36</td>
<td>1006</td>
</tr>
<tr>
<td>situational</td>
<td>112</td>
<td>37</td>
<td>7</td>
<td>156</td>
</tr>
<tr>
<td>Total</td>
<td>1246 (72%)</td>
<td>375 (22%)</td>
<td>103 (6%)</td>
<td></td>
</tr>
</tbody>
</table>

The key comparison is that between null subjects on the one hand, and the combined sum of the overt subjects, which are realized either in SV or VS order.

The following is a typical example, taken from the Emirati corpus that was used in the study. Parentheses identify subjects which are Ø in the Arabic original, except for inflectional form.

(68) a. *wu ya iθ-θaani* and came def-other

b. *Ø rigad* slept

illustrated and embedded in a theoretical framework based on this corpus in a number of studies, so it offers a well-defined jumping-off point.

Arabic dialects are quite uniform in their use of overt and null subjects. In order to gauge at least a general idea about the distribution of null and overt verbal subjects in the current dialect area, two extracts were counted totalling 3,270 words from the Nigerian Arabic tracking corpus set 2: pp. 10–13 from GR21AEJX and pp. 3–6 from 107SHRIF. The former has large narrative segments spoken by single speakers whereas the latter is very conversational with frequent turn taking. The same criteria for coding a null or overt subject was used for this as for the Emirati data (e.g. only verbs are coded which occur in a syntactic position which allows a free choice of overt/null subject).

These basic results are intriguing. There are 396 qualifying verb tokens in the texts. Overt subjects account for 111 tokens or 28% of the total, Ø-V (null subject) for 285 or 72%. These NA tracking statistics are a virtual copy of the Emirati data.
c. \( wu \; \hat{o} \; ya-t \)
   and \( \text{came-f} \)

   (Al-Rawi 1990: 122)

a. ‘Then the next one came.

b. (he) lay down.

c. and (she) came.’

In (68a) a new subject (underlined) is introduced, an overt subject. In (68b) this same subject continues as the subject of the following verb, and is realized as a null subject. In (68c) as well there is a null subject, but in this case it does not continue the same subject sequence, but rather references a subject which has been last used in the discourse some six clauses back. The FSG suffix -\( t \) shows that the subject of (68c) is not the subject of (68b), and it alerts the listener to find a previously-mentioned feminine suffix/referent.

Looking at the deployment in actual discourse of the complex paradigms compacted in Table 2a above gives a rather different take on what their function is. It is not so much to mark contrasts in a paradigm, as to mark differentiated subject reference in discourse. Note, however, that in Emirati Arabic discourse (see also n. 48 for WSA), nearly 75% of all verbal clauses have no overt subjects. For the vast majority of Arabic verbs in discourse, the only overt marker of subject is the verbal inflection. Moreover, this marker in the vast majority of cases flags a referent whose identity is previously established in discourse and is not in the clause where the verb occurs.

This verbal inflection is fleetingly short. Perceiving an inflection alone is a considerable task. Trying to calculate its referential function in discourse is probably an order or two more difficult. As noted above in this section, all evidence points to ancestral Nubi being exposed to normal Arabic discourse. From discourse, ancestral Nubi speakers would have had to have learned not only the form of the inflections as in Table 2a above, but also their discourse function. Indeed, without a grasp of the function, the form would be quite useless. From this perspective, the relatively massive loss of verbal inflection in Nubi is thus not so much a failure to learn the paradigms properly — ancestral Nubi, quite unlike contemporary Arabicists and general linguists, would certainly never have been exposed to a well organized pedagogical explanation of an Arabic verb paradigm — as it was a failure to learn what the subject inflections were used for. This failure is all the more understandable given that in the vast majority of cases the overt subject which the inflection references is found at a remove from the verb the inflection occurs on.

This exposition shifts the focus to the role of ur-creole speakers as language learners and SLA perspectives, to which I now turn. Before doing so, in passing,
the discussion around the loss (or failure to acquire) the anaphoric function of verbal subject inflections constitutes the last outcome of Arabic morphology described in (4f) above: the traceless disappearance of Arabic morphology.

4.2 The basic variety (BV) + target shift, an SLA model

The question of prior pidginization leads to a second domain of explanatory research which has gained ground in creole studies in recent years, and that is the interpretation of creoles as language learner varieties, hence invoking second language acquisition theory. This is a large sub-discipline in and of itself, and I restrict myself here to one model which creolists have integrated into their critical purview, namely the idea of a Basic Variety (BV), as developed by Klein & Perdue (1997).

On the basis of a large corpus study, Klein & Perdue argue that in so-called ‘untutored’ second language learning, i.e. not formal classroom learning, a consistent Basic Variety emerges across a range of L1 and target language (TL) pairs. Their results are so consistent that they identify a number of grammatical, semantic and pragmatic attributes of the variety. These include the complete lack of inflection in the BV, the use of few prepositions, the lack of complementizers and subordination, the use of few quantifiers, as well as a series of constraints which explains observed usage, for instance, that word order is sensitive to the agentivity status of the nouns (1997: 315).

Though Klein & Perdue disavow any relation between pidginization and BV (1997: 340), the attributes of BV do in fact share characteristics of pidginization, at least that described by McWhorter (2002), for instance, the lack of inflections, which is what is of immediate interest here.49

The BV has been taken up as a foil against which developments in creoles can be explained. As far as inflections go, many ancestral creole populations appear to share with the subjects who formed the basis of the BV description if not the complete lack of inflectional morphology in their creole or ur-creole variety, then at least a severely reduced amount of it. The BV thus dovetails with a reconstructed early creole development and if inflectional morphology is inherently difficult, as the BV would suggest it is, then the language learning perspective of early creole/untutored second language learners gives potentially a common explanatory reason.

49. Nubi does not agree in any obvious way with any of the attributes of BV. Nubi for instance, has a large number of prepositions (Wellens 2005: 218–22), has complementizers, manipulates sentential word order via the impersonal verb construction (Stem 2, 45) as well as via a grammaticized topicalization construction, and so on.
The relevance of the BV to creole genesis has been noted by more than one author; for instance, Becker & Veenstra (2003) observe that French creoles have a markedly simpler inflectional system than does French. French creoles, moreover, are interesting in that they divide into two broad groups, those in which the verb has one form (Martinique, Dominica), and those in which it has two (Louisiana Creole, Mauritian/Morisyen, Haitian to a limited degree, Sylvain 1979: 103).

Dominica: manž
Mauritian: manže and manž

Becker & Veenstra note that the double forms derive from the French verb paradigm itself, the long form for instance used inter alia in the infinitive *manger*, the short form in in the present singular (*je mange*, etc., 2003: 287). As with most of the Nubi forms described above, while the short/long distinction derives historically from the lexifier source, they have been completely refunctionalized in those Creole French varieties where they occur. The conditions of their occurrence are various. In Mauritian, for instance, the short form is used when the verb selects an adjacent, following object or locative, and the long form is used when what follows the verb is either null, or a constituent not selected by the verb (e.g. a temporal adverb (292).

Becker & Veenstra relate the French Creole developments to what Baker (1990) introduced as a target shift, the target being the lexifier language. Following Baker, early French Creole speakers, for instance in Haiti, would have approximated towards a variety of French, and hence would have acquired a short/long distinction. At a certain point however — in the case of Haitian the crucial event is seen as the start of a sugar-cane-based economy after 1680 which required a large importation of slaves — the number of new French learners became so large, and the social distance from those from whom they could learn a variety of French was so distant, that the speakers of the early creole had to take linguistic developments into their own hands. At this point, termed the ‘target shift’, the target was no longer French, but rather the emerging creole as we know it today, the BV becoming the starting point for further developments.

This model elegantly explains (1) the formal similarities between the lexifier and the creole and (2) predicts that there will be significant grammatical differences, even between morphemes which have the same etymological origin, since once the target is removed, the form of morphemes is set, but their function is not. It also explains the sharply reduced morphology of the creole, since it has a BV in its history.

The results which were described for Nubi, particularly those in 3.3 above, seem to replicate certain predictions of this model. What Becker & Veenstra’s application of the BV to the Creole French model explain is the discrepancy between the French values of morphemes and their values as crystalized in different Creole
French varieties. The BV, as it were, effaces the original French values, and opens the way for their refunctionalization in creole. The Nubi forms described in 3.3, have an analogous genesis, a refunctionalization of morphemes from Arabic in the emerging Nubi.

There are, however, significant differences between Nubi and the Creole French situation as described by Becker and Veenstra. One is an interpretive difference. Becker & Veenstra suggest that the BV is present in both the pre-target shift and the post-target shift variety. For them the BV after the target shift ‘… resulted in the acceptance of the basic variety as a separate norm…’

Even within their own model this interpretation is problematic. The emergence of the dual-form verb in Creole French with its refunctionalized values, a post-target shift product, presupposes that the two forms were already well-profiled as inflectional elements, whatever their pre-target-shift value was. A fixed inflectional morphemic value, however, is a post BV development in the model of Perdue & Klein. In Becker & Veenstra’s data, it is not the inflectional status of the short/long contrast as morphemes, but rather their specific morphemic values which is at issue. The post-target shift would merely have continued treating them morphemically, but in a different way. This in any case appears to be the situation by and large for Nubi. Most Nubi morphemes are presaged as morphemes in Arabic, only in Nubi they often have a refunctionalized value. Such a situation, in fact, is very much like that Winford (2006) postulates for Haitian Creole.50 TMA markers originally were derived from French, but after the target shift, substratal, as well as internally motivated innovations began to set Haitian Creole apart.

However, Becker & Veenstra have another important observation on the BV, which Perdue & Klein (1997) do not develop and which will be developed as central to understanding the genesis of Nubi. As noted, Perdue & Klein point out that the BV has no inflections. Looking more closely at their sample, however, this has to be qualified. The summary of the BV in Perdue & Klein (1997) is an idealization, a generalization over a detailed, longitudinal sample of 22 speakers learning five European languages (see summary in Becker & Veenstra 2003: 288). This extensive study (Dietrich et al. 1995) sampled via the collection of narratives of new immigrants to European countries for up to a 31-month period, in up to five different recording sessions. The BV appears to be based on the initial state of these speakers. Even at this initial point, however, while it appears largely correct that speakers had no functional inflections, they did have inflections. As the speakers progressed, the formal inflections increased in number, even if in most cases at the end of the sampling period most speakers had not yet attained the TL

50. Similarly, Kouwenberg (2006: 206) assumes that the lexifier, not a BV, is an early, pre-target shift TL.
values. In the very careful 1995 study, it is shown that at different sampling points, the inflectional forms appeared to have different values. For instance, Noyau et al. (1995) describe in detail a longitudinal study of four learners of French extending over 29 months. Two of these are speakers of Moroccan Arabic, one considered a slow learner, the other a fast learner.

For the fast learner they note that from the first conversation, at 14 months into the person’s immigration to France from Morocco, he already used up to nine different verbs forms, e.g. dorm, edormi, ladorm, lidormi (Dietrich 1995: 171), without clear structural or pragmatic values, though they note that verbs describing processes, such as ‘dormir’ are those most prone to have more variants. At 21 months, there is a tendency for V-e to indicate a transition from one state/event to another, V-Ø an extensive action/state (1995: 173) and at 33 months, V-e is used for past singular events, V-Ø habitual, as in (69).

(69) 1 don-e la cle
he give-e the key
‘He gave me the key.’

il don-Ø rien
he give nothing.’

At 43 months an auxiliary system is emerging, and for the first time differentiated person clitics (as opposed to third person only forms) appear. For the first time, V-e forms are used for stative past situations, i.e. a tense differentiation is developing (1995: 183).

Even while underscoring the difficulty of detecting the systematicity in the verbal inflections, such as occur, they do describe a tendency of the verbal inflections to develop in three stages (1995: 184):

(70) BV developmental stages of French verbs, SLA corpus
Inherent temporal features of situations
Aspectual difference (imperfective vs. perfective)
Temporal relations

Looking at other languages in the study, it is not even strictly correct that individuals in the initial observation session did not have any discrete inflectional material. Describing the acquisition of German by an Italian, there is a contrast between haben and habe (Dietrich 1995: 81, 82)

(71) haben termin mit doctor vierzehn uhr
have appointment with doctor 14 hour
(presumably, ‘I have an appointment with the doctor’).
Whereas (71) reflects a BV common infinitival form (*haben*), (72) already is half correct in that *habe* is only a finite verb form (though ‘incorrectly’ 1sg, not 3sg).

The study of Dietrich et al. was longitudinally orientated, and as the goal was to assess progress towards the TL, generalizations in the stages towards that goal were the main focus of the research. Variation as such was not the main point.

With the caveat that a full criticism requires a full-scale analysis of the data in variationist terms, two key points seem to be clear. First, while it may be correct that the BV does not have inflections, it does have what correspond in the L1 to inflected forms, e.g. *dorm*, *dormi*, *haben*, *habe* etc. Secondly, from the very beginning the untutored language learners are trying to develop systematic inflectional systems where the inflections contribute in some fixed and ascertainable way to the development of discourse. From these two observations one derives the following hypothesis as regards the evolution of Nubi.

From the very beginning ancestral Nubi were confronted with normal Arabic discourse, so that among their tasks was that of making sense of the rich Arabic inflectional system. Rather than a hindrance to be dispensed with and begin anew in a BV or in a pidginized form of Arabic, the major task was how to integrate the inflections into the emerging Nubi. The Nubi response was multifarious, and at this point defies a simple categorization.

When surveying the entire range of Nubi data covered it is possible to imagine a genesis in which at one point variation analogous to (71/72) occurred throughout the southern Sudan. Many forms were thrown up, until those which we know today as Nubi became established. Much of the inflectional material fell by the wayside, for various individual reasons, one instance, the loss of person, number, gender anaphore discussed in the previous section, 4.1. Other inflectional material had various fates, some winding up as frozen lexemic material, some refunctionalized into new but similar morphemes, and some replicating the original Arabic. Against the BV model which emphasizes a tabula rasa of inflectionless forms initially, the development of Nubi morphology would have been constrained from the beginning by what ancestral Nubi heard in discourse, the rich morphology representing an analytic challenge. In this, morpheme frequency was important.\(^{51}\)

\(^{51}\) A further problem in too literal an application of the BV to creole genesis, as noted in Becker & Veenstra (2003: 299), lies in the unknown social factor which surrounded emergent creole communities. If BV speakers in Europe do not advance far beyond the basic variety, it is no doubt due to the fact that their social networks allow them to thrive in their native language.
Thinking in terms of a target shift, it is difficult if not unrealistic, in the case of Nubi, to divide the world into a pre-target shift era and a post-target shift era. Some Nubi structures in fact copy in principle an Arabic target, plural marking for instance (3.2.1). Others work analogously to the French Creole short/long forms. The modal markers gi- and bi- for instance, are refunctionalized morphemically in a way the short/long forms are (3.3.1). Still other Arabic morphemes never quite made it in Nubi. A discrete distribution of possessive pronouns is discernible (3.1.1), but they are of no functional value in Nubi. In Nubi, target shifts seemed to be going on all the time, with different linguistic features shooting in different directions. In 4.3 I attempt to put some order into these developments.

It is, moreover, difficult to ascertain precisely when the demographic target shift would have taken place, where it did, whether it took place uniformly throughout the southern Sudanese camps, and so on. Still, the demographic statistics from Schweinfurth (1918) do suggest that already by 1874, only 23 years after the first permanent settlement in the southern Sudan from the North, it would have been impossible to assume a dominant Arabic target.

With these points in mind, one might attempt to redefine the idea of ‘target’ environment for Nubi in two ways. First, in chronological terms it could well be that the post target demographics existed from the era of the earliest settlements, and that throughout their short history, ancestral Nubi were faced with the task of working out a viable grammar with restricted Arabic input. By the same token, paradoxically, ancestral Nubi always developed in the shadow of Arabic, at least up to 1888; the pre-target phase was always lurking in the camps as well. This follows from what we know of the basic demographics of the camps of the southern Sudan — there were always fluent Arabic speakers in them — and from linguistic developments which match Arabic, the development of plurals, for instance. Phrasing the situation in Baker’s terms, pre- and post target environments existed side by side, simultaneously in the case of Nubi.52

The creole communities presumably did not have this luxury. They needed a common language. The dynamic thereby created is, apparently, of a different order from the dynamics of the move, or lack thereof, from BV to TL.

52. Whether one can develop a chronology of Nubi development with the help of the eyewitness reports and basic demographic data of the southern Sudan in the second half of the nineteenth century remains to be seen. My guess is that the stabilization of Nubi in its present-day form began in earnest with the advent of the Mahdi in 1884, when the populations in the camps began to polarize and factionalize, and the camp populations were turned in on themselves for survival.
4.3 Explanatory perspectives

From the descriptive account of the origins of Nubi morphology, and from the critical assessment of an application of creole-internal theories of origins, as well as one SLA perspective using the BV as a point of orientation, it should be clear that it is difficult to order all of the data within one comprehensive perspective, at least at this stage in the study of Nubi. What I will do in this section is define four factors, each of which independently offers a perspective for understanding Nubi genesis. The importance of any one of these is clearly highlighted when set against one data set or another, but the greater challenge of integrating them into a coherent model remains open.

4.3.1 Lexifier language is important
Contrary to the thrust of Siegel (2008: 81), quoted in 2.3 above, the role of the lexifier in guiding the development of Nubi morphology is readily discernible in the developments described in 3.1.2, 3.2 and 3.3. It is simply impossible to gauge the development of Nubi morphology without understanding its background in Sudanic and Egyptian Arabic. Even where there is no or little morphemic correspondence at all, as with the development of Nubi’s stress morphemes, the role of the lexifier in setting up the Nubi developments is striking. In this one case, Arabic as it were threw up formal material for ancestral Nubi to work with, and to a degree corresponding semantic as well (3.5.1). Ancestral Nubi took this input and came up with a rather unusual application of contrasting suprasegmental morphemes.

4.3.2 Discourse and frequency
This article has stressed the importance of discourse in understanding how Nubi developed. There are, unfortunately, no documented ancestral Nubi texts, though even if these existed it would still be relevant to try to reconstruct the type of discourse which ancestral Nubi were exposed to. This follows from my assumption, which has been argued for in various places in the article (e.g. 2.1, 4.1), that ancestral Nubi would have been exposed to normal spoken Arabic, and that they would have had to have extracted the basis of their emerging grammar from what they heard.

It follows from this assumption that the frequency of morphemic models in the Arabic discourse which ancestral Nubi would have heard played a crucial role in favouring one construction over another, the lower the frequency the less likely it would have been for a form to have become established in Nubi. In most cases the models were Arabic morphemes, but in one interesting example (3.4, 3.5), the model is in the phonological domain.

High frequency alone is not a sufficient condition guaranteeing morphemization in Nubi, as the fate of possessive morphemes discussed in 3.1.1 shows.
However, frequency is important in two ways. First it allows one to give a richer descriptive account of observed discrepancies in the data. Possessive morphemes are a case in point. They suggest that frequency alone can explain subtle differences in properties of the emerging creole, in this case, why certain nouns in Nubi end in /i ~ i/ or /u ~ a/.

Secondly it allows one to give greater weight to historical explanations which otherwise are phrased in categorical terms. In this regard, frequency plays a filtering role, which can be thought of in ‘if … then’ terms. It does not explain why a certain Arabic source is morphemicized; rather it says, if a form or morpheme should be morphemicized in Nubi, then the frequency of the form will play a crucial role in determining how the process unfolds (pace Erker & Guy 2012). A good example here is the morphemicization of stress. Why stress was morphemicized in Nubi may never be answered. How it was deployed, on the other hand, was suggested to depend in large part on the stress patterns which the ancestral Nubi are hypothesized to have heard. For instance, the most frequent Arabic stress pattern marking stress on the first syllable of a stem, the basic stem in the sense of 3.5.1, is converted to the Nubi stress pattern which marks the most frequent Nubi stem, which is either indicative or imperative. The less frequent Nubi impersonal Stem 2 (44), on the other hand, matches a less frequent Arabic stress-shifted stem, whose stress is always shifted towards the end of the word in Arabic.

Similarly, the data deriving from the personal pronouns are suggestive of a developmental filter: ancestral Nubi were on the way to morphologizing the personal pronoun possessive suffixes, but at some point the development was stopped. Coupled with the frequency data, the ‘arrested development’ forms discussed in 3.1.1 suggest that these forms would have been the springboard from which an Arabic-like suffixal system would have developed.

Frequency in ancestral Nubi, as noted at the beginning of this section, cannot be investigated directly, but it can be investigated on a reconstructed basis, using contemporary dialectal corpus material which plausibly mimics what might have been heard. The background tracking statistics in all cases did indeed suggest a good correlation, or in one case (3.5.1, (62–66)), lack thereof, between what Nubi morphemes became established, and what their relative discourse frequency might have been.

53. In one specific discussion of the role of frequency in creolization, DeGraff (1999: 509) invokes a drift towards less markedness as a major factor which, when combined with a certain degree of frequency, will select the ultimate form. I have used frequency in a different way, simply to explain why certain Nubi forms should appear at all from within a certain contrastive set: frozen possessive pronouns on certain nouns, the basic and more frequent verb Stem 1 vs. Stem 2, an -u on verbs, etc. How, if at all, markedness plays a role in these cases is well beyond the bounds of this paper.
4.3.3 The external world takes precedence over the internal world of the text

This is a factor which has been touched in the description, but which is too large a topic to treat exhaustively here. One of the effects of this factor is seen in the discussion around the loss of verbal inflection in Nubi in 4.1. There it was argued that a major reason for the loss of subject-marking inflection in Nubi was the inability of ancestral Nubi to discern the internal discourse immanent functionality of the inflections.

This can be contrasted with the noun and adjective plural marking which do carry over from Arabic to Nubi (3.2.1). As has been noted in various theoretical contexts (Levelt 1991: 197, 239; Myers-Scotton 2002: 195, 209; Plag 2008: 119), plural marking applies to an inherent property of the noun which relates directly to what it represents in the external world. Similarly, it could be contended that tense and aspect marking, also morphologicized in Nubi (3.3.2), reference directly the nature of the events in the world which they describe.

One can similarly account for the absence of the definite article in Nubi on these grounds, by the fact that in Arabic it marks previous mention in discourse or frame-based identity, i.e. text-internal attributes.54

At this point, it needs to be stressed, the complicating interplay of determining factors begins to come into play. For instance, high frequency seems not to offset the difficulty of functionalizing discourse anaphore. Affixal subject and object anaphors abound in Arabic — except for the 3msg perfect, every verb form has them overtly. However, only one suffix, -*kum*, survived functionally in Nubi.55 On a purely descriptive basis, one might claim that ‘worldliness’, as this section might be called, takes precedence over frequency. One might further adduce the evidence from 3.1.2, the second person conspiracy, to support this position. Second person forms are situational, and hence have a text-external, situational reference.

However, thinking through this formulation further, worldliness exists in the Arabic verb paradigm in the first and second person forms,56 yet these never established themselves in Nubi either. What may be suggested is that the Arabic

---

54. I leave aside the issue whether genericness, one of the functions of definite article marking in Arabic, is an internal or external property.

55. It might be noted in passing that in the only study which has been carried out on de-creolization of Juba Arabic — Nubi is not co-territorial with Arabic and therefore is not in a continuum relationship with Arabic — there does not appear to be a propensity for second person forms to appear before other person markers (Mahmud 1979). Hopefully further study will shed more light on the issue.

56. In fact, first and second person forms have a number of discourse functions; they reference the interlocutors, but they also can anaphorically mark interlocutors in indirect speech, and the second person and first person plural may have a generic function.
verb is simply too ‘noisy’ to be learned without intensive and regular contact with fluent speakers. There are so many competing forms, that establishing a discourse anaphoric basis of single inflections, second person inflections for instance, is not easy. Either the general principle is learned, and applied to all forms, or the paradigm is abandoned and an alternative is needed. I return to this point directly at the end of the next section.

4.3.4 Form in discourse

In 4.2 above, a major criticism of employing the BV as a model for creole development is that the SLA literature itself documents many instances of untutored learners trying to come to terms with inflectional morphology, even at the earliest documented stages of their learning. Moreover, the use of inflections is at least in some cases consistent enough that functionalization stages can be discerned (70), even if these do not correspond to TL systems.

One interpretation which covers both the observed SLA data and certain Nubi developments is the simple observation that form is important. Language learners, or language developers in the case of Nubi, look for regularity cues in the discourse environment. In a number of cases this factor is trivially included in others, since morphemes by definition usually have a form. There are therefore three cases which are particularly interesting in the data presented in this paper. The most interesting in this respect is the correspondence defined in 3.1.1, (5), repeated here.

(73) al-*h/ʕV > IV

This is a regular correspondence, yet on the Nubi end it is based on form alone. It appears that ancestral Nubi built this regularity on the basis of two formal properties: a perception of the special phonological properties of the pharyngeal inputs, along with the fact that a noun is often accompanied by the definite article.

A second obvious domain of development where simple form provides a crucial model is the emergence of stress morphemes from phonologically-defined stress patterns.

A third, more complicated case can be linked to the unresolved issue of the failure to functionalize the Arabic verb paradigm. If, as was suggested above in 4.3.3, a simpler alternative was needed, the imperative verb fits this description in two ways. The imperative is a well-profiled form, aC(V)CVC.57 A stem can always be identified on the basis of the initial aC, a situation contrasting with the Arabic

57. In a general way, whether Nubi has a vowel between C_C or not, e.g. amsuku ‘grab’, aCC vs. ašurubu ‘drink’, aCVC follows in part from inherited epenthetic vowel environments found in WSA. For instance, a vowel is inserted in the context CCsonorant > CVCsonorant e.g. NA āšarah ‘drink.imp’ (Owens 1993: 31–8).
The morphologization of an Arabic creole

imperfect (Table 2a above) where the stem consonant is masked by any of a number of consonants, and by a vowel.

Furthermore, the imperative (see Table 2a), can provide a final vowel, if needed, from its inflectional resources, and apparently it was, hence the final -u discussed in detail in 3.1.2.3. In the verb sample, 177 of the 244 verbs end in a final vowel, the majority of these in -u. What appears to be the motivating factor is a tendency, not a constraint as noted above, towards -V-final words. Investigating the source of this phonological tendency is a task in and of itself. In this case it appears that substratal influence does play an important role, in that a number of languages in the region, e.g. Dinka, do not allow final voiced stops or fricatives, while final voiceless stops can be unreleased or omitted altogether (Andersen 1987: 3, 5). This may well be what lies behind the common loss of final obstruents in Nubi, e.g. be ‘house’ < *beet, iná ‘there’ < *hinaak, bari ‘cold, be cold’ < *baarid, etc.; see also (29) above.58 In the case of Nubi verbs, those from an Arabic source which already end in a vowel, usually -a, tend to keep their vowel, and in fact, as can be seen in the statistical treatment in Figure 1, the fact of being vowel final in the Arabic source is the second most important factor explaining the form of the Nubi verb. Overall, the combination of these two factors — an initial aC drawn from the imperative, and a tendency to favour a -V-final word which in the majority of verbs was created with a final -u, also drawn from the imperative paradigm — leads to a well-profiled verbal form class.

I would note in passing that the imperative stands at the lower threshold of discourse realism. In the Emirati data presented in Table 3 above, for instance, imperative verbs account for only 1.2% of all verbs in the corpus. As noted in 4.3.2, however, frequency is a filter which always interacts with other factors.59

58. And no doubt other factors as well. One prominent group of substrate languages, the Central Sudanic Moru-Madi-Mangbetu group for instance, has only open, V-final syllables (Tucker & Bryan 1966:31, Vorbichler 1971:170). As pointed out by a reviewer, Juba Arabic often lacks these changes, a difference which could well be explicable as post-creolization influence from Sudanic Arabic.

59. As a word of caution, my impression is that most spoken corpora probably underrepresent imperative forms (as well as many other constructions, such as direct questions) since they are drawn from narratives, rather than intimate personal conversations where imperatives are more common. This is certainly the case in the corpora cited in this paper, from Nigerian and Emirati Arabic. I would, however, point out that both texts chosen for the detailed background data part of this paper are very informal. Neither are interviews, and both are free conversations recorded among close friends (and both can be accessed with audio and transcription at the site given in the bibliography).

As can be seen in Table 1, the imperative is based on the imperfect verb; where the imperfect has a person prefix, the imperative has an a-. Accordingly one can look at the tracking statistics
The summary in this section again demonstrates how discrete factors can operate in isolation, and also how they interplay with others identified in Section 4.3. Form alone is important, and as seen in (5/73) can operate independently of morphemic value to explain unusual Nubi historical lexico-phonology. By the same token, it also interacts with morphology, both original Arabic morphology and resulting Nubi morphology, to create the Nubi verb form class.

4.4 Summary

The following Table 4 gives a summary of the morphological elements both the Arabic and Nubi which have been treated in this article. These are classified according to their morphemic status in both the lexifier and in Nubi.

Explanations for the outcomes in Nubi are summarized in 4.3. However, while each of the factors identified in 4.3 appear to apply very well to one or sometimes more than one set of data, none would appear in and of themselves sufficient. For instance, 4.3.3 does not explain why there is agreement between noun and adjective,\footnote{In fact, in Plag’s summary (2008: 128), agreement is, if anything, predicted not to occur in a creole, being an instance of contextual inflection. Many other aspects of Nubi do not fit Plag’s generalization, intended to show the correspondence between creoles and early stages of interlanguage. To mention but two points, the Nubi verbal noun is distributed contextually, limited to non-finite positions, and in verb chains there is stem agreement (Owens 1977: Chapter 6).} nor does it ostensibly explain why the Nubi relative clause clitic of the imperfect verb vs. other forms which are well-represented in the corpora, the perfect (see Table 1) and the active participle. The statistics are as follows, with ‘Rawi’ and ‘Oral’ representing two independent samples of Emirati Arabic data.

<table>
<thead>
<tr>
<th></th>
<th>Rawi</th>
<th>Oral</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impf</td>
<td>225</td>
<td>688</td>
<td>259</td>
</tr>
<tr>
<td>Perfect</td>
<td>203</td>
<td>220</td>
<td>185</td>
</tr>
<tr>
<td>AP</td>
<td>20</td>
<td>62</td>
<td>18</td>
</tr>
</tbody>
</table>

These statistics do track the rank order of the origin of Nubi verbs, imperfect (in imperative guise) > perfect > active participle. However, measured against these statistics both the perfect and active participle origins are well underrepresented in Nubi. Moreover, other corpora, e.g. the heavily narrative-orientated data from Dahlgren (1998: 234) show perfects to dominate. All in all it appears that more than statistical predilection is needed to account for the emergence of the imperative as the verb stem of choice in Nubi.

As a final addendum to this discussion, further to the development of the Nubi verb from Arabic imperatives, one may, circularly, want to speculate that in the highly militaristic camp environment where Nubi grew, imperatives were in fact frequent and therefore became the verbal model. Alternatively, one might claim that imperatives were seized upon as foreigner talk forms, which eventually became the norm. Neither of these ideas have independent support, however (see 4.1).
Table 4. Types of transformations from Arabic to Nubi

<table>
<thead>
<tr>
<th>Arabic →</th>
<th>Nubi</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Forms with meaning</td>
<td>forms without meaning: *al-ħ/iV noun declension, certain -u/a kinship/body part nouns &lt; *possessive pronouns</td>
</tr>
<tr>
<td>2. Forms with meaning</td>
<td>forms with ‘statistically relevant’ meanings, imperative verb, -u suffix ‘category indicators’</td>
</tr>
<tr>
<td>3. Forms with meaning</td>
<td>form with identical meaning: Nubi plural, relative clause clitic</td>
</tr>
<tr>
<td>4. Forms with meaning</td>
<td>form with new, related but not wholly predictable meaning: TMA markers, -kum</td>
</tr>
<tr>
<td>5. Forms with no meaning</td>
<td>forms with meaning: Nubi suprasegmental morphemes</td>
</tr>
<tr>
<td>6. Forms with meaning</td>
<td>lost (not acquired), e.g. verbal subject inflection</td>
</tr>
</tbody>
</table>

perfectly matches that of Arabic, and it may be hard-pressed to account for other data mentioned in passing but not integrated in the current paper (see n. 60).

5. Nubi morphology, creole languages, and Arabic

The development of Nubi is striking in particular because its morphological origins are on the one hand so transparent, but on the other, so different from their Arabic source. It is precisely the differences which present the classical challenge of accounting for creole genesis economically and within a general framework of development. It has, however, been argued here (and indeed elsewhere; Owens 1980, 1989, 1991a, 2001) that Nubi defies easy categorization within existing theories of creole genesis. The substratal debate, one of the most interesting and widespread in creole studies, is of little direct relevance; prior pidginization, as argued above in 4.1, has no explanatory value in the development of Nubi morphology; a feature pool perspective says little of interest for Nubi genesis; whatever its inherent value, the features of the Bioprogram fail to match Nubi structure. In the context of the present paper, which emphasizes the dominant, even if non-deterministic role of Sudanese and Egyptian Arabic, SLA perspectives are potentially relevant. A constant issue accompanying them, however, is the extent to which off-the-shelf theories elucidate individual creole histories. One promising perspective, for instance, is target shift theory, which provides a certain orientation (4.2). However the model developed in the strong form by Becker & Veenstra, whereby the starting points

Beyond the non-correspondence with the off-the-shelf lists introduced by Plag, it is not clear that an interlanguage approach makes any interesting predications as to how the developments described in (5) and (11) came about.
are reduced to a BV minimum before elaborating in a post target variety, does not work for Nubi. If interlanguage stages are suggestive of how creoles developed (Plag 2008), even a brief look at Nubi indicates potential problems (see n. 60).

Within Arabic historical linguistics, Versteegh (1984, 2004) has attempted to integrate Nubi within a broad framework of the development of Arabic dialects, arguing that all have undergone a stage of simplification relative to Classical Arabic. Aside from the problematic historical and historical linguistic postulates on which these ideas rest (Owens 2009), Versteegh has yet to address the question of why Nubi morphology as a system is so massively different from Arabic dialects, whereas Arabic dialects hardly differ in their morphological structure and function from Classical Arabic.61 While one can agree with Versteegh that processes of language transmission are important (2004: 355), historical studies ultimately need to agree on the status of the products which such reconstructed processes lead to. In the case of Nubi, among the Araboid languages (Owens 2001: 352), Nubi is quite unique, offering little in the way of historical models for other linguistic developments in Arabic. In historical structural terms, Nubi is to Arabic what Saramaccan is to English or Haitian Creole is to French (though see below).

This closing discussion points to a final observation. This is that unless one will argue that Nubi is not a creole language,62 all the evidence from the development of Nubi morphology points to the need for a broad typologization of creole languages, with Nubi perhaps left as an isolate. Two distinctive features were illustrated in detail in this paper. It has more morphology than most creoles, and it is transparently derivable from its lexifier. Typologization in and of itself explains nothing, but at least it allows one to carry on a discourse about Nubi within the framework of creole genesis, which is where it belongs.63 To make Nubi more

---

61. And similarly, McWhorter (2007: Chapter 7), who sees the history of Arabic as one of simplification due to contact, where ‘Nubi represents a pole on one end of a cline, upon which the new Arabics occupy an intermediate point’ (2007: 191).

Against this view, in Owens (2009, 2013b) it is argued that the traditional view of Arabic as divisible into two broad historical varieties, one akin to Classical Arabic, the other the modern dialects, lacks historical linguistic foundation, and that a sharp division between the two cannot be drawn within a classical comparative linguistic framework. Nubi, on the other hand, represents a clear linguistic break with any variety of Arabic.

62. As Crowley (2008: 83) ambiguously notes.

63. A different perspective is that of Kihm (2011: 74), who recognizes the problematic status of Nubi as a creole language. Quite correctly, I believe, he concludes that creoles need to be integrated into the local histories from which they emerged. Whether his solution, to group Nubi together with Cape Verdian Creole into a ‘scattered Sprachbund’ is a best one, requires critical treatment.
transparent to the by now long and distinguished tradition of creole language studies, one should perhaps, as a reader astutely suggests, think of Nubi as a creole mesolectal by birth. In this case, as Don Winford suggests, the analogy in the previous paragraph becomes Nubi is to Arabic as Reunionais is to French. Which analogy is the more relevant opens up a new paradigm of discussion for Nubi, for comparative creole, and for Arabic itself.

Received: 2/5/13
Revised: 7/17/13
Accepted: 12/30/13

References


Bakir, Murtadha. 2010. Notes on the verbal system of Gulf Pidgin Arabic. *Journal of Pidgin and Creole Languages* 25(2). 201–228. DOI: 10.1075/jpcl.25.2.01bak


Klein, Wolfgang & Clive Perdue. 1997. The basic variety (or: couldn’t natural languages be much simpler?). *Second Language Research* 13. 301–47. DOI: 10.1191/026765897666879396


Mufwene, Salikoko. 2010. SLA and the emergence of creoles. *Second Language Acquisition* 32. 359–400. DOI: 10.1017/S027226311000001X


Smith, Norval. 2006. Very rapid creolization in the framework of the restricted motivation hypothesis. In Claire Lefebvre, Lydia White & Christine Jourdan (eds.), L2 acquisition and creole genesis, 49–66. Amsterdam: John Benjamins. DOI: 10.1075/lald.42.05smi


**Online resource**

Owens, J. & J. Hassan. In their own voices, in their own words. A sociolinguistically-based corpus of Nigerian Arabic. [http://www.arabistik.uni-bayreuth.de/de/projects/a_Idiomaticity__lexical_realignment__and_semantic_change_in_spoken_arabic/index.html](http://www.arabistik.uni-bayreuth.de/de/projects/a_Idiomaticity__lexical_realignment__and_semantic_change_in_spoken_arabic/index.html)

**Appendix. Basic statistics, sample of 244 Nubi verbs**

Note that not all variables add up to 244, the difference reflecting uncertainty in the classification of some features. The word 'source' here refers to the Arabic source of the Nubi form.

- *-u* final verb
  - *-u*: 124
  - non-*u*: 120

Transitivity
- Transitive verbs: 177
- Intransitive: 65

Imperative verbal source
- Imperative source: 147
- Non-imperative: 69
- Unclear: 17

From verbal source in Arabic
- Yes: 211
- No: 20

From -V final verb in Arabic
- Yes: 39
- No: 200

From perfect verb
- Yes: 36
- No: 204
Author’s address

Jonathan Owens
Bayreuth International Graduate School of African Studies
University of Bayreuth
95440 Bayreuth
jonathan.owens@uni-bayreuth.de