CHAPTER 22

PIDGINS AND CREOLES

MAURO TOSCO AND STEFANO MANFREDI

22.1 ARABIC-BASED PIDGINS AND CREOLES: WHAT THEY ARE AND WHAT THEY ARE NOT

The subject of this article is languages derived from the drastic restructuring of Arabic. They can be spoken either as first or second languages. Typically, pidginization results from insufficient exposure and consequent imperfect learning of a superstrate language in a highly multilingual context. When use of the variety is not discontinued, the original incipient pidgin may become extended or stabilized. Further nativization of the variety leads to its creolization, while continued exposure to the superstrate language may lead to a process of decreralization.

Many and widely different Arabic-based varieties have been called at various stages pidgins or creoles (PCs). In an attempt to classify these linguistic varieties, we distinguish two groups of PCs: the Sudanic PCs and the immigrant pidgins in Arab countries. This geographical opposition turns out to be both structural and historical. A few Sudanic pidgins have developed in the course of time into stabilized pidgins and creoles, while the recent and contemporary varieties spoken by immigrant communities in a few Arab countries bear all the marks of incipient varieties. Moreover, Sudanic PCs developed and are attested out of the core of the Arab-speaking world (in present-day Chad, Southern Sudan, and East Africa), where the direct presence of speakers of the superstrate (Arabic) was minimal or indirect; they therefore adhere to the canonical conditions for the emergence of PCs (see Map 22.1).

* Although often inconsistent and etymologyizing, we retain the authors’ transcriptions throughout. On the other hand, we modify and underline the authors’ glosses.
and could date back to the 8th century and the Islamization of Uzbekistan or the 14th and Timur’s (Tamerlane’s) empire. Connections with other Arabic dialects are uncertain, apart from a certain resemblance with the varieties of Northern Mesopotamia.

Owens (2001) argues that peripheral Arabic is an "Araboid language" that has undergone so many structural changes that it can no longer be regarded as a form of Arabic—being rather a "mixed language." The mixed nature of Central Asian Arabic is disputed by, for example, Zaboronski (2008), who prefers to stress the Arabic character of these varieties. Still, the profound impact of Uzbek, Tajik, and to a lesser extent Turkmen as well as, in Afghanistan, Dari, and Farsi has been tremendous and has induced, among others:

- The change of basic sentence word order into subject–object–verb (SOV)
- The introduction of postpositions alongside the inherited Arabic prepositions
- The introduction of a present perfect based on the Arabic active participle followed, for the first and second person, by the ending -in and the object pronominal affix; for example, zorb-in-ak "you have hit" (Zimmermann 2002: 46)
- The frequent use of "light verbs" in compounds, such as saw "to do" and zarab "to hit," on the model of Tajik and Uzbek
- The loss of the Arabic definite article and the introduction of an indefinite article "fad (= fara)"
- The introduction of the construction Possessor + Possessum followed by a pronominal affix, such as duk aadami miltu “that man’s nationality” (= that man his nationality) (Zaboronski 2008: 429)

As in PCs and non-native Arabic, pharyngealization is lost (but the pharyngeal fricatives are retained).

Certainly, these "Araboid varieties" are not PCs. According to Owens (2001: 353), "credalization involves a greater degree of simplification relative to the source language(s) than does the development of a mixed language; whereas the precise source of many creole structures is opaque, that of mixed languages is relatively transparent." This is certainly true of Central Asian Arabic, where the Turkish and Iranian source is evident. Is Central Asian Arabic, then, a mixed language? Certainly not in the strong sense of a language where two grammars functionally coexist. Cases of such a "double grammar in a single language" exist, although they are exceedingly rare: Michif (Rakker 1997) and Media Lengua (Muysken 1997) are probably the best examples. It seems instead possible to use the label mixed language in a diachronic sense, as defining a language that has undergone deep structural changes and even a complete typological metathesis as a result of the impact of a second language of a radically different type. In such a weaker meaning, Central Asian Arabic (and maybe other peripheral varieties heavily restructured by contact) can certainly be considered a mixed language.
22.1.2 Pidgins and Creoles versus Non-native Arabic

A lower degree of structural simplification and alteration is attested in non-native varieties of Arabic used as interethnic lingua francas. They certainly share with Arabic-based PCs a number of features. These non-native varieties of Arabic are probably more common a phenomenon than our scanty data may suggest. The complex situation of Arabic in Chad has been described and discussed several times, most recently by Miller (2009); Ferguson (1970) reports on Ethiopia. We shall concentrate here on Eritrea, where the use of Arabic as an interethnic medium has been studied by Simeone-Senelle (2000), making in passing a few parallels with the better investigated case of Chad.

In both cases, Arabic as an interethnic lingua franca is a third variety of Arabic, distinct from both the local native dialect(s) and "official" Arabic. Great differences are caused by the starkly different ethnic and linguistic picture of the two countries: in Chad, Arabic is today the major interethnic medium (spoken by maybe 60% of the total population). In Eritrea, the role of Arabic, negligible as a native language, is also apparently very minor as an interethnic medium.

As a consequence, the role of the local spoken dialects is stronger in Chad than in Eritrea; in the latter, as in a typical interethnic medium, the native language of the speakers plays a bigger role in shaping Arabic than in Eritrea Arabic /s/ is preserved by native speakers of Saho (which has /s/ in its inventory) and is replaced by /s/ by Afars (who have no /s/ in their native language); /t/ is preserved in the Arabic as spoken by Saho and often replaced by /t/ or palatalized by "Afars. Everywhere (and, as will be seen, in the Arabic-based PCs), pharyngealization is generally lost; Simeone-Senelle (2000: 157) notes that, when preserved, "emphasis" is rather realized as pharyngealization rather than being replaced by ejective consonants, as is the case in the neighboring Semitic languages. Interdental are either realized as alveolar stops or fricatives.

In both countries, a certain amount of morphological simplification is the rule; however—and this is a crucial difference—in contrast to PCs, verbal inflection and derivation are largely preserved. Thus, in Eritrea the first-person singular of the imperfective coalesces with the 1SG. and md-ndiger "I cannot." Likewise preserved are two series of personal pronouns (independent and bound) and a certain number of Arabic plural patterns. Influence from the native languages of the speakers is somewhat stronger in syntax, where, for example, the verb tends to appear in final position, as in most Ethiopian languages.

2 In both Eritrea and Chad Arabic is one of the two (de jure or de facto) official languages of the country (alongside Tigrinya in Eritrea and French in Chad). In both countries, therefore, official and written Arabic is part of the picture. Moreover, in both countries Arabic is spoken as a native language in Eritrea by just 10,000 speakers (9% of the total population of the country) according to Simeone-Senelle (2000: 155) and by a sizable part of the population (10%) in Chad.

3 de Ponge (1999): Arabic surpasses French as an interethnic medium any local language as well as at least as an oral medium.

22.2 Some General Features of Arabic-Based Pidgins and Creoles

Many features that go under the general label of "simplification" are common to any restructured variety of Arabic, while others are common to most. It is of the foremost importance here that the comparison be established with the lexifying dialect and not with the Classical or Standard language (although influence from the latter may be found in the pidgin due to the well-known phenomenon of diglossia).

Pending a more detailed analysis, the following is a reasonable list.

In phonology:

- Pharyngealization is universally lost; this involves both the pharyngeal fricatives and the pharyngealized fricatives and stops. The former are usually reduced to /t/ or /h/ (with further reduction to /h/ possible), respectively. The pharyngealized stops and fricatives /t/, /d/, /l/, and /h/ are reduced to the corresponding plain sounds.
- The velar fricatives /x/ (I) and /h/ (I) usually merge with /k/ and /g/.
- The uvular stop /q/ is usually changed into /k/ or /g/.
- Gemination and vowel length are usually lost.
- Interdental /t/ and /d/ (both only when present in the lexifying dialect) are either changed into dental stops /t/, /d/, or into fricatives /s/, /z/.

In morphosyntax:

- The productive Semitic and Arabic root-and-pattern morphology is lost as a productive mechanism at both the inflectional and derivational level. Although inflectional morphology in general is also greatly reduced, one notes in expanded pidgin (Juba Arabic) and creoles (Kiub) the partial retention of inflectional plural morphemes as well as the development of elaborated Tense-Mood-Aspect (TAM) marking and of the morphological use of suprasegmentals.
- The definite article /h/ (and its local variants, especially /h/) is lost.
- Analytical expression of the genitival phrase is generalized, and the Arabic "construct state" (Possessive Article-Possessor) is lost as a productive device.
- Only one series of pronouns (the independent ones) is preserved, while possessive and object suffix pronouns are lost.
- Sentential word order tends to be SVO and phrasal word-order Head-Modifier, although variation is attested (especially at the phrasal level) in unstable pidgins and contact varieties.

As expected, the lexical stock is greatly reduced, and its place is taken by semantically transparent compound expressions.
22.3 Early Sources for “Pidginized” Arabic

Thomason and Elgibali (1986) are the only source on the so-called Maridi Arabic. It was found in 1982 by Aisa Elgibali in the Kitab al-masa'il wa'l-mawsil (Book of the Roads and Kingdoms) by the 11th-century Andalusian geographer Abu 'Ubaid al-Bakri. The book, parts of which have been lost, was written in 1068 and is based upon literature and reports from travelers and geographers from the 10th century (al-Bakri himself never left Andalusia). The book is generally considered an important source for the history of West Africa, the trans-Saharan trade, and the Ghana Empire. In the short passage of interest here (which, it must be noted, is lacking in most editions), Al-Bakri provides second-hand data on the Arabic spoken by “Blacks” in an unknown locality called Maridi, which Thomason and Elgibali propose to locate in central Mauritania, that is, in the westernmost part of the Sahara. It is interesting, on the other hand, that the passage is found within a description of Aswan, in Southern Egypt. Kaye (1985) suggests instead that Maridi is the homonymous town in modern Southern Sudan (in a Zande-speaking area, where Juba Arabic is nowadays the main lingua franca). This hypothesis is discussed and refused by Thomason and Elgibali, who do not think that the Arabs could have reached such a southern latitude (close to the present-day border with the Congo Democratic Republic) by the 10th–11th century.

The short specimen (just 10 sentences, but 3 of them are identical) consists of a folks tale. Features typical of pidginization are the absence of an article, the uninjected verbal forms, and the presence of a preverbal aspectual particle written ʾiṯ, which Thomason and Elgibali connect to modern-day ʾiṯ of both Juba Arabic and Kanuri. One interesting phonological feature of the language is actually the shift of the Arabic letter ʾiṯ (variably realized in the Arabic dialects as ʾiṯ, ʾy, ʾiṯ, or ʾiṯ) to the letter ʾiṯ (the latter shift is actually attested in modern dialects of Upper Egypt and Sudan, which could be relevant for the localization of Maridi). Noteworthy is the unexpected presence of what looks like a dual personal pronoun, the lack of an existential element (like fī in other Arabic-based PCs) and, in syntax, the order adjective–noun. The lexical material is entirely Arabic, apart from a form written ʾiṯ and possibly to be interpreted as “people.” Thomason and Elgibali (1986: 326) tentatively connect it with Songhay -koi “person who,” but a much closer parallel can be found in Eastern Daju (Nilo-Saharan, Eastern Sudanic; Nuba Mountains) kīr “man; person.” The same word has entered Kordofanian Baggar Arabic as kīk (Manfredi 2010: 88). This is therefore a further element supporting a Sudanese location for Maridi.

Thomason and Elgibali interpret al-Bakri’s scanty data as evidence of an early Arabic-based pidgin or trade jargon spoken by Berbers. (Berbers are obviously not considered “black” by Arabs, but Thomason and Elgibali suggest that the speakers could have been Berberized “black.”) As a result, in Thomason and Elgibali’s analysis of the text, comparison with Berber varieties features prominently. On the other hand, the authors do not take into sufficient account the fact that the text is claimed to be the reproduction of what a “dignitary from Aswan” remembered of his visit to Maridi, the black population there, and “their miserable Arabic (arabiyathum al-murzia).” What we have is therefore not a “language” but its caricature: the “bad Arabic” of foreigners is seen through the lens of the native speakers.

Similar cases of “broken” Arabic in literary works are of course not unknown in various European literatures. The data found in the Italian literature of the 16th century have been studied several times; Contini (1994) provides a useful summary and a few sound conclusions. Ludovico di Vartena, a traveler from Bologna who lived in the Arab world for many years, gives in his Itinerario (1510) a certain number of sentences in Arabic. According to Contini, these are to be interpreted as instances of the author’s imperfect learning of Arabic, rather than as specimens of an Arabic pidgin. The case of La Zingara, a comedy by G. A. Giancarli (1545), whose protagonist (a gipsy woman) speaks a mixture of the Mediterranean Lingua Franca and of “Arabic” is possibly different. Contini argues that at least a part of the data can be taken to represent an incipient pidgin, possibly based upon an Egyptian variety of Arabic.

22.4 Sudanic Pidsins and Creoles

Following Tosco and Owens (1993), we adopt the definition Sudanic PCs as a linguistically homogenous group of pidgins and creoles that have a common origin in the southern Sudan. In the first half of the 19th century Arab slave traders penetrated southern Sudan from the north setting up military camps (locally known as zarah, SG zaraha, which literally means “cattle enclosure”) inhabited by a heterogeneous population ranged among different Nilotic groups (Mahmud 1983; Miller 1984, 2006; Owens 1985, 1990). Due to the asymmetrical linguistic intercourses between the Arabic-speaking traders and the slave population, a pidginized form of Arabic arose as a lingua franca. This pidginized variety represented the ancestor of all the contemporary Sudanic PCs, and for that reason Tosco and Owens label it “Common Sudan PC Arabic” (253).

The lexicon of the Common Sudan PC Arabic was a mixture of Sudanic and Egyptian dialects (with a predominance of the former dialectal subtype), while its substratum was composed of a number of Nilotic languages, such as Bari, Dinka, and Nuer. Given their common historical origin, all the Sudanic PCs share a large number of linguistic features both in domain of phonology and in that of morphosyntax, even though, due to the later geographical dispersion, Sudanic PCs also show interesting structural divergences. For the time being, the Sudanic PCs include four varieties: Turku, Borong Arabic, Kanuri, and Juba Arabic. On structural grounds, these can be divided into an eastern branch including Juba Arabic and Kanuri and a western branch with Turku and Borong Arabic.
22.4.1 Turku and Bongor Arabic

Turku is the name of an Arabic pidgin once spoken in western Chad. The scanty data concerning this Arabic variety were gathered by Gaston Muraz, a French medical officer who worked in western Chad at the beginning of the 20th century. Owens and Tosco (1993) published a descriptive and comparative study of Turkua based on this account. Turkua was possibly the first pidginized variety that split off from the Common Sudanic PC Arabic spoken in southern Sudan. Bongor Arabic (Luffin 2007) is instead the name generally given in Chad to a pidgin Arabic spoken in the southwestern part of the country in and around the town of Bongor, the capital of the region of Mayo-Kebbi Est. (Other Turkua-like varieties could well be found in Chad but have not been reported.) Bongor Arabic should not be confused with the arabe schadien referred to in Section 22.1.

In 1879, following the formal abolition of slavery by the Turco-Egyptian government, a Nubian trader known as Rabih withdrew with his slave soldiers into present-day Chad and eventually established himself in the region of Borno. According to Owens and Tosco (1993: 183), the Arabic variety that Rabih’s army brought to Chad achieved a sufficient degree of stability becoming a common means of communication for the foreign soldiers and other African populations of western Chad as well. We have no information as to whether Turkua creolized, but Bongor Arabic (see 22.2.1.4) has clear structural affinities with the Arabic variety described by Muraz.

Most of the lexicon and much of the grammar of Turku derive from Sudanic Arabic. In line with the other Arabic PCs, Turkua presents a reduced phonology in which secondary consonant realizations and long vowels are absent. Also Western Sudanic Arabic dialects (Owens 1993b: 86) generally lack pharyngeal sounds, but, contrary to Turkua, they are far from being morphologically less complex than other Arabic dialects. Like other Sudanic PCs, Turkua has only number as a morphological category. Turkua verbs are uninflected forms, which derive in large part from Arabic imperatives. As in Juba Arabic and in Khimbi, verbs often present a final -if. In addition, fi is used as an existential copula. Another important typological feature that Turkua shares with the other Sudanic PCs and with Western Sudanic Arabic dialects is represented by “exceed” comparative constructions using the verb fit(u) “to pass,” as in the following example from Tosco and Owens (1993: 211):

(1) inte awain fit kaddabel
    you bad pass K.

“You (SG) are worse than Kadabel.”

In common with Turkua, Bongor Arabic has intakum as 2PL personal pronoun (which consists of the 2PL independent pronoun intu and the 2PL bound pronoun -kum), while in eastern Sudanic PCs we find the form intakum (from the 2SG independent pronoun itu/itu and the 2PL bound pronoun -kum).

It is important to note that the influence of the Western Sudanic Arabic dialects is much stronger in Turkua than in other Sudanic PCs. For instance, differently from Khidja and Juba, the numeral sequence of tens in Turkua is 10a–11a, as in Nigerian Arabic (Tosco and Owens 1993: 230). Furthermore, the possessive marker in Turkua is dina, which derives from the Western Sudanic form hana. On this account, Turkua and Bongor Arabic are generally put together into the western branch of the Sudanic PCs.

Finally, it should be noted that, due to the prolonged coexistence with Chadian dialects, the structures of the pidginized variety of Bongor show instances of a steady pidginization. For example, in contrast to Turkua and other Sudanic PCs, Bongor Arabic distinguishes between independent (e.g., dina, inti, ku) and bound (e.g., -y, -k(), -há) pronouns (Luffin 2007: 638).

22.4.2 Kinubi

The eruption of the Mahdist revolt in 1884 signaled the end of Turco-Egyptian authority in southern Sudan. In 1888, a military expedition was sent by the Mahdi, Muhammad Ahmad, in an attempt to enlarge his authority to the southern territories cut off the southernmost province, Equatoria, from the northern Sudan. The governor of the province, the German Eduard Schnitzer (better known as Emin Pasha), was forced to flee to Uganda with those slave soldiers who had remained loyal to the central government. The army first moved to Wadelai, just north of Lake Albert. Later on, the troops were co-opted into the British King’s African Rifles, and they subsequently moved to Kenya and Tanzania (Heine 1982: 12; Owens 1990: 220; Luffin 2005a: 28). This series of population displacements gave rise to the current dialectal varieties of Kinubi,6 the only Arabic creole known so far.

Unlike Turkua, the dislocation of Schnitzer’s army to Uganda caused a decisive break of the Common Sudan PC Arabic from both its Arabic lexis and its Nilo-Saharan substratum. As a consequence, the variety once used as an interethnic means of communication in the southern Sudan was rapidly nativized by the children born in Uganda. The implications of such a process concern not only the structural stabilization and the grammatical expansion of the new creolized variety but also the identity of the creole speakers. Given that Nubi communities are now surrounded by a majority of Swahili or Lango speakers, Kinubi has become a fundamental marker of the Nubi ethnolinguistic identity. At the present time, a large amount of descriptive data is available on three Kinubi geographical

---

6 The glossary of “(K)-nubi” derives from the Sudanic Arabic word nubbi, nubi, which generally means “slave.” The term was then modified by the prefix K- that, among other things, marks glossosynonym in Swahili (Kaye 1994: 146; Luffin 2005a: 33). (K)-nubi is also referred to as “Nubi” (Welmers 2005) — although we prefer to use the glossary “(K)-nubi,” in opposition to the ethnosynonym “Nubi.”
varieties: (1) the Kenyan dialects of Mombasa (Luffin 2004, 2005a, 2005b, forthcoming) and (2) Kihera (Heine 1982; Owens 1977, 1989, 1990, 1993a, 2001; Khannis and Owens 2007) and (3) the Kimwili variety of Bombo in Uganda (Wellens 2005).

Regarding dialectal differences within Kimwili, Luffin (2005a: 75–76) proposes some possible isolosses for the identification of the Kimwili variety of Mombasa compared with the dialects of Kibera and Bombo. These are mainly related to some conservative features (e.g., presence of the plural marker -dt) as well as to the lower incidence of phonotactic process such as apocope and epenthesis. Wellens (2005: 179) also observes some diatopic variation among the Nubi communities of Uganda. She also notes that Kimwili speakers of Northern Uganda use the passive form of the auxiliary when the main verb is passive (e.g., arjía sebú úo—"he was left again," also noted for Kibera in Kenya; Owens 1977), while in the southern part of Uganda only the main verb takes the passive form (e.g., diría futú úo "he was passed again"). Another syntactic difference concerns the different position occupied by the negative marker ma; if in Uganda ma tends to occur sentence-finally, in Kenya it often precedes the verb (Luffin 2005a: 216–217; Wellens 2005: 250–253). Furthermore, Kimwili dialects also differ with regard to the different degree of interference from their Bantu substratum. For example, Mombasa Kimwili has integrated from Swahili three noun markers m-, wa-, ma- (Luffin 2005a: 125), while in the Kimwili variety of Bombo we find only the two markers m- and wa- (Wellens 2005: 75).

Kimwili phonologies are rather similar to those of other Sudanic PCs. As in Turkana and Juba Arabic, Kimwili generally lacks the Arabic pharyngealized consonants. On the other hand, it integrated secondary realizations such as ɣ, q, g, ʃ, from Bantu languages (Luffin 2005a: 58; Wellens 2005: 45). Kimwili presents a reduced five-vowels system (a, e, i, o, u); as in Turkana and Juba Arabic, stress is lexically distinctive as in the opposition between siba “seven” and aidi “morning” (Owens 1985: 145). In addition, Kimwili derives an infinitive and a passive voice by means of stress shift: ła:kúku “drink,” infinitive lásákúku, passive lásákúku (Owens 2001: 362). As far as Ugandan Kimwili is concerned, Wellens (2005: 54) proposes a four-way stress contrast that additionally distinguishes a gerund form: kásítu “wash,” gerund kástítu, infinitive kástítu, passive kástítu. An alternative analysis for stress–tone distinction in Kimwili has been proposed by Gussenhoven (2006: 218), for whom Kimwili “has obligatory, culminating, metrically, bound accent, with only a single tone being inserted in the accent location.” Following this analysis, Kimwili would represent an intermediary typology between tone languages and stress-accent languages.

Kimwili and the other Sudanic PCs typically display the order Head-Modifier. The head noun is followed by either the indefinite article wádi or a definite article dê. Pronominal possessors, adjectives, and numerals follow (Wellens 2005: 133):

(a) mísá wádi kwéisi
    woman one good
    "a good woman"

One of the most discussed features of Kimwili and of other Sudanic PCs is represented by the large number of verbs ending in -u#. The explanation proposed by Owens (1985) for this phonomorphological feature is that the Arabic imperative plurals ending in *-u* were the main morphological source for the analogical development of the pidginized verbal patterns. In Owens's diachronic reconstruction, Kimwili -u# subsequently generalized its morphological role to that of a verbal particle, Versteegh (1984), for his part, considers the final -u# as a transitivity marker derived from the Arabic 3rd SG masculine bound pronoun *-hu*; if Luffin (2005a: 265–267) sticks to Owens's position in his description of the Kimwili of Mombasa, Wellens (2005: 138–145) chooses to share Versteegh's hypothesis. In particular, she argues that in Ugandan Kimwili inherently transitive verbs with only one participant occur more often without -u# because they have a low degree of transitivity. Wellens (2005: 331–345) gives a diachronic explanation for his thesis, pointing out that Kimwili verbs derived from Arabic singular imperatives (with or without an object suffix) (Wellens 2005: 141):

(3) ya nyereki dê gi-áku(ú)
    CONJ child DEF PROG-cat
    “Thus the child (was) eating”

Kimwili and Juba Arabic share an innovative system of TAM marking based on the use of preverbal particles. Kimwili has three basic TAM markers:

- The preverbal particle bi- derives from the imperfective-indicative marker *-bi* found in Egyptian and Sudanic dialects; in Kimwili it generally expresses a future tense (Heine 1982: 53–55; Wellens 2005: 153–156).
- The particle gi-/ge provides a progressive meaning to the verb since it finds its origin in the phonological reduction of the active participle "gáu'd" (*-q-l-d' sít”), which is also used in Sudanic Arabic for expressing a progressive aspect (Luffin 2005a: 279; Wellens 2005: 148).
- The morphologically independent marker *kan* derives from the perfective 3rd singular masculine person of the verb "be" *kuan*, and in Kimwili it adds an anterior (past-before-past) meaning to both marked and unmarked verbs (Luffin 2005a: 180; Wellens 2005: 153).

According to the semantic distinction between static and nonstatic verbs, unmarked static verbs such as dju "want" and drufu "know" express a simple present, while unmarked nonstatic verbs such as kásí : "break" and gum "get up" have a perfective reference (Owens 1977: 105; Wellens 2005: 146). A few examples illustrating the most common TAM references expressed by Kimwili are as follows:

Simple present (unmarked static verb; Heine 1982: 35):

(4) sàgi dju báa al dàna bì dê
    sSG want house REL sSG buy DEF
    "She likes the house I bought."
Past (unmarked nonstative verb; Luffin 2005a: 282):

5. nābi wāsul mombasā bādir
Nubi arrive Mombasa early
Les Nubi arrivèrent tôt à Mombasa [Nubi arrived early in Mombasa]."

Future (bi-marked nonstative verb; Wellens 2005: 153):

6. kwēis āna kimān bi-wēnusa ša
good 1SG also FT-talk little
"Good, I will also talk a little bit."

Simple/generic present (gi-marked nonstative verb; Luffin 2005a: 285):

7. māra tāi gi-kēlem nūbi
woman my PROG-speak Kinubi
"Ma femme parle kinubi [My wife speaks Kinubi]."

Past-before-past (kan-marked existential copula fi; Wellens 2005: 157):

8. kan fi rīgā wāt fi rīyaqā na
ANT EXS man one in Rīyaqā there
"There was a man in Rīyaqā there."

gi-marked passive verb; Wellens 2005: 179

9. gi-nyakämā anāši
PROG-capture.PSV people
"The people were captured."

Imperfect (kan with gi-marked verb; Luffin 2005a: 294):

10. ūwo kan gi-so
3SG ANT PROG-do
"Il faisait [He was doing]."

Counterfactual conditional (kan with bi-marked verb in the apodosis; Wellens 2005: 160)

11. kan kan āfā bi-ābiru so sušā
if ANT 3SG know people his ANT 1PL IRB-try do what
"if you had known his people, what could we have tried to do?"

Finally, it is important to remark that, as in many other creole languages, Kinubi expresses a future progressive through the combination of the future/irrealis and progressive markers (Luffin 2005a: 281):

12. ūwo bi-gi-ja
3SG FT-PROG-cone
"Il sera en train de venir [He will be coming]."

22.4.3 Juba Arabic

Despite the partial regression of Arabic in the southern Sudan at the turn of the 20th century, the linguistic heterogeneity of the region contributed to the maintenance of the Common Sudanic PC Arabic as an interethnic means of communication. At the same time, the previously pidginized variety of Arabic started to be nativized by children born of interethnic couples. This situation gave rise over time to an "expanded pidgin" or, in other words, an intermediary variety between pidginss and creoles that has become the native language for only a few of its speakers (Baiker 2006: 139). This variety is commonly known as Juba Arabic and, even if it still represents the second or third language of many southern Sudanese groups, it also became the majority native language of the Juba urban center and is a "national" language according to the Sudanese People Liberation Movement's language policy (although English is the only official language of South Sudan).

Moreover, the prolonged coexistence of Juba Arabic with Sudanese (Standard) Arabic caused this expanded pidgin to be consistently influenced by its lexifier. This situation is described as a post-creole continuum, and it generally results in increased structural affinity between creole languages and their lexifiers (Versteegh 1993: 65-68; Kaye and Tosco 2001: 94-97). Similar to diglossic situations in modern Arabic dialects, the degree of structural interference from Sudanese (Standard) Arabic to Juba Arabic varies a great deal according to sociolinguistic variables such as the speakers' residence and their type and degree of education. As a consequence, Juba Arabic is marked by a higher degree of individual variation than the more stable Kinubi. From a sociolinguistic point of view, it is also important to note that, unlike Kinubi, Juba Arabic does not represent an exclusive marker of ethnic identity. On the contrary, it furnishes an inclusive basis of identification for all southern Sudanese people regardless of their different ethnolinguistic backgrounds.

A great deal of what is known of Juba Arabic (both along the linguistic and the sociolinguistic dimensions) is due to the painstaking work of Catherine Miller (1984, 1985,

7 The glosses of some Arabic (adverb juba) refers to Juba, the capital city of southern Sudan. Previously, Juba Arabic was also referred to as Binbshib Arabic, from the Turkish word binbo "major" (the military rank; Miller 1995: 179).

Given their numerous structural affinities, and the fact that there is a high degree of mutual intelligibility between the varieties, Juba Arabic and Kinubi may be said to represent together the eastern branch of the Sudanic PCs (Tosco and Owens 1993: 250). Apart from the numerous phonological similarities, we can note that, as in Kinubi, the analytic passive exponent of Juba Arabic is tuvi six (from the Egyptian *bita'a, later introduced in Sudanese Arabic in competition with the still more common bagg). Furthermore, as in Kinubi, transitive verbs often end with a final-iu (Miller 1993: 152-154). As an expanded pidgin, Juba Arabic displays typical features of creole languages. For example, Juba Arabic has a prototypical passive construction. This is expressed by a transitive clause in which the patient occupies the subject position, the passive verb is marked by stress shift, and an optional agent follows the preposition ma with (Manfredi and Petrollino forthcoming):

(13) John kuta geni fi sijin (ma jis) J put.PSV stay in prison (by army) "John was imprisoned (by the army)."

Another purely syntactic strategy for passivization involves no morphological marking on the verb and the patient in postverbal position. No agent is expressed, and the construction can be considered impersonal (cf. (9); Manfredi 2013):

(14) afu ana min arabi forgive sSG from Arabic "I was exempted from (the) Arabic exam."

"[They exempted me from the Arabic exam.]"

Furthermore, in line with many creole languages, Juba Arabic uses the bare verb gile "say" as complementizer of verbs of speaking and knowing (Miller 2001: 470).

(15) bniyya de ma be-druu gale jand de gi-kabasu uo girl DEF NEG IRK-know say people DEF PROG-betray sSG "La fille ne sait pas qu'ils vont la tromper."

"[The girl doesn't know that they will fool her]."

It is also to be noted, on the other hand, that Juba Arabic and Kinubi display important grammatical divergences that may be ascribed to the prolonged contact between the Sudanese expanded pidgin and its lexifier. One of the most striking discrepancies between Kinubi and Juba Arabic concerns the use of TMA markers: Juba Arabic possesses the same TMA markers as Kinubi, but it does not allow the sequence of bi- and ge-gi-for expressing a future progressive. That is, because the morpheme bi-, apart from introducing a future tense, also correlates with the habitual and the progressive aspects in Juba Arabic parallel to Sudanese (Standard) Arabic, such as bi-nil "he goes" (Tosco 1995: 458). In view of the fact that bi- also marks counterfactual conditional clauses (see (11) for counterfactual conditional in Kinubi), in Juba Arabic it generally correlates with an unreal aspect rather than with a future tense as in Kinubi.

Second, it is a fact that both Juba Arabic and Kinubi developed innovative grammatical functions compared with their common lexifier. This notwithstanding, if the creolization of Kinubi resulted in a large, independent expansion of its grammatical structures, the influence of Sudanese Arabic on Juba Arabic halts any similar process. For instance, in Juba Arabic the particle ya can be used either as a vocative marker (e.g., ya za "Hey man!"; Manfredi and Tosco forthcoming) or as a presentative focus marker (e.g., uo ya bhlu fi yaganda de "he buys them in Uganda", where the verb bhlu is in focus, Manfredi and Tosco forthcoming). In Kinubi, on the other hand, ya can be used either as nonverbal copula occurring between the subject and the predicate (e.g., dandi maris de ya jidi têna "Cet enfant Marius est notre grand-père"; Luffin 2005a: 184; see also (3)) or as a contrastive focus marker (e.g., ndira laiki ya ma gi-diqo lu "it was not your mama who beat you", where the negative operator ma is in focus, Wells 2005: 172).

As a final remark, Juba Arabic and Kinubi also differ to a degree with regard to their lexicon. As other pidgin languages, Juba Arabic compensates its lexical gaps through the lexicalization of Arabic morphosyntactic sequences. In Kinubi, these complex lexical items have been gradually replaced with Swahili or Luganda borrowings. For example, if in Juba Arabic the relative clause mdira râgîlî mà tou (lit. "the woman whose man died") has been lexicalized with the meaning of "widow" in Ugandan Kinubi the same semantic reference is expressed by the Luganda loan mawându (Behnstedt and Woidich 2010: 49).

### 22.5 IMMIGRANTS PIDGINS IN ARAB COUNTRIES

A number of incipient pidgins are attested among immigrant communities in the Eastern half of the Arab-speaking world (Holes, "Orality"). Such varieties have been so far documented in the Gulf (Smart 1990; Nesb 2008; Bââdir 2008, 2010), Iraq (Avram), and Lebanon (Bizar 2005, 2010). All share typical features of contact varieties or, to use Avram's (2010) label, "pre-pidgin.

#### 22.5.1 Gulf Pidgin Arabic

Gulf Pidgin Arabic (GPA) is the first documented variety with Arabic as a lexifier that emerged in recent years. Following Smart's (1990) early account, GPA has received
recently a good description by Nesb (2008), which is based on fieldwork in the Omani town of Buraimi, on the border with the United Arab Emirates (UAE), while its verbal system has been analyzed by Bakir (2008, 2010), with data from Qatar.

The emergence of a pidginized variety of Arabic in the Gulf States (from Kuwait in the North to Oman, including Saudi Arabia) is the result of the oil boom and ensuing economic development of the area, and can be traced back to the 1970s of the past century. Nesb (2008: 21) draws attention to the numerical weight of the immigrant communities in this part of the Arab world (up to over 80% of the total population in the UAE), their diverse linguistic background (although immigrants from the Indian peninsula are a clear majority), and a “politics of exclusion” (although their permanence in the community is often very long immigrants cannot ever really integrate in the host community and very rarely intermarry with locals). All these factors are conducive to the emergence of a restructured variety of the language of the host community and its possible stabilization as a structured variety.

Smart (1990) coined the term “Gulf Pidgin.” In his short sketch this emergent variety is seen through Arab eyes, his data being written material of a jocular variety (mainly cartoon captions) in newspapers from the United Arab Emirates. The result is probably more akin to Arabic foreign talk than to any actual speech, but certainly the influence of foreigner talk was important in the genesis of GPA. The influence of the substratum is heavily felt in the shift of Arabic fi to /fi/ in native speakers of Sinhala, Tagalog, Javanese, and Chavacano (who do not have /fi/ in their inventories; Nesb 2008: 32). Against the general trend to Head-Modifier phrasal order, many speakers produce sentences with Modifier-Head order, especially in the case of possessives akhi “sister’s daughter” and ana ukti “my sister,” an order which Nesb (2008: 54) says to be quite common among Sinhala speakers. The lexicon is overwhelmingly Arabic (amounting, according to Nesb 2008: 27, to more than 95% of the total). Arabic phrases are often interpreted as nouns, as in the frequent case of tismit (Gulf Arabic “what’s your name?” from ti-smit-ik “what-name-your.M.SG”) for “name.”

The only nominal series is made of the Arabic singular personal pronouns ana “I,” inte “you (SG),” and both huwa (Arabic “he”), and hija (Arabic “she”), but without any evident gender opposition. The only plural personal pronoun is nafardt (“persons” in Arabic), but the use of two pronouns is also possible (such as ana huwa “I he” to say “we”; Nesb 2008: 52).

The genitive construction employs the particle mal (from the Gulf Arabic possessive morpheme ma’dal); a few examples of pronominal possessive affixes, such as binti (Arabic binti “my daughter”) seem lexicalized and interpreted as bare nouns. Again, the order Modifier-Head is occasionally used with genitive constructions with mal, as in mal ana sadqa “my friend” (Nesb 2008: 63). Negation is mainly expressed through ma’tal, although a specific negation for nonimperative verbs, ma, is widely used (although ma’tal tends to spread to verbal negation too). Nesb (2008) does not deal with the independent negation “no.”

The whole question of the verb–noun distinction in GPA is complex as well summarized by Nesb (2008: 83–85). In general, only one lexeme is taken from any single Arabic root, and in principle nouns can be used as verbs; a few, such as kalum “to speak,” or tali “to study, learn,” are much more frequently used as verbs than as nouns. Still, etymological nouns and adjectives are verbalized through the use of the light verb sawa “to do,” and only etymological nouns tend to enter into possessive constructions with mal. On the basis of these criteria, Nesb thinks that a verb–noun distinction is possible in GPA. Arabic verbs appear to enter GPA either under the form of the (Gulf) Arabic imperative singular or of the imperfective third masculine singular, without any clear ratio: yerid “wait” (<MSG MPF) always appears in the form of the Arabic imperfective and a few others, such as yijit “sit,” overwhelmingly do. Still others, such as ruk “go” almost never appear in the Arabic imperative form, while others, such as yafif “run” or yaqit “see,” are strongly favored in the “imperfective.” In still other cases, such as yishel “to work,” the preference for the imperfective form can be due to the desire to avoid an initial /st/ cluster (cf. also Bakir 2010: 208, who thinks that prefixation of y-V- is basically phonologically determined).

The verbal system does not have real TAM markers, either in the form of affixes or discrete, dedicated morphemes: rather, the adverbs anam “first” > “before” and bidan “then, later,” though retaining their use as adverbs, are often used in a preverbal position to mark anteriority and posteriority, respectively (Nesb 2008: 65). The same function can be taken over by other adverbs too, such as bira “tomorrow” and tumi “yesterday” (Bakir 2010: 211). Bakir (ibid., 212) mentions the postverbal use of kalaas “done, finished” as a marker of completed action:

(16) Tattuk kalasa kalam sill  
cook CMLP meat take
“When the cooking is done, I raise the meat (from the pan).”

Possibility is expressed by yimkun or mukun and necessity by baazim (Bakir 2010: 213). Intention may be expressed through the use of ruah “to go” (Bakir 2010: 214):

(17) Tanu ruuh kallal baabaa  
SG PT speak master
“Th’l talk to Master.”

In contrast to Sudanic PCs, fi is a general copula, as in (Bakir 2010: 215, 216, respectively):

(18) fi mokkil
COP problem
“There is a problem.”
Still, *fi* is used in a much wider range of meanings; many cases point to an interpretation as a progressive marker or as a marker of factuality (Nees 2008: 89, 90, respectively):

(19)  
<table>
<thead>
<tr>
<th>تُست</th>
<th>مُعْنٍ</th>
</tr>
</thead>
<tbody>
<tr>
<td>fi</td>
<td>majmun</td>
</tr>
</tbody>
</table>

"Are you crazy?"

As mentioned already, a propos the verb–noun distinction, the verb *sawwi* "do" is used as a light verb, as in *sawwi māt* "kill" ("make die"); *sawwi arīs* "marry" ("make bride"); *sawwi madījn* "clean" ("make clean"); *sawwi sulī* "ask" ("make question"). GPA has not stabilized yet, and both inter- and intrapersonal variation remains huge. Following Winford (2006), Bakir (2010: 223) considers GPA to be a "stage 2 pidgin," which has acquired some simple rules for predication and solutions are not idiosyncratic.

### 22.5.2 "Romanian Arabic Pidgin"

Avram (2010) is a short description of an Arabic-based contact variety used in Romanian-serviced oil camps in Iraq in the period between 1974 and 1990, when the outbreak of the first Gulf War marked its death. The actors in the contact situation were the Romanian workforce and both Egyptian and Iraqi Arabs. The language itself is Arabic, but there is a sizable amount of English among the approximately 150 words that, according to the author, make up the basic vocabulary of the language. They include basic nouns and verbs, such as *work, sleep, speak,* and *give.* English-derived material may be used pronominally, as in the case of *pipol* (English *people*) employed as a personal pronoun for all persons. It is not so much the use of a common noun as a pronominal that is relevant here but rather the use of a nonprimary lexifier (English; Gulf Pidgin Arabic uses in the same context the Arabic plural

* A "Stage 2 pidgin" (or prototypical pidgin) is "characterized by a clear though rudimentary grammatical organization, in other words, regular though simple rules of predication..." (Winford 2006: 298).

* Pre-pidgins can be identified, following Winford’s (2006: 298) definition of "stage 1 pidgins," as varieties characterized by "very minimal syntactic structure, many of which lack either arguments or predicates."

#### 22.5.3 “Pidgin Madam”

Under the label of “Pidgin Madam,” Bizri (2005, 2010) discusses the simplified Arabic used by Sri Lankan housemaids in Lebanon. Again, the contact situation is basically bilingual: Lebanese Arabic versus either Sinhalese or Tamil. The Sri Lankan languages do not contribute at all to the lexicon, which is exclusively Arabic with a sprinkling of English (again, no is the negative particle) and French (e.g., *bonjour* and other greetings; *merci, bêbê*) elements, which belong to the Arabic lexifying register: for example, *bonjour* and *bônson* are part of Beirut’s speech and are actually more neutral than any Arabic greeting except the informal *marhaba* (although they are more common among Christians and the middle class; Germanas 2007). Bizri (2005: 54) rightly notes that in Pidgin Madam there is *moins d’influence du substrat que dans les pidgins traditionnellement étudiés* ("less substratal influence than among the traditionally-studied pidgins"). The influence of the substrate is basically limited to the phonology (thus, vowel length is retained), but, as noted by the author, a few modifiers (adjective, determiner, demonstrative) precede their head, and this may reflect the word order of the substrate.

A striking feature of the language is the abundance of morphologically complex forms of the lexifier that enter it: verbs are not taken from imperative singular forms only but to a large extent from the imperfective. The actors in the communication event are generally the landlady (the madam that gives the variety its name) and the maid: this is therefore basically a women’s language. Not surprisingly, many verbal forms have the Arabic feminine (generally, second- or third-person singular) affix: thus, *ruuh* "go" (rather than masculine *ruuu*), *neем* "sleep" (*m* "she slept"). Likewise, adjectives generally enter the language in their feminine singular form. Object pronouns and even

* A "Stage 2 pidgin" (or prototypical pidgin) is "characterized by a clear though rudimentary grammatical organization, in other words, regular though simple rules of predication..." (Winford 2006: 298).

* Pre-pidgins can be identified, following Winford’s (2006: 298) definition of "stage 1 pidgins," as varieties characterized by "very minimal syntactic structure, many of which lack either arguments or predicates."
modal markers are often incorporated in the verbal form: be-t-tebbiri "love, like" (< "she loves me"), tebbirni "talk" (< "you (F) talk to me."). Reanalysis of noun phrases (NP) as bare nouns is likewise common: s-ams-o "name" (< "what's his name?"). Unlike Romanian Arabic Pidgin, Pidgin Madam has a copula fi used in a very large variety of contexts (Bizri 2005: 65):

(22) hayda fi polis
this COP police
"ça il y a police" > "le policier qui était là-bas [the police who was there]"

(23) ana kulla fi gran
SG all COP business
"Ma tout il y a affaires" > "Toutes les affaires qui m'appartiennent [everything that pertains to me]"

This seems to imply that, if anything, there is even more latitude for variation in Pidgin Madam than in the other varieties. Again, everything points to a pre-pidgin rather than to a prototypical or more stabilized variety.

22.5.4 Conclusions on Arabic-Based Incipient Pidgins

It is not clear to what extent these simplified varieties are used by Arab natives in their interaction with foreigners: Bizri (2005: 66) writes of "une permission d’inventer donnée par ‘Madam à la domestique,’ (‘a leave to improvise given by the Madam to the maid’)" while Avram (2010: 21) speaks of "significant inter-speaker variation, due to the influence of the speakers’ first language, i.e. Romanian, Egyptian Arabic and Iraqi Arabic respectively." Still, the language he describes is consistently the one used by Romanians, and the very name he uses to define it, "Romanian Arabic Pidgin," points to the users of this variety and not to any significant role played by their native language. Rather than true contact languages, it seems we are dealing here with varieties having Arabic as their target, but with two different inputs. Pidgin Madam is based on fairly normal Lebanese Arabic in terms of morphological structure; if not in syntax, lexicon, and maybe tempo of speech. In the case of Romanian Arabic Pidgin, on the other hand, some input in terms of a restructured foreign Arabic used by natives seems more probable. According to Bizri, the basic principle at work in Pidgin Madam is mimécie, "mimétisme de la langue des maîtres." If this is so, the Pidgin is the exclusive domain of the foreigners when speaking with Arabs; in the case of Romanian Arabic Pidgin (as in Gulf Pidgin Arabic), use on the part of the natives is possible. In both cases, and different from "true" pidgins, further use on the part of a multilingual community of non-Arabs is excluded.

22.6 The Relevance of Arabic-Based Pidgins and Creoles for Arabic Linguistics and General Creolistics

Generally speaking, Arabic-based PCs have engendered very little attention among either Arabicist or creolists. On one hand, Arabic PCs were simply not considered part of Arabic studies, and, even after Versteegh (1984) brave proposal to use pidginization as a cornerstone of Arabic historical dialectology, they have received little consideration. On the other side, general linguists working on theories of pidginization and creolization have generally limited their attention to European-based PCs. In this regard, Jonathan Owens (2001), in an attempt to draw further attention to the structural relevance of the Arabic PCs for contact linguistics, aptly calls Kinubi "the orphan of all orphans." It is also true that during the last decade our knowledge of Arabic PCs progressed a great deal thanks to new descriptions of Kinubi varieties (Luffin 2005a; Wellems 2005) and to innovative studies on immigrant pidgins in Arab countries (Bizri 2005, 2010; Bakir 2008, 2010; Neiss 2008; Avram 2010). This notwithstanding, Arabic-based PCs still represent a marginal sphere of research.

Pidginization and creolization have been called on in the birth of the modern Arabic dialects (a topic not germane to the present article), most notably by Versteegh. Versteegh (1984, 2004) argues that the modern dialects arose through a stage of simplified pidginization followed by elaboration and creolization. At a later stage, the creoles came under the influence of the standardized form of the target language, thereby losing their most "deviant" features. Versteegh's hypothesis has been generally met with skepticism by Arabic scholars (e.g., Holes 1995, 2004; Fischer 2006; Owens 2009 [2006]), Catherine Miller (2002), emphasizing again the importance of Arabic-based PCs for historical Arabic linguistics, instead stresses that the analysis of the ongoing pidginization-creolization process in Southern Sudan could have a theoretical relevance for historical dialectology in respect of the fact that Arabic dialects always spread in contact environments.

As to general linguistics, the inclusion of Arabic-based PCs would greatly benefit creolists in their assumptions on pidginization and creolization, as their study may contribute to the long-standing discussion concerning the definitions of incipient pidgin, stable pidgin, expanded pidgin, creole, and post-creole. The different sociolinguistic contexts that gave rise to Arabic PCs may shed new light on the interdependence between contact situations and their linguistic outcomes and help clarify a number of questions related to pidgin and creole genesis.

Substratum accounts of creole genesis do not fare well with Arabic-based PCs: a list of words of "African" and unknown origin in Kinubi (Pasch and Thedwall 1987: 141-144) suffices to reveal how limited the impact of the substrate has been in the lexicon
most words in the list are actually recent loans from Swahili and would rather qualify as adstrate. Grammatical influence is equally limited (notwithstanding the special, but in fact limited, role played by Bari and studied by Bureng 1984). Universalist approaches have been adopted for the analysis of the verbal system of Kinubi (Onwes 1993a) and Juba Arabic (Tocci 1995). Recently, it has been shown that Kinubi falls squarely into the "creole type" as defined by Balcker et al. (2011). Juba Arabic is thereby included as well.

Onwes (2001: 368) proposes to analyze Kinubi in the light of the "restructuring" process that involved its Arabic lexifier. Adopting this perspective, Onwes reduces the role played by substrate and universal ("no-strate") factors in creole genesis and concentrates on the innovative aspects of the Kinubi grammar compared with Sudanic Arabic. More in particular, he argues that restructuring in Kinubi allows for an expansion of the tripartite nature of the creole-origin hypotheses (i.e., superstrate, substrate, and no-strate), since the "restructuring" hypothesis is "a consequence of the inability of the other approaches to fully account for the origins of creole structures" (ibid).

Recently, both Kinubi (Lufkin forthcoming) and Juba Arabic (Manfredi and Petrali forthcoming) have been included in the Atlas of Pidgin and Creole Language Structure (APICS). APICS gathers comparable synchronic data on the grammatical and lexical structures of a large number of PCs in an attempt to present the geographical distribution of the most relevant features for creole linguistics. Thus, the presence of two Sudanic Arabic-based PCs will give the opportunity to compare their typological proximity with non-Arabic-based PCs and to finally enlighten their superstrate-substrate against their no-strate features.

References


Behnstedt, Peter and Manfred WoDich. 2010. Wortlist des arabischen Dialekte. Band II-


