

Arabic pidgins and Arabic colloquials: a syntactic comparison

Joost Kremers

University of Nijmegen, The Netherlands
joostkremers@fastmail.fm

May 13, 2005

In this paper, I will look at the syntactic structure of an Arabic (stable) pidgin, Turku, and compare it with the syntax of a modern Arabic colloquial, specifically Cairene Arabic. With this comparison, I would like to investigate the syntactic feasibility of Versteegh's (1984) proposal that the modern Arabic colloquials developed from pidginized and creolized varieties of Classical Arabic. For the comparison, I will use a syntactic model based on Chomsky (1995). I will introduce the relevant aspects of this model in this paper.

Versteegh (1984) claims that the modern Arabic colloquials developed through a process of pidginisation, creolisation and decreolisation: his idea is that the Arabic spoken by the Arab conquerors was pidginised when the conquered people needed to communicate with their conquerors. In the following step, the pidgins that had developed became the input for new-born children of mixed parents, who as a result developed creole varieties of Arabic. After that, continued contact with speakers of the "original" Arabic led to a process of decreolisation, in which the creole varieties slowly lost many of their creole properties.

In this paper, I argue that from a syntactic point of view the process Versteegh describes may not be theoretically impossible, but it is actually quite unlikely. The reason for this is twofold: first, the striking similarities of the (main-stream) Arabic colloquials, (reference)¹ and second, the large differences that I will argue must exist between the colloquial dialects and the pidgins that Versteegh claims they derive from. The latter fact makes it unlikely that in all the places where an Arabic pidgin developed, the decreolisation process led to such uniform results.

In order to demonstrate the major differences that exist between the syntax of Arabic colloquials and Arabic pidgins, I will compare Cairene Colloquial Arabic with Turku, an Arabic-based pidgin, spoken in Chad and perhaps in the Central African Republic in the 19th and early 20th century. The pidgin was originally described in a work written by one Gaston Muraz, a colonial medical commander, around the year 1930. I will base my discussion on the paper by Tosco and Owens (1993) (henceforth T&O), who have analysed the (unpublished) book by Muraz.

Although Turku is obviously not an example of the pidgins that Versteegh claims must have existed in the centuries after the Arabic conquests, we can nonetheless take it as a reasonable model of those pidgins, because the syntax of the Arabic colloquials resembles the syntax of the spoken Arabic at the time of the conquests.² Another reason of course to take a modern pidgin as model is that we

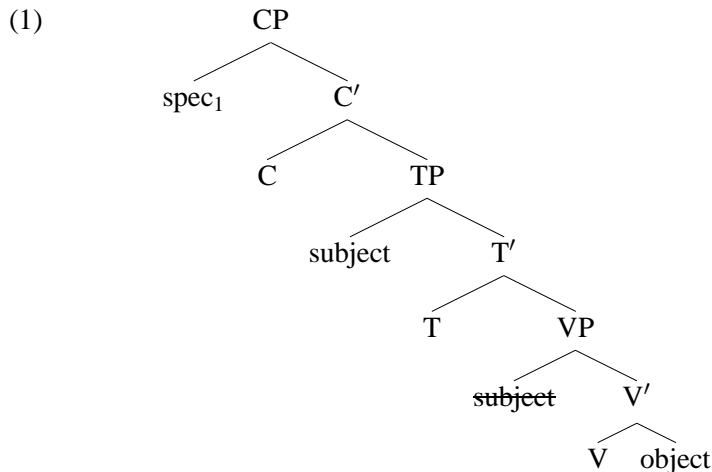
¹With "main-stream", I refer to the dialects spoken in what is now seen as the Arabic World. On the periphery of the Arabic World, one can find languages that appear to have a strong Arabic influence, but in many ways differ drastically from main-stream colloquials.

²There is much discussion in the literature whether this variety of Arabic was basically what we now know as Classical Arabic, or whether it was a dialect (or group of dialects) that differed from Classical Arabic. For the present point, this matter is irrelevant, because we can safely assume that the general syntactic structure of Classical Arabic, modern colloquials

have no data on Arabic pidgins that may have been spoken shortly after the conquests.
 (outline of the paper)

1 The syntactic model

The syntax model that I use in this paper is based on the minimalist model developed in Chomsky (1995) and subsequent work. In this model, the basic clause structure can be represented as in (1):



In this structure, V indicates the position of the lexical verb. *Subject* and *object* indicate the positions of the noun phrases that express the subject and the object of the verb. It is assumed that the subject originates in the VP, and then moves to the position indicated.³

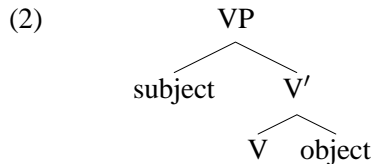
T carries the features for tense and aspect. In many interpretations of the model, T is split up into two or more different heads, with at least one for tense and one for aspect and possibly others as well. (See for example Cinque 1999 and Giorgi and Pianesi 1997.) Often T is used as a short-hand for the more elaborate structure. The tense and aspect features can be marked in the sentence by independent elements (as is the case in many creoles), or by affixes on the verb stem, like in many European languages. Sometimes a tense or aspect marker is actually “fused” with the verb stem, as is the case in Arabic. The assumption is that if such markers appear on the verb stem, the lexical verb V moves to the position of T. Apart from housing tense and aspect features, T also mediates in verb-subject agreement. The subject moves from the position in which it is generated, the specifier position of VP (spec,VP for short), to spec,TP, which creates a configuration in which the verb in T and the subject in spec,TP can agree.

The head C is the location of the complementizer (in the case of embedded clauses) and of sentence-introducing particles, such as Arabic *'inna*. Its specifier position (spec₁ in the tree in (1)) can for example be filled by a fronted (topicalized) element. The C element is taken to express the *force* of the clause, i.e., whether the clause is declarative, interrogative, imperative etc. Like the head T, C is often considered to be a shorthand for a number of different force markers, such as Foc (focus markers), Top (topic markers), Q (question words), etc.

and possible ancient colloquials was by and large identical, especially when compared to the syntactic breakdown and restructuring that normally takes place in pidginisation and creolisation processes.

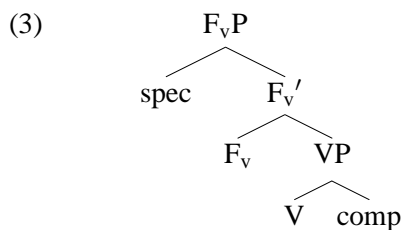
³In fact, this is a somewhat simplified version of the usual model, which claims the presence of a transitivisor or predicator *v*. The details of the model are not relevant here, so I will use the simplified version.

The assumption is that the structure in (1) is present in the clause in any natural, native language. Pidgins, however, are characterised by the lack of a consistent verbal system for marking tense and aspect (Holm, 2000). One way to describe this in the current model would be to say that everything above VP, the so-called functional shell, is simply absent:



However, if it is true that the structure in (1) is universal, it seems impossible that the functional heads C and T are just not present. In Kremers (2003), I take a somewhat different view of the functional shell around lexical items, which could explain the incongruity. Following Chomsky (1995), I assume that a functional shell contains a number of features. Features are formal syntactic properties such as tense, aspect, mood, number, gender, and for nouns case, definiteness, etc. In principle, each feature can project its own functional head, which will always take place in a fixed order. However, it is not *necessary* for each feature to project a separate head. It is also possible that two or more features are projected onto a single head.

Another way of phrasing this is to say that the functional shell minimally contains one head, with a number of features. If necessary, this head can be split, to form multiple heads, each projecting its own syntactic feature. What we see in pidgin languages then, is a combined head that expresses all the features of the functional complex. That is, one should see the structure of the verb phrase as follows:



In (3), the only two remaining heads are V, the lexical verb, and F_v, the verbal functional complex. Both these heads can, if necessary, project extra heads. Making this assumption, we can say that the difference between natural, native languages and pidgins is that a native language splits its functional heads in a consistent manner, whereas a pidgin does not generally split the functional head, or does so on a completely ad hoc basis.

2 A case study: Turku

Let us now take a look at Turku, to make the above remarks more concrete. In section 7 of their paper, T&O discuss the tense marking on the verbs in Turku. They indicate that there is no formal indication of tense or aspect in the clause in Turku, not with verbal affixes, nor with preverbal markers. They claim that the only consistent exception is the imperative, which generally lacks a subject, which they claim can be considered a formal marker of the imperative.⁴

⁴Although their paper contains an example of an imperative with a 2nd person pronoun, and it also contains an example of a verb with an understood but not expressed 3rd person a subject, which is obviously not an imperative. The lack of the subject, then, cannot be seen as a formal marker of the imperative.

The following pair of examples can serve to illustrate the lack of tense marking:⁵

- (4) a. máta úman béji shreb búta dá
when they come drink pond this
'when did they come drink from this pond?' (p. 216)
- b. hú béji sálam ána
he come greet me
'he is coming to greet me' (p. 216)

Note that these and all other examples are taken from Muraz' work; the translations are from T&O, who have translated Muraz' original translations from French.

In (4), the verb *béji* 'to come' is used in two different tenses: in the past tense in (4a), and in the present (progressive) in (4b). There is no formal marking on the verbs, however. The following example shows that the future tense does not require any special marking either:

- (5) a. hu máshi gidám anína
he walk in front us
'he is walking in front of us' (p. 194)
- b. batán íntukum máshi fi frans
again you.PL walk to France
'you will return to France' (p.194)

(5a) contains an example with the verb *máshi* 'to walk' in the present progressive, (5b) has the same verb with a future tense meaning. Note that the word *batán* cannot be taken as a marker of the future tense. The word *again* can well be used with a past tense verb (cf. *then I went to France again*). Presumably, in the current context (which is absent from T&O) *batán* would have been enough to indicate the future tense, but this does not mean it should be taken as a formal marker.

Another element that is consistently absent from verbs in Turku is person marking, or subject agreement.⁶ The examples given above may serve to show this: (4a) has a third person plural subject, (4b) and (5a) both have a third person masculine singular subject, and (5b) has a second person plural subject. The verb form is the same in the (a) and (b) examples of both, however. For another example, take the following:

- (6) a. fishan úman árfu íntukum áfya
so they know you well
'so that they know you are well' (p. 195)
- b. ána árfu same (...)
I know well (...)
'I know well (that...)' (p. 265)

In (6), we see the verb *árfu* used with a third person plural subject and a first person singular subject, without any formal marking on the verb.

However, T&O also make clear that there are some markers in Turku that do unequivocally designate aspect or tense. The most obvious of these is *gaed/gahed*, which indicates "an imminent, continuous or recurrent action" (p. 216):

⁵All the page references given with the examples are to Tosco and Owens (1993).

⁶Object agreement is also consistently absent, but Arabic colloquials have no object agreement either, so I will not consider it here.

- (7) a. yaoda mátra gáhed béji
 behold rain CONT come
 'there the rain is coming' (p. 216)
- b. al-nédem gáed béji
 PART-man CONT come
 'the man who is coming' (p. 216)

It should be noted that the particle *gahed* is certainly not obligatory. The bare verb stem can occasionally also express a continuous or imminent action.

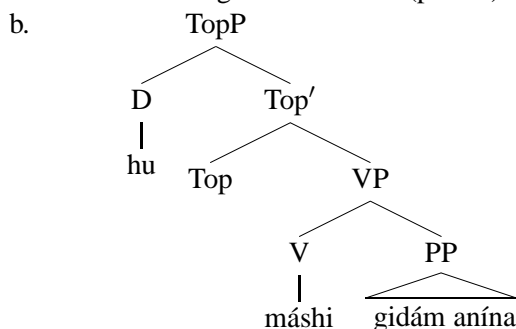
T&O mention a few other particles, *kulyum* 'always' and *yaoda* 'behold', and elements like *amis* 'yesterday', *ambukra* 'tomorrow' etc., but I would consider them adverbials. They are not purely functional markers that indicate tense or aspect. They obviously have a descriptive content of their own. There is, however, one other particle which occurs twice as a separate word and three times as (presumably) a bound morpheme in Muraz' text. This is the preverbal particle *be*:

- (8) doktór be-shúfu al-dóro
 doctor FUT-see PART-want
 'the doctor will see those who want (to be seen)' (p. 217)

When this marker *be-* occurs, the verb is translated with a future tense. We should note, however, that just like *gahed*, this particle is apparently not used obligatorily, because the source also contains examples of future tense clauses that lack this marker.

We can explain this variation in the manner described above. We can assume that pidgin speakers have available to them the same clausal structure as do speakers of a native language, if only because they are native speakers of some language themselves. If we then make the additional assumption that the functional complex is collapsed into one head, and only split into multiple heads when required, we can say that in a phrase such as (9a), the functional features do not project:

- (9) a. hu máshi gidám anína
 he walk in front us
 'he is walking in front of us' (p. 194)

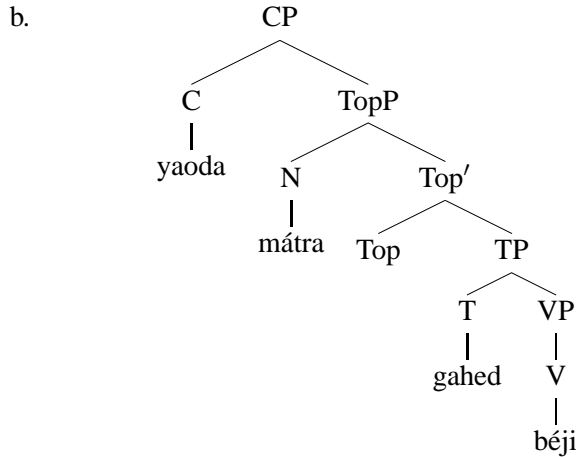


In (9b), I have placed the subject in a topic position rather than in spec,VP,⁷ in analogy with the example in (10). It is not unlikely that the subject in pidgins generally is a topic: topics have a tendency to be subjects, and pidgins lack the syntactic and morphological machinery needed to distinguish topics and (syntactic) subjects properly.

If we now take an example with the marker *gahed* we can say that the structure contains a T projection:

⁷One would normally assume the presence of *pro* in spec,VP in this case, but I leave it out for convenience.

- (10) a. yaoda mátra gáhed béji
 behold rain CONT come
 ‘there the rain is coming’ (p. 216)



Here, the element *gáhed* occupies the T position. This is an important point: under general assumptions, if a natural language projects T in one clause, it will do so in every clause.⁸ Pidgins are different, however. A head such as T can be projected on an *ad hoc* basis: in other words, it can be present in one clause and absent in another.

Another observation that speaks in favour of the assumption that functional features are only projected on an *ad hoc* basis is the fact that Turku does not show case. This is seen, for example, in the pronominal system. Table 1 lists the pronouns that T&O give:

person	singular	plural
1	ána	anína
2	inté / tu	íntukum / íntokum / íntekum
3	hú / ú / hia	úmun / úman

Table 1: Personal pronouns

The data show that these forms are used for both subject and object, and also after prepositions. In other words, there are no case distinctions, as would be expected from a pidgin:

- (11) a. ána túrdu ínte
 I expell you
 ‘I’ll send you out’ (p. 195)
- b. wédi ána fi forsambo
 send me to Fort Archambault
 ‘Send me to Fort Archambault’ (p. 195)
- c. hu máshi gidám anína
 it walks in front us
 ‘it is going in front of us’ (p. 194)

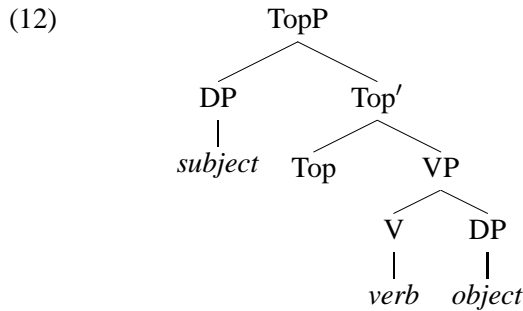
In generative syntax models, case is assigned either by independent agreement heads (AgrS for subject agreement, AgrO for object agreement),⁹ or by functional projections that are not specific to

⁸Or at least in every clause of the same type.

⁹There is a distinction between so-called structural case, which comprises nominative and accusative, and inherent case,

case but perform a sort of double role. T, for example, is generally regarded as the head that assigns nominative case. In either version of the model, however, the assumption that functional heads are not present predicts that case will not be marked.

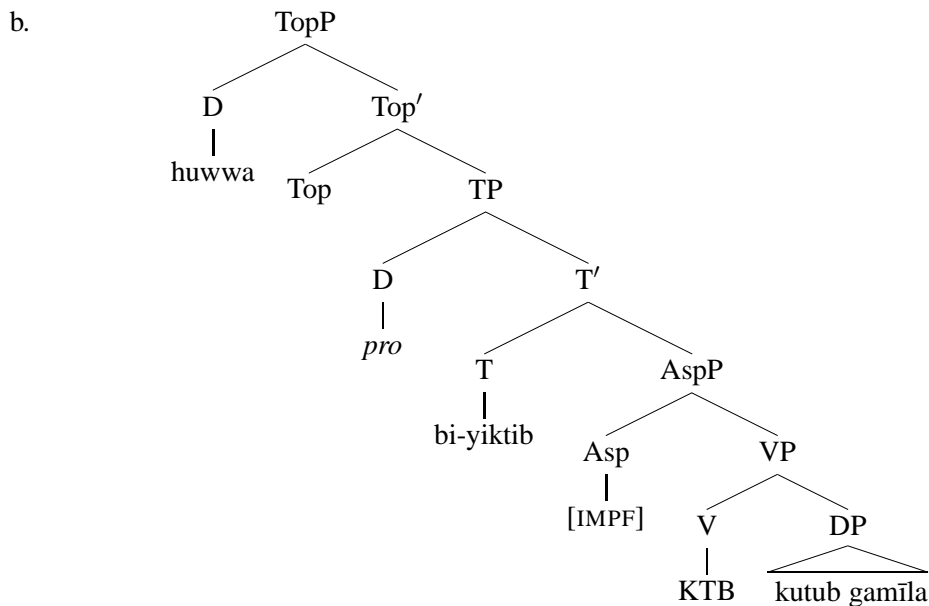
To summarise, the basic clause structure for Turku is (12):



3 Modern Arabic colloquials

Having developed a model of the clause structure of Turku, let us now look at Cairene Arabic, being a model for the modern colloquials. The first thing to note is that the clause structure must be more elaborate.

- (13) a. huwwa bi-yiktib kutub gamīla
 he HAB-he.writes books beautiful
 'he writes beautiful books'



First of all, colloquial Arabic has consistent tense/aspect marking. This means that a T head is consistently present. In all likelihood, T is not the only functional head, because there are actually two functional morphological elements on the verb: the prefix *bi-* and the imperfective marking consisting

which comprises the other cases that exist in language. (Although it is often argued that genitive must be considered a structural case as well.) Only structural case is assigned by Agr heads.

of the personal prefix and the vowel pattern.¹⁰

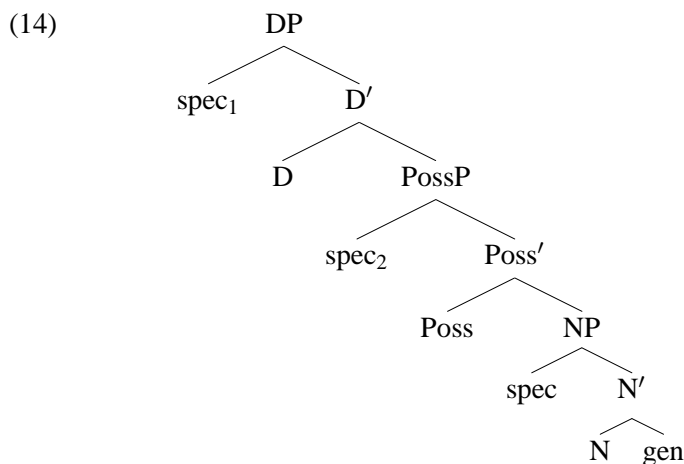
Because the verb stem, the imperfective marking and the tense prefix form one phonological complex, it is said that the verb in V moves to Asp and then to T, so that the morphological elements can be joined into one form. Such movement is typical for language, but it is generally absent in pidgins. In Turku, movement cannot take place for the simple reason that the heads to which V would move are not projected. Even in the cases where T or some other head is projected (as is the case in the example in (10) where T is present in the form of *gahed*) no movement takes place, since the verb is still morphologically independent from the element in T.

Another indication of the existence of functional heads is the fact that pronouns in colloquial Arabic take different form depending on their case: nominative pronouns are independent forms, which accusative (and genitive) pronouns are suffixes. This suggests that they are assigned case, which requires the presence of functional heads.

We see, then, that the differences between the clause structure of Turku and that of colloquial Arabic is substantial. However, we also see that the clause structure of Turku is very basic, broken down to the absolute minimum. This means that it is not entirely inconceivable that reconstruction of the clause structure — through creolisation and decreolisation — leads to something close to an Arabic colloquial. To see why this process is nonetheless unlikely to have taken place, we must look at the domain of the noun phrase.

4 The noun phrase

The noun phrase, like the verb phrase, is composed of a lexical base plus a functional shell. The lexical base in this case is N, the noun. A typical example of the functional shell of a noun phrase is (14):



The head D stands for the determiner, i.e., it is a projection of the definiteness feature. Poss is the marker of possession: if a noun is possessed (i.e., if it has a genitive/possessive complement) this head represents the possessive marker. There are other features in the noun phrase that are often assumed to project heads. These are case, number and gender.¹¹ For some, demonstratives form another class

¹⁰The exact labelling of these elements is a topic that I will not go into here. I will just assume that the imperfective form is an aspect marker and that the prefix *bi-* is a tense prefix, although it is obvious that the matter is not that simple.

¹¹Classifiers are also sometimes considered to be separate functional projections. Others, however, equate the classifier with D.

of functional projections, while others claim they are D elements.

We can describe the functional shell of the noun phrase in the same way as the functional shell of the clause: it contains a number of syntactic features, which are all able to project but do not necessarily do so. In a pidgin, we predict that some features may be present, but not in a consistent manner.

If we look at the Turku data again, we in fact see this pattern. There are a number of syntactic features that can be expressed in noun phrases, and we see them turn up occasionally, but by no means obligatorily or consistently.

Perhaps the most salient functional element in the noun phrase is the determiner. Turku has something that looks like a determiner, the element *da*.¹² T&O remark that it is not entirely clear whether this element should be seen as a definite determiner or as a demonstrative: Muraz sometimes translates *da* as a determiner, with French *le, la*, and sometimes as a demonstrative *ce, cette*:

- (15) a. fi híle da mardán num fí wál
in village DEF sick sleep exist Q
'is there sleeping sickness in the village?' (p. 206)
- b. álme da awán
water DEF bad
'this water is bad' (p. 207)

Like all functional elements, *da* is optional. This is demonstrated by the following example, which directly precedes (14a) in the original text, and in which the same word *híle* 'village' is used without *da*:

- (16) bojíni fí geríb híle wála
tsetse exist near village Q
'are there tsetse flies near the village?' (p. 202/207)

Here, the noun *híle* must be definite, since it is the only sensible interpretation of the sentence, but this definiteness is not marked. T&O do note that indefinite nouns are never modified by *da*.

But more important than its exact meaning is the placement of *da*. It always appears at the end of the noun phrase, following even relative clauses:¹³

- (17) cár al-gáhed béji dá
month PART-CONT come DEF
'the coming month' (p. 208)

The placement of *da* is quite remarkable. The general idea on noun-final determiners is that they appear as suffixes on the noun, and therefore could not follow any other material, such as adjectives or relative clauses. (See, e.g., Cinque, 1996) There is, however, an interesting parallel here with the clause-final interrogative marker that we met in the previous section. It is often claimed (see, for example, Abney (1987) or Szabolcsi (1994) for such proposals) that there is a parallel between functional elements in the noun phrase and in the clause. D would then be the parallel of C.¹⁴

If we assume that the element *wála* is a force marker and that it appears in C position, we can relate the positioning of *da* to the positioning of *wála*. In an abstract way, these elements are of a

¹²T&O indicate that there is one instance of *dolda*.

¹³On the marker *al-*, see below, section 4.3.

¹⁴There are also claims that D is in fact the parallel of T, and that there is a case projection, sometimes abbreviated as K, that is the noun-phrase equivalent of C.

similar type, and there is a tendency in Turku to place elements of this type at the end of the phrase that they are part of. This is, of course, a tendency that is markedly different from Classical Arabic and the modern Arabic colloquials, which show a strong tendency for placing such elements before the phrase they modify.

4.1 Possessives

Turku has a possessive marker *ana*. Practically all possessives are marked with it. T&O say that there are only two examples of an unmarked possessor:¹⁵

- (18) a. hīle ítokum
village you.PL
'your (pl) village' (p. 206)
- b. kanamáye áshra fí bakán ána
goats ten exist place my
'there are ten goats at my place' (p. 211)

The possessive phrase with *ana* follows after the possessed noun:

- (19) límpo ana nádem wáid
tax of man one
'the tax of one man' (p. 205)

In Kremers (2003) I argue that there are basically three types of genitive/possessive constructions. The first one uses an overt Poss head. The second makes use of a covert Poss head that assigns (usually overt) genitive case. This genitive case can be marked by a case ending, or by a preposition-like dummy case marker. The third method consists of using a predicative genitive marker.¹⁶

The data are very limited, which means that a proper analysis is not possible. However, we can get some indication of the type to which *ana* belongs. Let us first take a quick look at colloquial Arabic, before discussing *ana*.

Colloquial Arabic dialects usually have two types of possessive/genitive marking. The first is generally known as the construct state. In this, the head noun (i.e., the possessed noun) takes a special form. The noun does not take a definite article, and words with a feminine ending change this ending from *-a* to *-it*. The modifying noun (the possessor) immediately follows the head noun. One can say that it is assigned abstract genitive case. Abstract, because it does not show on the noun itself.

In Kremers (2003) I analyse the construct state as a structure containing a Poss head, which also contains definiteness features. This Poss head has some sort of affixal phonological form, which causes the head noun to move to it. Poss also assigns genitive case to the possessor.

The second possessive construction that colloquial Arabic has uses a predicative genitive marker. In Egyptian Arabic, this is *biā^c*. There are two characteristics of predicative genitive markers. First,

¹⁵There is, however, another example in their paper:

- (i) shīli rangáye laam wadín ana pokter anína
take basket meat other of porter our
'take some baskets of meat of our porters' (p. 203)

It is not impossible, though, that the form *anína* is ambiguous between a personal and a possessive pronoun, cf. section 4.2.

¹⁶I will not go into the different characteristics and properties of each here. For an extensive discussion, see Kremers (2003).

they sometimes agree with the head noun in gender and number, and second, they can be used across a copula. Egyptian *btā^c* shows both characteristics.

There is at least one example in the corpus that indicates that *ana* is a predicative possessive marker:

- (20) *ágaba fádal da ana nas híle*
 rest remain DEF of people village
 ‘the rest that remains belongs to the villagers’ (p. 208)

The subject in (20) is *ágaba fádal da*, which contains a noun, *ágaba*, plus a relative clause (here unmarked) consisting of only a verb, *fádal*. The predicate is the sequence *ana nas híle*, ‘of the villagers’.¹⁷ The head of this predicate is *ana*, the possessive marker. This construction is very similar to the following structure in Egyptian Arabic:

- (21) *il-kitāb da btā^c-i*
 the-book this of-me
 ‘this book is mine’

In Kremers (2003) I claim that this construction is indicative of a predicative genitive marker. For this reason, I will assume that *ana* is of this type.

4.2 Possessive pronouns

Turku has a set of possessive pronouns. T&O give a table of these, which I quote in table 2.¹⁸

person	singular	plural
1	anai / anahi	?
2	ana-ki	ana-kum
3	ana-hu	ana-hum / anam

Table 2: Possessive pronouns

The possessive pronouns always appear after the noun:

- (22) a. *gesh ana-ki*
 straw your
 ‘your straw’ (p. 205)
 b. *pokter anah-i wáid*
 porter my one
 ‘one of my porters’ (p. 204)

There are in principle a number of ways to classify possessive pronouns such as those in table 2, depending on their behaviour and properties. A first possibility is that these pronouns are genitive forms of the personal pronoun. However, in a pidgin, which lacks case marking, such a classification is unlikely.

An alternative is to assume that the personal pronouns are in fact D elements. In English, for example, the possessive pronouns *my*, *your*, *his*, *her*, etc. are of this type: they take the position of

¹⁷The noun combination *nas híle* can be considered a compound. T&O show that these are not uncommon in Turku.

¹⁸No form has been attested for the first person plural. As I indicated earlier, it may be the case that the form *anína* is ambiguous between a personal and a possessive pronoun.

the determiner. This means for one that the determiner and the possessive pronouns are mutually exclusive: they cannot co-occur.

This is true for English, but it does not seem to be true for Turku, given the following example:

- (23) kantón anaki da
district your DEF
'your district' (p. 204)

This example contains both a possessive pronoun and a definite determiner.¹⁹ It seems implausible, then that we must really analyse the possessive pronouns as D elements.

The final possibility is that these possessive pronouns are predicative possessive markers, similar to the English *mine, yours, his, hers*, etc. This makes sense for at least two reasons. First of all, we have already seen that the possessive marker *ana* is best analysed as a predicative possessive marker. The forms in table 2 are obviously related to *ana*. Second, a predicative possessive marker is not a functional element, contrary to a possessive D element, which is more in line with the nature of a pidgin.

4.3 Relative clauses

A relative clause in Turku can be marked with the formal marker *al*, but does not have to be. It does not seem to be the case, however, that *al* only appears when the head noun is definite, nor is it the case that it *always* appears when the head noun can be considered definite. Furthermore, as already noted above, the definite article or demonstrative *da* can appear after the relative clause, if the head noun is definite.

A typical relative clause is (24):

- (24) cár al-gáhed béji dá
month AL-CONT come DEF
'the coming month' (p. 208)

In this example, both *al* and *da* are present. Given the fact, however, that *da* also appears as a definite marker in noun phrases that do not have a relative clause, I will assume that *da* does not have any special function when a relative clause is present.

The only relative marker, then, is *al*. To get a more complete picture of this element, it is important to note that the same element also introduces headless relative clauses. T&O give the only two examples found in the corpus:

- (25) a. doktór be-shúfu al-dóro
doctor FUT-see AL-want
'the doctor will see those who want (to be seen)' (p. 207)
b. fishan árfu al-dóro sáo ásker
so that know AL-want make soldier
'so that he knows who can become a soldier' (p. 207)

Furthermore, T&O give an example of *al* marking an adjective, the only token in the corpus:

¹⁹We have seen that the element *da* sometimes has the value of a demonstrative, and it true that in some languages the demonstrative and a possessive D element can co-occur (e.g., in German *dieses meines Buch* 'this my book'), but since the translation of the current example does not use a demonstrative, I will assume that *al* is simply a determiner here.

- (26) kútu al-awán sík
 put *al-bad* apart
 ‘put the bad ones apart’ (p. 207)

The adjective here is used without a noun, that is, it is used as a substantive. In order to do this, the adjective must apparently be marked with *al-*.²⁰ In this way, *al* differs from *illi* in colloquial Arabic, which cannot be used as in (26).

As T&O note, all relative clauses have a subject or locative extraction site, and do not have a resumptive pronoun.²¹ It is not certain what conclusion must be drawn from that fact. The data show that a subject pronoun can sometimes be dropped if the referent is obvious from the context. It could be that that is what happens in (subject) relative clauses.

Alternatively, it is possible that the relative clause in Turku behaves more like relative clauses in European languages. In for example English, the relative marker is (generally) a *wh*-element that moves from the extraction site to clause-initial position, generally assumed to be spec,CP:

- (27) this book, [_{CP} which_i [_C \emptyset] I wrote t_i two years ago] . . .

English has another structure: instead of using a *wh* relative marker, it can use *that*. Here it is assumed that there is a phonologically null operator, usually designated with *OP*, that moves from the extraction site to spec,CP, whereas the C position is filled with *that*.²²

- (28) the book [_{CP} OP_i [_C that] I wrote t_i two years ago] . . .

The structure that colloquial Arabic dialects use is somewhat different, although based on the same elements. The relative marker *illi* is in C position, like the English element *that*. There is, however, no extraction: the operator that is generated in the relative clause is instantiated as a resumptive pronoun in the position in which it is generated. To take an Egyptian example:

- (29) il-kitāb [_{CP} \emptyset [_C illi] katabt-u min sanatēn] . . .
 the-book that I.wrote-it before two-years . . .
 ‘the book that I wrote two years ago’

From the data it seems that the Turku construction most resembles the English structure with *that*. Like English *that* it can be left out, and there is no resumptive pronoun. If we furthermore take into account that *al* is not restricted to nor obligatory with definite head nouns, like English *that*, we can see that the parallel is very strong.²³

²⁰Cf. English, where an adjective alone can also not be used substantively: it requires the addition of *one*.

²¹The only exception to both statements is the following example:

- (i) nas mardan-ín ma badoró-um
 people sick-PL not want-them
 ‘as for the sick people, we don’t want them’ (p. 194)

This example is typically not pure pidgin, but contains elements of colloquial Arabic. Among these are the plural ending on *mardanín*, the verbal prefix *ba-* and suffix pronoun *-um*. Since Tosco and Owens (1993, 217) note that Muraz’ data show variation between an acrolect (Sudanese Arabic) and a basilect (Turku), which are sometimes mixed, I will not take such data into account.

²²Note that *that* can also be left out: *the book OP \emptyset I wrote t_i two years ago. . .*

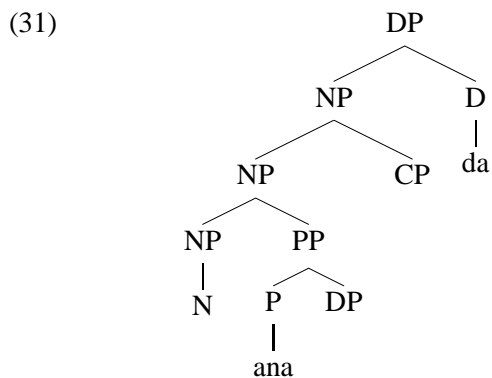
²³Although we must not forget that the scantness of the evidence does not allow us to draw too strong conclusions.

4.4 Noun-phrase structure

Let us now look at the structure that is needed to describe the phenomena discussed in the previous sections. We have seen that the order of the noun phrase is the following:²⁴

(30) noun - possessor - relative clause - determiner

I will assume that the structure does not overtly project a Poss head, since there is no strong evidence that there is one, and because it seems unlikely that a pidgin would. The possessor then is just a complement of the noun, marked with *ana*. The relative clause is adjoined to the noun, forming the NP. Then the determiner takes this NP as a complement, creating the following abstract structure:

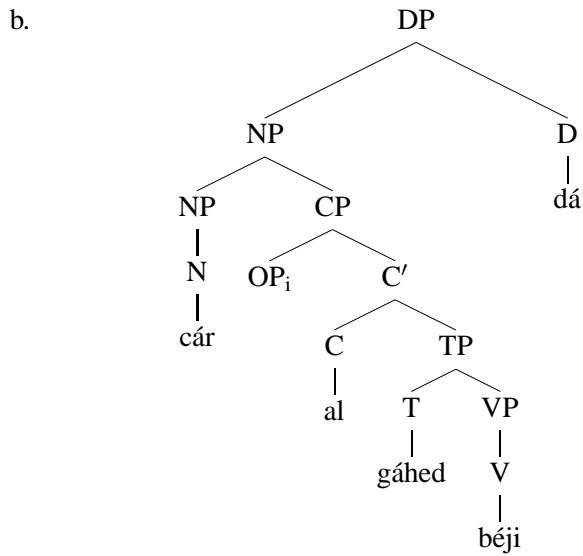


The reader will notice that the tree in (31) is left-branching. This is the only way in which the structures presented here can be described. The structure of the noun phrase in Turku is such that all types of modifiers are placed after the noun. This is not only true for the modifiers discussed in this paper, but also for other types, such as adjectives and numerals.

If we apply this to some concrete examples, we get the following:

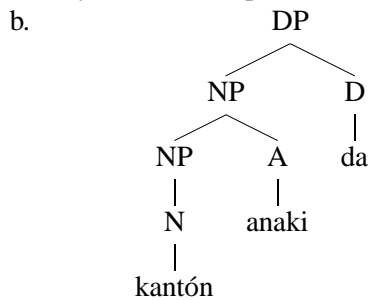
(32) a. *cár al-gáhed béji dá*
 month AL-CONT come DEF
 'the coming month' (p. 208)

²⁴This structure is taken from T&O, who actually present a somewhat more complicated structure, taking also adjectives and numerals into account. T&O actually put the possessor after the relative clause, but they give no example on which this ordering is based.



(32) contains a noun phrase with a relative clause. The clause itself is right-branching, but the noun phrase is left-branching, which gives the somewhat odd-looking tree structure.

- (33) a. kantón anaki da
 district your DEF
 'your district' (p. 204)

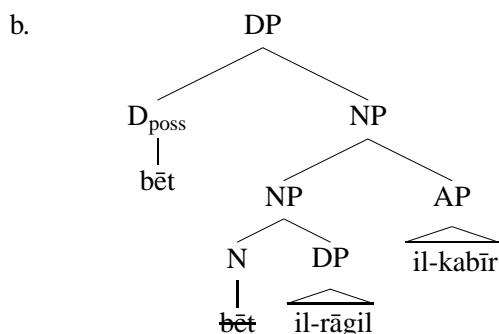


(33) is simpler in structure. The predicative possessive (which I have labeled A for *adjective*) is adjoined to the NP. Again we see that the structure is left-branching.

If we take a noun phrase from Egyptian Arabic for comparison, we see a number of differences:²⁵

- (34) a. bēt il-rāgil il-kabīr
 house the-man the-big
 'the man's big house'

²⁵The structure given here is based on the structure that I assume in Kremers (2000b) for Standard Arabic.



The major difference between Turku and colloquial Arabic is the fact that Turku is predominantly left-branching, whereas colloquial Arabic is predominantly right-branching.²⁶ This left-branching property of Turku, however, is not a typical pidgin feature. Pidgins tend to be right-branching languages. (reference)

Another difference between Turku and colloquial Arabic is the use of a Poss feature. In a construct state in colloquial Arabic, such as the example given here, the head noun moves to the D position. This D position contains not only a [DEF] feature, but also a [POSS] feature, hence the subscript *poss*.²⁷ Turku makes no use of any syntactic Poss feature. It uses a predicative marker to express possession.

I could also quickly mention the fact that Turku has no agreement between the head noun and the adjective.²⁸ I will not go into this matter here. (See Kremers (2000a) for a discussion of adjective agreement in Arabic.)

5 A pidgin/creole origin of Arabic colloquials

In order to understand if Arabic colloquials could have developed from creoles and ultimately pidgins, we need to establish if the pidgin structure could in principle develop into the structure of the colloquial given in the previous section.

There are some substantial differences between the two structures. The colloquial structure projects a number of syntactic features onto separate heads, more specifically, tense and aspect. Furthermore, I have assumed that *v* is present in the structure. Apart from the presence of these heads, which do not (obligatorily) occur in the pidgin structure, there is another difference between the two structures. Arabic colloquials make use of movement: the verb moves to incorporate tense and aspect features, the subject and the object move so that they can be assigned case. This movement is in fact the way in which the model describes complex word forms. A verb form can contain markers for tense, aspect and agreement because it can move.

It is in principle not impossible that a pidgin structure develops into something more elaborate. In fact, a creolization process is exactly that. A pidgin can project syntactic features more or less at random. In a “native” language, that is, a language which unlike a pidgin has native speakers, the features that project independent heads and the conditions under which they do so are fixed. For example, in an English main clause, T is always projected. Furthermore, in interrogative clauses, C

²⁶To be more precise: in Turku, heads follow their complements, whereas in (colloquial) Arabic, heads precede them. In both languages, specifiers and adjuncts always follow their heads.

²⁷More accurately, this position should be labeled Poss_{def}, because I believe it is primarily a Poss head, which is also marked for definiteness.

²⁸T&O remark that there are cases of such agreement in the corpus, but they show the characteristics of acrolect influences.

is always projected. A creolization process can be described as fixing these characteristics of the language: which features can project independent heads, and under which conditions do they do so.

Theoretically, then, it is not impossible that modern Arabic colloquials have developed through a creolization/decreolization process. However, if we look at the actual language forms, we must say that such a development is in fact highly unlikely.

In the previous sections, we have seen that the structure of Turku differs considerably from its source language (colloquial) Arabic. The clausal structure of Turku is typical for a pidgin: it lacks obligatory expression of functional elements, it lacks any kind of agreement and case marking, all of which are present in the superstrate. In the noun phrase, Turku shows much substrate influence. The noun-phrase structure is not typical for pidgins, especially in its left-branching character, but seems strongly influenced by substrate languages. We see, then, that Turku is in fact a mix between typical pidgin strategies and substrate influence.

A creole based on such a pidgin would most likely show very similar features.²⁹ Now considering the large area in which Arabic colloquials have developed, one can only assume that the pidgins and creoles on which these colloquials are supposedly based must have shown an enormous variety. Like Turku, all of these pidgins/creoles would have shown a mix of typical pidgins/creole properties and substrate influences, but all in a very different manner.

We know that the modern Arabic colloquials are all very similar in structure. In fact, the clausal structure of Egyptian Arabic given in (13) is typical for colloquials in general. This presents an important incongruity in the pidginization/creolization theory: how is it that a group of languages with such a striking structural similarity has grown out of what must have been such a diverse group of creoles. This becomes even more surprising if we take into account that not only the creole base would have differed, but also the decreolization processes that took place in the various countries must have differed. This would be an extra source for divergence between the various colloquials.

To make these comments more concrete, let us take a short look at the verbal system. A typical creole verbal system (see, e.g., Holm, 2000) is characterized by an invariable verbal stem and a small set of verbal prefixes, basically a grammaticalization of what we have seen for Turku. A typical colloquial Arabic verbal system has two verb stems, which are conjugated for person. One of these verb stems has verbal prefixes.

These differences between the two systems are very obvious, but there are a few others which are less obvious. First, the prefixes of a creole verbal system can usually be combined. However, the prefixes of the verbal system of an Arabic dialect cannot be combined. An imperfective verb has only one prefix, it cannot have two.

Another difference is that the unmarked verb form in a creole (that is, the form without any prefixes) has no fixed temporal reference. As Holm (2000, 175–178) makes clear, the tense interpretation of an unmarked verb depends on the context and on the lexical meaning of the verb.³⁰

If we equate the verbal prefixes of the imperfective in colloquial Arabic, with the verbal prefixes of creoles, as I have done here for the sake of argument, then we could consider the perfective of colloquial Arabic the parallel of the unmarked verb form of creoles. However, we see that the use of the perfective in colloquial Arabic is rather different from the use of the unmarked verb in creoles.

There is another aspect that is important. As I just stated, all Arabic colloquials have verbal

²⁹In fact, the Arabic creole Nubi (reference!) is (presumably?) based on pidgin varieties that show much similarity with Turku, and the two languages are indeed very alike.

³⁰That is, in a narrative context which is known to be in the past, past tense does not have to be marked. Furthermore, unmarked forms of stative verbs tend to be interpreted as present tense, because a state is ongoing, whereas unmarked forms of eventive verbs tend to be interpreted as past tense, because an event takes place at a specific point in time.

prefixes that combine with the imperfective. In this, they are structurally very similar.³¹ However, the specific prefixes that are used, that is, their phonological forms, differ from one dialect to another.

The difference in phonological form indicates that the various dialects developed relatively independently. This is not unexpected, given the large area and the lack of means of mass communications, such as we have today. The structural similarity indicates that the development of the different dialects, although independent, took place in very similar fashion.

It does not seem clear, however, how such a large number of independent developments could have resulted in such similar structures for the various colloquials if their starting points, that is, the pidgins/creoles that supposedly formed the input for the development, were most likely so very different.

This combination of independent developments and similar structural output seems to be possible only if *both* the circumstances *and* the starting points of the developmental processes at various locations were similar to a large extent. This is much more consistent with a situation such as that proposed by (mohammed!), who argues that the modern colloquials are the result of a partially successful attempt at unguided second language learning of Arabic by the original inhabitants of the conquered lands.

The input for such a process of unguided second language learning is the Arabic spoken by the Arab conquerors. Their Arabic was by any assumption much more uniform than a large collection of creole languages would have been.³²

6 Summary

In this paper, I have given an analysis of the Arabic-based pidgin Turku. I have shown that the structure of Turku differs considerably from colloquial Arabic. In the clause, Turku has typical pidgin-like features. They also show up in the noun phrase, but there, the influence from substrate languages is more clear.

I have described the structure of the pidgin with a chomskian model, which makes the extra assumption that each grammatical feature can project a functional head, but does not have to. In this way, the structure of Turku can be described as a special case of a more general grammatical structure, rather than as something completely different from “regular” languages.

Finally, I have argued that it is unlikely that all modern Arabic colloquials have a pidgin/creole origin. The structural similarities suggest that the starting point for the development of the colloquials was much more uniform than a large collection of creoles would likely be.

References

- Abney, S. (1987). *The English Noun Phrase in its Sentential Aspect*, PhD thesis, Cambridge, Mass: The MIT Press. MIT Working Papers in Linguistics.
- Chomsky, N. (1995). *The minimalist program*, Cambridge, Mass: The MIT Press.
- Cinque, G. (1996). The ‘antisymmetric’ programme: theoretical and typological implications, *Journal of Linguistics* **32**: 447–464.

³¹Though not identical. Some dialects have two prefixes, others may use three.

³²The presence of pidgin/creole-like features in the modern colloquials is not surprising in this development. It is often claimed that one stage in a process of unguided second language learning is a so-called *basic variety*, which shares many features with pidgins, such as lack of flexion and agreement, the absence of functional markers, etc. (references)

- Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*, New York: Oxford University Press.
- Giorgi, A. and Pianesi, F. (1997). *Tense and Aspect: from semantics to morphosyntax*, New York/Oxford: Oxford University Press.
- Holm, J. (2000). *An introduction to pidgins and creoles*, Cambridge: Cambridge University Press.
- Kremers, J. (2000a). Concord. ms. University of Nijmegen, draft.
- Kremers, J. (2000b). A recursive linearization approach to Arabic noun phrases. ms. University of Nijmegen.
- Kremers, J. (2003). *The noun phrase in Arabic: a minimalist approach*, PhD thesis, University of Nijmegen. LOT Dissertation Series 79.
- Szabolcsi, A. (1994). The noun phrase, in F. Kiefer and K. Kiss (eds), *The syntactic structure of Hungarian*, Vol. 27 of *Syntax and Semantics*, San Diego: Academic Press, pp. 197–174.
- Tosco, M. and Owens, J. (1993). Turku: a descriptive and comparative study, *Sprache und Geschichte in Afrika* **14**: 177–267.
- Versteegh, K. (1984). *Pidginization and creolization: the case of Arabic*, Vol. 33 of *Current Issues in Linguistic Theory*, Amsterdam: John Benjamins.