How Different are Arabic Dialects from Each Other and from Classical Arabic?

The differences between Standard Arabic and the dialects and between the Arabic dialects themselves have raised many questions about the relation between the spoken dialects and Standard Arabic and between the dialects (Owens 2006). Some proposals have gone as far as to suggest that some dialects, such as the Maghrebi dialects, are pidginized (subsequently creolized and decreolized) forms of Arabic (Versteegh 1984). Other proposals have assumed a sedentary/Bedouin split with the assumption that the latter are more conservative and likely have been transmitted through native first language acquisition rather than second language acquisition as has been proposed for western dialects (McWhorter 2007). For theoretical and computational linguistics, the similarities and differences between the Arabic varieties have the potential to shed important light on language variation and related issues of parameters and language change, and also contribute to the growing field of language and dialect distance (dialectometry). However, to engage these issues it is imperative to have systematic and in-depth descriptions and analyses of various patterns in different varieties. In this paper, we will report on an-going project that aims to provide such descriptions and analyses. The descriptions are based on a variety of data sources, elicited patterns and narratives from native speakers, published sources, and electronic corpora some of which developed by our research team. We will take sentential negation as case study but in the process we will also discuss clause structure, word order, position of the verb, polarity licensing, and the position of the subjects, all prominent topics in Arabic linguistics and theoretical linguistics in general.

In all varieties, there are two sentential negative patterns that in general occur either in the context of verbal or non-verbal predicates. Restricting our attention to the frequent patterns, Standard Arabic has the largest set of sentential negative markers (*laa*, *lan*, *lam*, *laysa*, *maa*) while the dialects are restricted to three (*maa*, *muš/miš/maš/maši/muu/mub*, *laa*) or just two (*maa* and *maš/maši*). Though the size of inventories differs the syntax appears to be the same. In all varieties, the sentential negative marker seems to be sensitive to the dependency between negation, tense, and the predicate. We will argue that in all the dialects under study (Egyptian, Jordanian, Moroccan, Qatari, and Yemeni), the negative marker occupies a position that dominates tense which in turn dominates (possibly indirectly) the projection of the predicate (Shlonsky 1997, Soltan 2007). The main differences, however, are limited to four main options or parameters:

(i)   The ability of negation to attract tense without the verb.
(ii)   The type of lexical category that can merge with negation.
(iii)   The option of leaving the pronominal subject negation (in TP).
(iv)   The ability of negation to carry agreement.

In Standard Arabic, negation can host tense without the verb (yielding *lam* and *lan*) or with the verb (yielding *maa*), it can carry agreement (yielding *laysa*), and the pronominal subject can remain lower (yielding *maa+NP* clusters). In the dialects, only lexical categories can merge with negation, hence only *maa* or *ma-š* in tensed contexts but in some dialects the pronominal subject can remain lower (yielding the *muu*, *mii*, and *mub* negatives in Gulf dialects) but in other dialects it cannot (Benmamoun and Al-Asbahi 2013).

Interestingly,   the language with the largest inventory and which allows for using both options of a parameter is Classical Arabic. For example, both merger with tense only and with tense and the verb are allowed, hence both *lam yaktub* and *maa katab* are available. However, in all varieties there are at least two forms of negation, a dependent form and an independent form which has to do with clause structure and the interaction between the predicate, tense, and negation. As we will demonstrate this system has been reproduced and continues to be reproduced through the process of grammaticalization, which supports the arguments for a shared syntactic core. This has two important implications. First, that the spoken dialects are all descendants of Classical Arabic or a related variety. Second, at a deeper syntactic level, the major dialects, including Maghrebi dialects, are fundamentally similar. The clause structure is the same and the parameters across the dialects seem to be similar as well. The differences have more to do with the options that a variety may take, which is typical of how language works crosslinguistically. There do not seem to be any syntactic arguments for the pidginization hypothesis, though the transmission of Arabic as L2 has resulted in some syntactic changes, particularly at the level of word order as we will discuss. Clause structure and how the functional categories interact with each other have been more resilient which is not what one would expected were pidginization a determining factor.