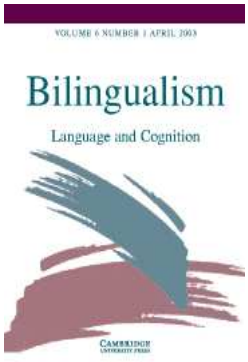


# Revisiting Arabic diglossic switching in light of the MLF model and its sub-models: the 4-M model and the Abstract Level model

NAIMA BOUSSOFARA-OMAR

Bilingualism: Language and Cognition / Volume 6 / Issue 01 / April 2003, pp 33 - 46
DOI: 10.1017/S1366728903001032, Published online: 12 May 2003

Bilingualism: Language and Cognition 6 (1), 2003, 33–46 © 2003 Cambridge University Press DOI: 10.1017/S1366728903001032 

# Revisiting Arabic diglossic switching in light of the MLF model and its sub-models: the 4-M model and the Abstract Level model*

NAIMA BOUSSOFARA-OMAR
*The University of Kansas*

*The goal of this paper is twofold. First, I wish to discuss two sets of problematic cases that arose when I applied Myers-Scotton's (1993) Matrix Language Frame model of codeswitching to Arabic diglossic switching (Boussofara-Omar, 1999). The first set involves the co-occurrence of system morphemes from both varieties of Arabic within a single CP. The second set concerns CPs in which the word order is that of the dialect but the system morphemes are from Standard Arabic, and CPs in which clashes occur between the subcategorization restrictions of the two varieties of Arabic participating in diglossic switching. The Matrix Language Frame model neither predicts nor offers explanations for either case. Second, in an effort to provide an explanation to their occurrence, I revisit the same problematic sets in light of Myers-Scotton's sub-models: the 4-M model and the Abstract Level model (Myers-Scotton and Jake, 2000, 2001), the latest refined version of the MLF model.*

## 1. Introduction

The majority of previous studies on the alternate use of the two varieties of Arabic was devoted to dividing the Arabic continuum into 'clearly marked' middle varieties. Some researchers (e.g. Blanc, 1960; Badawi, 1973; Meiseles, 1980) identified a constellation of intermediate levels that range from Modern Standard Arabic or a more traditional Classical Arabic to a 'plain colloquial', 'vernacular' or 'colloquial of the illiterate'. Others (El-Hassan, 1977; Mitchell, 1980; Mahmoud, 1986) posit a single intermediate variety, Educated Spoken Arabic (ESA) that is "created and maintained by the constant interplay of written and vernacular Arabic" (Mitchell, 1980, p. 13). These efforts were predominantly impressionistic in nature. They failed to provide a coherent theoretical framework to handle what Meiseles (1980, p. 120) calls the "uncharted sea of intermediate shades" and to explain the nature of the structural constraints on the 'mix' of the two varieties of Arabic.

In an effort to develop the notion that what is often called the 'third' language or 'middle' varieties in Arabic is, in fact, best analyzed as Arabic diglossic switching between Classical Arabic/Modern Standard Arabic (CA/MSA, which I subsume under the term fuṣḥaa) and the dialect (Walters, 1996a, b), I applied (Boussofara-Omar, 1999) Myers-Scotton's (1993) Matrix Language Frame (MLF) model of codeswitching. The study investigated the structural constraints on diglossic switching between fuṣḥaa and Tunisian Arabic through a linguistic analysis of 17 public political speeches (recordings lasting approximately 14 hours) that the former President of Tunisia, Habib Bourguiba, delivered between the years 1956 and 1968.

Analysis of the 'third' language and 'middle' varieties in terms of the MLF model shows that there is no variety conventionalized as either a 'third' language or as a 'middle' variety. It demonstrates that what is being conventionalized is really a pattern of switching between the two varieties of Arabic with the dialect as the matrix language/variety (ML) into which elements (i.e. content morphemes) of fuṣḥaa are embedded in ML + EL (embedded language) constituents. The data examined in light of the MLF model indicate that the 'mix' of the two varieties of Arabic is governed by structural constraints, and hence that the switching patterns are systematic and predictable.

The efforts to apply the MLF model to Arabic diglossic switching, however, gave rise to two sets of problematic switching patterns. The two sets pose a potential problem to the MLF model in the sense that it neither

* I gratefully acknowledge financial support from the University of Kansas. Thanks go to Keith Walters, Pieter Muysken and especially to the reviewers for their comments on an earlier draft of this paper.

Address for correspondence
The University of Kansas, Department of African & African-American Studies, 9 Bailey Hall, 1440 Jayhawk Boulevard, Lawrence, Kansas 66045-7574, USA.
**E-mail**: *nomar@ku.edu*

predicts their occurrence nor provides a satisfactory explanation for them. The first set involves utterances in which one finds system (grammatical) morphemes from both varieties of Arabic within a single CP, a switching pattern that, according to MLF model (Myers-Scotton, 1993), and, more precisely, the System Morpheme Principle, cannot occur. The second problematic set involves utterances in which the word order and subcategorization restrictions are those of one variety (Tunisian Arabic), but in which one finds system morphemes from the other variety that is participating in diglossic switching (i.e. fuṣḥaa), another switching pattern that the MLF model neither predicts nor allows.

In this study, I seek to re-examine the applicability of the MLF model and its new sub-models to Arabic diglossic switching, with a special focus on the two sets of problematic cases mentioned earlier. As a frame of reference, I review, in section 2, the major claims of the MLF model, outline its aspects relevant to the present study, and focus on the refinements brought to the MLF model, as articulated by Myers-Scotton and Jake (2000, 2001) in the 4-M model and the Abstract Level model. In section 3, I first discuss each set in terms of the MLF model, and then revisit each in the light of the 4-M model and the Abstract Level model, respectively. Section 4 concludes the study by laying out a framework for rethinking our conceptualization of diglossia and its consequences for linguistic variation and change in a post-diglossic situation.

## 2. Relevant aspects of the Matrix Language Frame model

The early version of the MLF model is fully explicated in Myers-Scotton (1993). In this paper, I focus only on the major aspects of the model and its two sub-models of particular relevance to the present study.

### *2.1. The original version of the MLF model*

In her early version, Myers-Scotton (1993) argues that at the heart of the MLF model lies the notion that two interrelated hierarchies underlie the structuring of sentences containing codeswitching (CS) in ML + EL constituents: (i) the matrix language versus the embedded language; and (ii) system morphemes versus content morphemes. The languages participating in CS are both activated but they are claimed to play differential roles. The ML plays the dominant role in the sense that it determines the morpho-syntax of ML + EL constituents. In other words, it supplies the system morphemes and directs morpheme order in ML + EL constituents. While only the ML supplies the system morphemes (e.g. plural affixes, complementizers and determiners), content morphemes (e.g. nouns and adjectives) may come either from the ML or the EL. Two testable hypotheses are directly derived from these distinctions:

1. *The Morpheme Order Principle* In ML + EL constituents consisting of singly-occurring EL lexemes and any number of ML morphemes, surface morpheme order (reflecting surface syntactic relations) will be that of the ML.
2. *The System Morpheme Principle* Within ML + EL constituents, all active system morphemes are from only one of the languages participating in CS, the ML.
(Myers-Scotton, 1993, p. 83)

If supported, Myers-Scotton and Jake (2001) argue, these hypotheses would empirically demonstrate the value of the ML as a theoretical construct. While the MLF model seems to handle a significant amount of data from 'classic' CS[1] and Arabic diglossic switching, as I have demonstrated in Boussofara-Omar (1999), it seems that the two basic principles that sustain the MLF fall short of explaining the problematic data under discussion in this paper. With the extension of the MLF model, Myers-Scotton and Jake (2001, p. 87) continue to argue that "[t]he ML is a theoretical construct universally applicable to all CS data sets, regardless of the specific language pairs involved".

The 4-M model and the Abstract Level model seem to fall short of providing coherent explanations for the problematic cases that the MLF model either disallowed or allowed but did not explain. Both sub-models seem to raise some issues that need to be addressed, as I will demonstrate in sections 3 and 4. Let me first explain the adjustments that Myers-Scotton and Jake (2000, 2001) brought to the MLF model to handle 'atypical' CS and language contact data better.

### *2.2. The MLF model adjusted*

In the literature on codeswitching, some 'atypical' CS data (as Myers-Scotton and Jake (2001, p. 94) term them) were reported. One of the most salient was double morphology (i.e. a case in which a system morpheme from EL duplicates one from the ML, e.g. double pronoun, double infinitive and double plural), which, according to Myers-Scotton and Jake (2001, p. 97), the MLF allows but does not explain.

In order to address the 'atypical' cases that occur in classic CS, as well as other switching configurations reported in studies on language contact situations (that the MLF model neither predicts nor explains), Myers-Scotton first states (1999, p. 14) that "[t]he MLF

[1] Myers-Scotton and Jake (2000, p. 1070) define 'classic code-switching' as "the use of morphemes from two or more linguistic varieties in the same intrasentential clause (CP), with the grammatical frame derived from only one of the participatory varieties".

model applies specifically to 'classic' CS; an extended version of the model explains the structure of other language contact phenomena". To address problematic lexical structure (which Myers-Scotton calls 'complex') in language contact situations, Myers-Scotton and Jake (2000, 2001) offer an extended version of the MLF model. They add two sub-models: the 4-M model and the Abstract Level model, both of which emphasize the notion that CS is based on abstract cognitive processes. In other words, what they are arguing is that CS cannot be investigated solely by explaining surface configurations but by "go[ing] beyond explaining the observed behavior of CS itself [and] investigating the linguistic knowledge that underlies CS" (Myers-Scotton and Jake, 2001, p. 84).

### *2.3. Identifying the ML: unit of analysis*

Because the ML plays a fundamental role in framing codeswitched utterances, its identification is important. The determination of the ML remains a problematic issue despite Myers-Scotton's numerous re-articulations of the unit of structural analysis in the MLF and the 4-M models.

Although the MLF model was initially formulated to explain structural constraints on intra-sentential code-switching, the unit of analysis in determining the ML was the discourse. Myers-Scotton (1993, p. 232) states that "[t]he ML is the language with the higher frequency of morphemes in a discourse sample in which CS occurs". The question of how large a discourse sample should be remains vague as Myers-Scotton (1993, p. 68) suggests that "a discourse sample must mean more than one sentence". The question "How much more?" remains unanswered. In an effort to be more specific, she re-articulates her definition of discourse suggesting that "[a] discourse sample of at least two sentences (within the same turn or across speakers) is a minimum" (1995, p. 238). In the latest refined version of the MLF, however, Myers-Scotton and Jake (2001, p. 89) argue: "Although the MLF was originally a model of intra-sentential CS, it has become clear that the appropriate unit of analysis is the bilingual CP, not the sentence. This is so because the CP is the highest unit projected by lexical items." In their opinion, the notion of sentence is not accurate enough although "it is primarily within a sentence that structural constraints are more than trivial because it is within such bilingual constructions that the grammars of the two languages are in contact" (Myers-Scotton and Jake, 2001, pp. 88–89). However, the CP as the structural unit of analysis raises more issues than it solves in identifying the ML, especially in cases where both languages/varieties participating in switching supply system morphemes within the same CP.

### *2.4. The 4-M model*

The 4-M model, Myers-Scotton (Myers-Scotton and Jake, 2000, 2001) argues, is primarily a model that explains how morphemes are activated and accessed during speech production. The main argument is centered on the notion that different lemmas underlying different types of morphemes become salient at different levels in speech production. Morphemes no longer fall into two types (content and system morphemes) but four: one type of content morphemes and three types of system morphemes. One type of system morphemes is called 'early system' morphemes. The other type comprises two kinds of late system morphemes ('bridge' and 'outsider system' morphemes).

Early system morphemes are activated at the lemma level. Hence, they are 'conceptually activated'. This feature indicates that "except for the surface-level phonological form, the critical information necessary for the form of the morphemes is available at the lemma level" (Myers-Scotton and Jake, 2000, pp. 1062–1063). Early system morphemes may be in the same lemma as the content morphemes or in a different lemma. They are realized within the maximal projection of the content morphemes that choose them. An example of an early system morpheme is the definite article, which the head indirectly elects to add the notion of definiteness and therefore to complete the semantic/pragmatic features of the speaker's intentions (Myers-Scotton and Jake, 2000, 2001). Late system morphemes, on the other hand, contain grammatical information. They are 'structurally assigned' and their "morphosyntactic information necessary for their projection is not available until directions are sent to the formulator to assemble larger constituents" (Myers-Scotton and Jake, 2000, p. 1063).[2] Examples of late system morphemes are the genitive/possessive *of* in English, the possessive -'*s*, and the third person present tense -*s*.

What is significant in the distinction made between early and late system morphemes is their differential access and election. According to Myers-Scotton and Jake, early system morphemes are activated and accessed along with the content morphemes at the lemma level.

[2] The MLF is a production-based CS model. It claims that the explanation for surface CS patterns lies in language production processes, which are lexically based. The grammatical structures that are elected are contained in lemmas, which are abstract entries in a speaker's mental lexicon (Levelt, 1989). Most researchers in speech production agree on a three-stage model of speech production (e.g. Kempen and Hoenkamp, 1987; Levelt, 1989): (1) Conceptualizer, (2) Formulator, (3) Articulator. The conceptualizer is the stage at which messages are at a representational level and prepositional in nature, with no linguistic formulation. The formulator, on the other hand, is the stage at which non-linguistic messages are transformed into a linguistic form so that they are actually produced at the next stage, i.e. the articulator.

They are realized within the maximal projection of the content morpheme that chooses them. Late system morphemes, on the other hand, are not activated at the lemma level. Their grammatical information is not projected until the formulator gathers all the constituent structures of maximal projections "although their slots are prepared in relevant entries in the mental lexicon" (Myers-Scotton, 1999, p. 21).

The refinement of the classification of system morphemes seems to contradict the fundamental premise that the ML and EL are differentially activated and that the ML plays the more dominant role in structuring the frame of an ML + EL utterance morpho-syntactically. One of the most important implications of the premise is that the ML system morphemes are more salient, and hence more easily accessed and elected, than the EL system morphemes. The question is: How can an 'early' system morpheme from the EL be accessed before a ML system morpheme when the model clearly claims that it is the ML that directs the morpheme order and supplies the (late) system morphemes? If, as Myers-Scotton and Jake (2000, 2001) argue, the EL content morpheme elects an EL early system morpheme at the conceptual level long before any ML system morphemes are accessed, elected and realized at the surface structure, is it not going to constrain the choice/election of the ML system morphemes? In other words, does this not contradict the claim that the ML constrains the role of the EL when supplying not only singly occurring elements but also complex constituents? According to both versions of the MLF model, even EL islands (i.e. constituents that are well formed according to EL requirements) are claimed to be under the ML control because they are part of larger ML + EL constituents. In other words, every EL element has to be checked for 'sufficient' congruence before being inserted. How, then, can EL early system morphemes be elected before the ML system morphemes are accessed? Furthermore, the early system morpheme notion does not seem to help explain instances where the ML and EL system morphemes that co-occur signal different functions (see section 3).

### *2.5. The Abstract Level model*

The Abstract Level model, the second sub-model that Myers-Scotton and Jake add to the MLF model in the refined version, is based on one major premise, which states that all lemmas include three levels of abstract lexical structure. They are:

1. the level of lexical-conceptual structure (activation of language specific semantic/pragmatic features between the conceptualizer and the mental lexicon),
2. the level of predicate-argument structure (mapping thematic structure onto grammatical relations), and
3. the level of morphological realization patterns/ realization of grammatical relations on the surface.

Myers-Scotton (1999, p. 19) argues that "when two languages are in contact, the three levels of lexical structure can be viewed individually and can be accessed independently of each other". The Abstract Level model, Myers-Scotton argues, allows for levels of lexical structure to be split and recombined between the languages involved in the switching. The result is a 'composite' ML. For example, a lexical item may demonstrate predicate-argument structure from one variety but may show morphological realization patterns from the other variety. Viewing the abstract lexical structure in terms of three levels helps explain patterns found in language contact phenomena. Jake and Myers-Scotton (in press) argue that two major factors give rise to a composite ML. The first is that in 'classic' CS, speakers have full access to the grammars of the languages involved: "[i]n some contact situations, speakers produce bilingual CPs but they do not have full access to the single variety that is the desired ML" (Myers-Scotton and Jake, 2000, p. 107). In other words, speakers do not have full access to the grammars of the languages involved in language contact situations. The second factor is operative "when speakers have divided loyalties" and the dominant language for a community is 'ambiguous' (Jake and Myers-Scotton, in press).

I have a problem understanding three vague concepts that Myers-Scotton and Jake put forward as arguments to explain the circumstances under which a composite ML arises. First, what is sufficient or insufficient access to a language? What evidence attests to the fact that, in language contact situations, speakers do not have 'full' access to the ML? What hinders speakers' access to the grammar of a language that they have chosen as an ML for framing their bilingual utterances? To acknowledge that the notion of 'insufficient' access/'insufficient' proficiency is 'imprecise' and that it represents some kind of a 'plague' in research on bilingualism, as Myers-Scotton (1999, p. 14) does, does not offer a plausible answer to the questions. The second problem is the notion of divided loyalties. If there were such a sociolinguistic factor, I would argue that speakers of classic CS have divided loyalties as well. Why should it be more accentuated in language contact situations? In the case of Arabic, at least, it is hard to demonstrate that speakers have more/less loyalty towards any of the two varieties

in the absolute. Different values, e.g. social, pragmatic, economic or symbolic, are ascribed to each variety, depending on who uses them, when and where they are used and to whom they are addressed. Symbolically, both varieties represent shared history, shared culture, shared identity and shared language within each Arab country and across the Arab world. Classical Arabic and Modern Standard Arabic are what make an Arab, and the dialect is what makes a Tunisian, a Moroccan or a Lebanese. The fuṣḥaa is nobody's native language/variety in Arabic-speaking communities but it is every Arab's variety/language. The third problem concerns the idea that the dominant language for a community is unambiguous in classic CS but 'sometimes' ambiguous in language contact situations. One may wonder when and why those situations of ambiguity arise. The fact of the matter remains that such assumptions raise many more issues than they solve. The vagueness and randomness of these three sociolinguistic factors contrast with the linguistic claim that the combination of structure from two varieties in language contact situations is not random (Myers-Scotton and Jake, 2001, p. 97). It seems to me that the calls for research to focus on the intersections between grammatical and sociolinguistic factors (e.g. Treffers-Daller, 1994) to explain switching patterns have yet to be responded to. In fact, serious attention to the interplay between linguistic and extra-linguistic factors in shaping CS patterns would encourage much-needed rethinking of our models in order to conceptualize the dialectical relationship between the two types of factors.

Furthermore, in her new ML Recognition Maxim, Myers-Scotton (1999, p. 23) suggests:

> It is a mistake to identify the ML with the notion of 'language'. Rather, the ML is best characterized as the source of grammatical structure for the bilingual CP. It happens that in 'classic' CS the ML is a single language, but this is not the case in bilingual CPs in other contact phenomena where grammatical structure comes from more than one variety.

Quite clearly, this argument demonstrates that the MLF model, as it was originally articulated and recently fine-tuned, is not as universal as it is claimed to be. To argue that the ML is not identifiable as a language but rather a source of grammatical structure in language contact situation, in order to account for the notion of composite ML, raises two major issues. The first relates to the question of how one might constrain such cases of 'split' or 'composite' ML. The second issue is that it weakens the significant pivotal concept of hierarchies upon which the model is based; I mean the ML/EL hierarchy and the content/system morpheme hierarchy. In a composite ML, it is more than one language that participates in structuring ML + EL constituents. Does this mean that the ML/EL hierarchy is suspended when the composite ML arises? Does it imply that the notion of interaction between the ML/EL and content/system morphemes hierarchy is also suspended when the composite ML arises? Does it mean that the EL is 'in charge' when a composite ML is created? How many 'early' system morphemes can be elected before a complete turnover occurs? I will return to these questions when I re-examine the second set of problematic cases that occur in my data, in section 4.

## 3. Potential problematic cases to the MLF model

As mentioned earlier, the data that I analyzed in a previous study in terms of the MLF model, clearly showed the pre-eminence of Tunisian Arabic (TA; the matrix variety) over fuṣḥaa (the embedded variety) in directing the order of the morphemes and supplying the system morphemes for ML + EL constituents.

In its applicability to Arabic diglossic switching, the model provides a theoretical frame to relate the study of a set of practices (mostly characterized as 'third' language or 'middle' varieties in previous studies on Arabic) to larger issues in the domain of language production. A model like the MLF, although challenging in its application to a diglossic situation, offers a novel perspective on analyzing the nature of switching patterns between two historically related varieties of the same language, while unveiling the special and complex nature of diglossia in Arabic speech communities. Robust as it is, the MLF model has provided a frame to account for a significant amount of data in the corpus I studied, except for two sets of cases that are the focus of the next two sections, section 3 and section 4.

*A note on the data and its presentation*. The data upon which this study is based consist of fourteen hours of tape-recorded public political speeches (seventeen speeches in all, nearly five hundred transcribed pages) that former President of Tunisia, Habib Bourguiba, delivered between the years 1956 and 1968. They are unscripted, and predominantly unprepared. They range over planned, semi-planned and truly unplanned extemporaneous discourses, and over highly formal, semi-formal and informal styles. The data were copied from tapes and records during the summer of 1995, at the Centre de Documentation Nationale in Tunis. The speeches that I collected were not the 'translated' or 'bleached' version of Bourguiba's speeches (into 'good' fuṣḥaa), as published by the Tunisian Ministry of Information.

In each example, the utterance under discussion is presented in the top line as uttered. (I assume that one can speak fuṣḥaa without using case endings, an assumption that some researchers on Arabic might reject.) In the line immediately beneath the linguistic forms, glosses

are provided using the abbreviations listed below.[3] In the third line, forms are marked as belonging to the dialect (TA), fusḥaa (F) or both (TA/F). To assist readers who are unfamiliar with fusḥaa and/or Tunisian Arabic, I also offer the versions of the utterances as they would appear if uttered in the respective varieties.

### *3.1. Illustration of ML + EL constituents*

Of relevance to the discussion of the problematic cases in the present study is the illustration of the pre-eminence of the ML (Tunisian Arabic) over the EL (fusḥaa) in setting the frame for ML + EL constituents. In the corpus, the system morpheme principle is strongly supported for the ML + EL constituents. Except for the two problematic sets, TA provides system morphemes (e.g. determiners, inflections) while fusḥaa provides content morphemes (e.g. verbs, nouns), as illustrated in (1).

Actual utterance

| (1) | il-ʕanaaṣir | illi | yi-nsajm-u |
|---|---|---|---|
| | DEF-parts | which | 3 MAS-match-PL |
| | TA-F | TA | TA-coll F-TA |
| | maʕa | baʕḍ-hum | |
| | with | few-them | |
| | F | TA-TA | |

'The elements that go with each other.'

In (1), note first the dialectal definite article 'il-' in 'il-ʕanaaṣir' and not the fusḥaa 'al-'. Note also how the colloquialized fusḥaa verb 'yi-nsajm-u' is inflected with the third person plural morpheme '-u' in accordance with TA agreement rule although 'ʕanaaṣir' is an inanimate fusḥaa noun. Also of interest in this example are the dialectal 'yi-' prefixed to the verb in 'yi-nsajm-u', rather than the fusḥaa verbal inflection 'ta-', and the morpheme '-hum' 'them' in 'baʕḍ-hum' (not the fusḥaa '-ha'), which refers back to 'ʕanaaṣir' (a fusḥaa inanimate plural noun).

Under the provision of the MLF model, EL system morphemes do occur but only in EL islands (i.e. constituents consist of EL morphemes and are well formed according to the grammatical rules of the EL).

Indeed, in the corpus, fusḥaa verb stems inflected with fusḥaa tense affixes occur only in EL islands, e.g. frozen expressions, and quotations from the Qur'an, as illustrated respectively in examples (2) and (3).

[3] Abbreviations used in glosses:

| | | | |
|---|---|---|---|
| ACC | Accusative | MAS | Masculine |
| DEF | Definite article | NEG | Negative |
| DEM | Demonstrative | PASS | Passive |
| FEM | Feminine | PERF | Perfect aspect |
| FUT | Future tense | PL | Plural |
| IMP | Imperfect | POSS | Possessive |
| SG | Singular | VN | Verbal noun |
| SUB | Subjunctive | | |

Actual utterance

| (2) | haðɛɛya | yu-ðkar | fa |
|---|---|---|---|
| | DEM | 3 SG MAS IMP PASS-mention | so |
| | TA | F | F |
| | yu-ʃkar | | |
| | 3SG MAS IMP PASS-thank | | |
| | F | | |

'This [should] be mentioned and thanked for.'

Of relevance in example (2) is the passive imperfect prefix 'yu-' in 'yu-ðkar' and 'yu-ʃkar', whose dialectal equivalent is 'yit-' as in 'yit-ðkar' and 'yit-ʃkar' in TA.

Actual utterance

| (3) | wa | lan | ta-staṭiiʕ-uu | ʔan |
|---|---|---|---|---|
| | and | NEG FUT | 2PL IMP-be able-SUB | that |
| | F | F | F | F |
| | ta-ʕdil-uu | | | |
| | 2PL IMP-be just-SUB | | | |
| | F | | | |
| | wa | law | ḥarist-um | |
| | and | if | try-2PL IMP | |
| | F | F | F | |

'And you will not be able to treat your wives equally even though you tried your best.'

(Qur'an, Surah IV:129)

Example (3) shows that inflectional tense markers come from fusḥaa. Note the fusḥaa subjunctive '-uu' in both fusḥaa imperfect verbs 'tastaṭiiʕ-uu' and 'taʕdil-uu' because of the presence of the fusḥaa negation future marker 'lan' 'will not' in the first case and 'ʔan' 'that' in the second. Observe also how the fusḥaa verbs are inflected with fusḥaa tense prefix 'ta-' and fusḥaa tense suffix indicating second person imperfect plural morpheme '-tum'.

### *3.2. Co-occurrence of system morphemes from both varieties of Arabic within a single CP*

As noted earlier, in section 2, the ML hypothesis clearly states that the ML sets the morpho-syntactic frame in ML + EL constituents. It also implies that frame building occurs before the insertion of content morphemes. Setting the frame means (i) specifying the morpheme order which, according to the Morpheme Order Principle, should not violate that of the ML in ML + EL constituents, and (ii) supplying the syntactically relevant morphemes in elements containing morphemes from the participating languages (i.e. in ML + EL constituents).

In the early version of the MLF model, the System Morpheme Principle predicts that all syntactically relevant system morphemes come from the ML in ML + EL constituents. The cases that I will discuss next are counter-examples to the System Morpheme Principle because they pattern according to a configuration that it disallows. Let me first offer the examples, then describe

the nature of system morphemes involved and how they are problematic, and, finally, re-visit them in the light of the 4-M model.

The first problematic set occurs four times in the data, only three instances of which are reported here, in (4.1), (5.1) and (6.1).The three examples involve an utterance in which system (grammatical) morphemes come from both varieties, i.e. a single head (a verb stem) has affixes from both the ML and the EL marking different functions (unlike double morphology), a configuration that the System Morpheme Principle, as articulated in the MLF, disallows.

In (4.1) and (5.1), the forms 'ma-ʔa-ðunnu-ʃ' and 'ma-ʔa-ʕtaqid-ʃ' consist of the TA discontinuous negation marker 'ma- . . . -ʃ' and the fuṣḥaa marker of first person singular imperfect tense 'ʔa-' and not the dialectal 'n -' as illustrated in (4.3) and (5.3). Both system morphemes are affixed to the fuṣḥaa verbs 'ðanna' in (4.1) and 'ʔi-ʕtaqada' in (5.1). It is important to note that the contrast 'ʔa-'/'n-' has served to classify the Maghreb (North African) dialect cluster as one group apart from the Machraq (Middle Eastern) dialect cluster. The first person marker 'ʔa-' in both verbs is unambiguously fuṣḥaa.

Actual utterance

(4.1) ma-ʔa-ðunnu-ʃ kɛɛnu
NEG-1SG IMP think-NEG were
TA - F - F - TA TA
'I do not think they were.'

Fuṣḥaa

(4.2) laa ʔa-ðunnu ʔanna-hum kɛɛnu
NEG 1SG IMP-think that-they were

Tunisian Arabic

(4.3) ma-n-ðunnu-ʃ kɛɛnu
NEG-1SG IMP-think were

Actual utterance

(5.1) ma-ʔa-ʕtaqid-ʃ
NEG-1SG IMP believe-NEG
TA - F - F - TA
'I don't believe.'

fuṣḥaa

(5.2) laa ʔa-ʕtaqid ʔanna-hum kɛɛnu
NEG lSG IMP-believe that-they were
'I don't believe that they were.'

Tunisian Arabic

(5.3) ma-n-ðunnu-ʃ kɛɛnu
NEG-lSG IMP think-NEG were
'I don't think they were.'

Also of interest in example (4.1) is the use of the verb 'ðanna' in the constituent 'ma-ʔa-ðunnu-ʃ'. When examining the surface structure of the constituent 'ʔa-ðunnu', it is unambiguously fuṣḥaa because of the presence of the 1st person singular imperfect marker. However, the subcategorization rule in place is clearly dialectal. The fuṣḥaa verb 'ðanna' and its dialectal cognate 'yðun' have different subcategorization rules. The verb 'ðanna' cannot be analyzed as a fuṣḥaa (EL) content morpheme because, according to the Blocking Hypothesis, when EL content morphemes are not congruent with ML morphemes in regard to subcategorization restrictions, they are blocked.[4] The fuṣḥaa verb subcategorizes for a complementizer 'ʔanna' 'that', which calls for a cliticized masculine plural pronoun '-hum', while the dialectal cognate calls for a verbal phrase (with a finite verb), as is illustrated in (4.2) and (5.3), respectively. And, indeed, that does seem to be the case in (4.1). In this constituent, TA supplies the negation marker and it is its subcategorization rule that is realized; fuṣḥaa, on the other hand, supplies the markers of tense, mood, and subject agreement. Both varieties are participating grammatically in framing the constituent, in other words, TA and fuṣḥaa are supplying system morphemes within the same CP.

In their refinement of the notion of congruence, Myers-Scotton and Jake (2000) argue that when congruence is 'seriously' disrupted (because the EL element is not well integrated into the ML frame), two compromise strategies are possible. They are EL islands[5] and 'bare forms'[6]. Neither strategy comes into play here to explain the clash in subcategorization restrictions that occurs in the examples under discussion. The constituent 'ma-ʔa-ðunnu-ʃ' is not an EL island since it contains a system morpheme from TA. Neither is it a bare form because it is inflected with system morphemes from both varieties. Similarly, in (5.1), the verb 'ʔi-ʕtaqada' is unambiguously a fuṣḥaa content morpheme affixed with a fuṣḥaa 1st person singular in the imperfect tense that is inserted in the constituent, but it clearly obeys the dialectal subcategorization rule, just like example (4.1).

Now consider example (6.1). Of interest here is the form 'ma-sa-ta-qif-ʃ' 'will not cease', in which the verb is inflected with system morphemes from TA and fuṣḥaa. Note how the TA discontinuous negation marker 'ma-. . . -ʃ', the fuṣḥaa future marker 'sa-' and the fuṣḥaa imperfect marker 'ta-' are affixed to the verb stem 'waqafa' 'to cease/stop'. The 2nd person imperfect prefix

[4] The Blocking Hypothesis states: "In ML + EL constituents, a blocking filter blocks any EL content morpheme which is not congruent with the ML with respect to three levels of abstraction regarding subcategorization" (Myers-Scotton, 1993, p. 120).

[5] EL islands are constituents that are well-formed according the EL grammar. According to Myers-Scotton and Jake, they are one of the two compromise strategies that come into play when congruence between the ML frame and EL element (to be inserted) is disrupted.

[6] Bare forms are the second compromise strategy. They are "lexemes that occur in a mixed constituent frame prepared by the ML, but the EL form is missing some or all of the required ML system morphemes" (Myers-Scotton and Jake, 2001, p. 106).

'ta-' in 'ta-qif' is unambiguously fuṣḥaa and not TA because it carries a short vowel in accordance with fuṣḥaa, not a long vowel, as it does in TA, as illustrated in (6.2) and (6.3). It is important to note that the verb 'waqafa' is the same in both varieties, except for the syllabic structure. It is 'waqafa' in fuṣḥaa and 'wquf' in TA, with the consonant cluster, a hallmark of Maghrebi dialects.

Actual utterance

(6.1) ma- sa-ta-qif-ʃ
NEG- FUT-3SG stop IMP-NEG
TA F-F-F/TA-TA
'It will not cease [to be].'

Fuṣḥaa

(6.2) lan ta-qif-a
FUT NEG 3FEM SG-cease-SUB

Tunisian Arabic

(6.3) muuʃ bɛɛʃ tɛɛ-qif
NEG FUT 3 FEM SG-cease

In (6.3), note that, in the future tense, TA does not use the discontinuous negation marker but a single particle 'muuʃ' in front of the future marker 'bɛɛʃ'. It should be pointed out that, in addition to the future bound morpheme 'sa-' (used in the actual utterance in the data, see (6.1), fuṣḥaa uses a particle 'sawfa' to also indicate future. Had the speaker used it in combination with the TA discontinuous negation marker 'ma-. . .-ʃ', it would have inevitably resulted in unacceptable utterances, as illustrated in (6.4).

(6.4) *ma- sawfa ta-qif-ʃ
NEG FUT 3 FEM SG-cease
TA F F
'It will not cease [to be].'

The question is: how to account for the co-occurrence of system morphemes from both varieties in examples (4.1), (5.1) and (6.1)? The co-occurrence of system morphemes from TA and fuṣḥaa within the same CP is a counter-example to the System Morpheme Principle as originally articulated in the MLF model (Myers-Scotton, 1993), which predicts that all syntactically active system morphemes must come from the ML.

In refining the model, Myers-Scotton re-adjusts her claim by arguing that the principle "only predicts that only some, not all, system morphemes must come from the ML" (Myers-Scotton, 2001, p. 94). Her re-articulation of the MLF System Morpheme Principle was a way to fine-tune her notion of how system morphemes are accessed, and how one type of system morphemes is indeed accessed and elected along with the content morphemes that choose them before the ML system morphemes are put in place. The 4-M model predicts and allows early system morphemes to come from EL, as I mentioned earlier.

### *3.3. Revisiting the first problematic set in light of the 4-M model*

As noted earlier, the examples under analysis are different from double morphology (e.g. infinitives, plurals and pronouns doubly-marked). Unlike double morphology, my problematic cases display the co-occurrence of system morphemes from both varieties but signaling different functions (e.g. TA negation marker and fuṣḥaa future marker affixed to a verb). It is some kind of what I call 'dual morphology'.

Myers-Scotton's (1993, p. 135) original discussion of double morphology suggests that 'mistiming' or 'misfiring' provides some kind of answer to double morphology cases. Myers-Scotton and Jake (2001, p. 94) re-articulate the account of double morphology by suggesting that in ML + EL constituents the EL system morphemes are "accessed along with the content morphemes before the ML fills in the frame with its version of [the doubled system morpheme]". These system morphemes are elected at the conceptual level and hence they are called 'early' system morphemes. Myers-Scotton (1999, p. 22) explains that structurally-assigned system morphemes (i.e. 'late' system morphemes) "only arise at a later point in the production process (although their slots are prepared in relevant entries in the mental lexicon) to set the frame".

Let us re-examine example (6.1) in terms of the 4-M model. In this instance, are we to consider the fuṣḥaa future and tense/person morphemes as early system morphemes indirectly elected by the verb 'waqafa' and the TA negation marker as a late system morpheme structurally-assigned in the production process at the functional level? If we examine the constituent 'ma-sa-ta-qif-ʃ' in the discourse in which it occurs, it is clear that TA is the matrix, in the sense that it supplies all the system morphemes and directs the word order. What evidence indicates that the EL system morphemes, i.e. the future marker 'sa-' and the tense/aspect marker 'ta-', are accessed early?

In their effort to explain why the Arabic frame allows very few English verbs with Arabic inflections, Jake and Myers-Scotton (1997) and Myers-Scotton and Jake (2001) argue that verbs in Arabic are not 'stored' as verb roots without the tense/aspect specification, at the conceptual level. In other words, lemmas underlying verbs in Arabic do not contain just information on the verb stem like other languages, e.g. English. Tense/aspect morphemes in Arabic are not in separate lemma entries, and hence they are elected with the verbs. Then the question is: why are fuṣḥaa tense/aspect markers not consistently and systematically activated along with fuṣḥaa verbs? Except

for the four problematic cases under discussion in this paper, verbs (ML or EL verbs) carry TA tense markers in ML + EL constituents in the corpus.

According to the MLF model, the ML is the dominant language. In other words, its system morphemes are more salient, more easily accessible, first in line to be elected to participate in building the frame. If 'some' EL system morphemes (called 'early system' in the 4-M model morphemes) are accessed earlier, will they not constrain and control the nature of ML system morphemes that are structurally assigned at a later stage? It is not possible to determine the ML in a CP if both languages supply the system morphemes. What is, then, the role of the ML as the language/variety in charge? Is it still the language that directs the word order and supplies some of the system morphemes? Or is it the language that supplies more system morphemes? It is hardly possible to determine the ML in the constituent 'ma-sa-ta-qif-ʃ' (or in 'ma-ʔa-ðunnu-ʃ', and 'ma- ʔa-ʕtaqid-ʃ', examples (4.1) and (5.1), for that matter) with system morphemes from fuṣḥaa and TA.

If the fuṣḥaa (EL) future marker, as in (6.1), is conceptually accessed and directly elected by the verb, it is going to constrain the choice of the late system morpheme that is necessary to structure this constituent (i.e. TA negation markers). Had the speaker accessed the fuṣḥaa future particle 'sawfa' instead of the bound morpheme 'sa-' (both of which express future), it would have been impossible for him to avoid an ungrammatical utterance because he would have had only two options to express the negation: either as a TA 'ma- . . . -ʃ' or a fuṣḥaa 'lan', each of which would have resulted in an unacceptable utterance; see examples (6.4) (repeated here for convenience) and (6.5).

(6.4) *ma- sawfa ta-qif-ʃ
NEG FUT 3 FEM SG-cease
TA F F
'It will not cease [to be].'

(6.5) *lan- sawfa ta-qif-(a)
NEG FUT 3 FEM SG-cease
TA F F
'It will not cease [to be].'

It has been argued in the CS literature that in some languages stems and affixes are inseparable at the production level. In other words, they are selected and retrieved as wholes (Butterworth, 1983, 1989). This argument has been put forward especially for analytic languages, like English. Levelt (1989), for instance, suggests that different types of languages have different types of lexical entry in a mental lexicon. For some languages (e.g. agglutinative languages, Turkish, for instance) stems and affixes are stored separately whereas in other languages they are stored, and hence retrieved, as full words. This argument may suggest a first possible interpretation of example (6.1). That is, the lemma for verb stem provides to the formulator a lexical item with a future marker attached to the verb as a single constituent. The constituent is provided to fill the slot prepared by the TA matrix. Tense and person affixes (e.g. the fuṣḥaa future marker and 2nd person prefix) demonstrate a strong attachment to their heads (i.e. the fuṣḥaa verb stem). In other words, the future and 2nd person markers may be situated at the same lemma location as the verb stem and, therefore, they are accessed and retrieved simultaneously, though after the ML has set the frame. Hence, they are elected under the control of the ML. What evidence attests to this hypothesis in the literature on language contact situations? Undoubtedly, further studies on how abstract entries of a root-pattern language are represented in the mental lexicon are needed.

While keeping in mind the notion of differential saliency and activation of system morphemes, I suggest a second, more plausible, explanation. Because of the nature of the frame, which requires that system morphemes must come from the matrix variety (here, TA), a slot for the TA negation marker is set (for it to be accessed and put in place later). The abstract entry of the discontinuous negation marker 'ma - . . . -ʃ' clashes with the abstract entry of the TA future marker 'bɛɛʃ' during the online evaluation of the utterance. Their co-occurrence, when realized at a later stage, would inevitably have resulted in a grammatically unacceptable constituent, i.e. *ma-bɛɛʃ + verb -ʃ; see example (6.6) below. It is important to recall that the TA discontinuous negation marker is used only with the perfect and imperfect tenses, but not with the future, as noted earlier.

(6.6) *ma- bɛɛʃ tɛqif-ʃ
NEG- FUT 3SG FEM IMP-cease-NEG
'It did not cease [to be].'

Because the TA discontinuous negation marker was elected, the speaker is, in a way, constrained in his selection of the next system morpheme (i.e. the future marker). He has three possible options: a fuṣḥaa bound morpheme 'sa-', a fuṣḥaa particle 'sawfa' and a TA particle 'bɛɛʃ'. But only one, i.e. the fuṣḥaa bound morpheme marker, is permissible for repairing such a wrong selection, and save the utterance. The negation marker is a 'first-in-line' system morpheme since TA is the ML, and therefore it is part and parcel of the frame-building stage. Tense markers are 'second-in-line'. But because the EL is also activated, although at a lesser rate, its system morphemes are likely to be accessed and come to the rescue to repair some erroneous system morpheme elections. The ML (i.e. TA) controls the grammatical frame and, consequently, it also controls the nature of

the EL system morpheme elected. Recall that it is 'sa-', the bound morpheme, and not the particle 'sawfa', that is accessed to rescue the speaker, repair the erroneous election and save the utterance. It is the morpheme that is more closely bound with the verb that comes to the rescue during online production.

In his explanation of the double morphology phenomenon, Azuma (1991) interprets the doubling phenomenon as evidence supporting the resistance of the frame to change once it is set. I concur with Azuma on the notions of frame resistance and repair. I, too, think that, once the frame is planned, it is not possible to change it. It is evaluated during online production and repaired via the EL when the need arises. This explanation is in unison with the two key hierarchies that sustain the MLF model. It is the ML that sets the morpho-syntactic frame in ML + EL constituents. The EL is activated to a lesser degree and hence supplies content morphemes, but also system morphemes to come to the rescue when the need arises.

### *3.4. Problems in word order and subcategorization rules*

In this section, I examine the second set of problematic cases. In my analysis, I discuss two types: word order and clashes in subcategorization restrictions between the two varieties participating in diglossic switching. In each set, however, the utterance is composed of grammatical projections from both varieties, fuṣḥaa and TA.

#### *TA word order but fuṣḥaa system morphemes*

This problematic set involves cases where all the surface forms (system and content morphemes) come from one variety, but the word order is from the other variety.

According to the System Morpheme Principle and the Morpheme Order Principle, one would expect the system morphemes and the word order to be those of the ML. Consider examples (7.1) and (8.1). In these examples, the dialect (TA) directs the word order, although all the surface forms come from fuṣḥaa.

Actual utterance

(7.1) liʔanna laa ya-nhaḍ
because NEG 3MAS SG IMP-advance
F F F
al-ʔadab
DEF-literature
F

illa bi-t-talaaquḥ
only with-DEF-exchange
'Because literary production does not progress unless [there is] exchange.'

(8.1) liʔanna taʕayyar sullam
because 3FEM SG PERF-change system
F F F
al-qiyam
DEF-values
F
'Because the system of values has changed.'

In (7.1), note the fuṣḥaa negation marker 'laa' (not the TA 'ma-'), the fuṣḥaa verbal inflection 'ya-' in 'ya-nhaḍ' (not the TA verbal inflection 'yi-'), and the fuṣḥaa definite article 'al-' in the NP 'al-ʔadab' (not the TA 'il-'). In (8.1), note the fuṣḥaa definite article 'al-' in 'al-qiyam', and the fuṣḥaa verb 'taʕayyar' inflected with the third person singular imperfect prefix 'ta-'. A colloquialized form of 'taʕayyar' in a ML + EL in which TA is the matrix variety, as I have demonstrated in Boussofara-Omar (1999), would have resulted in an initial consonant cluster 'tʕayyar' through the deletion of the vowel in the first syllable, according to TA syllabic structure.

Let me take (7.1) to discuss the problem. When examined closely and discussed in terms of word order, example (7.1) clearly obeys the TA word order, as illustrated in (7.2), and not fuṣḥaa syntactic specifications, as illustrated in (7.3). Examples (7.1) and (8.1) involve the pattern [because + VP + NP], i.e. the dialectal pattern as illustrated in (7.2),whereas the fuṣḥaa order is [because + NP + VP], as illustrated in (7.3).

TA word order: TA subordinator + VP + NP

(7.2) ʕla xaaṭar ma yi-tqaddam
because NEG 3MAS SG IMP-advance

il-ʔadab
DEF-literature
'Because literary production does not progress . . . '

Fuṣḥaa word order: Fuṣḥaa subordinator + NP + VP

(7.3) liʔanna al-ʔadab- (a) laa
because DEF-literature- (ACC) NEG
F F F

ya-nhaḍ
3MASC SG IMP-advance
F
'Because literary production does not progress . . . '

The MLF neither predicts nor allows the combination of structures from two varieties. This case is a clear counter-example to the Morpheme Order Principle since the word order is not that of the variety that supplies the system morphemes. This problematic set may also be

discussed in terms of clashes in subcategorization rules between fuṣḥaa and TA.

***Clashes in subcategorization restrictions***

Another set of problematic cases involves clashes between the subcategorization restrictions of the two varieties. What is problematic in this set is that the subcategorization rules in place are not those of the variety that is realized at the level of the surface structure.

Bentahila and Davies (1983) were among the first researchers who raised the issue of subcategorization when they questioned Poplack's (1980) Equivalence Constraint and its validity as far as their Moroccan Arabic/French data were concerned. Their 'Subcategorization Constraint' postulates that "in structures exhibiting switching . . . all items must be used in accordance with their own language-particular subcategorization restrictions" (Bentahila and Davies, 1983, p. 301). Similarly, from a lexically-based perspective, Azuma (1991, p. 7), for example, suggests that the main verb provides a "planning frame" and that the subcategorization of the main verb is always maintained. Muysken (1991), on the other hand, speaks in terms of 'categorical equivalence'. The insertion of content words (content morphemes in Myers-Scotton's terms) is done according to the requirements of the frame planned by the matrix language. In Myers-Scotton's view, early proposals on subcategorization (e.g. Bentahila and Davies, 1983; Azuma, 1991; Muysken, 1991) do not seem to be specific enough. She argues, for example, that they fall short of providing the motivation for the subcategorization restrictions that, in her own model are expressed specifically in terms of the notion of congruence couched in the Blocking Hypothesis (Myers-Scotton, 1993, p. 121; cf. fn. 4 above).

In (9.1), we have a case in which a verb from fuṣḥaa and its cognate from the dialect (TA) subcategorize for different categories of complements. As in the earlier examples, I provide the constituent as it was actually uttered as well its equivalent utterance in TA, (9.2), for comparison.

Actual utterance

| (9.1) | ya-ziid | ya-ftaḥ | ʔamaama |
|---|---|---|---|
| | 3MAS IMP-increase | 3MAS-open | before |
| | F | F | F |
| | n-naaʃiʔa | ʔabwaab | kabiira |
| | DEF-youth | doors | big |
| | F | F | F |

'It increasingly opens up big doors [opportunities] for the young.'

TA utterance

| (9.2) | y-ziid | y-ḥill | quddaam |
|---|---|---|---|
| | 3MAS IMP-increase | 3MAS IMP-open | before |
| | TA | TA | TA |
| | ʃ-ʃubbɛɛn | bibɛɛn | kbaar |
| | DEF-youth | doors | big |
| | TA | TA | TA |

The problem that example (9.1) poses is that the morphemes (system and content morphemes) at the surface structure come from fuṣḥaa but the subcategorization rule and word order in place are those from TA. The verbs 'zaada;' 'increase' and 'fataḥa' 'open' are inflected with the fuṣḥaa prefix 'ya-', indicating 2nd person in the imperfect; the fuṣḥaa plural in 'ʔabwaab' ('bibɛɛn' in TA) is in accordance with the fuṣḥaa pattern 'ʔafʕaal'. It is also interesting to note the feminine singular marker '-a' in the adjective 'kabiira', clearly indicating that the fuṣḥaa agreement rule of inanimate plurals is in place. The TA plural form of 'kabiira' is 'kbaar', irrespective of the animacy of the head noun, as illustrated in (9.2) and (9.4):

(9.3) TA masculine animate plural agreement

| | | |
|---|---|---|
| | rjɛɛl | kbaar |
| | men | old |

'old men'

(9.4) TA masculine non-animate plural agreement

| | | |
|---|---|---|
| | bibɛɛn | kbaar |
| | doors | big |

'big doors'

In (9.1), the utterance is fuṣḥaa in its surface realization but dialectal in its subcategorization rules and word order. In the example, the fuṣḥaa verb 'yaziid' is followed by a finite verb 'yaftah' marked for tense and person, while it normally subcategorizes for either a prepositional phrase or a noun phrase (with the noun being a 'maṣdar' verbal noun, VN, in this instance) as illustrated in (9.5) and (9.6) respectively.

Fuṣḥaa: Option 1

| (9.5) | ya-ziid(u) | fi | fatḥ(i) |
|---|---|---|---|
| | 3MAS SG-increase | in | opening/VN |
| | ʔabwaab | kabiira | |
| | doors | big | |

'It increasingly opens up big doors [opportunities].'

Fuṣḥaa: Option 2

| (9.6) | ya-ziid(u) | fatḥ(a) |
|---|---|---|
| | 3MAS SG-increase | opening/VN(ACC) |
| | ʔabwaab | kabiira |
| | doors | big |

Unlike the fuṣḥaa verb 'zaada', the TA 'zɛɛd' subcategorizes for a finite verb that is marked for tense and

person, as illustrated in (9.2). If we assume that in (9.1) fuṣḥaa is the ML and that the speaker calls for a TA lexeme (the verb 'zɛɛd') which is elevated, or fuṣḥaized, through the insertion of the prefix 'ya-' for third person singular imperfect, the major issue of subcategorization clash between the two varieties remains unresolved. According to the Blocking Hypothesis, an EL lexeme is accessed only if it is congruent with its ML counterpart, given the fact that the slot to be filled by the EL must satisfy the requirements of the ML, in this case, fuṣḥaa.

In addition to the clash involving the verb 'zaada', example (9.1) contains a second clash in subcategorization rules between fuṣḥaa and TA. The verb 'fataḥa', unlike its dialectal cognate 'ḥall' does not call for a prepositional phrase as its object. It subcategorizes for a NP. This structure clearly conforms to TA subcategorization restrictions.

Similarly, examples (7.1) and (8.1) above are instances of clashes in subcategorization rules between TA and fuṣḥaa. Unlike TA, the fuṣḥaa subordinator 'liʔanna' is subcategorized as requiring an NP (not a VP as it occurs in the actual utterance) or a cliticized pronoun attached to the conjunction, followed by a VP.

I do not think that it is possible to even try to argue that in (7.1), (8.1) or (9.1) fuṣḥaa is the ML since it is not framing the sentence (word order and subcategorization rules are dialectal). Neither is it possible to argue that TA is the ML since fuṣḥaa supplies the system morphemes. To simply say that it is a composite ML is to stretch the MLF model to account for instances that it has neither predicted nor allowed to occur. How can these instances be predicted and explained in a principled manner? The composite ML construct, as it was articulated, cannot be constrained. Are we to call any utterance in which both languages/varieties participate in either the morphology or the syntax of an utterance a composite ML? As mentioned in section 2.2, I have problems understanding the sociolinguistic factors that Myers-Scotton (1999) invokes to account for the circumstances under which the composite ML arises.

In their explanation of what they call composite ML structures that occurred in their data, Jake and Myers-Scotton (in press) provide the following example (their (6b.20)), among others:

(10.1) **He sends to my professor telling him**
he sends to my professor telling him
ʔiða ʔaana I deserve an A.
if I I deserve an A
'He sends [it] to my professor telling him if I deserve an A.'

In their analysis of this utterance, they suggest that it has a CP that would otherwise be well-formed in English, except that the verb *send* lacks an object (*he sends to my professor*). Further, they state: "[i]t is clear from context that a document is sent" (Jake and Myers-Scotton, in press, ms. p. 13). It is clear to me that this utterance is structured according to Arabic grammar where the object does not necessarily have to be realized at the surface structure. It is implied and understood. In Arabic, the utterance 'He sends to my professor telling him' reads in fuṣḥaa, Palestinian Arabic and Tunisian Arabic as illustrated in (10.2) (10.3) and (10.4), respectively.

Fuṣḥaa

| | | | |
|---|---|---|---|
| (10.2) | yabʕaθ | li-ʔustað-ii | ya-quul la-hu . . . |
| | send 3 IMP | to-professor-1SG POSS | 3MASIMP-say to-him |

'He sent to my professor saying to him . . .

Palestinian Arabic

| | | | | |
|---|---|---|---|---|
| (10.3) | huwwi | y-waddi | la-ʔustaðii | y-qul-l-u . . . |
| | he | send 3SG PERF | to-professor-my | 3SG IMP-say-to him |

'He sent a letter to my professor, he says in it . . . '

Tunisian Arabic

| | | | |
|---|---|---|---|
| (10.4) | yibʕaθ | l-ʕustað-ii | y-qul-l-u . . . |
| | send 3SG PERF | to-professor-1SG POSS | 3MAS IMP-say-to-him |

'He sent to my professor saying [he says] to him . . . '

Their example is similar to (9.1) in my data. Both English and Palestinian Arabic participate in structuring the utterance. English provides system and content morphemes but the subcategorization rules are in accordance with Palestinian Arabic.

It is important to note that, with respect to examples (7.1), (8.1) and (9.1), eleven Egyptian, Lebanese, Syrian and Tunisian university-educated native speakers of Arabic judged these types of utterances acceptable when read to them even though they violate the canons of prescriptive fuṣḥaa. One cannot argue they are just performance errors. Nor can they be seen as the speaker's 'failure' to speak fuṣḥaa proficiently. Because all the morphemes (system and content) are from fuṣḥaa, the utterances may easily be perceived as fuṣḥaa when used in spoken discourse. One may wonder whether "we are [not] positively forced to ask ourselves not just what MSA [fuṣḥaa] is, but what it is for these people", to borrow Parkinson's (1993, p. 71) words. The lexical (i.e. content morphemes) and morphological (i.e. system morphemes) 'flags' seem to supersede the syntactic constraints in speakers' judgment of what constitutes spoken fuṣḥaa and the dialect.

## 4. Conclusion

In this paper, I have sought to revisit Arabic diglossic switching in terms of the MLF model, as it was originally articulated (Myers-Scotton, 1993) and later

refined (Myers-Scotton and Jake, 2000, 2001) with the addition of the 4-M model and the Abstract Level model. My purpose has been to lay out a framework for thinking about language change and variation in Arabic.

Although there has been ample discussion of the varieties of Arabic in various Arab countries, there has been little focus on how these varieties might be influencing one another. The general tendency has been to perceive the influence as unidirectional, i.e. from fuṣḥaa to the dialect. I believe that the constant leakage and the great overlap between the two varieties spurred by the dramatic social changes (i.e. more widespread literacy and hence higher access to fuṣḥaa) that took place in the Arab world seem to facilitate the flow both ways. Moreover, the growing practice of switching between the two varieties of Arabic may eventually give rise to a national spoken standard that is dialectal in its underlying structure but fuṣḥaa in its surface realization, in each Arab country.

The cases discussed in this paper, although problematic to the MLF and its sub-models, suggest that the 'mix' of fuṣḥaa and the dialect involves more than what has been termed a 'middle' language or 'middle' varieties, whose contours are 'fuzzy' and whose structure is fluid. The complex interactions of fuṣḥaa and the dialect cannot be merely framed within either the Fergusonian paradigm[7] or the continuum notion. The 'mix' of fuṣḥaa and the dialect in the oral speech of educated Arabs is as structurally constrained, and socially motivated (study in progress), as switching between any two pairs of languages. In other words, combining structures from the two varieties of Arabic is not random. The combinations are predictable.

My efforts to consider the nature of diglossic switching in the light of Myers-Scotton's MLF model and its sub-models represent an attempt to grapple with the sterile debate on whether there is only one 'middle' language or numerous 'middle' varieties. Although this study shows that the MLF model does not seem to handle all the data of the corpus, it provides a fertile ground for, first, rethinking our understanding of Educated Arab speakers' alternate use of both varieties of Arabic; second, testing the model in a language contact situation; and, finally, suggesting some refinements that need to be brought to the MLF model and its sub-models in order to offer explanations for cases that they neither predict nor explain.

[7] What I am referring to is Ferguson's (1959) early idealized characterization of diglossia when suggesting that the two varieties of Arabic are in complementary distribution functionally. It is idealized in the sense that it does not reflect the constant flux and ever-increasing leakage between the two varieties. For further references, see Mahmoud (1986).

## References


Azuma, J. (1991). Processing and intra-sentential codeswitching. PhD dissertation. University of Texas, Austin.

Badawi, S. (1973). *Mustawayat al-lugha al-'arabiyya al-mu'aṣra fi miṣr*. Cairo: Dar al-Ma'arif.

Bentahila, A. & Davies, E. E. (1983). The syntax of Arabic-French codeswitching. *Lingua*, 59, 301–330.

Blanc, H. (1960). Style variation in spoken Arabic: A sample of inter-dialectal educated conversation. *Harvard Middle Eastern Monographs*, 3, 79–161. Cambridge, MA: Harvard University, Center for Middle Eastern Studies.

Boussofara-Omar, N. (1999). Arabic diglossic switching: An application of Myers- Scotton's MLF model. PhD dissertation, University of Texas, Austin.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (ed.), *Language production, development, writing and other language process* (vol. II), pp. 257–295. London: Academic Press.

Butterworth, B. (1989). Lexical access in speech production. In D. Marslen-Wilson (ed.), *Lexical representation and process*, pp. 108–135. Cambridge: Cambridge University Press.

El-Hassan, S. A. (1977). Educated Spoken Arabic in Egypt and the Levant: A critical review of diglossia and related concepts. *Archivum Linguisticum*, 8, 112–132.

Ferguson, C. A. (1959). Diglossia. *Word*, 15, 325–340.

Jake, J. L. & Myers-Scotton, C. (1997). Codeswitching and compromise strategies: Implications for lexical structure. *International Journal of Bilingualism* 1, 29–39.

Jake, J. L. & Myers-Scotton, C. (in press). Second generation shifts in socio-pragmatic orientation and codeswitching patterns. In A. Rouchdy (ed.), *Language contact and language conflict in Arabic*. Richmond, Surrey: Curzon Press.

Kempen, G. & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: Cambridge University Press.

Mahmoud, Y. (1986). Arabic after diglossia. In J. Fishman (ed.), *The Fergusonian Impact* (vol. I), pp. 239–251. Berlin: Mouton de Gruyter.

Meiseles, G. (1980). Educated Spoken Arabic and the Arabic language continuum. *Archivum Linguisticum*, 11(2), 118–148.

Mitchell, T. F. (1980). Dimensions of style in a grammar of Educated Spoken Arabic. *Archivum Linguisticum*, 11(2), 89–106.

Muysken, P. (1991). Needed: A comparative approach. In *Papers for the Symposium on codeswitching in Bilingual Studies: Theory, Significance and Perspectives* (Barcelona, 21–23 March 1991), pp. 253–272. Strasbourg: European Science Foundation.

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon. [2nd ed., with a new Afterword, 1997, Oxford: Oxford University Press.]

Myers-Scotton, C. (1995). A lexically based model of codeswitching. In L. Milroy & P. Muysken (eds.), *One*

*speaker, two languages: Cross-disciplinary perspectives on codeswitching*, pp. 233–256. Cambridge: Cambridge University Press.

Myers-Scotton, C. (1999). Putting it all together: The matrix language and more. In B. Brendemoen, E. Lanza, & E. Ryen (eds.), *Language encounters across time and space*, pp. 13–28. Oslo: Novus Press.

Myers-Scotton, C. & Jake, J. L. (2000). Four types of morphemes: Evidence from aphasia, codeswitching and second language acquisition. *Linguistics*, 38, 1053–1100.

Myers-Scotton, C. & Jake, J. L. (2001). Explaining aspects of codeswitching and their implications. In. J. Nichol (ed.), *One mind, two languages: Bilingual language processing*, pp. 91–125. Oxford: Blackwell.

Parkinson, D. (1993). Knowing Standard Arabic: Testing Egyptians' MSA abilities. In M. Eid & C. Holes (eds.), *Perspectives on Arabic Linguistics V*, pp. 47–74. Amsterdam: Benjamins.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino español: Toward a typology of codeswitching. *Linguistics*, 18, 581–618.

Treffers-Daller, J. (1994). *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin: Mouton de Gruyter.

Walters, K. (1996a). Intrasentential codeswitching in diglossic settings and its implications for linguistic variation and language change. In J. Arnold, R. Blake, B. Davidson, S. Schwenter & J. Solomon (eds.), *New Ways of Analyzing Variation 23 Proceedings*, pp. 401–416. Stanford, CA: Center for the Study of Language and Information.

Walters, K. (1996b). Diglossia, linguistic variation, and language change. In M. Eid (ed.), *Perspectives on Arabic Linguistics VIII*, pp. 157–197. Philadelphia & Amsterdam: John Benjamins.

**Received May 27, 2002 Revision accepted November 15, 2002**