Constructing a reference speakers of standard language in diglossia

variation in the use of case and mood inflection in Formal Spoken Arabic

Andreas Hallberg

In the diglossic situation of Arabic there are no native speakers of the standard language. In formal situations where people are expected to use the standard variety, there is thus no reference speaker to style shift towards. Drawing on the concept of outgroup referee design (Bell 1984), it is argued that speakers construct a reference speaker to function as target of style shifts for formal situations. The use of case and mood inflection (CMI), a salient Standard Arabic feature in 17 televised news-interviews (38,000 words) is analyzed, revealing similarities in the reference speaker is constructed by the interviewees. The data indicates that the constructed reference speaker, the target of formal style shift, use CMI in words with enclitic pronouns and where it is orthographically represented, and not in words with definite article. These patterns effectively put a ceiling level for CMI style shifts far below what is prescribed by overt linguistic norms.

In English, as well as in many other languages, the standard variety is associated with the socioeconomically privileged class from which it derives prestige and association with upward social mobility. When paying more attention to their speech and shift to a more formal style, speakers use a higher proportion of standard features, reflecting their social stratification (Labov 2006:237ff). In the diglossic situation of Arabic there are no native speakers of the standard variety. This has two important implication for our theoretical understanding of formal speech in. First, SA is not associated with any socioeconomic class. It therefore lacks the prestige to make it the target of socially motivated linguistic adaptation and change (Ibrahim 1986; Owens 2001; Al-Wer 2013). Sociolinguistic change in Arabic speaking communities generally entails adoption of features of the local or regional dominant group, independently of how these features relate to SA. Several instances have been document of how variants that correspond to SA, but that are stigmatized by local norms, are abandoned for the sake of a locally prestigious, but non-standard variant. For example, (Holes 1987) showed how the underprivileged Baharna ethnic group in Bahrain tend to switch from [y] in the traditional dialect to [ʤ] of the politically dominant Arab group, despite this resulting in speech more divergent from SA. (See Al-Wer 2013 for further discussion.) On the other hand, in formal, public situations, speakers are expected to use SA, and react to this expectation by introducing SA features in their speech. Linguistic variation in Arabic can thus be described as occurring in two largely independent dimensions: one is relating to prestige in the traditional sociolinguistic sense of social stratification of linguistic features, and the other relating to standardness and formality (Owens 2001:452). This latter dimension is often referred to as the diglossic continuum (Hary 1996; Meiseles 1980).

The second implication of the lack of native speakers of SA is that for speakers acting in formal situations, there is no real-world reference speaker to function as a model or target for style shifts. In previous research on formal styles of spoken Arabic, the reference speaker has ha implicitly been characterized by prescriptively correct, traditional Standard or Classical Arabic, so that speakers are assumed to operate on a span of variation ranging from completely non-standard Arabic to completely prescriptively correct Standard Arabic, as the language is described in traditional grammars (Badawı̄ 1973; Hary 1996; see Boussafara-Omar 2006 for an overview).

There are reasons to question this characterization of the upper limit of style shift in Arabic. First, if prescriptively correct SA is the target form, we expect some speakers to reach this target. This is however virtually unheard of in extemporaneous speech. (See Parkinson (1994a), further discussed below, for an example.) Second, sociolinguistic theory suggests that people shift to converge with other speakers, whether individual or groups , not with the abstract rules presented to them in grammar instruction (Bell 1984; Labov 1972). The notion that sparks of Arabic style shift towards prescriptively correct speech is in this article tested on data on case and mood inflection (CMI) in formal speech. Drawing on Bells’s model of intra-speaker linguistic variation as audience design (1984; 2002), style shifts is assumed to be, at its core, convergence with a reference speakers. The characteristic of the reference speaker thus determines the direction of the style shift, that is, what linguistic features are adopted, complete convergence with the reference speaker represents its outermost limit.

CMI is a salient SA feature and a marker of linguistic correctness (Bohas et al. 1990:50; Haeri 2003:40; Uhlmann 2012:104) and is therefore likely to play an important role in the construction of the reference speaker. Traditional grammars and widespread educational practices prescribe complete use of CMI on all nouns and adjectives and on verbs in the imperfect stem. There are, however, indications that the constructed reference speaker does not employ CMI consistently. In formal spoken situations, where speakers are expected to use SA, is used only sporadically, even by speakers whose speech is predominantly SA in terms of lexicon, syntax, and phonology. Parkinson (1994a:207–8) describes this use of case and mood inflectional forms as speakers “sprinkling them through their text at a certain rate to give the flavor of fuṣḥā [SA] without making it so difficult on themselves they would not be able to speak at all.” He further suggests that even though speakers use CMI at different rates, they do so under

a similar set of constraints, patterning their sprinkling in a particular way, with some forms (defined both phonologically, morphologically and syntactically) being more likely to receive vowels [CMI] than others.

This article tests this hypothesis, with the assumption that any such constraints reflect the constructed reference speakers, or put differently, reflect the target form of style shifts towards SA. In so doing, this paper aims to answer two basic questions:

How much does the use of CMI vary between speakers?

Is the use of CMI constrained, probabilistically or absolutely, by morphological or syntactic factors?

The rest of this article is organized as follows. First, the variable under investigation, CMI, is discussed with regards to its role in Formal Spoken Arabic, and previous research on this variable is reviewed. Thereafter the method is presented, including the corpus collection, the coding scheme. Methodological problems concerning the overlapping of standard and non-standard forms and lexicalized forms of CMI are also discussed. Data on the use of CMI is then presented and analyzed according to variation between speakers and correlations with morphosyntactic parameters and orthographic representation. The results from this analysis are then discussed in terms of relation to the constructed SA referee. The main findings and conclusions are summarized in the final section.

Diglossia and outgroup referee design

A useful framework for understanding formal style shift in diglossia is the audience design model of stylistic variation, and more specifically the concept of outgroup referee design (Bell 1984; Bell 2002). In this model, stylistic shifts are viewed as always being related to an audience. In the typical case, the audience is the addressee, and the speaker style shifts to converge with them. Speakers may also shift in a way that deviates from the addressee, towards a third party, or “referee”. This referee group may be one with which the speaker identifies, ingroup referee design, which is generally done to signal distance and exclusion toward the addressee. The third-party referee group may also be one with which the speaker does not identify, outgroup referee design. In this case, the speaker “lay claim to a speech and identity which are not their own but which hold prestige” (Bell 1984:188). Such referee design is typically short-lived, as they derive their effect from a contrast with speech behavior. Style shift towards to an outgroup reference group is associated with certain difficulties, since speakers “have to overcome ignorance both of a target speech community to which they may have no access, and of a target variety which they may never have heard spoken natively” (Bell 1984:190). The larger the distance to the reference group, the more limited and vague the knowledge of their speech, making shifts toward them “partial and imperfect”, but this may be enough to signal identity with the reference group (Bell 1984:190). This partial shift, in effect, often means that an invariable feature in the speech of the reference group is turned into a variable feature in the style shift (Bell 1984:191).

Formal speech in diglossic communities is an example of long-term and institutionalized outgroup referee design (Bell 1984:189). The lack of a real-life representative is however not sufficiently accounted for in the model. Bell (1992:330) mentions in passing that the reference group for the High variety in diglossia is “distanced either by space or time”. This distinction is developed in Snow’s (2013) categorizes diglossic communities as a) modern diglossia, where the High variety is based on geographically distanced language community; b) revived diglossia, where the High variety is based on a historical state of the language that has been revived and reconstructed by scholarly efforts; and c) traditional diglossia, where the High variety is the classical language of a premodern empire and has been in continual use only for literary purposes and for spoken interaction (cf. Hudson 2002). Examples of modern diglossia include Switzerland, with the Standard German High variety, and Hong Kong, with the Standard Chinese High variety. Here reference group is are geographically distant, but their speech is well known and readily available through mass-media and, at least by some members of the community, also by direct contact. In revived and traditional diglossia, no referee or model speaker is available. Arabic is a case of traditional diglossia, and for speakers acting in a situation where they are expected to use SA (the High variety) there is no reference group, no real-life speaker, toward which the they can style shift. The reference speaker must therefore be constructed.

The characteristics of the constructed SA are not clear, as speakers may draw on a variety of sources when constructing it, including textual and aural language input of various forms, as well as formal language instruction. It is also not clear if and to what extent the constructed referee differs between speakers. That is, speakers of SA may have different model speakers towards which thy converge, style shifting in slightly different directions.

Formal Spoken Arabic

Speech in formal public situations, where speakers are expected to use SA, is characterized by a mixture of standard and non-standard features. This register will here be referred to as Formal Spoken Arabic, defined as the most formal register available to a speaker in unscripted speech. Linguists have struggled to define and describe these mixed forms of Arabic with its high degree of variability and complex interactions between SA and vernacular forms. One approach has been to divide forms of speech into “levels”, defined by varying proportions of SA and vernacular features or by associations with certain domains, situations, and/or group of speakers (Badawı̄ 1973; Blanc 1964; Meiseles 1980). One version of the level-approach has been to propose one stable “middle” variety with its own grammar, used for formal or semi-formal oral interaction (El-Hassan 1977; El-Hassan 1978; Mitchell 1980). Other studies have approached variation in Formal Spoken Arabic by quantitatively comparing speakers acting in the same formal public setting, in radio broadcast discussion programs (Schulz 1981), academic panel discussions (Mejdell 2006), or by instructing informants to use SA in an interview (Parkinson 1994a; Parkinson 2003). These studies show a large degree of individual variation on all variables studied, so that speakers acting in the same situation differ in what SA features they employ and how much, and in which non-standard features they avoid.

A particularly salient feature in style shifts is CMI. CMI has a highly symbolic value as a marker of standardness, inculcated by traditional grammar writing and educational practices. In the early stages of codification of Arabic in the ninth century, an ideology of linguistic correctness developed according to which speech without CMI was seen as incorrect, ungrammatical, incomplete, or even unintelligible (Ayoub 2006:629f; Versteegh 1983; Bohas et al. 1990:50), and CMI, and especially case inflection, continues to this day to be the focal point of Arabic language instruction in Arabic speaking countries (Uhlmann 2012:104). Indeed, ‘grammar’ (Ar. qawāʿid) is for many people synonymous to the system of CMI (Haeri 2003:40; Ibrahim 1983:512).

Despite the importance ascribed to CMI in the language community, few speakers have good command of it. It is exceedingly rare even for highly educated speakers of Arabic to be able to correctly and consistently use inflectional forms in a manner complying with the traditional prescriptive grammar (Beeston 1970:53; Kaye 1972:43; Ibrahim 1983:511; Saiegh-Haddad et al. 2016:3; Parkinson 1994b; Parkinson 1994a). For syntactic disambiguation, CMI is almost completely superfluous, since case roles are determined by word order and verbal agreement, and mood by sets of particle preceding the verb (Corriente 1971; Corriente 1973; Holes 2004:17, 173). Furthermore, non-orthographic CMI is generally assumed not to be phonologically encoded in silent reading (Stetkevych 2006:84; Bateson 1967:81f; Saiegh-Haddad et al. 2016:3; Taouka et al. 2004:50), and morphological case accordingly play only a marginal role in the parsing of written text (Hallberg et al. 2020). Speakers of Arabic therefore have very little practical incentive to master the systems of CMI, other than for oral performances in formal public settings.

We currently have only limited knowledge of to what extent and in what grammatical contexts CMI is used in Formal Spoken Arabic and how it varies between speakers. Meiseles (1977) and Magidow (2012) investigated CMI by analyzing selected examples sentence from radio and television broadcast. This method is problematic due to the variation in the use of CMI even within individual speakers, making it difficult to draw conclusions from isolated examples. Schulz (1981) investigated occurrences of a number of standard and non-standard features, including CMI, in discussion programs in discussion programs in Egyptian radio. He found wide variation in the use of CMI. Of the 49 speakers in his corpus, 29 never inflected verbs for mood and 17 speakers never inflected nominals for case. For speakers who did use case inflection it was more common in the accusative, largely due to speakers producing the accusative -an ending. He attributed this to this ending being orthographically represented (see below). Elgibali (1985) counted occurrences of case inflection in material from Egypt and Kuwait representing the five diglossic levels proposed by Badawı̄ (1973). He found progressively more case endings going up the levels, and more in the formal than in the informal situation in each level. The study does, however, have several methodological problems, particularly with the selection of material. For example, no distinction is made between unscripted speech, reading aloud, and recitation of the Quran. Parkinson (1994a) provides detailed descriptions of SA oral performances by four speakers in Cairo who in an interview were asked to speak SA. These speakers were hand-picked from a larger pool of speakers to represent the width of variation in all the collected material. The speaker with the lowest level of education (one year in high school) added CMI more or less at random, seemingly without understanding their function as grammatical markers. On the other extreme, one of the four speakers used CMI consistently, in a performance that was “utterly unique” in its extreme adherence to prescriptive grammar (Parkinson 1994a:207). Parkinson describes this performance as one which “could only come from someone who is somewhat withdrawn from an engaged role in the more day to day society around him” (Parkinson 1994a:209). The other two speakers used CMI to different degrees but with similar patterns of distribution: both used CMI at higher rates on words with enclitic pronouns and where the inflection is orthographically represented, and hardly ever on words with definite article.

This article builds on these studies by testing these results on a larger corpus and by using a more detailed coding scheme, allowing for a more fine grained analysis of variation in CMI.

# Method

## Corpus collection

The material of this study consists of 17 interviews with political public figures, broadcast in the program Liqāʾ al-yawm [Today’s meeting] on the Al Jazeera news channel during 2010 and 2011. The program has the format of a traditional ‘news interview’ (Heritage 1985; Clayman et al. 2002) and is as such an unscripted but highly formal situation. The speakers are assumed to use the most formal register available to them in extemporaneous speech. From the initial pool of 118 programs broadcast during 2010 and 2011, interviews with speakers of Syrian (5), Palestinian (7), and Egyptian Arabic (5) were included in the corpus. All these speakers are male. (Of all the 118 interviews broadcasts during 2010 and 2011, only two were with females.) Ages range from 50 to 81 years (mean 65.1, ). The corpus includes some prominent figures, such as Muḥammad Mursī of the Muslim Brotherhood in Egypt, who shortly after this interview was elected president of Egypt; Sallām Fayyāḍ, prime minister of Palestine 2007–2013; and Walīd al-Muʿallim, Syrian foreign minister from 2006 to the time of writing. The programs are publicly available as video on Al Jazeera’s YouTube channel[[1]](#footnote-1) or as audio on the Al Jazeera news site.[[2]](#footnote-2) See the appendix for the list of speakers and URLs to the published recordings. Transcripts of the interviews were retrieved from the Al Jazeera news site, transliterated, adapted to the CHAT format (MacWhinney 2000) for ease of analysis, and edited to accurately reflect the recording.

Annotation

Nominals and imperfect verbs (the form where mood inflection applies) were annotated for realization of CMI (inflected, uninflected, ambiguously inflected, incorrectly inflected, inaudible, or not applicable) and for morphological and syntactic parameters determining the prescribed form of the CMI. Nominals were annotated for (a) realization case inflection; (b) syntactic role; (c) head/attribute; (d) definiteness; and (e) declension. Verbs were annotated for (a) realization of mood inflection; (b) syntactic position; (c) person; (d) gender; (e) number; and (f) conjugation. Nominals were annotated for declension following Ryding (2005:184ff). Verbs were annotated for conjugation only to the extent to which mood inflection is affected, giving three declensions: standard (e.g., yaḏhab ‘go’), “hollow” (stems with medial root consonant w or y realized as a long vowel, e.g., yaqūl ‘say’), and “defective” (stems with final root consonant w or y realized as a long vowel, e.g., yaʾtī ‘come’). Defective verbs were further divided into three categories depending on the final vowel (-ā, -ī, or -ū). Tokens were also segmented for proclitic, stem, and inflectional suffix (e.g., wa-yaḏhab-ūn ‘and-3M.go-PL.IND’).

Some categories of words and expression are systematically uninflected for case and mood, or are otherwise methodologically problematic. These categories were excluded from annotation but marked up for their respective category according to the following: (a) titles and forms of address (e.g., ustaḏ, appr. ‘Mister’); (b) proper nouns; (c) repetitions, corrections, and reformulations; (d) formulaic expressions (e.g., al-ḥamd-u li-lāh ‘praise-NOM to-God’); (e) words with non-standard stems (e.g., ḥaga ‘thing’) or non-standard affixes (e.g., b-fakkir ‘IND-1SG.think’); (f) numerals above 10; (g) inaudible words; and (h) quotations (including reuse of formulations in the interviewer’s question).

Furthermore, adverbs were excluded from all analyses. Many adverbs are derived from adjectives with the suffix -an, which is commonly analyzed as accusative case inflection. It is, however, for several reasons best regarded as derivation rather than inflection. First, it changes the syntactic category of the word from noun or adjective to adverbs. Second, many frequent adverbs formed with ‑an, such as jiddan ‘very’ and ayḍan ‘also’, do not occur in the un-derived form (\*jidd, \*ayḍ). Third, when adverbs are formed productively, the ‑an suffix is used consistently, as opposed to when the ending indicates a case role in the argument structure of the verb. There are, for example, 44 tokens in the present corpus of adverbs formed from adjectives with final ‑ī/-iyy (the so called nisba adjectives, e.g., juzʾiyy-an ‘partly’), and in none of these is the -an ending omitted. When this same ending is used to mark a case relation it is often omitted, as detailed below.

A number of checks were performed to locate and correct erroneously annotated words. All inflectional prefixes and suffixes were checked for legal combinations with annotation tags, and lists of word stems were generated for each annotated declension and conjugation and manually checked for erroneous members.

Ambiguous forms of inflection

A complication in the annotation of CMI is the overlap between SA word forms that encode mood or case, and forms of the vernacular substrate that do not. The nominal suffix -īn, for example, encodes masculine, plural, and non-nominative case in SA, but only masculine and plural in non-standard varieties. It is therefore often impossible to assess whether this suffix on a given word is used as a realization of SA or non-standard morphology, especially since speakers switch between the two systems in unpredictable ways. Such overlapping forms were coded as having ambiguous CMI. The most frequent instances of such ambiguous forms of case inflection are -a or -i preceding the definite article l- on the following word; -īn in the sound masculine plural declension; --ū in ‘the five nouns’ declension with annexation (e.g., axū-na ‘our brother’); -ayn/-ēn in the dual declension; the enclitic possessive pronouns -ak, -ik, and -u(h) (2ms., 2mf, and 3ms), where the vowel can be interpreted either as SA CMI or as part of the non-standard enclitic pronoun. The most frequent forms with ambiguous mood inflection are (a) -a preceding the definite article l- on the following word; (c) -u in 3m.pl.; (d) -i in 2f.s.; (e) the enclitic possessive pronouns -ak and -u(h) (2ms. and 3ms.); and (b) -ø in the jussive mood. Regarding the latter point, the null-ending is shared by the non-standard verb form and the SA jussive inflected form.Technically, therefore, all tokens with a null-ending are ambiguous. This would, however, render all possible realizations of verbs either inflected or ambiguously inflected, making any analysis of occurrences of CMI on verbs impossible. The choice was therefor made to annotate the null-ending on verbs in the indicative and subjunctive positions as uninflected for mood, rather than ambiguously inflected.

Tokens annotated as ambiguous as to CMI were excluded from most analysis. However, some categories only have forms that are inflected, or ambiguously inflected, but not uninflected. For these categories, there are therefore no negative observations of inflection, and any count of CMI after exclusion of ambiguous forms, would necessarily give 100% inflected forms. To mitigate this, several analyses were performed on a dataset with these categories filtered out. This dataset consists of a subset of 13,808 tokens and will in the following be referred to as the disambiguated dataset.[[3]](#footnote-3) Furthermore, words with incorrect CMI, that are of a word class or in morphological situations where CMI is not applicable, or with inaudible CMI were excluded in the analyses below. These together make up 4.5% of the original data.

In some situations, the principles for coding a word form as ambiguously inflected differed between speakers, depending on their dialectal substrate. For example, in Egyptian Arabic an epenthetic vowel is inserted after the word stem to break up consonant clusters resulting from the addition of an enclitic pronoun (Aboul-Fetouh 1969; Woidich 2006). In the word dars‑a‑na ‘lesson‑our’, for example, the epenthetic vowel a occupies the same place as a potential case marking vowel, between the stem and the enclitic pronouns, and is often indistinguishable from it. Such tokens were coded as ambiguous, but only for speakers of Egyptian Arabic.

Statistical analysis

A database of annotated tokens was generated from the transcripts using a script written in R (R Core Team 2013). In addition to annotated information, the script extracted a number of other variables for each token, including proclitic, word stem, suffix, speaker, beginning and end time of the utterance, and preceding and succeeding word. Significance tests were perform on generalized regression models on non-aggregated data with the lme4 package in R (Bates et al. 2015), with significance levels calculated with the lmerTest package (Kuznetsova et al. 2017). Models were performed with speaker intercept and slope as random effect in a maximal random effects structure (Barr et al. 2013), with speaker slope removed as random effect where models did not converge.

Counting only speech by the interviewees, the corpus comprises a total 5 hours and 22 minutes of continuous speech and 38,000 words. The database includes 15,000 are nominals and 3,500 are imperfective verbs annotated for statistical analysis.

Analysis

Overall rates of inflection and general characteristics of speech styles

The rate of CMI varies considerably between speakers in the corpus, even though speakers perform under very similar circumstances. Rates of CMI for speakers as calculated from the disambiguated dataset are listed in table 1, with speakers ordered from top to bottom by rate of case inflection. The use of case inflection ranges from 0.2% (one token) to 42.2%, clustering around the lower end, with 13 of the 17 speakers performing below the mean of 7.5%. Mood inflection is used at higher rates on average: one speaker inflects verbs for mood at a rate of 71%, three speakers around 25%, and the rest below the mean of 9.9%, while three speakers do not inflect a single word for mood. Rates of inflection of case and mood are correlated ( using Spearman’s rank coefficient). Speakers who use more of one type of inflection reliably also use more of the other, and they accordingly rank in similar ways on both measures. Note that the speaker with the highest rates of CMI, Tayzīnī, while being something of an extreme, is still far from prescriptive use of CMI. Note also that non-standard lexemes were excluded from annotation. A speaker who use a large proportion of non-standard words can therefore potentially score high on this measure of CMI, as indeed is the case for Abū Majd, described presently.

Table 1: Rates of case and mood inflection by speaker

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | case |  | mood |  |
|  | % infl. | n | % infl. | n |
| Tayzīnī | 42.2 | 516 | 68.5 | 92 |
| Badīʿ | 18.3 | 712 | 23.8 | 231 |
| al-Qaddūmī | 15.6 | 475 | 5.5 | 109 |
| Abū Majd | 10.6 | 595 | 26.2 | 149 |
| Mursī | 7.4 | 539 | 24.3 | 136 |
| Kayālī | 6.9 | 635 | 2.1 | 141 |
| Fayyāḍ | 4.6 | 819 | 2.8 | 144 |
| al-Muʿallim | 4.1 | 441 | 2 | 102 |
| al-Barādiʿī | 3.9 | 664 | 3.1 | 261 |
| ʿAbd al-Qādir | 3.9 | 645 | 0.7 | 147 |
| al-Xuḍarī | 2 | 842 | 0.5 | 214 |
| Hilāl | 1.9 | 646 | 0 | 69 |
| al-ʾAsʿad | 1.7 | 526 | 0.8 | 130 |
| Šallaḥ | 1.6 | 1204 | 2.5 | 326 |
| Ġalyūn | 1.5 | 675 | 5.8 | 121 |
| ʿArīqāt | 0.4 | 711 | 0 | 151 |
| al-Miṣrī | 0.2 | 461 | 0 | 98 |
|  |  |  |  |  |
| Mean | 7.5 |  | 9.9 |  |
| Median | 3.9 |  | 2.5 |  |
| Total |  | 11106 |  | 2621 |

All speakers, accept one, use an unmistakably formal register that can be described as SA with occasional use of CMI, and some admixture of non-standard features. The formal register is most clearly seen in the predominant use of standard variants of frequent words, such as the standard yurīd ‘want’ instead of the non-standard bəddu (Palestinian, Syrian) or ʿāyəz (Egyptian), and in the use of SA phonological forms, such as [q] (qāl ‘say’), instead of the non-standard [ʾ] (ʾāl).

Three speakers stand out in the corpus in using styles that deviate from this overall picture. The first is Tayzīnī, Professor of philosophy at Damascus university and a distinguished Marxist theoretician. He is an outlier in terms of the very high rates of CMI and the complete lack of dialectal features in his speech, making for a style that is strikingly bookish. He also uses some strikingly archaic linguistic features, for example in (1), with the archaic, or Quranic, negative particle ʾin. It is coupled with consistent use of CMI except preceding the sentence final pause.

wa-huwa inna mā taqūl-ūna-hu min mā yaḥduṯ-u fī sūrīya in huwa illā natījat-u muʾāmarat-in xārijiyya  
and-it that what 2MPL.say-it of what 3MS.happens-IND in Syria not it but result-NOM conspiracy-GEN foreign  
‘And that is: that which you say about what is happening in Syria is that it is nothing but the result of a foreign conspiracy.’ (Tayzīnī, 20:14)

On the other end of the spectrum is al-Miṣrī, a Palestinian businessman and former minister and MP. In the interview he produces only two instances of case inflection (one of which is in a sound masculine plural and therefore not included in table 1) and no instance of mood inflection. His speech is dominated by Palestinian Arabic features, with occasional admixture of SA, as opposed to the other way around, making him the only speaker in the corpus who does not use an unmistakably formal register. He uses, for example, the non-standard Palestino-Syrian pseudo-verb biddu ‘want’ (Cowell 2005:384) 49 times in the interview and the SA variant yurīd only twice. The other sixteen speakers together use biddu or ʿāyiz 30 times in total. While his speech stands out with a high proportion of non-standard features, his infrequent use of CMI is by no means exceptional: two other speakers make no use of mood inflection and five others inflect words for case at a rate less than two percent.

Finally, Abū Majd, minister of media in Egypt, stands out in his frequent and drastic switches between saliently standard and no-standard forms. As can be seen in table 1, he uses a relative high rate of CMI, but he combines this with heavy use of saliently non-standard words, often in the same utterance. In (2), for example of this, the first part of the utterance is saliently non-standard Egyptian Arabic (marked with bold face in the example), and the second part is SA, with standard phonology and with CMI on two of the three words.

ana miš āyiz qināʿ yukallim-u qināʿ-an  
I not want mask 3MS.address-IND mask-ACC  
‘I do not want a mask talking to a mask.’ (Abū Majd, 24:03)

These three speakers illustrate the wide range and variability in the use of CMI in individual styles of Formal Spoken Arabic. Despite this variability, there are some striking similarities in how speakers distribute the CMI that they do use in different morphosyntactic situations.

Inflection by syntactic role

The most obvious possible pattern of CMI to be tested is one of syntactic position, so that one specific case or mood is inflected more than others. This is however not the case. Except for the accusative, where the effect of orthographic representation comes into play, syntactic role plays little to no role for the distribution of CMI. With regards to mood, verbs in the disambiguated dataset are inflected from the indicative at a rate of 8.7% (), and for the subjunctive at 8.9% (). The difference is not significant (). The jussive and the imperative moods do not differ from non-standard forms in term of inflectional endings and were not included in the disambiguated dataset. (For the special case of jussive inflection of hollow verbs, see below.) For nominals, words are more often inflected for accusative than for the nominative or genitive. In the disambiguated dataset, 4.8% () of nominals in nominative position, 4.7% () in genitive position are inflected for case, and 16.1% () in accusative position are inflected for case. The much higher rate of inflection in the accusative is significantly different from inflection in both the nominative () and the genitive (). Indefinite accusatives in the highly frequent triptote declension take case inflection that is orthographically represented (see table ¿tbl:caseinflection?). If these indefinite nouns are excluded, however, the difference in case inflection disappears: 5.85% () for nominative, 4.0% () for genitive, and 4.7% () for accusative, with none of these differences being significant. Further evidence that the higher rate of accusative inflection is an effect of orthography rather they syntax is presented below. Over all, then, if the effect of orthography is controlled for, the different cases and moods are inflected at similar rates.

Words favored for inflection

Enclitic pronouns

Words with enclitic pronouns show rates of CMI for above the overall average. Parkinson (1994a) observed for two of his four speakers (those not representing either end of the spectrum of variation) that words with enclitic pronouns tend to be more often inflected for case and mood than other words. This pattern is confirmed in the current study. It is in fact quite striking and highly consistent across speakers. Rates of CMI in words with and without enclitic pronouns in the disambiguated dataset are listed in table 2. Words without enclitic pronoun are inflected occasionally, at 5.33% and 7.31% for case and mood respectively. Words with enclitic pronouns are much less frequent with only a couple of hundred tokens, or 4% of the total data. There words are however inflected at very high rates: 40.4% and 41.03% for case and mood respectively. The differences in rates of inflection is significant for both case () and mood (). The effect is very consistent among individual speakers, as illustrated in figure 1. Except for a few speakers at the lower end, all speakers drastically increase their rates of CMI in words with enclitic pronouns, as compared to other words.

Table 2: Rates of inflection in words with and without enclitic pronoun

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | w/o enclitic pron. |  | with enclitic pron. |  |
|  | % infl. |  | % infl. |  |
| Case | 5.33 | 10664 | 40.5 | 442 |
| Mood | 7.31 | 2504 | 41.03 | 117 |

Figure 1: Rate of inflection in words with and without enclitic pronoun.

Figure 1: Rate of inflection in words with and without enclitic pronoun.

There are further indications, apart from the high rates of inflection, that speakers feel an particular expectation to produce CMI in words with enclitic pronouns. The first such indication is these words are over-represented for prescriptively incorrect CMI. Words with enclitic pronouns make up almost half of words with incorrect CMI, 30 of 68 tokens, but only 6.3% of the total data. Speaker thus seem to be under some pressure to inflect these words for case, . A second indication comes from one particular speaker, al-Asʿad, founder and commander of the Free Syrian Army. He begins the interview with consistent use of CMI on words with enclitic pronouns, on seven words in close succession. He then produces two incorrect forms, and thereafter abandons this way of speaking, producing only seven instances of CMI thereafter, evenly distributed in the rest of 23 minute interview. This speakers appears to bee constructing a reference speaker with constant use of CMI on words with enclitic pronouns, but finds himself lacking the proficiency to shift towards this aspect of the reference speaker, and abandons it. As seen in figure 1, several other speakers are very successful in sifting towards a similarly constructed referee.

Orthographic CMI

A second category of words that have a high rate of CMI in speech are those where the inflectional ending has an orthographic representation. In writing, CMI is mostly absent due to the defective nature of the Arabic writing system, in which short vowels and nunation are represented only with optional diacritics. Forms of CMI can thus be categorized as being orthographic or non-orthographic, depending on whether the inflectional ending is represented by an addition or change in the letter sequence of the word, or only by diacritics. Orthographic CMI occurs in six situation for case and four situations for mood. These situations are highly diverse, representing various combinations of case and mood with inflectional paradigms, definiteness, grammatical person, and properties of the word stem. These situations are listed in table 3. Because of this diversity, orthographic and non-orthographic CMI can often not be straightforwardly compared in terms of the rate of realization of the inflection, without conflating this effect with other variables. Furthermore, some forms of orthographic CMI are very scarce in the data. Nevertheless, the counts of inflected and uninflected tokens with orthographic CMI give a clear indication that tokens with orthographic CMI are inflected at high rates in speech, as compared to the overall rates discussed above. Types of orthographically marked CMI with 100 or more tokens (A1, A2, and B1 in table 3) all have rates of inflection of 40% and above, to be compared to the overall speaker average of 7.5%.

Table 3: Rates of orthographic case and mood inflection. The inflected and uninflected forms are separated by slash.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | %infl. |  |
| A. | case |  |  |
| 1. | Indefinites triptote accusative[[4]](#footnote-4) | 45.5 | 497 |
|  | (bayt-an/bayt بيتا/بيت ‘house’) |  |  |
| 2. | Sound masculine plural nominative | 39.6 | 101 |
|  | (miṣriyy-ūn/miṣriyy-īn مصريون/مصريين ‘Egyptians’) |  |  |
| 3. | Triptotes with stem-final ʾ and enclitic pronoun | 84.2 | 19 |
|  | (mabdaʾ-u-na, mabdaʾ-a-na, mabdaʾ-i-na/mabdaʾ-na |  |  |
|  | مبدؤنا, مبدأنا, مبدئنا ‘principle-our’) |  |  |
| 4. | Dual nominative | 41.2 | 17 |
|  | (dawlat-ān/dawlat-aynēn دولتان/دولتين ‘state-DUA’) |  |  |
| 5. | ‘The five nouns’ accusative or genitive in annexed form | 33.3 | 3 |
|  | (ʾax-ī, ʾax-ā/ʾax-ū أخي, أخا/أخو ‘brother’) |  |  |
| 6. | Defective accusative indefinite | 0.0 | 1 |
|  | (qāḍiy-an/qāḍiy قاضيا/قاض ‘judge’) |  |  |
|  |  |  |  |
| B. | mood |  |  |
| 1. | 2/3mpl. in indicative | 59.2 | 162 |
|  | (yaḏhab-ūn(a)/yaḏhab-u يذهبون/تذهبوا ‘go.3ms’) |  |  |
| 2. | 1s., 1pl., 2ms., 3ms., and 3ms. hollow jussive | 97.7 | 43 |
|  | (yakun/yakūn يكون/يكن ‘was.3ms not’) |  |  |
| 3. | 2/3d. indicative | 100.0 | 2 |
|  | (yaḏhab-ān, yaḏhab-ā يذهبان, يذهبا ‘go.2md’) |  |  |
| 4. | 2fs. indicative | n/a | 0 |
|  | (taḏhab-īn/taḏhab-i تذهبين/تذهبي ‘go.2fs’) |  |  |

The pattern of high rates of inflection of orthographic CMI is most clearly shown in indefinite accusative nominals of the triptote declension (A1 in table 3), for which there is more data than in the other categories. For indefinite triptotes in nominative and genitive positions, where the case inflection is non-orthographic, 4.6% () of tokens are inflected for case. For tokens in accusative position, where case inflection is orthographic, 32.9% () of tokens are inflected. This difference is significant (). Indeed, every single speaker have a higher rate of case inflection in the accusative than in the other two cases for this type of word. Note that since adverbs are not included, frequent lexicalized forms such as kaṯīran ‘a lot’ or abadan ‘ever’ are not included in these numbers.

In the analysis above, we do not only compare the existence of an orthographic representation, but also syntactic role (nominative/genitive vs. accusative), and it is therefore possible that the increased rate of inflection is due to the syntactic role of the accusative, rather then by the presence of an orthographic representation of the inflection. For certain types of word stems, however, the orthographic representation of the accusative marking ‑an suffix is canceled. This provides a useful testing ground to further establish the effect of orthography on the rate of case inflection. The orthographic representation of the accusative marking suffix ‑an is canceled on word stems with the feminine ending ‑a(t) ة, and on nouns with stem-final -aʾ أ, or -āʾ اء. For tokens with a prescribed accusative inflectional ending -an, the rate of inflection in words where this ending is orthographically represented was compared to words where it is not. This comparison isolates the effect of orthography, since the phonological form of the ending and the syntactic role is the same for all words being compared. For tokens with an orthographic accusative -an inflection, the suffix is realized at a rate of 45.47% (). For tokens with a non-orthographic accusative -an inflection, the corresponding rate is 9.96% (). The difference is significant (). This effect is highly consistent across speakers, as illustrated in figure 2. Thus, in the same morphological and syntactic situation, where the accusative inflection is phonologically identical, the rate of inflection in speech is highly affected by the availability of an orthographic representation of the inflectional ending.

Figure 2: Rates of inflection for words with a prescibed accusative ‑an ending, with and without an orthographic representation.

Figure 2: Rates of inflection for words with a prescibed accusative ‑an ending, with and without an orthographic representation.

A particularly interesting case of orthographic CMI is inflection in the jussive mood of the so called hollow verbs, verbs with a medial root consonant w or y (B2 in table 3). This form seems to have a special status in the system of CMI in formal spoken Arabic in being consistently applied by all speakers. In this class of verbs, the middle root surfaces as a stem-internal long vowel, as illustrated in (3). In the jussive mood, which most importantly occurs after the past negative particle lam, the long vowel is shortened, as in (4). This is the only form of CMI that affects the internal form of the stem.

yaqūl(-u)  
3MS.say(-IND)  
‘he says’

lam yaqul  
NEG.PAST 3MS.say  
‘he did not say’

There are 43 instances in the corpus of hollow verbs in a position where they prescriptively take a shortened jussive form, represented in fourteen of the seventeen speakers. These tokens are inflected for jussive with the shortened form in all instances accept one, reproduced in (5). Here, the verb tasīr ‘proceed’ is in jussive position after the negation lam, but it is not produced by the speaker in its shortened jussive form tasir, as would be prescriptively correct.

al-qaḍāya l-iqtṣādiyya kull-ha lam tasīr bi-manṭiq iqtiṣādi ʿilmi  
DEF-matters DEF-economic all-3MS NEG.PAST 3FS.proceed with-logic economic scientific  
‘The economic matters did not proceed in a logical, economical, and scientific manner.’ (Kayālī, 12:04)

There is a second, synonymous way to negate past tense verbs in SA, with the particle mā and a verb in the perfect stem, as in (6). This form of past tense negation, while prescriptively corrects, is also used in non-standard varieties of Arabic and is therefore stylistically less formal, and less frequent in SA [Van Mol (2006); 264]. It does, however, provide speakers an alternative to the jussive inflected form and way to opt out of this form of CMI.

mā qāl  
NEG.PAST 3MS.say  
‘he did not say’

Nevertheless, for speakers who do use the jussive form of past tense negation, which is the majority of speakers in the present corpus, the jussive inflection is applied very consistently. This could, however, not be statistically tested against other forms of jussive inflection, since in other words the jussive inflected form it coincides with the non-standard form and is therefore ambiguous.

Over-all, then, there is a clear correlation between orthographic representation of CMI and the rates of realization of CMI in speech.

Words disfavored for inflection

Above, two situations where CMI is favored were identified. Equally important for a description of CMI in formal spoken Arabic are situations that are disfavored for inflection. These are, however, more difficult to detect due to the over-all low rates of inflection. For mood, no position that is disfavored for inflection could reliably be identified. One candidate is the subjunctive ‑a inflection on verbs of the defective paradigm (e.g. yantahiy‑a ‘3MS.end-SUB’ and yanmuw‑a ‘3MS.grow‑SUB’). There are 43 tokens in the corpus of defective verbs with prescribed ‑a ending, none of which are inflected for mood, indicating that this inflected form may not be part of Formal Spoken Arabic. However, this zero rate of inflection was not significantly different from the 9.6% () rate of inflection of tokens in other declension with prescribed subjunctive -a inflection ().

For nominals, on the other hand, words with definite article (a)l- stand out as clearly disfavored for case inflection. Nominals were coded for four types of definiteness: (a) definite article; (b) construct state; (c) indefinite; and (d) enclitic pronoun. Examples if these types of definiteness together with percentages of case inflection in the disambiguated dataset are listed in table 4. As was shown above, words with enclitic pronoun are strongly favored for case marking. Words with definite article are, on the other hand, clearly disfavored. Only 1.15% of tokens with the definite article in the disambiguated dataset are inflected for case, making them the type of definiteness with the lowest rate of inflection. The rate of inflection in words with definite article is different from that of each of the three other types of definiteness (). Words with definite article make up almost half of all nominals in the corpus, meaning that the fact that these are disfavored for case marking in and of itself accounts for a large part of the overall lack of case inflection in the corpus.

Table 4: Percentage of case inflection by type of definiteness

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | % infl. | n | example |  |
| Definite article | 1.1 | 5136 | al-bayt(-u) | ‘DEF-house(-NOM)’ |
| Construct state | 5.4 | 1533 | bayt(-u) aḥmad | ‘house(-NOM) Ahmed’ |
| Indefinite | 11.7 | 3994 | bayt(-un) | ‘house(-NOM)’ |
| Enclitic pronoun | 40.5 | 441 | bayt(-u)-na | ‘house(-NOM)-our’ |

The patterns of words with definite article being disfavored for case inflection is consistent across speakers, as illustrated in figure 3. The figure shows how for all speakers who make some use of case inflection, the rate of inflection is drastically reduced in words with definite article as compared to other types of definiteness. Tayzīnī, the speaker with by far the highest rate of CMI, inflects case on tokens with definite article at a rate of 10.8%. In fact, of the 59 tokens of inflected case on words with the definite article in the disambiguated dataset, around half, 29 tokens, are produced by this one speaker. While this is a high rate as compared to the other speakers, it is far below his rates in other types of definiteness. Excluding Tayzīnī from the data, only 0.007% of tokens with definite article in the disambiguated dataset are inflected for case.

Figure 3: Rates of case inflection by definiteness.

Figure 3: Rates of case inflection by definiteness.

Discussion

This paper set out to answer two questions about variation in CMI in Formal Spoken Arabic: how it varies between speakers and whether it is limited by absolute or probabilistic constraints. These questions were posed in order to characterize the constructed SA reference speaker. The results aligns with previous research in showing considerable inter-speaker variation in the use of CMI. It adds to it by providing evidence of a ceiling level of CMI of around 50%, and in demonstrating that CMI is highly determined by morphosyntactic parameters.

For the morphosyntactic distribution of CMI, three distinctive patterns were found. First, diacritic CMI is very rare in nominals with definite article. This has far reaching consequences, since these words constitutes a large part of natural speech: 65.6% of all nominal tokens, and 53.2% of the total number of tokens (nominal and verbal) are of this category.

This restriction by itself, therefor, sets a ceiling level for the global rate of CMI at below 50%. This pattern is remarkable in that it is in stark contrast with prescriptive grammar as a consistent pattern of omission of CMI. There is to the best of my knowledge no indications in Arabic grammars and teaching materials that words with definite article are not to be inflected for case in speech, and speakers of Arabic, such as those appearing in the present corpus, likely receive no explicit instruction to omit CMI on these words. On the contrary, CMI, and particularly case inflection, is typically presented as an indispensable, core part of SA grammar and is a focal point of Arabic language instructing (Bohas et al. 1990:50; Haeri 2003:40; Uhlmann 2012:104; Ibrahim 1983:152). This indicates that the constructed SA reference speaker that represents the direction of formal style shift in Arabic does not feature diacritic CMI on words with definite article. On other words, a style shift to a formal register does not entail using diacritic case inflection on words with definite article.

The second way in which the use of CMI is structured in formal spoken Arabic is that all speakers, to the degree that they employ CMI, do so at much higher rates in words with orthographic CMI and on words with enclitic pronouns. There preference to use orthographic CMI is most likely an effect of reading. Given the fact that Formal Spoken Arabic is a highly specialized register with low density of communication providing limited aural input, we may assume that speakers draw heavily on the written register, to which speakers are more often exposed, when construction a SA reference speaker. In written SA, CMI is only represented in a few limited situations, but there with complete consistency, a pattern of CMI that is projected on the constructed reference speaker. Furthermore, speakers most likely develop a higher degree of proficiency in CMI in situations where it is orthographically marked, seeing to the dominance of written language in the SA input. When speaking, they draw on this proficiency to include some CMI in their speech, resulting in a disproportionate amount of orthographic CMI.

The preference to use CMI on words with enclitic pronouns is not as easily explained. One possible explanation for this may be reanalysis of aural input. One of the sources of input for the construction of the reference speaker may be prescriptively correct renderings of SA, for example in news broadcasts, voice-overs, and recitation. In these forms of SA, omission of diacritic CMI is mandatory preceding a breath pause (Holes 2004:62; Mitchell 1990:99; Nelson 2001:29). Forms such as balad-ø ‘[a] country’ or al-balad-ø ‘DEF-country-ø’, without CMI, are therefore very common in these forms of Arabic, and furthermore occur in perceptually salient positions at the end of clauses or phrases preceding a pause. Speakers constructing a reference speaker of SA may draw on such performances and generalized the pausal omission of CMI to also include other, non-pausal positions, giving, for example, the consistent omission of CMI for words with definite article discussed above. The prescriptive omission of CMI is, however, not applied on words with enclitic pronoun, so that forms such as balad‑ø-na ‘country-our’, with enclitic pronoun and omission of CMI, never occur in prescriptively correct renderings. There is therefore no omission to be generalized, and omission of CMI on these words is not projected to the constructed reference speaker.

Summary and conclusion

In this article, style shifts towards SA were interpreted as outgroup referee design (Bell 1984). It was argued that in the diglossic language situation, since the outgroup SA referee does not exist, this speaker is constructed to function as a target for style shifting. Data on the use of use CMI, a salient SA feature, was presented and interpreted according to this framework. The results indicate that the constructed referee, the imagined model speaker of SA, does not use CMI on words with definite article, but does so on the less frequent categories of words with enclitic pronoun and on words where CMI is orthographically represented. This puts a ceiling for possible style shift using CMI and furthermore indicates that speakers draw primarily on aural and textural input for constructing the referee, rather than on formal language instruction and prescriptive grammar. Indeed,

# Appendix

Below are listed the name, nationality, and occupation or title of speakers in the corpus, and URLs to the recording of the interview. Interviews are available either as video on Al Jazeera’s YouTube channel, or as audio on their webpage (last confirmed access 27 September 2019).

1. Ḥātim ʿAbd al-Qādir. Palestine. Minister of Jerusalem affairs.  
   [*https://www.youtube.com/watch?v=kR4cQZjZPQU*](https://www.youtube.com/watch?v=kR4cQZjZPQU)
2. Aḥmad Abū l-Majd. Egypt. Jurist; liberal Islamist thinker; former minister of youth and minister of media.  
   [*https://www.youtube.com/watch?v=qyoL92NpSN8*](https://www.youtube.com/watch?v=qyoL92NpSN8)
3. Ṣāʾib ʿArīqat. Palestine. Minister of negotiations.  
   [*http://www.aljazeera.net/audioplayer/42ef7b37-9c74-49ff-8368-ef4c6b41bcb2*](http://www.aljazeera.net/audioplayer/42ef7b37-9c74-49ff-8368-ef4c6b41bcb2)
4. Riyāḍ al-Asʿad. Syria. Founder and commander of the Free Syrian Army; former general, Syrian army.  
   [*https://www.youtube.com/watch?v=8arRIIt8PTU*](https://www.youtube.com/watch?v=8arRIIt8PTU)
5. Muḥammad Badīʿ. Egypt. Supreme guide of the Muslim Brotherhood.  
   [*https://www.youtube.com/watch?v=qlUOAXUnOFI*](https://www.youtube.com/watch?v=qlUOAXUnOFI)
6. Muḥammad al-Barādiʿī. Egypt. Founder of National Association for Change; Nobel Peace Price laureate, former director general of IAEA.  
   [*https://www.youtube.com/watch?v=Phi-NPJXUK4*](https://www.youtube.com/watch?v=Phi-NPJXUK4)
7. Salām Fayyāḍ. Palestine. Prime minister.  
   [*http://www.aljazeera.net/audioplayer/a0f24635-4a55-471a-968f-f096bdc2ed2a*](http://www.aljazeera.net/audioplayer/a0f24635-4a55-471a-968f-f096bdc2ed2a)
8. Burhān Ġalyūn. Syria. Head of the Syrian National Council.  
   [*https://www.youtube.com/watch?v=zkQxo\_ejsJc*](https://www.youtube.com/watch?v=zkQxo_ejsJc)
9. ʿAlī d-Dīn Hilāl. Egypt. Media Secretary, Naitional Democratic Party; former minister of youth.  
   [*https://www.youtube.com/watch?v=ht4j0rQrzYU*](https://www.youtube.com/watch?v=ht4j0rQrzYU)
10. Munīb al-Miṣrī. Palestine. Businessman; former minister of public works, Jordan; former MP, Fath.  
    [*https://www.youtube.com/watch?v=imWxttjk3u8*](https://www.youtube.com/watch?v=imWxttjk3u8)
11. Muḥammad Kayālī. Syria. Former head of military police.  
    [*https://www.youtube.com/watch?v=XvCWHsVmHgk*](https://www.youtube.com/watch?v=XvCWHsVmHgk)
12. Walīd al-Muʿallim. Syria. Minister of foreign affairs.  
    [*http://www.aljazeera.net/audioplayer/ff65db75-e6d5-4b72-9d54-f5431bac5de9*](http://www.aljazeera.net/audioplayer/ff65db75-e6d5-4b72-9d54-f5431bac5de9)
13. Muḥammad Mursī. Egypt. Chairman of the Freedom and Justice Party; affiliated with the Muslim Brotherhood.  
    [*https://www.youtube.com/watch?v=FJCZ7pR1IMg*](https://www.youtube.com/watch?v=FJCZ7pR1IMg)
14. Farūq al-Qaddūmī a.k.a. Abū Luṭf. Palestine. Cofounder of Fath.  
    [*http://www.aljazeera.net/audioplayer/0bfd7ece-f358-4c64-ba68-e6fc7887d157*](http://www.aljazeera.net/audioplayer/0bfd7ece-f358-4c64-ba68-e6fc7887d157)
15. Ramaḍān Šallaḥ. Palestine. General secretary of Islamic Jihad.  
    [*https://www.youtube.com/watch?v=aDJ3\_GO13jg*](https://www.youtube.com/watch?v=aDJ3_GO13jg)
16. Ṭayyib Tayzīnī. Syria. Proffesor of Philosophy, Damascus University, Marxist thinker.  
    [*https://www.youtube.com/watch?v=tf3XoeGpBr8*](https://www.youtube.com/watch?v=tf3XoeGpBr8)
17. Jamāl al-Xuḍarī. Palestine. Independent MP; Head of the Popular Committee Against Siege of Gaza.  
    [*https://www.youtube.com/watch?v=rZcd7vAx-hY*](https://www.youtube.com/watch?v=rZcd7vAx-hY)

# References

Aboul-Fetouh, Hilmi M. 1969. *A morphological study of Egyptian Colloquial Arabic*. The Hague: Mouton.

Al-Wer, Enam. 2013. Sociolinguistics. In J. Owens (ed) *The Oxford handbook of Arabic linguistics*, 244–263. Oxford: Oxford University Press.

Ayoub, Georgine. 2006. Laḥn. In C. H. M. Versteegh (ed) *Encyclopedia of Arabic language and linguistics*, 628–634. Leiden: Brill.

Badawı̄, as-Saʿı̄d Muammad. 1973. *Mustawayāt al-ʿarabiyya al-muʿāṣira fı̄ miṣr [the levels of contemporary Arabic in Egypt]*. Cairo: Dār al-maʿārif bi-Miṣr.

Barr, Dale J., Levy, Roger, Scheepers, Christoph, & Tily, Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68 (3):255–278.

Bates, Douglas, Mächler, Martin, Bolker, Ben, & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67 (1):1–48.

Bateson, Mary Catherine. 1967. *Arabic language handbook*. Washington, D.C: Center for Applied Linguistics.

Beeston, A. F. L. 1970. *The Arabic language today*. London: Hutchinson.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13 (2):145–204.

Bell, Allan. 1992. Hit and miss: Referee design in the dialects of new zealand television advertisements. *Language & Communication* 12 (3):327–340.

Bell, Allan. 2002. Back in style: Reworking audience design. In P. Eckert & J. R. Rickford (eds) *Style and sociolinguistic variation*, 139–161. Cambridge: Cambridge University Press.

Blanc, Haim. 1964. Stylistic variation in spoken Arabic: A sample of interdialectal educated conversation. In C. A. Ferguson (ed) *Contributions to Arabic linguistics*, 79–156. Cambridge, Mass.: Harvard University Press.

Bohas, Georges, Guillaume, Jean-Patrick, & Kouloughli, Djamel Eddine. 1990. *The Arabic linguistic tradition*. London: Routledge.

Boussafara-Omar, Naima. 2006. Diglossia. In C. H. M. Versteegh (ed) *Encyclopedia of Arabic language and linguistics*, 629–37. Leiden: Brill.

Clayman, Steven, & Heritage, John. 2002. *The news interview: Journalists and public figures on the air*. Cambridge: Cambridge University Press.

Corriente, Federico. 1973. Again on the functional yield of some synthetic devices in Arabic and semitic morphology (a reply to J. Blau). *The Jewish Quarterly Review* 64 (2):154–163.

Corriente, Federico C. 1971. On the functional yield of some synthetic devices in arabic and semitic morphology. *The Jewish Quarterly Review* 62 (1):20–50.

Cowell, Mark W. 2005. *A reference grammar of Syrian Arabic*. Washington, DC: Georgetown University Press.

Elgibali, Alaa. 1985. *Towards a sociolinguistic analysis of language variation in Arabic: Cairene and Kuwaiti dialects*. Pittsburgh: Unpublished Ph.D. thesis.

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.

El-Hassan, S. A. 1978. Variation in the demonstrative system in educated spoken arabic. *Archivum Linguisticum* 9:32–57.

Haeri, Niloofar. 2003. *Sacred language, ordinary people: Dilemmas of culture and politics in Egypt*. New York: Palgrave Macmillan.

Hallberg, Andreas, & Niehorster, Diederick C. 2020. Parsing written language with non-standard grammar: An eye-tracking study of case marking in arabic. *Reading and Writing: an interdisciplinary journal*.

Hary, Benjamin. 1996. The importance of the language continuum in Arabic multiglossia. In A. Elgibali & E.-S. M. Badawi (eds) *Understanding Arabic: Essays in contemporary Arabic linguistics in honor of El-Said Badawi*, 69–90. Cairo: American University in Cairo Press.

Heritage, John. 1985. Analyzing news interviews: Aspects of the production of talk for an overhearing audience. In T. A. van Dijk (ed) *Handbook of discourse analysis*, 95–117. London: Academic Press.

Holes, Clive. 1987. *Language variation and change in a modernising arab state: The case of bahrain*. London; New York: Kegan Paul International.

Holes, Clive. 2004. *Modern Arabic: Structures, functions, and varieties*. Washington, DC: Georgetown University Press.

Hudson, Alan. 2002. Outline of a theory of diglossia. *International Journal of the Sociology of Language*:1–48.

Ibrahim, Muhammad H. 1983. Linguistic distance and literacy in Arabic. *Journal of Pragmatics* 7 (5):507–515.

Ibrahim, Muhammad H. 1986. Standard and prestige language: A problem in Arabic sociolinguistics. *Anthropological Linguistics* 28 (1):115–126.

Kaye, Alan S. 1972. Remarks on diglossia in Arabic: Well-defined vs. ill-defined. *Linguistics* 10 (81):32–48.

Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82 (13):1–26.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2006. *Principles of linguistic change: Social factors* Digital print. Malden, MA: Blackwell.

MacWhinney, B. 2000. *The childes project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.

Magidow, Alexander. 2012. Explaining inconsistent case marking in spoken formal Arabic. *Zeitschrift für arabische Linguistik* (55):62–99.

Meiseles, Gustav. 1977. Restitution of “word-endings” in Modern Literary Arabic. *Israel Oriental Studies* 7:173–195.

Meiseles, Gustav. 1980. Educated Spoken Arabic and the Arabic language continuum. *Archivum Linguisticum* 11:118–148.

Mejdell, Gunvor. 2006. *Mixed styles in spoken Arabic in Egypt: Somewhere between order and chaos*. Boston, MA: Brill.

Mitchell, Terence Frederick. 1980. Dimensions of style in a grammar of Educated Spoken Arabic. *Archivum Linguisticum* 11:89–106.

Mitchell, Terence Frederick. 1990. *Pronouncing Arabic*. Oxford: Clarendon.

Nelson, Kristina. 2001. *The art of reciting the Qurʾan*. Cairo; New York: The American University in Cairo Press.

Owens, Jonathan. 2001. Arabic sociolinguistics. *Arabica* 48 (4):419–469.

Parkinson, Dilworth B. 1994a. Speaking fuṣḥā. In Y. Suleiman (ed) *Arabic sociolinguistics: Issues and perspectives*, 179–211. London; New York: Routledge.

Parkinson, Dilworth B. 1994b. Testing native speakers: Implications for teaching Arabic to non-native speakers. In E. N. McCarus, R. M. Rammuny, & D. B. Parkinson (eds) *Investigating Arabic: Linguistic, pedagogic and literary studies in honor of Ernest N. McCarus*, 191–203. Columbus, Ohio: Greyden Press.

Parkinson, Dilworth B. 2003. Verbal features in oral Fuṣḥa performances. *International Journal of the Sociology of Language* 163 (1):27–41.

R Core Team. 2013. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ryding, Karin C. 2005. *A reference grammar of Modern Standard Arabic*. Cambridge, UK: Cambridge University Press.

Saiegh-Haddad, Elinor, & Schiff, Rachel. 2016. The impact of diglossia on voweled and unvoweled word reading in Arabic: A developmental study from childhood to adolescence. *Scientific Studies of Reading* 20 (4):1–14.

Schulz, David Eugene. 1981. *Diglossia and variation in formal spoken Arabic in Egypt*. Ann Arbor: Unpublished Ph.D. thesis.

Snow, Don. 2013. Revisiting Ferguson’s defining cases of diglossia. *Journal of Multilingual and Multicultural Development* 34 (1):61–76.

Stetkevych, Jaroslav. 2006. *The Modern Arabic Literary Language: Lexical and stylistic developments*. Washington DC: Georgetown University Press.

Taouka, Miriam, & Coltheart, Max. 2004. The cognitive processes involved in learning to read in Arabic. *Reading and Writing* 17 (1):27–57.

Uhlmann, Allon J. 2012. Arabs and Arabic grammar instruction in Israeli universities: Alterity, alienation and dislocation. *Middle East Critique* 21 (1):101–116.

Van Mol, Mark. 2006. Evolution of MSA: The case of some complementary particles. In D. B. Parkinson & S. Farwaneh (eds) *International journal of corpus linguistics*, 135–147.

Versteegh, Kees. 1983. Arabic grammar and corruption of speech. *Al-Abhath* 31:139–60.

Woidich, Manfred. 2006. Cairo Arabic. In C. H. M. Versteegh (ed) *Encyclopedia of Arabic language and linguistics*, 323–333. Leiden: Brill.

1. <https://www.youtube.com/user/aljazeerachannel> [↑](#footnote-ref-1)
2. <http://www.aljazeera.net> [↑](#footnote-ref-2)
3. The following categories of words were excluded in the disambiguated dataset: for nominals (a) ‘the five nouns’ with annexed noun or pronoun (*abū-k ‘father-your’*); (b) sound masculine plural (*miṣriy-ūn/-īn* ‘Egyptians-m.pl.NOM/-m.pl.non-NOM’); (c) dual (*miṣriy-ān/-ayn* ‘Egyptians-dua.NOM/-dua.non-NOM’); and (d) words with 2m.s., 2f.s., or 3m.s. enclitic pronoun (*-ak, -ik, -u* with variants), and for verbs (a) 2f.s. (*taḏhabiin* ‘you go’); (b) 2m.pl. (*taḏhabuun* ‘you go’); (c) 3m.pl. (*yaḏhabuun* ‘they go’); (d) jussive position (*lam yaḏab* ‘he did not go’); and (e) imperative (*iḏhab* ‘go.2m.s.’). [↑](#footnote-ref-3)
4. Excluding stems with final *tāʾ marbūta* ـة, *aʾ* ـأ, or *āʾ* ـاء, for which the orthographic CMI is canceled. [↑](#footnote-ref-4)