



No. CCLS-11-02

**Title: Conventional Orthography for Dialectal Arabic  
(CODA) Version 0.1 – July 2011**

**Authors:** Nizar Habash, Mona Diab and Owen Rambow

# Conventional Orthography for Dialectal Arabic (CODA)

Nizar Habash, Mona Diab and Owen Rambow

Version 0.1 - July 2011

## What is CODA?

Arabic dialects have no standard published orthographies since there are no Arabic dialect academies nor is there a large edited body of dialectal literature that follows the same spelling standard. CODA is a conventional orthography for Modern Standard Arabic and all of the Arabic dialects that aims at filling this gap designed primarily for the purpose of developing computational models of Arabic dialects. The following are some of the principles of design for CODA.

- CODA for **MSA** is standard correct Arabic spelling.
- CODA addresses three orthographic issues:
  - simple **typos** such as transposed letters (such as كبير vs. كبير) or missing spaces such as (انت رجل vs. انت رجل) or misspelled MSA words (such as فناه جميله vs. فناة جميلة in the appropriate context) ;
  - intentionally **creatively spelled words** and speech effects (such as كتيبيير);
  - inconsistent **choices in dialectal spelling** due to lack of standards.

CODA addresses all of these, and all of these need to be corrected to the CODA standard. Most of this manual is about the third type of errors.

- CODA abstracts away from **phonological** variations (between a dialect and MSA, among dialects, and between a dialect and its sub-dialects/accents) by relating the word form to its closest MSA cognate when possible (such as يكتب could be /yiktib/ in Egyptian corresponding to /yaktub/ in MSA). That said, relating etymological variants is not the main goal of CODA, especially when change is extensive and unrecognizable to a lay person, for example we don't want to exchange لمالن by مش.
- CODA preserves dialectal **morphology** (e.g., dialectal clitics حتقول instead of ستقول).
- CODA preserves dialectal **syntax**, i.e. there is no change of word order.
- CODA should be easy to learn and **write** to achieve high inter-annotator agreement; the more CODA looks like what a dialect speaker may write, the better.
- CODA is not a purely phonological representation; however, text in CODA can be **read** perfectly in dialectal Arabic given the specific dialect and a pronunciation guide.
- CODA is an **ad hoc convention**. There are numerous decisions that could have been made differently especially when it comes to the phonology/orthography interface. In particular, common dialectal words that may be exceptions are specified in an appendix. CODA is intended to be as faithful as possible to the dialects but also to be readable by all Arabic literate readers. But it is a balancing act, trying to preserve the dialectal variation while achieving interannotator consistency, hence we err on the side of caution and try to relate the cognates of MSA words to their MSA forms.
- CODA is intended as a **Pan-Arabic convention**. This is necessary to address the dialectal continuum problem where sub-dialects that are geographically under the influence of two highly distinctive dialectal centers may reflect aspects from both, e.g.,

Eastern Jordan and Western Iraq or even Iraqi as a dialect in-between Levantine and Gulf.

These principles make CODA comparable to English spelling (a bit phonological, a bit historical, lots of exceptions). *Perhaps this is not the most perfect spelling system for any Arabic dialect, it is however a necessary step toward creating common conventions and eventually standards. It also fills in a gap between not writing the dialects at all in a common convention and creating very different dialect-specific standards.*

## CODA Guidelines and the LDC Guidelines

These guidelines are inspired by the [LDC guidelines for transcribing Levantine](#) and Iraqi Arabic (Buckwalter and Maamouri, 2004). They differ from them in the following ways:

1. Whereas the LDC guidelines are for transcription, and thus focus more on phonological variants of sub-dialect in some case, CODA is intended for general purpose writing in a way that abstracts from these variations when possible.
2. CODA is intended/designed as a pan-Arabic convention. We extend the LDC guidelines to cover Egyptian Arabic (for which we profited from the LDC work on CallHome Egyptian) and refer to some specifics of Tunisian and Moroccan (to be extended in later versions).
3. CODA fills in some gaps not addressed in the LDC guidelines, specifically related to morphology.

In the rest of this manual the color red is used to mark an incorrect spelling that needs to be addressed. And green is used to highlight some corrections.

## I. General Spelling

A. Correct any typos and creative choices in spelling. These include missing, added, repeated, and transposed letters; and merged words and split words. Beware of letters that do not connect as they may appear unconnected.

- رئيس الوزراء -> رئيس الوزراء
- رئيسالوزراء -> رئيس الوزراء
- الفلسطينين -> الفلسطينيين
- الفلسط ينيين -> الفلسطينيين
- الفلسطيينين -> الفلسطينيين
- ما عنده كلمة منيحة يحكيها -> معندو كلمي منيحا يحكيها
- يا الله شو حكيو كتييييير -> يا الله شو حكيه كثير
- محمدعبد الغفارمحمد -> محمد عبد الغفار محمد

B. Do not change punctuation by adding/deleting or replacing any punctuation marks.

C. There are general rules of spelling, which we present below. These typically follow MSA style of writing and include directions for common dialectal morphemes. Common dialectal words are spelled exceptionally and are provided in a list at the end of this document.

## II. Letters and Sounds

**A. Diacritics (التشكيل):** ignore all diacritics, including dialectal nunation (see last three examples below). All dialectal variations in short/no vowels (that are never long under any context) are ignored. Do not add or delete existing diacritics.

- بيكتب may be read as /biyiktib/ or /byoktob/, leave as is
- بلاد /bilaad/ /blaad/
- مثلن -> مثلاً
- غصبين -> غصب
- على كلن -> على كل

**B. Alif Maqsura/Ya ( ي / ي ):** spell final ي/ي correctly: ى for /a/ and ي for /i/

- السيد على محمد ابو شنب -> السيد علي محمد ابو شنب
- الكلام ده على مين؟ -> الكلام ده علي مين

### C. Consonant-Consonant and Vowel-Vowel sound difference

- If the word has a direct cognate in MSA, use MSA spelling. A cognate is an etymologically related form, or أصل مشترك. A direct cognate will share all of phonological consonants, ignoring differences shown in Table 1 for example (كثيرا كثير). In a few cases, the dialectal cognate may have additional or deleted consonants such as (جاج/دجاج نص/نصف) that are readily recognizable as cognates from context. Note that the word may not appear in MSA as a word, but it is related via the same root. *It is likely a lot of forms will be in Lisan Al3arab anyway*, e.g. ظلمة/ظلام.
- If you are not sure if the word has a known cognate, check with the annotation appendix for exceptions that may reveal some less known cognate to your dialectal variant-- how to spell سنيل `annoying'? (see answer below)
- If the word is not in the exception list and has no MSA cognate, spell as pronounced.

MSA/CODA orthography	Dialectal pronunciation variants	Examples of dialect words written in CODA
ق	ك ئ ج غ	طريق /Tarii'/ /Tariidj/
ك	تش	كلب /t\$alb/
ث	ست	كثير /ktiir/
ذ	دزظض	كذب /kezeb/ /kidb/
ض	ظزد	ضابط /Zaabit/
ظ	ضزد	ظلمة /Dalma/
ص	س	صيف /sayf/ pronounced like Eng safe

ط	ت	يا لطيف /ya latiif/
س	ص	فستان /fuSTaan/
ت	ط	فستان /fuSTaan/
ع	ح	كعك/كحك
ل	ن	برتقال/برتقان/بردآن
ج/ز	ز/ج	زوج for husband which you may see as جوز
ج	Egyptian, Levantine, Gulf and MSA pronunciations	جميل /gamiil/ or /jamiil/ or /yamiil/

Table 1

Notice that these rules apply over all dialects. So, the fact that the Levantine and Egyptian words for “lie” are /kezeb/ and /kidb/, respectively, they are both written with the MSA cognate: كذب.

This approach will also cause some words in the same dialect to become **homographs**: e.g. ثقيل in Levantine can be /t'iil/ ‘heavy’ or /sa'iil/ ‘annoying’. These words will be listed in the appendix on special words. Please notify us of any forms you may encounter to add.

The examples in Table1 are an almost complete list of well understood cognate cases. It is not fully complete.

For a word to be written as its MSA cognate, we need the following conditions to apply:

- (1) The dialect word and its meaning IN CONTEXT do not exist together in MSA.
- (2) The dialect word and MSA cognate word MEAN the same thing in CONTEXT.
- (3) The two words have etymological phonological transformations appearing in Table 1  
e.g. شغل/work exists in MSA -- so we do not change it to عمل.  
e.g., طريضة/table does not exist in MSA but there is no word meaning table in Arabic that satisfies (3), so, we do not write it

If a word is encountered that satisfy conditions 1 and 2 but not 3 and is very close or you believe is related but the information is missing in Table 1, please ask your lead so that we can determine whether its transformation should be added or not.

We make a general preference to preserve long vowels in base words even if they shorten in different contexts.

- كاتب [kaatib]
- كاتبين [katbiin] كتيبين

- كاتبينها [katbinha] **كتبينها**
- تقولها [tquulha] `you say it'
- تقول لها [tqulha] `you tell her'

Long dialectal vowels that are short in their MSA cognate are ignored.

- الوليدة رمت الكرة [liwliida rmaat ilkuura] `the girl threw the ball' [TUN]

MSA words that use the Dagger Alif (الألف الخنجرية) are not included in this exception: هذا/الرحمن.  
No dialect word will use Dagger Alif unless the MSA cognate also uses it.

- Do not mark the distinction among the diphthong /ay/, the long high front /ii/ and the long mid front /ee/ vowels:
  - دير
    - الدير/الزور/الراهبات **دير** [deer] (read as IPA not English, it rhymes with *dare*)
    - دير (صب) المي **دير** [diir] (read as IPA; it rhymes with English *deer*)
  - في
    - تحت الشجرة في في [fii]..[fay] (Levantine: there is a shadow under the tree)
- Similarly, do not mark diphthong /aw/, the long high back /uu/ and the long mid back /oo/ vowels:
  - دور
    - دور حوالين البيت **دور** [duur] (read as IPA not English, it rhymes with *food*)
    - اوقف عال**دور** [door] (read as IPA; it rhymes with English *door*)
    - جاب لي دور [dawar]
    - دور عال**دورا** [dawwar]

## D. Hamza

1. Always write the Hamza above or below Alif and the Alif Madda; EXCEPT THAT, you should ignore the distinction between Hamzat-Wasl and Real Hamza (i.e. همزة الوصل (وهمزة القطع): this only applies to word-initial Alifs; ignoring them means do not add/delete Hamza above/below these (and only these) Alifs. An exception to this exception is the definite article (never put a Hamza on its Alif).
  - a. امام ، إخراج ، إحتفال [all correct in CODA]
  - b. سأل ، سأل [NOT word-initial, so we must pick the correct form in context]
  - c. رأس ، رأس [NOT word-initial, so we must pick the correct form in context]
  - d. مآسي [Alif Madda is always corrected]
2. If the Hamza consonant changes into a vowel or is deleted (together with its vowel possibly) but the word is still relatable to an MSA cognate, write the MSA cognate:
  - a. **لؤلؤ ، هوا ، سما ، بئر ، مايل ، رأس ، ولاد ، مرا/مراة**
  - b. **لؤلؤ ، هوا ، سما ، بئر ، مائل ، رأس ، أولاد ، إمراة**
3. Some verbs may show more radical change, but they should be spelled using the MSA cognate nonetheless:
  - a. **أرأيت/أريناها**
  - b. **قرأ / قرأت / قرأناها**

NOTE: Word-final Hamzas that may vary by case in MSA do not vary in CODA (see section on clitics below). For example, MSA may have بهاءها // بهاؤها // بهانها but the dialects only have بهاءها.

### E. Foreign words and place names (Modified from LDC)

- Most foreign words and place names already have established MSA spellings (e.g., Washington واشنطن, Los Angeles لوس انجلوس). In cases where the MSA spelling has regional variants, follow the Egyptian spelling since it confuses g/j/dj and does not depend on Perso-Arabic extensions as in Iraqi:
  - "garage" (Levantine: كراج -- contrast with Egyptian: جراج)
  - "congress" (Levantine: كونغرس -- contrast with Egyptian: كونجرس)
  - for /p/ use ب (e.g., "Pam" بام),
  - for /č/ use تش (e.g., "Chang" تشانج),
  - for /v/ use ف (e.g., "Vivian" فيفيان),
  - for /ž/ use ج (e.g., "Mirage" ميراج),
  - for /g/ use ج (e.g., "Gilbert" جيلبرت).

## III. Word: Stems, Affixes and Clitics

The basic word consists of stem and obligatory affixes (some of which are nil). To the basic words, clitics can be added (all optional). Example: فحتكاتبيها `so you will correspond with her'.

enclitics (optional)	suffixes (obligatory)	Base Stem (obligatory)	prefixes (obligatory)	proclitics (optional)
ها	يـ	كاتب	تـ	فحـ

### A. Stem and Affixes

*Nouns and Adjectives* have specific rules for affixes:

- Ta Marbuta is always ة and never ه at end of word. Inside a word (after clitics) it may be ت or ا. The spelling of Ta Marbuta is not affected by pronunciation:
  - سيارة سميرة جديدة in LEV [sayyaarit samiira jdiide]
  - سيارتي // سيارتنا
  - معلمته `his teacher'
  - معلمها `she taught him' شافية 'she is seeing' the gerund deverbal form using the active participle
  - غزالي - غزالة if the word is meant to be `a female deer'. It should be noted that غزالي is a correct spelling of `my deer'
  - Note that some words introduce a Ta-marbuta variant only in construct state (معنى/معناة؛ جوا/جواة؛ برا/براة؛ مدام/مدامة, e.g., إضافة): Egyptian Arabic only has مدامته not مدامه for `his wife'

- Dialectal noun suffixes include +ين (both dual and plural) and +ات also. Note that +ين does not vary depending on case and construct state as in MSA (ين // ون // ان // ا // و // ي).

Verbs are spelled as in MSA with the following exceptions:

- Added vowels for geminates (مضاعف):
  - مد / مديت / مديتوا
- Plural /uu/ affixes are spelled with +وا in word final spelling and +و in word medial spelling
  - كتبوا / كتبوا / كتبوها / كتبوها
  - [TUN] نكتبوا / نكتبوها
- Feminine /ii/ affixes are spelled with extra +ي
  - كتبتني / تكتبي

## B. Clitics

There are **four** general rules for spelling clitics:

1. **All single letter particles are cliticized** (attached) while multi-letter clitics are not attached. The only exception is the definite article +ال
  - a. البيت / بالبيت / بهالبيت / عهالبيت
2. Pronominal enclitics (الضمائر المتصلة) and the negation particle (+ش) attach word finally but nothing else. CODA does not attach the prepositional phrase (ل+ضمير) to the word preceding it.
  - a. بيتنا / بيتكم
  - b. ما كتبناش
  - c. ضرب له سلام // حكى لها قصة
3. We follow MSA morphophonemic writing as in writing +ال almost always as +ال regardless if the stem starts with a sun/moon letter and regardless of any reduced pronunciations:
  - a. القمر /il'amar/
  - b. الشمس /i\$amis/
  - c. الجبال /lijbaal/
  - d. الاولاد /liwlaad/
4. The general rule of clitics is that when we attach the clitics to the word, we do not change the base word spelling even if there is a phonology change in either stem or clitic, e.g., +ال+بيت is بالبيت. There are some exceptions:
  - a. +ال+ل become لل (but +ال+ب remains بال)
    - i. +ال+البيت // للبيت
  - b. +ة becomes ت or ا
    - i. معلمة+هم // معلمتهم // معلماهم
  - c. الف واو الجماعة is deleted
    - i. كتبوا+ها // كتبوها
  - d. Alif Maqsura is turned into Alif or Ya depending on the word
    - i. حكى+هم // حكاهم



ii. على+هم // عليهم

A common error is to apply the shadda spelling rule across clitic-stem boundary.

- مسك+كم // مسككم NOT مسككم
- شبه+ها // شبهها NOT شبهها
- سمن+ني // سمنني NOT سمنني

## C. Dialectal Clitics not in MSA

The following is a list of dialectal (non-MSA) clitics and some comments:

- The demonstrative h+ appears before Al; it can be in-between it and other clitics
  - هالبيت / لهالبيت / وبهالبيت
- The progressive b+/d+ attaches to imperfect verb. Make sure to always spell the base word correctly (the example with pluses is only to explain the prefixes; do not write the pluses in CODA)
  - ب+تكتب // بتكتب
  - ب+يكتب // بيكتب NOT بيكتب even if pronounced as such
  - ب+اكتب // بأكتب NOT بأكتب even if pronounced as such; Note Hamza!
  - ب+نكتب // بنكتب NOT منكتب even if pronounced as such
  - د+اكتب // داكتب [IRQ]
  - ك+نكتب // كنكتب [MOR]
- The future marker H+/g+ attaches to imperfect verb. Make sure to always spell the base word correctly:
  - ح+امشي // حامشي NOT حمشي
  - غ+يكتب // غيكتب
  - NOTE: Egyptian Arabic has two common variants: +ح and +ه. ONLY USE +ح.
- Plural /uu/ clitics are spelled with وا+
  - شافكوا
- Feminine /ii/ clitics are spelled with extra ي+
  - شافوكي
  - This particular pronoun (+كي) has another form (+ك):
    - شافك // شافوكي /\$aafik//\$afuuki/
    - ما شافوكيش // ما شافوكيش
- The 3rd person pronouns (ه ها هم هن) are always written with an ه even if not pronounced or assimilated
  - كتابه /ktaabo/ /ktaabuh/
  - كتابها /ktaabha/ /ktaaba/
  - كتبوها /katabuuha/ /katabuwa/

- كتبه /katabuu/
- كتبتياها /katabtiha/ /katabtiya/
- Other single letter prepositions and particles:
  - Shared with MSA: ل / ب / ك / ف / و / etc.
    - + و is ALWAYS attached: البيت NOT البيت
  - +ع 'on/at'
    - ع+البيت // عاليبيت
  - +ت 'so'
    - ت+يكون // تيكون
  - +ش 'what'
    - ش+د+تقول // شدتقول
  - +ش 'not'
    - ما كتبتش /katabi\$/ `he did not write'
    - ما كتبووش /katabuu\$/ `**they** did not write'

Note: the 3rd person masculine singular pronominal clitic (+) changes form before هوش to +ش:

  - كتبه LEV:/katabo/ EGY:/katabu<sup>h</sup>/ `he did not write **it**'
  - كتبه LEV:/katabuu/ EGY:/katabuu<sup>h</sup>/ `they did not write **it**'
  - ما كتبهوش LEV:/kataboo\$/ EGY:/katabhuu\$/ `he did not write **it**'
  - ما كتبوهوش LEV:/katabuhoo\$/ EGY:/katabuhuu\$/ `they did not write **it**'

Note: The Egyptian variant +شي of +ش should be written as +ش.
- Particles that SHOULD NOT BE cliticized
  - mA of Negation. Always add a space between it and word even though the -\$ is cliticized.
    - ما كتبتش NOT ما كتب+ش // ما كتبتش
    - ما ظلش NOT مظلش
  - Indirect object pronouns (+ل pronoun) is also separated
    - قول له NOT قلله
    - قلت لك NOT قلناك
  - Any particles longer than one letter. The only exception is Al.
    - عم بيكتب NOT عميكتب ، عميكتب
    - Note: عم always takes +ب with verb
    - رح يمشي NOT رحيمشي

#####

#### #### Common Dialectal Words ####

- Please refer to CODA Tables on google docs for more organized sets of words  
<https://spreadsheets.google.com/ccc?key=0AixsNT2QVxgydFFudHpISUtzNUJUV2FtbXMwdV9IeVE&hl=en#gid=0>

- These tables will be continuously updated.
- There may be lots of variants in sub-dialects and pronunciations or other common spellings.

The general guideline are as follows:

- Ignore distinctions that are the result of short vowels and common consonantal replacement that lead to MSA-like forms.
- If no MSA-like form exist, choose form from city sub-dialect.
- Consonantal deletion does not count as a replacement. In some case, a large number of sub-dialectal variants are grouped together exceptionally (see demonstratives).

CODA	Possible Variants (Not Complete)	English	Example	Dialects
أتاري	اثاري	it turned out		LEV
برضو	برضه / بردو / برده	also		LEV, EGY
عشان	علشان, مشان	in order to		LEV, EGY
إمبارح	إمبيرح	yesterday		LEV, EGY
بركبى	بلكى	may be		LEV
برا	بره	outside / outside of	كان بيستاني برا البيت براه للبيت	LEV, EGY
براة [اضافة only]	برات / بریت	outside of	كان بيستاني براه البيت براته للبيت	LEV
جوا	جوه	inside / inside of	كان بيستنى جوا البيت جواه للبيت	LEV, EGY
جواة [اضافة only]	جوات	inside of	كان بيستنى جواة البيت جواته للبيت	LEV
بس		only, enough, just		LEV, EGY
مش	موش	not		LEV, EGY
كمان	كمانا, كمانه	also		LEV, EGY
نيال		how lucky..		LEV

انو	اني، انهو، انهبي	which	من أنو جامعة	LEV
انهبي	انهو	which	من أنهبي جامعة	EGY
هسع		now		LEV
هلق	هلاً	now		LEV
دلوقتي		now		EGY
راح	ح		verbal particle	LEV

### Pronouns

أنا	اني	I		LEV, EGY
إحنا	نحننا	we		LEV, EGY
إنت	اتنا	you m.s.		LEV, EGY
إنتبي	انت	you f.s.		LEV, EGY
إنتوا	اتتو	you p.		LEV, EGY
هو	هوه / هوا	he / it		LEV, EGY
هبي	هيه / هي	she / it		LEV, EGY
هم	همه / هما	they		LEV, EGY

ي+ ني		me/my		LEV, EGY
نا+	ني+	us/our		LEV, EGY
ك+	كا+	your m.s.		LEV, EGY
ك+ كي		your f.s.		LEV, EGY
كم+		your p.		LEV, EGY

كن+		your p.	ambiguous!	LEV, EGY
كوا+		your p.		LEV, EGY
ه+	هو+ و	his, him		LEV, EGY
ها+	هي+ ا	her		LEV, EGY
هم+		them, their		LEV, EGY
هن+	هن+ ون	them, their	ambiguous!	LEV, EGY

ده	دا	this/that m.s.		EGY
دي		this/that f.s		EGY
دول	دولك، دولهم	these/those		EGY
هذا	هذا هاد هاذ هاذ، هيدا، هيدا	this m.s.		LEV, TUN
هذي	هادي، هذي، هيدي	this f.s.		LEV, TUN
هذول	هذول، هذوله، هذولا	these		LEV
هذاك	هذاك، هذاك	that m.s.		LEV
هذيك	هظيك، هذيك، هذيك	that f.s		LEV
هذوك	هذولاك، هذوك، هظوك	those		LEV
زي		like		
هي هي+ ضمير هيان+ ضمير	هاي	this there	/hayy/	
أبوا	أبوه، أبون، أبوا	yes, indeed		LEV, EGY
إمبلى	إمبلا، مبلى			

لا	/la' / and /la''a/	no		LEV, EGY
لا	la	not/neither-nor	No Hamza	LEV, EGY
له	lah	no way!		LEV
كده	كدهه ، كدهو	like that		EGY
هيك	هيك، هيكه، هيكي	like that		LEV
هون	هان, هانا	here		LEV
هنا		here		EGY
هناك	هنيك	there		LEV,EGY

اللي	الي، ياللي	who, which, whom	The Alif-Lam is treated as a definite article in cliticization: باللي، عالي، للي، ... هاللي، لهاللي	LEV, EGY
يا اللي	يللي، ياللي		/yalli/ can be one of two things, a variant of illi or ya+ill. They are spelled differently	LEV, EGY
تبع	تبعه/تبعون		collapse the variant forms since they are optional	LEV
بتاع، بتوع				EGY
ديال				MOR
حق				GLF
مال				IRQ

فين		where		EGY, LEV
-----	--	-------	--	----------

وين		where		LEV
مين		where	also allow “ من وين ”	LEV
مين	من	who		LEV, EGY
منو	منى [this is only fem]	who		IRQ
ليش		why		LEV
ايش	ويش	what		LEV
قديش	اديش	how much/many	Don't confuse with “ قد ايش ”	LEV
قديه		how much/many	Don't confuse with “ قد ايه ” in EGY	LEV
ايه		why		EGY
ليه		why		EGY
شو		what, what!		LEV
لشو		why		LEV
ايمتي	ايمتي، امن	when		LEV, EGY

فيه	فيه	there is, in		
فيه	فيو	in it		
ما فيش		there is not		
ما فيهوش		not in it		
معليش		never mind		
معليه		never mind		
حدا حد			do not convert to واحد	

حدش				
-----	--	--	--	--

## Prepositions

مع			معاي / معيك / معكون / معو / معايا / معانا	with		LEV, EGY
معى		معنا				
معك	معك	معكم				
معه	معه	معهم				
+بد				want (verboid)	ب+ود : From MSA : +ضمير	LEV
بدي		بدنا				
بدك	بدك	بدكم				
بده	بدها	بدهم				
+إل					write to make less ambiguous with ل +ضمير.	
إلى		إلنا				
إلك	إلك	إلكم				
إله	إله	إلهم				

بالله	يلا، يالله		yalla (hurry up) do not confuse with “يا الله” (oh, my god/godness)	
إن شاء الله	...انشاله انشالله			

مدام		wife	هي مدام بن علي	LEV, EGY
مدامة [إضافة only]	مدامت	wife	هي مدامتو لبن علي هي مدامة بن علي	LEV, EGY
زلمة		man, guy		LEV



فم	ثم, تم	mouth	unrecognizable cognate	LEV
بق	بؤ	mouth		EGY
دغري		straightforward		LEV, EGY
زاكې		delicious	زاكې يعنى لذيذ	
كوبس		good		EGY
منيح		good	see lexical variant مليح	LEV
مليح		good	see lexical variant منيح	LEV
أكمَن	اِكم	a bunch	هالاكمَن contraction	LEV
ألو		hello? (on phone)		LEV, EGY
أوضة		room		
أوكې		OK		
يد	ايد			
يديڭ+ضمير يديڭ	ايدىڭ+ضمير ايدىڭ		These cases are similar also for eyes, feet, etc. (body parts)	
لسه	لسا	still		
لسات+ضمير				
شوي		a bit	although it has an MSA explanation...	
شوية		a bit		
طيب	طب	ok, so		

Common confusing cases

ثقیل	تقیل/تقیل /t'iil/	heavy		LEV
	سئیل /sa'iil/	annoying		LEV
رجل	/rijjaal/رجال	man		LEV
رجال	/rjaal/رجال	men		LEV

Homophone Variants:

أنا	/ani/ (أني)			LEV
أنو	/ani/ (أني)			LEV

#### MSA RULE APPLIES

The following are cases where using an MSA form applies although the words may seem different:

شيء	شيء، إشي	something		LEV
ماء	مي، مية ماي، مائة	water		EGY,LEV
تعالوا	تعال، تعالي، تعالوا	come here		

#### Numbers

Although we follow MSA spelling on numbers we do not apply MSA syntax rules. For example, MSA expects that the gender of the number is the reverse of the noun in some cases:

ثلاث نساء وثلاثة رجال

But in the dialects, this is not the case: ثلاث ستات وثلاث رجال

The number ثلاثة is used in the dialect with unspecified objects or simple counting (one, two, three, go!).

Some dialect also drop the final consonant on some numbers: خمستعش for خمسة عشر. We use the MSA for that form. Also, here ignore gender flipping: خمسة عشر ولد/بنت.

صفر		0		ALL
نصف	نص	1/2		ALL
واحد		1		
إثنين	تئين	2		

ثلاثة ثلاث	ثلت، ثلاثة	3	ثلاث اولاد راحوا عالسوق	
أربعة أربع		4		
خمسة خمس		5		
ستة ست		6		
سبعة سبع		7		
ثمانية ثمان	تمنية، ثمانية	8		
تسعة تسع		9		
عشرة عشر		10		
أحد عشر	حد عشر ابدعش ...، احدعش	11		ALL
إثنا عشر	تناش تتعشر طنعش، اثنعش	12	USE nominative always (never اثنى عشر )	ALL
ثلاثة عشر	ثلاثطعشر، ثلاثتعشر ثلاثتبعش	13		ALL
أربعة عشر	أربعطعشر، ربعتعشر أربعتبعش	14		ALL
خمسة عشر	، خمسطعشر خمستعشر خمسبعش	15		ALL
ستة عشر	سسطعشر، ستتعشر، ستبعش	16		ALL
سبعة عشر	سبعتعشر، سبعطعشر، سبعبعش	17		ALL
ثمانية عشر	، ثمانطعشر	18		ALL

	ثمانتعشر تمنطعش ثماتعش			
تسعة عشر	تسعطعشر، تسعتعش، تسعتعشر	19		ALL
مئة	ميت، مية	100		ALL
مئتين	ميتين،			
ثلاثمئة	تلميت تلمية			
اربعمئة				
خمسمة				
ستمئة				
سبعمة				
ثمانمة				
تسعمة				
ألف				
آلاف	تالاف		ثمان الاف /taman talaaf/	

Also Follow MSA Rules in spelling names of the days of the week

Sunday يوم الأحد

Monday يوم الإثنين

Tuesday يوم الثلاثاء

Wednesday يوم الأربعاء

Thursday يوم الخميس

Friday يوم الجمعة

Saturday يوم السبت