# ArabLEX

# برج المعاجم

# Tower of Lexicons

# *ArabLEX*

# Arabic Full Form Lexicon

# معجم اللغة العربية الكامل

## 1. Introduction

### 1.1 Synopsis

**The CJK Dictionary Institute** (CJKI) is pleased to announce the release of the **Comprehensive Arabic Full Form Lexicon** (*ArabLEX*), covering approximately **538 million entries** inflected, conjugated and cliticized wordforms. This includes both general vocabulary as well as fully inflected Arab and non-Arab proper nouns (personal and place names).

*ArabLEX* is not only comprehensive in coverage, but is also rich in morphological, grammatical, phonological, and orthographical attributes. In addition, it maps all unvocalized forms to their vocalized counterparts and to the lemma, and provides precise phonemic (and eventually phonetic) transcriptions. These features can significantly contribute to the training of language models for NLP, deep learning and AI for such applications as speech technology (TTS and ASR) and machine translation. It aims to serve as the ultimate resource for Arabic natural language processing.

This document includes specification and data samples for all four *ArabLEX*:

| | | |
|---|---|---|
| **DAG** | Database of Arabic General Vocabulary | 88 million entries |
| **DAN** | Database of Arabic Names | 218 million entries |
| **DAF** | Database of Arabic Foreign Names | 226 million entries |
| **DAP** | Database of Arabic Place Names | 6 million entries |

### 1.2 Approximate Quantities

Currently the DAG module (general vocabulary) of *ArabLEX* is estimated at 88 million entries. If we include proper nouns (DAF, DAP and DAN), the total is approx 538 million. In future expansion, we plan to include negated forms, periphrastic elatives, compound verb tenses and other categories, and thus expect the total entries to exceed one billion.

### 1.3 Database Structure and Schema

Each of the four modules (DAG, DAF, DAP and DAN) are stored in four data tables:

(1) **Canonical**, (2) **Morphological**, (3) **Orthographical**, and (4) **Phonological**. The database is stored in 15 data tables interlinked by IDs and SubIDs within each module. Each module consists of the following functionally distinct data tables:

**Data Tables for Each Module**

| | | DAG | DAP | DAF | DAN |
|---|---|---|---|---|---|
| 1 | Canonical | | DAP-C | DAF-C | DAN-C |
| 2 | Morphological | DAG-M | DAP-M | DAF-M | DAN-M |
| 3 | Orthographical | DAG-O | DAP-O | DAF-O | DAN-O |
| 4 | Phonological | DAG-P | DAP-P | DAF-P | DAN-P |

(1) **Canonical**: This is the primary table that provides basic data such as the lemmata in several formats, romanization, ID links to other modules, and other fields as necessary such as gender codes and frequency.

(2) **Morphological**: Comprehensive morphological attributes (max over 20) and wordforms (often thousands) for each lemma, including inflections (possessives, plurals), declensions (case endings), conjugations (tense and person), and cliticized forms. It is the core, often the largest, component of *ArabLEX*.

(3) **Orthographical**: Various attributes such as orthographical and spelling variants, including vocalized and unvocalized variants for each lemma and their wordforms.

(4) **Phonological**: Various phonological attributes, such as phonemic and phonetic transcriptions and syllabification for all lemmata and their wordforms (inflections, conjugations, etc.) ideal for speech technology.

NOTE: *ArabLEX* is constantly evolving and expanding. Some modules don't necessarily have all the fields shown in the field descriptions.

## 1.4 More Information

**Detailed whitepaper**
https://www.cjk.org/wp-content/uploads/ArabLEX.pdf

**White paper summary**
https://www.cjk.org/wp-content/uploads/ArabLEX_summary_.pdf

**Academic paper**
https://www.cjk.org/wp-content/uploads/Halpern-LREC2022Paper.pdf

**Detailed sample**
https://www.cjk.org/wp-content/uploads/2020/11/ArabLEX.xls

**The cover:** *The Burj Khalifa,* the tallest building in the world, soars 832 meters above the skies of Dubai. This is symbolic of *ArabLEX*, which is the largest Arabic lexicon ever created.

## 2. DATABASE OF ARABIC GENERAL VOCABULARY (DAG)

### 2.1 Synopsis

DAG is a **comprehensive full form lexicon of Arabic general vocabulary** that contains a rich set of morphological, grammatical, and phonological attributes. Proper nouns are in principle excluded as they are given in the modules DAN, DAF and DAP (described below).

### 2.2 More Information

**Website:** http://www.cjk.org/data/arabic/nlp/arabic-full-form-lexicon/
**White paper:** ArabLEX.pdf
**Data sample:** ArabLEX.xls

### 2.3 Approximate Quantities

| Type | Lemmas | Inflected forms |
|---|---|---|
| **Nouns** | 14,293 | 30,679,202 |
| **Adjectives** | 6,228 | 23,874,505 |
| **Verbs** | 9,509 | 33,377,031 |
| **TOTAL** | 30,030 | 87,930,738 |

### 2.4 The Morphological Component (DAG-M)

#### 2.4.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **POS** | **part of speech code**<br>[V] verb<br>[N] noun<br>[A] adjective |
| 4 | **SP** | **POS subclassification**<br>**verbs**<br>[T] transitive<br>[I] intransitive<br>[D] ditransitive<br>[B] transitive & intransitive<br>**adjectives**<br>[P] positive<br>[E] elative |

| | | |
|---|---|---|
| | | [e] elative with non-elative meaning |
| 5 | **ARAB_V** | vocalized cliticized Arabic entry |
| 6 | **ARAB_BW** | vocalized cliticized Arabic entry in Buckwalter transliteration |
| 7 | **ARAB_U** | unvocalized cliticized Arabic entry |
| 8 | **LEMMA_V** | lemma in vocalized Arabic |
| 9 | **LEMMA_U** | lemma in unvocalized Arabic |
| 10 | **GEN** | **gender code for stem**<br>[M] masculine<br>[F] feminine |
| 11 | **NUM** | **number code for stem**<br>[S] singular<br>[D] dual<br>[P] plural<br>[B] broken plural |
| 12 | **CASE** | **case ending code for nouns and adjectives**<br>[ACU] accusative<br>[GEN] genitive<br>[NOM] nominative |
| 13 | **PER** | (person code) **person, number, gender**<br>[000] zero person<br>[1SC] first person singular<br>[2SM] second person singular masculine<br>[2SF] second person singular feminine<br>[3SM] third person singular masculine<br>[3SF] third person singular feminine<br>[1PC] first person plural<br>[2PM] second person plural masculine<br>[2PF] second person plural feminine<br>[2DC] second person dual<br>[3PM] third person plural masculine<br>[3PF] third person plural feminine<br>[3DM] third person dual masculine<br>[3DF] third person dual feminine<br>[XSM] singular masculine<br>[XDM] dual masculine<br>[XPM] plural masculine<br>[XSF] singular feminine<br>[XDF] dual feminine<br>[XPF] plural feminine |
| 14 | **DEF** | **definiteness**<br>[D] explicitly definite by cliticizing of article Al<br>[d] definite by adding pronomial enclitics<br>[I] indefinite<br>[G] first term of genitive construct (Idhafa, e.g. baytu in baytu-l-kaatibi) |
| 15 | **RAT**<br>(under construction) | **rationality** (animate or not)<br>[R] rational<br>[I] irrational |

| | | |
|---|---|---|
| 16 | **FORM** | **verb form** (see cavehelp.htm §12)<br>[01] Form I<br>[02] Form II<br>[03] Form III<br>[04] Form IV<br>[05] Form V<br>[06] Form VI<br>[07] Form VII<br>[08] Form VIII<br>[09] Form IX<br>[10] Form X<br>[Q1] Form QI<br>[Q2] Form QII<br>[Q3] Form QIII<br>[Q4] Form QIV |
| 17 | **TYPE** | **code for verb conjugation type** (see cavehelp.htm §13)<br>[S] sound<br>[Z] hamzated<br>[G] geminate<br>[A] assimilated<br>[H] hollow<br>[D] defective<br>[W] doubly week |
| 18 | **TENSE** | **code for verb tense**<br>[01] active perfect indicative<br>[02] active imperfect indicative<br>[03] active imperfect subjunctive<br>[04] active imperfect jussive<br>[05] active imperfect imperative<br>[06] active imperfect energetic I<br>OMITTED because obsolete<br>[07] active imperfect energetic II<br>OMITTED because obsolete<br>[08] passive perfect indicative<br>[09] passive imperfect indicative<br>[10] passive imperfect subjunctive<br>[11] passive imperfect jussive<br>[12] active participle<br>[13] passive participle<br>[14] verbal noun<br>[52] future indicative<br>[53] future passive |
| 19 | **PROC_V** | prefix or proclitic in vocalized Arabic |
| 20 | **PROC_U** | prefix or proclitic in unvocalized Arabic |
| 21 | **STEM_V** | stem in vocalized Arabic |
| 22 | **STEM_U** | stem in unvocalized Arabic |
| 23 | **ENC_V** | suffix or enclitic in vocalized Arabic |
| 24 | **ENC_U** | suffix or enclitic in unvocalized Arabic |

| | | |
|---|---|---|
| 25 | **ROOT** | root of the entry |

### 2.4.2 Data Sample

The fields in the sample below are a subset of the fields in the morphological component. There are many other fields useful for natural language processing. IMPORTANT: Only a small subset of available fields is shown here. Be sure to check out the full sample at ArabLEX.XLS for all fields.

**Morphological Component (nouns)**

| SUBID | POS | ARAB_V | ARAB_BW | LEMMA_V | GEN | NUM | CASE | PER | DEF | ROOT |
|---|---|---|---|---|---|---|---|---|---|---|
| 0001 | N | كَاتِبٌ | kaAtibN | كَاتِبٌ | M | S | NOM | 000 | I | كتب |
| 0002 | N | كَاتِبُ | kaAtibu | كَاتِبٌ | M | S | NOM | 000 | G | كتب |
| 0003 | N | كَاتِبِي | kaAtibiy | كَاتِبٌ | M | S | NOM | 1SC | d | كتب |
| 0004 | N | كَاتِبُكَ | kaAtibuka | كَاتِبٌ | M | S | NOM | 2SM | d | كتب |
| 0005 | N | كَاتِبُكِ | kaAtibuki | كَاتِبٌ | M | S | NOM | 2SF | d | كتب |
| 0006 | N | كَاتِبُهُ | kaAtibuhu | كَاتِبٌ | M | S | NOM | 3SM | d | كتب |
| 0007 | N | كَاتِبُهَا | kaAtibuhaA | كَاتِبٌ | M | S | NOM | 3SF | d | كتب |
| 0008 | N | كَاتِبُنَا | kaAtibunaA | كَاتِبٌ | M | S | NOM | 1PC | d | كتب |
| 0009 | N | كَاتِبُكُمْ | kaAtibukumo | كَاتِبٌ | M | S | NOM | 2PM | d | كتب |
| 0010 | N | كَاتِبُكُنَّ | kaAtibukun~a | كَاتِبٌ | M | S | NOM | 2PF | d | كتب |
| 0011 | N | كَاتِبُكُمَا | kaAtibukumaA | كَاتِبٌ | M | S | NOM | 2DC | d | كتب |
| 0012 | N | كَاتِبُهُمْ | kaAtibuhumo | كَاتِبٌ | M | S | NOM | 3PM | d | كتب |
| 0013 | N | كَاتِبُهُنَّ | kaAtibuhun~a | كَاتِبٌ | M | S | NOM | 3PF | d | كتب |
| 0014 | N | كَاتِبُهُمَا | kaAtibuhumaA | كَاتِبٌ | M | S | NOM | 3DM | d | كتب |
| 0015 | N | كَاتِبُهُمَا | kaAtibuhumaA | كَاتِبٌ | M | S | NOM | 3DF | d | كتب |
| 0016 | N | كَاتِبٍ | kaAtibK | كَاتِبٌ | M | S | GEN | 000 | I | كتب |
| 0017 | N | كَاتِبِ | kaAtibi | كَاتِبٌ | M | S | GEN | 000 | G | كتب |
| 0018 | N | كَاتِبِي | kaAtibiy | كَاتِبٌ | M | S | GEN | 1SC | d | كتب |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0019 | N | كَاتِبِكَ | kaAtibika | كَاتِبٌ | M | S | GEN | 2SM | d | كتب |
| 0020 | N | كَاتِبِكِ | kaAtibiki | كَاتِبٌ | M | S | GEN | 2SF | d | كتب |
| 0021 | N | كَاتِبِهِ | kaAtibihi | كَاتِبٌ | M | S | GEN | 3SM | d | كتب |
| 0022 | N | كَاتِبِهَا | kaAtibihaA | كَاتِبٌ | M | S | GEN | 3SF | d | كتب |
| 0023 | N | كَاتِبِنَا | kaAtibinaA | كَاتِبٌ | M | S | GEN | 1PC | d | كتب |
| 0024 | N | كَاتِبِكُمْ | kaAtibikumo | كَاتِبٌ | M | S | GEN | 2PM | d | كتب |
| 0025 | N | كَاتِبِكُنَّ | kaAtibikun~a | كَاتِبٌ | M | S | GEN | 2PF | d | كتب |
| 0026 | N | كَاتِبِكُمَا | kaAtibikumaA | كَاتِبٌ | M | S | GEN | 2DC | d | كتب |
| 0027 | N | كَاتِبِهِمْ | kaAtibihimo | كَاتِبٌ | M | S | GEN | 3PM | d | كتب |
| 0028 | N | كَاتِبِهِنَّ | kaAtibihin~a | كَاتِبٌ | M | S | GEN | 3PF | d | كتب |
| 0029 | N | كَاتِبِهِمَا | kaAtibihimaA | كَاتِبٌ | M | S | GEN | 3DM | d | كتب |
| 0030 | N | كَاتِبِهِمَا | kaAtibihimaA | كَاتِبٌ | M | S | GEN | 3DF | d | كتب |

## Morphological Component (verbs)

| SUBID | POS | SP | ARAB_V | ARAB_BW | LEMMA_V | PER | FORM | TYPE | TENSE | ROOT |
|---|---|---|---|---|---|---|---|---|---|---|
| 0229 | V | T | كَتَبَ | kataba | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0230 | V | T | كَتَبَنِي | katabaniy | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0231 | V | T | كَتَبَكَ | katabaka | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0232 | V | T | كَتَبَكِ | katabaki | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0233 | V | T | كَتَبَهُ | katabahu | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0234 | V | T | كَتَبَهَا | katabahaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0235 | V | T | كَتَبَنَا | katabanaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0236 | V | T | كَتَبَكُمْ | katabakumo | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0237 | V | T | كَتَبَكُنَّ | katabakun~a | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0238 | V | T | كَتَبَكُمَا | katabakumaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0239 | V | T | كَتَبَهُمْ | katabahumo | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0240 | V | T | كَتَبَهُنَّ | katabahun~a | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0241 | V | T | كَتَبَهُمَا | katabahumaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0255 | V | T | وَكَتَبَ | wakataba | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0256 | V | T | وَكَتَبَنِي | wakatabaniy | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0257 | V | T | وَكَتَبَكَ | wakatabaka | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0258 | V | T | وَكَتَبَكِ | wakatabaki | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0259 | V | T | وَكَتَبَهُ | wakatabahu | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0260 | V | T | وَكَتَبَهَا | wakatabahaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0261 | V | T | وَكَتَبَنَا | wakatabanaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0262 | V | T | وَكَتَبَكُمْ | wakatabakumo | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0263 | V | T | وَكَتَبَكُنَّ | wakatabakun~a | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0264 | V | T | وَكَتَبَكُمَا | wakatabakumaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0265 | V | T | وَكَتَبَهُمْ | wakatabahumo | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0266 | V | T | وَكَتَبَهُنَّ | wakatabahun~a | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0267 | V | T | وَكَتَبَهُمَا | wakatabahumaA | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0268 | V | T | فَكَتَبَ | fakataba | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0269 | V | T | فَكَتَبَنِي | fakatabaniy | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0270 | V | T | فَكَتَبَكَ | fakatabaka | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |
| 0271 | V | T | فَكَتَبَكِ | fakatabaki | كَتَبَ | 3SM | 01 | S1 | 01 | كتب |

## 2.5 The Orthographical Component (DAG-O)

### 2.5.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |

| | | |
|---|---|---|
| 5 | **VAR_V** | variant in vocalized Arabic |
| 6 | **VAR_U** | variant in unvocalized Arabic |

### 2.5.2 Data Sample

Under construction.

## 2.6 The Phonological Component (DAG-P)

### 2.6.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **CARS** | cliticized Arabic entry in phonemic transcription (CARS) |
| 6 | **IPA** | phonetic transcription in the International Phonetic Alphabet (IPA) |
| 7 | **XSAMPA** | phonetic transcription in X-SAMPA |

### 2.6.2 Data Sample

**Phonological Component (nouns)**

| SUBID | ARAB_V | CARS | IPA | XSAMPA |
|---|---|---|---|---|
| 0001 | كَاتِبٌ | kátibun | ˈkaː.ti.bun | "kaː.ti.bun |
| 0002 | كَاتِبُ | kátibu | ˈkaː.ti.bu | "kaː.ti.bu |
| 0003 | كَاتِبِي | kátibi̱ | ˈkaː.ti.bi | "kaː.ti.bi |
| 0004 | كَاتِبُكَ | kātíbuka | kaː.ˈti.bu.ka | kaː."ti.bu.ka |
| 0005 | كَاتِبُكِ | kātíbuki | kaː.ˈti.bu.ki | kaː."ti.bu.ki |
| 0006 | كَاتِبُهُ | kātíbuhu | kaː.ˈti.bu.hu | kaː."ti.bu.hu |
| 0007 | كَاتِبُهَا | kātíbuha̱ | kaː.ˈti.bu.ha | kaː."ti.bu.ha |
| 0008 | كَاتِبُنَا | kātíbuna̱ | kaː.ˈti.bu.na | kaː."ti.bu.na |

| | | | | |
|---|---|---|---|---|
| 0009 | كَاتِبُكُمْ | kātíbukum | kaːˈti.bu.kum | kaː."ti.bu.kum |
| 0010 | كَاتِبُكُنَّ | kātibukúnna | kaː.ti.buˈku.nːa | kaː.ti.bu."ku.nː.a |
| 0011 | كَاتِبُكُمَا | kātibúkumạ | kaː.tiˈbu.ku.ma | kaː.ti."bu.ku.ma |
| 0012 | كَاتِبُهُمْ | kātíbuhum | kaːˈti.bu.hum | kaː."ti.bu.hum |
| 0013 | كَاتِبُهُنَّ | kātibuhúnna | kaː.ti.buˈhu.nːa | kaː.ti.bu."hu.nː.a |
| 0014 | كَاتِبُهُمَا | kātibúhumạ | kaː.tiˈbu.hu.ma | kaː.ti."bu.hu.ma |
| 0015 | كَاتِبُهُمَا | kātibúhumạ | kaː.tiˈbu.hu.ma | kaː.ti."bu.hu.ma |
| 0016 | كَاتِبٍ | kátibin | ˈkaː.ti.bin | "kaː.ti.bin |
| 0017 | كَاتِبِ | kátibi | ˈkaː.ti.bi | "kaː.ti.bi |
| 0018 | كَاتِبِي | kátibi̱ | ˈkaː.ti.bi | "kaː.ti.bi |
| 0019 | كَاتِبِكَ | kātíbika | kaːˈti.bi.ka | kaː."ti.bi.ka |
| 0020 | كَاتِبِكِ | kātíbiki | kaːˈti.bi.ki | kaː."ti.bi.ki |
| 0021 | كَاتِبِهِ | kātíbihi | kaːˈti.bi.hi | kaː."ti.bi.hi |
| 0022 | كَاتِبِهَا | kātíbihạ | kaːˈti.bi.ha | kaː."ti.bi.ha |
| 0023 | كَاتِبِنَا | kātíbinạ | kaːˈti.bi.na | kaː."ti.bi.na |
| 0024 | كَاتِبِكُمْ | kātíbikum | kaːˈti.bi.kum | kaː."ti.bi.kum |
| 0025 | كَاتِبِكُنَّ | kātibikúnna | kaː.ti.biˈku.nːa | kaː.ti.bi."ku.nː.a |
| 0026 | كَاتِبِكُمَا | kātibíkumạ | kaː.tiˈbi.ku.ma | kaː.ti."bi.ku.ma |
| 0027 | كَاتِبِهِمْ | kātíbihim | kaːˈti.bi.him | kaː."ti.bi.him |
| 0028 | كَاتِبِهِنَّ | kātibihínna | kaː.ti.biˈhi.nːa | kaː.ti.bi."hi.nː.a |
| 0029 | كَاتِبِهِمَا | kātibíhimạ | kaː.tiˈbi.hi.ma | kaː.ti."bi.hi.ma |
| 0030 | كَاتِبِهِمَا | kātibíhimạ | kaː.tiˈbi.hi.ma | kaː.ti."bi.hi.ma |

## Phonological Component (verbs)

| SUBID | ARAB_V | CARS | IPA | XSAMPA |
|---|---|---|---|---|
| 0229 | كَتَبَ | kátaba | ˈka.ta.ba | "ka.ta.ba |
| 0230 | كَتَبَنِي | katábani̱ | kaˈta.ba.ni | ka."ta.ba.ni |

| 0231 | كَتَبَكَ | katábaka | ka.ˈta.ba.ka | ka."ta.ba.ka |
|---|---|---|---|---|
| 0232 | كَتَبَكِ | katábaki | ka.ˈta.ba.ki | ka."ta.ba.ki |
| 0233 | كَتَبَهُ | katábahu | ka.ˈta.ba.hu | ka."ta.ba.hu |
| 0234 | كَتَبَهَا | katábahạ | ka.ˈta.ba.ha | ka."ta.ba.ha |
| 0235 | كَتَبَنَا | katábanạ | ka.ˈta.ba.na | ka."ta.ba.na |
| 0236 | كَتَبَكُمْ | katábakum | ka.ˈta.ba.kum | ka."ta.ba.kum |
| 0237 | كَتَبَكُنَّ | katabakúnna | ka.ta.ba.ˈku.nːa | ka.ta.ba."ku.nːa |
| 0238 | كَتَبَكُمَا | katabákumạ | ka.ta.ˈba.ku.ma | ka.ta."ba.ku.ma |
| 0239 | كَتَبَهُمْ | katábahum | ka.ˈta.ba.hum | ka."ta.ba.hum |
| 0240 | كَتَبَهُنَّ | katabahúnna | ka.ta.ba.ˈhu.nːa | ka.ta.ba."hu.nːa |
| 0241 | كَتَبَهُمَا | katabáhumạ | ka.ta.ˈba.hu.ma | ka.ta."ba.hu.ma |
| 0255 | وَكَتَبَ | wakátaba | wa.ˈka.ta.ba | wa."ka.ta.ba |
| 0256 | وَكَتَبَنِي | wakatábanị | wa.ka.ˈta.ba.ni | wa.ka."ta.ba.ni |
| 0257 | وَكَتَبَكَ | wakatábaka | wa.ka.ˈta.ba.ka | wa.ka."ta.ba.ka |
| 0258 | وَكَتَبَكِ | wakatábaki | wa.ka.ˈta.ba.ki | wa.ka."ta.ba.ki |
| 0259 | وَكَتَبَهُ | wakatábahu | wa.ka.ˈta.ba.hu | wa.ka."ta.ba.hu |
| 0260 | وَكَتَبَهَا | wakatábahạ | wa.ka.ˈta.ba.ha | wa.ka."ta.ba.ha |
| 0261 | وَكَتَبَنَا | wakatábanạ | wa.ka.ˈta.ba.na | wa.ka."ta.ba.na |
| 0262 | وَكَتَبَكُمْ | wakatábakum | wa.ka.ˈta.ba.kum | wa.ka."ta.ba.kum |
| 0263 | وَكَتَبَكُنَّ | wakatabakúnna | wa.ka.ta.ba.ˈku.nːa | wa.ka.ta.ba."ku.nːa |
| 0264 | وَكَتَبَكُمَا | wakatabákumạ | wa.ka.ta.ˈba.ku.ma | wa.ka.ta."ba.ku.ma |
| 0265 | وَكَتَبَهُمْ | wakatábahum | wa.ka.ˈta.ba.hum | wa.ka."ta.ba.hum |
| 0266 | وَكَتَبَهُنَّ | wakatabahúnna | wa.ka.ta.ba.ˈhu.nːa | wa.ka.ta.ba."hu.nːa |
| 0267 | وَكَتَبَهُمَا | wakatabáhumạ | wa.ka.ta.ˈba.hu.ma | wa.ka.ta."ba.hu.ma |
| 0268 | فَكَتَبَ | fakátaba | fa.ˈka.ta.ba | fa."ka.ta.ba |
| 0269 | فَكَتَبَنِي | fakatábanị | fa.ka.ˈta.ba.ni | fa.ka."ta.ba.ni |

| | | | | |
|---|---|---|---|---|
| 0270 | فَكَتَبَكَ | fakatábaka | fa.ka.ˈta.ba.ka | fa.ka."ta.ba.ka |
| 0271 | فَكَتَبَكِ | fakatábaki | fa.ka.ˈta.ba.ki | fa.ka."ta.ba.ki |

# 3. DATABASE OF ARABIC PLACE NAMES Part I (DAP-C)

## 3.1 Synopsis

This **Arabic-English place name database** provides worldwide coverage of common place names, given in standard MSA orthography.

## 3.2 More Information

**Website:** http://www.cjk.org/data/arabic/proper/dapna/

## 3.3 Approximate Quantities

| Type | Quantities |
|---|---|
| **Total entries** | 21,000 |
| **Bilingual entries** | 7,000 |
| **Monolingual entries** | 14,000 |

## 3.4 The Canonical Component (DAP-C)

### 3.4.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique ID |
| 2 | **LEMMA_V** | vocalized Arabic place name |
| 3 | **LEMMA_BW** | vocalized Arabic place name in Buckwalter transliteration |
| 4 | **LEMMA_U** | unvocalized Arabic place name |
| 5 | **ENGLISH** | English equivalent |
| 6 | **REGION** | **indicates if place is in the Middle East**<br>[M] country or region in Arab League<br>[m] country or region is one of the following countries<br>- Afghanistan<br>- Iran<br>- Israel<br>- Pakistan<br>- Turkey |

### 3.4.2 Data Sample

| ID | LEMMA_V | LEMMA_BW | LEMMA_U | ENGLISH | REGION |
|---|---|---|---|---|---|
| P008289 | عَسْقَلَان | EasoqalaAnu | عسقلان | Ashkelon | m |
| P016507 | أَبْهَا | >abohaA | أبها | Abha | M |
| P017340 | اَلْقَاهِرَةُ | AaloqaAhirapu | القاهرة | Cairo | M |
| P018064 | قِطَاعُ غَزَّة | qiTaAEu gaz~ap | قطاع غزة | Gaza Strip | m |
| P018462 | اَلْعِرَاقُ | AaloEiraAqu | العراق | Iraq | M |
| P018961 | لَنْدَنُ | lanodanu | لندن | London | |
| P019537 | نِيُويُورْكُ | niyuwyuwroku | نيويورك | New York | |
| P019764 | أُوسَاكَا | >uwsaAkaA | أوساكا | Osaka | |
| P030090 | أَبُوظَبْيِ | >abuwZaboyi | أبوظبي | Abu Dhabi | M |

## 4. DATABASE OF ARABIC PLACE NAMES Part II (DAP-M, DAP-O, DAP-P)

### 4.1 Synopsis

**Part II of DAP** includes the **cliticized forms for the place names** in DAP Part I (DAP-C), both enclitics and proclitics (e.g. وَكَعَمَّانِ).

### 4.2 More Information

**Website:** http://www.cjk.org/data/arabic/nlp/arabic-full-form-lexicon/

### 4.3 Approximate Quantities

| Type | Quantities |
|---|---|
| **Seed names** | 21,000 |
| **Cliticized forms** | 6,400,000 |

### 4.4 The Morphological Component (DAP-M)

#### 4.4.1 Field Description

If the same schema (field structure) as DAG-M is required, all the fields shown in section 2.4 will be given. Many fields are irrelevant and will get a value of "-". Below are only the fields relevant to DAP.

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique foreign key to DAP-C |
| 2 | **SUBID** | unique identifier of cliticized form |
| 3 | **ARAB_V** | vocalized cliticized Arabic place name |
| 4 | **ARAB_BW** | vocalized cliticized Arabic place name in Buckwalter transliteration |
| 5 | **ARAB_U** | unvocalized cliticized Arabic place name |
| 6 | **LEMMA_V** | vocalized Arabic place name linked to DAP-C |
| 7 | **GEN** | **gender code**<br>[M] masculine<br>[F] feminine |
| 8 | **NUM** | **number code**<br>[S] singular<br>[D] dual<br>[P] plural<br>[B] broken plural |
| 9 | **CASE** | **case ending code**<br>[ACU] accusative<br>[GEN] genitive<br>[NOM] nominative |
| 10 | **PER2** | **person, number, gender for enclitics**<br>[000] zero person<br>[1SC] first person singular<br>[2SM] second person singular masculine<br>[2SF] second person singular feminine<br>[3SM] third person singular masculine<br>[3SF] third person singular feminine<br>[1PC] first person plural<br>[2PM] second person plural masculine<br>[2PF] second person plural feminine<br>[2DC] second person dual<br>[3PM] third person plural masculine<br>[3PF] third person plural feminine<br>[3DM] third person dual masculine<br>[3DF] third person dual feminine |
| 11 | **DEF** | **definiteness code**<br>[D] explicitly definite by cliticizing of article Al<br>[d] definite by adding pronomial enclitics |

| | | |
|---|---|---|
| | | [I] indefinite<br>[G] first term of genitive construct (Idhafa, e.g. *baytu in baytu-l-kaatibi*) |

### 4.4.2 Data Sample

| SUBID | ARAB_V | ARAB_BW | ARAB_U | LEMMA_V | GEN | NUM | CASE | PER | DEF |
|---|---|---|---|---|---|---|---|---|---|
| 0085 | وَمِصْرُ | wamiSoru | ومصر | مِصْرُ | F | S | NOM | 000 | G |
| 0086 | وَمِصْرِي | wamiSoriy | ومصري | مِصْرُ | F | S | NOM | 1SC | d |
| 0087 | وَمِصْرُكَ | wamiSoruka | ومصرك | مِصْرُ | F | S | NOM | 2SM | d |
| 0088 | وَمِصْرُكِ | wamiSoruki | ومصرك | مِصْرُ | F | S | NOM | 2SF | d |
| 0089 | وَمِصْرُهُ | wamiSoruhu | ومصره | مِصْرُ | F | S | NOM | 3SM | d |
| 0090 | وَمِصْرُهَا | wamiSoruhaA | ومصرها | مِصْرُ | F | S | NOM | 3SF | d |
| 0091 | وَمِصْرُنَا | wamiSorunaA | ومصرنا | مِصْرُ | F | S | NOM | 1PC | d |
| 0092 | وَمِصْرُكُمْ | wamiSorukumo | ومصركم | مِصْرُ | F | S | NOM | 2PM | d |
| 0093 | وَمِصْرُكُنَّ | wamiSorukun~a | ومصركن | مِصْرُ | F | S | NOM | 2PF | d |
| 0094 | وَمِصْرُكُمَا | wamiSorukumaA | ومصركما | مِصْرُ | F | S | NOM | 2DC | d |
| 0095 | وَمِصْرُهُمْ | wamiSoruhumo | ومصرهم | مِصْرُ | F | S | NOM | 3PM | d |
| 0096 | وَمِصْرُهُنَّ | wamiSoruhun~a | ومصرهن | مِصْرُ | F | S | NOM | 3PF | d |
| 0097 | وَمِصْرُهُمَا | wamiSoruhumaA | ومصرهما | مِصْرُ | F | S | NOM | 3DM | d |
| 0098 | وَمِصْرُهُمَا | wamiSoruhumaA | ومصرهما | مِصْرُ | F | S | NOM | 3DF | d |
| 0099 | وَمِصْرَ | wamiSora | ومصر | مِصْرُ | F | S | GEN | 000 | G |
| 0100 | وَمِصْرِي | wamiSoriy | ومصري | مِصْرُ | F | S | GEN | 1SC | d |
| 0101 | وَمِصْرَكَ | wamiSoraka | ومصرك | مِصْرُ | F | S | GEN | 2SM | d |
| 0102 | وَمِصْرَكِ | wamiSoraki | ومصرك | مِصْرُ | F | S | GEN | 2SF | d |
| 0103 | وَمِصْرَهُ | wamiSorahu | ومصره | مِصْرُ | F | S | GEN | 3SM | d |
| 0104 | وَمِصْرَهَا | wamiSorahaA | ومصرها | مِصْرُ | F | S | GEN | 3SF | d |
| 0105 | وَمِصْرَنَا | wamiSoranaA | ومصرنا | مِصْرُ | F | S | GEN | 1PC | d |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0106 | وَمِصْرَكُمْ | wamiSorakumo | ومصركم | مِصْرُ | F | S | GEN | 2PM | d |
| 0107 | وَمِصْرَكُنَّ | wamiSorakun~a | ومصركن | مِصْرُ | F | S | GEN | 2PF | d |
| 0108 | وَمِصْرَكُمَا | wamiSorakumaA | ومصركما | مِصْرُ | F | S | GEN | 2DC | d |
| 0109 | وَمِصْرَهُمْ | wamiSorahumo | ومصرهم | مِصْرُ | F | S | GEN | 3PM | d |
| 0110 | وَمِصْرَهُنَّ | wamiSorahun~a | ومصرهن | مِصْرُ | F | S | GEN | 3PF | d |
| 0111 | وَمِصْرَهُمَا | wamiSorahumaA | ومصرهما | مِصْرُ | F | S | GEN | 3DM | d |
| 0112 | وَمِصْرَهُمَا | wamiSorahumaA | ومصرهما | مِصْرُ | F | S | GEN | 3DF | d |
| 0113 | وَمِصْرَ | wamiSora | ومصر | مِصْرُ | F | S | ACU | 000 | G |
| 0114 | وَمِصْرِي | wamiSoriy | ومصري | مِصْرُ | F | S | ACU | 1SC | d |
| 0115 | وَمِصْرَكَ | wamiSoraka | ومصرك | مِصْرُ | F | S | ACU | 2SM | d |
| 0116 | وَمِصْرَكِ | wamiSoraki | ومصرك | مِصْرُ | F | S | ACU | 2SF | d |
| 0117 | وَمِصْرَهُ | wamiSorahu | ومصره | مِصْرُ | F | S | ACU | 3SM | d |
| 0118 | وَمِصْرَهَا | wamiSorahaA | ومصرها | مِصْرُ | F | S | ACU | 3SF | d |
| 0119 | وَمِصْرَنَا | wamiSoranaA | ومصرنا | مِصْرُ | F | S | ACU | 1PC | d |
| 0120 | وَمِصْرَكُمْ | wamiSorakumo | ومصركم | مِصْرُ | F | S | ACU | 2PM | d |
| 0121 | وَمِصْرَكُنَّ | wamiSorakun~a | ومصركن | مِصْرُ | F | S | ACU | 2PF | d |
| 0122 | وَمِصْرَكُمَا | wamiSorakumaA | ومصركما | مِصْرُ | F | S | ACU | 2DC | d |
| 0123 | وَمِصْرَهُمْ | wamiSorahumo | ومصرهم | مِصْرُ | F | S | ACU | 3PM | d |
| 0124 | وَمِصْرَهُنَّ | wamiSorahun~a | ومصرهن | مِصْرُ | F | S | ACU | 3PF | d |
| 0125 | وَمِصْرَهُمَا | wamiSorahumaA | ومصرهما | مِصْرُ | F | S | ACU | 3DM | d |
| 0126 | وَمِصْرَهُمَا | wamiSorahumaA | ومصرهما | مِصْرُ | F | S | ACU | 3DF | d |

## 4.5 The Orthographical Component (DAP-O)

### 4.5.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |

| | | |
|---|---|---|
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **VAR_V** | variant in vocalized Arabic |
| 6 | **VAR_U** | variant in unvocalized Arabic |

### 4.5.2 Data Sample

Under construction.

## 4.6 The Phonological Component (DAP-P)

### 4.6.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **CARS** | cliticized Arabic entry in phonemic transcription (CARS) |
| 6 | **IPA** | phonetic transcription in the International Phonetic Alphabet (IPA) |
| 7 | **XSAMPA** | phonetic transcription in X-SAMPA |

### 4.6.2 Data Sample

| SUBID | ARAB_V | CARS | IPA | XSAMPA |
|---|---|---|---|---|
| 0085 | وَمِصْرُ | wamíṣru | wa.ˈmiṣ.ru | wa."mis_-.ru |
| 0086 | وَمِصْرِي | wamíṣri̱ | wa.ˈmiṣ.ri | wa."mis_-.ri |
| 0087 | وَمِصْرُكَ | wamíṣruka | wa.ˈmiṣ.ru.ka | wa."mis_-.ru.ka |
| 0088 | وَمِصْرُكِ | wamíṣruki | wa.ˈmiṣ.ru.ki | wa."mis_-.ru.ki |
| 0089 | وَمِصْرُهُ | wamíṣruhu | wa.ˈmiṣ.ru.hu | wa."mis_-.ru.hu |

| 0090 | وَمِصْرُهَا | wamíṣruhạ | wa.ˈmiṣ.ru.ha | wa."mis_-.ru.ha |
|---|---|---|---|---|
| 0091 | وَمِصْرُنَا | wamíṣruna | wa.ˈmiṣ.ru.na | wa."mis_-.ru.na |
| 0092 | وَمِصْرُكُمْ | wamíṣrukum | wa.ˈmiṣ.ru.kum | wa."mis_-.ru.kum |
| 0093 | وَمِصْرُكُنَّ | wamiṣrukúnna | wa.miṣ.ru.ˈku.nːa | wa.mis_-.ru."ku.nːa |
| 0094 | وَمِصْرُكُمَا | wamiṣrúkumạ | wa.miṣ.ˈru.ku.ma | wa.mis_-."ru.ku.ma |
| 0095 | وَمِصْرُهُمْ | wamíṣruhum | wa.ˈmiṣ.ru.hum | wa."mis_-.ru.hum |
| 0096 | وَمِصْرُهُنَّ | wamiṣruhúnna | wa.miṣ.ru.ˈhu.nːa | wa.mis_-.ru."hu.nːa |
| 0097 | وَمِصْرُهُمَا | wamiṣrúhumạ | wa.miṣ.ˈru.hu.ma | wa.mis_-."ru.hu.ma |
| 0098 | وَمِصْرُهُمَا | wamiṣrúhumạ | wa.miṣ.ˈru.hu.ma | wa.mis_-."ru.hu.ma |
| 0099 | وَمِصْرَ | wamíṣra | wa.ˈmiṣ.ra | wa."mis_-.ra |
| 0100 | وَمِصْرِي | wamíṣrị | wa.ˈmiṣ.ri | wa."mis_-.ri |
| 0101 | وَمِصْرَكَ | wamíṣraka | wa.ˈmiṣ.ra.ka | wa."mis_-.ra.ka |
| 0102 | وَمِصْرَكِ | wamíṣraki | wa.ˈmiṣ.ra.ki | wa."mis_-.ra.ki |
| 0103 | وَمِصْرَهُ | wamíṣrahu | wa.ˈmiṣ.ra.hu | wa."mis_-.ra.hu |
| 0104 | وَمِصْرَهَا | wamíṣrahạ | wa.ˈmiṣ.ra.ha | wa."mis_-.ra.ha |
| 0105 | وَمِصْرَنَا | wamíṣrana | wa.ˈmiṣ.ra.na | wa."mis_-.ra.na |
| 0106 | وَمِصْرَكُمْ | wamíṣrakum | wa.ˈmiṣ.ra.kum | wa."mis_-.ra.kum |
| 0107 | وَمِصْرَكُنَّ | wamiṣrakúnna | wa.miṣ.ra.ˈku.nːa | wa.mis_-.ra."ku.nːa |
| 0108 | وَمِصْرَكُمَا | wamiṣrákumạ | wa.miṣ.ˈra.ku.ma | wa.mis_-."ra.ku.ma |
| 0109 | وَمِصْرَهُمْ | wamíṣrahum | wa.ˈmiṣ.ra.hum | wa."mis_-.ra.hum |
| 0110 | وَمِصْرَهُنَّ | wamiṣrahúnna | wa.miṣ.ra.ˈhu.nːa | wa.mis_-.ra."hu.nːa |
| 0111 | وَمِصْرَهُمَا | wamiṣráhumạ | wa.miṣ.ˈra.hu.ma | wa.mis_-."ra.hu.ma |
| 0112 | وَمِصْرَهُمَا | wamiṣráhumạ | wa.miṣ.ˈra.hu.ma | wa.mis_-."ra.hu.ma |
| 0113 | وَمِصْرَ | wamíṣra | wa.ˈmiṣ.ra | wa."mis_-.ra |
| 0114 | وَمِصْرِي | wamíṣrị | wa.ˈmiṣ.ri | wa."mis_-.ri |
| 0115 | وَمِصْرَكَ | wamíṣraka | wa.ˈmiṣ.ra.ka | wa."mis_-.ra.ka |

| | | | | |
|---|---|---|---|---|
| 0116 | وَمِصْرَكِ | wamíṣraki | wa.ˈmiṣ.ra.ki | wa."mis_-.ra.ki |
| 0117 | وَمِصْرَهُ | wamíṣrahu | wa.ˈmiṣ.ra.hu | wa."mis_-.ra.hu |
| 0118 | وَمِصْرَهَا | wamíṣraha | wa.ˈmiṣ.ra.ha | wa."mis_-.ra.ha |
| 0119 | وَمِصْرَنَا | wamíṣrana | wa.ˈmiṣ.ra.na | wa."mis_-.ra.na |
| 0120 | وَمِصْرَكُمْ | wamíṣrakum | wa.ˈmiṣ.ra.kum | wa."mis_-.ra.kum |
| 0121 | وَمِصْرَكُنَّ | wamiṣrakúnna | wa.miṣ.ra.ˈku.nːa | wa.mis_-.ra."ku.nːa |
| 0122 | وَمِصْرَكُمَا | wamiṣrákuma | wa.miṣ.ˈra.ku.ma | wa.mis_-."ra.ku.ma |
| 0123 | وَمِصْرَهُمْ | wamíṣrahum | wa.ˈmiṣ.ra.hum | wa."mis_-.ra.hum |
| 0124 | وَمِصْرَهُنَّ | wamiṣrahúnna | wa.miṣ.ra.ˈhu.nːa | wa.mis_-.ra."hu.nːa |
| 0125 | وَمِصْرَهُمَا | wamiṣráhuma | wa.miṣ.ˈra.hu.ma | wa.mis_-."ra.hu.ma |
| 0126 | وَمِصْرَهُمَا | wamiṣráhuma | wa.miṣ.ˈra.hu.ma | wa.mis_-."ra.hu.ma |

## 5. DATABASE OF ARABIC FOREIGN NAMES Part I (DAF-C)

### 5.1 Synopsis

This database covers non-Arab personal names, their Arabic equivalents, some Arabic script variants and English equivalents. All DAF data can be provided in vocalized or unvocalized formats.

### 5.2 More Information

**Website:** http://www.cjk.org/data/arabic/proper/dafna/

### 5.3 Approximate Quantities

The approximate number of entries is as follows:

| Type | Quantities |
|---|---|
| **Total entries** | 400,000 |
| **Seed names** | 223,367 |
| **Arabic variants** | 176,633 |

### 5.4 The Canonical Component (DAF-C)

#### 5.4.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique ID identifying entry block (unique on English) |
| 2 | **LEMMA_V** | vocalized non-Arab personal name |
| 3 | **LEMMA_BW** | vocalized non-Arab personal name in Buckwalter transliteration |
| 4 | **LEMMA_U** | unvocalized non-Arab personal name |
| 5 | **ENGLISH** | English original |
| 6 | **TYPE** | **type of name**<br>[G] given name<br>[S] surname<br>[GS] given name and surname |
| 7 | **GEN** | **gender of name**<br>[M] male<br>[F] female<br>[MF] male and female<br>[-] not applicable |
| 8 | **RS_FREQ** | frequency of Latin surname based on US Census |
| 9 | **RG_FREQ** | frequency of Latin given name based on Social Security statistics |

#### 5.4.2 Data Sample

| ID | LEMMA_V | LEMMA_BW | LEMMA_U | ENGLISH | TYPE | GEN | RS_FREQ | RG_FREQ |
|---|---|---|---|---|---|---|---|---|
| F086620 | هَالْبِرْن | haAlobiron | هالبرن | Halpern | S | - | 0004121 | - |
| F098222 | إِزَابِيلَا | <izaAbiylaA | إزابيلا | Izabella | G | F | - | 0025717 |
| F098611 | جَاك | jaAk | جاك | Jack | GS | MF | 0015256 | 0696625 |
| F100877 | جَانِيت | jaAniyt | جانيت | Janet | GS | MF | 0000437 | 0557605 |
| F107324 | جُولِيت | juwliyt | جوليت | Juliet | G | F | - | 0030202 |
| F170757 | بِيتِرْسُون | biytirosuwn | بيترسون | Peterson | GS | M | 0278297 | 0000756 |
| F193601 | شْمِيت | $omiyt | شميت | Schmidt | S | - | 0147034 | - |
| F204232 | سْمِيث | somiyv | سميث | Smith | GS | MF | 2442977 | 0004733 |
| F234748 | وِيلِيَام | wiyliyaAm | ويليام | William | GS | MF | 0013373 | 4133327 |

# 6. DATABASE OF ARABIC FOREIGN NAMES Part II (DAF-M, DAF-O, DAF-P)

## 6.1 Synopsis

**Part II of DAF** includes the **cliticized forms for the non-Arab personal names**. Some cliticized forms may be relatively rare but they are valid.

## 6.2 More Information

**Website:** http://www.cjk.org/data/arabic/nlp/arabic-full-form-lexicon/

## 6.3 Approximate Quantities

| Type | Quantities |
|---|---|
| **Seed names** | 223,367 |
| **Cliticized forms** | 226,784,907 |

## 6.4 The Morphological Component (DAF-M)

### 6.4.1 Field Description

If the same schema (field structure) as DAG-M is required, all the fields shown in section 2.4 will be given. Some fields are irrelevant and will get a value of "-". Below are only the fields relevant to DAF.

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique foreign key to DAF-C |
| 2 | **SUBID** | unique identifier of cliticized form |
| 3 | **ARAB_V** | vocalized cliticized non-Arab personal name |
| 4 | **ARAB_BW** | vocalized cliticized non-Arab personal name in Buckwalter transliteration |
| 5 | **ARAB_U** | unvocalized cliticized non-Arab personal name |
| 6 | **LEMMA_V** | vocalized non-Arab personal name linked to DAF-C |
| 7 | **GEN** | **gender code**<br>[M] masculine<br>[F] feminine<br>[C] common gender |
| 8 | **NUM** | **number code**<br>[S] singular<br>[D] dual<br>[P] plural<br>[B] broken plural |
| 9 | **CASE** | **case ending code**<br>[ACU] accusative<br>[GEN] genitive |

| | | |
|---|---|---|
| | | [NOM] nominative |
| 10 | **PER2** | **person, number, gender for enclitics**<br>[000] zero person<br>[1SC] first person singular<br>[2SM] second person singular masculine<br>[2SF] second person singular feminine<br>[3SM] third person singular masculine<br>[3SF] third person singular feminine<br>[1PC] first person plural<br>[2PM] second person plural masculine<br>[2PF] second person plural feminine<br>[2DC] second person dual<br>[3PM] third person plural masculine<br>[3PF] third person plural feminine<br>[3DM] third person dual masculine<br>[3DF] third person dual feminine |
| 11 | **DEF** | **definiteness code**<br>[D] explicitly definite by cliticizing of article Al<br>[d] definite by adding pronomial enclitics<br>[I] indefinite<br>[G] first term of genitive construct (Idhafa, e.g. *baytu in baytu-l-kaatibi*) |

### 6.4.2 Data Sample

| SUBID | ARAB_V | ARAB_BW | ARAB_U | LEMMA_V | GEN | NUM | CASE | PER2 | DEF |
|---|---|---|---|---|---|---|---|---|---|
| 0083 | وَجَاكُ | wajaAku | وجاك | جَاكُ | C | S | NOM | 000 | G |
| 0084 | وَجَاكِي | wajaAkiy | وجاكي | جَاكُ | C | S | NOM | 1SC | d |
| 0085 | وَجَاكُكَ | wajaAkuka | وجاكك | جَاكُ | C | S | NOM | 2SM | d |
| 0086 | وَجَاكُكِ | wajaAkuki | وجاكك | جَاكُ | C | S | NOM | 2SF | d |
| 0087 | وَجَاكُهُ | wajaAkuhu | وجاكه | جَاكُ | C | S | NOM | 3SM | d |
| 0088 | وَجَاكُهَا | wajaAkuhaA | وجاكها | جَاكُ | C | S | NOM | 3SF | d |
| 0089 | وَجَاكُنَا | wajaAkunaA | وجاكنا | جَاكُ | C | S | NOM | 1PC | d |
| 0090 | وَجَاكُكُمْ | wajaAkukumo | وجاككم | جَاكُ | C | S | NOM | 2PM | d |
| 0091 | وَجَاكُكُنَّ | wajaAkukun~a | وجاككن | جَاكُ | C | S | NOM | 2PF | d |
| 0092 | وَجَاكُكُمَا | wajaAkukumaA | وجاككما | جَاكُ | C | S | NOM | 2DC | d |
| 0093 | وَجَاكُهُمْ | wajaAkuhumo | وجاكهم | جَاكُ | C | S | NOM | 3PM | d |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0094 | وَجَاكُهُنَّ | wajaAkuhun~a | وجاكهن | جَاكَ | C | S | NOM | 3PF | d |
| 0095 | وَجَاكُهُمَا | wajaAkuhumaA | وجاكهما | جَاكَ | C | S | NOM | 3DM | d |
| 0096 | وَجَاكُهُمَا | wajaAkuhumaA | وجاكهما | جَاكَ | C | S | NOM | 3DF | d |
| 0097 | وَجَاكِي | wajaAkiy | وجاكي | جَاكَ | C | S | GEN | 1SC | d |
| 0098 | وَجَاكِكَ | wajaAkika | وجاكك | جَاكَ | C | S | GEN | 2SM | d |
| 0099 | وَجَاكِكِ | wajaAkiki | وجاكك | جَاكَ | C | S | GEN | 2SF | d |
| 0100 | وَجَاكِهِ | wajaAkihi | وجاكه | جَاكَ | C | S | GEN | 3SM | d |
| 0101 | وَجَاكِهَا | wajaAkihaA | وجاكها | جَاكَ | C | S | GEN | 3SF | d |
| 0102 | وَجَاكِنَا | wajaAkinaA | وجاكنا | جَاكَ | C | S | GEN | 1PC | d |
| 0103 | وَجَاكِكُمْ | wajaAkikumo | وجاككم | جَاكَ | C | S | GEN | 2PM | d |
| 0104 | وَجَاكِكُنَّ | wajaAkikun~a | وجاككن | جَاكَ | C | S | GEN | 2PF | d |
| 0105 | وَجَاكِكُمَا | wajaAkikumaA | وجاككما | جَاكَ | C | S | GEN | 2DC | d |
| 0106 | وَجَاكِهِمْ | wajaAkihimo | وجاكهم | جَاكَ | C | S | GEN | 3PM | d |
| 0107 | وَجَاكِهِنَّ | wajaAkihin~a | وجاكهن | جَاكَ | C | S | GEN | 3PF | d |
| 0108 | وَجَاكِهِمَا | wajaAkihimaA | وجاكهما | جَاكَ | C | S | GEN | 3DM | d |
| 0109 | وَجَاكِهِمَا | wajaAkihimaA | وجاكهما | جَاكَ | C | S | GEN | 3DF | d |
| 0110 | وَجَاكَ | wajaAka | وجاك | جَاكَ | C | S | ACU | 000 | G |
| 0111 | وَجَاكِي | wajaAkiy | وجاكي | جَاكَ | C | S | ACU | 1SC | d |
| 0112 | وَجَاكَكَ | wajaAkaka | وجاكك | جَاكَ | C | S | ACU | 2SM | d |

## 6.5 The Orthographical Component (DAF-O)

### 6.5.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |

| | | |
|---|---|---|
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **VAR_V** | variant in vocalized Arabic |
| 6 | **VAR_U** | variant in unvocalized Arabic |

### 6.5.2 Data Sample

Under construction.

## 6.6 The Phonological Component (DAF-P)

### 6.6.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **CARS** | cliticized Arabic entry in phonemic transcription (CARS) |
| 6 | **IPA** | phonetic transcription in the International Phonetic Alphabet (IPA) |
| 7 | **XSAMPA** | phonetic transcription in X-SAMPA |

### 6.6.2 Data Sample

| SUBID | ARAB_V | CARS | IPA | XSAMPA |
|---|---|---|---|---|
| 0083 | وَجَاكُ | wajáku | wa.ˈʒa:.ku | wa."Za:.ku |
| 0084 | وَجَاكِي | wajáki̱ | wa.ˈʒa:.ki | wa."Za:.ki |
| 0085 | وَجَاكُكَ | wajákuka | wa.ˈʒa:.ku.ka | wa."Za:.ku.ka |
| 0086 | وَجَاكُكِ | wajákuki | wa.ˈʒa:.ku.ki | wa."Za:.ku.ki |
| 0087 | وَجَاكُهُ | wajákuhu | wa.ˈʒa:.ku.hu | wa."Za:.ku.hu |
| 0088 | وَجَاكُهَا | wajákuha̱ | wa.ˈʒa:.ku.ha | wa."Za:.ku.ha |
| 0089 | وَجَاكُنَا | wajákuna̱ | wa.ˈʒa:.ku.na | wa."Za:.ku.na |
| 0090 | وَجَاكُكُمْ | wajákukum | wa.ˈʒa:.ku.kum | wa."Za:.ku.kum |

| | | | | |
|---|---|---|---|---|
| 0091 | وَجَاكُكُنَّ | wajākukúnna | wa.ʒaː.ku.ˈku.nːa | wa.Za:.ku."ku.nːa |
| 0092 | وَجَاكُكُمَا | wajākúkumạ | wa.ʒaː.ˈku.ku.ma | wa.Za:."ku.ku.ma |
| 0093 | وَجَاكُهُمْ | wajā́kuhum | wa.ˈʒaː.ku.hum | wa."Za:.ku.hum |
| 0094 | وَجَاكُهُنَّ | wajākuhúnna | wa.ʒaː.ku.ˈhu.nːa | wa.Za:.ku."hu.nːa |
| 0095 | وَجَاكُهُمَا | wajākúhumạ | wa.ʒaː.ˈku.hu.ma | wa.Za:."ku.hu.ma |
| 0096 | وَجَاكُهُمَا | wajākúhumạ | wa.ʒaː.ˈku.hu.ma | wa.Za:."ku.hu.ma |
| 0097 | وَجَاكِي | wajā́kị | wa.ˈʒaː.ki | wa."Za:.ki |
| 0098 | وَجَاكِكَ | wajā́kika | wa.ˈʒaː.ki.ka | wa."Za:.ki.ka |
| 0099 | وَجَاكِكِ | wajā́kiki | wa.ˈʒaː.ki.ki | wa."Za:.ki.ki |
| 0100 | وَجَاكِهِ | wajā́kihi | wa.ˈʒaː.ki.hi | wa."Za:.ki.hi |
| 0101 | وَجَاكِهَا | wajā́kihạ | wa.ˈʒaː.ki.ha | wa."Za:.ki.ha |
| 0102 | وَجَاكِنَا | wajā́kinạ | wa.ˈʒaː.ki.na | wa."Za:.ki.na |
| 0103 | وَجَاكِكُمْ | wajā́kikum | wa.ˈʒaː.ki.kum | wa."Za:.ki.kum |
| 0104 | وَجَاكِكُنَّ | wajākikúnna | wa.ʒaː.ki.ˈku.nːa | wa.Za:.ki."ku.nːa |
| 0105 | وَجَاكِكُمَا | wajākíkumạ | wa.ʒaː.ˈki.ku.ma | wa.Za:."ki.ku.ma |
| 0106 | وَجَاكِهِمْ | wajā́kihim | wa.ˈʒaː.ki.him | wa."Za:.ki.him |
| 0107 | وَجَاكِهِنَّ | wajākihínna | wa.ʒaː.ki.ˈhi.nːa | wa.Za:.ki."hi.nːa |
| 0108 | وَجَاكِهِمَا | wajākíhimạ | wa.ʒaː.ˈki.hi.ma | wa.Za:."ki.hi.ma |
| 0109 | وَجَاكِهِمَا | wajākíhimạ | wa.ʒaː.ˈki.hi.ma | wa.Za:."ki.hi.ma |
| 0110 | وَجَاكَ | wajā́ka | wa.ˈʒaː.ka | wa."Za:.ka |
| 0111 | وَجَاكِي | wajā́kị | wa.ˈʒaː.ki | wa."Za:.ki |
| 0112 | وَجَاكَكَ | wajā́kaka | wa.ˈʒaː.ka.ka | wa."Za:.ka.ka |

# 7. DATABASE OF ARABIC NAMES Part I (DAN-C)

## 7.1 Synopsis

Very comprehensive database of **Arab personal names and** romanized **name variants**

with a variety of supplementary information.

### 7.2 More Information

**Website:** http://www.cjk.org/data/arabic/proper/dan/
**White paper:** http://www.cjk.org/cjk/arabic/DAN3.pdf
**Research paper:** http://www.cjk.org/cjk/reference/danpaper.pdf
**Online demo:** http://cjki.dyndns.org/ante/ante.php (login "dandana")

### 7.3 Approximate Quantities

| Type | Quantities |
|---|---|
| **Seed names** | unvocalized: 76,000<br>vocalized: approx 100,000 |
| **Romanized variants** | 6,595,008 |

### 7.4 The Canonical Component (DAN-C)

#### 7.4.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique identifier for vocalized Arab personal name |
| 2 | **U_ID** | unique identifier for unvocalized Arab personal name |
| 3 | **SUBID** | identifies members of a variants group |
| 4 | **R_NAME** | romanized variant of ARABIC |
| 5 | **LEMMA_V** | vocalized Arab personal name |
| 6 | **LEMMA_BW** | vocalized Arab personal name in Buckwalter transliteration |
| 7 | **LEMMA_U** | unvocalized Arab personal name |
| 8 | **TYPE** | **type of name**<br>[G] given name<br>[S] surname<br>[GS] given name and surname<br>[U] unknown |
| 9 | **GEN** | **gender of name**<br>[M] male<br>[F] female<br>[MF] male and female<br>[U] unknown<br>[-] not applicable |
| 10 | **R_TYPE** | **type of romanization**<br>[I] IC (Intelligence Community) Standard<br>[V] variant of undetermined Romanization system |
| 11 | **R_FREQ** | web frequency as a string, not necessarily representative of web frequency as a name |

### 7.4.2 Data Sample

| SUBID | R_NAME | LEMMA_V | LEMMA_BW | TYPE | GEN | R_TYPE | R_FREQ |
|---|---|---|---|---|---|---|---|
| 001683 | Abdurrahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0002680000 |
| 001421 | Abderrahmane | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000603000 |
| 001647 | Abdulrahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000589000 |
| 001617 | Abdul Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000478000 |
| 001387 | Abdelrahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000348000 |
| 001420 | Abderrahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000206000 |
| 001366 | Abdel Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000136000 |
| 000946 | Abd Al Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000120000 |
| 001671 | Abdurahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000074000 |
| 001407 | Abderahmane | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000061900 |
| 001660 | Abdur Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000056400 |
| 001472 | Abdirahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000043700 |
| 001625 | Abdul Rehman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000040700 |
| 001013 | Abd Elrahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000035800 |
| 001665 | Abdur Rehman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000023600 |
| 001388 | Abdelrahmane | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000017300 |
| 000996 | Abd El Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000013970 |
| 000968 | Abd Ar Rahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000012100 |
| 001406 | Abderahman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000010000 |
| 001654 | Abdulrehman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000009200 |
| 001574 | Abdourahmane | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000008740 |
| 001681 | Abdurrachman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000008570 |
| 001674 | Abdurakhman | عَبْد ٱلرَّحْمَٰن | Eabod Aalr~aHom`n | G | M | V | 0000008000 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 001668 | Abdurachman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000007330 |
| 001514 | Abdolrahman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000007220 |
| 001481 | Abdirrahman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000007070 |
| 001573 | Abdourahman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000006600 |
| 001613 | Abdul Rachman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000005240 |
| 001542 | Abdorrahman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000004510 |
| 001645 | Abdulrachman | عَبْد اَلرَّحْمٰن | Eabod Aalr~aHom`n | G | M | V | 0000004350 |

# 8. DATABASE OF ARABIC NAMES Part II (DAN-M, DAN-O, DAN-P)

## 8.1 Synopsis

**Part II of DAN** that includes the **cliticized forms for Arab personal names**. Some encliticized forms may be relatively rare but they are valid.

## 8.2 More Information

**Website:** http://www.cjk.org/data/arabic/nlp/arabic-full-form-lexicon/
**White paper:** ArabLEX.pdf

## 8.3 Approximate Quantities

| Type | Quantities |
|---|---|
| **Seed names** | unvocalized: 76,000<br>vocalized: approx 100,000 |
| **Cliticized forms** | 218,215,875 |

## 8.4 The Morphological Component (DAN-M)

### 8.4.1 Field Description

If the same schema (field structure) as DAG-M is required, all the fields shown in section 2.4 will be given. Many fields are irrelevant and will get a value of "-". Below are only the fields relevant to DAN.

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | unique foreign key to DAN-C |

| | | |
|---|---|---|
| 2 | **SUBID** | unique identifier for cliticized forms |
| 3 | **ARAB_V** | vocalized cliticized Arab personal name |
| 4 | **ARAB_BW** | vocalized cliticized Arab personal name in Buckwalter transliteration |
| 5 | **ARAB_U** | unvocalized cliticized Arab personal name |
| 6 | **LEMMA_V** | vocalized Arab personal name linked to DAN-C |
| 7 | **GEN** | **gender code**<br>[M] masculine<br>[F] feminine<br>[C] common gender |
| 8 | **NUM** | **number code**<br>[S] singular<br>[D] dual<br>[P] plural<br>[B] broken plural |
| 9 | **CASE** | **case ending code**<br>[ACU] accusative<br>[GEN] genitive<br>[NOM] nominative |
| 10 | **PER2** | **person, number, gender for enclitics**<br>[000] zero person<br>[1SC] first person singular<br>[2SM] second person singular masculine<br>[2SF] second person singular feminine<br>[3SM] third person singular masculine<br>[3SF] third person singular feminine<br>[1PC] first person plural<br>[2PM] second person plural masculine<br>[2PF] second person plural feminine<br>[2DC] second person dual<br>[3PM] third person plural masculine<br>[3PF] third person plural feminine<br>[3DM] third person dual masculine<br>[3DF] third person dual feminine |
| 11 | **DEF** | **definiteness code**<br>[D] explicitly definite by cliticizing of article Al<br>[d] definite by adding pronomial enclitics<br>[I] indefinite<br>[G] first term of genitive construct (Idhafa, e.g. baytu in baytu-l-kaatibi) |

### 8.4.2 Data Sample

| SUBID | ARAB_V | ARAB_BW | ARAB_U | LEMMA_V | GEN | NUM | CASE | PER2 | DEF |
|---|---|---|---|---|---|---|---|---|---|
| 0089 | وَمُحَمَّدٌ | wamuHam~adN | ومحمد | مُحَمَّد | C | S | NOM | 000 | I |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0090 | وَمُحَمَّدُ | wamuHam~adu | ومحمد | مُحَمَّد | C | S | NOM | 000 | G |
| 0091 | وَمُحَمَّدِي | wamuHam~adiy | ومحمدي | مُحَمَّد | C | S | NOM | 1SC | d |
| 0092 | وَمُحَمَّدُكَ | wamuHam~aduka | ومحمدك | مُحَمَّد | C | S | NOM | 2SM | d |
| 0093 | وَمُحَمَّدُكِ | wamuHam~aduki | ومحمدك | مُحَمَّد | C | S | NOM | 2SF | d |
| 0094 | وَمُحَمَّدُهُ | wamuHam~aduhu | ومحمده | مُحَمَّد | C | S | NOM | 3SM | d |
| 0095 | وَمُحَمَّدُهَا | wamuHam~aduhaA | ومحمدها | مُحَمَّد | C | S | NOM | 3SF | d |
| 0096 | وَمُحَمَّدُنَا | wamuHam~adunaA | ومحمدنا | مُحَمَّد | C | S | NOM | 1PC | d |
| 0097 | وَمُحَمَّدُكُمْ | wamuHam~adukumo | ومحمدكم | مُحَمَّد | C | S | NOM | 2PM | d |
| 0098 | وَمُحَمَّدُكُنَّ | wamuHam~adukun~a | ومحمدكن | مُحَمَّد | C | S | NOM | 2PF | d |
| 0099 | وَمُحَمَّدُكُمَا | wamuHam~adukumaA | ومحمدكما | مُحَمَّد | C | S | NOM | 2DC | d |
| 0100 | وَمُحَمَّدُهُمْ | wamuHam~aduhumo | ومحمدهم | مُحَمَّد | C | S | NOM | 3PM | d |
| 0101 | وَمُحَمَّدُهُنَّ | wamuHam~aduhun~a | ومحمدهن | مُحَمَّد | C | S | NOM | 3PF | d |
| 0102 | وَمُحَمَّدُهُمَا | wamuHam~aduhumaA | ومحمدهما | مُحَمَّد | C | S | NOM | 3DM | d |
| 0103 | وَمُحَمَّدُهُمَا | wamuHam~aduhumaA | ومحمدهما | مُحَمَّد | C | S | NOM | 3DF | d |
| 0104 | وَمُحَمَّدٍ | wamuHam~adK | ومحمد | مُحَمَّد | C | S | GEN | 000 | I |
| 0105 | وَمُحَمَّدِ | wamuHam~adi | ومحمد | مُحَمَّد | C | S | GEN | 000 | G |
| 0106 | وَمُحَمَّدِي | wamuHam~adiy | ومحمدي | مُحَمَّد | C | S | GEN | 1SC | d |
| 0107 | وَمُحَمَّدِكَ | wamuHam~adika | ومحمدك | مُحَمَّد | C | S | GEN | 2SM | d |
| 0108 | وَمُحَمَّدِكِ | wamuHam~adiki | ومحمدك | مُحَمَّد | C | S | GEN | 2SF | d |
| 0109 | وَمُحَمَّدِهِ | wamuHam~adihi | ومحمده | مُحَمَّد | C | S | GEN | 3SM | d |
| 0110 | وَمُحَمَّدِهَا | wamuHam~adihaA | ومحمدها | مُحَمَّد | C | S | GEN | 3SF | d |
| 0111 | وَمُحَمَّدِنَا | wamuHam~adinaA | ومحمدنا | مُحَمَّد | C | S | GEN | 1PC | d |
| 0112 | وَمُحَمَّدِكُمْ | wamuHam~adikumo | ومحمدكم | مُحَمَّد | C | S | GEN | 2PM | d |
| 0113 | وَمُحَمَّدِكُنَّ | wamuHam~adikun~a | ومحمدكن | مُحَمَّد | C | S | GEN | 2PF | d |
| 0114 | وَمُحَمَّدِكُمَا | wamuHam~adikumaA | ومحمدكما | مُحَمَّد | C | S | GEN | 2DC | d |
| 0115 | وَمُحَمَّدِهِمْ | wamuHam~adihimo | ومحمدهم | مُحَمَّد | C | S | GEN | 3PM | d |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0116 | وَمُحَمَّدِهِنَّ | wamuHam~adihin~a | ومحمدهن | مُحَمَّد | C | S | GEN | 3PF | d |
| 0117 | وَمُحَمَّدِهِمَا | wamuHam~adihimaA | ومحمدهما | مُحَمَّد | C | S | GEN | 3DM | d |
| 0118 | وَمُحَمَّدِهِمَا | wamuHam~adihimaA | ومحمدهما | مُحَمَّد | C | S | GEN | 3DF | d |
| 0119 | وَمُحَمَّدًا | wamuHam~adFA | ومحمدا | مُحَمَّد | C | S | ACU | 000 | I |
| 0120 | وَمُحَمَّدَ | wamuHam~ada | ومحمد | مُحَمَّد | C | S | ACU | 000 | G |

## 8.5 The Orthographical Component (DAN-O)

### 8.5.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **VAR_V** | variant in vocalized Arabic |
| 6 | **VAR_U** | variant in unvocalized Arabic |

### 8.5.2 Data Sample

Under construction.

## 8.6 The Phonological Component (DAN-P)

### 8.6.1 Field Description

| No. | Field | Description |
|---|---|---|
| 1 | **ID** | uniquely identifies the lemma |
| 2 | **SUBID** | uniquely identifies inflected forms |
| 3 | **VARID** | **identifies variants of headword**<br>[00] no variants<br>[01] main form<br>[02] variant form |
| 4 | **ARAB_V** | vocalized cliticized Arabic entry |
| 5 | **CARS** | cliticized Arabic entry in phonemic transcription (CARS) |

| | | |
|---|---|---|
| 6 | **IPA** | phonetic transcription in the International Phonetic Alphabet (IPA) |
| 7 | **XSAMPA** | phonetic transcription in X-SAMPA |

### 8.6.2 Data Sample

| SUBID | ARAB_V | CARS | IPA | XSAMPA |
|---|---|---|---|---|
| 0089 | وَمُحَمَّدٌ | wamuḥámmadun | wa.mu.ˈħa.mːa.dun | wa.mu."X\a.mːa.dun |
| 0090 | وَمُحَمَّدُ | wamuḥámmadu | wa.mu.ˈħa.mːa.du | wa.mu."X\a.mːa.du |
| 0091 | وَمُحَمَّدِي | wamuḥámmadi̱ | wa.mu.ˈħa.mːa.di | wa.mu."X\a.mːa.di |
| 0092 | وَمُحَمَّدُكَ | wamuḥammáduka | wa.mu.ħa.ˈmːa.du.ka | wa.mu.X\a."mːa.du.ka |
| 0093 | وَمُحَمَّدُكِ | wamuḥammáduki | wa.mu.ħa.ˈmːa.du.ki | wa.mu.X\a."mːa.du.ki |
| 0094 | وَمُحَمَّدُهُ | wamuḥammáduhu | wa.mu.ħa.ˈmːa.du.hu | wa.mu.X\a."mːa.du.hu |
| 0095 | وَمُحَمَّدُهَا | wamuḥammáduha̱ | wa.mu.ħa.ˈmːa.du.ha | wa.mu.X\a."mːa.du.ha |
| 0096 | وَمُحَمَّدُنَا | wamuḥammáduna̱ | wa.mu.ħa.ˈmːa.du.na | wa.mu.X\a."mːa.du.na |
| 0097 | وَمُحَمَّدُكُمْ | wamuḥammádukum | wa.mu.ħa.ˈmːa.du.kum | wa.mu.X\a."mːa.du.kum |
| 0098 | وَمُحَمَّدُكُنَّ | wamuḥammadukúnna | wa.mu.ħa.mːa.du.ˈku.nːa | wa.mu.X\a.mːa.du."ku.nːa |
| 0099 | وَمُحَمَّدُكُمَا | wamuḥammadúkuma̱ | wa.mu.ħa.mːa.ˈdu.ku.ma | wa.mu.X\a.mːa."du.ku.ma |
| 0100 | وَمُحَمَّدُهُمْ | wamuḥammáduhum | wa.mu.ħa.ˈmːa.du.hum | wa.mu.X\a."mːa.du.hum |
| 0101 | وَمُحَمَّدُهُنَّ | wamuḥammaduhúnna | wa.mu.ħa.mːa.du.ˈhu.nːa | wa.mu.X\a.mːa.du."hu.nːa |
| 0102 | وَمُحَمَّدُهُمَا | wamuḥammadúhuma̱ | wa.mu.ħa.mːa.ˈdu.hu.ma | wa.mu.X\a.mːa."du.hu.ma |
| 0103 | وَمُحَمَّدُهُمَا | wamuḥammadúhuma̱ | wa.mu.ħa.mːa.ˈdu.hu.ma | wa.mu.X\a.mːa."du.hu.ma |
| 0104 | وَمُحَمَّدٍ | wamuḥámmadin | wa.mu.ˈħa.mːa.din | wa.mu."X\a.mːa.din |
| 0105 | وَمُحَمَّدِ | wamuḥámmadi | wa.mu.ˈħa.mːa.di | wa.mu."X\a.mːa.di |
| 0106 | وَمُحَمَّدِي | wamuḥámmadi̱ | wa.mu.ˈħa.mːa.di | wa.mu."X\a.mːa.di |
| 0107 | وَمُحَمَّدِكَ | wamuḥammádika | wa.mu.ħa.ˈmːa.di.ka | wa.mu.X\a."mːa.di.ka |
| 0108 | وَمُحَمَّدِكِ | wamuḥammádiki | wa.mu.ħa.ˈmːa.di.ki | wa.mu.X\a."mːa.di.ki |
| 0109 | وَمُحَمَّدِهِ | wamuḥammádihi | wa.mu.ħa.ˈmːa.di.hi | wa.mu.X\a."mːa.di.hi |

| | | | | |
|---|---|---|---|---|
| 0110 | وَمُحَمَّدِهَا | wamuḥammádihạ | wa.mu.ħa.ˈmːa.di.ha | wa.mu.X\a."mːa.di.ha |
| 0111 | وَمُحَمَّدِنَا | wamuḥammádinạ | wa.mu.ħa.ˈmːa.di.na | wa.mu.X\a."mːa.di.na |
| 0112 | وَمُحَمَّدِكُمْ | wamuḥammádikum | wa.mu.ħa.ˈmːa.di.kum | wa.mu.X\a."mːa.di.kum |
| 0113 | وَمُحَمَّدِكُنَّ | wamuḥammadikúnna | wa.mu.ħa.mːa.di.ˈku.nːa | wa.mu.X\a.mːa.di."ku.nːa |
| 0114 | وَمُحَمَّدِكُمَا | wamuḥammadíkumạ | wa.mu.ħa.mːa.ˈdi.ku.ma | wa.mu.X\a.mːa."di.ku.ma |
| 0115 | وَمُحَمَّدِهِمْ | wamuḥammádihim | wa.mu.ħa.ˈmːa.di.him | wa.mu.X\a."mːa.di.him |
| 0116 | وَمُحَمَّدِهِنَّ | wamuḥammadihínna | wa.mu.ħa.mːa.di.ˈhi.nːa | wa.mu.X\a.mːa.di."hi.nːa |
| 0117 | وَمُحَمَّدِهِمَا | wamuḥammadíhimạ | wa.mu.ħa.mːa.ˈdi.hi.ma | wa.mu.X\a.mːa."di.hi.ma |
| 0118 | وَمُحَمَّدِهِمَا | wamuḥammadíhimạ | wa.mu.ħa.mːa.ˈdi.hi.ma | wa.mu.X\a.mːa."di.hi.ma |
| 0119 | وَمُحَمَّدًا | wamuḥámmadan | wa.mu.ˈħa.mːa.dan | wa.mu."X\a.mːa.dan |
| 0120 | وَمُحَمَّدَ | wamuḥámmada | wa.mu.ˈħa.mːa.da | wa.mu."X\a.mːa.da |