

Intonational Variation in Arabic Corpus Coding System



All filenames are in the format: **XXXX-YYY-ZZ**

where: **XXXX** denotes speaker group
YYY denotes task (or target sentence/type)
ZZ denotes individual speaker

egca-fco-f1.wav

e.g. "egca"

e.g. "fco"

e.g. "f1" (f = female; m = male)

Country/speaker group codes:

<i>code</i>	<i>speaker group</i>	<i>age group (approx.)</i>	<i>recording location</i>
egca	Cairo, Egypt	18-30	Cairo, Egypt
irba	Baghdad, Iraq	18-30	Amman, Jordan
joka	Karak, Jordan	18-30	Karak, Jordan
kwur	Kuwaiti Urban	18-30	Kuwait City, Kuwait
mobi	Moroccan Arabic-Tamazight bilinguals	18-30	Casablanca, Morocco
moca	Casablanca, Morocco (non-Tamazight speakers)	18-30	Casablanca, Morocco
moco	Casablanca, Morocco (non-Tamazight speakers)	40-60	Casablanca, Morocco
ombu	Buraimi, Oman	18-30	Buraimi, Oman
syda	Damascus, Syria	18-30	Amman, Jordan
tuns	Tunis, Tunisia	18-30	Tunis, Tunisia

See <participants.xlsx> for speaker codes and metadata.

Task codes:

<i>code</i>	<i>task</i>	<i>task type</i>
fco	free conversation	unscripted
mp1 + mp2	map tasks (x 2)	unscripted
ret	folktale retold from memory	unscripted
sto	folktale read from script	scripted
sd1 + sd2	role play scripted dialogue (x 2)	scripted

Scripts used with each speaker group for scripted tasks are available as .xlsx files in <scripts.zip>.

Text transcriptions are available for the majority of unscripted audio files. See <text-checklist.xlsx> for a list of which sound files have a transcription file, and what levels of transcription are provided for each file.

Sentence types:

dec	declarative
whq	wh-question
ynq	yes-no question
coo	coordinated question
inf	answer to wh-question
con	answer to yes-no question
idf	answer to coordinated question
voc	vocative

The sentences are extracted from the sd1/sd2 role play scripted dialogue (which is also available in full). Further information is available at: <http://ivar.york.ac.uk/>