User's guide for the Tel-Aviv University audio database of Judeo-Baghdadi Arabic

How to cite:

Berrebi, Si, Eyal Marco, Areej Khourieh, and Ezer Rasin. 2023. *The Tel-Aviv University audio database of Judeo-Baghdadi Arabic*, version 1.0. Online access at: https://pcomplab.github.io/JBAdatabase.html.

Judeo-Baghdadi Arabic (JBA) is an endangered dialect of Arabic, spoken primarily in Israel by Jewish immigrants from Baghdad, Iraq. The dialect is distinct from the Muslim and Christian Arabic varieties spoken in Baghdad. Based on the 2022 yearly report of the Israeli Central Bureau of Statistics, almost 99% of the 42,200 immigrants from Iraq living in Israel today are over 65 years old.[1] The vast majority of second generation immigrants have not acquired the dialect, which is therefore unfortunately expected to go extinct in the decades to come.

Certain aspects of the dialect have previously been described by Blanc (1964), Mansour (1991), and Bar-Moshe (2019). This is the first public database of transcribed and annotated recordings from this dialect.

The database consists of recordings of a speaker born in Israel in 1950 and exposed primarily to JBA in early childhood. Elicitation sessions were conducted in Hebrew between October 2021 and June 2022. The database includes approximately 2,350 JBA productions, pruned from sessions as single words in isolation or as phrases. Each production is accompanied with the following information, given in columns (from left to right):

1. **Transcription**:
    a. Given in the International Phonetic Alphabet (IPA), with the exception of stress, which is indicated by an acute accent (V́), and the back unrounded open-mid vowel, which is indicated by [o] (rather than the IPA symbol [ɔ]).
    b. Pauses are marked with [P], hesitations/corrections with [M].
    c. Some words may appear more than once with different transcriptions, reflecting within-speaker variation.
    d. Generally, each word (i.e., each separated string in the transcription) carries one stress. Exceptions are *ma* (negation) and the prepositions *ʕal* 'on', and *min* 'from', which may appear with or without stress.
2. **Link to the recording.**
3. **Session number**.

---

4. **Session date**.
5. **Gloss**: An English gloss.
6. **Morphological segmentation**: a segmentation of the word into morphemes (when possible).
7. **Morphological gloss**: Morphological glosses follow the conventions below:

abbrev.	value	category
1	1st person	person
2	2nd person	person
3	3rd person	person
POSS	possessive	possessive determiner, pronoun and pronominal index slot
ACC	accusative	grammatical case relator
M	masculine	gender
F	feminine	gender
SG	singular	number
DU	dual	number
PL	plural	number
/	non-concatenative morphology, e.g., broken plurals	
SGT	singulative	nominal
IMP	imperative	mood
PFV	perfective	aspect
IPFV	imperfective	aspect
PROG	progressive	aspect
DET	determiner	pronominal
NEG	negative	
NEG.IMP	negative (imperative form)	
DO	direct object	pronominal index slot
IO	indirect object	pronominal index slot
CONJ	conjunctive	interpropositional relation

abbrev.	value	category
DISJ	disjunction	interpropositional relation
EXIST	existential	grammatical verb
CARD	cardinal	numeral
ORD	ordinal	numeral
RED	reduplicative	language game
PASS	passive	voice
DEM	demonstrative	determination
APTCP	active participle	
PPTCP	passive participle	
VOC	vocative	
DOBJ	double-object	
[ALL CAPS]	recent loanword/code switch	loanword
F.link	construct-state linking morpheme	
V, VI, VII, X	verbal template prefixes	
<VIII>	verbal template infixes	

In cases of syncretism in affixes (e.g., the prefix *t-* stands for both '2msg' and '3fsg'), the gloss corresponds to the intended meaning provided by the speaker.

8. **Syllable structure**.

Verbs are classified with respect to two additional properties:

9. **Binyan** (also *Wazan*; plural *Binyanim*) – the vocalic and prosodic template characteristic of Semitic verbal morphology, which may include affixes. In Judeo-Baghdadi Arabic, nine of the traditional Arabic Binyanim are used. Verbs are classified on the basis of the following, traditional enumeration of Binyanim (C stands for a consonant slot):

I. CaCaC

II. CaCCaC

III. CaaCaC[2]

IV. Not used in JBA

V. t-CaCCaC

VI. t-CaaCaC

VII. n-CaCaC

VIII. C<t>aCaC

IX. CCaCC - rare, and is not part of the published portion of the database yet

X. sta-CCaC

10. **Gizra** (plural Gzarot): the type of the root according to the traditional analysis of Semitic languages:
    a. "Strong triradical" (XYZ) roots: roots with three stable consonants that are not deleted across the paradigm. For example, [**katab**] 'he wrote'.
    b. Quadriradical roots: roots with four consonants. For example, [**lámlam**] 'he collected'.
    c. "Geminated" (XYY) roots, in which the second and third root consonants are identical and typically surface as a double consonant. For example, [**ħábbu**] 'they loved'.
    d. Glide-first roots (jXY; sometimes referred to as "weak" or "assimilated". For example, [n**wálad**] 'he was born'.
    e. Glide-second roots (XjY; sometimes referred to as "hollow"). For example, [**tˤaːɣ**] 'he flew'.
    f. Glide-third roots (XYj; sometimes referred to as "defective" or "lame"). For example, [**tˤáʕa**] 'he gave'.

Some roots belong to two root-types at the same time. For example, [**d͡ʒá**] 'he came' is classified as both glide-second and glide-third (XjY and XYj).

The corpus will be updated and corrected periodically. For questions, comments, or to report any errors you come across, please contact us at phoncomplab@gmail.com.

---

[2] Some verbs in this Binyan have different vowels. For example, [sóːləf] 'he chatted' and [béːrək] 'he blessed'.

References

Bar-Moshe, Assaf. 2019. *The Arabic Dialect of the Jews of Baghdad: Phonology, Morphology, and Texts.* Wiesbaden: Harrassowitz.

Blanc, Haim. 1964. *Communal dialects in Baghdad*. Cambridge: Harvard University Press.

Mansour, Jacob. 1991. *The Jewish Baghdadi dialect: studies and texts in the Judaeo-Arabic dialect of Baghdad.* Studies in the history and culture of Iraqi Jewry, 7.