

# CJKI Arabic Romanization System

## CARS

Jack Halpern <jack@cjki.org>  
The CJK Dictionary Institute, Inc.

### Summary

## 1. Introduction

### 1.1 Overview

The **CJKI Arabic Romanization System**, or **CARS**, is an innovative phonemic transcription system developed mainly for ease of use by learners and as an aid to linguists in analyzing the phonological structure of Modern Standard Arabic (MSA). The most important feature of CARS is its readability: its unambiguous encoding of phonological and prosodic information in an intuitive, easy to read set of symbols. The new system has several unique features not found elsewhere, especially the indication of word stress and vowel neutralization. Various other features distinguish CARS as an easy-to-use, highly accurate phonemic transcription system.

### 1.2 Orthographic Ambiguity

As is well known, unvocalized Arabic is highly ambiguous (كاتب can be pronounced in seven ways). But even fully vocalized is ambiguous because some letters are not pronounced:

1. أ ('*alif* 'waṣla) in مِنَ الْمُمْكِنِ *mina l-múmkini* 'possible' (rarely shown)
2. ل (*lām*) in assimilated articles as in اَلدَّخْلُ '*addákhlu* 'income'
3. final long vowels, as in أَنَا '*ána* 'I', هَذَا '*hádhā* 'this' (final *ā* is shortened)
4. final double consonants, as in حُبْ '*hub* 'love' (final double /b/ becomes single)
5. '*alif* in certain verb forms, as in كَتَبُوا '*kátabu* 'they wrote'

Word stress and vowel neutralization, both of great importance to correct pronunciation, are of course not shown at all. Looking at a word like كَتَبَ '*kataba* 'he wrote', there is no way to determine which syllable is stressed without knowing complex rules and exceptions. Neither is there a way to know that the final long vowel in words like هَذَا '*hádhā* 'this' and عَلَى '*éala* 'on' is neutralized (shortened).

### 1.3 Why romanization

Since one of the primary goals of CARS is to aid learners, let us pose a fundamental question: do learners of Arabic even need a romanization system? The answer is a definitive yes. Though some educators may argue otherwise, the ambiguity of even fully vocalized Arabic and the need to convey stress and neutralization indicate that romanization is indeed necessary. Even professional teachers wrongly assume that vocalized Arabic is *sufficient* for pronouncing Arabic correctly, a common but unfortunate misconception based on the unconscious knowledge that native speakers have of orthographic rules.

## 2. Description of System

The **CJKI Arabic Romanization System**, or **CARS**, is an innovative phonemic transcription system developed mainly for ease of use by learners. The most important feature of CARS is its readability: its unambiguous encoding of phonological and prosodic information in an intuitive, easy to read set of symbols. The new system, which is making its debut with the appearance of **The CJKI Arabic Learner's Dictionary**, includes a set of symbols that represent all Arabic phonemes unambiguously, and, for the first time, symbols indicating word stress and vowel neutralization.

Here is a quick glimpse of how CARS works:

الْحُكُومَةُ الْيَابَانِيَّةُ 'alḥukūmatu lyabāniyyatu the Japanese government

Long vowels, as in *kū*, are shown by a macron over the vowel. Word stress is indicated by the accent mark, as in *nīy* and *kū* above. For the first time in any system, neutralized long vowels are shown by a macron below the vowel (as in *a* above) to remind the reader that they are written, but *not* pronounced, as long vowels. The undertie (˘) is used as a liaison before words beginning with 'alif waṣla (ī).

CARS is an up-to-date, scientific transcription system that aims to be easy to learn. It comes in three flavors:

1. The **Standard CARS** symbol set aims to (1) unambiguously represent the Arabic phonemes, (2) accurately indicate word stress, and (3) explicitly indicate which vowels are shortened and which consonants are undoubled (*neutralization*).
2. **Extended CARS** adds three useful optional features by indicating case endings, velarized /a/, and syllable boundaries (see Section 4)
3. **Proxy CARS** is a set of substitute symbols to facilitate input from a standard ASCII keyboard (see Section 5).

CJKI offers a suite of tools (**CARS Tools**) to automatically convert Proxy CARS to Standard and Extended CARS, as well as tools to automatically convert from Arabic script to any of the three varieties of CARS. These tools are available for free download from the CJKI website.

Detailed explanations of CARS, word stress and neutralization can be found at:

1. Detailed white paper on CARS  
<http://www.cjk.org/cjk/arabic/cars.doc>
2. Academic paper on CARS  
[http://www.cjk.org/cjk/arabic/cars\\_paper.doc](http://www.cjk.org/cjk/arabic/cars_paper.doc)
3. Word Stress and Vowel Neutralization in Modern Standard Arabic  
<http://www.cjk.org/cjk/reference/index.htm>

### 3. Standard CARS

**3.1 CARS Symbols:** The table below shows the special symbols (not letters) used in CARS. **Proxy symbols** are shown in brackets.

Standard CARS Symbols

Symbol	Symbol Name	Proxy Symbol	Description
–	macron	<i>vv</i>	The <b>macron</b> ( <i>ā ī ū</i> ) [ <i>aa ii uu</i> ] indicates a long vowel, as in سَافَرْتُ <i>sāfartu</i> [ <i>saafa/rtu</i> ].
ˈ	acute accent	/	The <b>acute accent</b> ( <i>á í ú</i> ) [ <i>a/ i/ u/</i> ] represents a stressed syllable, as in أَأَنْجَ <i>'ánq</i> [ <i>'a/na_</i> ].
ˆ	macron + acute accent	<i>vv/</i>	The <b>macron plus acute accent</b> ( <i>á í ú</i> ) [ <i>aa/ ii/ uu/</i> ] indicates a long stressed vowel, as in أَأَسْرَرْتُ <i>'aṣṣarráfu</i> [ <i>'aSSarraa/fu</i> ].
–	macron below	<i>x_</i>	The <b>macron below</b> ( <i>a i u b d etc.</i> ) [ <i>a_ i_ u_ b_ d_ etc.</i> ] represents neutralization: either a long vowel that is shortened, as in أَأَنْجَ <i>ána</i> [ <i>'a/na_</i> ], or a double consonant that is undoubled, as in حُبَّ <i>hub</i> [ <i>Hub_</i> ].
.	underdot	<i>caps</i>	The <b>underdot</b> modifies certain consonants to represent others, especially the emphatics: i.e., ص <i>ṣ</i> , ذ <i>ḏ</i> , ط <i>ṭ</i> , ظ <i>ẓ</i> and ح <i>ḥ</i> . <b>Capitals</b> can be used as a proxy, as in [ <i>S, D, T, Z, H</i> ].
˘	undertie	~	The <b>undertie</b> indicates liaison between words resulting from the omission of 'alif waṣla (إ) in word initial position, as in فِي الصَّيْنِ <i>fī ṣṣīni</i> [ <i>fi_~Ssii/ni</i> ].
'	apostrophe	<i>none</i>	The <b>apostrophe</b> represents the <i>hamza</i> in all its written forms (أ، إ، ؤ، ء، ئ، آ) Thus أَ and آ, إ are represented by 'a, 'i and 'u, as in أَخَذَ <i>'ákhadha</i> [ <i>'a/khadha</i> ].
ε	epsilon	<b>E</b>	The Greek letter <b>epsilon</b> represents the letter <i>ḥayin</i> (ح), as in سَعِيدٌ <i>saēid</i> [ <i>saEiid</i> ].
˘	breve	^	The <b>breve</b> disjoins the letters <i>t, s, d</i> and <i>k</i> from a following <i>h</i> to show that they are distinct letters when not separated by a vowel; that is, <i>t<sup>h</sup>, s<sup>h</sup>, d<sup>h</sup></i> and <i>k<sup>h</sup></i> represent <i>ته, سه, ده</i> and <i>كه</i> , as opposed to <i>th, sh, dh</i> and <i>kh</i> , which represent <i>ث thā', س sīn, ذ dhāl,</i> and <i>خ khā'</i> . A <b>circumflex</b> can be used as proxy, as in [ <i>t^h s^h d^h k^h</i> ].

**3.2 Vowels and Diphthongs.** Short and long vowels are represented as shown below.

Short Vowels		
CARS	Proxy	Description
a	a	short /a/
i	i	short /i/
u	u	short /u/
á	a/	stressed short /a/
í	i/	stressed short /i/
ú	u/	stressed short /u/

Long Vowels		
CARS	Proxy	Description
ā	aa	long /a/
ī	ii	long /i/
ū	uu	long /u/
á	aa/	stressed long /a/
í	ii/	stressed long /i/
ú	uu/	stressed long /u/

The diphthongs are written as *ay* and *aw*, as in *بَيْت bayt* ‘house’ and *يَوْم yawm* ‘day’, while the glides are written as *iy* and *uw*, as in *عَرَبِيَّة erabīyya* ‘Arab’ and *فُتُوَّة futúwwa* ‘youth’. Nunation, *tā' marbūṭa* (ة), *'alif maqṣūra* (ى) and various other special vowels are represented strictly as they are pronounced, as in *بُيُوتٌ buyūtun*, *عَرَبِيَّة erabīyya* and *عَلَى ʿalā*.

**3.3 Consonants.** The consonants are represented by the following symbols, mostly following convention.

#### Plain Consonants

ب	b	bā'
ت	t	tā'
ج	j	jīm
د	d	dāl
ر	r	rā'
ز	z	zāy

س	s	sīn
ع	ε	εayin
ف	f	fā'
ق	q	qāf
ك	k	kāf
ل	l	lām

م	m	mīm
ن	n	nūn
ه	h	hā'
و	w	wāw
ي	y	yā'
ء	'	hamza

#### Underdotted Letters and Diagraphs

ص	ṣ	ṣād
ض	ḍ	ḍād
ط	ṭ	ṭā'
ظ	ẓ	ẓā'
ح	ḥ	ḥā'

ث	th	thā'
خ	kh	khā'
ذ	dh	dhāl
ش	sh	shīn
غ	gh	ghayn

*Shadda* is represented by duplicating the consonant that bears the *shadda*, as in *مُحَمَّد muḥammad*, written with a double *m*.

- 3.4 Consonant Assimilation.** When the article *أل* precedes a sun letter the assimilated pronunciation is shown explicitly by doubling the letter, as in *أَلصَّغِيرِ* 'aṣṣaghír.
- 3.5 Hamza.** The *hamza* (أ إ إِيَّ آ آءِ ؕ ؖ ؗ) is represented by 'a, 'i and 'u, and ' as in *أَخَذَ* 'ákhadha. The hamza is also shown at the beginning of words to remind the reader that it must always be pronounced.
- 3.6 Word Stress.** A major innovation of CARS is the use of an acute accent to indicate stress, as in *يَعْمَلُ* *yáemalu* 'he works'. Stress is of great importance to learners, as there is no way to predict it unless one knows a set of complex rules.
- 3.7 Neutralization.** Another major innovation of CARS is the indication of which vowels and consonants are *neutralized* (shortened) in actual pronunciation by placing a macron below the neutralized letter, as in *هَذَا* *hádhā* 'this' and *حُبٌّ* *hub̄* 'love'. Neutralization is a common phenomenon but ignored by Arabic teachers and dictionaries, misleading one to believe all long vowels are pronounced long.
- 3.8 Liaison and digraph separation** are explained in the table on page 3.
- 3.9** The Greek letter **epsilon** ε represents *ع* (as in *سَعِيدٌ* *saeīd*), to which it is similar, instead of the often confusing different types of apostrophes or such symbols as 3.

#### 4. Extended CARS

CARS is enhanced by several *optional* features referred to as **Extended CARS**. This consists of three symbols to explicitly indicate case endings, velarized /a/, and syllable boundaries. These are useful for pedagogical and lexicographic applications. **Proxy symbols** are shown in brackets

Extended CARS Symbols

Symbol	Symbol Name	Proxy Symbol	Description
	vertical bar	()	The <b>vertical bar</b> is an optional symbol that indicates case endings, as in <i>الصَّرَافُ</i> 'aṣṣarráf u ['ASSArrAA/f(u)], where the " u" indicates the beginning of a nominative case ending.
ḍ	turned alpha	Α	The <b>turned alpha</b> is an optional symbol that represents a velarized /a/ (similar to the vowels in <i>car</i> or <i>saw</i> ), as in <i>الصَّرَافُ</i> 'ḍṣṣarráf u ['ASSArrAA/fu]. This vowel can be short ( <i>v</i> ), long ( <i>v̄</i> ), stressed ( <i>v́</i> ), neutralized ( <i>v̇</i> ) or both long and stressed ( <i>v̄́</i> ).
.	middle dot	-	The <b>middle dot</b> is an optional symbol that indicates syllabification, as in <i>السَّرَافَةُ</i> 'aṣ-ṣarráf u ['AS-SAr-rAA/-fu], used on special occasions to indicate syllable boundaries.

Normally, it is not necessary to use the options provided by Extended CARS, but they can be used freely as necessary. Using them all together, as in

الصَّرَافُ الآلِيَّ 'aṣṣarrafu l'ālyy|u ATM

makes a CARS string look a bit intimidating because of the various symbols. In Standard CARS, this is written 'aṣṣarrāfu l'ālyy, which is more friendly.

## 5. Proxy CARS

**Proxy CARS** is a set of substitute symbols to facilitate input from a standard ASCII keyboard. These **proxy symbols** makes it easy to input, search and manipulate the symbols used in Standard CARS and Extended CARS. Proxy CARS is easy to write and hard to read, whereas Standard CARS is easy to read and hard to write.

**Extended CARS Symbols**

Symbol Example	Symbol Name	Proxy Example	Proxy Name	Function
<i>ā</i>	macron	<i>aa</i>	double vowel	long vowels
<i>á</i>	acute accent	<i>a/</i>	slash	stressed vowels
<i>ǎ</i>	macron + acute accent	<i>aa/</i>	double vowel + slash	long stressed vowels
<i>a̲</i>	macron below	<i>a_</i>	underscore	neutralization
<i>ş</i>	underdot	<i>S</i>	capital letter	emphatics etc.
<i>a_b</i>	undertie	<i>a~b</i>	tilde	liaison
<i>ḍh</i>	breve	<i>d^h</i>	circumflex	digraph disjunction
<i>ɖ</i>	turned alpha	<i>A</i>	capital A	velarization
<i>ε</i>	epsilon	<i>E</i>	capital E	<i>εayin</i>
<i>la·ka</i>	middle dot	<i>la-ka</i>	hyphen	syllabification

## 6. Sample Text

<b>Vocalized Arabic</b>	مُنْدُ مِئَاتِ السِّنِينَ، كَانَتْ تَعِيشُ فِي الْأَصِينِ أَرْمَلَةً فَاقِيرَةً مَعَ ابْنِهَا الصَّغِيرِ عَلَاءِ الدِّينِ.
<b>Standard CARS</b>	múndhu mi'áti <sub>ssinína</sub> , kánat taéishu fi <sub>ṣṣíni</sub> 'armálatun faqíratun máea <sub>bniha</sub> ṣṣaghíri ealá'i <sub>ddíni</sub> .
<b>Proxy CARS</b>	mu/ndhu mi'aa/ti~ssinii/na, kaa/nat taEii/shu fi_~SSii/ni 'arma/latun faqii/ratun ma/Ea~bniha_~SSaghii/ri Ealaa/'i~ddii/ni.
<b>Extended CARS</b>	múndhu mi'áti <sub>ssinín</sub>  a, kánat taéishu fi <sub>ṣṣín</sub>  i 'armála tun foqíra tun máea <sub>bniha</sub> ṣṣoghír i ealá'i <sub>ddín</sub>  i.
<b>English</b>	Hundreds of years ago, there lived in China a poor widow with her young son, Aladdin.