

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

by Donald Mastronarde, Center for the Tebtunis Papyri, University of California, Berkeley; prepared with the assistance of Todd Hickey, Brigit Flannery, and Jean Li.

Revised September 2009, with updates, and addition of Appendix on Arabic transliteration for papyrologists; please communicate comments, corrections, additions to djmastronarde@berkeley.edu. Additions December 2019 for added Unicode characters in New Athena Unicode version 5.006.

A set of Latin characters, some with diacritics, and a few additional symbols have been used in scholarly publications to represent ancient Egyptian texts in transliteration. The most common use of such transliteration is in the presentation of Demotic Egyptian texts from the Hellenistic and Roman eras, but a subset of this group is occasionally used to present texts written in Hieroglyphs.

Specialized fonts with ad hoc encodings have been used by scholars in their word-processing and publication. It is important to replace the previous encodings with Unicode, the international standard for encoding characters and symbols in all the world's writing systems.

There are currently two slightly different systems for transliteration, and the European and the traditional (acc. to James P. Allen, *Middle Egyptian*). Characters from both traditions are listed here.

The standard roman characters without diacritics are not discussed here, since they are already present in the Basic Latin block of Unicode. Those used in transliteration are: j, y, w, b, p, f, m, n, r, h, z, s, q, k, g, t, d.

Note that not all characters that are used in lowercase are to be found in these transliteration systems in uppercase. Capitals are used only in proper names, and some characters represent sounds that do not occur at the opening of a word.

Most of the characters with diacritics already have a code point for the precomposed form in Unicode. A few that do not can be handled either with Private Use Area code points (as originally done in a 2005 version of the customized input) or by using multiple code points for the decomposed form and allowing a good display to be handled by a font with OpenType ligatures. The latter method requires the use of applications that handle OpenType ligatures (in Mac OS X that means Word 2008 and not earlier versions, and Firefox 3 and not earlier versions and Safari 4 for OS X [but Safari 4's support for ligatures is only partial as of September 2009]).

Character name	example	Unicode	remarks
Latin small letter a with macron	ā	U+0101	used in older publications
Latin small letter c with caron	č	U+010d	used in German system(?)

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

Latin small letter c with caron and dot below	č̣	U+010d, U+0323 (or U+0063, U+030c, U+0323)	used in German system; PUA U+ec41 in New Athena Unicode font
Latin capital letter d with line below	Ḑ	U+1e0e	
Latin small letter d with line below	ḑ	U+1e0f	
Latin small letter g with caron	ǧ	U+01e7	
Latin capital letter H with dot below	Ḥ	U+1e24	
Latin small letter h with dot below (= dotted h)	ḥ	U+1e25	
Latin capital letter H with line below (= fourth h)	Ḧ	U+0048, U+0331	PUA U+ec40 in New Athena Unicode font
Latin small letter h with line below (= fourth h)	ḧ	U+1e96	
Latin capital letter H with breve below (= rounded h or third h)	Ḩ	U+1e2a	
Latin small letter h with breve below (= rounded h or third h)	ḩ	U+1e2b	
Latin capital letter H with circumflex below	Ḫ	U+0048, U+032d	PUA U+ec48 in New Athena Unicode font
Latin small letter H with circumflex below	ḫ	U+0068, U+032d	PUA U+ec47 in New Athena Unicode font
Latin capital letter Egyptological yod [or Latin capital letter I with Egyptological yod modifier]	ꜥ	U+0049, U+0357 or U+a7bc	See separate note below.
Latin small letter Egyptological yod [or Latin small letter I with Egyptological yod modifier]	ꜥ	U+0069, U+0357 or U+a7bd	See separate note below.
Latin small letter i with reversed breve below	ḭ	U+0069, U+032f	used in verb endings in some systems (temporarily PUA U+ec44 in New Athena Unicode font)
Latin capital letter K with dot below	Ḳ	U+1e32	

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

Latin small letter k with dot below	ḱ	U+1e33	
Latin small letter o with macron	ō	U+014d	used in older publications
Latin capital letter S with acute	Š	U+015a	used for transliteration of Hieroglyphic, but not Demotic, in traditional system
Latin small letter S with acute	ś	U+015b	used for transliteration of Hieroglyphic, but not Demotic, in traditional system
Latin capital letter S with caron (= shin)	Š	U+0160	
Latin small letter s with caron (= shin)	š	U+0161	
Latin capital letter T with line below	Ṭ	U+1e6e	
Latin small letter t with line below	ṭ	U+1e6f	
Latin small letter t with circumflex below	ṡ	U+1e71	
Latin small letter u with macron	ū	U+016b	
Latin small letter u with reversed breve below	ṹ	U+0075, U+032f	PUA U+ec45 in New Athena Unicode font
Latin capital letter Egyptological alef	ʾ	U+a722	Unicode 5.1
Latin small letter Egyptological alef	ʾ	U+a723	Unicode 5.1; formerly PUA U+ec42 in New Athena Unicode font
Latin capital letter Egyptological ayin	ʿ	U+a724	Unicode 5.1
Latin small letter Egyptological ayin	ʿ	U+a725	Unicode 5.1; formerly PUA U+ec43 in New Athena Unicode font

Punctuation frequently used in these transliteration systems:

Character name	example	Unicode	remarks
Hyphen	-	U+002d	

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

Double hyphen (equals sign)	=	U+003d	mark inserted before suffixes (alternative: double oblique hyphen)
Double oblique hyphen	≡	U+2e17	mark inserted before suffixes (alternative: double hyphen)
Dot	.	U+002e	written within word after determinative (alternative: middle dot)
Middle dot	·	U+00b7	written within word after determinative (alternative: dot)
Left corner bracket	⌈	U+2308	to be replaced in time by new U+2e42
Right corner bracket	⌋	U+2309	to be replaced in time by new U+2e43

NOTE on Egyptological yod:

Unicode has refused to approve a separate character for the yod modifier used with the letter I/i, and is never likely to do so. The Unicode Technical Committee (UTC) has recommended the use of an existing modifier within a OpenType ligature (the modifier may of course be drawn in some particular desired comparable shape when a precomposed glyph is created). Originally, UTC suggested that the scholarly community of experts choose between U+0313 COMBINING COMMA ABOVE [´] (this is the diacritic used for the smooth breathing of polytonic Greek, among other things) or U+0357 COMBINING RIGHT HALF RING ABOVE [ˆ]. No consensus emerged, however, and a third option has been proposed, U+0486 COMBINING CYRILLIC SMOOTH BREATHING. As of 2009 the UTC accepts that there are three options and still leaves it to the scholarly community to arrive at a consensus, if they wish to. The original Demotic input provided by the Center for the Tebtunis Papyri used a PUA code point. The 2009 version adopted the U+0357 code point for the modifier, but supplies the same precomposed glyph by ligature for I/i followed by either of the other two options just mentioned. **In version 3, 2019, the keyboard enters the newly-approved code points of Latin Capital Letter Glottal I and Latin Small Letter Glottal I.**

Sources:

Janet H. Johnson, *Thus Wrote ʿOnchsheshonqy: An Introductory Grammar of Demotic*, 3rd ed. [Studies in Ancient Oriental Civilization, no. 45] Chicago, pp. 1-8

James P. Allen, *Middle Egyptian: An Introduction to the Language and Culture of Hieroglyphs*, Cambridge 2000, pp. 14-19

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

and sample scholarly publications.

APPENDIX 1: Arabic transliteration for papyrologists

Systems of transliteration into roman characters for Arabic also rely on the use of various diacritics. Many of the combined characters used in standard systems are the same as are used in Demotic Egyptian transliteration, and New Athena Unicode has been revised (from version 3.5) to include the remainder.

For the transliteration equivalents, see the listing provided by IFAO at <http://www.ifao.egnet.net/publications/outils/polices/> or the wikipedia article at http://en.wikipedia.org/wiki/Romanization_of_Arabic.

Character name	example	Unicode	remarks
FROM IFAO LISTING			
Latin small letter a with macron	ā	U+0101	
Modifier right half ring = hamza	’	U+02be	added to NAU from version 3.5
Latin small letter t with line below	ṭ	U+1e6f	
Latin small letter g with caron	ǧ	U+01e7	
Latin small letter h with dot below	ḥ	U+1e25	
Latin small letter h with breve below	ḥ̆	U+1e2b	
Latin small letter d with line below	ḏ	U+1e0f	
Latin small letter s with caron (= shin)	š	U+0161	
Latin small letter s with dot below	ṣ	U+1e63	added to NAU from version 3.5
Latin small letter d with dot below	ḍ	U+1e0d	added to NAU from version 3.5
Latin small letter t with dot below	ṭ̣	U+1e6d	added to NAU from version 3.5
Latin small letter z with dot below	ẏ	U+1e93	added to NAU from version 3.5
Modifier left ring = ayin	‘	U+02bf	added to NAU from version 3.5

Representing Transliteration of Ancient Egyptian (and Arabic) in Unicode

Latin small letter g with dot above	ġ	U+0121	
Latin small letter u with macron	ū	U+016b	
Latin small letter I with macron	ī	U+012b	
ADDITIONAL CHARACTERS USED IN OTHER SYSTEMS			
Latin small letter h with line below	ḥ	U+1e96	
Latin small letter g with macron	ḡ	U+1e21	
Latin small letter t with diaeresis	ṫ	U+0074, U+0308	precomposed glyph and OpenType ligature instruction added in NAU 3.5
Latin small letter y with grave	ÿ	U+1ef3	

APPENDIX 2: four additional characters for Ugaritic transliteration

In Unicode version 12, the following characters have been added along with the glottal i forms referred to above:

Latin Capital Letter Glottal A	U+a7ba	shift-option q
Latin Small Letter Glottal A	U+a7bb	shift-option r
Latin Capital Letter Glottal U	U+a7be	shift-option o
Latin Small Letter Glottal U	U+a7bf	shift-option 0 [zero]