

Universal Multiple-Octet Coded Character Set  
 International Organization for Standardization  
 Organisation Internationale de Normalisation  
 Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Final proposal to encode Anatolian Hieroglyphs in the SMP of the UCS****Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)****Authors: Michael Everson and Deborah Anderson****Status: Liaison Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Replaces: N3236R (2007-05-01), N4264 (2012-05-02)****Date: 2012-07-15**

**1. Introduction.** Anatolian Hieroglyphs were first thought to have been used for the Hittite language, because they first appeared on personal seals from Hattusha, the capital of the Hittite Empire. While Hittites did make use of the characters on seals and on their monumental inscriptions, the characters were used as text primarily for the related language Luwian; a few glosses in Urartian and some divine names in Hurrian are known to be written in Anatolian Hieroglyphs. Most of the texts are monumental stone inscriptions, though some letters and accounting documents have been preserved inscribed on strips of lead. (See Figure 1.)




















**Figure 1.** Anatolia, showing where the Luwian, Hittite, Urartian, and Hurrian languages were spoken.


**2. Structure.** Anatolian Hieroglyphs have both syllabic and logographic components. Words may be represented entirely with logographs (in particular at the earlier period), or with phonetic complements along with the logogram in a variety of configurations. For the word *wawis* ‘cow’ for instance, the following spellings occur:



















 BOS     
     (BOS)*wa/i-wa/i-sa*     
    *wa/i-wa/i-sa*     
    BOS-*wa/i-sa*




Characters may be written left-to-right or right-to-left, and texts longer than two lines are mostly boustrophedon; as with Egyptian, one reads a line into the faces of the people and animals. Lines are often divided from one another by horizontal rules. Within a line, characters are written in vertical columns, mostly from top to bottom, though for aesthetic reasons they may be placed out of phonetic or logical order. For modern lexicographical use it is anticipated that straightforward linear left-to-right placement of characters will find the most favour for its practicality. (See Figure 1.)










**3. Punctuation and spacing.** In some texts, word division is indicated with the use of U+145B5  ANATOLIAN HIEROGLYPH A386 or its variant U+145B6  ANATOLIAN HIEROGLYPH A386A. To mark logograms, the characters U+145CE  ANATOLIAN HIEROGLYPH A410 BEGIN LOGOGRAM MARK and U+145CF  ANATOLIAN HIEROGLYPH A410A END LOGOGRAM MARK are optionally used, flanking the logogram as in   ‘wood’; see Figure 2. (In Marazzi and Laroche A410 is given as a single character but it has been split into two in order to permit the two characters to flank logograms productively.) The characters U+145F7  ANATOLIAN HIEROGLYPH A450 (which occurs very frequently) and U+144EF  ANATOLIAN HIEROGLYPH A209 (which occurs much more rarely) are occasionally used in what looks like a “space-filler” function, often at the end of a word. Often when hieroglyphs are written, the practice is to try to reproduce the layout of the original text, rather than to write text out in modern sentences. However, in such cases, U+200B ZERO-WIDTH SPACE and/or U+0020 SPACE may be used as needed. One combining character is used, namely, U+145B1  ANATOLIAN HIEROGLYPH COMBINING MARK A383 which attaches fairly productively to Anatolian characters (it always touches its base character). A similar character, U+145B2  ANATOLIAN HIEROGLYPH A383A is used as a personal determinative; it is a spacing character.



**4. Repertoire.** The repertoire in this proposal differs from that of N4264 in that it adds 22 characters, which are highlighted in yellow in the code chart below. Through the course of the history of the decipherment of Anatolian Hieroglyphs, a standardized catalogue of letters has been drawn up; the “Laroche set” of numbers 1–497 from Laroche 1960 form the basis for the repertoire, supplemented by additions 501–524 from Hawkins and Çambel 2000. The encoding proposed here is based on this standardized catalogue, despite the fact that several of the characters are now considered to be variants of each other. U+144D2  A183 is now known to be a variant of U+145DC  A423 which has a reading *ku*; U+1458D  A352, U+1458E  A353, U+1458F  A354, and (possibly) U+1463B  A519 are all variants of U+14586  A345 (a pitcher). Although this might *seem* to be “character duplication”, really it is not. The encoding of these characters will enable not only the representation of text in the Luwian language, but also the representation of the long history of Anatolian studies, where the catalogue entities have been distinguished in discussions of the decipherment. Scholars wishing to publish normalized texts will certainly wish to avoid the use of the “redundant” characters, or might choose to use them as indicative of the temporal or geographical provenance of a text. But documents relating to the decipherment itself distinguish them regularly, and that distinction must be maintained. The names list contains cross references from variants to the main character in each instance. This is considered vital information by specialists.

The characters and glyphs in this proposal largely reflect those found in Marazzi 1998, with some additions from Hawkins and Çambel 2000 (particularly pp. 617–623), and Herbordt 2005 (pp. 399-425). When a glyph was not included in Marazzi (such as \*11), the shape is usually taken from Laroche (as U+1440B  ANATOLIAN HIEROGLYPH A011). Some glyphs are amalgams of various sources, particularly where there are strong similarities between characters. A few characters may have different shape when used in its syllabic value (which often appears with a simplified glyph) *vis à vis* its logographic value. These have been encoded separately, given the semantic distinction, in “monumental” and a “cursive” forms:

U+1442B		A041	U+1442C		A041A
U+14448		A066	U+14449		A066A
U+1446F		A100	U+14470		A100A
U+14476		A104	U+14477		A104A
U+1447A		A105	U+1447B		A105A
U+1447E		A107	U+1447F		A107A
U+14484		A110	U+14485		A110A
U+144F6		A215	U+144F7		A215A
U+1457A		A336	U+1457A		A336A








A few characters are considered separate text elements in Marazzi, but were not assigned separate catalog numbers by Laroche. These have nevertheless been separately proposed for encoding and are distinguished usually by an “A” or “B” after the first element’s character number. For example, U+145F8  ANATOLIAN HIEROGLYPH A450A represents *a+ra/i* or *ra+a* and is composed of U+145F7  ANATOLIAN HIEROGLYPH A450 (*a*) + U+145B1  ANATOLIAN HIEROGLYPH A383. This is an extremely productive combining character which never stands alone, but which always joins to its base character, either following it or being attached below it. (Depending on the directionality of a line of text, “following it” may appear to be either on the left side (RTL) or the right side (LTR); see Figures 2, 6, and 7.) Since the encoding for Anatolian Hieroglyphs is based on the Laroche catalogue, and since three of its characters are formed of a base character plus the combining *ra/i*, canonical decompositions have been given in order to prevent multiple spellings:


U+144F0  A209A = U+144EF  A209 + U+145B1  ANATOLIAN HIEROGLYPH A383 COMBINING RA OR RI  
 U+145B9  A389 = U+145B8  A388 THREE + U+145B1  ANATOLIAN HIEROGLYPH A383 COMBINING RA OR RI  
 U+145F8  A450a = U+145F7  A450 + U+145B1  ANATOLIAN HIEROGLYPH A383 COMBINING RA OR RI

U+1457C  ANATOLIAN HIEROGLYPH A336B (deriving from  A378 +  A336) is an example of a complex character without a decomposition.




**5. Character names.** The character names for Anatolian hieroglyphs are somewhat complex. A single coherent set of character names does not exist, but it has been possible to devise one in working out this proposal.













**5.1 The name of the script.** As suggested in the introduction above, the names “Anatolian Hieroglyphs”, “Luwian Hieroglyphs”, and “Luvian Hieroglyphs” might be applied to this script; the name “Hieroglyphic Hittite” is a misnomer for the language, and “Hieroglyphic Luwian” refers to a variety of the Luwian language (as opposed to “Cuneiform Luwian”). Since the script is used for languages other than Luwian, and since that language may be spelled “Luwian” or “Luvian”, specialists have suggested that the name “Anatolian Hieroglyphs” is the best choice, even if this is clearly a syllabic system and “hieroglyphic” has often been used for a pictorial/logographic system.



**5.2 The use of catalogue numbers.** Catalogue numbers were assigned by Meriggi and Güterbock but these were superseded by the numbers 1 to 497 which Laroche assigned and published in 1960 in *Les hiéroglyphs hittites*. The Laroche numbers have been taken as standard by subsequent researchers; Hawkins and Çambel 2000 augment this list with their numbers 501 to 523 and additional characters are listed in Herbordt 2005. In a number of instances, Laroche lists variants within a given catalogue number, thus: Laroche  457 (1) is distinguished from Laroche  457 (2). The naming convention followed here for this is similar to that used by Egyptologists for Egyptian characters: the first is Laroche 457, the second Laroche 457A. (Not all of the variants listed in Laroche 1960 are distinguished in this way; Laroche  460 (1) and  460 (2) are simply glyph variants and only a single 460 is proposed for encoding.) There is one exception: the set Anatolian Hieroglyphs U+14574  A332A, U+14575  A332B, and U+14576  A332C maintain the names with “A”, “B” and “C” without any character named Anatolian Hieroglyph \*A332, because Marazzi had already identified with the names 332A, 332B, and 332C, and changing the pattern for \*332 could be confusing to users.

**5.2.1 Recommendations for expansion.** For future expansion based on newly-discovered characters, Anatolianists have two options: If the character is clearly based on an existing character, its catalogue number could be based on that. If the complete body of a horse were found, for instance, it could be given a catalogue number based on U+1446E  A099, such as \*99a. Other new signs could be added in the 500 series. It can be recommended that in scholarly publications the A- numbers be used for encoded

characters, and L- or H- or \*- numbers be used in publications for new discoveries which have not been standardized in the UCS.

**5.3 Annotations.** Scholars traditionally use Latin names to describe the characters, and in some of them the Latin is supplemented with full stops and plus signs in order to assist identification. The full stop indicates a group in which the elements are separate (as in A107B  BOS.MI) and the plus sign indicates that the elements form a graphic unit (as in A107  BOS+MI). Informative notes in the names list present these Latin glosses. In addition, characters may have one or more phonetic readings, so the readings are also presented in the notes. Though Latin names are capitalized and/or italicized in the handbooks, lowercase is maintained due to naming conventions for the annotations. The subscript “x” is identified in the annotations as “-x”, as in the annotation “= *syllabic ha-x*” for U+14561  ANATOLIAN HIEROGLYPH A314.

Some annotations list the character elements used in its graphic appearance. These annotations appear in the handbooks (Marazzi, CHLI and Herbordt). For example, U+1452D  ANATOLIAN HIEROGLYPH A267 is identified as the Latin logogram STELE, but is actually composed of A267A+A268 + LAPIS+SCALPRUM (the plus indicates the two are graphically joined). Both STELE and LAPIS+SCALPRUM are included in the annotation. Some characters may just contain the elements that appear graphically, as in the character U+1451C  ANATOLIAN HIEROGLYPH A250, which is annotated “MAGNUS.DOMUS”, indicating it is composed of two logograms (A363.A247 . with the full stop indicating the two elements are not graphically joined), or a logogram with a syllabogram (U+14505  ANATOLIAN HIEROGLYPH A227A, composed of a225+a278 + URBS+*li*). These characters are not decomposable. No comments were included in the annotations if a particular sign was used only in certain combinations; an example is  A300 which is usually preceded by  A045 or followed by  A488. This is a matter of orthography and specialists working with Anatolian Hieroglyphs would presumably already know this.

**5.4 Cross-references in the names list.** The cross references indicate the relationship of variant characters to the main character, such as U+14403  ANATOLIAN HIEROGLYPH A004 → U+144EC  ANATOLIAN HIEROGLYPH A207. This kind of annotation provides vital information to the user of the names list, namely, guidance as to the proper character to be used for a normalized text, since the variant forms have been deemed obsolete. This is essential information which should be retained in the names list. The cross-reference pattern is that used in Marazzi, which typically provides a cross-reference from an Empire (early) form to the later form, where the values listed (so A320 has a cross-reference to A165, A270 has a cross-reference to A070, etc.).

**5.5 Issues with Latin names and phonetic readings in the names list.** Earlier drafts of this proposal received criticism regarding the annotations, and their use has been minimized here.

**6. Numbers.** Further research needs to be done to ensure that a sufficient quantity of Anatolian numbers are encoded to be practical in use. At present numbers for 1, 2, 3, 4, 5, 8, 9, 10, 12, 100, and 1000 have been proposed for encoding.

**7. Text processing.** For non-linear display of Anatolian Hieroglyphs, the conventions of the Egyptological *Manual de codage* should be sufficient, and indeed it appears that software which can support Egyptian Hieroglyph rendering can also support Anatolian.

**8. Glyphs.** The fonts used in this proposal were designed by Michael Everson, based on glyphs in Massimiliano Marazzi, *Il Geroglifico Anatolico* (Napoli 1998) and on glyphs in Emmanuel Laroche, *Les Hiéroglyphes Hittites* (Paris 1960) and in Suzanne Herbordt, *Die Prinzen- und Beamtensiegel der*







14612;ANATOLIAN HIEROGLYPH A475;Lo;0;L;;;;;N;;;;;  
14613;ANATOLIAN HIEROGLYPH A476;Lo;0;L;;;;;N;;;;;  
14614;ANATOLIAN HIEROGLYPH A477;Lo;0;L;;;;;N;;;;;  
14615;ANATOLIAN HIEROGLYPH A478;Lo;0;L;;;;;N;;;;;  
14616;ANATOLIAN HIEROGLYPH A479;Lo;0;L;;;;;N;;;;;  
14617;ANATOLIAN HIEROGLYPH A480;Lo;0;L;;;;;N;;;;;  
14618;ANATOLIAN HIEROGLYPH A481;Lo;0;L;;;;;N;;;;;  
14619;ANATOLIAN HIEROGLYPH A482;Lo;0;L;;;;;N;;;;;  
1461A;ANATOLIAN HIEROGLYPH A483;Lo;0;L;;;;;N;;;;;  
1461B;ANATOLIAN HIEROGLYPH A484;Lo;0;L;;;;;N;;;;;  
1461C;ANATOLIAN HIEROGLYPH A485;Lo;0;L;;;;;N;;;;;  
1461D;ANATOLIAN HIEROGLYPH A486;Lo;0;L;;;;;N;;;;;  
1461E;ANATOLIAN HIEROGLYPH A487;Lo;0;L;;;;;N;;;;;  
1461F;ANATOLIAN HIEROGLYPH A488;Lo;0;L;;;;;N;;;;;  
14620;ANATOLIAN HIEROGLYPH A489;Lo;0;L;;;;;N;;;;;  
14621;ANATOLIAN HIEROGLYPH A490;Lo;0;L;;;;;N;;;;;  
14622;ANATOLIAN HIEROGLYPH A491;Lo;0;L;;;;;N;;;;;  
14623;ANATOLIAN HIEROGLYPH A492;Lo;0;L;;;;;N;;;;;  
14624;ANATOLIAN HIEROGLYPH A493;Lo;0;L;;;;;N;;;;;  
14625;ANATOLIAN HIEROGLYPH A494;Lo;0;L;;;;;N;;;;;  
14626;ANATOLIAN HIEROGLYPH A495;Lo;0;L;;;;;N;;;;;  
14627;ANATOLIAN HIEROGLYPH A496;Lo;0;L;;;;;N;;;;;  
14628;ANATOLIAN HIEROGLYPH A497;Lo;0;L;;;;;N;;;;;  
14629;ANATOLIAN HIEROGLYPH A501;Lo;0;L;;;;;N;;;;;  
1462A;ANATOLIAN HIEROGLYPH A502;Lo;0;L;;;;;N;;;;;  
1462B;ANATOLIAN HIEROGLYPH A503;Lo;0;L;;;;;N;;;;;  
1462C;ANATOLIAN HIEROGLYPH A504;Lo;0;L;;;;;N;;;;;  
1462D;ANATOLIAN HIEROGLYPH A505;Lo;0;L;;;;;N;;;;;  
1462E;ANATOLIAN HIEROGLYPH A506;Lo;0;L;;;;;N;;;;;  
1462F;ANATOLIAN HIEROGLYPH A507;Lo;0;L;;;;;N;;;;;  
14630;ANATOLIAN HIEROGLYPH A508;Lo;0;L;;;;;N;;;;;  
14631;ANATOLIAN HIEROGLYPH A509;Lo;0;L;;;;;N;;;;;  
14632;ANATOLIAN HIEROGLYPH A510;Lo;0;L;;;;;N;;;;;  
14633;ANATOLIAN HIEROGLYPH A511;Lo;0;L;;;;;N;;;;;  
14634;ANATOLIAN HIEROGLYPH A512;Lo;0;L;;;;;N;;;;;  
14635;ANATOLIAN HIEROGLYPH A513;Lo;0;L;;;;;N;;;;;  
14636;ANATOLIAN HIEROGLYPH A514;Lo;0;L;;;;;N;;;;;  
14637;ANATOLIAN HIEROGLYPH A515;Lo;0;L;;;;;N;;;;;  
14638;ANATOLIAN HIEROGLYPH A516;Lo;0;L;;;;;N;;;;;  
14639;ANATOLIAN HIEROGLYPH A517;Lo;0;L;;;;;N;;;;;  
1463A;ANATOLIAN HIEROGLYPH A518;Lo;0;L;;;;;N;;;;;  
1463B;ANATOLIAN HIEROGLYPH A519;Lo;0;L;;;;;N;;;;;  
1463C;ANATOLIAN HIEROGLYPH A520;Lo;0;L;;;;;N;;;;;  
1463D;ANATOLIAN HIEROGLYPH A521;Lo;0;L;;;;;N;;;;;  
1463E;ANATOLIAN HIEROGLYPH A522;Lo;0;L;;;;;N;;;;;  
1463F;ANATOLIAN HIEROGLYPH A523;Lo;0;L;;;;;N;;;;;  
14640;ANATOLIAN HIEROGLYPH A524;Lo;0;L;;;;;N;;;;;  
14641;ANATOLIAN HIEROGLYPH A525;Lo;0;L;;;;;N;;;;;  
14642;ANATOLIAN HIEROGLYPH A526;Lo;0;L;;;;;N;;;;;  
14643;ANATOLIAN HIEROGLYPH A527;Lo;0;L;;;;;N;;;;;  
14644;ANATOLIAN HIEROGLYPH A528;Lo;0;L;;;;;N;;;;;  
14645;ANATOLIAN HIEROGLYPH A529;Lo;0;L;;;;;N;;;;;  
14646;ANATOLIAN HIEROGLYPH A530;Lo;0;L;;;;;N;;;;;

### Primary bibliography

- Hawkins, John David, and Halet Çambel. 2000. *Corpus of hieroglyphic Luwian inscriptions*. Berlin & New York: Walter de Gruyter. ISBN 3-11-010864-X
- Herbordt, Suzanne. 2005. *Die Prinzen- und Beamtensiegel der hethitischen Grossreichszeit auf Tonbulln aus dem Nişantepe-Archiv in Hattusa*. Mit Kommentaren zu den Siegelinschriften und Hieroglyphen von J. David Hawkins. Mainz am Rhein: Verlag Philipp von Zabern. ISBN: 3-8053-3311-0
- Laroche, Emmanuel. 1960. *Les hiéroglyphes hittites. Première partie: L'écriture*. Paris: Éditions du Centre National de la Recherche Scientifique.
- Marazzi, Massimiliano, Natalia Bolatti-Guzzo, & Paola Dardano. 1998. *Il geroglifico Anatolico: sviluppi della ricerca a venti anni dalla sua "ridecifrazione". Atti del Colloquio e della tavola rotonda Napoli-Procida, 5-9 giugno 1995*. (Istituto Universitario Orientale, Dipartimento di Studi Asiatici, Series Minor; 57) Napoli: Istituto Universitario Orientale.

### Secondary bibliography

- Hawkins, John David. 2003. "Scripts and Texts" in *The Luwians*, ed. H. Craig Melchert. (Handbook of Oriental Studies, Section 1: The Near and Middle East, vol. 68). Leiden: Brill.
- Melchert, H. Craig. 1996. "Anatolian Hieroglyphs" in *The World's Writing Systems*, ed. Peter T. Daniels & William Bright. New York; Oxford: Oxford University Press. ISBN 0-19-507993-0
- Meriggi, Piero. 1962. *Hieroglyphisch-hethitisches Glossar*. Zweite, völlig umgearbeitete Auflage. Wiesbaden: Otto Harrassowitz.
- Payne, Annick. 2004. *Hieroglyphic Luwian*. (Elementa Linguarum Orientis; 3) Wiesbaden: Harrassowitz. ISBN 3-447-05026-8

### Acknowledgements

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.