Unlocking the Mayan Hieroglyphic Script with Unicode

Carlos Pallán Gayol (MAAYA Project, University of Bonn) and Deborah Anderson (SEI, Dept. of Linguistics, UC Berkeley) DH2016 – Krákow, Poland







Ancient Maya writing





"Landa's syllabary"

Códice Dresde, p. 50



Image source: SLUB Library, Dresden

Códice Madrid, p. 26



Image source: Museo de América, Madrid

Codex production



Image source: National Geographic

Examples of Logograms (aka "word signs"):



B'A:LAM b'a:lam "jaguar"



CHA:K *Cha:k* "rain god"



TU:N tu:n "stone"

Examples of syllabic spellings:



b'a-ka-b'a *b'akab'* "first of the land" (title)

b'a-la-ma

b'a:lam

"jaguar"

b'a-ki *b'a:k* "bone" (title)

ka-b'a *kab'* "earth" (land)

The Maya syllabary:



Th	e Maya	ı syllaba	ry:
. E I		ං ගි ිටි මබම	
• @		- CD2	
	ی ۲	C) S)	
•	60	: E	
	E I	Ð	
* (V) * (V) * (V)			, see the second
• @ \$? ?	G S S	ر الألغ
- A	o ge		E 5
• 🚱	•	(@); {[]]}	(H)
•	69 6		

Japanese Hiragana syllabary:

	а	1	u	е	ο					
ø	あ	UV.	う	え	お					
k	か	き	<	け	Ĵ					
s	さ	L	す	せ	そ					
t	te	ち	0	τ	٤					
n	な	10	ぬ	ね	Ø					
h	は	ひ	\$	^	æ					
m	ま	74	む	හ්	ŧ					
У	や		кþ	01B 001	ት					
r	Б	IJ	る	れ	3					
w	わ	ക		象	を					
		∕v (N)								



Infixation:



Conflation:



Superimposition:



Unicode Standard

• The Unicode Standard is the international standard used for the electronic interchange of text



Unicode Standard: communication



Unicode Standard: searching

G 🔄 - Google Search		× +						
€ 🛈 🔒 https://ww	ww.google	e.com/#q=ਙ					C	Q. Search
Google	ŋ							Q
	All	Maps	Images	Videos	Shopping	More 🔻	Search tools	
	About	: 1,570 resi	ılts (0.28 sec	onds)				
	වී (graph වු e	e gyptian emica.cor gyptian hie	l hieroglyr n/놀 roglyph i011a	oh i011a (a (U+13196)	U+13196) (@ Graphemica	@ Graphe	emica a	

Unicode Standard: archiving

- Stable format international standard
- Important for grant applications
- Provides long-term access
- Aids in discoverability



Getting a script into Unicode

- Write a Unicode proposal that describes the script in detail
- Get approval by two standards committees (Unicode TC and ISO group)





Getting a script supported in software

- After approval (2+ years): need fonts and software support
- Cf. Cherokee 4 years 12 years Cherokee Cherokee appears Cherokee Proposal in Chrome approved; written browser/ Apple published in products Unicode 3.0 1995 2005 2011 1999 CHEROKEE NATION

Getting a script into Unicode: New Developments – Univ. Shaping Engine

• Universal Shaping Engine (USE) should help reduce the time it takes for new scripts to be supported in fonts and applications (after approval)

12 years > 1-3 years?

Microsoft	Store ~	Products ~	Support	Posted Feb	ruary 23, 2	2015
The Fire Hose	The Offi	cial Microsoft Blog	Microsoft On t	he Issues	Next	Tra
Windo Unive Iangu	ows' (rsal S age s	global r haping scripts	each exp Engine 1	bands for	with	1
ଽୖୖ	<u>Š</u> Š	रिदेर्षि	দ্বিধ	ផតំ	Z	۲I

Getting a script into Unicode: Lengthy process

• Takes patience, commitment and dedication to see a script from first proposal until appearance in the Unicode Standard, fonts and software.



Script Encoding Initiative at UC Berkeley

- Helps user communities get scripts (and characters) proposed for Unicode
- Since 2002, has assisted in getting over 80 scripts into Unicode
- Keeps proposals moving forward



Meeting on Newa script in Kathmandu, Nepal, 2014; first proposal 2000, script published June 2016



Script Encoding Initiative at UC Berkeley

- In 2015 and 2016, set up meetings at UC Berkeley between Carlos Pallán and Unicode experts
- Discussed
 - Requirements of Mayan script
 - Ways to handle Mayan features in Unicode
- Funding received for 2017 from Unicode Consortium for Mayan work by C Pallán



Other Hieroglyphic Scripts in Unicode

• Egyptian hieroglyphs (1,071 chars. published in Unicode 5.2, 2009)



• Meroitic hieroglyphs (published in Unicode 6.1, 2012)

ᢞᢁᢅᢢᡘᢛ᠍ᢁ᠁

• Anatolian hieroglyphs (published in Unicode 8.0, 2015)



Unicode approaches relevant to Mayan: Egyptian hieroglyphs clustering

• Current default display of Egyptian hieroglyphs

• Desired layout



Unicode approaches relevant to Mayan: Egyptian hieroglyphs clustering

- Since 2009, Egyptologists have been using software that created images, not Unicode characters, to create clusters.
 - Not standardized
 - Not searchable
 - Not widely interchangeable
- In 2015, Universal Shaping Engine was released
 - Had a working model showing how clusters could be created with 3 new characters



Unicode approaches relevant to Mayan: Egyptian hieroglyphs clustering

• Egyptian clustering with new format characters (with Univ Shaping Engine)



Unicode approaches relevant to Mayan: Chinese-Japanese-Korean (CJK) descriptors

• CJK Ideographic Description Characters used to describe layout

Unicode approaches relevant to Mayan: Chinese-Japanese-Korean (CJK) descriptors

U+2FF1	U+2FF0	U+2FF8	2FF4	2FF3	2FF2	not available	not available	not available	not available	not available	not available
D ₁ D ₂	D ₁ D ₂	D ₂	D ₁	D ₁ D ₂ D ₃	D ₁ D ₂ D ₃				a▼ C	a b	av dv b► ie▲
关	侮	쇸	巷	鋊	衍	not attested?	not attested?	not attested?	not attested?	not attested?	not attested?
H		đ	M		E.	-ab	B	13			



Concordance tool



Optimizes annotating/creating Mayan textual atabases

•	10.39 0 Tot Records	9 al (Sorted)	Show All	Vew Record Delete Record	Find	- Mart								
	RDANCE . W	ew As: 🔳 📖	Previe											
Picture_EV	R al EVR_Code	EVR_variant	Picture_Th	Thompson_Code	T_vari.	Picture_Ma	Macri_Vail_Code	MV_V_	Picture_Zi	Zimmermann_Code	Z_V	Phonetic Value	Value_Thompson	Iconic_valu
	E-17	Concentration (1997)	0	T0526	and a second of		Y51			Z1337				17th DAY (KAB'AN
8	E-18		\odot	T0527	-	1	XQ8			Z1338				10th.DAY (ETZ'NA
	E-19		0	T0528			ZC1			Z1339		TU:N?	(Р) Ровадие	19th.DAY (KAWAK
0	E003											777	(UNK) Unknown, undeclohered	777
00	E004		0.4	T0287			HES		69	20050		CH'OK?/ch'o?	and should be	
000	E006		(alala)	T0069	v0t		33B		600	Z0044		je?/AKAN?	(P) Possible	
0::0	E010		0 0	T0140	v01	0'*'0	AMB	¥02	0:0	Z0063		la	(K) Known	
jā.	E012		1	T0137	v01		331.		1	Z0007				
ນນນ	E013	v01	000	T0136	v02		33F	v01		20065, 0078		ji .	(K) Known	
1	E013	v02	1	T0136	vOt	2 }	ZD2	v03		Znn		no	(K) Known	
880	E014			Tnn			000			Znn				
nnn	E015		ത്ത്തി	T0135	v02		3267			Znn				
	E017		ക്ക	T0124			323		ശരവ	Z0045		tzi	(K) Known	FOLIATION/SCR
1	E021		1	T0171			PCE		9	20015		assign UND		
S	E022		- 2					1				ne?	(P) Possible	SCROLL/CURL
Ser	E022.272		253	T1049	1	B	SCE			Z0150	_	UH?	(P) Possible	
-	E023			T0025	v05	Ĩ	AA1	V01		70081	-	ka	00 Known	
100	E024		(num)	10023	v03		161			70079				
*	E025		- u	10048	101	-	161	102		20072		NAH	(K) Known	
1	E026		8	10010		1	HEG		F	70001	-			
-i-	E027		1	10001	vot	1	004	vot	i	20001a			-	
13	E030		- Å	T0181			7111		2	200018		10	-	
1	6021		1º per	10101	-		163	-		20060		ja ni	W Kenne	-
	6031		1000	70110		-	102	-	L	20002		N'ANT.	(vy solen	
	E032		1 gine	10122	-	70	250		5	20030		n an	-	
- UII	E032;242			10122:05034	-	æ	250	wurt -		20038:1357		KAK	(A) Anden	
	E033			10126		-	32M	-	-7 m 7	20074		уа		
20.02	E033-003	566 E440		10026		23/9/26	201	-	2.3	0074.1301.0074	Z1301		2	
r	E034		1	101/1	v92	-	PCE		7	20015			-	
1	E035	101	I CON	10070	v01	-	32A	V02	Ť	20017	_	AJ??		
×	E0357	¥02	oro	10074	v05		32A	v01		20075		ma?		
4	E036		5	T0015	¥02		258	¥02	y .	Z0010a		77	(PR) Problematic	
÷.	E040		6	10033	v01		AMC?		E.	Z0030		K,OH	(K) Known	DROPLETS
00	E041		0	T0073			32A		00	Z0044a		he??	(PR) Problematic	
8	E043		8	T0013	v04	1	HE6			Z0001		U	(K) Kinown	
8	E044			T0012	v01		164		3	20025		AJ/a	-	
6	E045		QQ	T0013	+03	6	HEG		8	Z0002				
8	E047		3	T0015	v01		258		9	Z0010		77	(PR) Problematic	
8	E050		g	T0229_v01		1	AL1		1.000	???		a/AJ	(K) Known	
8	E051		S	T0147			PT3		8	Z0014		assign UND		
4	E053		10	T0346			266			???				

🛉 File Edit View Insert Format Recor	ds Scripts Tools Window Help							
171 2933 / 7 -j- Records Found (2	412 Sorted) Show All New Record Delete Record Find	Sort Sort						
Layout: MAAYA_Annotations2 View A	s: Preview							_
Concordance	Annotations Calendrics	Alm	anacs		Captions		lcons	5
EXTRACTOR	Add New Record DRE_CPG MAD_GK P	AR_JS	1226 FN	FONDS NATIONAL SUISSE SOWEDZERSCHER NATI FONDO NAZIONALE SVIZZ SWISS NATIONAL SCIENC		G Deutsche Forschungsgemeir	nschaft univers	sität <mark>bonn</mark>
			Press on the gray r	member boxes [a,b	olito see correspondir	ng givphs in the CUNCU	IRDANCE	Althou
Column_Code_ID	Section Cal_Cycle Cycle_Alt. Page T'ol/Reg. Frame		а	b	c	b	e	ſ
DRE_05b_B1	▼ Almanac 4b ▼ 05 b 7	Thompson_IDs	0759b	0025	0181			
Select Codex:	Cycle_type Glyph_Dates	Macri & Vail 2009	AP7	AA1	7111			
CON 2 DRE_CPG	Concordance Visual Std TUN	Zimm 1956 IDs	0706	0004	201			
Reading Order:	Members Members ranking (1-4) Context_Important& h_c 3.3.4	EVR_1961_IDs	0706	0081	0080			
03	significant: a,0,0 0,0,4	Type in this row to autocomplete all	322 :	023 .	030			
Glyph_block	Link to www.mayacodices.org (corresponding Frame):	Transcription p	e?	ka	ja			
SR CR GCY BY RY B1	http://mayacodices.org//text.asp?	Transliteration a	uto:					
	pe?-ka-ja	ransliteration mar	ualspe(h)kaj					
		Segmentation			Morph_Analysis	s		Tran
Column_Code_ID	Section Cal_Cycle Cycle_At. Page T'ol/Reg. Frame		а	b	С	d	е	ſ
DRE_05b_B2	Cycle type	Thompson_IDs	0668 ·	0102				
DRE_CPG V	Show Pagination Concordance Visual Std TDN	Macri & Vail 2009	MZ9 ·	1B2				
Reading	Nembers Members ranking (1-4)	Zimm_1956_ID5	0169 •	0061				
Order:	Context_Important2 b 3.4	EVR_1961_IDs Type in this row to	530 .	112				
04	Link to www.mayacodices.org	autocomplete all			_			_
Glyph_block	(corresponding Frame): http://mayacodices.org/feed.asp?	c	na	KI				
SR CR GCV BV RV	cha lá	Transliteration a	uto:					
	LI Id-N	Principal Contraction and	UNI Chaighja		Momb Analysis			Tran
		segmentation						
Column_Code_ID	Section Cal_Cycle Cycle_Alt. Page T'ol/Reg. Frame		а	b	с	d	e	ſ
DRE_05b_C1	D4-15 V Almanac 5b V 05 b 01	Thompson_IDs	0589	0765b				
Select Codex:	Show Pagination Glyph_Bates	Macri & Vail 2009	ZUF .	AP5				
Bag	concordance root to to to	Zimm_1956_IDs	1369	0707				
A Corder:	Context_Important2_b 4.4	EVR_1961_IDs	424	323				
CEL CON DE		autocomplete all						
05	Link to www.mayacodices.org							
Glyph_block	Link to www.mayacodices.org (corresponding Frame):	Transcription	0/j0	ch'o				
SR CR CCV BV RV	Link to www.mayacodices.org (corresponding Frame): http://mayacodices.org//exd.asp?	Transcription	0.j0 uto:	ch'o				
Gyph_block	Link to www.mayacodices.org (corresponding Frame): http://mayacodices.org/feed_asp2 0-Ch'0	Transcription Transliteration a Transliteration mar	0/j0 uto: ual: <i>j0ch</i> '	ch'o				

Towards OCR-compatible digital repositories

RASTER IMAGE (i.e. scan) static (i.e. non searchable)



D1

D2

D3

ENCODED TEXTUAL DATA

Dynamic (enables OCR recognition, usage of different fonts, querying, text-mining, etc.)

C2

117.260



C3

423/515

+176/204.031

D3

D1





D2



530.112 515/504.013

026.401

Dresden Codex page 30b, text

(examples in (LA)TEX by B. Delprat and S. Orevkov 2012)

Thank you - Questions?

Contact:

Carlos Pallán Gayol pallan.carlos@gmail.com Debbie Anderson: dwanders@berkeley.edu

This talk was supported by NEH Preservation and Access grant PR-50205-15 and the DFG (Deutsche Forschungsgemeinschaft)