



African Scripts on Computers and Devices: An Update

Deborah (Debbie) Anderson
Researcher, Dept. of Linguistics, UC Berkeley

LOC.
WORLD

#[LocWorldWide46](#) Africa 

Script Encoding Initiative

Dept. of Linguistics
UC Berkeley

- Department of Linguistics
University of California, Berkeley
- Scripts to Encode
• Progress Overview
- News & Presentations
• Press
- How to Donate
• Donors

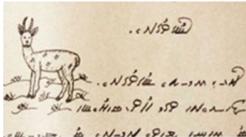
WELCOME TO THE SCRIPT ENCODING INITIATIVE

The Script Encoding Initiative (SEI), established in the [UC Berkeley Department of Linguistics](#) in April 2002, is a project devoted to the preparation of formal proposals for the encoding of scripts and script elements not yet currently supported in Unicode (ISO/IEC 10646).

[Unicode](#) is the universal computing standard specifying the representation of text in all modern software. To date, Unicode has largely focused on the major modern scripts, particularly those scripts most widely used in business. Some minority and historic scripts have already been encoded, as well as historic characters of the major modern scripts.



The goal of the SEI project is to fund the preparation of script proposals that will be successfully approved by the Unicode Technical Committee and WG2 (ISO/IEC 10646) without requiring extensive revision or involvement of the committee itself.



A secondary goal to encourage the creation of freely-available Unicode-conformant fonts. This will help to promote widespread adoption and implementation of the scripts.

By providing funding for proposal authors, drawn from faculty and graduate students as well as other experts, the Script Encoding Initiative represents a concerted effort to tackle the remaining scripts and remaining script issues. The project will be assisted by a Unicode Vice President to assure that the proposals meet requirements of the Unicode Technical Committee and of the international standards community. To date, the project has helped get over [70 scripts encoded](#).

The Script Encoding Initiative project is of world-wide importance, for minority and historic scripts. For a minority language, having its script included in the universal character set will help to [promote native language education, universal literacy, cultural](#)



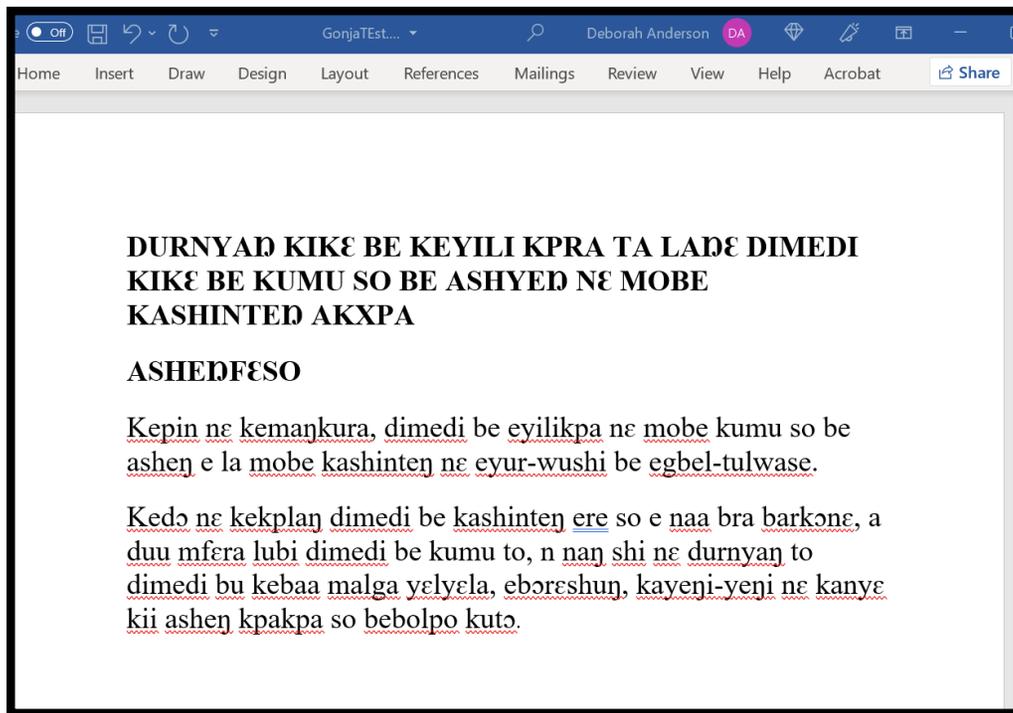
SEI website: <https://linguistics.berkeley.edu/sei>; Script images are of Garay script



Scripts of the World



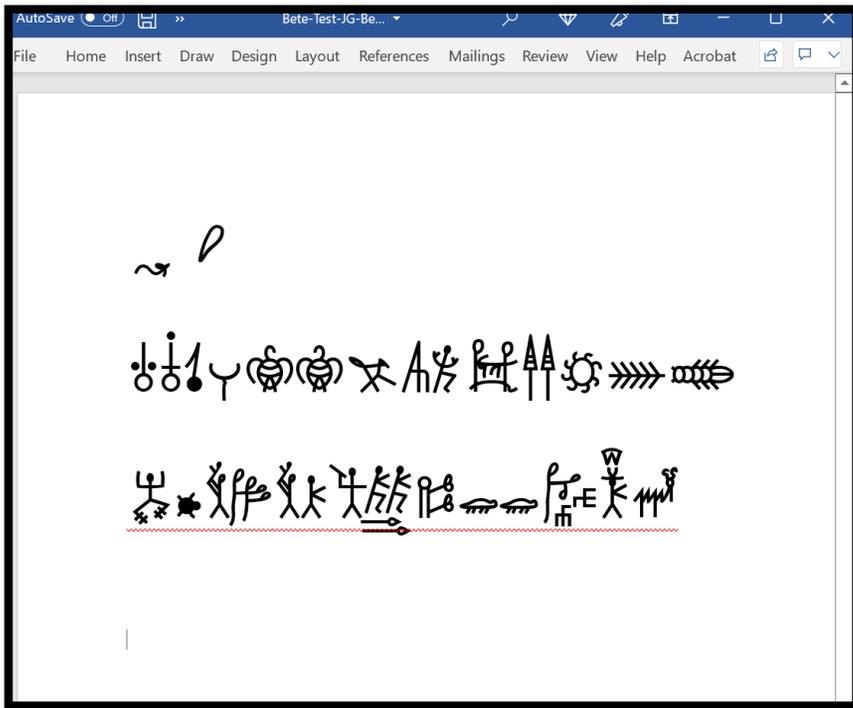
Sending text



Latin script; Gonja language (Ghana)



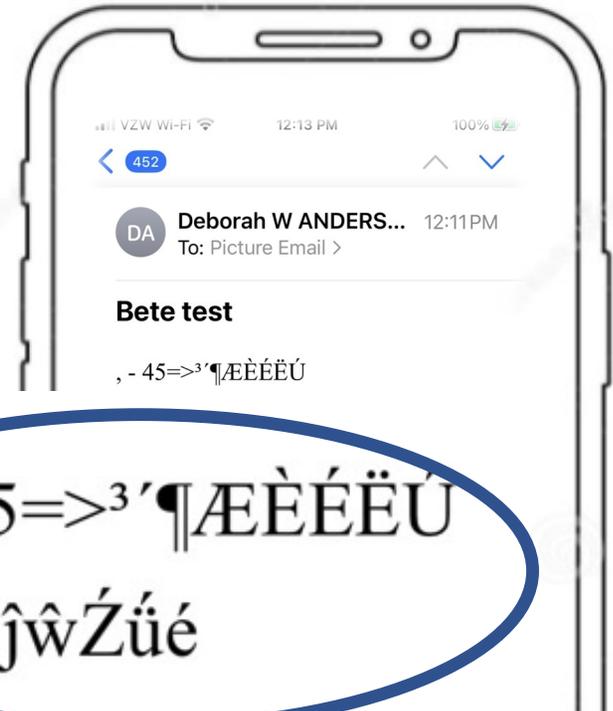
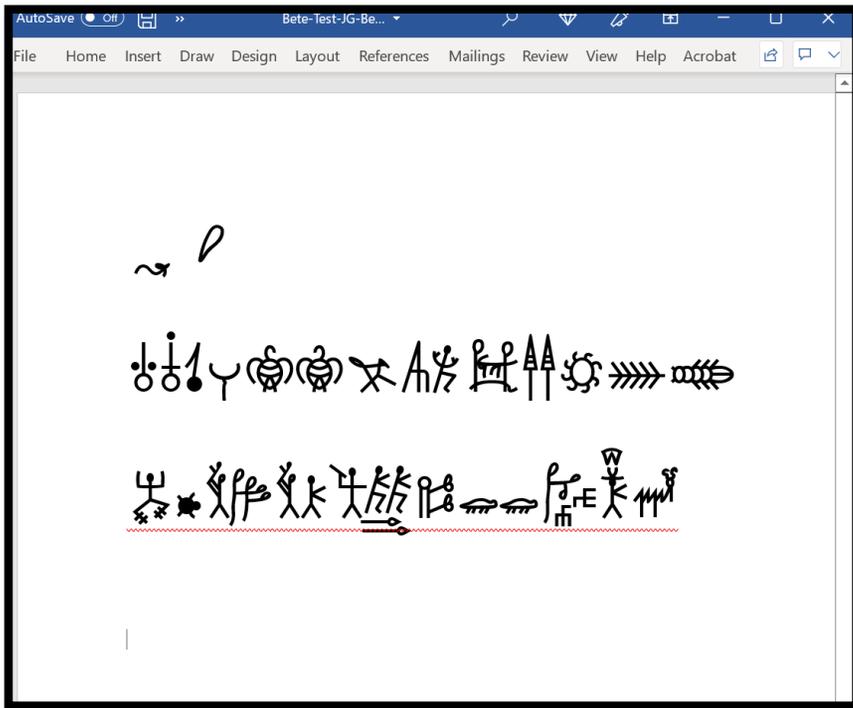
Sending text: The Problem



Bété script; Bété language (Côte d'Ivoire)



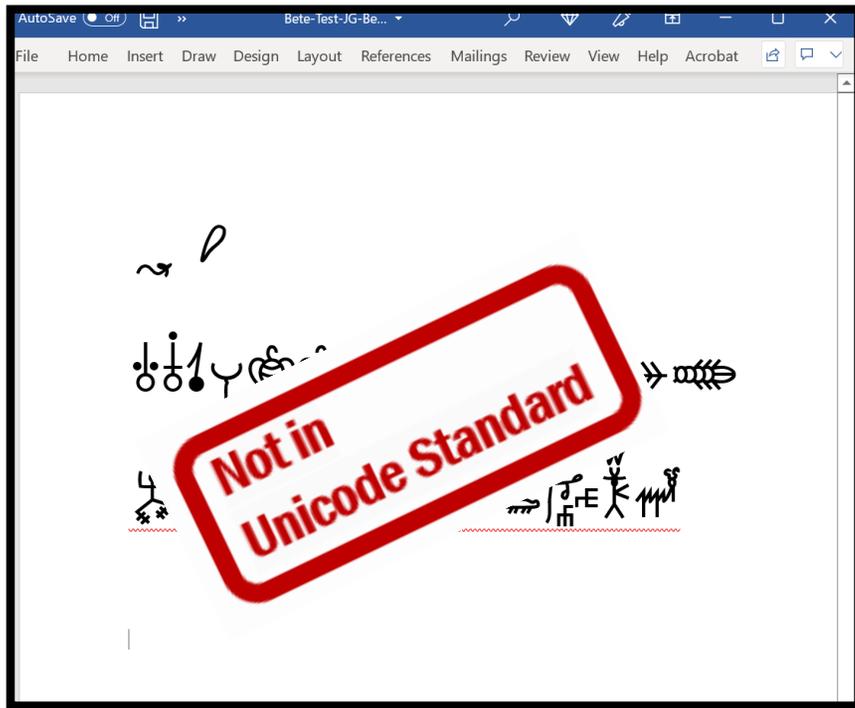
Sending text: The Problem



Bété script; Bété language (Côte d'Ivoire)



Sending text: The Problem

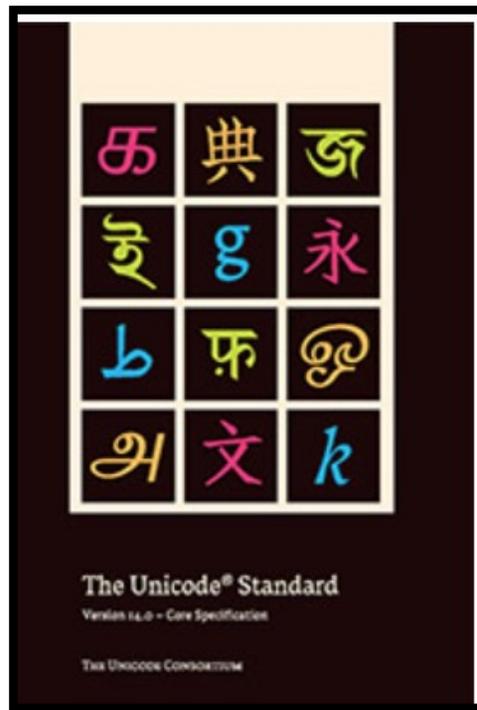


Bété script; Bété language (Côte d'Ivoire)

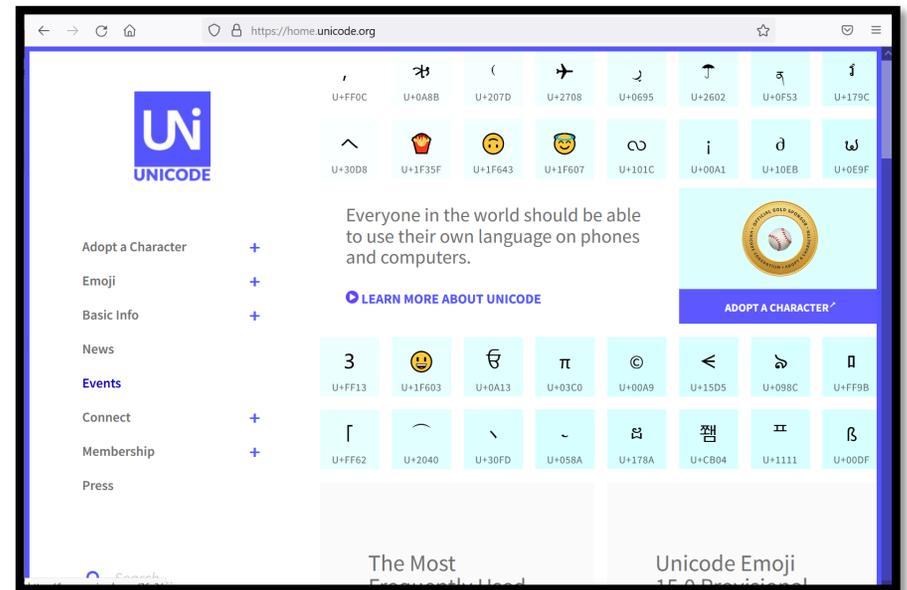


Unicode Standard

International character encoding standard



Print-on-demand publication of Unicode Standard



Unicode Website:
<https://home.unicode.org/>



What is Unicode?

Sultan NJOYA, Roi des Bamun, 1^{er} écrivain Camerounais



Le SHÜ-MOM, écriture inventée par le Sultan NJOYA en 1896

1 2 3 4 5 6 7 8 9 10 11 12
 a ka u ku é ré te o ngi i la pa ni rié lé
 13 14 15 16 17 18 19 20 21 22 23 24
 mé ta nda njem m su mu shi si sw sw kye ket nue nu
 25 26 27 28 29 30 31 32 33 34 35 36
 nzwe yo su yu ya sa kur pw ngé nté pü wü pé fé ru
 37 38 39 40 41 42 43 44 45 46 47 48
 lu mi ni rw nga na sho puo fu fom wa na li pi bo
 49 50 51 52 53 54 55 56 57 58 59 60
 ké mbèn ren men ma ti ki ken nkwen sw
 3=1 3=2 7=3 4=4 5=5 6=6 7=7 8=8 9=9 10=10

Bamum

	A6A	A6B	A6C	A6D	A6E	A6F
0	Ɀ A6A0	Ɀ A6B0	Ɀ A6C0	Ɀ A6D0	Ɀ A6E0	Ɀ A6F0
1	Ɀ A6A1	Ɀ A6B1	Ɀ A6C1	Ɀ A6D1	Ɀ A6E1	Ɀ A6F1
2	Ɀ A6A2	Ɀ A6B2	Ɀ A6C2	Ɀ A6D2	Ɀ A6E2	Ɀ A6F2
3	Ɀ A6A3	Ɀ A6B3	Ɀ A6C3	Ɀ A6D3	Ɀ A6E3	Ɀ A6F3

5377 1889 **FOUMBAN-CAMEROUN** 3777 1933



What is Unicode?

International character encoding standard

		Bamum					
		A6A	A6B	A6C	A6D	A6E	A6F
0							
	A6A0	A6B0	A6C0	A6D0	A6E0	A6F0	
1							
	A6A1	A6B1	A6C1	A6D1	A6E1	A6F1	
2							
	A6A2	A6B2	A6C2	A6D2	A6E2	A6F2	
3							
	A6A3	A6B3	A6C3	A6D3	A6E3	A6F3	

A6A0



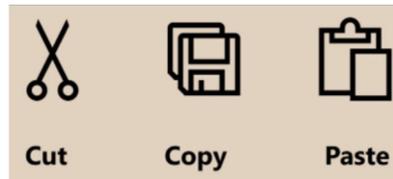
BAMUM LETTER A



Background: Unicode Basics

Once a script is in Unicode, the foundation is laid for users to:

- Copy and paste text



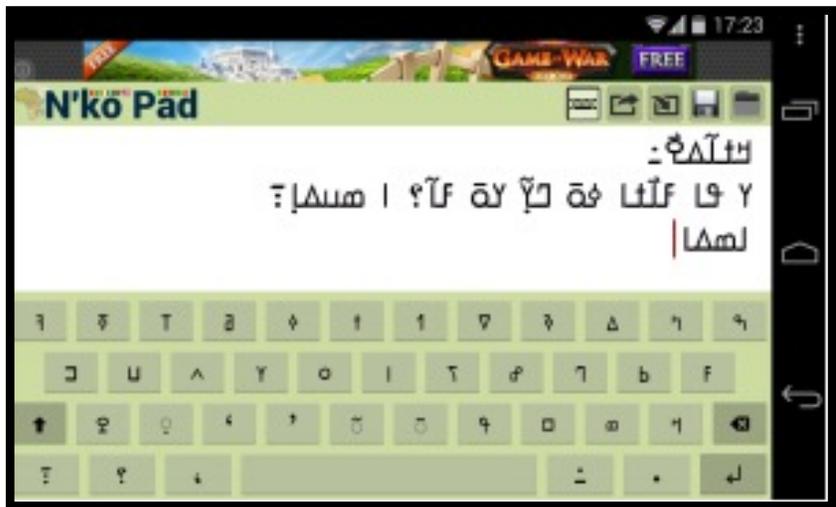
- Search on Internet, word-processing documents, email (etc.)



Background: Unicode Basics

Once a script is in Unicode, it can be

- used by modern language users and scholarly communities for digital communication



N'Ko



Coptic



Script Encoding Initiative

- Helps users get eligible scripts into Unicode Standard

Department of Linguistics
University of California, Berkeley

• Scripts to Encode
• Progress Overview

• News & Presentations
• Press

• How to Donate
• Donors

WELCOME TO THE SCRIPT ENCODING INITIATIVE

The Script Encoding Initiative (SEI), established in the [UC Berkeley Department of Linguistics](#) in April 2002, is a project devoted to the preparation of formal proposals for the encoding of scripts and script elements not yet currently supported in Unicode (ISO/IEC 10646).

Unicode is the universal computing standard specifying the representation of text in all modern software. To date, Unicode has largely focused on the major modern scripts, particularly those scripts most widely used in business. Some minority and historic scripts have already been encoded, as well as historic characters of the major modern scripts.



The goal of the SEI project is to fund the preparation of script proposals that will be successfully approved by the Unicode Technical Committee and WG2 (ISO/IEC 10646) without requiring extensive revision or involvement of the committee itself.



A secondary goal to encourage the creation of freely-available Unicode-conformant fonts. This will help to promote widespread adoption and implementation of the scripts.

By providing funding for proposal authors, drawn from faculty and graduate students as well as other experts, the Script Encoding Initiative represents a concerted effort to tackle the remaining scripts and remaining script issues. The project will be assisted by a Unicode Vice President to assure that the proposals meet requirements of the Unicode Technical Committee and of the international standards community. To date, the project has helped get over [70 scripts encoded](#).

The Script Encoding Initiative project is of world-wide importance, for minority and historic scripts. For a minority language, having its script included in the universal character set will help to promote native language education, universal literacy, cultural

<https://linguistics.berkeley.edu/sei>



Scripts in Unicode

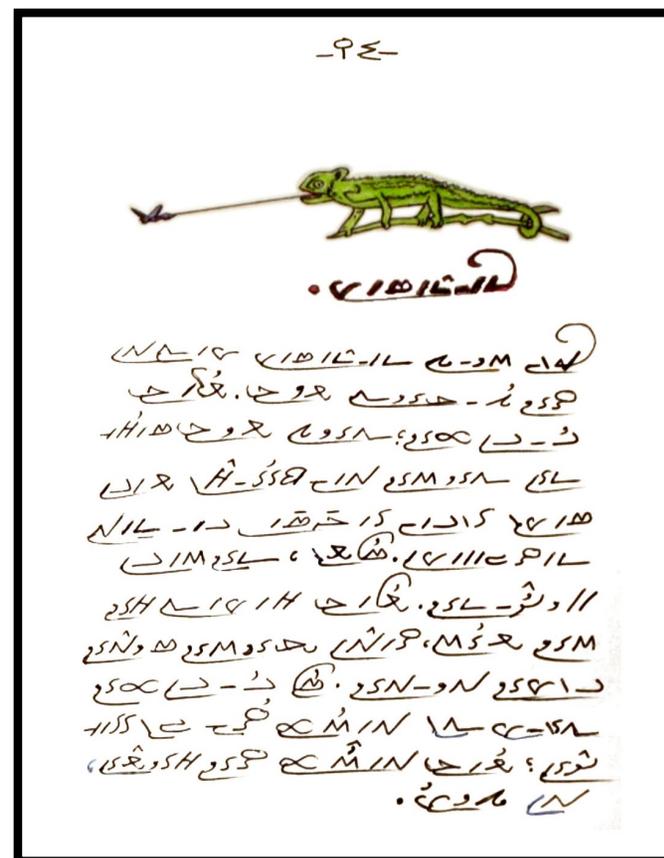
In Unicode: 159 scripts total



Unencoded Scripts

In Unicode: 159 scripts total

Not in Unicode: 130+ scripts



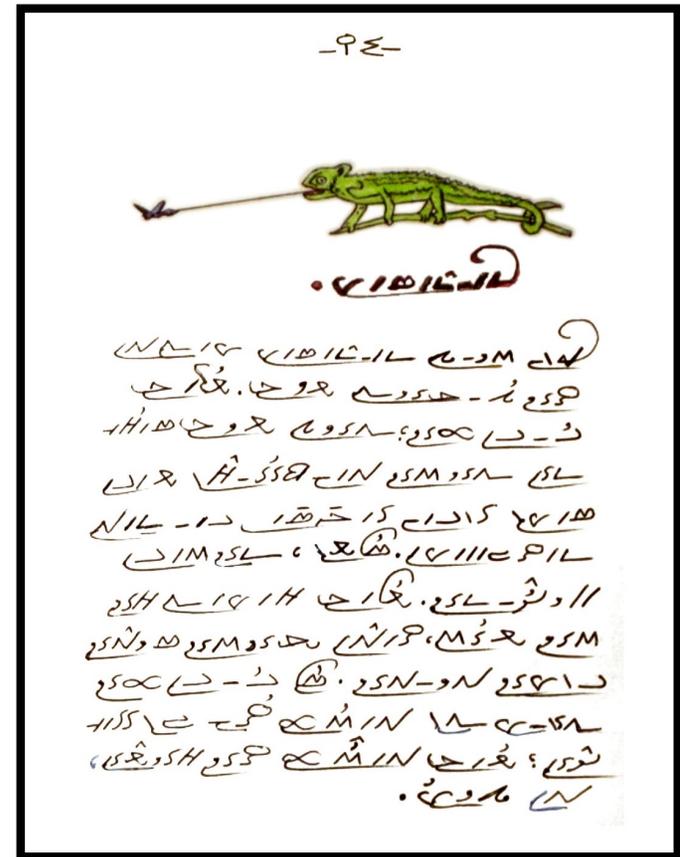
Garay script (not in Unicode yet)



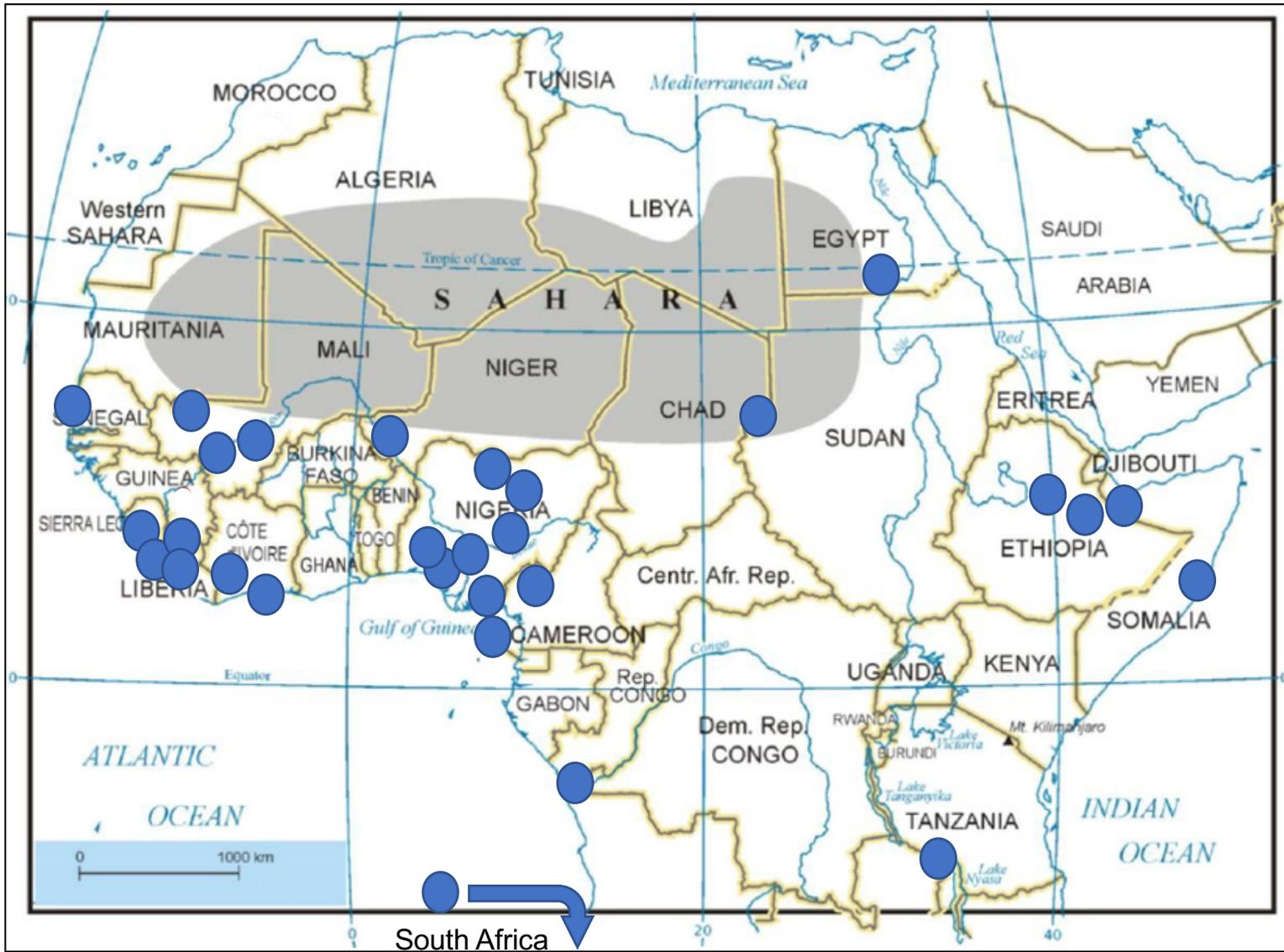
Unencoded Scripts

In Unicode: 159 scripts total
African scripts = 15 scripts

Not in Unicode: 130+ scripts
African scripts = 40 scripts
(7+ good candidates for Unicode)



Modern African Scripts not in Unicode



● Location of a script not in Unicode

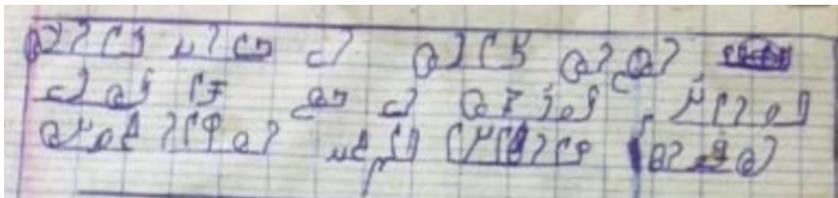


Steps to Getting a Script into Unicode



Steps to Getting a Script into Unicode

- Script is identified



Steps to Getting a Script into Unicode

- Script is identified
- Write a proposal, get review by committees, and if meets requirements, get approval



Steps to Getting a Script into Unicode

- Script is identified
- Write a proposal, get review by committees, and and if meets requirements, approval
- Published in Unicode

The Unicode[®] Standard



Steps to Getting a Script into Unicode

- Script is identified
 - Write a proposal, get review by committees, and and if meets requirements, approval
 - Published in Unicode
- The Unicode[®] Standard** } **at least 2 years
(usually more)**



Steps to Getting a Script into Unicode

- Script is identified
- Write a proposal, get review by committees, and and if meets requirements, approval
- Published in Unicode version
- Need fonts, keyboard, software updates, locale data

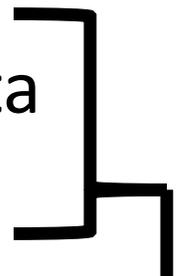


Steps to Getting a Script into Unicode

- Script is identified
- Write a proposal, get review by committees, and and if meets requirements, approval
- Published in Unicode version
- Need fonts, keyboard, software updates, locale data



can take years



Steps to Getting a Script into Unicode

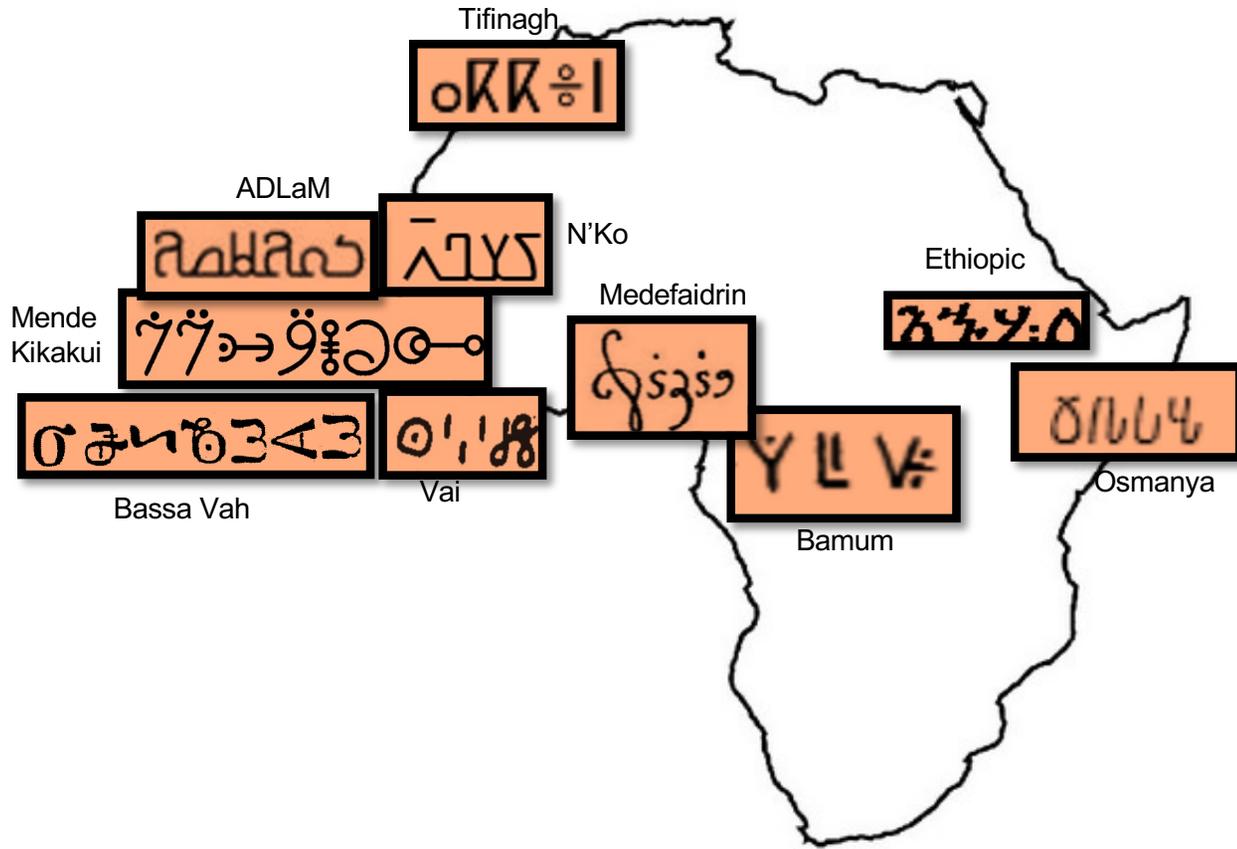
- Script is identified
- Write a proposal, get review by committees, and and if meets requirements, approval
- Published in Unicode version
- Need fonts, keyboard, software updates, locale data
- Accessible (hopefully!)



Status of African scripts and Outstanding Issues



Modern African Scripts in Unicode





	Supported across multiple platforms	Font	Notes
Ethiopic	✓	✓	
Osmanya		✓	
Tifinagh	✓	✓	
N'Ko	✓	✓	
Vai	✓	✓	
Bamum	✓	✓	LDML keyboard layout needed
Bassa Vah		✓	Character picker available
Mende Kikakui		✓	Limited web-based input methods available
ADLaM	✓	✓	
Medefaidrin		✓	



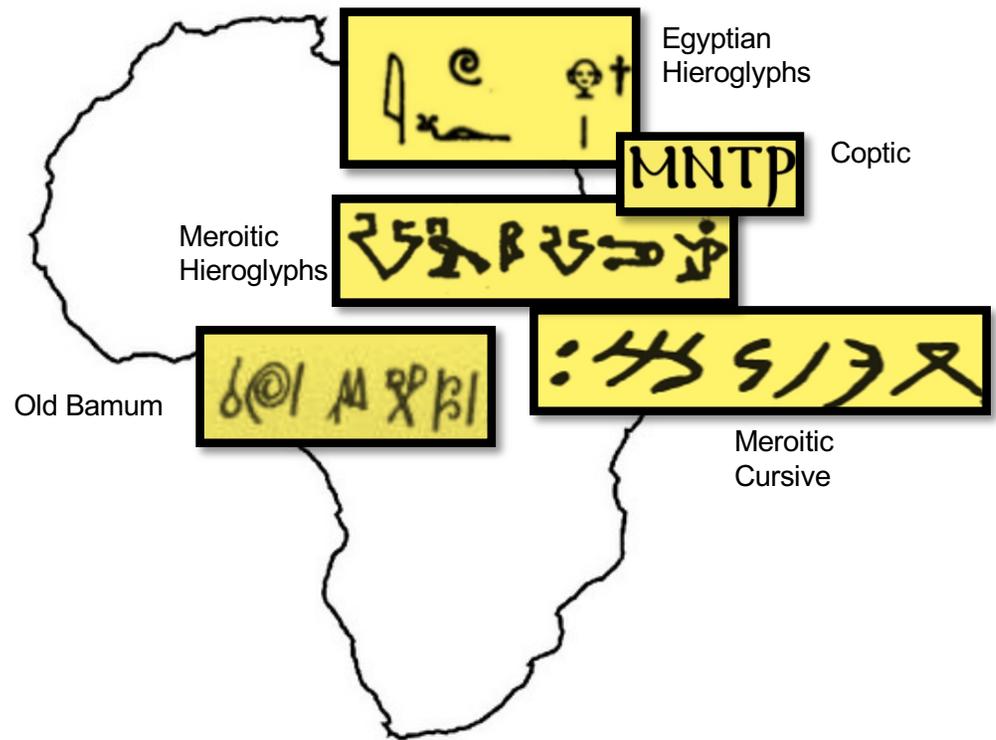
Historic African scripts in Unicode

Egypt and Nile Valley

- Egyptian Hieroglyphs
- Meroitic Hieroglyphs
- Meroitic cursive
- Coptic (now liturgical)

Central Africa

- Old Bamum



	Supported across multiple platforms	Font	Notes
Egyptian Hieroglyphs	✓	✓	New format controls being added; repertoire expansion underway
Meroitic Hieroglyphs		✓	
Meroitic Cursive		✓	
Coptic	✓	✓	
Old Bamum	✓	✓	



Source: Charles Riley and (for Coptic) Kamal Mansour



Issues for Scripts not in Unicode

- Need adequate printed material
- Need experts

Name	1	2	3	4	5	6	7	8	9	10	11	12	13
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1									
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								
Handwritten name	6	6	7	1	+								

Kpelle tax table



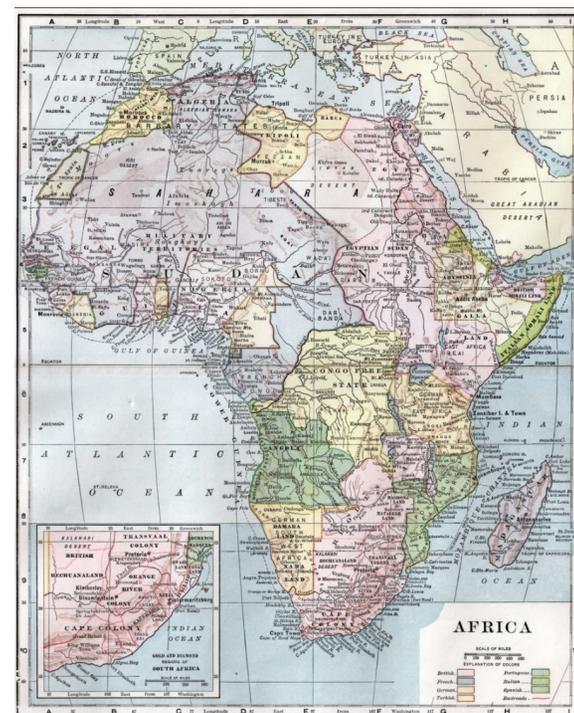
Creating Brand-New Scripts

- Often considered by users to be better suited to represent the sounds of their language (over the Latin or Arabic scripts)
- Reflects a community's unique linguistic identity and heritage



Creating Brand-New Scripts

- Often considered by users to be better suited to represent the sounds of their language (over the Latin or Arabic scripts)
- Reflects a community's unique linguistic identity and heritage
- Way to write indigenous languages of Africa, not tied to Africa's history of colonization



Map of Africa 1910 (Wikipedia)



Issues for Brand-New Scripts

- Committees will ask:
 - ❖ Number of users today?
 - ❖ How many publications?
 - ❖ Number of schools?



Two Brief Case Studies



I. Case Studies

– Ditema Tsa Dinoko (=Isibheqe Sohlamvu)

Key traits:

- Created in 2010s by linguists, designers, software programmers
- Used for southern Bantu languages in South Africa and neighboring countries
- Inspired by Sesotho mural art / Zulu beadwork
- No Unicode proposal



<https://isibheqe.org.za/>



I. Case Studies – Ditema Tsa Dinoko (=Isibheqe Sohlamvu)



7 | VN DIKAKANYO/TSELE LE TSELE April 2016

ISIGODLO SEZIFUNDISWA: LUGCINWA ULWAZI EMAGCEKENI AKWANGQONDONKULU

THEKWANE NDABA
isiZulu

KwaNgqondonkulu thina abafundi nezifundiswa sisendaweni lapho kukhona ixoxo elikhalayo emgodini, asilalele isexwayiso salo. KwaNgqondonkulu emagcekeni onke thina abangaphakathi senziwa ukuba sikholwe ukuthi umsebenzi wethu muhle. Iqiniso lisababa nxa kubonwe isimo sezifundiswa ezahlukaniswa nemindeni yazo. Kule ndaba kuzobekwa phansi izindlela ezintathu ulwazi lakwaNgqondonkulu lwebiwa ngazo. Okokuqala kungkusetshenziswa kwesiNgisi nakwesiBhunu emagcekeni emfundo ephakeme. Okwesibili kuyizinhlobo ezintathu zocwaningo ezicelwe ngabaphathi bezigodlo zifundiswa. Okwesithathu kuyizinhlobo zokucindezela zalo cwaningo ezifihliwe.

Ukucindezela sasetshenziswa isilungu ukucindezela abanengi. Lokhu kusenzeka emagcekeni ethu esimanje. Ikhulu abantu ezweni lethu bakhuluma izilimi zabo njengesithaca, isiLala, Sepedi, Xhosa, isiXhosa, Setswana, Sesotho, Tshivenda, isiBhunu, isiCamtho, SiSwati, isiNdebele, isiZulu nezinye eziningi, kodwa ke sisaduma isiNgisi nesiBhunu emnuthweni, komabonwakude, ekuqashweni nangaphakathi kwezombusazwe. Isilungu sesithathile zonke zezikhundla zokuzwakala kuleli zwe.

Yingakho ke kusondela kithi inkinga la.Njengabafundi asifuni neze ukuzilondolozela lo bazi esiluthalayo khona. Cha, kufuze silufake ezandleni zabakithi ekhaya balunwampule. Kodwa pho kunganikezwa njani lo bazi nxa lusesilungwini? Kuzobachazelwa njani nxa abazali nabomndeni wethu bangakazwike neze la magama athize enhlobo yesiNgisi esibandayo? "Intersectionality" kungathiwa kanjani ngesiZulu? Ne-"Decolonial project", olwalubasiza abelungu bohulumeni wobandlululo ukuba baqhube nokwazi ukubacindezela abantu kanjani. Esinye isibonelo salona silona luka-A.T. Bryant olwalucwaninga ngombuso wesifundazwe saKwaZulu. Lona lwabonwa yizikhulu zaseKoloni yase-'Natali'. Abelungu bona babefuna ukwazi ukuthi umbuso wakhiwe njani KwaZulu ukuze bethole indlela ethile ukuhlaselelana ngayo abantu bomdabu. Lolu cwaningo lungaphakathi kwenhlobo yolwazi ethiwa ngesilungu i-'Anthropology', kumbé ngesiZulu ethi ukucwaninga mayelana namasiko, umnotho, ubukhosi, umbuso, inkolo nezinye izinto zonke ezenziwa ngabantu bonke bomhlaba. Luyingozi lo cwaningo ngoba lusaceliwa yizinkampani ezifuna ukuqhamuka ezweni labantu.

Ngakho ke kungathiwa ukuthi amagceka akwaNgqondonkulu ayizigodlo zezifundiswa ngoba lo lwazi lwazo lungavumelwe ukuphuma kuleso sigodlo ukuba luhambe kuze lufike ezandleni zabakithi abangemayo imithuba yokufunda. Ngaphesvu kwalokho kwaNgqondonkulu kufunekile ucwaningo oluyokhombisa izindlela ezintsha ukweba umhlaba wabantu ngazo.

Ziningi izinhlobo zalo lwazi olukwazi ukufuna ukuphapha 'maphiko. Lo lwazi nalo cwaningo olwenziwa yizifundiswa luyinhlobo njengempuphu ebizwa kamnandi. Ezinye zalezo zinhlobo zibalulekile zimi phezulu nokho zisakhona ezinye zazo ezingudaka ngenxa yokusetshenzisela kwazo ngobukoloni. Esigabeni esthazo kuzokhunjiswa isibonelo salokho. Okunye kungucwaningo lolo olungakwazi ukusiza abantu bomhlaba ekukhuphakeni kwabo, njengokuchaza izitho zezilwanyane ezithi "thorax", "abdomen" nezinye njalo. Angithi sidlinda ukuluchitha lo lwazi kodwa ngenkathi iyakhula indlala yabakithi pho ngabe sishayane ngenduku ngezitho zezilwanyane? Sinemithwalo yabazali nayokhokho bethu esidinge siyithwale.

Njengoba ngishilo ngenhla, kade kwaqanjwe ucwaningo

I si go do Se si fu ndi oxa Lu go nxa U lu si E ma go ke ri a kw ngqo nso rku lu

Those of us who are within the walls of the university often forget or do not

https://issuu.com/vernacnews4/docs/final_6



II. Case Studies – Beria

Key traits:

- Created in 1950s by Adam Tajir for writing the Zaghawa language of Sudan (Darfur), Chad, and diaspora; in 2000, Siddick Adam Issa modified the script
- Glyphs derive from camel brands
- Unicode preliminary proposal
- Groups use script differently; need community to reach consensus



Images from proposal by Lorna Priest [Evans], SIL:
<https://www.unicode.org/L2/L2008/08265-beria-branding.pdf>



After a script is in Unicode...

- Need fonts
- Need keyboard (or other input method)
- Need locale data for Unicode Common Locale Data Repository (CLDR)
- May need software updates
- May need to lobby companies to support language and script



Addendum: Outstanding Issues for Latin and Arabic scripts

AFRICAN LANGUAGES WRITTEN WITH LATIN SCRIPT

- Fonts and keyboards may be needed containing all letters
- Diacritics may be not properly placed in font

Wọ́ n ní ẹ̀bùn

AFRICAN LANGUAGES WRITTEN WITH ARABIC SCRIPT

- Fonts and keyboards may be needed containing all letters



***For assistance on these issues, contact Translation Commons
Project: krista@translationcommons.org***



Summary

- Vibrant continent with large number of modern and historic languages and scripts – including many recently created scripts.
- Getting a script into Unicode (and onto computers/devices) is a multi-step process, taking several years
- For unencoded scripts or problems, the Script Encoding Initiative can assist!



Thank you

Questions?

Contact: Debbie Anderson
dwanders@berkeley.edu



Poster of World's Writing Systems

<http://www.worldswritingsystems.org/poster.html>

Funding provided by National Endowment for the Humanities
grant (Universal Scripts Project) PR-268710-20

Thanks to Charles Riley (Yale)



#LocWorldWide46 Africa 