

DOCUMENTING AND PRESERVING LANGUAGES WITH UNICODE

A TALK ON CHARACTER ENCODING, FONTS, AND KEYBOARDS

DEBORAH ANDERSON, SCRIPT AD HOC CHAIR AND LEAD, SCRIPT ENCODING INITIATIVE, UCB

ANDREW GLASS, CHAIR OF UNICODE CLDR KEYBOARD SUBCOMMITTEE AND PRINCIPAL
PROGRAM MANAGER AT MICROSOFT

UNICODE WEBINAR, MAY 16 2023



TODAY'S PRESENTATION

- Debbie Anderson: Basics of Unicode character encoding

Based on slide set by Peter Constable (and Ken Whistler)

- Andrew Glass: Fonts and keyboards



IMPORTANCE OF UNICODE

- Now is a critical time to document and preserve languages (and their scripts) due to disappearance of languages and the loss of written materials in those languages.

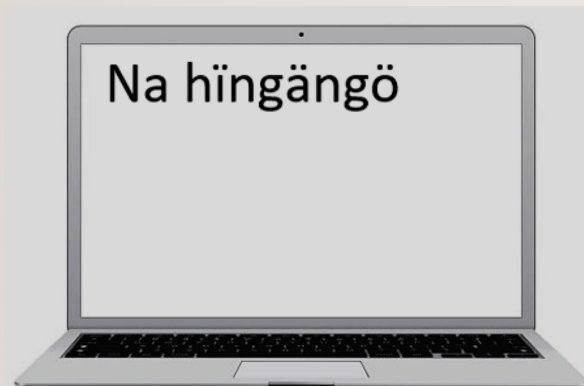


<https://www.endangeredlanguages.com/#/3/10.453/16.371/0/100000/0/low/mid/high/dormant/awakening/unknown>

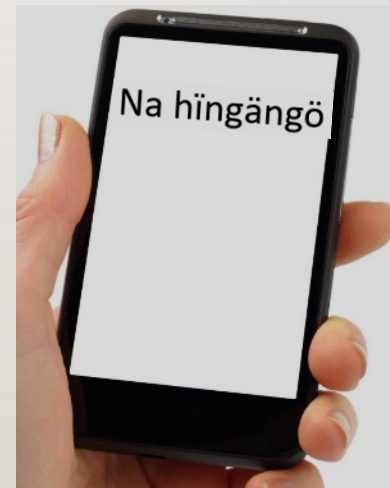


IMPORTANCE OF UNICODE

- Unicode underlies all electronic text communication today and hence is vital to preserving texts used to write languages (and to document them)



Sango language,
example from https://www.unicode.org/udhr/d/udhr_sag.html



IMPORTANCE OF UNICODE

- A critical first step is getting those characters that are used to write and describe languages into Unicode.

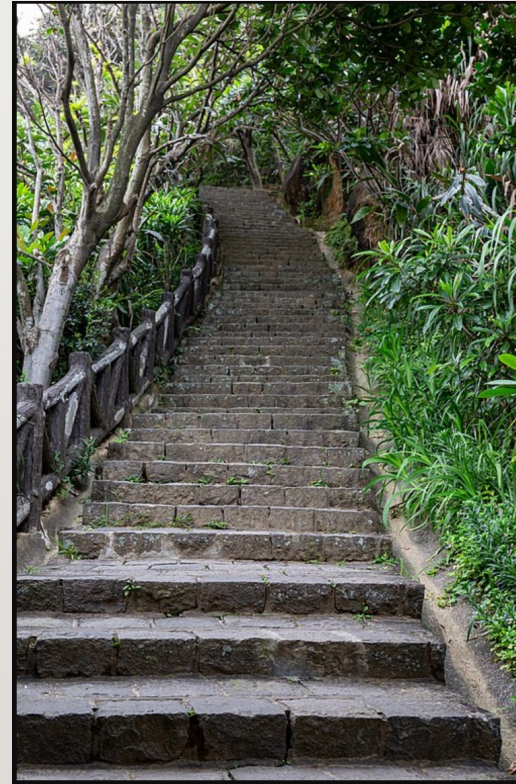


Image from: https://commons.wikimedia.org/wiki/File:20190415_Yehliu_geopark_stairs-2.jpg



THE SPECIAL-CHARACTER PROBLEM

- Linguists and language users work with all kinds of characters
 - International Phonetic Alphabet (IPA)
 - other phonetic systems and technical notation
 - transliterations
 - orthographies
 - scripts of living or extinct languages
 - developing orthographies

[aɪ p^hiː eɪ]

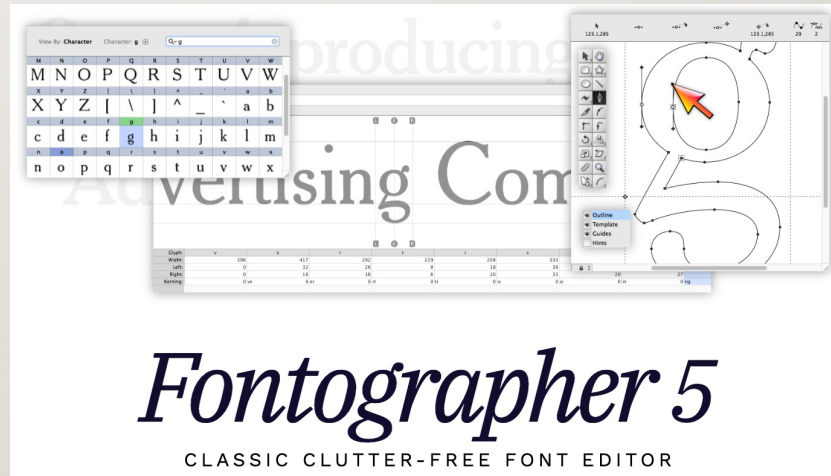


Bottom Image from Raymond Basquez Sr, Neal Ibanez & Myra Masiel-Zamora (2018) 'Atáaxum Alphabet. Great Oak Press, Pechanga Band of Luiseño Mission Indians.

<https://www.unicode.org/L2/L2022/22113r-two-latin-chars.pdf>; letter for Luiseno

THE SPECIAL-CHARACTER PROBLEM

- Language users' / Linguists' workaround: create custom-built fonts
 - change the shapes in the 'slots' to the shapes needed
 - custom fonts built with Fontographer, etc.



<https://www.fontlab.com/font-editor/fontographer/>





UNICODE –BASED STANDARD FONT

	008	009	00A	00B	00C	00D	00E	00F
0	XXX 0080	DCS 0090	NB SP 00A0	° 00B0	À 00C0	Đ 00D0	à 00E0	đ 00F0
1	XXX 0081	Pu1 0091	ı 00A1	± 00B1	Á 00C1	Ñ 00D1	á 00E1	ñ 00F1
2	BPH 0082	Pu2 0092	ç 00A2	² 00B2	Â 00C2	Ò 00D2	â 00E2	ò 00F2
3	NBH 0083	STS 0093	£ 00A3	³ 00B3	Ã 00C3	Ó 00D3	ã 00E3	ó 00F3
4	IND 0084	CCH 0094	¤ 00A4	´ 00B4	Ä 00C4	Ô 00D4	ä 00E4	ô 00F4
5	NEL 0085	MW 0095	¥ 00A5	µ 00B5	Å 00C5	Õ 00D5	å 00E5	õ 00F5
6	SSA 0086	SPA 0096	ı 00A6	¶ 00B6	Æ 00C6	Ö 00D6	æ 00E6	ö 00F6
7	ESA 0087	EPA 0097	§ 00A7	· 00B7	Ç 00C7	× 00D7	ç 00E7	÷ 00F7
8	HTS 0088	SOS 0098	¨ 00A8	¸ 00B8	È 00C8	Ø 00D8	è 00E8	ø 00F8
9	HTJ 0089	XXX 0099	© 00A9	¹ 00B9	É 00C9	Ù 00D9	é 00E9	ù 00F9
A	VTS 008A	SCI 009A	ª 00AA	º 00BA	Ê 00CA	Ú 00DA	ê 00EA	ú 00FA
B	PLD 008B	CSI 009B	« 00AB	» 00BB	Ë 00CB	Û 00DB	ë 00EB	û 00FB
C	PLU 008C	ST 009C	¼ 00AC	¼ 00BC	Ì 00CC	Ü 00DC	ì 00EC	ü 00FC
D	RI 008D	OSC 009D	½ 00AD	½ 00BD	Í 00CD	Ý 00DD	í 00ED	ý 00FD
E	SS2 008E	PM 009E	® 00AE	¾ 00BE	Î 00CE	Þ 00DE	î 00EE	þ 00FE
F	SS3 008F	APC 009F	— 00AF	¿ 00BF	Ï 00CF	ß 00DF	ï 00EF	ÿ 00FF

NAROM FONT “HACKED FONT”

	008	009	00A	00B	00C	00D	00E	00F
0	a 204C	d 204D	L 00A0	À 00C0	T 00C0	à 00E0	□ 00F0	
1	Ă 204E	‘ 2049	ğ 00A1	İ 00C1	ş 00C1	t 00E1	z 00F1	□ 00F1
2	ă 204A	’ 2048	Ÿ 00A2	Î 00C2	Ŵ 00C2	â 00E2	□ 00F2	
3	ǎ 2042	“ 2043	ÿ 00A3	Ŷ 00C3	Ŷ 00C3	- 00E3	□ 00F3	
4	Ț 204E	” 2040	Ț 00A4	Ț 00C4	Ț 00C4	ä 00E4	ô 00F4	
5	Ț 204A	Ț 2042	Ț 00A5	Ț 00C5	Ț 00C5	ε 00E5	□ 00F5	
6	B 204D	d 2043	H 00A6	¶ 00C6	§ 00C6	Z 00E6	ǎ 00F6	□ 00F6
7	b 2041	Ț 2044	h 00A7	Ț 00C7	Ț 00C7	ç 00E7	÷ 00F7	
8	Ț 00C8	ə 00C8	Ț 00A8	Ț 00C8	Ț 00C8	z 00E8	è 00F8	° 00F8
9	Ț 00C9	ə 00C9	Ț 00A9	Ț 00C9	Ț 00C9	é 00E9	ù 00F9	
A	Ț 00CA	Ț 00CA	Ț 00AA	Ț 00CA	Ț 00CA	ê 00EA	□ 00FA	
B	Ț 00CB	Ț 00CB	Ț 00AB	Ț 00CB	Ț 00CB	ë 00EB	û 00FB	
C	Ț 00CC	Ț 00CC	Ț 00AC	Ț 00CC	Ț 00CC	ü 00EC	□ 00FC	
D	Ț 00CD	Ț 00CD	Ț 00AD	Ț 00CD	Ț 00CD	í 00ED	ý 00FD	
E	Ț 00CE	Ț 00CE	Ț 00AE	Ț 00CE	Ț 00CE	î 00EE	þ 00FE	
F	Ț 00CF	Ț 00CF	Ț 00AF	Ț 00CF	Ț 00CF	ï 00EF	□ 00FF	



THE SPECIAL-CHARACTER PROBLEM

- “Hacked font” relies on font creator’s private knowledge

- can’t exchange data
- apps don’t behave right
- can’t switch fonts
- *no inter-operability*

Intended text (using nonstandard font)

Υriy tasαet.

How same text may appear to others



çri£ tasa?et.



THE SPECIAL-CHARACTER PROBLEM


- Solution: Unicode
 - a single, international standard
 - universal usage
 - inter-operability
 - comprehensive coverage
 - all the characters everyone needs



*that's the
goal, at
least*



THE UNICODE CHARACTER REPERTOIRE



NEWS, ANNOUNCEMENTS, RELEASE INFO, AND CALENDAR UPDATES
FROM THE UNICODE CONSORTIUM

TUESDAY, SEPTEMBER 13, 2022

Announcing The Unicode® Standard, Version 15.0

Version 15.0 of the Unicode Standard is now available, including the core specification, annexes, and data files. This version adds 4,489 characters.

ᐃ	ᐣ	ᐤ
1E4D5	1E4E5	1E4F5

LINKS OF INTEREST

- [What is Unicode?](#)
- [The Unicode Consortium](#)
- [Archived Announcements](#)

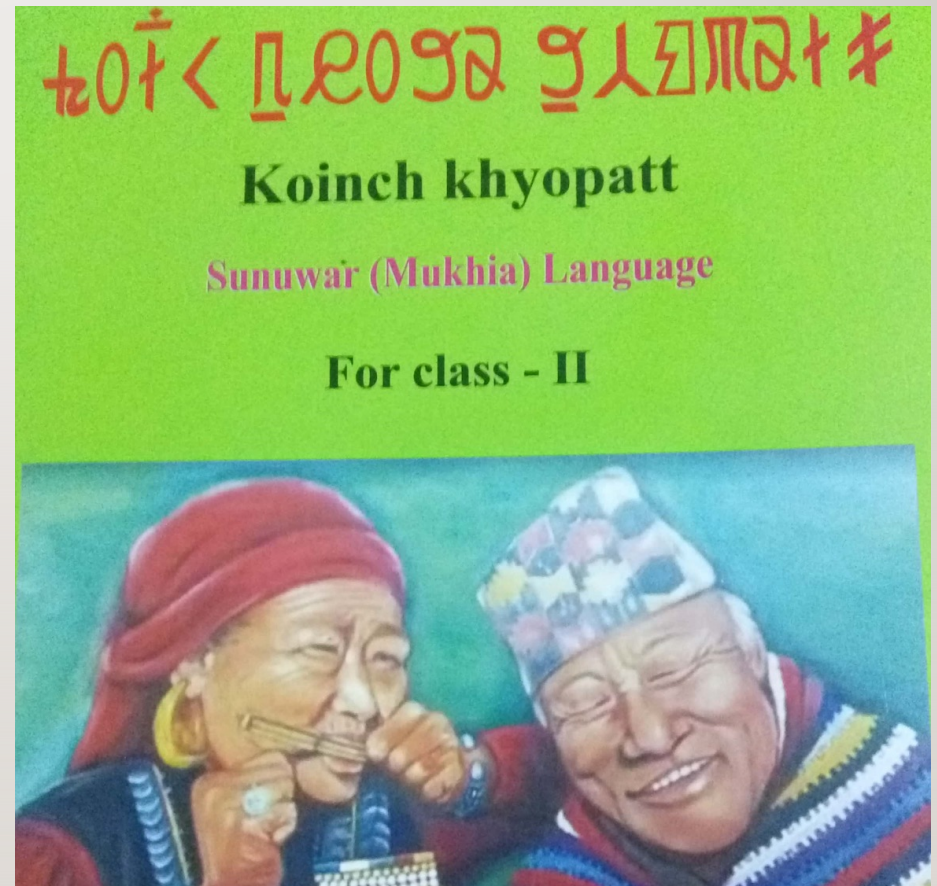
- Unicode 15.0 (the current version)
 - 149,186 characters
 - 161 scripts
 - Chinese characters: > 98,000
 - lots of symbols
 - room for 825,000 characters more!



<http://blog.unicode.org/2022/09/announcing-unicode-standard-version-150.html>

THE UNICODE CHARACTER REPERTOIRE

- Upcoming full version
 - Unicode 16.0 is scheduled to be published Sept. 2024
- Beyond 16.0: Work in progress
 - Many additional symbols and scripts, including Seal script, Jurchen, and Mayan Hieroglyphs



Sunuwar script (scheduled for Unicode 16.0)

THE BASICS: THE UNICODE CHARACTER REPERTOIRE (I)

- Referring to a Unicode character
 - two unique identifiers:
 - a name : **LATIN SMALL LETTER ESH WITH DOUBLE BAR**
 - a number — a “code point” **U+1DF0B**

(representative glyph  can be changed)



THE BASICS: THE UNICODE CHARACTER REPERTOIRE (2)

- Code charts (on Unicode website)
 - graphic chart
 - names list
 - may include some additional info about identity and purpose of character

Gurmukhi

	0A0	0A1	0A2	0A3	0A4	0A5	0A6	0A7
0		ਐ 0A10	ਠ 0A20	ਰ 0A30	ੀ 0A40			ੰ 0A70
1	ੳ 0A01		ਕ 0A21		ੳ 0A41			ੳ 0A71
2	ੳ 0A02		ਚ 0A22	ਲ 0A32	ੳ 0A42			ੳ 0A72

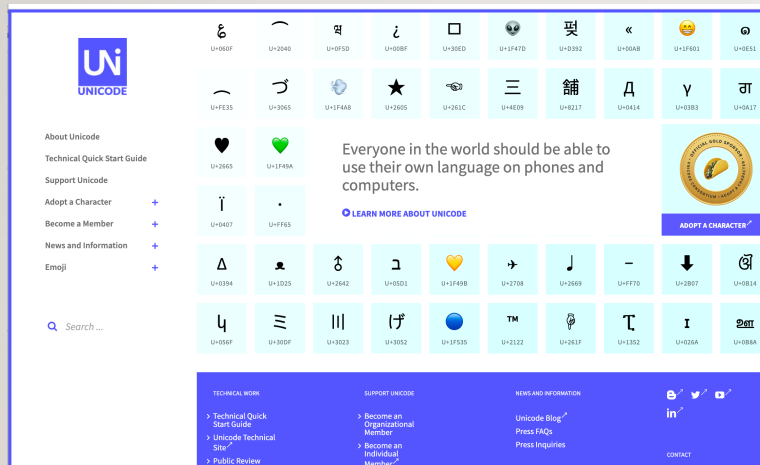
0A01	Gurmukhi	0A6C
Based on ISCII 1988		
Various signs		
0A01 ੳ GURMUKHI SIGN ADAK BINDI	0A32 ਲ GURMUKHI LETTER LA	
0A02 ੳ GURMUKHI SIGN BINDI	0A33 ਲ GURMUKHI LETTER LLA	
0A03 ੳ GURMUKHI SIGN VISARGA	= 0A32 ਲ 0A3C ੳ	
Independent vowels		
0A05 ਅ GURMUKHI LETTER A	0A34 ੳ <reserved>	
0A06 ਆ GURMUKHI LETTER AA	0A35 ਵ GURMUKHI LETTER VA	
0A07 ਇ GURMUKHI LETTER I	0A36 ਸ਼ GURMUKHI LETTER SHA	
0A08 ਈ GURMUKHI LETTER II	= 0A38 ਸ਼ 0A3C ੳ	
0A09 ਉ GURMUKHI LETTER U	0A37 ੳ <reserved>	
0A0A ਊ GURMUKHI LETTER UU	0A38 ਸ਼ GURMUKHI LETTER SA	
	0A39 ਹ GURMUKHI LETTER HA	
	Various signs	
	0A3C ੳ GURMUKHI SIGN NUKTA	
	* for extending the alphabet to new letters	

From <https://www.unicode.org/charts/PDF/U0A00.pdf>



Unicode Consortium website

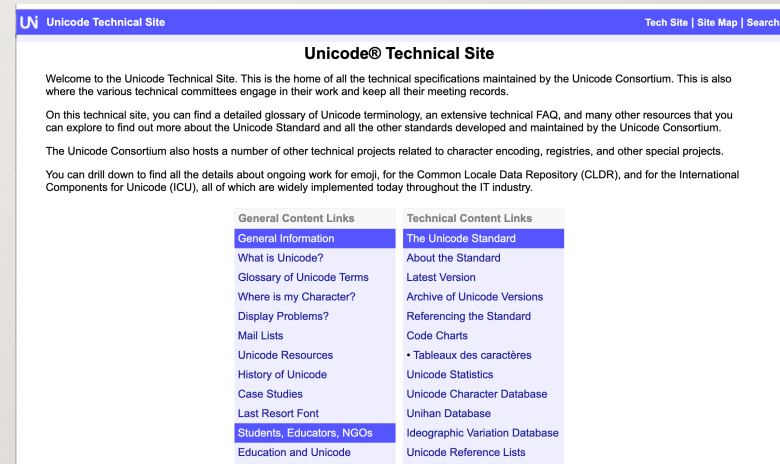
GENERAL WEBSITE



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

/ TECHNICAL WEBSITE

<http://unicode.org/main.html>

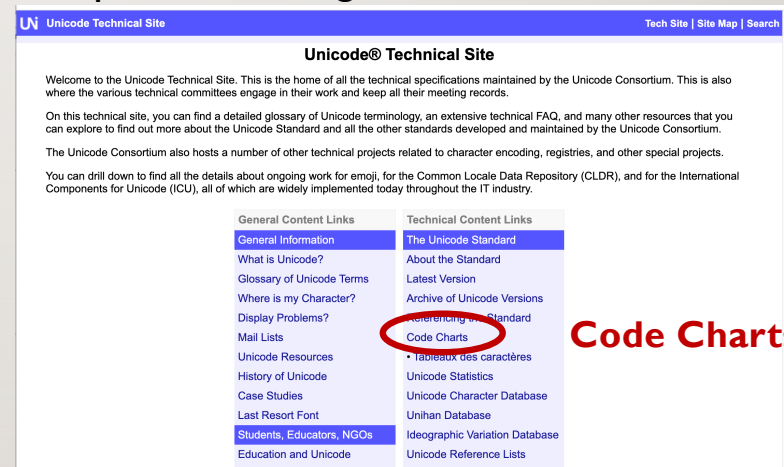


Unicode Consortium website

TECHNICAL SITE (LINK TO CODE CHARTS PAGE)

TECHNICAL SITE:

<http://unicode.org/main.html>



Unicode Consortium website

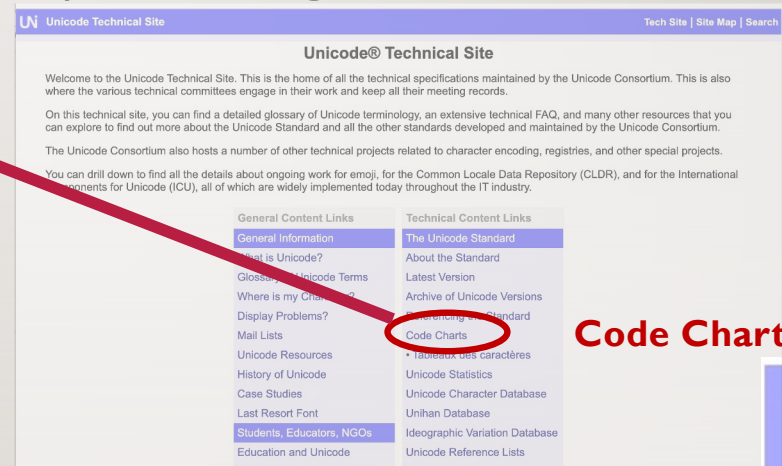
TECHNICAL SITE (LINK TO CODE CHARTS PAGE)

CODE CHARTS PAGE

<https://unicode.org/charts>

Unicode 15.0 Character Code Charts			
SCRIPTS SYMBOLS & PUNCTUATION NAME INDEX			
Find chart by hex code: <input type="text"/> <input type="button" value="Go"/> Help Conventions Terms of Use			
Scripts			
European Scripts	African Scripts	South Asian Scripts	Indonesian & Philippine Scripts
Armenian	Adlam	Ahom	Balinese
Armenian Ligatures	Bamum	Bengali and Assamese	Batak
Carian	Bamum Supplement	Bhaiksuki	Buginese
Caucasian Albanian	Bassa Vah	Brahmi	Buhid
Cypriot Syllabary	Coptic	Chakma	Hanunoo
Cypro-Minoan	Coptic in Greek block	Devanagari	Javanese
Cyrillic	Coptic Epact Numbers	Devanagari Extended	Kawi
Cyrillic Supplement	Egyptian Hieroglyphs	Devanagari Extended-A	Makasar
Cyrillic Extended-A	Egyptian Hieroglyph Format Controls	Dives Akuru	Rejang
Cyrillic Extended-B	Ethiopic	Dogra	Sundanese
Cyrillic Extended-C	Ethiopic Supplement	Grantha	Sundanese Supplement
Cyrillic Extended-D	Ethiopic Extended	Gujarati	Tagalog
Elbasan	Ethiopic Extended-A	Gunjala Gondi	Tagbanwa
Georgian	Ethiopic Extended-B	Gurmukhi	East Asian Scripts
Georgian Extended	Medefaidrin	Kaithi	Bopomofo
Georgian Supplement	Mende Kikakui	Kannada	Bopomofo Extended
Glagolitic	Merotic	Kharoshthi	CJK Unified Ideographs (Han) (35MB)
Glagolitic Supplement	Merotic Cursive	Khojki	CJK Extension A (6MB)
Gothic	Merotic Hieroglyphs	Khudawadi	CJK Extension B (40MB)
Greek	N'Ko	Lepcha	CJK Extension C (3MB)
Greek Extended	Osmanya	Limbu	CJK Extension D
Ancient Greek Numbers	Tifinagh	Mahajani	CJK Extension E (3.5MB)
Latin	Val	Malayalam	CJK Extension F (4MB)
Basic Latin (ASCII)	Middle Eastern Scripts	Masaram Gondi	CJK Extension G (2MB)
Latin-1 Supplement	Anatolian Hieroglyphs	Meetei Mayek	CJK Extension H (2.5MB)
Latin Extended-A	Arabic	Meetei Mayek Extensions	CJK Extension I (2.5MB)

<http://unicode.org/main.html>



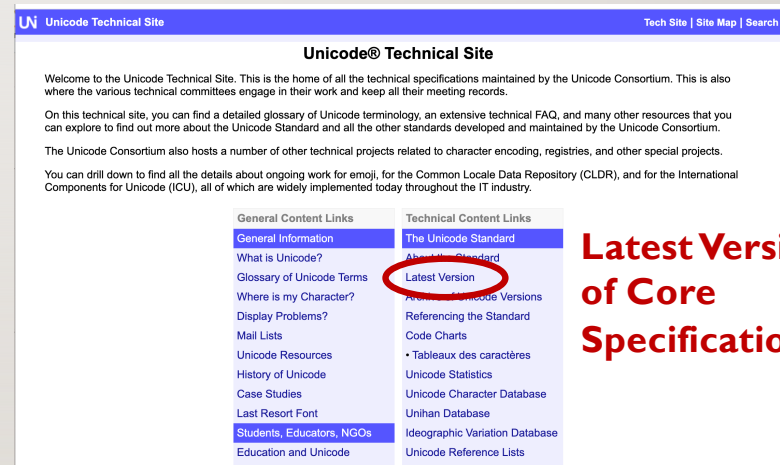
Code Charts



Unicode Consortium website

TECHNICAL SITE (LINK TO “LATEST VERSION”)

<http://unicode.org/main.html>



**Latest Version
of Core
Specification**



Unicode Consortium website

TECHNICAL SITE (LINK TO “LATEST VERSION”)

LATEST VERSION OF “CORE SPEC”

Unicode 15.0.0

Tech Site | Site Map | Search

Unicode® 15.0.0
2022 September 13 (Announcement)

This page summarizes the important changes for the Unicode Standard, Version 15.0.0. This version supersedes all previous versions of the Unicode Standard.

A. Summary

Unicode 15.0 adds 4,489 characters, for a total of 149,186 characters. These additions include 2 new scripts, for a total of 161 scripts, along with 20 new emoji characters, and 4,193 CJK (Chinese, Japanese, and Korean)

15.0.0 Core Specification

All Chapters and Appendices Together:
• Full Text pdf for Viewing (14 MB)

15.0.0 Front Matter

Title and Copyright
Contents
Unicode 15.0 Web Bookmarks
Preface

15.0.0 Chapters

- 1 Introduction
- 2 General Structure
- 3 Conformance
- 4 Character Properties
- 5 Implementation Guidelines
- 6 Writing Systems and Punctuation
- 7 Europe-I
- 8 Europe-II

“Core Spec” includes

- general introduction,
- conformance and implementation guidelines
- chapters on all the characters (arranged by script / class of characters)

<http://unicode.org/main.html>

Unicode Technical Site

Tech Site | Site Map | Search

Unicode® Technical Site

Welcome to the Unicode Technical Site. This is the home of all the technical specifications maintained by the Unicode Consortium. This is also where the various technical committees engage in their work and keep all their meeting records.

On this technical site, you can find a detailed glossary of Unicode terminology, an extensive technical FAQ, and many other resources that you can explore to find out more about the Unicode Standard and all the other standards developed and maintained by the Unicode Consortium.

The Unicode Consortium also hosts a number of other technical projects related to character encoding, registries, and other special projects.

You can drill down and find all the details about ongoing work for emoji, for the Common Locale Data Repository (CLDR), and for the International Components for Unicode (ICU), all of which are widely implemented today throughout the IT industry.

General Content Links

- General Introduction
- What is Unicode?
- Glossary of Unicode Terms
- Where is my Character?
- Display Problems?
- Mail Lists
- Unicode Resources
- History of Unicode
- Case Studies
- Last Resort Font
- Students, Educators, NGOs
- Education and Unicode

Technical Content Links

- The Unicode Standard**
- Latest Version
- Previous Unicode Versions
- Referencing the Standard
- Code Charts
- Tableaux des caractères
- Unicode Statistics
- Unicode Character Database
- UniHan Database
- Ideographic Variation Database
- Unicode Reference Lists

**Latest Version
of Core
Specification**



THE BASICS: THE UNICODE CHARACTER REPERTOIRE (3)

- Basic character identity
 - name and code point are immutable, but glyph can be changed (within limits)
- Many other properties that define complete identity and semantics
 - case, case mappings, general category, behavior for text segmentation, etc.



Image from <https://typeclasses.com/beginner-crash-course/map>



THE BASICS: THE UNICODE CHARACTER REPERTOIRE (4)

- Organization in Unicode code space
 - characters organized into blocks of related characters
 - typically, by script
 - characters for a writing system may be in multiple, non-contiguous blocks
 - punctuation may be shared across different scripts

Georgian

Georgian Extended
Georgian Supplement

Glagolitic

Glagolitic Supplement

Gothic

Greek

Greek Extended
Ancient Greek Numbers

Latin

Basic Latin (ASCII)
Latin-1 Supplement
Latin Extended-A
Latin Extended-B
Latin Extended-C
Latin Extended-D
Latin Extended-E
Latin Extended-F
Latin Extended-G
Latin Extended Additional
Latin Ligatures
Fullwidth Latin Letters
IPA Extensions
Phonetic Extensions
Phonetic Extensions Supplement

Linear A

Linear B

Linear B Syllabary
Linear B Ideograms



<https://www.unicode.org/charts/>

UNICODE DESIGN PRINCIPLES (I)

- Unification
 - unify characters within scripts
 - Same script, different languages: unify

cat, chat, gato, Katze

N.B. Unicode encodes *scripts*, not languages

- Different scripts: don't unify

ABCD ABΓΔ AБВГ



UNICODE DESIGN PRINCIPLES (2)

- Characters, not *glyphs*
 - character: unit of abstract, textual information
 - glyph: graphic image used for presentation of a character

character: LATIN SMALL LETTER A

Glyphs:



UNICODE DESIGN PRINCIPLES (3)

- Characters, not *glyphs*
 - characters : glyphs may not be 1:1

ARABIC LETTER HEH



- Unicode assumes applications will deal with display
 - font + rendering engine



UNICODE DESIGN PRINCIPLES (4)

- Characters may not be the same as *text elements in a writing system*

ç

< U+0063 LATIN SMALL LETTER C,
U+0303 COMBINING TILDE >

ch

< U+0063 LATIN SMALL LETTER C,
U+0068 LATIN SMALL LETTER H >



UNICODE DESIGN PRINCIPLES (5)

- Dynamic composition
 - complex text elements can be composed dynamically from sequences of characters



< U+0063 LATIN SMALL LETTER C,
U+0324 COMBINING DIAERESIS BELOW,
U+032A COMBINING BRIDGE BELOW,
U+0303 COMBINING TILDE,
U+0306 COMBINING BREVE,
U+0301 COMBINING ACUTE ACCENT >



UNICODE DESIGN PRINCIPLES (6)

- Support for legacy standard character sets
 - all standards in wide usage as of May 1993
 - required many compromises with other design principles
 - many “presentation” and pre-composed characters

U+00E1 “á” LATIN SMALL LETTER A WITH ACUTE



ALTERNATE REPRESENTATIONS (DYNAMIC COMPOSITION)

- Combining mark sequences
 - if two marks occupy similar space:
 - “stack” in order
 - different order of marks are significant

$a + \tilde{\circ} + \hat{\circ} \rightarrow \hat{\tilde{a}}$

$a + \hat{\circ} + \tilde{\circ} \rightarrow \tilde{\hat{a}}$



ALTERNATE REPRESENTATIONS (DYNAMIC COMPOSITION)

- Combining mark sequences
 - if two marks don't interact typographically:
 - different orders look the same
 - no meaningful difference

$a + \hat{\circ} + \circ \rightarrow \hat{a}$

$a + \circ + \hat{\circ} \rightarrow \hat{a}$



ALTERNATE REPRESENTATIONS - NORMALIZATION

- A text element may have several equivalent representations:
 - Canonically equivalent: considered the same

$$B + \ddot{A} \equiv B + A + \ddot{\circ}$$

- Compatibility equivalent: can mean the same in some, but not all, circumstances

$$2 + \frac{1}{4} \approx 2 + 1 + / + 4$$

Example from <https://www.unicode.org/versions/Unicode15.0.0/ch02.pdf>



GETTING PRACTICAL: ENCODING CHARACTERS AND SCRIPTS (I)

- What do you need to be able to use Unicode?
 - Encode characters and scripts
 - Process (takes at least 2 years)
 - Write a Unicode proposal
 - Review by the Script Ad Hoc (for non-emoji/non-CJK)
 - Approval by the Unicode Technical Committee
 - Publish in a version of Unicode

Unicode request for two BMP Latin characters L2/22-113R

Kirk Miller, kirkmiller@gmail.com

2022 July 09

This request is for two Latin letters used in basic literacy material. We request that they be added to the BMP.

The Luiseño community of California (Uto-Aztecan, ISO [LUI]) uses the casing pair <§ §> in their literacy and language-revival program. Myra Masiel-Zamora, Curator of Pechanga Cultural Resources in Temecula, has asked that we adopt <§ §> into Unicode to aid those efforts. These two letters are also used in descriptive Cupeño material (ISO [CUP]).

Thanks to Deborah Anderson of the Universal Scripts Project for her assistance.

Characters

§ A7CC LATIN CAPITAL LETTER S WITH DIAGONAL STROKE. Figures 1–2.
§ A7CD LATIN SMALL LETTER S WITH DIAGONAL STROKE. Figures 1–3.

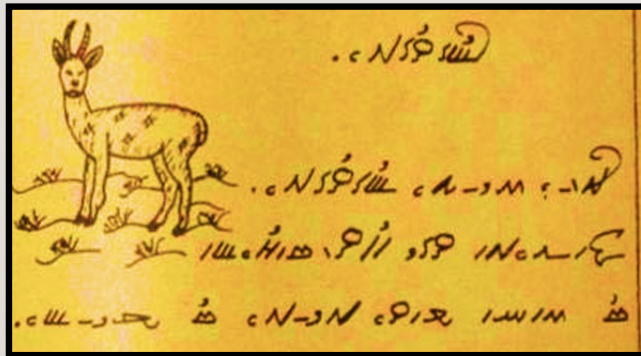
Properties

A7CC; LATIN CAPITAL LETTER S WITH DIAGONAL STROKE;
Lu; 0; L; ; ; ; N; ; ; A7CD;

Example from <https://www.unicode.org/L2/L2022/22113r-two-latin-chars.pdf>

GETTING PRACTICAL: ENCODING CHARACTERS AND SCRIPTS (2)

- Write a Unicode proposal, using proposal templates:
 - For new character additions (use document [L2/23-104](#))
 - For new scripts (use document [L2/23-105](#))



<https://www.unicode.org/L2/L2012/L2139-n4261-garay.pdf>

Garay script



GETTING PRACTICAL: ENCODING CHARACTERS AND SCRIPTS (3)

- Proposal review process by Script Ad Hoc:
 - Proposals should be complete
 - Proposals need to make a case why characters are needed
 - New scripts need to demonstrate usage with list of publications in script (not written by script creator); repertoire should be stable for several years



Garay script

<https://www.unicode.org/L2/L2012/12139-n4261-garay.pdf>

GETTING PRACTICAL: ENCODING CHARACTERS AND SCRIPTS (4)

- For a language with no orthography
 - Suggestion: Develop an orthography using Unicode characters

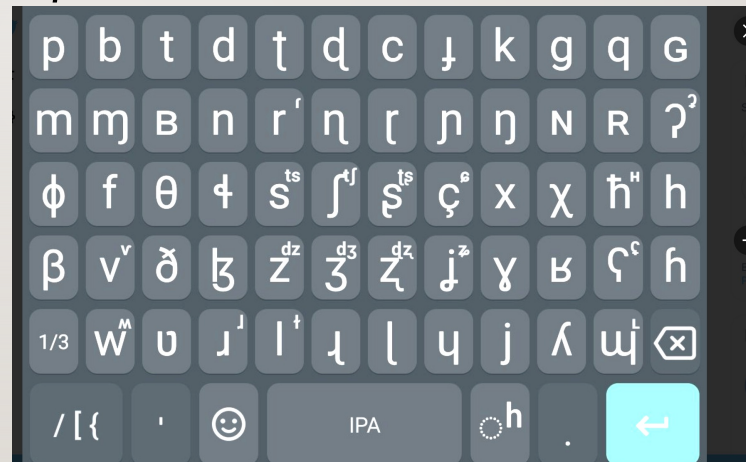
See [Unicode Technical Note #19](#) “Recommendations for Creating New
Orthographies”:

Unicode Technical Notes	
Unicode® Technical Note #19	
RECOMMENDATIONS FOR CREATING NEW ORTHOGRAPHIES	
Version	2
Authors	Deborah Anderson, with Rick McGowan and Ken Whistler (and incorporating comments by Lorna Priest)
Date	April 3, 2023
This Version	http://www.unicode.org/notes/tn19/tn19-2.html
Previous Version	http://www.unicode.org/notes/tn19/tn19-1.html
Latest Version	http://www.unicode.org/notes/tn19/



GETTING PRACTICAL: AFTER ENCODING CHARACTERS AND SCRIPTS

- What do you need to be able to use Unicode?
 - *Use Unicode characters and scripts*
 - **Fonts**
 - **Keyboards**
 - **Software support**



Gboard IPA keyboard



GETTING INVOLVED IN CHARACTER ENCODING

- Templates for script and character proposals:
 - <https://www.unicode.org/L2/L2023/23I04-addl-script-template-april2023.pdf>
 - <https://www.unicode.org/L2/L2023/23I05-new-script-template-april2023.pdf>
- Submitting character proposals: www.unicode.org/pending/proposals.html
- Guidelines on creating an orthography (from Unicode perspective):
 - <https://www.unicode.org/notes/tn19/>
- FAQs on character proposals: https://www.unicode.org/faq/char_proposal.html
- Script Ad Hoc description: <https://www.unicode.org/consortium/scriptadhoc.html>
- Unicode YouTube channel: <https://www.youtube.com/@unicode/about>



THANK YOU

- Support for Universal Script Project (/Script Encoding Initiative) comes from NEH grant PR-268710-20 and donations.
- Script Encoding Initiative Website: <http://linguistics.berkeley.edu/sei>
- For questions (after this webinar): please use Unicode Feedback form: <https://www.unicode.org/reporting.html>

Gunjala Gondi
script

