

# Unicode Fonts and Keyboards

Andrew Glass

# Unicode



Listening

Speaking

Human  
language

Reading

Writing

# Unicode



Listening

Speaking

Human

Reading

Writing

Computer

Storing

Transmitting

Computing

# Unicode



Human

Reading

Writing

Computer

Unicode

Storing

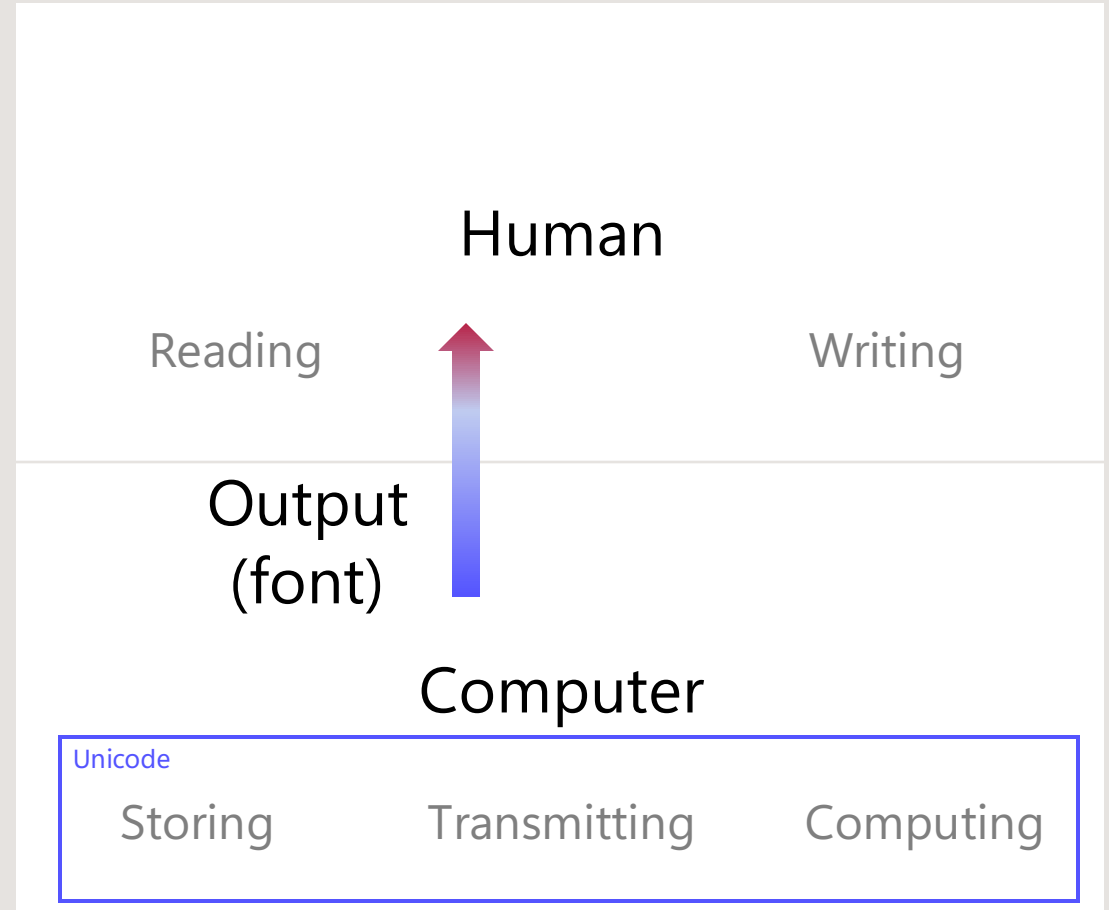
Transmitting

Computing

# Unicode



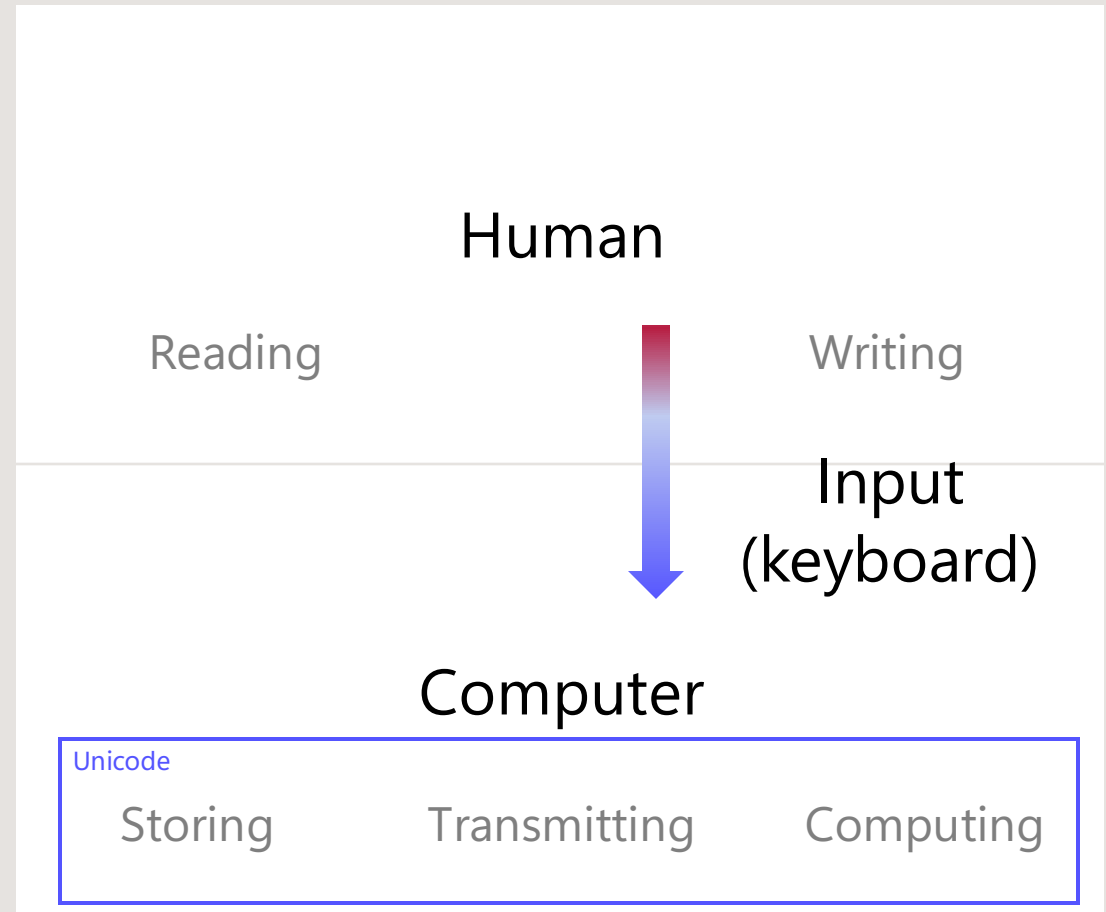
- Visualizing the orthographic needs of a language is the job of a font



# Unicode



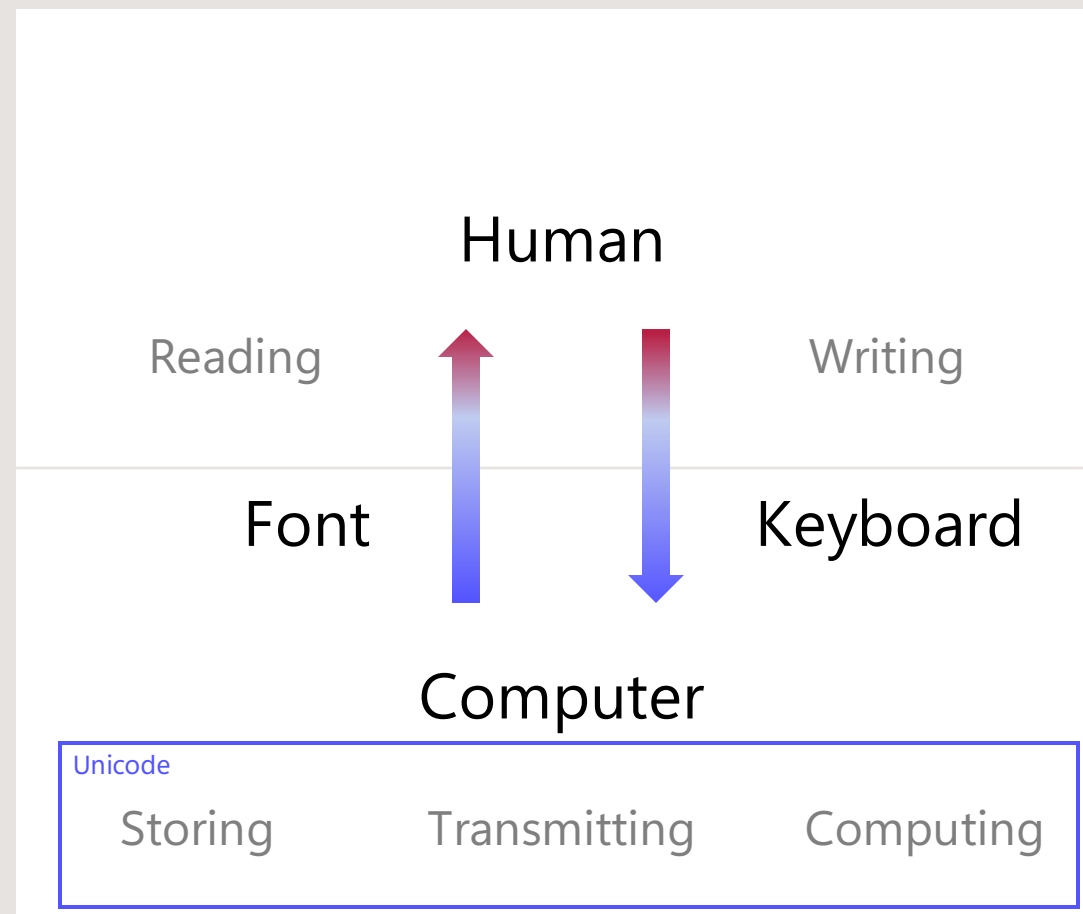
- Entering language data is the responsibility of a “Keyboard”



# Unicode



- Fonts and keyboards are complementary systems—both are essential to support the successful digitization of a language.
- They should be developed together.



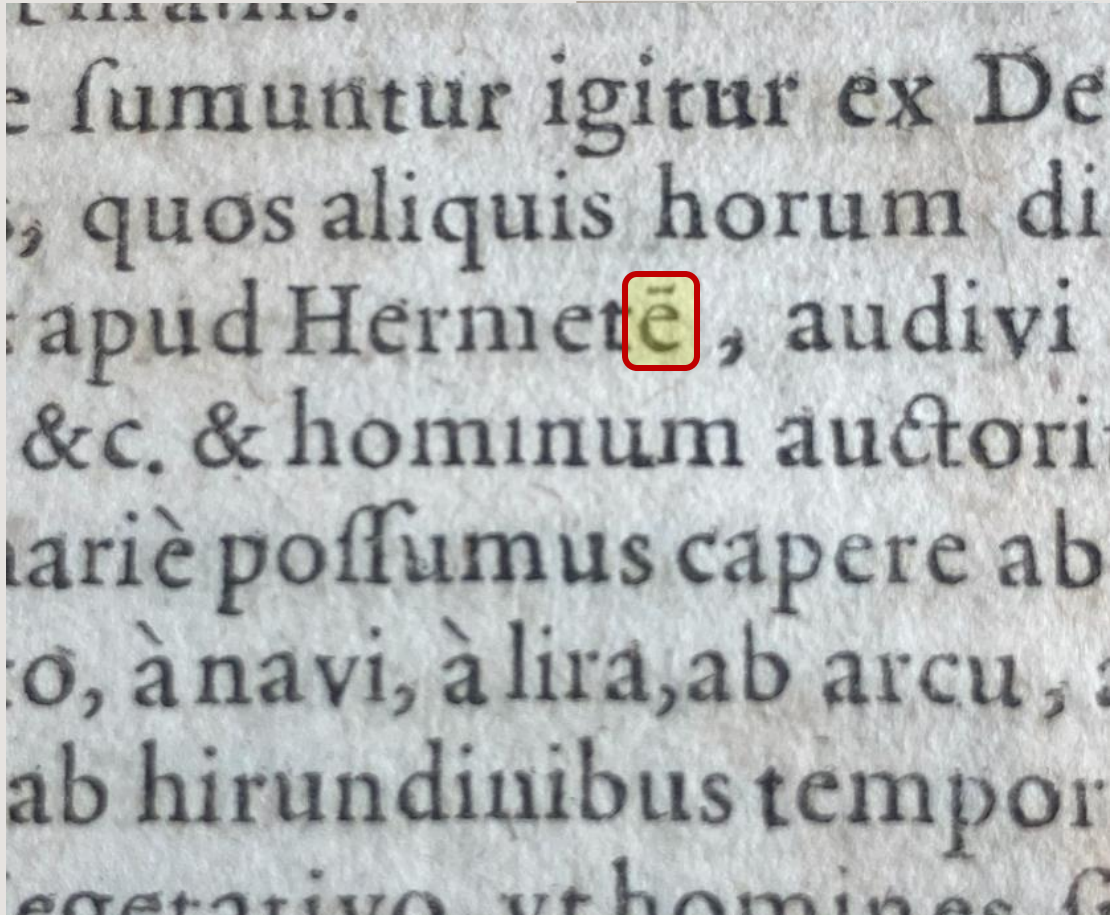
# Fonts



- Alternate forms • combining diacritics • ligatures • justification • line-breaking • pre-base reordering • stacking • superscription • positional alternates • variation selectors
- block formatting • mirroring • left-to-right • right-to-left • TTB LTR • TTB RTL •



# Fonts



- Character mapping

0100

Latin Extended-A

017F

	010	011	012	013	014	015	016	017
0	Ā 0100	Ð 0110	Ġ 0120	İ 0130	ı 0140	Ŏ 0150	Š 0160	Ů 0170
1	ā 0101	đ 0111	ġ 0121	ı 0131	Ł 0141	ő 0151	š 0161	ů 0171
2	Ă 0102	Ē 0112	Ģ 0122	IJ 0132	ł 0142	Œ 0152	Ț 0162	Ț 0172
3	ă 0103	ē 0113	ģ 0123	ij 0133	Ń 0143	œ 0153	ț 0163	Ț 0173
4	Ą 0104	Ĕ 0114	Ĥ 0124	Ĵ 0134	ń 0144	Ŕ 0154	Ť 0164	Ű 0174
5	ą 0105	ĕ 0115	ĥ 0125	ĵ 0135	Ņ 0145	ŕ 0155	ť 0165	ű 0175
6	Ć 0106	Ė 0116	Ħ 0126	Ƙ 0136	ņ 0146	Ŗ 0156	Ŧ 0166	Ŷ 0176
7	ć 0107	ė 0117	ħ 0127	ƙ 0137	Ň 0147	ŗ 0157	ŧ 0167	ŷ 0177
8	Ĉ 0108	Ę 0118	Ĩ 0128	κ 0138	ň 0148	Ř 0158	Ũ 0168	Ỳ 0178

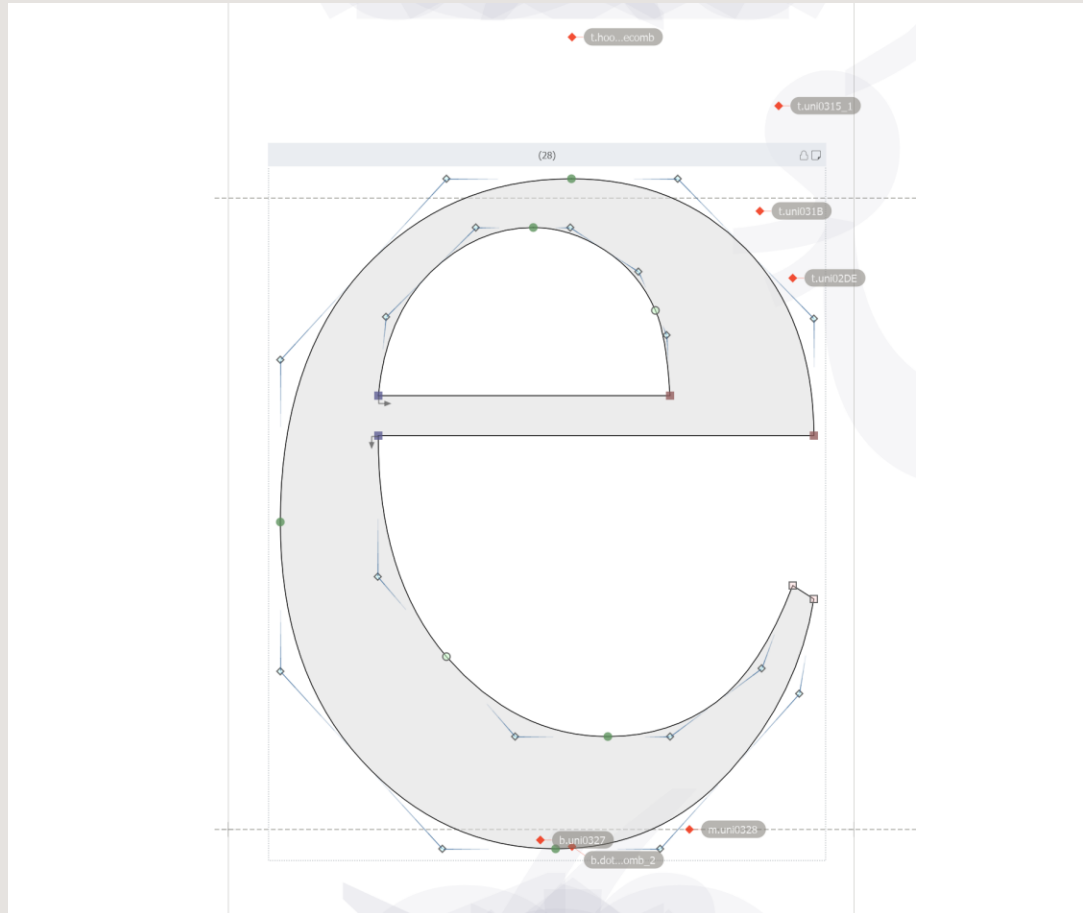
# Fonts

... in annis.  
... fumuntur igitur ex Deo  
... quos aliquis horum dii  
... apud Hermet<sup>e</sup>, audiui  
... c. & hominum auctoritati  
... riè possumus capere ab  
... , à navi, à lira, ab arcu,  
... b hirundinibus tempore  
... egativo, ut homines f...

- Character mapping

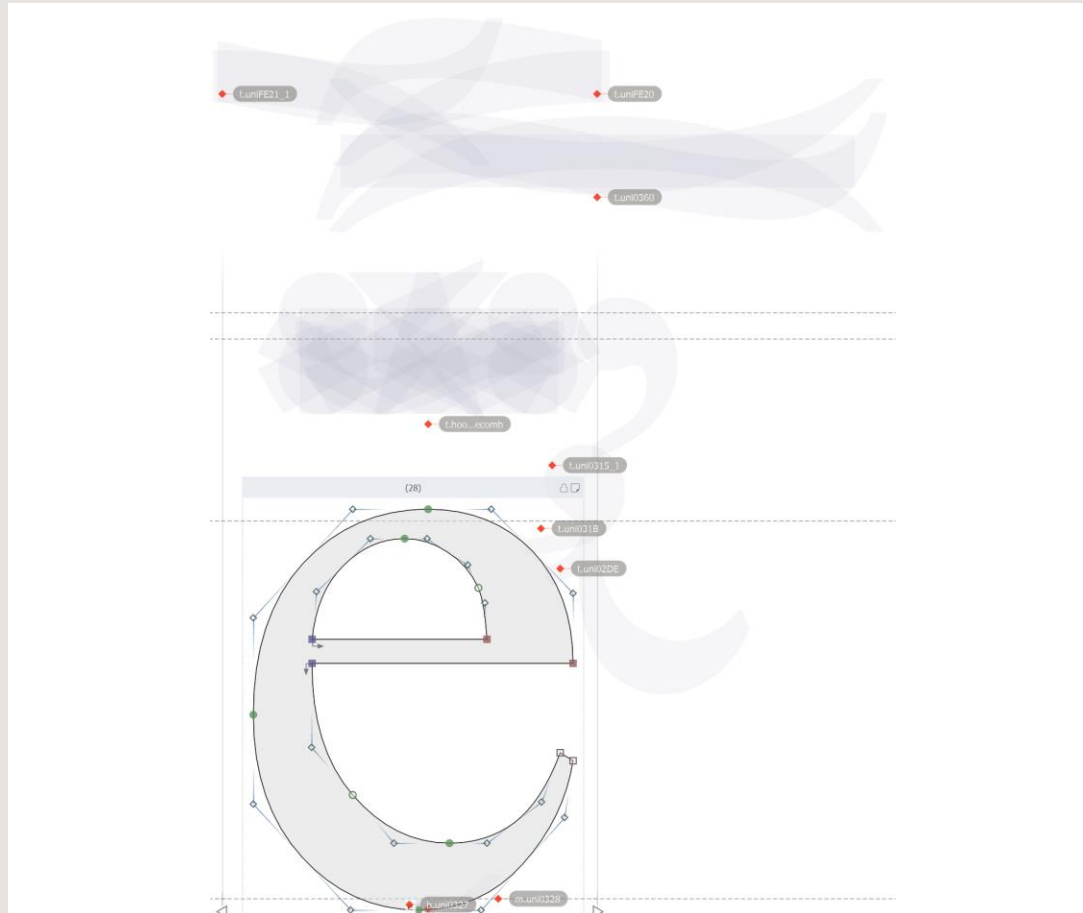
00FC	00FD	00FE	00FF	0100
ü	ý	þ	ÿ	Ā
0106	0107	0108	0109	010A
Ć	ć	Ĉ	ĉ	Č
0110	0111	0112	0113	0114
Đ	đ	Ē	ē	Ě
011A	011B	011C	011D	011E
Ě	ě	Ĝ	ĝ	Ğ
0124	0125	0126	0127	0128
Ĥ	ĥ	Ħ	ħ	Ĩ

# Fonts



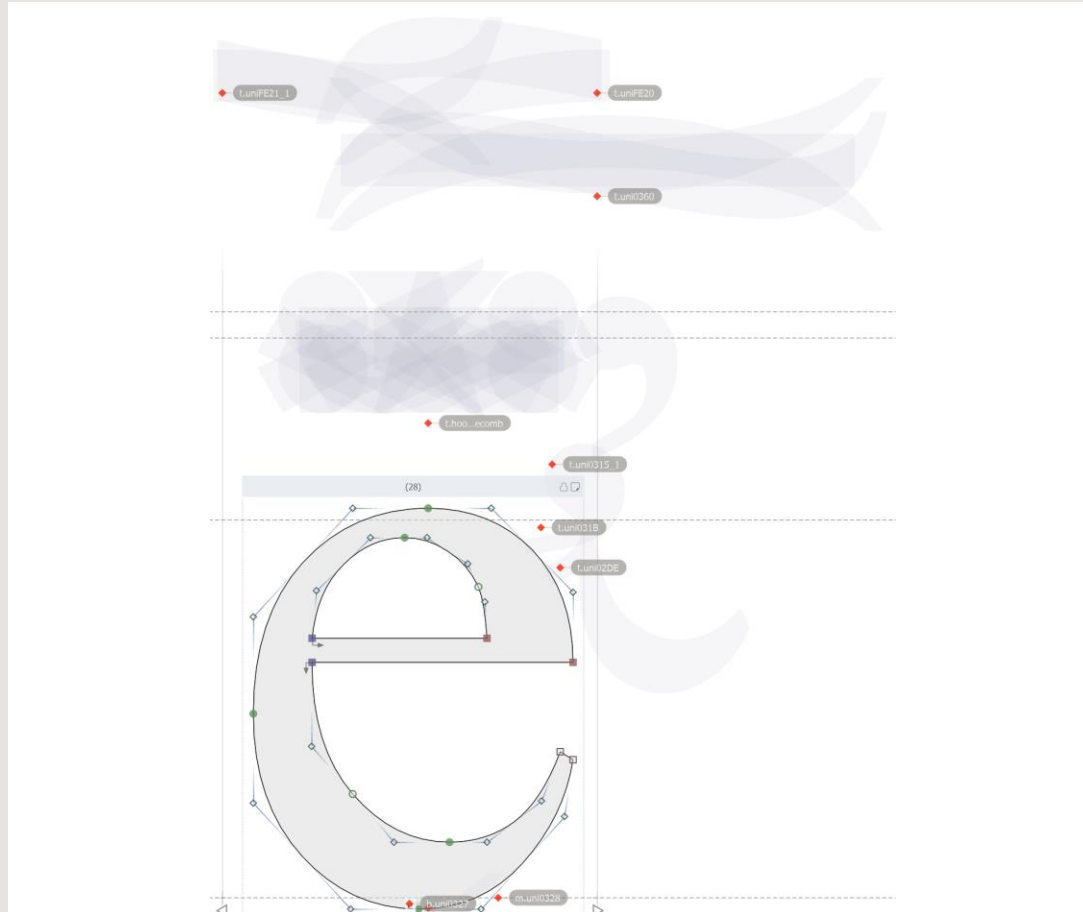
- Character mapping
- Vector outlines

# Fonts































- Character mapping
- Vector outlines
- Glyph positioning (GPOS)

# Fonts

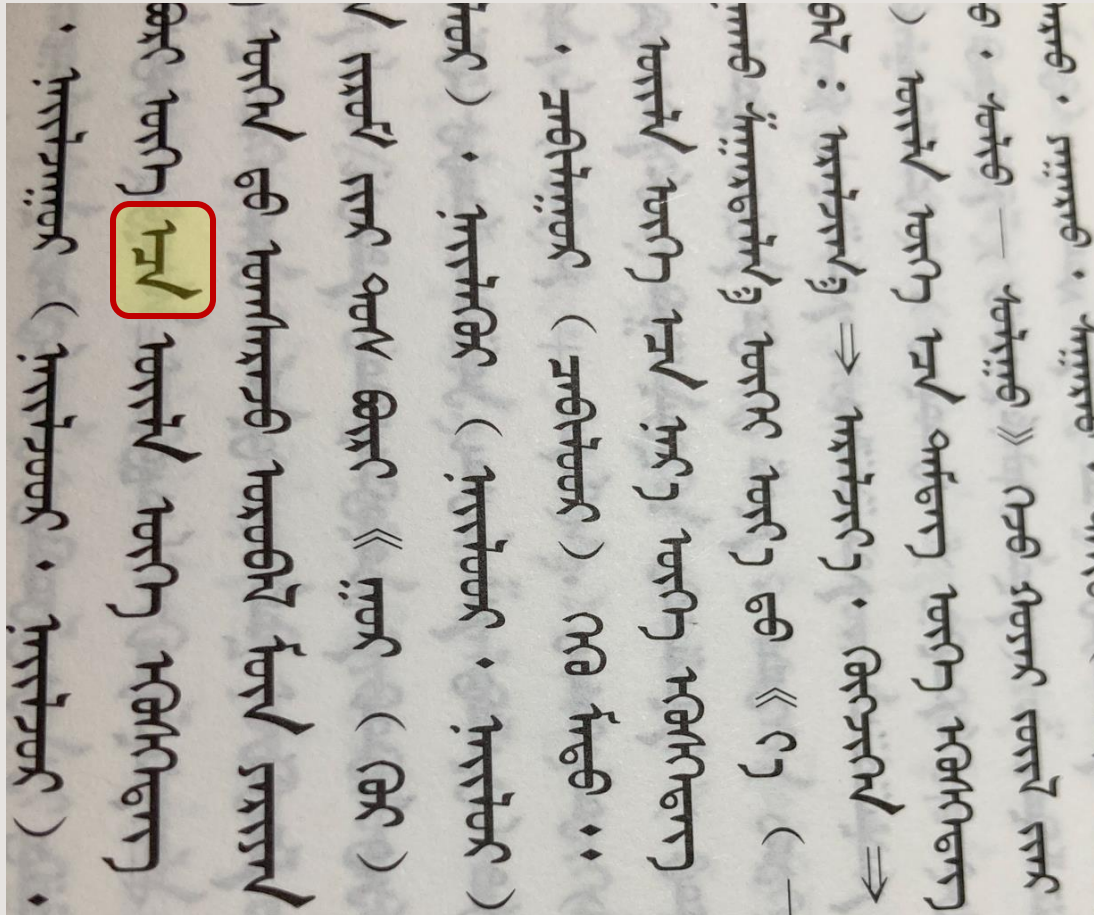


- Character mapping
- Vector outlines
- Glyph positioning (GPOS)

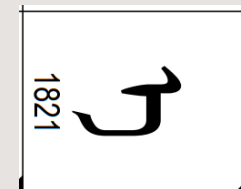
Combining Diacritical Marks

	030	031	032	033	034	035	036
0	 0300	 0310	 0320	 0330	 0340	 0350	 0360
1	 0301	 0311	 0321	 0331	 0341	 0351	 0361
2	 0302	 0312	 0322	 0332	 0342	 0352	 0362
3	 0303	 0313	 0323	 0333	 0343	 0353	 0363

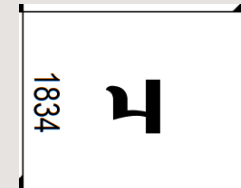
# Fonts



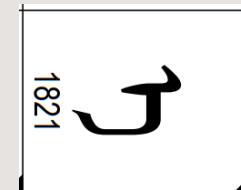
- Character mapping
- Vector outlines
- Glyph positioning (GPOS)
- Glyph substitution (GSUB)



MONGOLIAN LETTER E



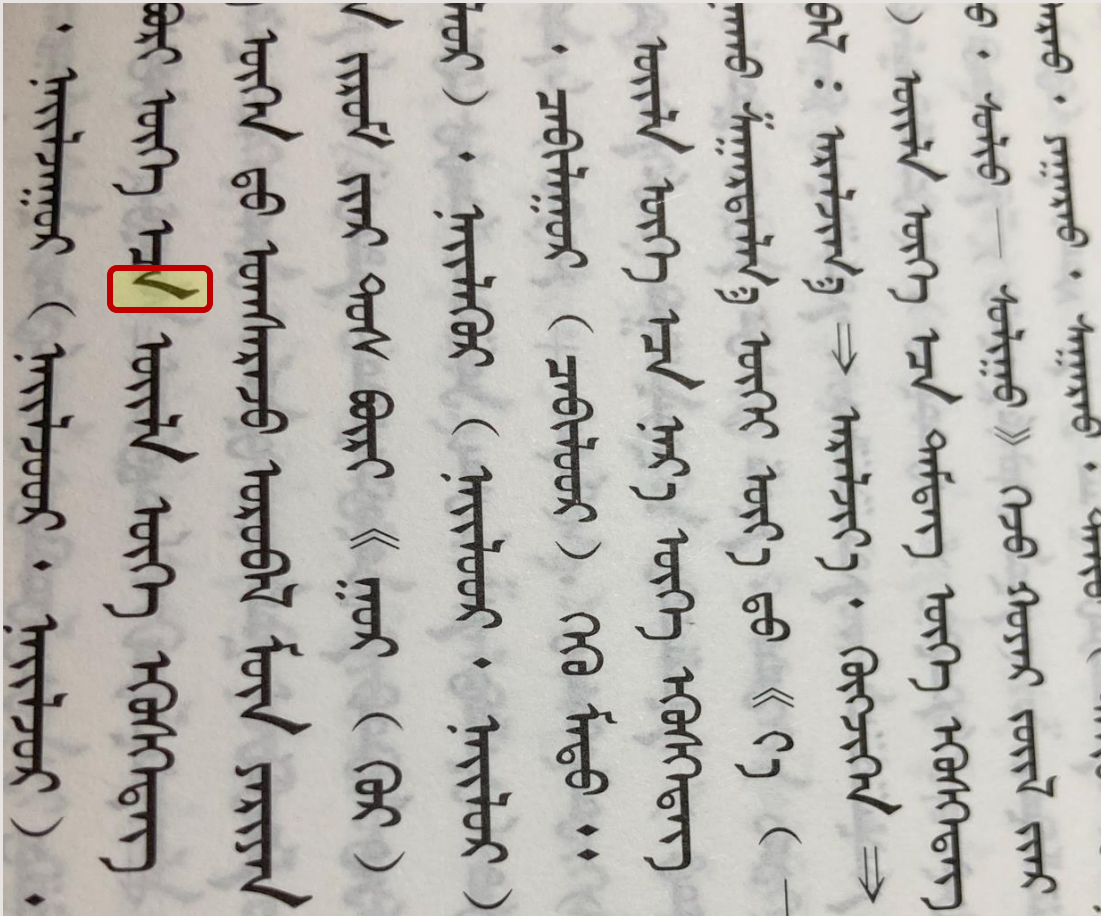
MONGOLIAN LETTER CHA



MONGOLIAN LETTER E

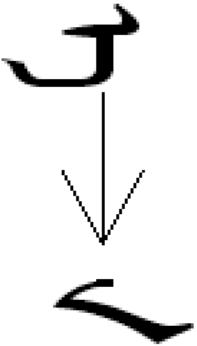


# Fonts



- Character mapping
- Vector outlines
- Glyph positioning (GP)
- Glyph substitution (GS)

1821	MONGOLIAN LETTER
1834	MONGOLIAN LETTER
1821	MONGOLIAN LETTER

	uni1821 -> uni1821.fina
---	-------------------------

# Reordering

a carpet-weaver (El.).  
kulic<sup>ü</sup> 1 कुलिषू (= कलچ) ।  
small disc-shaped cake made  
clarified butter, etc. kulic<sup>ü</sup>  
पूपविशेषो वर्तुलः m., id.  
kulic<sup>ü</sup> 2 कुलिचू, sg. gen. f. of 1  
kālacan कालचन् or kālacēn  
in the evening (K.Pr. 105  
cf. kāla, p. 434a, l. 21. —

Visual order



093F

DEVANAGARI  
VOWEL SIGN I



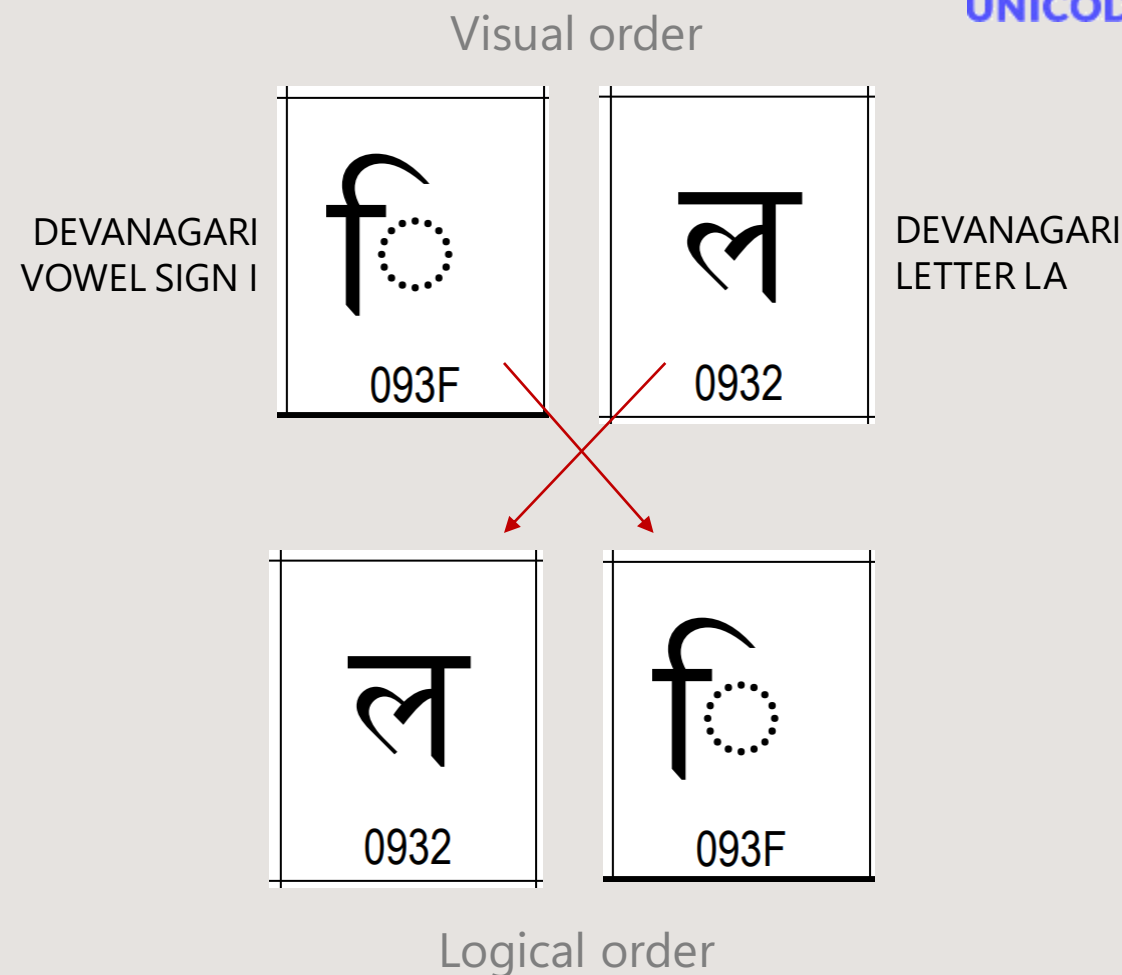
0932

DEVANAGARI  
LETTER LA



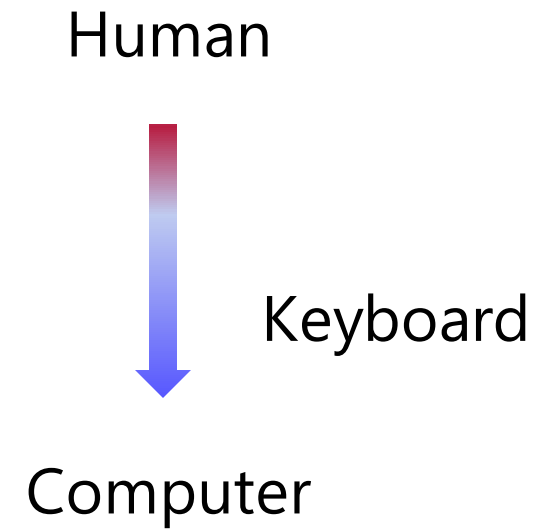
# Reordering

a carpet-weaver (El.).  
kulic<sup>ü</sup> 1 कुलिषू (= कलچ) ।  
small disc-shaped cake made  
clarified butter, etc. kulic<sup>ü</sup>  
पूपविशेषो वर्तुलः m., id.  
kulic<sup>ü</sup> 2 कुलिचू, sg. gen. f. of  
kālacan कालचन् or kālacēn  
in the evening (K.Pr. 105  
cf. kāla, p. 434a, l. 21. —)




# Keyboards

- The job of a keyboard is to get language into a computer



# Keyboards

- "Typewriters"
  - Key -> single codepoint



 -> e 0x0065



# Keyboards

- "Typewriters"

- Key -> single codepoint
- Key -> multiple codepoints

 ->  0xa984 0xa9bc





# Keyboards

- "Typewriters"
  - Key -> single codepoint
  - Key -> multiple codepoints
  - Modifier keys

Shift + E -> E 0x0045



# Keyboards

- "Typewriters"

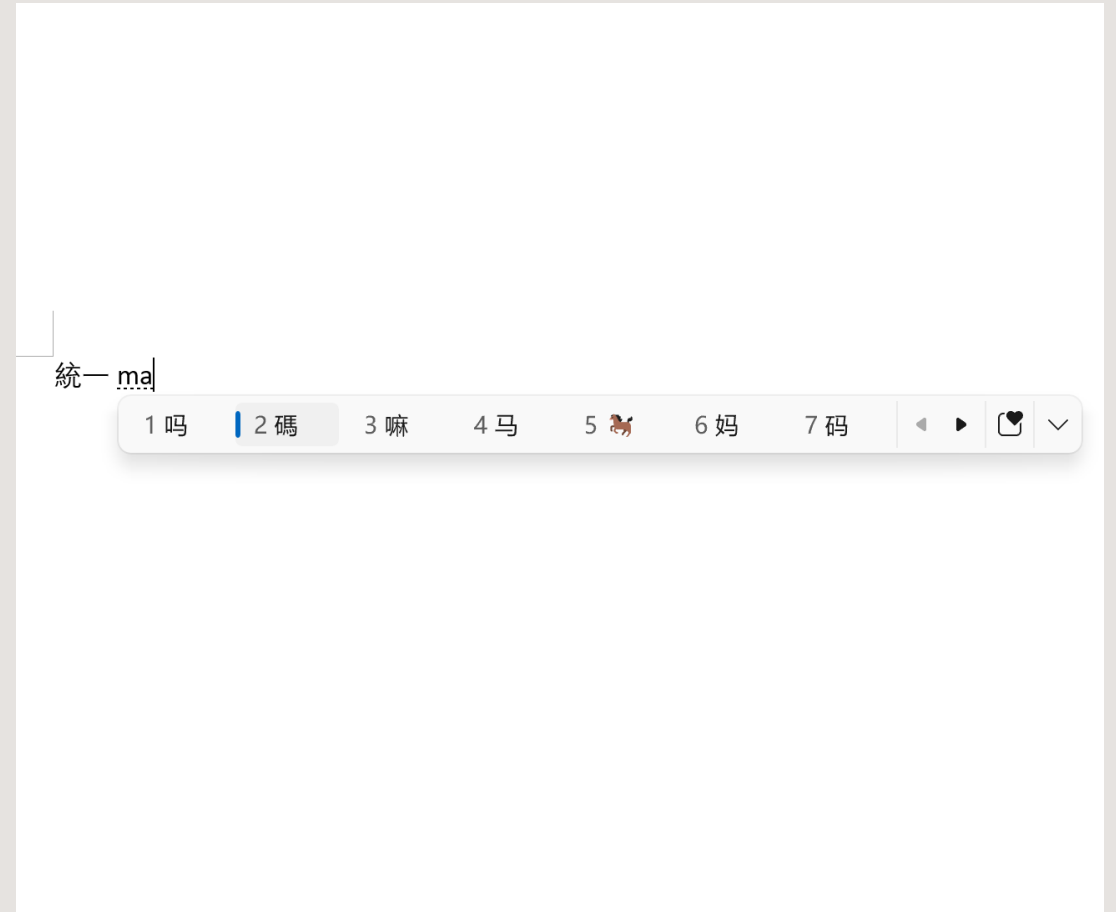
- Key -> single codepoint
- Key -> multiple codepoints
- Modifier keys
- Deadkeys

AltGr + - , E -> ë 0x0113



# Keyboards

- “Typewriters”
- IMEs



# Keyboards

- "Typewriters"
- IMEs
- Complex orthographies

മൊ



഼

200C 0DD9



മെ

~~200C-0DD9~~ 0D9A 0DD9



മൊ

0D9A ~~0DD9~~ 0DDC



# Keyboards

- "Typewriters"
- IMEs
- Complex orthographies

```
<transforms type='contextual'>
  <!-- independent vowels -->
  <transform contextbefore="\u{D85}" input="\u{DCF}" to="\u{D86}"/> <!-- ே
  <transform contextbefore="\u{D85}" input="\u{DD0}" to="\u{D87}"/> <!-- ை
  <transform contextbefore="\u{D85}" input="\u{DD1}" to="\u{D88}"/> <!-- ொ
  <transform contextbefore="\u{D8B}" input="\u{DDF}" to="\u{D8C}"/> <!-- ோ
  <transform contextbefore="\u{D8D}" input="\u{DD8}" to="\u{D8E}"/> <!-- ௌ
  <transform contextbefore="\u{D8F}" input="\u{DDF}" to="\u{D90}"/> <!-- ்
  <transform contextbefore="\u{D91}" input="\u{DCA}" to="\u{D92}"/> <!-- ௎
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D91}" to="\u{D93}"/>
  <transform contextbefore="\u{D94}" input="\u{DDF}" to="\u{D96}"/> <!-- ௐ
  <!-- dependent vowels -->
  <transform contextbefore="\u{DD8}" input="\u{DCF}" to="\u{DF2}"/> <!-- ௑
  <transform contextbefore="\u{DD9}" input="\u{DCA}" to="\u{DDA}"/> <!-- ௒
  <transform contextbefore="\u{DD9}" input="\u{200C}\u{DD9}" to="\u{DDB}"/>
  <transform contextbefore="\u{DD9}" input="\u{DCF}" to="\u{DDC}"/> <!-- ௓
  <transform contextbefore="\u{DDC}" input="\u{DCA}" to="\u{DDD}"/> <!-- ௔
  <transform contextbefore="\u{DD9}" input="\u{DDF}" to="\u{DDE}"/> <!-- ௕
  <!-- consonants EXPAND -->
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D9A}" to="\u{D9A}"/>
  <transform contextbefore="\u{200C}\u{DDB}" input="\u{D9A}" to="\u{D9A}"/>
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D9B}" to="\u{D9B}"/>
  <transform contextbefore="\u{200C}\u{DDB}" input="\u{D9B}" to="\u{D9B}"/>
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D9C}" to="\u{D9C}"/>
  <transform contextbefore="\u{200C}\u{DDB}" input="\u{D9C}" to="\u{D9C}"/>
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D9D}" to="\u{D9D}"/>
  <transform contextbefore="\u{200C}\u{DDB}" input="\u{D9D}" to="\u{D9D}"/>
  <transform contextbefore="\u{200C}\u{DD9}" input="\u{D9E}" to="\u{D9E}"/>
  <transform contextbefore="\u{200C}\u{DDB}" input="\u{D9E}" to="\u{D9E}"/>
```

ஹெ



ஹெ

200C 0DD9



ஹெ

~~200C-0DD9~~ 0D9A 0DD9



ஹெ

0D9A ~~0DD9~~ 0DDC

# Keyboards

- “Typewriters”
- IMEs
- Complex orthographies

h

ḥ

ḥ

ḥ



+	[RALT K_H]	>	ḥ	
ḥ	+	[RALT K_H]	>	ḥ
ḥ	+	[RALT K_H]	>	ḥ
ḥ	+	[RALT K_H]	>	ḥ



# Keyboards



- “Typewriters”
- IMEs
- Complex orthographies
- References

## **CLDR Keyboard subcommittee**

<https://cldr.unicode.org/index/keyboard-workgroup>

## **MSKLC**

<https://www.microsoft.com/en-us/download/details.aspx?id=102134>

## **SIL Keyman**

<https://keyman.com/developer/>

## **IME authoring**

<https://learn.microsoft.com/en-us/windows/apps/design/input/input-method-editors>

## **About Windows keyboards**

<https://kbdlayout.info/>

# Unicode Fonts and Keyboards

Andrew Glass