

The Latin Script LGR

Bill Jouris

The Cast

member origins:

- Serbia (Chair)
- Peru (Sub-chair – Variants)
- Germany (2)
- Sweden
- United States
- Egypt

What We're Looking At

Repertoire

Variants

The Latin Script is a Mess

(That's why it's taken us 7 years)

- 445 living languages
- 210 languages included
- 221 characters (glyphs)

Which Languages Included?

- Official languages (EGIDS 0-4)
 - National
 - State, provincial or regional
- Large populations and EGIDS 5 (substantial literature)
And over 1 million native speakers

Letters

a b c d e f g h i k l m n o p
q r s t v x y z
æ ə ə ε γ ı ı j ɔ œ ß þ u w ʒ

Excluded:

‘ ! | || †

Diacritics

Grave Accent 

Acute Accent 

Circumflex Above 

Tilde 

Macron 

Breve 

Caron 

Dot Above 

Diaeresis 

Double Acute 

Ring Above 

Hook Above 

Horn 

Cedilla 

Ogonek 

Dot Below 

Comma Below 

Circumflex Below 

Macron Below 

Variants

“What you see is what you get” . . . only works if what you see is what you *think* you see.

What you want:

- As a user, go where you expect to go
- As a registrant, having your customers/users reliably come to your site, not go somewhere else

A Small Slice

Any given language only uses a few diacritics.

- English: None!
- Spanish: Tilde, Diaeresis, Acute
- French: Cedilla, Acute, Circumflex, Grave, Diaeresis

Variants:

Is cañon noticeably different from cañon, if you don't know the Macron diacritic?

Is café noticeably different from café, if you don't know the Dot Above diacritic?

30 variations of Letter O alone!

Plus other languages

o ɔ ȯ ò ó ô õ ö
ø ō Ȫ ǒ Ǔ ɔ ǿ ɔ̇
ȯ ò̇ ó̇ ỏ̇ ố̇ ồ̇ ỗ̇ ỗ̇̂
ô̇ ớ̇ ờ̇ ỡ̇ ỡ̇̂ ợ̇ ờ̇̂

Greek: o ó Cyrillic: o ö ɵ

Armenian: օ

Hebrew: o Malayalam: O

Myanmar: o Oriya: O

Variants vs Confusables vs Different

What are they?

The “reasonably careful user”

.com

.COM

ALL CAPS

.coṃ

M with Dot below

.cõm

O with Horn

.çom

C with Cedilla

.cõ̃m

O with Circumflex and Tilde

.corn

C O R N

.COM

Cyrillic

Variants Process

Unicode Name		Unicode Code Point	Glyph	Glyph	Unicode Code Point	Unicode Name	Inspectors						Panel Inspection	
							MT	DT	BJ	HH	MS	ND	MM	
							1=Identical 2=Nearly Identical (*) 3=Distinguishable 4=Different 5=NA.						Confirmed Review	
Small Letter Thom	00FE	þ	þ	01A5	Small Letter P with Hook	3	3	3	3	3	4	3	Not a variant	
		þ	þ			3	3	2	2	3	3	2		
		þ	þ			3	3	4	3	3	4	2		

þ þ Times New Roman

þ þ Arial

þ þ Courier New

How Much is Enough?

00E8	Latin Small Letter E With Grave	è	ě	Latin Small Letter E With Hook Above	1EBB
		è	ě		
		è	ě		

A majority isn't

Too Much Information

Latin Small Letter T	t ↔ ƚ	Latin Small Letter L with Stroke
Latin Small Letter J	j ↔ ĵ	Latin Small Letter I with Ogonek
Latin Small Letter Eth	ð ↔ đ	Latin Small Letter D with Stroke
Latin Small Letter C with Acute	ć ↔ ċ	Latin Small Letter C with Hook
Latin Small Letter F	f ↔ ƒ	Latin Small Letter T with Hook

In-Script Variants

Schwa	ə	ə	Turned E
Iota	ɪ	ɪ	Dotless I
D with Caron	ď	đ	D with Hook
A with Breve	ă	ǎ	A with Caron
O with Diaeresis	ö	õ	O with Double Acute
Ligature AE	æ (<i>æ</i>)	œ	Ligature OE
K	k	ƀ	K with Horn

Cross-Script Variants

Related languages

- Cyrillic – Latin Variants 29 including:
 - Er p p P
 - Es with descender ç ç C with cedilla
- Greek – Latin Variants 18 including:
 - Nu v v V
 - Beta ß ß Sharp S
- Armenian – Latin Variants 7 including:
 - Seh g g G
 - Yiwn l l lota

Cross-Script Variants

Generic Symbols

l	o	c	o	Latin
l	o	c		Cyrillic
l	o			Hebrew
	o			Greek
	o			Armenian
o1	o	c	o	Myanmar
		c	o	Lao
o	o			Oriya

Oddities

Dot Above vs Acute (and Grave)

- Consonants

ć ↔ ċ ñ ↔ ñ̇ ź ↔ ż

- Vowel (Due only to Transitivity with Greek)

í ↔ i

- Vowels NOT Variants

é ↔ è è ↔ è̇ ì ↔ i

No Underlining

www.example.test

No diacritics

www.example.test

L with Dot below

www.example.test

E with Macron below

www.example.test

S with Comma below

www.example.test

L with Circumflex below

www.example.test

M with Cedilla

With Underlining

www.example.test

No diacritics

www.example.test

L with Dot below

www.example.test

E with Macron below

www.example.test

S with Comma below

www.example.test

L with Circumflex below

www.example.test

M with Cedilla